Investigating the relationship between competence and patient outcome with CBT

Article

Accepted Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

www.reading.ac.uk/centaur

CentAUR

Investigating the relationship between competence and patient outcome with CBT

Amanda Branson[a1] PhD., Roz Shafran PhD[b], Pamela Myles[a]

[a.] School of Psychology and CLS, University of Reading, Earley Gate, Reading,

Berkshire, RG14 6AL, United Kingdom Tel: +44 118 378 7531, Email:

a.branson@reading.ac.uk, r.shafran@ucl.ac.uk, pamelamyles@reading.ac.uk,

[b.] Child Health Centre, University College London, Gower Street, London, WC1E

6BT, United Kingdom. r.shafran@ucl.ac.uk

[a1] Corresponding Author

Requests for reprints should be sent to the corresponding author.

Abstract

Little is understood about the relationship between therapist competence and the outcome of patients treated for common mental health disorders. Understanding the relationship between competence and patient outcome is of fundamental importance to the dissemination and implementation of Cognitive Behavioural Therapy (CBT). The current study extends existing literature by exploring the relationship between CBT competence and patient outcome in routine clinical practice within the framework of the British Government's Improving Access to Psychological Therapies (IAPT) programme. Participants comprised 43 therapists treating 1247 patients over a training period of one year. Results found little support of a general association between CBT competence and patient outcome; however significantly more patients of the most competent therapists demonstrated a reliable improvement in their symptoms of anxiety than would be expected by chance alone, and fewer experienced no reliable change. Conversely, significantly more patients treated by the least competent therapists experienced a reliable deterioration in their symptoms than would be expected. The implications of these results for the dissemination and implementation of CBT are discussed.

Introduction

Further to recommending Cognitive Behavioural Therapy (CBT) for the treatment of anxiety and depression (NICE, 2009, 2011), the National Institute for Health and Care Excellence (NICE) guidelines state that therapists should be *'experienced in the treatment of the disorder and competent in the delivery of treatment provided'* (NICE, 2009, p116). Comprehensive training is considered necessary for the attainment of CBT competences, and delivery of specific evidence-based CBT interventions. The majority of studies indicate that CBT knowledge and skills can be successfully trained (e.g. Barnfield, Mathieson, & Beaumont, 2007; James, Blackburn, Milne, & Reichfelt, 2001; Keen & Freeston, 2008; McManus, Westbrook, Vazquez-Montes, Fennell, & Kennerley, 2010; Simons, et al., 2010). McManus and colleagues (2010) evaluated the outcomes of trainees enrolled on a Postgraduate Diploma in CBT between 1998 and 2009 (n=278) using a combination of audio recordings of therapy sessions, rated using the Cognitive Therapy Scale (CTS: Young & Beck, 1980), case reports and essays. Results demonstrated a significant improvement across assessment types indicating improved clinical knowledge and skill. Keen and Freeston (2008) explored the acquisition of competence on a comparable training programme (n=52) using video recordings of therapy sessions rated using the revised CTS (CTS-R: I. Blackburn, et al., 2001), essays and case reports. Results again demonstrated significant improvements. Simons, et al. (2010) provided community therapists with brief CBT training, and found not only did therapists' competence improve from pre-training baseline to six months post training but that these gains in competence were maintained 12 months after training.

The acquisition of CBT skills is a prerequisite for the dissemination and implementation of CBT. It is often erroneously assumed that training improves knowledge and skill, and thereby patient outcome (Okiishi, Lambert, Nielsen, & Olgles, 2003). However, patient outcome cannot be considered as an index of competence since cause and effect cannot be determined from this relationship (Fairburn & Cooper, 2011; Webb, DeRubeis, & Barber, 2010). Improvement in symptoms may be due to the efficacy of the intervention, the passage of time alone,

or patients may be differentially assigned to therapists for reasons not determined within the investigation (Muse & McManus, 2013).

