
Article:
Kellett, S., Stockton, C., Marshall, H. et al. (3 more authors) (2018) Efficacy of narrative reformulation during cognitive analytic therapy for depression: Randomized dismantling trial. Journal of Affective Disorders, 239. pp. 37-47. ISSN 0165-0327

https://doi.org/10.1016/j.jad.2018.05.070

Citation: Kellett, S., Stockton, C., Marshall, H., Hall, J., & Delgadillo, J. (*in press*). Efficacy of narrative reformulation during Cognitive Analytic Therapy for depression: A randomized dismantling trial. *Journal of Affective Disorders*.

Efficacy of Narrative Reformulation during Cognitive Analytic Therapy for Depression: A Randomized Dismantling Trial

Stephen Kellett

Centre for Psychological Therapies Research, University of Sheffield, UK

Sheffield Health and Social Care NHS Foundation Trust, UK

Corrie Stockton

South West Yorkshire NHS Foundation Trust, UK

Helen Marshall

South West Yorkshire NHS Foundation Trust, UK

Jo Hall

University of Sheffield, UK

&

Jaime Delgadillo

University of Sheffield, UK

Correspondence:

Correspondence concerning this article should be addressed to Dr Stephen Kellett, Consultant Clinical Psychologist, University of Sheffield, Sheffield S10 2TN, UK. Electronic mail may be sent to s.kellett@sheffield.ac.uk

Abstract

Background: Narrative reformulation (NR) is a component of cognitive analytic therapy (CAT) that is assumed to increase client engagement and improve clinical outcomes. This study set out to test these claims. **Methods:** A single-blind randomized and controlled dismantling trial investigated treatment outcomes for moderate to severely depressed patients receiving CAT in Primary Care. Ninety-five participants were randomized to either full-CAT (n=52) or CAT minus narrative reformulation (CAT-NR, n=43). Treatment duration in both arms was 8-sessions and was matched apart from the omission of the NR in the CAT-NR arm. The primary outcome measure was the Patient Health Questionnaire (PHQ-9), with secondary outcome measures of anxiety, functioning, helpfulness and the therapeutic alliance. Outcomes were assessed at screening, every treatment session and at 8-weeks follow-up. **Results:** Growth curve modelling found that NR did not enhance the efficacy of CAT for depression. There were no significant differences between groups in terms of attendance, adverse events, anxiety, functioning, helpfulness or therapeutic alliance. There were large within-group effect sizes (d_+>1.5), as CAT with or without NR produced significant reductions in depression ($p<0.01$). **Limitations:** The primary outcome were assessed via self-report and the follow-up period was brief. **Conclusions:** These results suggest that NR may be redundant when treating depression with CAT. Whilst a brief 8-session version of the CAT model appears suitable for treating depression in Primary Care, further research regarding the need for NR is indicated.

Keywords: cognitive analytic therapy, deconstruction trial, depression

Whilst a large evidence base validates the efficacy of certain psychotherapies for depression (Cuijpers, Berking, Andersson, Quigley, Kleiboer, & Dobson, 2013), evidence identifying the specific and active ingredients of these effective treatments is scarce (Cuijpers, van Straten, Andersson, & van Oppen, 2008). The appropriate methodologies for identifying the active constituents of psychotherapies are deconstruction trials (Carrico & Antoni, 2008) or component analyses (Bell, Marcus & Goodlad, 2013). Such methodologies offer the opportunity to isolate causal relationships between therapy ingredients and outcomes (Borkovec & Sibrava, 2005) and allow conclusions to be drawn about the clinical need for specific treatment components or techniques (Czaja, Schultz, Lee, & Belle, 2003). 'Dismantling' trials assess the efficacy of a treatment when a specific component is removed, whilst 'additive' trials assess treatments with a specific component added (Ahn & Wampold, 2001). For example, an early landmark dismantling trial by Jacobson et al. (1996) found that the behavioural activation component of cognitive behaviour therapy (CBT) produced equivalent outcomes to full CBT. This study supported the application of behavioural activation as a stand-alone depression treatment (Mazzucchelli, Kane & Rees, 2009). Meta-analytic evidence illustrates that dismantling trials rarely yield any significant differences between study arms, whereas additive designs tend to yield small (but improved) treatment outcomes (Bell & D'Zurilla, 2009).

Cognitive analytic therapy (CAT) is a relational, integrative and time-limited psychotherapy informed by cognitive and psychodynamic theory/methods, which was specifically designed for use in pressured public services (Ryle & Kerr, 2002). Whilst CAT was initially developed in the UK, it is now practiced internationally, most notably in Ireland, Spain, Italy, Australia and Greece (Ryle, Kellett, Hepple & Calvert, 2014). Treatment is typically delivered in 8, 16 or 24 session versions of the model allocated according to patient complexity (Ryle & Kerr, 2002). The analytical aspect of CAT concerns the concept of

reciprocal role dynamics (Ryle & Kerr, 2002), in which the patient inhabits and enacts (both in and out of therapy) unhelpful, limited and stereotyped roles, which are often formed as a result of early developmental abuse/neglect. Roles can be self-self, self-other and other-self with the patient reversing between both ends of a reciprocal role (e.g. both criticising of self/others and eliciting criticism from others and also feeling put down/humiliated by the actions of others; summarised as a criticising to humiliated reciprocal role). The cognitive aspect of the model (Ryle & Kerr, 2002) concerns procedural sequences of aim, thought, feeling, action and consequences that result from reciprocal role enactment. The cognitive and analytic aspects of the CAT model were initially theoretically integrated via the object relations procedural sequence model (Ryle, 1991) and then further assimilated via the multiple self-states model (MMSM; Ryle, 1997).

CAT differs from cognitive-behavioural therapy for depression by taking a relational approach to symptomatology, working with the past, analysing enactments within the therapeutic relationship and associated analysis of habitual relationship patterns (Ryle & Kellett, 2017). The closest version of CBT to CAT would be schema therapy, as there is an emphasis during schema therapy on family history, developmental trauma and interpersonal relationships creating and maintaining distress (Young, Klosko & Weishaar, 2003). A model of how change is accomplished and achieved during CAT for depression has been recently developed using task analysis (Sahu, Kellett & Hardy, 2017). This model shows that change is typically achieved on the foundation stone of improved self-awareness of depressogenic reciprocal roles and associated procedures, through the development of an 'observing self' (Ryle & Kerr, 2002). The first tool used in CAT to facilitate greater self-awareness is narrative reformulation (NR). NR specifies the developmental origins of dysfunctional relational roles and patterns, highlighting possible enactments of such dysfunctional role procedures in the therapeutic relationship, identifying target problem procedures underlying/maintaining the

presenting problem, predicting reactions to termination and defining goals (Hamill, Reid & Reynolds, 2008). NR is presented in letter form, read to patients at the completion of assessment and patients are asked to review and then add to the NR (often as the first between-session task) and a final version is agreed (Ryle & Kellett, 2017).

NR has been proposed to be a central aspect of the CAT clinical method (Ryle & Kellett, 2017). Whilst NR only takes one session to deliver, preparation (and associated clinical supervision) time is significant due to the complexity of the task (Denman, 2001). The validity of reformulation during CAT to accurately reflect reciprocal roles and target problem procedures has previously been illustrated in a case study using the core conflictual relationship theme method and structural analysis of social behaviour - cyclic maladaptive pattern assessments (Bennett & Parry, 1998). NR can evoke both strong positive and negative emotional responses in patients (Rayner, Thompson & Walsh, 2011). A negative emotional reaction to NR would be viewed as a potential enactment of a reciprocal role (i.e. the patient experiences NR as criticising and feels humiliated), rather than the NR being inherently wrong in itself (Ryle & Kellett, 2017). NR has been widely assumed to be therapeutic in itself by lifting morale, strengthening therapeutic alliances via compassion, raising hope and focussing the treatment (Ryle, 1990; Ryle & Kerr, 2002). Qualitative evidence suggests that NR can help patients to feel accepted/understood, enhancing connections with self, the therapist and the therapy (Hamill et al. 2008). However, evidence concerning the symptomatic impact of NR is mixed. Two small studies (Evans & Parry, 1996; Shine & Westacott, 2010) failed to find any effect of NR on symptom amelioration. There has been evidence presented of sudden (beneficial) symptomatic change following NR in a case series of CAT for obsessive morbid jealousy (Curling, Kellett & Tottterdell, 2018) and during the single case experimental design evaluations of dissociative identity disorder, (Kellett, 2005), paranoid personality disorder (Kellett & Hardy, 2014), sex addiction (Kellett, Simmonds-Buckley & Totterdell, 2016) and

obsessive morbid jealousy (Curling, Kellett, Totterdell, Parry, Hardy & Berry, 2017). However, these small N studies are likely to be limited by selection biases and may not be representative of wider clinical populations.

