eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

The acceptability, effectiveness, and durability of cognitive analytic therapy: Systematic review and meta-analysis

Craig Hallam[1], Melanie Simmonds-Buckley[2]* (iD),
Stephen Kellett[2,3] (iD), Beth Greenhill[1,4] (iD) and Andrew Jones[1]

[1]University of Liverpool, UK
[2]University of Sheffield, UK
[3]Sheffield Health and Social Care NHS Foundation Trust, UK
[4]North West Boroughs NHS Foundation Trust, UK

Objectives. This paper sought to conduct a meta-analysis of the effectiveness and durability of cognitive analytic therapy (CAT) and assess the acceptability of CAT in terms of dropout rates.

Design. Systematic review and meta-analysis.

Methods. PROSPERO registration: CRD42018086009. Searches identified CAT treatment outcome studies eligible to be narratively synthesized. Pre–post/post-follow-up effect sizes (ESs) were extracted and synthesized in a random-effects meta-analysis. Variations in effect sizes were explored using moderator analyses. Dropout rates were extracted. Secondary analyses synthesized between-group ES from trials of CAT.

Results. Twenty-five studies providing pre–post CAT treatment outcomes were aggregated across three outcome comparisons of functioning, depression, and interpersonal problems. CAT produced large pre–post improvements in global functioning (ES = 0.86; 95% CI 0.71–1.01, N = 628), moderate-to-large improvements in interpersonal problems (ES = 0.74, 95% CI 0.51–0.97, N = 460), and large reductions in depression symptoms (ES = 1.05, 95% CI 0.80–1.29, N = 586). All these effects were maintained or improved upon at follow-up. Limited moderators of CAT treatment effect were identified. CAT demonstrated small–moderate, significant post-treatment benefits compared to comparators in nine clinical trials (ES = 0.36–0.53; N = 352). The average dropout rate for CAT was 16% (range 0–33%).

Conclusions. Patients with a range of presenting problems appear to experience durable improvements in their difficulties after undergoing CAT. Recommendations are provided to guide the further progression of the CAT outcome evidence base.

*Correspondence should be addressed to Melanie Simmonds-Buckley, Cathedral Court, University of Sheffield, Sheffield S1 2LT, UK (email: m.simmonds-buckley@sheffield.ac.uk).

Practitioner points

- Large pre–post reductions in global functioning and depression outcomes and moderate–large reductions in interpersonal problems are evident after CAT.
- The effects of CAT appear durable, and interpersonal functioning significantly improves over follow-up time.
- CAT produces small–moderate benefits compared to trial comparators.
- CAT appears to be an engaging psychotherapy that maintains patients in treatment.

Access to and the provision of psychological therapies in public services is underpinned by the philosophy of 'evidence-based' practice (EBP; McHugh & Barlow, 2012) to enable patients to be treated with empirically supported psychotherapies (Tolin, McKay, Forman, Klonsky, & Thombs, 2015). The 'hourglass model' (Salkovskis, 1995) defines the three-stage cyclical process enabling EBP: (1) completion of initial practice-based evidence (PBE) studies, (2) generation of methodologically robust clinical efficacy trials, and (3) implementation of evidenced-based psychotherapies in services, with associated audit and evaluation. Barkham, Stiles, Lambert, and Mellor-Clark (2010) advocated equipoise in relation to EBP and PBE, to encourage a balanced critique of the evidence base for any psychotherapy. Public service commissioning demands good evidence when deciding on funding, and so, cognitive analytic therapy (CAT) is required to demonstrate that it is an acceptable, safe, and effective intervention (Marriott & Kellett, 2009). A discrepancy between the large amount of PBE versus small amount EBP evidence has been previously highlighted for CAT (Calvert & Kellett, 2014). This places CAT in a vulnerable position of being widely practised, but currently with a lack of confidence and clarity around the acceptability, efficacy, and effectiveness of the approach (Margison *et al.*, 2000).

   In response, Ryle, Kellett, Hepple and Calvert (2014) completed an initial quantitative synthesis of CAT evidence via a meta-analysis of eleven pre–post study effect sizes (ESs) and demonstrated moderate-to-large effects. There were, however, five critical limitations that the current review sought to address. First, it was unclear why and which measures were used, thus limiting the conclusions able to be drawn about specific effects (Kızılırmak, Özdemir, & Öngen, 2015). Analysing the impact of CAT on specific outcomes would provide insight into the types of change CAT can produce. Second, aggregated follow-up comparisons nor moderators of CAT treatment effect were evaluated. Therefore, understanding the durability of CAT and also the features of the model which contribute to more beneficial effects was indicated. Third, no within-group correlations were performed when calculating the pre–post ES – hence, completing correlations would improve the reliability of the ES produced. Fourth, no assessment of study quality, nor any quantitative analysis on the impact study quality on CAT treatment effects was completed. As study quality can inflate ES estimates (Cuijpers, Donker, Weissman, Ravitz, & Cristea, 2016), this study also sought to quality assess the studies and complete a moderator analysis of the role of study quality. Fifth, the use of pre–post ES without control group comparisons (e.g., passive and active controls in clinical trials) meant it was not possible to specifically attribute change to CAT, because treatment effects could not be separated from effects of time. The number of randomised controlled trials (RCTs) of CAT conducted since the Ryle *et al.* (2014) review, although not sufficient for a comprehensive disorder-specific synthesis, did offer the opportunity to conduct a preliminary evaluation of CAT treatment effects relative to control groups in the RCT evidence base.

   To summarize, the main aims of the present review were to provide a contemporary quantitative synthesis of the CAT evidence base by producing a well-

conducted meta-analysis of pre–post and follow-up CAT outcomes, moderators of ES observed, and produce an initial evaluation of the efficacy of CAT. The study also sought to differentiate the effect of CAT on different clinical outcomes: global functioning, interpersonal problems, and depression. Due to limitations of uncontrolled pre–post ES, conclusions about the effectiveness of CAT throughout this review in the PBE studies should be considered in the context that patients undergoing CAT have experienced improvements in their difficulties, but the improvements cannot be specifically attributed to CAT. As such, a secondary aim was to provide an initial estimate of between-groups treatment effects on overall outcomes when CAT is compared to any comparator from the available controlled trials. Finally, CAT has been presented previously as an engaging psychotherapy that creates low dropout rates across diagnoses (Calvert & Kellett, 2014), and so, this study sought to assess the average dropout rate for CAT.

Method

The review was preregistered on PROSPERO (CRD42018086009) and is reported according to PRISMA guidelines (Moher *et al*., 2015).

Study selection

Electronic searches of four databases (PsycINFO, MEDLINE, CINAHL, and Web of Science) were conducted using the search term *cognitive analytic* (allowing for wildcard variations modified for each database), replicating the previous CAT review (Calvert & Kellett, 2014). No search limiters were applied, with the searches identifying studies of CAT treatment outcome published up until 2019 (final search was performed on 23 August 2019). Reference lists of identified studies and the list of published CAT studies on the Association of Cognitive Analytic Therapy (ACAT) website were cross-referenced to identify any additional studies not identified by the database search. Finally, key authors were emailed to enquire about further studies of CAT and authors' publication lists were checked. After duplicates were removed, the primary reviewer screened titles and abstracts and reviewed the identified full texts against the inclusion and exclusion criteria. An independent reviewer (trainee clinical psychologist) screened 25% of the full-text articles and checked 25% of the included studies to ensure they met eligibility criteria, with discrepancies resolved through discussion (rater agreement = 81%). Overall, consensus on study eligibility queries was achieved via discussion amongst all members of the review team.

Eligibility criteria

Articles were eligible for inclusion if they were a treatment outcome study that reported pre- and post-treatment scores on a validated outcome measure (i.e., means and standard deviations [SDs]) for adults (mean sample age 16+) treated with CAT for any psychiatric diagnosis or problem related to psychological factors. Studies were included if they used either a RCT, a non-randomised controlled trial, an uncontrolled (pre–post) design, or a case-series design. Both group and individual delivery of CAT were eligible, with no restriction on treatment setting. Studies that combined CAT with another treatment modality (e.g., CAT combined with CBT) were excluded, as were the CAT consultancy

studies and all the single-case experimental design studies. Included studies were required to be available in English language and be published or have undertaken a peer-review process to ensure a minimum standard of study quality and replicable search procedures. Unpublished dissertations and conference papers were excluded. For inclusion in the primary quantitative synthesis, pre–post treatment outcomes were required to have been assessed using a validated measure of global functioning, interpersonal difficulties, and/or depression. A measure was considered validated if adequate psychometric properties had been demonstrated in a published study. Studies were excluded if they did not provide sufficient information to calculate ES (i.e., means and *SD*). For inclusion in the secondary between-groups synthesis, studies were required to have used a RCT design and assessed post-treatment between-groups outcomes on one of the specified outcome categories or any other psychiatric outcome measure for CAT versus any comparator condition.

Outcomes

Primary analyses

The three outcomes of interest were self-report measures of *global functioning* (e.g., Symptom Checklist-Revised-90 [SCL-R-90], Clinical Outcomes in Routine Evaluation-Outcome Measure [CORE-OM], Brief Symptom Inventory [BSI]), *interpersonal difficulties* (e.g., Inventory of Interpersonal Difficulties [IIP], Persons Related to Others Questionnaire-2 [PROQ-2]), and *depression* (e.g., Beck Depression Inventory [BDI/BDI-II], Hospital Anxiety and Depression Scale [HADS-D], Patient Health Questionnaire-9 [PHQ-9]). Originally, additional outcomes relating to anxiety measures and remission rates were included in the protocol; however, insufficient eligible studies assessed these outcomes, so the analyses were not performed. To ensure each study contributed only one ES per outcome, the most frequently used measure was selected when multiple measures of one outcome were reported. Dropout rates were also evaluated as a proxy for CAT treatment acceptability. Dropout was determined as percentage of patients who completed treatment according to the definition used in the original study.

Secondary analyses

Due to limited comparability between types of outcomes assessed in the eligible RCTs, between-groups comparisons assessed overall treatment outcome as measured by any of the outcome measures from the primary analysis or a disorder-specific outcome measure (if one of the primary outcomes measures was not assessed) rather than across each outcome separately (as specified in the original protocol). ES from individual studies that assessed more than one of the three primary outcomes was first aggregated in a mini meta-analysis to capture the aggregated treatment outcome and ensure only one ES per study was included.

Quality assessment

The Downs and Black (1998) 27-item quality checklist was used to assess methodological quality as it incorporates randomized and non-randomized study designs. This tool assesses five aspects of methodological quality – reporting (10 items); internal validity – confounding (6 items); internal validity – bias (7 items); external validity (3 items); and power (1 item). Due to reported difficulties in applying the item relating to power, this

item was modified to a yes/no response to indicate whether the study had adequate power (0.8) to detect a pre–post moderate ES (0.6; based on lower bound confidence interval of previous CAT meta-analysed aggregated ES) at α < .05 (O'Connor *et al.*, 2015). Studies were therefore assessed on a scale of 0–28 (higher scores indicated higher quality) and categorized according to the following quality thresholds; poor (<14), fair (14–18), good (19–23), and excellent (24–28). The primary reviewer assessed all included studies, with 30% double-rated by an independent rater (trainee clinical psychologist). Inter-rater agreement was assessed with Cohen's kappa statistic (k), interpreted as .21–40 indicating fair agreement, .41–.60 as moderate agreement, .61–.80 as substantial agreement, and .81–1.0 as almost perfect agreement (Cohen, 1960; Landis & Koch, 1977). There was substantial agreement between the raters (k = 0.77). Any discrepancies between raters were discussed and a consensus reached to produce an agreed quality score for each study.