The small amount of research examining whether therapist competence is a correlate of patient outcome is mixed. In one of the most robust studies, Strunk, Brotman, DeRubeis, and Hollon (2010) investigated associations between therapist competency ratings and symptom change in patients presenting with depression (n=60) in the context of a treatment trial. Results indicated that competence predicted session-to-session symptom change early in treatment, and also predicted evaluator-rated end of treatment depressive symptom severity. Recently, Ginzburg and colleagues found that therapist competence explained 48% of variance in the outcomes of patients treated for social anxiety disorder (Ginzburg, Bohn, Höfling, & Weck, 2012).

Conversely, Huppert, Bufka, Barlow, Gorman, and Shear (2001) found no effect of therapist competence on patient outcome within a treatment trial for Panic Disorder. In a meta-analytic review of the competence-outcome literature, competence was found to be unrelated to patient outcome (Webb, et al., 2010). The aggregated effect size across 17 studies was close to zero; however studies focusing specifically on the relationship between competence and treatment of depression had larger effect sizes than those focusing on anxiety disorders. This may be because the Cognitive Therapy Scale (CTS: Young & Beck, 1980) was developed for the treatment of depression and its validity in assessing CBT competences in other areas is not established (Webb, et al., 2010).

An additional reason for the mixed research findings may be due to limitations associated with study design. Much of the research within the field has been conducted within the framework of a randomised controlled trial. Treatment trials typically recruit only those therapists who exceed pre-determined criteria for competence (e.g. Huppert, et al., 2001; Trepka, Rees, Shapiro, Hardy, & Barkham, 2004), therefore the available range of competence is subject to a restriction of range (Brown, et al., 2013; Wilborg, Knoop, Wensing, & Bleijenberg, 2012). No study has yet examined the association between competence and patient outcome in a large sample of trainee therapists delivering CBT in routine clinical practice. The

English 'Improving Access to Psychological Therapies' (IAPT) programme provides an ideal opportunity to examine the association between therapist competence and patient outcome. This £400 million national initiative provides comprehensive, curriculum-led therapist training in CBT, combining University based theoretical and experiential learning, with service provision. Trainees receive extensive clinical supervision from experienced supervisors who undergo additional specific CBT supervision training as part of the programme. All supervisors are accredited by the British Association for Behavioural and Cognitive Psychotherapies (BABCP). All patients complete weekly routine outcome measurement of depression and anxiety symptoms.

Another possible explanation for the equivocal findings regarding clinical competence and patient outcomes is the view that competence in the delivery of a specific intervention (such as CBT) is largely irrelevant; the critical determinant of outcome is the therapist (Okiishi, Lambert, Eggett, Nielsen, & Dayton, 2006). In particular it has been suggested that there are some 'supershrinks' whose patients improve quicker, show a greater reduction in symptoms, and reach recovery more often than others, whilst factors such as case load, therapist theoretical orientation and gender exert no influence (Okiishi, et al., 2003). If there are such 'supershrinks' or 'metacompetent' therapists, identifying their characteristics could have important implications for the treatment of patients, in terms of referral to therapists most likely to offer improved outcome, and for selecting applicants for training programmes (Heinonen, Lindfors, Laaksonen, & Knekt, 2012; Okiishi, et al., 2006). Previous studies have examined the influence of characteristics such as age, gender, years of, and type of experience (e.g. Huppert, et al., 2001; James, et al., 2001). However, research to date has been unable to pinpoint the intrinsic qualities of the effective therapist (Okiishi, et al., 2003).

The aim of the current study was to explore the relationship between patient outcomes and both CBT clinical competence and therapist characteristics within the IAPT training programme. It was hypothesized that (1) Trainee CBT therapists would improve in competence during training (2) there would be an association between therapist competence and patient outcomes, (3) Therapist characteristics (age,

gender, years of experience and academic attainment) would be associated with patient outcome.

Methods

Training course

The IAPT CBT postgraduate diploma at the University of Reading is a year-long training programme combining theoretical learning with clinical supervision and routine clinical practice, in partnership with National Health Service (NHS) IAPT service sites. The University of Reading IAPT CBT diploma was amongst the first to be accredited by the BABCP.