In view of the inconclusive evidence outlined above, this study sought to isolate and test the efficacy of NR during CAT, due to (a) NR being potentially draining of therapist time (Denman, 2001), (b) unresolved questions concerning the clinical utility of NR, (c) the lack of depression trials within the CAT evidence base (Calvert & Kellett, 2014) and (d) the extant NR evidence base being wholly based on small studies. The primary aim of this study was to assess (in a suitably powered study) whether NR is a specific active ingredient of CAT. In order to achieve this aim, the study employed a dismantling deconstruction trial methodology (Bell, Marcus, & Goodlad, 2013). Therefore, treatment outcomes for depressed participants receiving treatment as usual (i.e. full-CAT), were contrasted with those who received CAT minus its narrative reformulation component (i.e. CAT-NR). The research is novel as the CAT evidence base does not contain any deconstruction trials. Study hypotheses were as follows: (1) participants in the full-CAT arm would achieve better depression outcomes and (2) experience a better therapeutic alliance and find therapy more helpful.

Method

Participants

Ethical and research governance approvals were obtained from the English National Health Service (NRES reference number: 10/H0405/53) and the trial was registered (CT reference number: 10/H0405/53). Participants were recruited and treated in a single Improving Access to Psychological Therapies (IAPT) service in the United Kingdom (UK), set in a socio-economically deprived community. IAPT services in the UK provide evidence-based

psychological therapies in Primary Care for anxiety and depression, using a stepped-care treatment model (Clark, 2011). General medical practitioners had initially diagnosed depression in potential participants, identified the need for a 'talking treatment' and referred them to the IAPT service. IAPT staff then allocated depressed patients to the trial after a screening appointment, with trial recruitment taking place via an initial study suitability meeting. The initial screening appointments were conducted by Psychological Wellbeing Practitioners, who are graduate workers in IAPT services that provide guided self-help interventions based on cognitive-behavioural theory. National curricula and associated assessment and treatment competency frameworks are available (UCL, 2015). PWPs are therefore specifically trained in the assessment of common mental health problems and make treatment recommendations as a routine aspect of their role (Firth, Barkham, Kellett & Saxon, 2015). Trial screenings were completed by the research team and if the participant was suitable for the trial then they were provided with a study information leaflet explaining the dismantling methods of the study (in lay person's language). In brief, participants were told that if they consented to participate in the trial they would be randomised to one of two arms (i.e. full treatment or treatment without a feedback letter from the therapist), and they would be blind to their allocation. No participants refused to participate in the trial after reading the information leaflet.

Figure 1 details the flow of participants through the stages of the study and documents the reasons for trial exclusion. In total, n=125 were screened for study suitability, with n=95 (76%) randomised (26 males and 69 females with an age range of 19-65) and n=30 (24%) excluded. Inclusion criteria were that the trial screening interview had to identify the presence of depression (interview conducted using DSM-IV criteria; American Psychiatric Association, 1994) and participants needed to have clinical case-level depression symptoms (i.e. score 10-21) on the Patient Health Questionnaire (PHQ-9, Spitzer, Kroenke & Williams, 1999).

Exclusion criteria were: a PHQ-9 score <10, not meeting DSM-IV diagnostic criteria for depression, significant ongoing risk issues (i.e. actively suicidal or currently self-harming), having a co-morbid anxiety disorder which was the primary reason for referral to treatment, reluctance to engage in psychotherapy, previous in-patient admission, significant amount of previous contact with mental health services (e.g. received services two or more times without any noted clinical change), visual impairment, non-English speaking, history of overdoses/self-injury and current dependent substance use. Potential participants with a comorbid anxiety disorder were therefore excluded; this decision was informed by the recent systematic review of the CAT evidence base which highlighted the specific need to produce depression-specific clinical trials (Calvert & Kellett, 2014).

Research design

A single-blind randomised controlled deconstruction trial methodology was applied in this study (Bell, Marcus, & Goodlad, 2013). Participants were allocated to two treatment arms by a researcher using parallel-group 1:1 randomisation stratified by therapist, using Graphpad (2005). Suresh's (2011) review of online randomisation resources for clinical trials found that Graphpad produced unbiased randomisation sequences. Once randomized, trial therapists offered 8-session CAT treatments and also conducted an 8-week follow-up review. The treatment study arms were either (1) full-CAT or (2) CAT-NR. In total, 52 (54.7%) participants were randomised to full-CAT and 43 (45.3%) were randomised to CAT-NR. Individual randomisation sequences were created for each therapist to minimise therapist effects and participants were blind to their allocation.

Treatment model, fidelity and competence

Treatment in both arms consisted of eight (50-minute) CAT sessions following the three-phase theoretical structure of CAT; reformulation, recognition and revision (Ryle & Kerr, 2002).

Apart from the NR manipulation, both treatment arms were matched and identical in terms of the elements considered fundamental to CAT. That is, an assessment phase (including completion of the psychotherapy file; Ryle, 1990) leading to a sequential diagrammatic reformulation (SDR), associated patient self-monitoring of identified roles and patterns (recognition), and a final stage of change-focussed work (revision), culminating in the exchange of goodbye letters by therapist and client at the last session. Figure 2 describes the core similarities and key difference between the study arms. The arms were matched in terms of the 8-session treatment contract. Seven (five female and two male) clinical psychologists and clinical psychology trainees in their final year of training delivered CAT following a treatment manual (Stockton, 2012). None of the trial therapists were fully accredited CAT therapists. A weekly 2-hour supervision group was provided and all patients were reviewed each week to ensure treatment integrity. A Consultant Clinical Psychologist and accredited CAT psychotherapist, supervisor and trainer facilitated all supervision groups. Competency was assessed during supervision using the Competency in Cognitive Analytic Therapy measure (CCAT; Bennett & Parry, 2004) from session audiotapes to encourage competency.

Due to the focus of the research, treatment fidelity checks were performed on audiotapes of session 3 in both arms (i.e. the NR session in the full-CAT arm and the SDR session in the CAT-NR arm). The coder (a qualified CAT therapist) was blind to allocation. In 100% of cases in the full-CAT arm, at session 3 an NR was read to the participant, discussed within the session and the participant asked to take the reformulation home in order to re-read, reflect upon and add to in terms of any missing or misplaced content. In 100% of cases in the CAT-NR arm at session 3, the session included the construction of an SDR followed by the agreement of a between-session task which involved self-monitoring of the patterns and roles summarised by the SDR. Ten recordings of third sessions from the CAT-NR and full-CAT arm were randomly selected for double-rating by another CAT therapist; agreement was 100%.

Additionally, eight participants (8.4%) had one treatment session rated for CAT competence by the author of the CCAT measure. Overall, the delivery of CAT treatment was deemed satisfactory, with an average rating across therapists of 25 (range 18-40). The cut-off score for CAT treatment fidelity is 20 (Bennett & Parry, 2004). An RCT of CAT for personality disorder reported an average CCAT score of 22, with a range of 13-38 (Clarke, Thomas & James, 2013).

Outcome measures; description and timing

Attendance was recorded via accessing service records. Adverse effects were recorded in terms of emergency psychiatric team input and in-patient admissions. The following self-report measures were taken at screening, at each of the eight treatment sessions and again at 8-weeks follow-up. The Patient Health Questionnaire-9 (PHQ-9; Kroenke, Spitzer, & Williams, 2001), Generalised Anxiety Disorder-7 (GAD-7; Spitzer et al., 2006) and Work and Social Adjustment Scale (WSAS; Mundt, Marks, Shear & Greist, 2002) were taken before each session and the WAI-S (WAI-S; Tracey & Kokotovi, 1989) and the HAT (HAT; Llewelyn, 1988) were taken after the session.

Patient Health Questionnaire-9 (PHQ-9)

This 9-item measure is a validated case-finding tool for major depressive disorder (Kroenke, Spitzer, & Williams, 2001). Scores on the PHQ-9 range from 0-27, a score of ≥10 is used as a cut-off to identify cases with diagnosable depression symptoms (Wittkampf et al., 2007). The PHQ-9 is based on DSM criteria for symptoms of major depression (Lowe et al., 2004) and is widely recognised as a valid and reliable measure of depressive symptoms (Martin et al., 2006).

Generalized Anxiety Disorder-7 (GAD-7)

This 7-item measure screens for the presence of generalized anxiety disorder and other anxiety problems such as panic disorder and post-traumatic stress. Summed scores range between 0

and 21 with a recommended cut-off of ≥8 for the diagnosis of a clinically important anxiety disorder (Spitzer et al., 2006). The GAD-7 is recommended as a useful tool to assess and to monitor change in anxiety symptom severity (Gyani et al., 2013).

Work and Social Adjustment Scale (WSAS)

The WSAS measures the impact of mental health difficulties on aspects of daily functioning (Mundt, Marks, Shear & Greist, 2002). Scores on the WSAS range from 0 to 40, a higher score on the scale indicates a higher level of impaired functioning (Mataix-Cols et al., 2005). The WSAS has been previously used as a measure of functional impairment in Primary Care (Gyani et al., 2013).

Working Alliance Inventory-Short (WAI-S)

This measure assesses the strength of the therapeutic alliance across three subscales: agreement on therapy tasks, goal agreement and quality of the therapeutic bond (Tracey & Kokotovi, 1989). The WAI-S has been found to have a good internal consistency (Smits et al, 2014) and predictive validity (Busseri & Tyler, 2003).