Data extraction

An a priori extraction tool was designed to code data on the following criteria: (1) *methodological characteristics* (study design, study quality, assessment of therapist's competence using CCAT [yes/no], mean CCAT score, and indication of therapist competence [CCAT> 20 competent CAT/CCAT < 20 incompetent CAT], (2) *intervention characteristics* (number of sessions, format [group/individual], treatment setting, and therapist qualification [CAT qualified/trainee], *patient characteristics* (age, gender [% male], and specified difficulties), and *outcomes* (pre- and post-treatment means, SDs, and dropout rates). Data on whether studies assessed adverse effects (hospital admissions, reliable deterioration rates) during CAT were also extracted. Where data were not reported, the information was requested from authors by email.

Effect sizes

Pre–post effect sizes

For the primary analyses, pre–post ES was calculated for global functioning, interpersonal difficulties, and depression outcomes. ES was calculated by subtracting the mean post-treatment score from the mean pre-treatment score and dividing by the pre-treatment standard deviation (*SD*). To account for the violation of independence, calculations for pre–post ES require an estimate of the correlation between pre- and post-scores (r). Where pre–post correlations were reported, the actual value of r was used to calculate individual study ES. In the absence of a reported pre–post correlation, an imputed value of 0.6 was used based on the median within-group correlation extracted from 811 measures of pre–post clinical trial arms (Balk *et al.*, 2012) and sensitivity analyses were conducted to evaluate the effect of different imputed r values (see moderator and sensitivity analyses section). ES was converted to Hedges g using the *J*-correction to account for small study sample biases (Hedges & Olkin, 1985). If a study had used a RCT or controlled trial design, pre–post ES was calculated for the CAT treatment arm only. Positive ES indicated symptom reductions after treatment (negative ES indicated symptom deterioration) and was interpreted as 0.2 indicating a small effect, 0.5 indicating a moderate effect, and 0.8 indicating a large effect size (Cohen, 1992).

Between-groups effect sizes

For the secondary analyses, pre-/post-control group effect sizes were calculated for included studies that had used a RCT design to calculate the effect of CAT versus comparator groups. ES was calculated by subtracting the mean pre–post change of the comparison group from the mean pre–post change in the CAT group and dividing by the pooled pre-treatment *SD* to account for pre-treatment group differences (Morris, 2008). Where there was a follow-up assessment, post-treatment to follow-up control group ES was calculated using the same procedure. All ESs were adjusted using the small study J-correction (Hedges & Olkin, 1985). Positive ES indicated a treatment effect in favour of CAT.

Data synthesis

Effect sizes were synthesized via a random-effects meta-analysis with inverse variance weighting (based on the DerSimonian & Laird, 1986 estimator) using Meta-Essentials workbooks (Suurmond, van Rhee, & Hak, 2017). Overall, pooled treatment estimates, 95% confidence intervals (CI), and prediction intervals (PI) were produced using the inverse of the variance to weight effect estimates. Between-study heterogeneity was assessed using the I^2 statistic to indicate the percentage of variation and the accompanying Q statistic to assess significance. Study heterogeneity was grouped into low (25%), moderate (50%), and high (75%) thresholds (Higgins, Thompson, Deeks, & Altman, 2003).

Moderator and sensitivity analyses

In the primary analyses, the anticipated between-study heterogeneity was explored using pre-specified subgroup and meta-regression analyses. Subgroup analysis was used to explore four categorical variables: study type [PBE/RCT], therapist experience [CAT qualified/not qualified], format [1:1/group], and diagnosis [personality disorder/other]. Due to the range of presentations treated, diagnosis classifications were collapsed into two groups: (1) cases (or over half of included cases) presenting with a personality disorder and (2) all other diagnoses. Meta-regression was used to explore four continuous variables: study quality, gender [% male], mean age, and treatment duration [number of sessions]. To account for multiple testing, the alpha threshold for significance was adjusted to $p < 0.0125$ ($\alpha = .05/4$) for between-subgroup differences and meta-regression beta-coefficients, respectively. A minimum of 10 studies were required to perform moderator analyses (Cochrane Collaboration, 2011).

Due to typically low rates of reporting for pre–post *r* values, sensitivity analyses were employed to investigate the effect of imputing different values for *r* in ES calculations on the overall estimated CAT treatment effect (Borenstein, 2009). Four separate overall CAT treatment effects were aggregated using study ES calculated with the value of *r* imputed as 0.0, 0.25, 0.5, and 0.75, respectively (Balk *et al.*, 2012). Where possible, ES for studies that reported a true value for *r* was also aggregated and compared to the aggregated imputed ES (using *r* = 0.6 from the main analyses) for the same studies to assess the potential bias that might be present in the findings in the absence of knowing the true *r* values.

Publication bias

Publication bias was assessed via three approaches: (1) Funnel plots of study ES plotted against standard errors (SE) were visually inspected for asymmetry (indication of reporting

biases in the included studies); (2) Egger's regression was used to statistically test for the presence of publication bias (Egger, Smith, Schneider, & Minder, 1997); and (3) 'Trim and Fill' imputation was conducted to estimate treatment effects adjusted for publication bias (Duval & Tweedie, 2000).

Results

Study selection

The search strategy produced a combined total of 28 CAT outcome articles (Figure 1). After the removal of duplicates, a total of 576 articles were screened with 100 identified for full-text review. Thirty articles were initially deemed eligible for inclusion, although two studies were subsequently excluded, as they were reports on the same data of another included study. This left $k = 28$ articles for qualitative synthesis, with $k = 25$ studies eligible for the primary pre–post quantitative synthesis and $k = 9$ RCTs for the secondary between-groups quantitative synthesis.

Study characteristics

Characteristics of the included studies are presented in Table 1. Of the 28 included studies, $k = 10$ (36%) were RCTs and the remaining $k = 18$ (64%) were PBE studies (pre–post design $k = 10$; case series $k = 8$). Mean study quality was 12.68 ($SD = 4.50$; maximum score 28), ranging between 5 and 21 (see supplementary materials for full study quality ratings). Overall study quality was poor, with only 11 studies classified as fair and three as good. The RCTs of CAT generated higher ratings of methodological quality (mean = 16.6, $SD = 3.75$), than the PBE studies (mean = 10.22, $SD = 3.44$). Study quality by sub-domain highlighted that studies scored highest in terms of methodological reporting (with the exception of reporting adverse events). Study quality was lowest for internal validity related to confounding in selection biases, particularly in terms of lack of randomization and lack of adequate adjustment for confounding in the analyses. In terms of power, less than half ($k = 12$) of the studies were sufficiently powered to detect a moderate pre–post effect.

All studies bar one were conducted in public health services. CAT was typically delivered individually ($k = 26$; 93%) or occasionally in groups ($k = 2$; 7%). Treatment duration ranged from 5 to 30 sessions. CAT was employed in the standard eight-session ($k = 2$; 7%), 16-session ($k = 10$; 35%), or 24-session ($k = 8$; 28%) versions, with $k = 5$ (18%) studies using a combination of treatment versions. A six-session ($k = 1$), 12-session ($k = 1$), and 20-session ($k = 1$) versions were used in the remaining studies ($k = 1$ study did not report treatment duration; total studies exceed number of included studies as one study included separate 16- and 24-session CAT groups). In $k = 13$ studies, therapists were qualified with the remaining $k = 15$ studies either comprising of trainee CAT therapists or therapists without accredited as a CAT practitioner. Just $k = 5$ studies (18%) reported competency ratings, with all reporting mean Competence in Cognitive Analytic Therapy (CCAT) scores exceeding the competence cut-off (>20; Bennett & Parry, 2004).

The age of patients ranged between 16 and 49 years. Presenting problems include mixed diagnoses ($k = 8$), personality disorders (PD; $k = 5$), complex trauma ($k = 3$), anxiety/panic disorder ($k = 2$), anorexia nervosa ($k = 2$), depression ($k = 1$), obsessive–compulsive disorder (OCD; $k = 1$), bipolar disorder ($k = 1$), psychosis ($k = 1$), morbid jealousy ($k = 1$), serious mental disorders ($k = 1$), chronic pain ($k = 1$), and poor diabetes
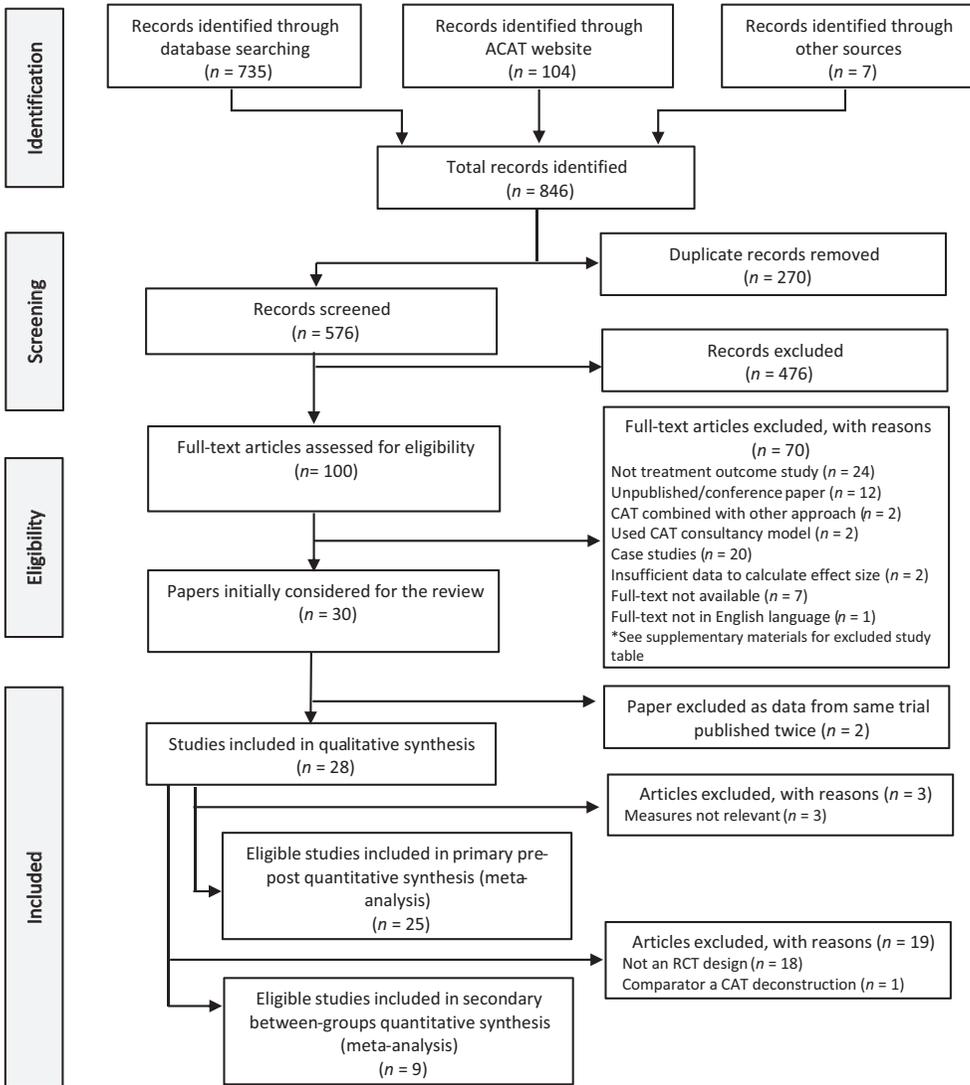
Figure 1. PRISMA flow chart of study selection.

management ($k = 1$). Twenty studies reported follow-up data, with follow-ups ranging between 1 and 24 months (mean = 7.5 months). Dropout rates ranged between 0 and 39% (mean = 15%). Adverse events were reported in seven studies ($k = 2$ studies reported incidents of hospital admissions; $k = 5$ had no serious adverse events). Rates of reliable deterioration after CAT were available in $k = 16$ studies; $k = 11$ studies reported no reliable deterioration, while $k = 5$ studies reported incidences of deterioration ranging between 6 and 20% (mean = 11.4%).