The training follows a national curriculum comprising three modules which encompass: the core principles of CBT, treatment for anxiety and treatment for depression. Details of the curriculum can be found at: http://www.iapt.nhs.uk/silo/files/national-curriculum-for-high-intensity-cognitive-behavioural-therapy-courses.pdf. Over the duration of the programme, therapists receive a minimum of 300 hours of teaching and 35 hours of clinical supervision, they must also assess, formulate and carry out a full course of CBT for a minimum of eight patients presenting with depression or an anxiety disorder. Entry criteria require therapists to be graduates, with extensive general mental health clinical experience. On successful completion of the course, therapists are eligible for accreditation by the BABCP.

Participants

Therapists

All therapists who were enrolled on the IAPT CBT training programme at the University of Reading between 2008/09 and 2011/12, and were employed by one of five participating IAPT National Health Service sites were invited to take part (N=75). Forty three therapists consented to be involved.

The mean age of therapists was 38.6 (range=24.8-54.4, SD=8.29), 29 (67.4%) were female. All therapists were graduates; of those who gave information on the subject studied (*n*=41), 19 (46.3%) studied psychology, 7 (17.1%) studied nursing, 10 (24.4%) studied another social science, and 5 (12.2%) studied other subjects. Four (9.8%) achieved a first class pass in their degree, 26 (63.4%) were awarded an Upper Second (2:1), and 11 (26.8%) were awarded a Lower Second (2:2). The mean amount of pre-training experience was 10.42 years (range=2-28.0, SD=7.28).

Chi-Square test found that the distribution of males to females did not differ significantly across service sites (χ^2(1, 43) = 1.38, *p*=.71), nor did the distribution of degree award classifications (χ^2(6, 40) =10.93, *p*=.09). One-Way ANOVAs found there to be no significant differences in age (*F*(4,38)=0.61, *p*=.66) or years of experience (*F*(4,38)=0.31, *p*=.87) across services.

Patients

IAPT follows a Stepped Care model of treatment delivery; following clinical assessment, patients receive either CBT or low intensity interventions, and may transition from low intensity interventions to CBT or vice-versa. At each point of contact, patients complete the IAPT Minimum Data set (IAPT, 2011a, 2011b), which includes measurement of symptoms of anxiety and depression. These data are stored in electronic patient management systems (IAPTus or PCMis).

Patients were included in the study if their current course of treatment was CBT and they had not been 'stepped-up' from low intensity treatment. This approach was taken since evidence suggests a positive relationship between step-up rates and recovery (David M Clark, Gyani, Layard, & Shafran, 2013) the inclusion of patients who had previously received step-2 interventions could therefore bias results. Furthermore this approach allowed for direct comparison with national data (Gyani, Shafran, Layard, & Clark, 2011).The entire course of therapy must also have been provided by one participating therapist. Patients must have received two or more sessions for which outcome data were available; and must have concluded their treatment; if the end of treatment status was '*declined treatment*', '*unsuitable*', or '*dropped out*', data were included, so long as a minimum of two sessions were

recorded for which outcome data were available. Change scores for 1247 patients meeting the above criteria were included in the analysis. Data were collected in two phases – during the 12 month training period ('during') (n=793) and the six months after the training period ('post') (n=454).

Measures

Therapist

Demographic data: participant age, gender and experience were collected from student records.

Clinical Skill: *Cognitive Therapy Scale Revised* (CTS-R: I. Blackburn, et al., 2001) was used to assess clinical skill. The CTS-R is a 12-item scale designed to measure therapist competence; items are measured on a 7-point Likert scale ranging from incompetent (0) to expert (6). The CTS-R has been shown to have adequate inter-rater reliability and high internal consistency (I. Blackburn, et al., 2001). Scores on the CTS-R range from 0-72, the threshold for competence is 36, i.e. any score falling below 36 is classified as a fail.