Helpful Aspects of Therapy (HAT)

This measure gathers quantitative and qualitative information on the patient's perceptions of the helpfulness during treatment sessions (Llewelyn, 1988). The HAT asks participants to describe both the most and least helpful aspects of the therapy session and then to rate the helpfulness/ unhelpfulness of the session (Llewelyn, 1988). Participants rate how helpful or hindering the session is on a 1 to 9 Likert Scale with 1 being extremely hindering and 9 being extremely helpful.

Analysis strategy

Primary outcome definition and sample size calculation

The primary outcome was the PHQ-9 depression score at 8-weeks follow-up. Assuming a 'medium' effect size of f=.25, a significance level of alpha=.05 with two study arms providing data at three time points (NR, termination and follow-up) a total sample size of N=24 per group enabled 80% power to test for differences between full-CAT and NR-CAT. The sample of N = 95 therefore was adequately powered for the primary analysis.

Primary analysis

Analyses were conducted using SPSS v.24. Longitudinal multilevel modelling (MLM) was used to examine changes in depression (PHQ-9) symptoms over time, and to compare the rate of change (growth trends) between treatment groups at 8-weeks follow-up. A two-level model was applied; where session-by-session PHQ-9 scores (level 1) were nested within each case (level 2), including random intercepts and random slopes at level 2. This method appropriately modelled the repeated (and therefore autoregressive) outcome measurements, and had the additional advantage of being able to model growth trends over missing data-points, thus enabling best use of all available data following intention-to-treat principles. Following standard MLM guidelines (Singer & Willett, 2003), an unconditional (no predictors) model was initially examined to determine the variance explained at each level. We then examined alternative trends used to model changes over time (e.g., linear, quadratic, cubic and log-linear trends), and selected the best fitting model using −2 log likelihood ratio tests. After initial model checking, the main analysis applied a two-level linear growth model, controlling for pre-treatment PHQ-9 severity (mean-centred), entering the following predictors: group (CAT-NR vs. full-CAT), and a group * time interaction term. This interaction term was the main hypothesis test, since it compared linear changes over time between treatment groups. The

main MLM analysis was repeated in a sensitivity analysis additionally controlling for treatment sessions (mean-centred), to assess the robustness of the results, in case there were any systematic differences in treatment duration across groups.

Secondary analyses

The main MLM analysis was repeated including all data points up to session 4 (the session after the NR component was delivered), and then up to the end of the acute phase treatment (session 8). These MLM analyses were repeated to examine between-group differences in secondary outcome measures (GAD-7, WSAS). MLM was also used to compare linear growth trends in WAI-S and HAT scores between groups, but taking session 1 scores as covariates (since these measures were not completed pre-treatment) and examining changes after session 3 (since this is when NR took place, and these measures were completed at the end of the session). To assess treatment acceptability, three subsamples were created and compared between groups; (a) those who attended all sessions (termed 'full treatment completers'), (b) those who had attended 4-7 sessions (termed 'partial treatment completers') and (c) those who had attended less than 4 sessions ('dropouts'). Means, paired sample t-tests and associated effect sizes pre-post and screening-to-follow-up PHQ-9 outcomes were also calculated. For all effect size calculations, the Cohen (1988) guidance was used of $d_+ > 0.20$ being a small effect, $d_+ > 0.50$ being a medium effect and $d_+ > 0.80$ being a large effect.

In order to compare individual treatment response rates in the arms, five depression outcome categories were created on pre-post and screening-to-follow-up PHQ-9 comparisons. The reliable change index (RCI; Jacobson & Traux, 1991) assessed whether change on PHQ-9 scores were beyond chance variation or regression to the mean. Outcome categories were 'recovered' (reliable improvement plus change from a clinical case to a non-clinical case), 'improved' (reliable improvement), 'positive clinical change' (change from a clinical case to a

non-clinical case), 'stasis' (a non-reliable change in either direction) or 'deteriorated' (reliable deterioration). McNemar tests examined differences between the outcome rates achieved by CAT-NR versus full-CAT, and were only employed if cell size was greater than 10 in each treatment arm sub-sample. Treatment response was not analysed on an intention-to-treat basis, as patients who had dropped out prior to session 1 (n=5) and those lost to follow-up (n=28) were excluded from this analysis.

Results

Screening, suitability and treatment acceptability

Pathways through treatment for the full-CAT and CAT-NR arms are summarised in the CONSORT diagram (see Figure 2). Reasons for exclusion (n=30; 24%) were mainly due failure to meet diagnostic threshold for depression (n=10) and/or having a co-morbid anxiety disorder which was the primary reason for referral (n=10). Ninety-five patients (76%) were randomised into either full-CAT (n=52; 54.8%) or CAT-NR (n=43; 45.2%). In the full-CAT arm 34% were unemployed or in receipt of a health related benefit in comparison to 32% in the CAT-NR arm. In terms of attendance, in the full-CAT arm, 35 participants (67.3%) were full treatment completers, 7 (13.5%) partially completed treatment and 10 (19.2%) dropped-out. Fifteen participants (28.9%) in the full-CAT treatment arm were lost to follow-up, with n=37 (71.1%) completing the follow-up assessment. In the CAT-NR arm, 27 patients (62.8%) fully completed treatment, 10 (23.3%) partially completed treatment and 5 (13.9%) dropped-out. In the CAT-NR arm, 13 (30.2%) patients were lost to follow-up, with n=30 (69.8%) completing follow-up. At follow-up, 3 patients (n=2 from the full-CAT arm and n=1 from CAT-NR arm) were referred onto further therapy (referred to CBT in each case). No participants in either arm needed emergency psychiatric team input or in-patient admission during their contact time within the trial. The study arms did not significantly differ in terms

of dropout (χ^2 (1, N=95)=.468, p= .494), treatment completion (χ^2 (1, N=95)= .468, p= .494), lost to follow-up (χ^2 (1, N=95)=.022, p= .883) or completed follow-up rates (χ^2 (1, N=95)=.022, p=.883).

Randomisation checks and demographic characteristics

Randomisation checks assessed for any potential demographic or symptomatic differences prior to treatment (see Table 1). There were no age differences between the full-CAT (M=43.3, SD=11.8) and CAT-NR arms (M=39.3, SD=11.1). Gender was equally distributed across full-CAT (24.5% male; 75.5% female) and CAT-NR (22.0% male; 78.0% female). Depression (PHQ-9) was equally moderately-severe at initial screening in the full-CAT (M=15.7, SD=5.1) and CAT-NR (M=14.3, SD=5.4) arms. There were no significant differences between the arms at screening in terms of anxiety, functioning, or initial session differences in session helpfulness or the working alliance ratings.

Efficacy of narrative reformulation

The MLM analysis results are summarised in Table 2 (full regression outputs are available in a supplementary appendix). There PHQ-9 model group * time interaction term was not statistically significant at treatment session 4 (B = 0.17, SE = 0.35, p = .63), at post-treatment (B = −0.07, SE = 0.20, p = .74) or at 8-weeks follow-up (B = −0.07, SE = 0.17, p = .67). The same results were obtained in a sensitivity analysis controlling for treatment duration (interaction term: B = −0.07, SE = 0.17, p = .67, not shown in table). Similarly, no significant effects were observed for any of the secondary outcomes at any of the measurement points.

The effect of CAT for depression in full-CAT and CAT-NR

Table 3 reports the means, within-group effect sizes and t-test results for full-CAT and CAT-NR on pre-post and screening-to-follow-up comparisons. Both treatment arms produced large

effect sizes (d_+>1.5; Cohen, 1988); highly significant differences on pre-post and screening-to-follow-up comparisons were apparent.

Individual outcomes in full-CAT and CAT-NR

Individual depression outcome category rates for full-CAT and CAT-NR are reported in table 4. The recovery rate (i.e. a reliable and clinically significant improvement in depression) did not significantly differ between full-CAT (34.60% at termination and 38.50% at follow-up) and CAT-NR (44.20% at termination and 34.90% at follow-up). At the end of treatment, significantly more participants in the CAT-NR arm had experienced a clinically significant reduction in depression (scores below diagnostic cut-off) and at follow-up significantly more participants in the full-CAT arm had a stasis outcome (no statistically reliable change).

Discussion

This research was novel in attempting to isolate and examine the clinical efficacy of an assumed active ingredient of CAT (i.e. narrative reformulation) during the treatment of depression in Primary Care. The study used a dismantling methodology in a suitably powered trial to achieve its aims, being a relatively rare example of dismantling studies in the field of depression treatment. A particular strength of the study was that the adherence checks confirmed that the experimental component (NR procedure) was indeed present in the intervention arm and then absent in the control arm. Participants were moderately to severely depressed at intake according to the PHQ-9 categorisations and randomisation was effective at yielding balanced and equivalent groups prior to the intervention. Therefore, subsequent comparisons of the two study groups were based on solid methodology, rationale and associated evidence of treatment adherence.