Primary meta-analysis of pre–post CAT treatment effects

Meta-analytic comparisons were conducted on pre–post treatment and from post-treatment to follow-up global functioning, interpersonal problems, and depression

Table 1. Characteristics of included studies

Study First Author (Year)	Study type (design)	Treatment setting (presenting problem)	N	Mean age (SD) [range]	Sex (% male)	Therapist CAT qualified (mean CCAT score)	Treatment duration (format)	Measures	Follow-up duration (months)	Dropout rate%	Adverse events (% RD)	Quality score
[1] Brockman et al. (1987)[b]	RCT	Hospital clinic (mixed)	30	NR	NR	No (N/A)	12 (1:1)	GHQ BDI	8–22[a]	12%	NR (NR)	7 (Poor)
[2] Clarke and Llewelyn (1994)	PBE (case series)	NHS clinic (complex trauma)	7	27 (10.05) [19–48]	0%	Yes (N/A)	16 (1:1)	SCL-90-R BDI	3	14%	NR (0%)	7 (Poor)
[3] Evans and Parry (1996)	PBE (case series)	Not specified (mixed)	4	32 (8.12) [24–42]	0%	Yes (N/A)	13–20 (1:1)	SCL-90-R	None	0%	NR (0%)	5 (Poor)
[4] Dunn et al. (1997)	PBE (pre–post)	Hospital clinic (mixed)	164	NR	42%	No (N/A)	16 (1:1)	SCL-90-RIIP BDI	None	18%	NR (NR)	11 (Poor)
[5] Fosbury et al. (1997)[b]	RCT	Diabetes clinic (poor diabetes control)	10	30.5 (10.6) [18–55]	30%	Not specified (N/A)	16 (1:1)	IIP	3 & 9	33%	NR (NR)	17 (Fair)
[6] Clarke and Pearson (2000)	PBE (case series)	Outpatient (complex trauma)	4	32.50 (13.96) [22–53]	100%	Yes (N/A)	16 (1:1)	SCL-90-R BDI	None	0%	NR (NR)	6 (Poor)
[7] Ryle and Golynkina (2000)	PBE (pre–post)	Hospital clinic (BPD)	27	34.3 (7.5)	41%	Majority no (N/A)	24 (1:1)	SCL-90-R IIP BDI	18	13%	2 hospital admissions (0%)	11 (Poor)
[8] Wildgoose et al. (2001)	PBE (case series)	NHS clinic (BPD)	5	39.40 (7.13) [30–47]	40%	Yes (N/A)	16 (1:1)	SCL-90-R IIP	9	0%	NR (20%)	6 (Poor)
[9] Birtchnell et al. (2004)	PBE (pre–post)	Psychotherapy service (mixed)	31	39.0 (10.1) [22–61]	24%	No (N/A)	16 (1:1)	CORE PROQ-2	3	22.5%	NR (NR)	9 (Poor)
[10] Chanen et al. (2008)[b]	RCT	Youth clinic (BPD)	41	16.3 (0.8)	17%	Yes (N/A)	24 (1:1)	SOFAS	12 & 24	29.3%	No adverse effects (NR)	16 (Fair)
[11] Marriott and Kellett (2009)	PBE (pre–post)	Psychotherapy service (mixed)	ST: 38 MT: 27	NR	ST:42% MT:26%	Yes (N/A)	ST − 16 MT −24 (1:1)	BSI IIP BDI-II	None	NR	NR (NR)	12 (Poor)
[12] Tzouramanis et al. (2010)	PBE (pre–post)	Community mental health centre (panic disorder)	128	33.4 (8.9)	21%	NR (N/A)	NR (1:1)	BDI	12	16%	NR (NR)	6 (Poor)
[13] Clarke et al. (2013)[b]	RCT	Specialist PD clinic (personality disorder)	38	36 (9.5) [19-59]	28%	Yes (22)	24 (1:1)	CORE IIP	18	20%	No diff in hospital admissions (9-11%)	16 (Fair)
[14] Kellett et al. (2013)	PBE (pre–post)	NHS clinics (BPD)	17	29.99 (8.77)	18%	Yes (34.35)	24 (1:1)	CORE	6	11%	NR (6%)	15 (Fair)

Table 1. (*Continued*)

Study First Author (Year)	Study type (design)	Treatment setting (presenting problem)	N	Mean age (SD) [range]	Sex (% male)	Therapist CAT qualified (mean CCAT score)	Treatment duration (format)	Measures	Follow-up duration (months)	Dropout rate%	Adverse events (% RD)	Quality score
[15] Calvert et al. (2015)	PBE (pre–post)	Psychotherapy service (complex trauma)	157	34.65 (10.67) [18–64]	0%	No (N/A)	24 (group)	BSI IIP HADS-d	None	22%	NR (7%)	14 (Fair)
[16] Darongkamas et al. (2017)	PBE (pre–post)	NHS clinic (mixed)	47	41 [16–69]	35%	No (NA)	8, 16 & 24 [mode = 16] (1:1)	CORE IIP	None	NR	NR (0%)	9 (Poor)
[17] Evans et al. (2017)[b]	RCT	Psychotherapy service (bipolar disorder)	9	48.33 (9.84)	22%	Yes (34.16)	24 (1:1)	CORE BDI-II	6 &12	11%	No adverse events (0%)	18 (Fair)
[18] Meadows and Kellett (2017)	PBE (pre–post)	Primary care service (anxiety)	11	37 (10.7) [24–57]	41%	No (N/A)	6 (1:1)	PHQ-9	1	9%	NR (0%)	13 (Poor)
[19] Kellett et al. (2018)	RCT	Primary care service (depression)	CAT: 52 CAT-NR: 43	43.3 (11.8) 39.3 (11.1)	25% 22%	No (25)	8 (1:1)	PHQ-9	2	19% 14%	No adverse events (0%)	19 (Good)
[20] Taylor et al. (2018)	PBE (case series)	NHS clinics (psychosis)	7	26.71 (6.40) [19–34]	57%	No (29)	5–30 (1:1)	SOFAS	2	14%	No adverse events (14%)	15 (Fair)
[21] Curling et al. (2018)	PBE (case series)	NHS clinic (obsessive morbid jealousy)	3	49.3 (11.72) [36–58]	33%	Yes (N/A)	16–24 (1:1)	BSI IIP BDI-II	4	0%	NR (0%)	14 (Fair)
[22] Tyrer and Masterson (2019)	PBE (case series)	NHS clinic (mixed – all had low mood)	6	41.67 (18.76) [26–73]	17%	Yes (N/A)	16 (1:1)	CORE	3	0%	NR (0%)	8 (Poor)
[23] Baronian and Leggett (2019)	PBE (pre–post)	Pain management service (chronic pain)	53	48.1 (12.52)	25%	Yes (N/A)	8 (1:1)	CORE	None	5.2%	NR (0%)	14 (Fair)
[24] Kellett et al. (2019a)[b]	RCT	High secure hospital (serious mental disorder)	10	38.70 (9.38) [18–65]	100%	No (N/A)	16 (group)	PROQ-2	2	10%	NR	21 (Good)
[25] Kellett et al. (2019b)	PBE (case series)	Private sector CAT clinic (mixed anx/dep or PD)	7	NR	NR	Yes (N/A)	16 or 24 (1:1)	CORE	None	30%	NR (0%)	9 (Poor)

Continued

Table 1. (*Continued*)

Study First Author (Year)	Study type (design)	Treatment setting (presenting problem)	N	Mean age (SD) [range]	Sex (% male)	Therapist CAT qualified (mean CCAT score)	Treatment duration (format)	Measures	Follow-up duration (months)	Dropout rate%	Adverse events (% RD)	Quality score
Studies included in secondary quantitative between-groups synthesis only												
[26] Treasure et al. (1995)[b]	RCT	Outpatient eating disorder clinic (anorexia nervosa)	14	NR	NR	NR (N/A)	20 (1:1)	Morgan & Russell scale[c]	12	29%	NR (NR)	19 (Good)
[27] Dare et al. (2001)[b]	RCT	Outpatient eating disorder clinic (anorexia nervosa)	21	27.2 (7.6)	0%	NR (N/A)	24 (1:1)	Morgan & Russell scale[c]	12	38%	2 hospital admissions (NR)	17 (Fair)
[28] Ameri et al. (2017)[b]	RCT	Psychiatric centre (OCD)	12	37.78 (5.38) [31–44]	50%	No (N/A)	16 (1:1)	Y-BOCS[c]	3	25%	NR (NR)	16 (Fair)

Notes. BDI-I/II = Beck Depression Inventory; BPD = borderline personality disorder; BSI = Brief Symptom Inventory; CAT = cognitive analytic therapy; CAT-NR = CAT without narrative reformulation component; CCAT = Competence in Cognitive Analytic Therapy measure; CORE = Clinical Outcomes in Routine Evaluation; GHQ = General Health Questionnaire; HADS-D = Hospital Anxiety and Depression Scale-Depression subscale; IIP = Inventory of Interpersonal Problems; MT = medium-term CAT; N/A = not applicable; N = number of patients; NR = not reported; PBE = practice-based evidence; PD = personality disorder; PHQ-9 = Patient Health Questionnaire-9; PROQ-2 = Person's Relating to Others Questionnaire; RCT = randomized controlled trial; RD = reliable deterioration; SCL-90-R = Symptom Checklist-90-Revised; SOFAS = Social and Occupational Functioning Assessment Scale; ST = short-term CAT.
[a]Insufficient data to calculate follow-up effect sizes.; [b]RCT studies included in secondary between-groups quantitative synthesis.; [c]Measure not relevant to outcome classification used in primary pre–post quantitative synthesis.