Trainees submitted three self-selected audio recordings of CBT therapy sessions (CTS-R1, CTS-R2, and CTS-R3) which are submitted towards the end of modules 1, 2 and 3 respectively. Recordings are marked by experienced CBT therapists, who are accredited by the British Association for Behavioural and Cognitive Psychotherapies (BABCP). Markers also receive one full-day of marker training annually. Inter rater reliability was explored through the blind rating of 9 recordings (3 x CTS-R1, CTS-R2 and CTS-R3) by six markers using one-way random effects Intra-Class-Correlations (ICCs) with absolute agreement (Shrout, 1998); ICCs for CTS-Rs1-3 were all within the moderate range (ICCs=0.71 $p<0.001$, 0.72 $p<0.001$ and 0.62 $p<0.01$ respectively).

Recordings are each marked by a single marker; furthermore In accordance with routine assessment on University courses, all recordings scoring less than 38 or more than 49, and a random sample were moderated by an experienced, accredited

CBT therapist (n=1; CTS-R1: 33 (77%); CTS-R2: 25 (60%); CTS-R3: 26 (63%)). A proportion of those in each range were also scrutinised by the external examiner, including all recordings scoring less than 38 or more than 49[1].

Patient

Patient Health Questionnaire (PHQ-9): The PHQ-9 (Kroenke, Spitzer, & Williams, 2001) measures the nine DSM-IV (American Psychiatric Association, 1994) diagnostic criteria for depression on a 0-3 point Likert style scale; It has been found to be both reliable and valid in the facilitation of recognition and diagnosis of depression within primary care; and can also be used to monitor symptom change over the course of treatment. A score greater than 9 indicates clinically significant symptoms of depression (IAPT, 2011b).

Generalized Anxiety Disorder Scale (GAD-7): This seven item self-report scale is both reliable and valid as a tool for screening for and assessing the severity of GAD in clinical settings (Spitzer, Kroenke, Williams, & Löwe, 2006). It has also been found to measure symptoms of panic disorder, social anxiety disorder and post-traumatic stress disorder moderately well (IAPT, 2011b). The GAD-7 is measured on a 0-3 point Likert scale, a score greater than 7 indicates clinically significant symptoms of anxiety (IAPT, 2011a).

Results

Preliminary analyses

A one-way ANOVA conducted to investigate whether mean CTS-R scores differed across service sites was found to be non-significant ($F_{(4,39)}=.726$, MSE=56.88, *p*=.58), and no post-hoc comparisons (Bonferroni corrected) approached significance (*p*=.93 to *p*=1.0), suggesting that training outcomes were comparable across service sites.

---

[1] Further details regarding the ratio of work scrutinised by the external examiner were not available.

Acquisition of competence

CTS-R scores are presented in Table 1. Trainees who failed at first attempt were afforded one opportunity to submit further recording, however the current data set reflects first submissions only.   Three therapists ultimately went on to fail the course (7%), two withdrew from training (4.7%), 23 passed (53.5%) and 15 were awarded a merit (34.9%), no therapists achieved a distinction.

[Insert Table 1 here]

Patterns of change over time

CTS-R scores were normally distributed. A one-way repeated measures ANOVA was found to be significant (F(2,78)=5.34, p=.007). Planned comparisons (repeated Bonferroni corrected) performed to understand the significant ANOVA identified a significant increase in scores from CTSR1 (M=35.51, SD=7.69) to CTSR2 (M=38.95, SD=6.26), p=.008, d=.49, and a non-significant decrease from CTS-R2 to CTS-R3 (M=37.48, SD=5.96) p=.50, d=.24.

Reliability of change in competence

Whilst results demonstrate that overall levels of competence increased over the duration of training, they do not explain the pattern of change for individual trainees. Importantly, what proportion of trainees who were rated as not competent on CTS-R1 moved to competency by CTS-R3, and was the change reliable? As suggested by McManus et al (2010), Jacobson and Truax (1991) approach to calculating statistically significant and reliable change was applied to trainees' CTS-R scores to determine whether scores had: *a)* changed sufficiently so that change is unlikely to be due to chance alone, and *b)* crossed the threshold of competence, as determined by the scale's developers (50%).