No significant differences in depression outcomes were found between groups at any stage of the therapy. This would strongly indicate that NR is not an active ingredient of CAT treatment for depression. This suggests clinically that the efficiency of CAT could be improved (when working with depression) by removing the NR component and moving straight onto co-constructing the sequential diagrammatic reformulation (SDR) after the initial assessment. The current findings also support the available meta-analytic evidence that dismantling trials are typically unlikely to yield any significant differences between study arms (Bell et al., 2013). Both treatment arms produced significant reductions in depression, and so fill an identified gap in the CAT evidence base (Calvert & Kellett, 2014). There was no evidence that CAT with or without NR increased the likelihood of clinical deterioration; and the majority of patients recovered, improved or had a clinical positive change. On secondary outcomes, no significant differences were found between the two treatment arms following NR, at treatment termination or at follow-up in terms of anxiety outcomes, functional impairment, the therapeutic alliance or ratings of the helpfulness of therapy. These results present a real challenge to CAT orthodoxy of NR being both therapeutic in itself and also strengthening the working alliance (Ryle & Kerr, 2002). The results are consistent with other empirical evidence that NR has little impact on perceived helpfulness or the working alliance (Evans & Parry, 1996; Shine & Westacott, 2010). An improvement on any replication efforts would be the addition of a measure of readiness to change (such as the URICA; Bergly, Stallvik, Nordahl & Hagen, 2014) to assess motivation as a potential moderator of NR during CAT.

Sub-group analyses of the outcome data revealed some intriguing differences between the arms, albeit in smaller samples, so the results need to be interpreted with some caution. Participants who did not receive a narrative reformulation were more likely to have experienced a clinically significant reduction in depression by the end of treatment. This effect may have been due to the CAT-NR group immediately progressing onto mapping the sequential

diagrammatic reformulation (SDR) following assessment and therefore entering the recognition and then revision stages (i.e. the active change work) of CAT earlier (Ryle & Kerr, 2002). However, this difference was not maintained at follow-up, with a significantly greater proportion of CAT-NR participants remaining in stasis in terms of depression outcome. This alternatively suggests that NR may serve some protective function regarding the long-term management of depression, which is important considering the often high rates of relapse observed (Steinert, Hofmann, Kruse, & Leichsenring, 2014). Nevertheless, these subgroup analyses are limited by missing data (at 8-week follow-up), whereas the primary MLM analyses are better suited to cope with missing data and confirmed that there were no significant differences in depression changes over time. In terms of safety, then no participants in either arm were admitted to an in-patient unit due to a deterioration in their depression or were referred to the psychiatric emergency team. Similarly, on the primary outcome measure, there were no reliable deteriorations at end of treatment or at follow-up in either of the arms of the trial. These results would imply that brief 8-session CAT can be safely delivered in Primary Care.

This study suggests that this brief 8-session version of CAT is also 'long enough' to be effective at treating depression, despite being considerably shorter than typical 16 or 24-session CAT treatment protocols (Ryle & Kerr, 2002). The CAT literature has previously lacked sufficient evidence to support the application of this brief version of the model (Calvert & Kellett, 2014). Effective but brief (efficient) psychological interventions are in demand in public services due to the need for ensuring and maintaining high patient throughput in pressured clinical services (Clark, 2009). The large within-group effect sizes observed in this study (d_+>1.5) were larger than the Ryle et al. (2014) meta-analytic mean effect size (d_+ 0.83) for CAT. This is probably due to the Ryle et al. (2014) meta-analysis being mainly based on personality disorder outcomes and therefore based in 'harder to help' populations. For

moderate-to-severe depression, 16 sessions of high intensity CBT is recommended for IAPT services (NICE, 2009), but the acceptability of CBT for some patients can be poor (Milosevic, Levy, Alcolado & Radomsky, 2015). Despite the brevity of the CAT intervention, the recovery rates in this trial (34.9% to 44.2%) are comparable to rates observed across the controlled trials of CBT for depression (e.g. 49%, 95% CI: 0.32–0.66; Cuijpers et al., 2014). It is, however, possible that this recovery rate may be influenced by spontaneous remission in some cases. Spontaneous remission from major depression has been previously documented in meta-analyses of data from control group participants (e.g., 20% reported by Posternak and Miller, 2001). The development of a brief 8-session CAT model for depression potentially increases service plurality and patient choice in IAPT services in the UK. Given the acknowledged similarities between CAT and CBT, then CBT and schema therapy for depression have been compared to show that equivalent outcomes were achieved by both therapies, and in this trial both treatments delivered weekly sessions for 6 months, followed by monthly therapy sessions for 6 months (Carter et al. 2013). Furthermore, the dropout rate for both treatment arms combined was low (i.e. 16.8%) and treatment completion rates were high (i.e. 67.3% in full-CAT and 62.8% in CAT-NR attended all eight sessions). This mirrors the CAT evidence base of consistently high treatment retention rates (Calvert & Kellett, 2014), and also compares favourably with the dropout rate for CBT outcome studies (26.2%; Fernandez, Salem, Swift, & Ramtahal, 2015).

NR is acknowledged as a very time consuming process in the context of pressured clinical services (Denman, 2001). Despite NR taking up just one clinical session, this obscures the amount of preparation time and supervision required. Without strong evidence for superiority of NR, the significant amount of time spent on NR may not be necessary to improve treatment outcomes when treating depression. Previous studies have suggested that NR may be an important treatment component when working with patients with highly complex

presentations (Kellett, 2005; Kellett & Hardy, 2014; Kellett et al. 2016; Curling et al. 2017; Curling et al. 2018). However, previous studies have not applied a dismantling method, and therefore such claims are currently founded on weak evidence and in small samples. It should also be noted that despite the trial being registered in a public database, registration was retrospective and this may be seen as a limitation. The authors acknowledge that such retrospective registration does offer a strong assurance against biases such as selective outcome reporting. In this regard, we note that the study findings are indeed contrary to our initial hypotheses, which were specified a-priori and subjected to scientific review ahead of conducting the trial, as part of the ethics approval process. The study protocol, thus did go through a scientific review process and the planned analyses reported here were planned and recorded publicly within the NHS approval process. The original protocol did not plan growth curve modelling (which is the main analysis presented in this paper, as recommended by peer reviewers), and that addition to the results replaces the ANCOVA that was originally planned.

A number of limitations should be considered when interpreting the present results. Previous authors have raised some concerns about the PHQ-9 measure in terms of lacking sensitivity to change and being somewhat brief and limited in terms of depth and breadth of assessment of depressive symptoms (Manea, Gilbody & McMillan, 2015). The trial was conducted in an IAPT service and such services use the PHQ-9 as one of two primary outcome measures for depressed patients (Clark, 2009), and so it was decided to retain the measure in order to increase the external validity of the trial. IAPT services use a pre-specified and nationally consistent battery of outcomes measures (i.e. the PHQ-9, GAD-7 and the WSAS) and therefore despite potential participants with comorbid anxiety disorders being excluded, the GAD-7 was used a standard aspect of IAPT outcome monitoring. The fact that the primary outcome measure (and the majority of secondary outcomes) were based on self-report and the follow-up period was short are acknowledged as study weaknesses. The battery of measures in

the trial would have benefited from the addition of a measure of symptom insight (Lincoln, Lüllmann & Rief, 2007). Recommendations suggest that depression trials should measure treatment outcomes independently and also 18-months post treatment completion (NICE, 2009). Extensive trial selection criteria were also employed, which reduces the generalisability of the findings (Rothwell, 2006) and particularly the decision to exclude participants with a comorbid anxiety disorder. It is acknowledged that in Primary Care setting that depression and anxiety often occur as a comorbid condition (Das-Munshi et al. 2008), but as highlighted in the method, the decision to exclude was based on the attempt to improve the focal evidence base in relation to CAT treatment of depression. The study did not exclude potential participants with psychotic, bipolar or psychotic presentations, because the context of the service in which the trial was conducted (i.e. an IAPT service) is commissioned to deliver evidenced-based psychological therapies for anxiety and depression (Clark, 2009). Therefore, potential participants with these forms of comorbidity would be treated in another part of the service (i.e. Secondary Care). Finally, therapists in the study were not fully accredited CAT practitioners, although they were trained and closely supervised by an expert and fully accredited CAT practitioner. Controlling for personality pathology would have been a useful addition to the method, because there is evidence in IAPT services that personality difficulties supress outcomes and recovery rates (Goddard, Wingrove & Moran, 2015). Regardless of training, supervision support and evidence presented regarding model adherence and fidelity, and it is unclear whether therapists' training status may have affected the results (Roth, Pilling & Turner, 2010). Whilst dropout rates were low in each arm, the reasons for dropout were not collected and this would be a valuable addition to future studies.