| Study name	ES	Weight
(a) GLOBAL FUNCTIONING (pre-post)		
Taylor (2018)	0.23	3.38%
Clarke (2000)	0.33	6.05%
Calvert (2015)	0.37	7.21%
Evans (2017)	0.45	3.99%
Clarke (2013)	0.70	5.95%
Ryle (2000)	0.70	5.56%
Wildgoose (2001)	0.76	2.37%
Birtchnell (2004)	0.82	5.65%
Kellett (2013)	0.84	4.81%
Evans (1996)	0.86	4.07%
Dunn (1997)	0.87	7.22%
Chanen (2008)	0.92	5.74%
Clarke (1994)	1.03	3.54%
Brockman (1987)	1.06	5.33%
Marriott (2009) ST-CAT	1.08	5.69%
Kellett (2019b)	1.11	1.65%
Tyrer (2018)	1.12	2.37%
Marriott (2009) MT-CAT	1.15	5.03%
Darongkamas (2017)	1.21	5.87%
Curling (2018)	1.26	2.70%
Baronian (2019)	1.48	5.82%
Total	**0.86**	**100%**
Test for overall effect: $Z = 11.87$ ($p = <.0001$)		
Heterogeneity: $Q = 67.47$; $df = 20$ ($p <.0001$); $I^2 = 70\%$		
(b) INTERPERSONAL FUNCTIONING (pre-post)		
Kellett (2019a)	0.12	6.25%
Calvert (2015)	0.39	13.01%
Fosbury (1997)	0.44	5.88%
Dunn (1997)	0.71	13.60%
Birtchnell (2004)	0.72	9.46%
Clarke (2013)	0.72	9.69%
Marriott (2009) ST-CAT	0.77	9.77%
Darongkamas (2017)	0.82	9.75%
Curling (2018)	1.00	6.53%
Marriott (2009) MT-CAT	1.16	7.55%
Ryle (2000)	1.22	7.38%
Wildgoose (2001)	1.67	1.14%
Total	**0.74**	**100%**
Test for overall effect: $Z = 8.08$ ($p = <.0001$)		
Heterogeneity: $Q = 30.59$; $df = 11$ (($p = .001$); $I^2 = 64\%$		
(c) DEPRESSION SYMPTOMS (pre-post)		
Calvert (2015)	0.49	10.31%
Brockman (1987)	0.61	8.83%
Evans (2017)	0.65	6.28%
Meadows (2017)	0.69	3.15%
Ryle (2000)	0.85	8.29%
Tzourmanis (2010)	0.85	10.14%
Dunn (1997)	0.89	10.34%
Clarke (2000)	0.99	4.11%
Clarke (1994)	1.32	4.88%
Curling (2018)	1.44	1.75%
Kellett (2018) CAT-NR	1.46	8.40%
Kellett (2018) Full-CAT	1.48	8.74%
Marriott (2009) ST-CAT	1.63	7.90%
Marriott (2009) MT-CAT	1.75	6.87%
Total	**1.05**	**100%**
Test for overall effect: $Z = 9.17$ ($p = <.0001$)		
Heterogeneity: $Q = 64.56$; $df = 13$ ($p < .0001$); $I^2 = 80\%$		
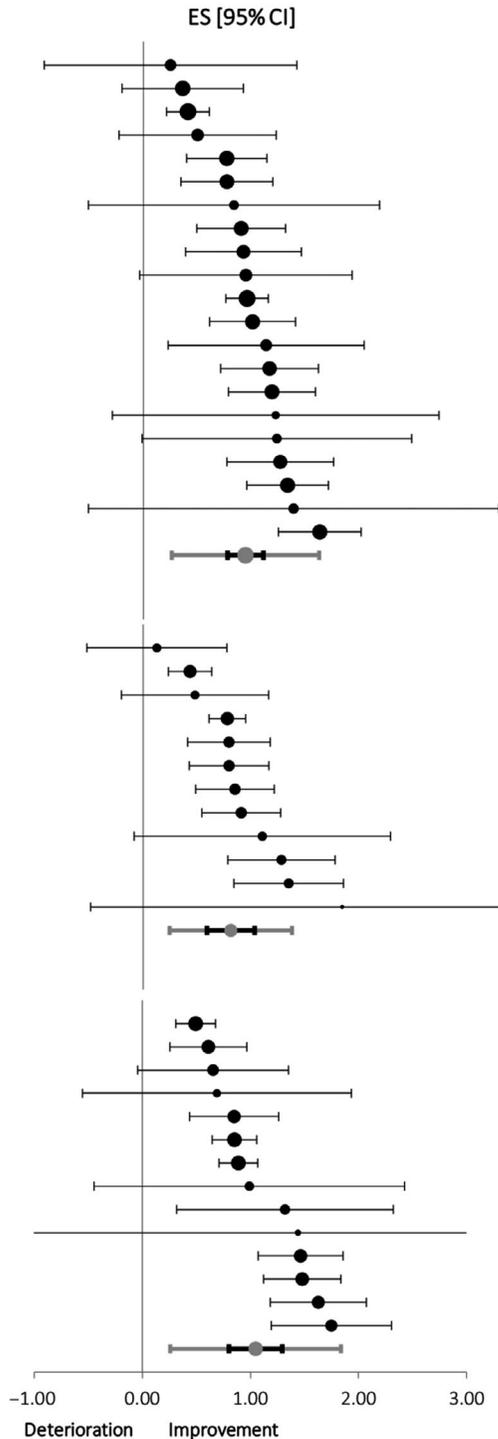
Figure 2. Forest plots of pre–post CAT treatment effect sizes (ES) for (a) global functioning, (b) interpersonal functioning, and (c) depression outcomes [grey error bars represent prediction intervals for overall aggregate CAT treatment estimates].
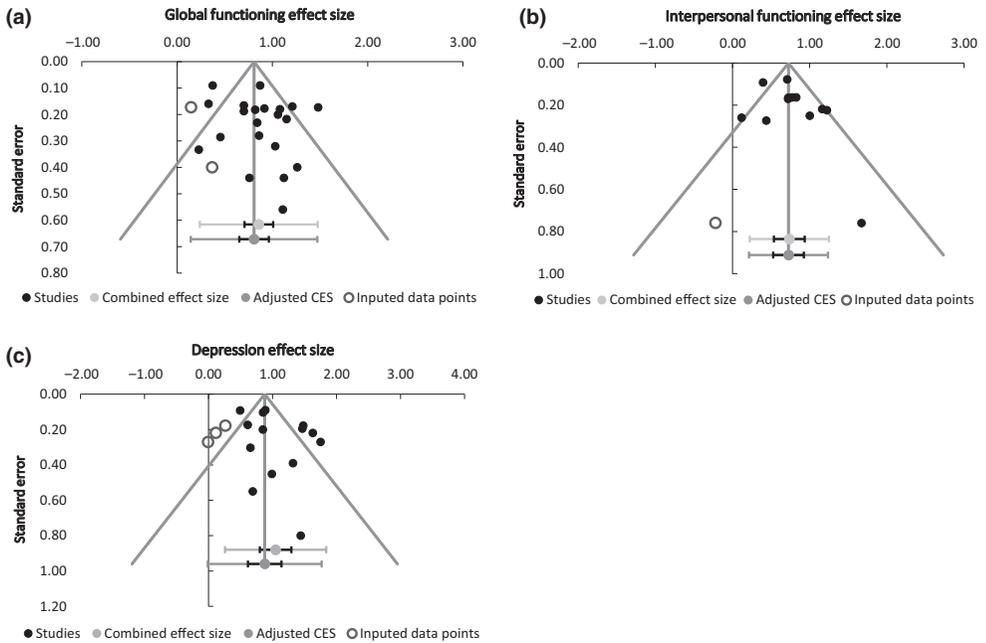
Figure 3. Funnel plots for distribution of studies reporting pre–post CAT treatment outcomes for (a) global symptoms, (b) interpersonal functioning, and (c) depression symptoms [black and grey error bars represent confidence and prediction intervals, respectively].

outcomes. Three studies [26–28 in Table 1] were not eligible for inclusion due to outcomes being assessed with measures that were not relevant to the specific outcome classifications (i.e., outcomes were assessed with disorder-specific measures; anorexia nervosa ($k = 2$) and OCD ($k = 1$) presentations). The remaining $k = 25$ studies evaluated CAT based on outcomes of global functioning, interpersonal problems, and/or depression symptoms. The Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) was the most commonly used measure of global functioning ($k = 8$), followed by the Symptom Checklist-90-Revised (SCL-90-R; $k = 6$). All but $k = 2$ studies that assessed interpersonal difficulties used the Inventory of Interpersonal Problems (IIP; $k = 9$). The most common depression measure was the Beck Depression Inventory (BDI-I or II; $k = 9$).

Effect of CAT on global functioning

Pre–post and follow-up comparisons. Twenty-one comparisons (extracted from 20 studies) totalling $N = 628$ participants provided pre–post treatment outcomes of global functioning. Figure 2a presents the pooled effect size (ES), showing large, significant improvements in global functioning following CAT (ES = 0.86; 95% CI 0.71–1.01; 95% PI 0.24–1.48; Z = 11.87; $p < .001$). There was significant, moderate between-study heterogeneity ($I^2 = 70\%$; Q = 67.47, $p < .001$). End of treatment to follow-up data were provided in 11 comparisons ($N = 146$ participants). There was no significant change in functioning at follow-up, indicating the changes in global functioning were maintained (ES = 0.12; 95% CI −0.04 to 0.28; 95% PI −0.14 to 0.37; Z = 1.64; $p = .100$). There was

minimal, non-significant between-study heterogeneity detected ($I^2 = 14\%$; Q = 11.67, $p = .308$).

Publication bias.   Visual inspection of the funnel plot in Figure 3a and statistical testing using Egger's regression did not indicate substantial asymmetry in study distribution for reporting of pre–post global functioning outcomes (B = 0.62, t(20) = 0.58, $p = .567$). Trim and fill imputation accounted for two smaller missing studies with a minimal to small CAT treatment effects, producing a slightly reduced overall pooled treatment estimate (ES = 0.81; 95% CI 0.65–0.96) that was still representative of a large effect. Taken together, the analyses suggest minimal impact of reporting bias on the pooled treatment estimate.

Effect of CAT on interpersonal difficulties

Pre–post and follow-up comparisons.   Twelve comparisons (extracted from 11 studies) totalling $N = 492$ participants provided pre–post interpersonal change outcomes. The pooled ES demonstrated moderate-to-large, significant reductions in interpersonal difficulties (Figure 2c; ES = 0.74, 95% CI 0.54–0.94; 95% PI 0.22–1.25; Z = 8.08; $p < .001$). There was significant, moderate between-study heterogeneity ($I^2 = 64\%$; Q = 30.59, $p = .001$). End of treatment to follow-up data were provided in seven comparisons ($N = 94$ participants). There was a small, significant effect in favour of further reductions in interpersonal difficulties over the follow-up period (ES = 0.38; 95% CI 0.15–0.61; 95% PI 0.15–0.61; Z = 4.08; $p < .001$). Between-study heterogeneity was not detected ($I^2 = 0\%$; Q = 5.77, $p = .450$).