A pre-requisite of the reliable change index is the test-retest reliability of a measure, however this statistic is not currently available for the CTS-R. Due to the overlap between the CTS-R and the original CTS in terms of items and scoring, the mean of

four CTS reliability coefficients reported by Vallis, Shaw, and Dobson (1986) were used *(.81)*. Applying this to Jacobson and Truax's (1991) formula and adapting the procedure of McManus et al (2010) to include only trainees who scored below competence at CTS-R1 (*n*= 18) the criterion for reliable change was set as 4.50. Using this criterion, 42.11% (n=8) showed a reliable improvement, no students reliably deteriorated, while the remaining 57.89% did not change reliably. In relation to point *b* (crossed the threshold of competence), of those who demonstrated reliable improvement on CTS-R3, 75% (*n*=6) also passed the threshold of competence (36). Whilst the remaining two trainees showed a reliable improvement from CTS-R1 to CTS-R3, they remained below competence at CTS-R3.

Patient outcomes

Change in symptoms of anxiety and depression were calculated by subtracting final GAD-7 and PHQ-9 scores from first scores, a negative score reflects a reduction in symptoms. Jacobson and Truax's (1991) method of calculating statistically reliable change was used to determine the reliability of change in symptoms of depression and/or anxiety. The criterion for reliable change was calculated to be 2.12 on the PHQ-9 and 2.48 on the GAD-7; as outcome measure scores are integers, a change equal to or greater than 3 was required on each measure in order for change to be considered reliable. Individual reliable change indices greater than 1.96 are likely to occur by chance alone in only 5% of cases (Evans, Margison, & Barkham, 1998; Jacobson & Truax, 1991).

Combined reliable change on the PHQ-9 and GAD-7 was calculated using the approach taken by Gyani, et al. (2011) in their evaluation of Year 1 IAPT patient data. To demonstrate combined reliable change, patients should reliably improve on the PHQ-9 and/or GAD-7, if reliable improvement occurred on only one measure, there should be no deterioration on the other. Using this approach, 66.7% (n=529) of patients showed reliable improvement in their symptoms of anxiety and depression, and 5.5% (n=44) showed reliable deterioration; the equivalent figures for the six month post-training time point were 68.7% (n=312) and 3.3% (n=15). Gyani, et al. (2011) reported that 63.8% of patents reliably improved nationally in the first year. One-sample Wilcoxon Signed Ranks tests found that the percentage of

patients showing reliable improvement did not differ significantly from the national sample during training (*p*=.18), in the post training period the difference approached significance (*p*=.05).

Effect of time on patient outcome

Mann-Whitney U tests revealed no significant differences in reliable improvement by time (during training *Mdn*=67.1, or six months post training *Mdn*=75.0, *p*=.85), however the proportion of patients reliably deteriorating was significantly lower in the six months post training period (*Mdn=0.00)* than during training *(Mdn=0.05)* (U=501.00, *z*=02.01, *p*=0.04).

Effect of competence on patient outcome

To test hypothesis 2, associations between therapist competence and patient outcomes were explored. Mean competency ratings (CTS-Rs 1-3) were calculated for each therapist, this mean rating was applied both to the during-training and post-training analyses. Mean PHQ-9 and GAD-7 reliable change, and proportion of patients reliably improving, or deteriorating during each time frame (during and post training) were calculated for each therapist. Therapists treating less than five patients during the time frame (during and/or post) were excluded. .

Spearman Rho correlations between mean CTS-R scores and patient outcome variables were in the modest range (*r*=.070 to *r*=.267) both during and post-training, and none were significant (see Table 2).

[Insert Table 2 here]

There is a risk that treatment response could exert an influence on the competence-outcome relationship. CTS-R1 was conducted in the first session, therefore correlational analyses were repeated using CTS-R1 to 3 scores independently, allowing us to explore whether stage of treatment or treatment response affected the competence-outcome relationship. To reduce the risk of type one error, individual difference variables were excluded, and alpha was set at .01. Spearman Rho

correlations were found to be predominantly modest (*r=.056 to r*=.284) and non-significant (*p*=.08 to *p*=.97); only one correlation was in the moderate range, but it did not reach significance at the .01 level (*r*=.388, *p*=.03).