To conclude, brief CAT for depression appears to be a safe and effective treatment option in IAPT and Primary Care services. This sufficiently powered and controlled study has revealed little evidence that NR specifically enhanced depression outcomes, the working

alliance or perceived helpfulness of the therapy. Findings have major clinical and service implications for the delivery of brief CAT and the role and function NR in the treatment of depression. It would appear appropriate to now progress onto a head-to-head trial of CBT versus CAT for depression in IAPT services. This would be particularly useful if the Goddard et al. (2015) evidence could be used to select patients with personality issues accessing IAPT services. This research also serves as a prompt for future research focal to deconstructing the efficacy of case formulation during other psychotherapies. There may be a tendency to 'overvalue' formulation which, despite clinician approval, is an activity that is not particularly well-grounded in scientific evidence (Bieling & Kuyken, 2003). Overall, these findings suggest that a brief 8-session CAT intervention can be effective for treating depression and future controlled research is indicated regarding the role, impact and function of case formulation.

Conflicts of interest: None declared. The Association of Cognitive Analytic Therapists (ACAT) provided the funding for the CCAT analysis.

Data availability: The data is available for other researchers to use in collaboration with the research team. Data access should be requested by contacting the corresponding author.

References

Ahn, H., & Wampold, B. E. (2001). Where oh where are the specific ingredients? A meta-analysis of component studies in counselling and psychotherapy. Journal of Counselling Psychology, 48, 251-257. doi: 10.1037//0022-0167.48.3.251

American Psychiatric Association (1994). Diagnostic and Statistical Manual of Mental Disorder (4th edition). Washington DC: American Psychiatric Association.

Bell, E. C., Marcus, D. K., & Goodlad, J. K. (2013). Are the parts as good as the whole? A meta-analysis of component treatment studies. Journal of Consulting and Clinical Psychology, 81, 722-736. doi: 10.1037/a0033004

Bennett, D. (1994). Readiness to change the impact of reformulation: a case example of cognitive analytic therapy. International Journal of Short-Term Psychotherapy, 9, 83-91.

Bennett, D., & Parry, G. (2004). A measure of psychotherapeutic competence derived from cognitive analytic therapy. Psychotherapy Research, 14, 176-192. doi: 10.1093/ptr/kph016

Bennett, D., & Parry, G. (1998). The accuracy of reformulation in cognitive analytic therapy: A validation study. Psychotherapy Research, 8, 84-103. doi: 10.1080/10503309812331332217

Bergly, T.H., Stallvik, M., Nordahl, H., & Hagen, R. (2014). The predictive validity of the URICA in a sample of patients in substance use treatment. Addictive Disorders & Their Treatment, 13, 170–178. doi: 10.1097/ADT.0000000000000041

Bieling, P.J., & Kuyken, W. (2003). Is cognitive case formulation science fact or science fiction? Clinical Psychology Science and Practice, 10, 52–69. doi: 10.1093/clipsy.10.1.52

Borkovec, T. D., & Sibrava, N. J. (2005). Problems with the use of placebo conditions in psychotherapy research, suggested alternatives, and some strategies for the pursuit of the placebo phenomenon. Journal of Clinical Psychology, 61, 805-818. doi: 10.1002/jclp.20127

Busseri, M. A., & Tyler, J. D. (2003). Interchangeability of the Working Alliance Inventory and Working Alliance Inventory, Short Form. Psychological Assessment, 15, 193-197. doi: 10.1037/1040-3590.15.2.193

Calvert, R., & Kellett, S. (2014). Cognitive analytic therapy. A review of the outcome evidence base for treatment. Psychology and Psychotherapy: Theory, Research and Practise, 87, 253-277. doi: 10.1111/papt.12020

Carter, J.D., McIntosh, V.V., Jordan, J., Porter, R.J., Frampton, C.M., & Joyce, P.R. (2013). Psychotherapy for depression: a randomized clinical trial comparing schema therapy and cognitive behavior therapy. Journal of Affective Disorders, 151, 500-505. doi: 10.1016/j.jad.2013.06.034.

Carrico, A.W., & Antoni, M.H. (2008). Effects of psychological interventions on neuroendocrine hormone regulation and immune status in HIV-positive persons: a review of randomized controlled trials. Psychosomatic Medicine, 70, 575-584. doi: 10.1097/PSY.0b013e31817a5d30

Clark, D.M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. International Review of Psychiatry, 23, 318-327. doi: 10.3109/09540261.2011.606803

Clarke, S., Thomas, P., & James, K. (2013). Cognitive analytic therapy for personality disorder: randomised controlled trial. British Journal of Psychiatry, 202, 129-134. doi: 10.1192/bjp.bp.112.108670

Cohen, J. (1988). Statistical power analysis for the behavioural sciences (2nd edition). United States of America: Lawrence Erlbaum Associates.

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. Journal of Consulting and Clinical Psychology, 76, 909-922. doi: 10.1037/a0013075

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. Canadian Journal of Psychiatry, 58, 376-385. doi: 10.1177/070674371305800702

Cuijpers, P., Karyotaki, E., Weitz, E., Andersson, G., Hollon, S. D., & van Straten, A. (2014). The effects of psychotherapies for major depression in adults on remission, recovery and improvement: a meta-analysis. Journal of Affective Disorders, 159, 118-126. doi: 10.1016/j.jad.2014.02.026

Curling, L., Kellett, S., Totterdell, P., Parry, G., Hardy, G., & Berry, K. (2017). Treatment of obsessive morbid jealousy with cognitive analytic therapy: An adjudicated hermeneutic single-case efficacy design evaluation. Psychology and Psychotherapy, Theory, Research and Practice. doi: 10.1111/papt.12151.

Curling, L., Kellett, S., & Totterdell, P. (2018). Cognitive analytic therapy for obsessive morbid jealousy; a case series. Journal of Psychotherapy Integration.

Czaja, S. J., Schultz, R., Lee, C. C., & Belle, S. H. (2003). A methodology for describing and decomposing complex psychosocial and behavioural interventions. Psychology and Aging, 18, 385-395. doi: 10.1037/0882-7974.18.3.385

Das-Munshi, J., Goldberg, D., Bebbington, P.E., Bhugra, D.K., Brugha, T.S., Dewey, M.E., Jenkins, R., Stewart, R., & Prince, M. (2008). Public health significance of mixed anxiety and depression: beyond current classification. British Journal of Psychiatry, 192, 171–177. doi: 10.1192/bjp.bp.107.036707

Denman, C. (2001). Cognitive-analytic therapy. Advances in Psychiatric Treatment,7, 243-252. doi: 10.1192/apt.7.4.243

Evans, J., & Parry, G. (1996). The impact of reformulation in cognitive analytic therapy with difficult-to-help clients. Clinical Psychology and Psychotherapy, 3, 109-117. doi: 10.1002/(SICI)1099-0879(199606)

Fernandez, E., Salem, D., Swift, J. K., & Ramtahal, N. (2015). Meta-analysis of dropout from cognitive behavioral therapy: Magnitude, timing, and moderators. Journal of Consulting and Clinical Psychology, 83, 1108. doi: 10.1037/ccp0000044

Firth, N., Barkham, M., Kellett, S., & Saxon, D. (2015). Therapist effects and moderators of effectiveness and efficiency in psychological wellbeing practitioners: A multilevel modelling analysis. Behaviour Research and Therapy, 69, 54-62. doi: 10.1016/j.brat.2015.04.001

Goddard, E., Wingrove, J., & Moran, P. (2015). The impact of comorbid personality difficulties on response to IAPT treatment for depression and anxiety. Behaviour Research and Therapy, 73, 1-7 doi.org/10.1016/j.brat.2015.07.006

Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: lessons from year one of IAPT. Behaviour Research and Therapy, 51, 597-606. doi: 10.1016/j.brat.2013.06.004

Hamill, M., Reid, M., & Reynolds, S. (2008). Letters in cognitive analytic therapy: The patient's experience. Psychotherapy Research, 18, 573–583. doi 10.1080/10503300802074505

Jacobson, N. S., Dobson, K. S., Traux, P. A., Addis, M. E., Koerner, K., Gollan, J. K., Gortner, E., & Prince, S. E. (1996). A component analysis of cognitive-behavioural treatment for depression. Journal of Consulting and Clinical Psychology, 64, 295-304.