Publication bias.   Visual inspection of the pre- to post-treatment funnel plot of interpersonal difficulty ES (see Figure 3b) suggested some slight asymmetry in the distribution of studies; however, statistical testing with Egger's regression did not detect significant reporting bias (B = 0.28, t(11) = 1.18, $p = .265$). Trim and Fill imputed data for one missing smaller study with a small deterioration effect after CAT, resulting in minimal change in the CAT ES estimate (ES = 0.73, 95% CI 0.53–0.92).

Effect of CAT on depression

Fourteen comparisons (extracted from 12 studies) totalling $N = 586$ participants provided pre–post treatment data for depression outcomes. The pooled ES demonstrated large, significant reductions in depression symptoms (Figure 2c; ES = 1.05, 95% CI 0.80–1.29; 95% PI 0.26–1.84; Z = 9.17; $p < .001$). There was significant, large between-study heterogeneity ($I^2 = 80\%$; Q = 64.56, $p < .001$). Comparisons between outcomes at end of treatment and follow-up were available from 8 comparisons ($N = 193$ participants). There was minimal change at follow-up indicating depression symptomatology remained stable (ES = 0.13; 95% CI −0.08 to 0.35; 95% PI −0.28 to 0.55; Z = 1.47; $p = .071$). Between-study heterogeneity was low and not significant ($I^2 = 41\%$; Q = 11.86, $p = .105$).

Publication bias.   Visual inspection of the pre- to post-treatment funnel plot of depression ES (see Figure 3c) suggested there was some asymmetry in the distribution of

studies, indicating larger effects of CAT on depression outcomes were more likely to be reported by small N studies. Egger's regression did not detect a statistically significant influence of reporting bias (B = 0.72, t(13) = 0.60, p = .561). Trim and Fill imputed three studies with minimal to small CAT treatment effects producing a reduced effect estimate of 0.88 (95% CI 0.62–1.14) suggesting some impact of reporting bias, albeit still representing a large effect.

Moderator analyses

Meta-regressions (Table 2) and subgroup analyses (Table 3) investigating moderators of CAT treatment effects explored the significant between-study heterogeneity identified in the pre–post treatment comparisons for each outcome. Variations in treatment effects for global functioning, interpersonal problems, and depression symptoms were not explained by differences in participant age or gender. Initial analyses suggested smaller ESs were associated with longer CAT treatment for global functioning and higher study quality for interpersonal problems; however, after accounting for multiple testing both effects were no longer significant at the Bonferroni-adjusted p-value (< .0125). Analysis of ESs for categorical subgroups found no significant differences based on the presence of PD cases versus other presentations, or whether the study was practice-based or an RCT. Effects for global functioning and depression symptoms were not associated with variations in ESs for different formats of CAT, or when therapy was delivered by a qualified CAT therapist. However, larger ESs for interpersonal difficulties were associated with individual treatment delivered by CAT qualified therapists, explaining 53% and 20% of the observed variance, respectively (although after adjustment, the effect of therapist qualification fell just short of the Bonferroni-corrected significance threshold). Moderate-to-large heterogeneity was still present in over half the subgroups.

Sensitivity analyses

Sensitivity analyses explored the impact of imputed pre–post correlation values (r) on overall CAT treatment estimates and are presented in Table 4. For each outcome, the magnitude of the CAT treatment effect with different r imputations ranged between moderate and large effects in favour of a significant improvement following CAT. This suggests imputation of missing values had a minimal to low impact on the overall conclusions that could be drawn. Depression outcome had a wider range of potential effect magnitude (0.60–1.30), than global functioning (0.72–0.97) and interpersonal difficulties (0.57–0.87), thus indicating interpretations may be slightly less reliable for depression outcomes. In addition, eight studies reported r for global functioning scores enabling a comparison of pooled treatment estimates using the true value or an imputed value (0.6) of r. The pooled treatment estimate was 0.94 (95% CI 0.54–.1.33; Z = 5.65; p < .001) when using an imputed estimate of r (0.6) and 0.98 (95% CI 0.65–.1.30; Z = 7.05; p < .001) when using the true study value of r.

Secondary meta-analysis of RCT between-groups treatment effects

Meta-analytic comparisons were conducted to aggregate the effect of CAT compared to a comparator condition in RCTs at post-treatment, and where available, at follow-up. One RCT could not be included as CAT was compared to a dismantled version of CAT [19]. In the remaining nine RCTs, CAT was compared to another intervention in k = 3 studies,

Table 2. Meta-regression analyses of pre–post CAT treatment outcomes

Outcome	Variable		k	B-coefficient	95% CI	SE	p	R^2 (%)
Global functioning	Quality	(rating 0–28)	21	−0.01	−0.06 to 0.04	0.02	.666	1.22
	Gender	(% of males)	19	0.00	−0.01 to 0.00	0.00	.294	7.90
	Mean age	(16 − 49 years)	16	0.01	−0.01 to 0.04	0.01	.191	11.24
	Treatment duration	(5–30 sessions)	20	−0.03	−0.06 to 0.00	0.01	**.029***	21.78
Interpersonal functioning	Quality	(rating 0–28)	12	−0.06	−0.11 to −0.01	0.02	**.015***	36.71
	Gender	(% of males)	12	0.00	−0.01 to 0.01	0.00	.586	2.51
	Mean age	(31–49 years)	9	0.02	−0.03 to 0.07	0.02	.271	10.74
	Treatment duration	(16–24 sessions)	12	0.01	−0.04 to 0.06	0.02	.572	2.80
Depression	Quality	(rating 0–28)	14	0.03	−0.03 to 0.08	0.03	.314	8.48
	Gender	(% of males)	13	0.00	−0.01 to 0.01	0.01	.656	1.69
	Mean age	(27–49 years)	10	0.02	−0.03 to 0.06	0.02	.475	5.31
	Treatment duration	(8–24 sessions)	12	−0.02	−0.07 to 0.03	0.02	.411	7.48

Note. CI = confidence interval; k = number of comparisons; R^2 = percentage of variation explained; SE = standard error.
*Significant at $p < .05$ threshold.; **Significant at Bonferroni-adjusted $p < .0125$ threshold for multiple testing. Positive effect size indicates in favour of pre–post improvement.

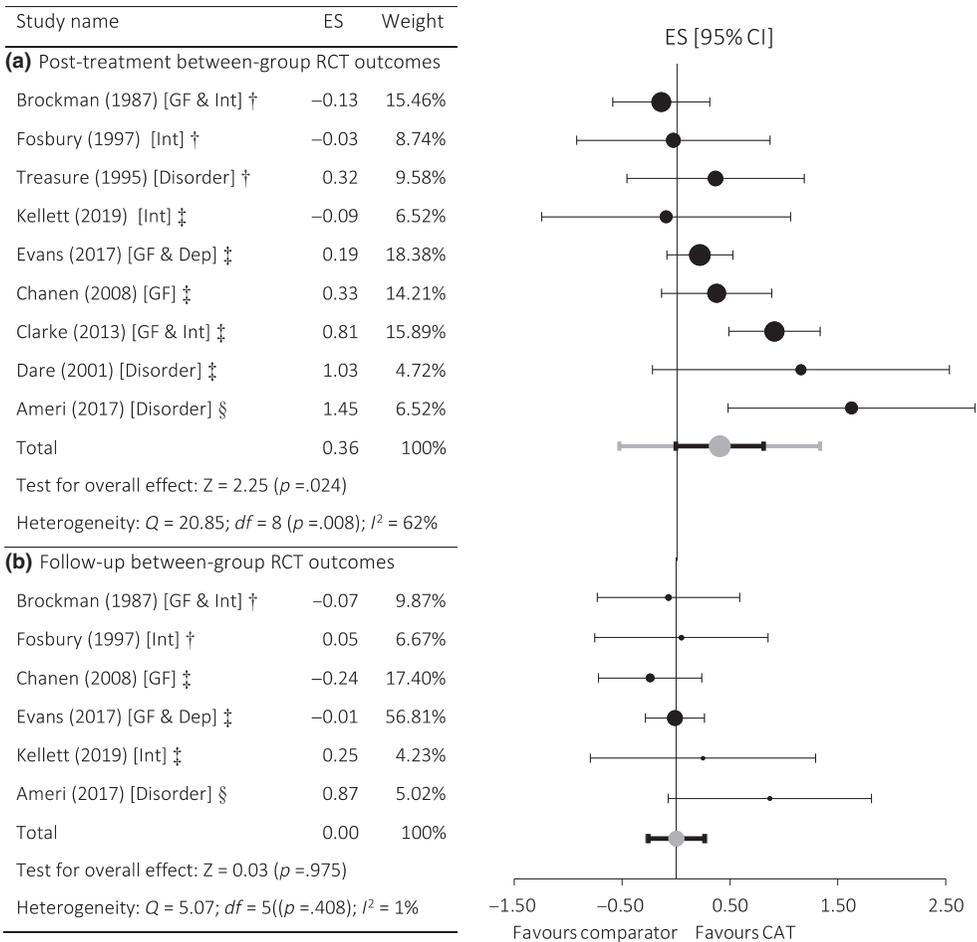**Table 3.** Subgroup analysis of pre–post CAT treatment effect outcomes

Outcome	Variable	Subgroup	k	ES (g)	95% CI	I^2 (%)	p (between subgroups)	R^2 (%)
Global functioning	Study type	PBE	17	0.87*	0.70 to 1.04	75*	.641	0.32
		RCT	4	0.80*	0.55 to 1.05	23		
	Therapist CAT qualified	Qualified	15	0.90*	0.73 to 1.06	57*	.500	0.67
		Not qualified	6	0.78*	0.49 to 1.07	84*		
	Format	1:1	20	0.89*	0.76 to 1.03	54*	–	–
		Group	1	0.37*	–	–		
	Diagnosis	Include PD cases	6	0.81*	0.70 to 0.91	0	.531	0.58
		Other diagnoses	15	0.88*	0.68 to 1.07	79*		
Interpersonal functioning	Study type	PBE	9	0.80*	0.61 to 0.98	68*	.118	8.01
		RCT	3	0.48*	0.12 to 0.84	50		
	Therapist CAT qualified	Qualified	6	0.94*	0.74 to 1.14	22	**.015***	19.81
		Not qualified	5	0.59*	0.39 to 0.79	69*		
	Format	1:1	10	0.79*	0.67 to 0.90	24	**<.001****	52.71
		Group	2	0.36*	0.19 to 0.54	2		
	Diagnosis	Include PD cases	4	0.76*	0.24 to 1.28	75*	.892	0.06
		Other diagnoses	8	0.72*	0.55 to 0.89	61*		
Depression	Study type	PBE	10	1.03*	0.76 to 1.30	78*	.864	0.05
		RCT	4	1.08*	0.60 to 1.56	83*		
	Therapist CAT qualified	Qualified	7	1.22*	0.89 to 1.56	59*	.259	1.98
		Not qualified	6	0.95*	0.59 to 1.30	87*		
	Format	1:1	13	1.11*	0.89 to 1.32	70*	–	–
		Group	1	0.49*	–	–		
	Diagnosis	Include PD cases	1	0.85*	–	–	–	–
		Other diagnoses	13	1.07*	0.82 to 1.31	81*		

Notes. CI = confidence interval; ES = effect size; k = number of comparisons; PBE = practice-based evidence; PD = personality disorder; R^2 = percentage of variation explained; RCT = randomized controlled trial.
*Significant at p < .05 threshold.; **Significant at Bonferroni-adjusted p < .0125 threshold for multiple testing of between subgroups. Positive effect size indicates in favour of pre–post improvement.