There is also a risk that initial symptom severity could exert an influence on the competence-outcome relationship; patients who begin treatment with higher baseline PHQ-9 and/or GAD-7 scores are likely to experience a greater change in symptoms than those who commence treatment at a lower baseline (Gyani, et al., 2011). A Spearman Rho correlation matrix was generated to explore whether patients' first scores on the PHQ-9 and GAD-7 were related to CTS-R scores, correlations were week and non-significant for the each measure respectively (*r*=.014, *p*=.62; *r*=.024, *p*=.41).

Further analyses were conducted to explore relationships between therapist competence and patient outcome. Mixed effects models were applied, with random intercept term for therapist, modelling the effect of CTS-R on change in a) PHQ-9 and b) GAD-7 scores, considering both linear and quadratic relationships. The linear effect of competence on change in depressive symptoms was found to be non-significant (F(1,25.33)=2.47, p=.13, Var (u_{0j}) = 0.94, z=1.70, p=.09); similarly the the quadratic effect also failed to reach significance (F(1,24.93)=4.29, p.05, Var (u_{0j}) = 0.81, z=1.59, p=.11).

In a similar pattern, the linear effect of therapist competence on change in symptoms of anxiety was found to be non-significant (F(1,27.60)=3.12, p=.09; Var (u_{0j}) = 0.96, z=1.97, p=.05), as was the quadratic model (F(1,27.45)=2.39, p=.13, Var (u_{0j}) = -0.01, z=1.97, p=.05). These results indicate that therapist competence was not a driving factor in change in patients' symptoms of anxiety.

Differences in patient outcome between the most and least competent therapists

To see if outcomes differed between the most and least competent therapists, therapists were grouped by mean CTS-R score into one of three competence groups. Therapists scoring in the top 10% (n=4, M=47.7, SD=2.71) were placed in

the "top group", therapists scoring in the bottom 10% (n=4, M=29.16, SD=0.57) were placed in the "bottom" group, the remaining therapists were in the mid-range group (n=35, M=36.98, SD=3.77).

Chi-squared analyses indicated no significant differences between CTS-R competence groups for reliable change on the PHQ-9 (χ^2(4)=8.49, p=.08, V=.08), and no standardised residuals reached significance (see Table 3). Reliable change on the GAD-7 differed significantly between CTS-R competence groups (χ^2(4)=13.94, p=.007, V=.10) (see Table 4). Investigation of standardised residuals suggested that significantly fewer patients than would be expected experienced no reliable change in their symptoms of anxiety if treated by the 'top' group (n=80, expected=98.7, Z=-1.9) and more patients than would be expected experienced a reliable improvement, although the standardised residual didn't quite reach significance (n=164, expected=143.2, z=1.7). Conversely, when treated by the 'bottom' group, significantly more patients than would be expected experienced a reliable deterioration in their symptoms (n=8, expected=3.9, z=2.1).

[Insert Table 3 and Table 4 here]

Effect of therapist characteristics on competence and patient outcome

To test hypothesis 3, Spearman Rho correlations were used to explore associations between therapist characteristics (age, gender, degree classification and years of experience) and (i) clinical competence (CTS-R) and (ii) patient outcome (mean reliable change on the PHQ-9 and GAD-7, and proportions of patients showing a reliable improvement or deterioration in their condition). Alpha was set at .01 to reduce the risk of spurious correlations whilst minimising the risk of Type II error.

Table 3 shows that correlations between CTS-R (clinical competence) and therapist variables were predominantly in the modest range, both during training and in the subsequent six months. No correlations reached significance at the .01 level, however in the post-training phase a moderate, negative correlation was observed between therapist age and CTS-R (r=-.479, p=.02), older therapists performed less well than their younger counterparts.

In a broadly similar pattern, correlations between therapist variables and patient outcomes failed to reach significance at the .01 level, and were predominantly in the modest range (r=.012 – r=.303) both during and post training.