Jacobson, N.S., & Traux, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology, 59, 12-19. doi: org/10.1037/0022-006X.59.1.12

Kellett, S. (2005). The treatment of dissociative identity disorder with cognitive analytic therapy: Experimental evidence of sudden gains. Journal of Trauma and Dissociation, 6, 55-81. doi: 10.1300/J229v06n03_03

Kellett, S., & Hardy, G. (2014). Treatment of paranoid personality disorder with cognitive analytic therapy: a mixed methods single case experimental design. Clinical Psychology and Psychotherapy, 21, 452-464. doi: 10.1002/cpp.1845

Kellett, S., Simmonds-Buckley, M., & Totterdell, P. (2016). Testing the effectiveness of cognitive analytic therapy for Hypersexuality Disorder: an intensive time-series evaluation. Journal of Sex and Marital Therapy. doi: 10.1080/0092623X.2016.1208129

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. Journal of General Internal Medicine, 16, 606–613. doi: 10.1046/j.1525-1497.2001.016009606.x

Lincoln, T.M., Lüllmann, E., & Rief, W. (2007). Correlates and long-term consequences of poor insight in patients with schizophrenia. A systematic review. Schizophrenia Bulletin, 33, 1324–1342. doi: 10.1093/schbul/sbm002

Llewelyn, S. P. (1988). Psychological therapy as viewed by clients and therapists. British Journal of Clinical Psychology, 27, 223-237. doi: 10.1111/j.2044-8260.1988.tb00779.x

Lowe, B., Kroenke, K., Herzog, W., & Grafe, K. (2004). Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). Journal of Affective Disorders, 81, 61-66. doi: 10.1016/S0165-0327(03)00198-8

Manea, L., Gilbody, S., & McMillan, D. (2015). A diagnostic meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) algorithm scoring method as a screen for depression. General Hospital Psychiatry, 37, 67-75. doi: 10.1016/j.genhosppsych.2014.09.009

Martin, A., Winfried, R., Klaiberg, A., & Braehler, E. (2006). Validity of the brief Patient Health Questionnaire mood scale (PHQ-9) in general population. General Hospital Psychiatry, 28, 71-77. doi:10.1016/j.genhosppsych.2005.07.003

Mataix-Cols, D., Cowley, A. J., Hankins, M., Schneider, A., Bachofen, M., Kenwright, M., Gega, L., Cameron, R., & Marks, I. M. (2005). Reliability and validity of the Work and Social Adjustment Scale in phobic disorders. Comprehensive Psychiatry, 46, 223-228. doi: 10.1016/j.comppsych.2004.08.007

Mazzucchelli, T., Kane, R., & Rees, C. (2009). Behavioural activation treatments for depression in adults: a meta-analysis and review. Clinical Psychology: Science and Practise, 16, 383-411. doi: 10.1111/j.1468-2850.2009.01178.x

Milosevic, I., Levy, H.C., Alcolado, G.M., & Radomsky, A.S. (2015). The treatment acceptability/adherence scale: moving beyond the assessment of treatment effectiveness. Cognitive Behaviour Therapist, 44, 456-469. doi: 10.1080/16506073.2015.1053407

Mundt, J. C., Marks, I. M., Shear, K., & Greist, J. H. (2002). The Work and Social Adjustment Scale: a simple measure of impairment in functioning. British Journal of Psychiatry, 180, 461-464. doi: 10.1192/bjp.180.5.461

NICE Clinical Guideline 90. (2009). Depression in Adults. The treatment and management of depression in adults (update). London, UK: National Institute for Health and Clinical Excellence.

Posternak, M. A., & Miller, I. (2001). Untreated short-term course of major depression: a meta-analysis of outcomes from studies using wait-list control groups. Journal of Affective Disorders, 66, 139-146. doi: 10.1016/S0165-0327(00)00304-9

Rayner, K., Thompson, A. R., & Walsh, S. (2011). Clients' experience of the process of change in cognitive analytic therapy. Psychology and Psychotherapy: Theory, Research and Practise, 84, 299-313. doi: 10.1348/147608310X531164

Roth, A. D., Pilling, S., & Turner, J. (2010). Therapist training and supervision in clinical trials: implications for clinical practice. Behavioural and Cognitive Psychotherapy, 38, 291-302. doi: 10.1017/S1352465810000068

Rothwell, P.M. (2006) Factors that can affect the external validity of randomized controlled trials. Trials; 1(1): e9. doi: 10.1371/journal.pctr.0010009

Ryle, A. (1990). Cognitive analytic therapy: active participation in change. Chichester: Wiley.

Ryle, A. (1991). Object relations theory and activity theory: A proposed link by way of the procedural sequence model. British Journal of Medical Psychology, 64, 307-316.

Ryle, A. (1997) The structure and development of borderline personality disorder: a proposed model. British Journal of Psychiatry, 170, 82– 87.

Ryle, A., & Kerr, I.B. (2002). Introducing cognitive analytic therapy. Principles and practise. Chichester: Wiley.

Ryle, A., Kellett, S., Hepple, J., & Calvert, R. (2014). Cognitive analytic therapy at 30. Advances in Psychiatric Treatment, 20, 258-268. doi: 10.1192/apt.bp.113.011817

Shine, L., & Westacott, M. (2010). Reformulation in cognitive analytic therapy: Effects on the working alliance and the client's perspective on change. Psychology and Psychotherapy: Theory, Research and Practise, 83, 161-177. doi: 10.1348/147608309X471334

Singer, J., & Willett, J. (2003). Applied longitudinal data analysis: Modeling change and event occurrence. New York: Oxford University.

Spitzer, R. L., Kroenke, K., & Williams, J. B. W. (1999). Validation and utility of the self-report version of PRIME-MD: The PHQ primary care study. The Journal of the American Medical Association, 282, 1737-1744. doi:10.1001/jama.282.18.1737

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. Archives of Internal Medicine, 166, 1092-1097. doi: 10.1001/archinte.166.10.1092

Smits, D., Luyckx, K., Smits, D., Stinckens, N., & Claes, L. (2015). Structural characteristics and external correlates of the Working Alliance Inventory-Short Form. Psychological Assessment, 27, 545-551. doi: 10.1037/pas0000066

Steinert, C., Hofmann, M., Kruse, J., & Leichsenring, F. (2014). Relapse rates after psychotherapy for depression – stable long-term effects? A meta-analysis. Journal of Affective Disorders, 168, 107-118. doi: 10.1016/j.jad.2014.06.043

Stockton, C. (2012). Treatment manuals for depression dismantling study. University of Sheffield, UK.

Suresh, K.P. (2011). An overview of randomization techniques: An unbiased assessment of outcome in clinical research. Journal of Human Reproductive Science, 4, 8-11. doi: 10.4103/0974-1208.82352.

Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1, 207-210. doi: 10.1037/1040-3590.1.3.207

University College London (2015). National Curriculum for the Education of Psychological Wellbeing Practitioners. 3rd edition.

Wittkampf, K. A., Naeije, L., Schene, A. H., Huyser, J., & van Weert, H. C. (2007). Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. General Hospital Psychiatry, 29, 388-395. doi: 10.1016/j.genhosppsych.2007.06.004

Young, J.E., Klosko, J.S., & Weishaar, M. (2003). Schema Therapy: A Practitioner's Guide. Guilford Publications: New York.

Figure 1

CONSORT Diagram

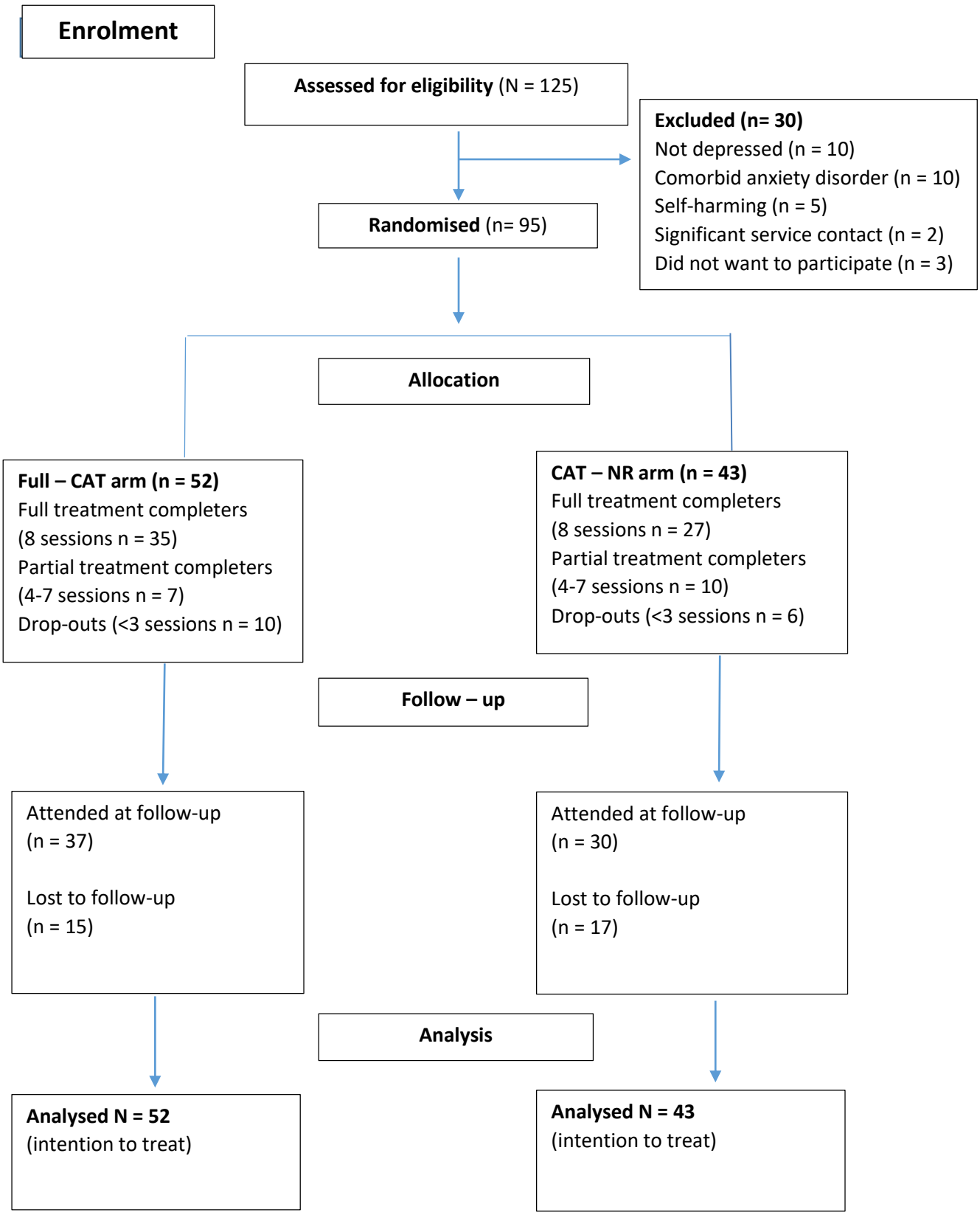

Figure 2

Description of treatment pathway in full-CAT and CAT-NR treatment arms.