Table 4. Sensitivity analysis of imputed pre–post correlations (*r*) on aggregated CAT treatment effect

Outcome	*k*	Main analysis estimates 0.6	Imputed value of *r* 0.0	0.25	0.5	0.75
Global functioning	21	0.86 (0.71–1.01)	0.72 (0.53–0.92)	0.75 (0.58–0.93)	0.81 (0.66–0.97)	0.97 (0.81–1.13)
Interpersonal functioning	11	0.74 (0.54–0.94)	0.57 (0.35–0.79)	0.61 (0.41–0.81)	0.68 (0.49–0.88)	0.87 (0.62–1.12)
Depression	14	1.05 (0.80–1.29)	0.66 (0.49–0.83)	0.77 (0.58–0.96)	0.94 (0.72–1.17)	1.30 (0.99–1.61)

Study name	ES	Weight
(a) Post-treatment between-group RCT outcomes		
Brockman (1987) [GF & Int] †	−0.13	15.46%
Fosbury (1997) [Int] †	−0.03	8.74%
Treasure (1995) [Disorder] †	0.32	9.58%
Kellett (2019) [Int] ‡	−0.09	6.52%
Evans (2017) [GF & Dep] ‡	0.19	18.38%
Chanen (2008) [GF] ‡	0.33	14.21%
Clarke (2013) [GF & Int] ‡	0.81	15.89%
Dare (2001) [Disorder] ‡	1.03	4.72%
Ameri (2017) [Disorder] §	1.45	6.52%
Total	0.36	100%
Test for overall effect: Z = 2.25 (*p* =.024)		
Heterogeneity: *Q* = 20.85; *df* = 8 (*p* =.008); *I*² = 62%		
(b) Follow-up between-group RCT outcomes		
Brockman (1987) [GF & Int] †	−0.07	9.87%
Fosbury (1997) [Int] †	0.05	6.67%
Chanen (2008) [GF] ‡	−0.24	17.40%
Evans (2017) [GF & Dep] ‡	−0.01	56.81%
Kellett (2019) [Int] ‡	0.25	4.23%
Ameri (2017) [Disorder] §	0.87	5.02%
Total	0.00	100%
Test for overall effect: Z = 0.03 (*p* =.975)		
Heterogeneity: *Q* = 5.07; *df* = 5((*p* =.408); *I*² = 1%		

† CAT compared to another intervention; ‡ CAT compared to treatment as usual; § CAT compared to no treatment.
Abbreviations: GF: Global functioning outcome; Int: Interpersonal Problems outcome; Dep: Depression outcome; Disorder: Disorder specific outcome; RCT: Randomised controlled trial; ES: Effect size; CI: Confidence interval

Figure 4. Forest plots of between-group effect sizes (ES) from randomized controlled trial (RCT) outcomes of CAT versus a comparator condition at (a) post-treatment and (b) follow-up [grey error bars represent prediction intervals for overall aggregate estimates].

treatment as usual (TAU) in $k = 5$ studies, and no treatment in $k = 1$ study (see supplementary materials for between-group comparator characteristics). Outcomes assessed were disorder-specific outcomes ($k = 3$ [Anorexia Nervosa $k = 2$; OCD $k = 1$]), interpersonal problems ($k = 2$), global functioning ($k = 1$), or a combination of global functioning, interpersonal problems, and/or depression outcomes ($k = 3$). Dropout rates for CAT during clinical trials ranged between 0 and 38% (mean = 23%) compared to 6-44% (mean = 26%) for comparators.

Post-treatment and follow-up comparisons
Nine RCT comparisons totalling $N = 352$ participants evaluated post-treatment CAT outcomes compared to a control condition (CAT $N = 181$; control $N = 171$). The pooled

ES presented in Figure 4 indicates a small, significant treatment effect in favour of CAT (ES = 0.36, 95% CI 0.00–0.72; 95% PI- 0.48–1.19; Z = 2.25; p = .024). There was significant, moderate between-study heterogeneity (I^2 = 62%; Q = 20.85, p = .008). The pooled ES including only the six comparisons of post-treatment CAT outcomes versus TAU/no treatment conditions showed a moderate effect in favour of CAT (ES = 0.53, 95% CI 0.01–1.05, 95% PI −0.49 to 1.55, Z = 2.63, p = .009, I^2 = 64%, Q = 13.99, p = .016). Six follow-up treatment arm comparisons provided follow-up data on outcomes for CAT versus any control conditions for N = 175 participants (CAT N = 88; control N = 87). The pooled ES demonstrated there was no difference in follow-up outcomes (Figure 4; ES = 0.00, 95% CI −0.26 to 0.26; 95% PI- 0.27–0.28; Z = 0.03; p = .975). Between-study heterogeneity was not detected (I^2 = 1%; Q = 5.07, p = .408). The pooled follow-up ES for the four TAU/no treatment comparisons only was ES = 0.06 (95% CI −0.56 to 0.68; 95% PI- 0.88 to 1.00; Z = 0.31; p = .758, I^2 = 40%, Q = 5.00, p = .172).

Discussion

CAT is defined through its relational and collaborative approach, in working with the past, using enactments within the therapeutic relationship and associated analysis of habitual relationship patterns, via analysis of reciprocal role dynamics and intra- and interpersonal procedures (Ryle & Kellett, 2018). The present study has built on previous reviews of the effectiveness of CAT (Calvert & Kellett, 2014; Ryle *et al*., 2014) to provide a contemporary quantitative synthesis of the state of the CAT treatment outcome evidence base. This was achieved by specifying CAT pre–post treatment effects for global functioning, interpersonal difficulties, and depression, and also quantifying the longer-term effects of CAT for the first time. The uncontrolled ES found was comparable to the pre–post ES found in the previous quantitative review (d = 0.83; Ryle *et al*., 2014). Moderators of CAT treatment effects were explored, alongside consideration of the impact of largely PBE studies on the confidence in overall treatment estimates. An aggregated controlled CAT effect size has also been estimated to provide a preliminary indication of CAT effectiveness, in the context of the acknowledged limitations of uncontrolled pre–post treatment effects. CAT continues to be used to treat a variety of typically severe and complex psychological disorders; 12 separate diagnoses were present in the current review with evaluations of CAT for disorders including psychosis, morbid jealousy, and chronic pain recently completed. However, less than 25% of studies assessed adverse event rates during CAT, and so, this needs to become a more widely assessed safety outcome.

The findings overall show patients who undergo CAT experience improvements across a range of clinical difficulties, seeing moderate-to-large pre–post reductions in global symptoms (g = 0.86), interpersonal difficulties (g = 0.74), and depression (g = 1.05) and small–moderate beneficial effects compared to comparators (g = 0.36–0.53). Although the pre–post improvements cannot be specifically attributed to CAT, the superior between-groups outcomes suggest CAT is effective. Evidence of outcome reductions was maintained at medium-/long-term follow-up in functioning and depression, and it is noteworthy that interpersonal difficulties significantly reduced over follow-up time. Dropout was low (15%) suggesting that CAT is an acceptable to patients.

Overall, 28 CAT treatment studies were synthesized, with 25 studies included in a synthesis of pre–post outcomes and nine studies in a RCT between-groups synthesis. This encompassed an additional N = 12 studies (including four RCTs) that had been published since the CAT reviews in 2014 called for urgent development of the evidence base, and in

particular the conduct of RCTs. The increase in the number of RCTs (i.e., 36% of studies were RCTs) indicates that the CAT evidence base is developing. However, there is still a tendency towards favouring PBE-style evaluations typically in complex clinical populations. It is encouraging that CAT has developed translations of the model in terms of group delivery (Calvert, Kellett, & Hagan, 2015), a consultancy version for patients in secondary care that community teams struggle to engage (Kellett *et al.*, 2019c), an 8-session version suitable for step 3 delivery in Improving Access to Psychological Therapies (IAPT) services (Kellett *et al.*, 2018), forensic CAT (Kellett *et al.*, 2019d), and a psychoeducational version for step 2 delivery in IAPT services (Meadows & Kellett, 2017).

Assessment of the specific impact of CAT on interpersonal difficulties was particularly indicated, as CAT presents itself as a relationally informed psychotherapy, capable of conceptualizing and changing interpersonal processes (Ryle & Kellett, 2018). In terms of pre–post change, then large improvements in global functioning and depression and a moderate improvement in interpersonal difficulties were observed following CAT. However, the follow-up meta-analysis revealed little further change in functioning and depression (i.e., neither significant improvements nor deterioration indicating that treatment gains were maintained), but with significant improvements in interpersonal distress occurring over the follow-up period. This suggests that the relational nature of CAT (Ryle & Kellett, 2018) is impacting on interpersonal dynamics usefully over time. The results concerning interpersonal change therefore provide support for the underlying theory and approach of CAT.

The high presence of heterogeneity between studies warranted exploration through meta-regression and subgroup analyses; however, limited moderators of CAT treatment effect were identified. While it is important to remember moderator analysis cannot be used to infer causality, it provides insight into potential treatment effects (Cochrane Collaboration, 2011ollaboration, 2011). ESs across all outcomes did not differ according to participant age, gender, diagnosis, or for the outcomes in RCT versus PBE studies. However, it is likely these analyses were hampered by small sample sizes and the limited number of studies within subgroup analyses creating a lack of power to detect small effects (Cochrane Collaboration, 2011ollaboration, 2011). All outcomes showed a trend of larger effects for individually delivered CAT compared to group formats; however, only interpersonal outcomes had multiple studies of group CAT to enable subgroup comparisons. One-to-one CAT produced larger effects, whereas group CAT outcomes were significantly more modest. This may be a reflection of the relative lack of theoretical work concerning group delivery of the model, as only two group CAT studies were available and were in complex patient groups (trauma/forensic) that could be confounding the effect. Smaller global functioning effects were associated with longer treatment contracts, which is possibly an artefact of more complex clinical presentations being offered the 24-session version of the model in routine services (Marriott & Kellett, 2009).

Limitations

The findings need to be considered alongside methodological limitations. Many of the limitations of the current review reflect the limitations of the evidence base, rather than the methods used in this study. CAT is clearly frequently used in the treatment of highly varied patient groups, and this limited the specificity of the review, particularly through the lack of bodies of disorder-specific CAT outcome research. CAT has generated most evidence around the treatment of borderline personality disorder. Restricting the search to peer-reviewed studies may have overlooked eligible studies in the grey literature. The

review would have been improved through more consistent efforts to access the grey literature. The review was limited by the wide variety of measures used in the CAT outcome literature, so the creation of outcome classification clusters could be challenged and may have introduced bias (Puhan, Soesilo, Guyatt, & Schünemann, 2006). There was a lack of studies comparing CAT to active treatments, and so, the results cannot be used to comment reliably on the comparative efficacy of CAT (Bucher, Guyatt, Griffith, & Walter, 1997).