Discussion

The current study reports on training data of 43 CBT therapists, and the outcomes of 1247 patients treated for anxiety and/or depression during and after training. The IAPT minimum dataset has provided an ideal opportunity to explore the effect of therapist competence on patient outcomes within routine clinical practice.

The data showed that competence in the delivery of CBT improved over the duration of a yearlong, national curriculum-led training programme. CTS-R scores improved significantly over time, passing the threshold of competence. Furthermore, 50% of those who performed below competence on their first assessment passed their third assessment, 75% of whom did so reliably. This finding adds further support to the existing evidence that it is possible to train the majority of therapists, from diverse backgrounds, to competence in delivering CBT in a yearlong training programme. Between modules two and three a small, non-significant deterioration was observed. This is not necessarily surprising; whilst the trainees have gained experience during training, they remain relative novices and as such high scores would not be expected in the majority of cases. It should be noted that twenty five therapists passed the threshold of competence on their first submission. This is as would be expected at this stage, the threshold score can be qualified as 'advanced beginner'; it would be expected that therapists would exceed this level, moving towards "competence" (score in the 40s) prior to the completion of training, with few therapists achieving scores in the range of "proficiency" (scoring in the 50-60 range) (I. M. Blackburn, James, Milne, & Reichelt, 1999).

Limited support was found for the hypothesis that there would be an association between CBT competence and patient outcome. Across time frames no associations between CTS-R and reliable improvement in symptoms of anxiety (GAD-7) or depression (PHQ-9) approached significance. This may reflect limitations in generic competency measures such as the CTS-R to detect the discrete competences

associated with the delivery of specific CBT protocols (Fairburn & Cooper, 2011; Webb, et al., 2010). Similarly, research has found the CTS to lack inter-rater reliability (Dimidjian, et al., 2006; Jacobson, et al., 1996). However when raters are trained together, reliability can be improved (J. Barber, P, Sharpless, Klostermann, & McCarthy, 2007), within the current study, Intra-class correlations were acceptable, suggesting that raters were reliable. Disorder specific measures (see for example: D M Clark, von Consbruch, Hinrichs, & Stangier, 2007) may offer a more reliable estimate of competence; it has however been suggested that judgements of competence may be more reliably made where raters are experts in the given treatment modality (J. P. Barber, Crits-Christoph, & Luborsky, 1996), indeed raters in the current study were experts in the field of CBT for depression and anxiety. This finding therefore lends support to the view that patient outcome should not be used as a metric of competence for any disorder (Fairburn & Cooper, 2011).

It is important to note that despite the failure to find a general association between competence and patient outcome, when extreme mean competence values were explored, significant differences emerged in the frequency of patients reliably improving and deteriorating. Significantly more patients of the most competent 10 percent of therapists demonstrated a reliable improvement in their symptoms of anxiety (but not depression), and fewer experienced no reliable change. Conversely, more patients treated by the weakest 10% of therapists demonstrated a reliable deterioration in their symptoms of anxiety (but not depression) than would be expected. Therefore, whilst some authors support the view of competence being a continuum of increasing skill (e.g. J. Barber, P, et al., 2007; Manring, Beitman, & Dewan, 2003), a more appropriate approach may be to consider that the majority of therapists, of average competence are 'good enough' (Yager & Kay, 2003), whilst a small number of, highly effective therapists, or '*supershrinks'* (Okiishi, et al., 2003) may achieve superior outcomes. Equally, competence may be a more critical determinate of outcome for some patients than for others.

A further explanation for the lack of association between competence and patient outcomes may be found in a definitional disparity between competence and performance. Performance, argue Bein and colleagues (2000), is *'the level of skill evidenced in a particular session or case'*, whilst competence is *'a general capacity,*

across cases, to perform skilfully' (p 128). Whilst competence reflects a stable state of being, performance is prone to measurement error, potentially varying substantially across therapy sessions or cases. In their evaluation of the generalisability of postgraduate therapist training, Keen and Freeston (2008) found substantial variability in trainee performance, arguing that to obtain competence scores with less measurement error, ratings of 19 therapy sessions over the course of training are necessary.