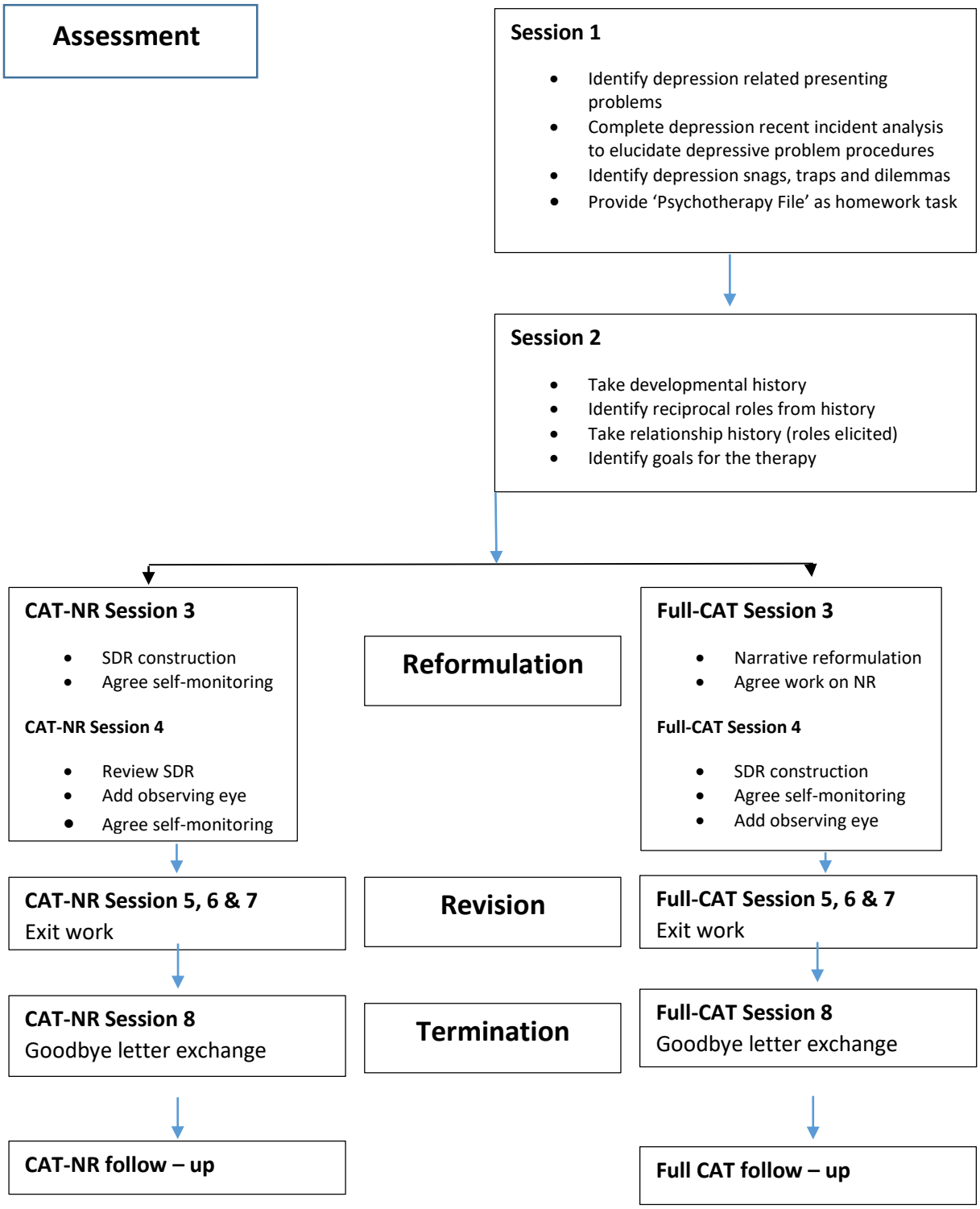

Table 1.
Pre-treatment full CAT and CAT-NR demographic and clinical characteristics

Characteristic	Full CAT (n = 52)	CAT-NR (n = 43)	t-test[a]	p
Age in years, mean (SD)	43.3 (11.8)	39.3 (11.1)	$t(88) = 1.643$	.104
Female (%)	75.5	78.0	χ^2 (1, N = 90) = .080	.777
Measures				
PHQ-9 mean (SD)	15.7 (5.1)	14.3 (5.3)	$t(91) = 1.323$	.189
GAD-7 mean (SD)	13.0 (6.1)	11.2 (5.6)	$t(91) = 1.508$	.135
WSAS mean (SD)	19.4 (8.6)[c]	17.7 (6.9)[d]	$t(88) = 1.046$	.299
WAI-S mean (SD)[b]	63.0 (15.0)[e]	64.0 (12.5)[f]	$t(81) = -.345$	.731
HAT mean (SD)[b]	7.5 (1.8)[g]	6.9 (2.2)[h]	$t(47) = .919$	.363

[a]Chi-square used to assess for significant gender differences between the treatment arms; [b] process measures necessarily taken at first therapy session; [c]N=49; [d]N=41; [e]N=43; [f]N=40; [g]N=26; [h]N=23. PHQ-9 = Patient Health Questionnaire-9; GAD-7 = Generalised Anxiety Disorder-7; WSAS = Work and Social Adjustment Scale; WAI-S = Working Alliance Inventory-Short Form; HAT = Helpful Aspects of Therapy

Table 2. Means, effect sizes, and multilevel modelling results for full CAT and CAT-NR on primary and secondary outcomes (N = 95)

	Within-group outcomes and analyses				Between-group analyses (Full CAT vs CAT-NR)		
Sample and measures	Full-CAT (n = 52) Mean (SD)	ES	CAT-NR (n = 43) Mean (SD)	ES	ES	Fixed effects for group * time B, SE [95% CI]	p
PHQ-9							
Session 4	12.10 (6.11)	0.71	10.47 (6.27)	0.70	0.26	0.17, 0.35 [-0.53, 0.87]	0.63
End of treatment	8.85 (6.62)	1.35	7.23 (5.90)	1.30	0.26	-0.07, 0.20 [-0.47, 0.33]	0.74
Follow-up	8.80 (7.52)	1.36	7.47 (5.85)	1.26	0.20	-0.07, 0.17 [-0.41, 0.26]	0.67
GAD-7							
Session 4	10.50 (5.31)	0.41	8.07 (5.79)	0.56	0.44	0.14, 0.33 [-0.51, 0.79]	0.66
End of treatment	7.72 (5.56)[b]	0.87	5.65 (4.89)	1.00	0.40	-0.04, 0.18 [-0.39, 0.31]	0.81
Follow-up	7.97 (6.46)	0.83	5.93 (5.11)	0.95	0.35	-0.05, 0.15 [-0.35, 0.26]	0.77
WSAS							
Session 4	19.02 (16.23)	0.05	16.07 (7.94)	0.23	0.23	1.76, 0.91 [-0.04, 3.57]	0.06
End of treatment	13.50 (9.47)	0.68	11.51 (8.36)	0.89	0.22	0.42, 0.33 [-0.23, 1.07]	0.20
Follow-up	13.35 (10.40)	0.70	12.24 (7.92)	0.78	0.12	0.26, 0.30 [-0.35, 0.86]	0.40
WAI-S							
Session 3	66.45 (13.24)	0.23	70.10(10.74)	0.49	0.30	-1.43, 1.13 [-3.68, 0.82]	0.21
End of treatment	72.60 (9.50)	0.64	73.76 (8.18)	0.78	0.13	0.02, 0.41 [-0.79, 0.82]	0.97
Follow-up	72.57 (9.62)	0.64	74.27 (8.24)	0.82	0.19	0.02, 0.41 [-0.79, 0.82]	0.97
HAT							
Session 3	7.82 (1.48)	0.18	7.89 (1.38)	0.45	0.05	-0.77, 0.51 [-1.78, 0.24]	0.14
End of treatment	8.08 (1.32)	0.32	7.54 (2.42)	0.29	0.28	-0.03, 0.14 [-0.32, 0.26]	0.82
Follow-up	8.08 (1.35)	0.32	7.57 (2.44)	0.30	0.26	-0.03, 0.14 [-0.32, 0.26]	0.82

Notes: ES = effect size; B = regression coefficient; SE = standard error; 95% CI = confidence intervals for B

Figure 3

Full-CAT versus CAT-NR session-by-session depression (PHQ-9) mean scores.