The number of included studies within outcomes and within-study sample sizes was small, which can inflate ESs and provide inaccurate evaluations of between-study heterogeneity (Inthout, Ioannidis, Borm, & Goeman, 2015). In addition, subsequent moderator analyses were subject to low power and insufficient subgroups to be able to reliably detect variation in effects (Guolo & Varin, 2017). For example, the number of studies of adults within specific diagnoses was low, especially when compared to similar meta-analyses for other treatment modalities (Tolin *et al.*, 2015). The methodological quality of included studies was generally sub-optimal, and there were indications that for some outcomes, lower quality studies might have produced larger effects. Poor study quality is commonly criticized for contributing inflated ESs (Altman, 1994). The lack of studies employing the CCAT (i.e., 16%) means that the certainty with which CAT was actually being delivered is questionable. Assessing dropout by calculating an average is problematic as precision and variance in study estimates is not accounted for. The amount of studies assessing serious adverse events was low (23%).

Finally, the type of the evidence available means that ESs may have been susceptible to uncontrolled error thus biasing their interpretation. As the CAT evidence base has a strong reliance on PBE, pre–post rather than between-group ESs were used in the primary analysis. Critics argue such uncontrolled ESs should be avoided in meta-analyses, as they may index change caused by external factors (Cuijpers, Weitz, Cristea, & Twisk, 2017). To control for violation of independence in pre–post ESs alongside limited reporting of the correlation of pre–post scores, a fixed correlation value was used when data were missing. However, when this fixed value is considerably different from the true correlation, it has been shown to inflate ES estimates (Cuijpers *et al.*, 2017). While attempts were made to manage these biases through the use of Hedges *g* correction for small sample sizes, accounting for pre–post correlation and assessing the impact of different imputed values, it does raise questions as to whether these samples were suitable to be combined.

Research, clinical, and organizational implications

Researchers still need to strive to generate evidence, particularly using randomized and controlled methods, and produce CAT evidence for specific diagnoses. CAT needs to generate more evidence for its acceptability and effectiveness with common mental health problems. The evidence base comprises mostly of one-to-one CAT delivery, and more evaluations of group CAT are indicated. It is essential that CAT outcome studies consistently report adverse event rates. Competency assessment should be the norm rather than the exception in PBE- and EBP-style studies. Evaluation of CAT is hampered by a lack of consensus on types of outcome used to evaluate treatment, largely it appears due to the complexity and variety of disorders treated. Use of a generic measure of symptomatic/functioning, such as CORE-OM (Barkham, Gilbert, Connell, Marshall, & Twigg, 2005), would enable more robust and widespread comparisons across studies. The more frequent use of the Personality Structure Questionnaire (PSQ; Pollock, Broadbent,

Clarke, Dorrian, & Ryle, 2001) is also indicated as this is based on the CAT model and has recently been cross-culturally validated (Berrios, Kellett, Fiorani, & Poggioli, 2016).

Given the frequency of PBE-style CAT studies, it would be useful for future pre–post studies to report correlations between pre–post scores as standard. Short- and long-term follow-up needs to be routinely built into the design of any future CAT PBE or EBP studies to clarify the true durability of CAT treatment effects. Although the current study provides a basic commentary on dropout rates as a proxy for treatment acceptability, the evidence base would benefit from a meta-analysis of dropout from routine service delivery and clinical trials of CAT (see Imel, Laska, Jakupcak, & Simpson, 2013 for an example). Future studies need to enable increased patient choice, and so, completion of patient preference trials is indicated and is underway (Kellett *et al*., 2019a). The present meta-analysis was based on nomothetic psychometric outcomes. However, as change in ideographic 'target problems' specific to the patient is at the core of evaluating change in the model (Ryle & Kerr, 2002), a meta-analysis of the large CAT single-case experiential design literature is also indicated.

Both CAT accredited therapists and therapists without CAT accreditation produced statistically similarly large treatment effects for symptomatic and functioning changes. However, reductions in interpersonal difficulties when CAT was delivered by unaccredited therapists were moderate and were significantly lower than the large effects observed for accredited CAT therapists (albeit just short of significance after adjusting for multiple testing). Given that the implementation of methods to help patients change their interpersonal roles and procedures is thought to play a crucial role in the benefits experienced (Ryle & Kellett, 2018), this implies that the interpersonal work of CAT needs to be supported through formal training and associated clinical supervision.

Conclusions

The aims of the study were met, and a contemporary and methodologically improved meta-analysis of the CAT evidence base was produced. The findings add to a growing body of evidence suggesting relationally informed therapies can be beneficial in reducing psychological distress and interpersonal difficulties (Fonagy, 2015; Jakobsen, Hansen, Simonsen, & Gluud, 2011). The results of this review highlight an emerging, but still relatively small diagnostically specific CAT evidence base. Despite methodological limitations, the within-groups and preliminary between-groups findings taken together suggest CAT is useful across a range of clinical presentations. Patients particularly appear to benefit interpersonally, and the findings reported here do support the commissioning of CAT for disorders with a significant interpersonal element. In conclusion, these findings should provide sufficient impetus for a coordinated research strategy to move the CAT evidence base forward in a targeted and productive manner.

Conflicts of interest

All authors declare no conflict of interest.

Data availability statement

Data sharing is not applicable to this article as no new data were created or analysed in this study.

References

References marked with an asterisk indicate CAT treatment studies included in the meta-analysis.

Altman, D. G. (1994). The scandal of poor medical research. *British Medical Journal*, *308*, 283. https://doi.org/10.1136/bmj.308.6924.283

*Ameri, M. A., Yousefi, A., Rad, S. A., & Khorshidi, F. R. (2017). The effectiveness of cognitive analytic therapy on the severity of symptoms in patients with obsessive-compulsive disorder. *International Journal of Information and Education Technology*, *7*, 212–216. https://doi.org/10.18178/ijiet.2017.7.3.868

Balk, E. M., Earley, A., Patel, K., Trikalinos, T. A., & Dahabreh, I. J. (2012). *Empirical assessment of within-arm correlation imputation in trials of continuous outcomes. Methods Research Report AHRQ Publication No. 12 (13)-EHC141-EF*. Rockville, MD: Agency for Healthcare Research and Quality. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23326900

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. *British Journal of Psychiatry*, *186*, 239–46. https://doi.org/10.1192/bjp.186.3.239

Barkham, M., Stiles, W. B., Lambert, M. J., & Mellor-Clark, J. (2010). Building a rigorous and relevant knowledge base for the psychological therapies. In M. Barkham, G. E. Hardy & J. Mellor-Clark (Eds.), *Developing and delivering practice-based evidence: A guide for the psychological therapies* (pp. 21–61). Chichester, UK: John Wiley & Sons Ltd. https://doi.org/10.1002/9780470687994.ch2

*Baronian, R., & Leggett, S. J. E. (2019). Brief cognitive analytic therapy for adults with chronic pain: A preliminary evaluation of treatment outcome. *British Journal of Pain*, *14*(1), 57–67. https://doi.org/10.1177/2049463719858119

Bennett, D., & Parry, G. (2004). A measure of psychotherapeutic competence derived from cognitive analytic therapy. *Psychotherapy Research*, *14*, 176–92. https://doi.org/10.1093/ptr/kph016

Berrios, R., Kellett, S., Fiorani, C., & Poggioli, M. (2016). Assessment of identity disturbance: Factor structure and validation of the personality structure questionnaire in an Italian sample. *Psychological Assessment*, *28*, e27–e35. https://doi.org/10.1037/pas0000143

*Birtchnell, J., Denman, C., & Okhai, F. (2004). Cognitive analytic therapy: Comparing two measures of improvement. *Psychology and Psychotherapy: Theory, Research and Practice*, *77*, 479–92. https://doi.org/10.1348/1476083042555398

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed.) (pp. 221–35). New York, NY: Russell Sage Foundation.

*Brockman, B., Poynton, A., Ryle, A., & Watson, J. P. (1987). Effectiveness of time-limited therapy carried out by trainees. Comparison of two methods. *British Journal of Psychiatry*, *151*, 602–10. https://doi.org/10.1192/bjp.151.5.602

Bucher, H. C., Guyatt, G. H., Griffith, L. E., & Walter, S. D. (1997). The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology*, *50*, 683–91. https://doi.org/10.1016/S0895-4356(97)00049-8

Calvert, R., & Kellett, S. (2014). Cognitive analytic therapy: A review of the outcome evidence base for treatment. *Psychology and Psychotherapy: Theory, Research and Practice*, *87*, 253–77. https://doi.org/10.1111/papt.12020

*Calvert, R., Kellett, S., & Hagan, T. (2015). Group cognitive analytic therapy for female survivors of childhood sexual abuse. *British Journal of Clinical Psychology*, *54*, 391–413. https://doi.org/10.1111/bjc.12085

*Chanen, A. M., Jackson, H. J., McCutcheon, L. K., Jovev, M., Dudgeon, P., Hok, P. Y., . . . McGorry, P. D. (2008). Early intervention for adolescents with borderline personality disorder using cognitive analytic therapy: Randomised controlled trial. *British Journal of Psychiatry*, *193*, 477–84. https://doi.org/10.1192/bjp.bp.107.048934

*Clarke, S., & Llewelyn, S. (1994). Personal constructs of survivors of childhood sexual abuse receiving cognitive analytic therapy. *The British Journal of Medical Psychology*, *67*, 273–89. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7803319

*Clarke, Sue, & Pearson, C. (2000). Personal constructs of male survivors of childhood sexual abuse receiving cognitive analytic therapy. *British Journal of Medical Psychology*, *73*, 169–77. https://doi.org/10.1348/000711200160408

*Clarke, S., Thomas, P., & James, K. (2013). Cognitive analytic therapy for personality disorder: Randomised controlled trial. *British Journal of Psychiatry*, *202*, 129–34. https://doi.org/10.1192/bjp.bp.112.108670

Cochrane Collaboration (2011). *Cochrane handbook for systematic reviews of interventions*. In J. P. T. Higgins, & S. Green (Eds.) (version 5). Retrieved from www.handbook.cochrane.org

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46. https://doi.org/10.1177/001316446002000104

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–9. https://doi.org/10.1037/0033-2909.112.1.155

Cuijpers, P., Donker, T., Weissman, M. M., Ravitz, P., & Cristea, I. A. (2016). Interpersonal psychotherapy for mental health problems: A comprehensive meta-analysis. *American Journal of Psychiatry*, *173*, 680–7. https://doi.org/10.1176/appi.ajp.2015.15091141

Cuijpers, P., Weitz, E., Cristea, I. A., & Twisk, J. (2017). Pre-post effect sizes should be avoided in meta-analyses. *Epidemiology and Psychiatric Sciences*, *26*, 364–8. https://doi.org/10.1017/S2045796016000809

*Curling, L., Kellett, S., & Totterdell, P. (2018). Cognitive analytic therapy for obsessive morbid jealousy: A case series. *Journal of Psychotherapy Integration*, *28*, 537–55. https://doi.org/10.1037/int0000122

*Dare, C., Eisler, I., Russell, G., Treasure, J., & Dodge, L. (2001). Psychological therapies for adults with anorexia nervosa. *British Journal of Psychiatry*, *178*, 216–21. https://doi.org/10.1192/bjp.178.3.216

*Darongkamas, J., Newell, A., Hewitt-Moran, T., & Jordan, S. (2017). *Clinical outcomes of cognitive analytic therapy delivered by trainees*. *Reformulation*, *Winter*, 38–42.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*, 177–88. https://doi.org/10.1016/0197-2456(86)90046-2

Downs, S., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomised studies of health care interventions. *Journal of Epidemiololgy and Community Health*, *52*, 377–84. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1756728/pdf/v052p00377.pdf

*Dunn, M., Golynkina, K., Ryle, A., & Watson, J. P. (1997). A repeat audit of the cognitive analytic therapy clinic at Guy's Hospital. *Psychiatric Bulletin*, *21*, 165–8. https://doi.org/10.1192/pb.21.3.165

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–63. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*, 629–34. https://doi.org/10.1136/BMJ.315.7109.629

*Evans, J., & Parry, G. (1996). The impact of reformulation in cognitive-analytic therapy with difficult-to-help clients. *Clinical Psychology and Psychotherapy*, *3*, 109–17. https://doi.org/10.1002/(SICI)1099-0879(199606)3:2<109:AID-CPP65>3.0.CO;2-U

*Evans, M., Kellett, S., Heyland, S., Hall, J., & Majid, S. (2017). Cognitive analytic therapy for bipolar disorder: A pilot randomized controlled trial. *Clinical Psychology and Psychotherapy*, *24*, 22–35. https://doi.org/10.1002/cpp.2065

Fonagy, P. (2015). The effectiveness of psychodynamic psychotherapies: An update. *World Psychiatry*, *14*, 137–50. https://doi.org/10.1002/wps.20235

*Fosbury, J. A., Bosley, C. M., Ryle, A., Sönksen, P. H., & Judd, S. L. (1997). A trial of cognitive analytic therapy in poorly controlled type I patients. *Diabetes Care*, *20*, 959–64. https://doi.org/10.2337/diacare.20.6.959

Guolo, A., & Varin, C. (2017). Random-effects meta-analysis: The number of studies matters. *Statistical Methods in Medical Research*, *26*, 1500–18. https://doi.org/10.1177/0962280215583568

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press. https://doi.org/10.1016/C2009-0-03396-0

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*, 557–60. https://doi.org/10.1136/BMJ.327.7414.557

Imel, Z. E., Laska, K., Jakupcak, M., & Simpson, T. L. (2013). Meta-analysis of dropout in treatments for posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology*, *81*(3), 394–404. https://doi.org/10.1037/a0031474

Inthout, J., Ioannidis, J. P. A., Borm, G. F., & Goeman, J. J. (2015). Small studies are more heterogeneous than large ones: A meta-meta-analysis. *Journal of Clinical Epidemiology*, *68*, 860–9. https://doi.org/10.1016/j.jclinepi.2015.03.017

Jakobsen, J. C., Hansen, J. L., Simonsen, E., & Gluud, C. (2011). The effect of interpersonal psychotherapy and other psychodynamic therapies versus "treatment as usual" in patients with major depressive disorder. *PLoS One*, *6*, e19044. https://doi.org/10.1371/journal.pone.0019044

Kellett, S., Bee, C., Aadahl, V., Headley, E., & Delgadillo, J. (2019a). Patient preference trial of cognitive behavioural versus cognitive analytic guided self-help in IAPT: Study protocol. *Behavioural and Cognitive Psychotherapy*. Manuscript submitted for publication.

*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 and Psychotherapy*, *20*, 216–25. https://doi.org/10.1002/cpp.796

*Kellett, S., Easton, K., Cooper, M., Millings, A., Simmonds-Buckley, M., & Parry, G. (2019b). *A mobile app to enhance relational awareness during cognitive analytic therapy: Results from a mixed methods case series*. Manuscript submitted for publication.

Kellett, S., Ghag, J., Ackroyd, K., Freshwater, K., Finch, J., Freear, A., . . . Simmonds-Buckley, M. (2019c). Delivering cognitive analytic consultancy to community mental health teams: Initial practice-based evidence from a multi-site evaluation. *Psychology and Psychotherapy: Theory, Research and Practice*, 1–27. https://doi.org/10.1111/papt.12221

*Kellett, S., Hall, J., & Compton Dickinson, S. (2019d). Group cognitive analytic music therapy: A quasi-experimental feasibility study conducted in a high secure hospital. *Nordic Journal of Music Therapy*, *28*, 224–55. https://doi.org/10.1080/08098131.2018.1529697

*Kellett, S., Stockton, C., Marshall, H., Hall, J., Jennings, C., & Delgadillo, J. (2018). Efficacy of narrative reformulation during cognitive analytic therapy for depression: Randomized dismantling trial. *Journal of Affective Disorders*, *239*, 37–47. https://doi.org/10.1016/j.jad.2018.05.070

Kızılırmak, P., Özdemir, O., & Öngen, Z. (2015). The most critical question when reading a meta-analysis report: Is it comparing apples with apples or apples with oranges? *Anatolian Journal of Cardiology*, *15*, 701–8. https://doi.org/10.5152/akd.2014.5665

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159. https://doi.org/10.2307/2529310

Margison, F. R., Barkham, M., Evans, C., McGrath, G., Clark, J. M., Audin, K., & Connell, J. (2000). Measurement and psychotherapy: Evidence-based practice and practice-based evidence. *British Journal of Psychiatry*, *177*, 123–30. https://doi.org/10.1192/bjp.177.2.123

*Marriott, M., & Kellett, S. (2009). Evaluating a cognitive analytic therapy service; Practice-based outcomes and comparisons with person-centred and cognitive-behavioural therapies. *Psychology and Psychotherapy: Theory, Research and Practice*, *82*, 57–72. https://doi.org/10.1348/147608308X336100

McHugh, R. K., & Barlow, D. H. (2012). *Dissemination and implementation of evidence-based psychological interventions*. Oxford, UK: Oxford University Press.

*Meadows, J., & Kellett, S. (2017). Development and evaluation of Cognitive Analytic Guided Self-Help (CAT-SH) for use in IAPT services. *Behavioural and Cognitive Psychotherapy*, *45*, 266–84. https://doi.org/10.1017/S1352465816000485

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . Whitlock, E. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*, 1. https://doi.org/10.1186/2046-4053-4-1

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*, 364–86. https://doi.org/10.1177/1094428106291059

O'Connor, S. R., Tully, M. A., Ryan, B., Bradley, J. M., Baxter, G. D., & McDonough, S. M. (2015). Failure of a numerical quality assessment scale to identify potential risk of bias in a systematic review: A comparison study. *BMC Research Notes*, *8*, 224–30. https://doi.org/10.1186/s13104-015-1181-1

Pollock, P. H., Broadbent, M., Clarke, S., Dorrian, A., & Ryle, A. (2001). The Personality Structure Questionnaire (PSQ): A measure of the multiple self states model of identity disturbance in cognitive analytic therapy. *Clinical Psychology and Psychotherapy*, *8*, 59–72. https://doi.org/10.1002/cpp.250

Puhan, M. A., Soesilo, I., Guyatt, G. H., & Schünemann, H. J. (2006). Combining scores from different patient reported outcome measures in meta-analyses: When is it justified? *Health and Quality of Life Outcomes*, *4*, 94. https://doi.org/10.1186/1477-7525-4-94

*Ryle, A., & Golynkina, K. (2000). Effectiveness of time-limited cognitive analytic therapy of borderline personality disorder: Factors associated with outcome. *British Journal of Medical Psychology*, *73*, 197–210. https://doi.org/10.1348/000711200160426

Ryle, A., & Kellett, S. (2018). Cognitive analytic therapy. In W. J. Livesley & R. Larstone (Eds.), *Handbook of personality disorders: Theory, research, and treatment* (pp. 489–511). New York, NY: Guilford Press.

Ryle, A., Kellett, S., Hepple, J., & Calvert, R. (2014). Cognitive analytic therapy at 30. *Advances in Psychiatric Treatment*, *20*, 258–68. https://doi.org/10.1192/apt.bp.113.011817

Ryle, A., & Kerr, I. B. (2002). *Introducing cognitive analytic therapy: Principles and practice. Introducing cognitive analytic therapy: Principles and practice*. Chichester, UK: John Wiley & Sons. https://doi.org/10.1002/9780470713587

Salkovskis, P. (1995). Demonstrating specific effects in cognitive and behavioural therapy. In M. Aveline & D. Shapiro (Eds.), *Research foundations for psychotherapy practice* (pp. 191–228). Chichester, UK: Wiley.

Suurmond, R., van Rhee, H., & Hak, T. (2017). Introduction, comparison, and validation of meta-essentials: A free and simple tool for meta-analysis. *Research Synthesis Methods*, *8*, 537–53. https://doi.org/10.1002/jrsm.1260

*Taylor, P. J., Perry, A., Hutton, P., Tan, R., Fisher, N., Focone, C., . . . Seddon, C. (2018). Cognitive analytic therapy for psychosis: A case series. *Psychology and Psychotherapy: Theory, Research and Practice*, *28*, 537–55. https://doi.org/10.1111/papt.12183

Tolin, D. F., McKay, D., Forman, E. M., Klonsky, E. D., & Thombs, B. D. (2015). Empirically supported treatment: Recommendations for a new model. *Clinical Psychology: Science and Practice*, *22*, 317–38. https://doi.org/10.1111/cpsp.12122

*Treasure, J., Todd, G., Brolly, M., Tiller, J., Nehmed, A., & Denman, F. (1995). A pilot study of a randomised trial of cognitive analytical therapy vs educational behavioral therapy for adult

anorexia nervosa. *Behaviour Research and Therapy*, *33*, 363–7. https://doi.org/10.1016/0005-7967(94)00070-Z

*Tyrer, R., & Masterson, C. (2019). Clients' experience of change: An exploration of the influence of reformulation tools in cognitive analytic therapy. *Clinical Psychology & Psychotherapy*, *26*, 167–74. https://doi.org/10.1002/cpp.2339

*Tzouramanis, P., Adamopoulou, A., Bozikas, V., Voikli, M., Zagora, C., Lombtzianidou, M., & Garyfallos, G. (2010). Evaluation of cognitive-analytic therapy (CAT) outcome in patients with panic disorder. *Psychiatrike*, *21*, 287–93.

*Wildgoose, A., Clarke, S., & Waller, G. (2001). Treating personality fragmentation and dissociation in borderline personality disorder: A pilot study of the impact of cognitive analytic therapy. *British Journal of Medical Psychology*, *74*, 47–55. https://doi.org/10.1348/000711201160795

Supporting Information

The following supporting information may be found in the online edition of the article:

Table S1. Rejected full-text studies and reason for rejection.

Table S2. Methodological quality ratings for included studies rated using the modified Downs & Black scale.

Table S3. Study characteristics for between-groups trial CAT and comparator arms.