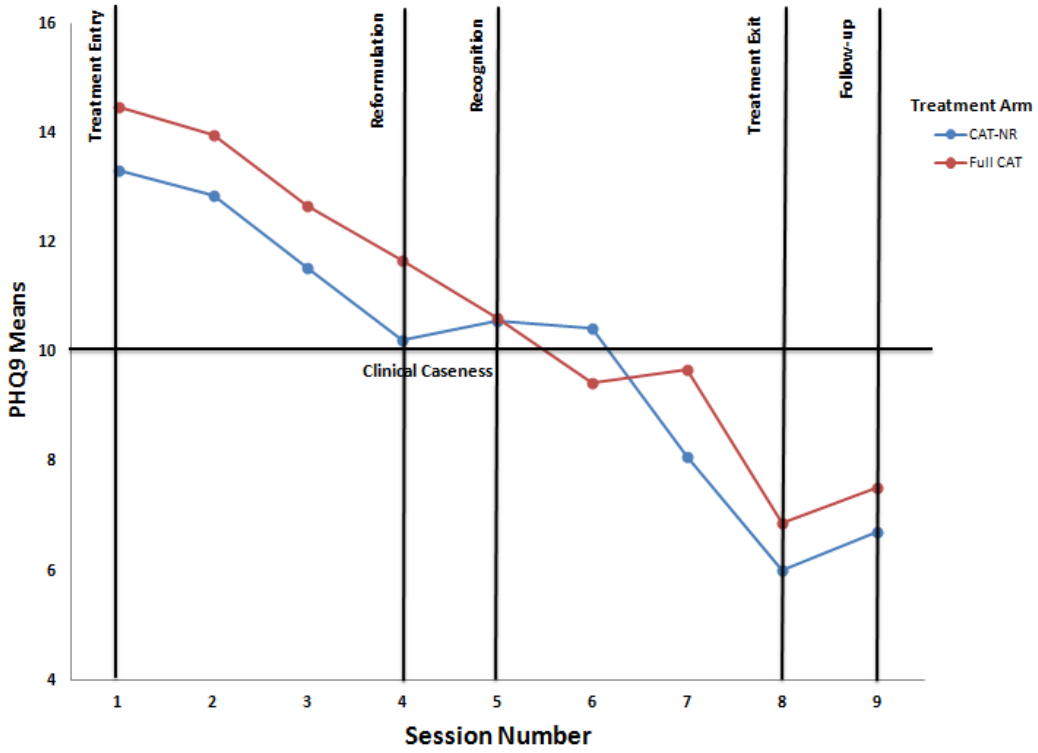

Figure 4

Full-CAT versus CAT-NR session-by-session depression (PHQ-9) linear growth trends and 95% confidence intervals.

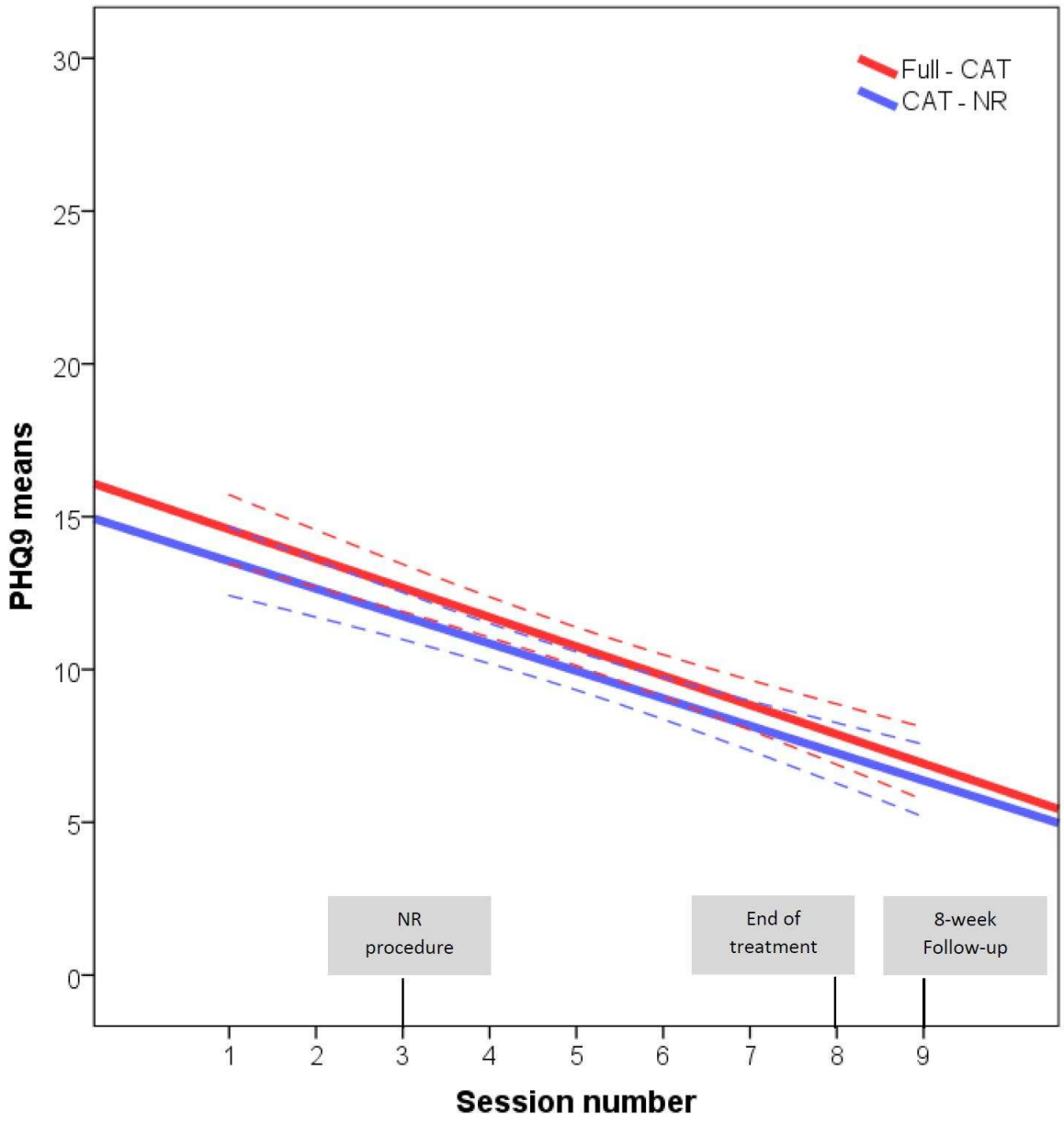

Table 3

Pre-post and screening-to-follow-up PHQ-9 comparisons.

Group	Session 1 mean (SD)	Session 8 mean (SD)	Effect size (d)	t-value
Full-CAT	15.7(5.1)	6.9(5.4)	d = 1.68	t(38)= 9.939 p = <.001*
CAT-NR	14.3(5.4)	6.0(4.7)	d = 1.63	t(35)= 6.933 p = <.001*
	Screening mean (SD)	Follow-up mean (SD)	Effect size (d)	t-value
Full-CAT	15.7(5.1)	7.5(7.4)	d = 1.29	t(35)= 8.515 p = <.001*
CAT-NR	14.3(5.4)	6.7(6.4)	d = 1.28	t(29)= 5.939 p = <.001*

* p = significant < 0.001

Table 4.
Recovery rates full-CAT and CAT-NR at the end of treatment and at follow-up

	Within-group outcomes		Between-group outcomes (Full CAT vs CAT-NR)	
	Full-CAT (n; %)	CAT-NR (n; %)	McNemar value (n = 95)	p
End of treatment PHQ-9 outcome status				
Recovered	18 (34.6)	19 (44.2)	0.60	.44
Improved	24 (46.2)	20 (46.5)	0.00	1.00
Clinically significant change	29 (55.8)	33 (76.7)	8.31	0.04*
Stasis	17 (32.7)	9 (20.9)	2.52	.112
Deteriorated	0 (0.0)	0 (0.0)		
Follow-up PHQ-9 outcome status				
Recovered	20 (38.5)	15 (34.9)	1.31	.25
Improved	26 (50.0)	17 (39.5)	0.00	1.00
Clinically significant change	28 (53.8)	28 (68.1)	3.35	.067
Stasis	18 (34.6)	13 (30.2)	6.61	.001*
Deteriorated	0 (0.0)	0 (0.0)		

Note. *p<.05