Relatedly, although evidence-based treatments are protocol driven, the competent therapist will adapt therapy to the needs of the patient. Individual therapy sessions may therefore not *a)* reflect the adaptive nature of a course of treatment, or *b)* provide the opportunity for therapists to display their true skill in all fields under evaluation. Thus, whilst it is acknowledged that the self-selection of therapy sessions by trainees may introduce a bias, it could also be argued that self-selection adds 'fairness' in that it allows students to choose examples of work that they believe demonstrate the competences required, a process which may also support learning. The process of reflecting on and critiquing one's own work, and experiencing a feedback loop encourages learners to view their work in relation to key criteria, which may act as a learning tool (Kolb, 1984). It is possible therefore that modular evaluation and self-selection is simply not sufficient to glean an accurate measure of competence, rather it offers a relatively unstable measure of performance.

There was little support for the final hypothesis that there would be an association between therapist characteristics (age, gender, degree performance and experience) and patient outcome. Within the broad sample of HI therapists (n=115), from which the current sample were drawn, a range of variables, in particular prior academic attainment, and personality, were associated with training outcomes. A positive association was found between CTS-R scores and degree performance; and also between the personality dimension of agreeableness (NEO PI-R: Costa & McCrae, 2006) and CTS-R performance (Branson & Shafran, 2015). Whilst apparently important in terms of training, therapist characteristics do not appear to be central to the relationship between therapist and treatment outcomes.

The current study overcomes some limitations of previous studies by linking competence to outcomes with a large sample of patients treated over a three-year period in routine clinical practice. The naturalistic design may be seen as an advantage of this study in that it enables exploration of real-life data from routine clinical practice; arguably the results of treatment offer a true measure of the effectiveness of a treatment or treatment framework. Nevertheless, the uncontrolled, retrospective nature of the study is a limitation. The archival nature of the patient dataset means that we cannot know if there were any systematic differences in the allocation process, how accurately data were entered, whether missing data were missing at random, or if certain therapists failed to input session by session data routinely or strategically. There is also a question of generalizability; the IAPT service sites are routine in the UK but not elsewhere. Whether the findings apply to other routine clinical service settings remains to be answered. Importantly, it must be acknowledged that the current study does not take into consideration the fact that therapy outcomes are likely to be impacted by many factors other than therapist competence (Trepka, et al., 2004), such as those associated with the patient. Data recorded by the IAPT computer systems did not provide patient demography or case history, and patient self-report symptomology (PHQ-9 and GAD-7) was used as a proxy for diagnosis and complexity, which is not ideal. Comorbid symptoms of anxiety may complicate CBT for depression (Singer, Dobson, & Dozois, 2008); and patients with chronic difficulties may require a highly skilled therapist (Moore & Garland, 2003; Riso & Newman, 2003). Indeed, evidence suggests that chronicity may moderate the relationship between competence and outcome (Strunk, et al., 2010).

Further limitations centre on the use of self-selected therapy sessions, and the stage of therapy within which competency ratings were made. It is argued that self-selection can result in a restriction of range (J. Barber, P, et al., 2007; Webb, et al., 2010), and artificial inflation of true competence (Fairburn & Cooper, 2011). If future research is going to use therapist selected sessions, it will be important to understand whether these sessions are representative of a therapists' work overall. A Related issue focuses on the use of mean competency ratings in some analyses. Whilst this approach allows for some reduction of measurement error, the approach does not discriminate between therapists who may maintain relatively average

scores across training and those who perhaps started from a lower baseline and demonstrated greater improvement over the three ratings. This may have complicated the relationship between competence and patient outcome.

In regard to the stage of therapy, to ensure that any competence-outcome relation is a true reflection of competence and subsequent outcome, competence should be measured as early as possible, perhaps even at intake since large gains are often observed between intake and session one. It was only possible to do this for the first recording, data were not available regarding the stage of therapy for recordings two and three, we therefore cannot be sure about the directionality of the competence-outcome relationship for the subsequent sessions. Despite these limitations, this is one of only a few studies to explore the relationship between CBT competence and patient outcome within routine clinical practice, and the first to examine the question within the Improving Access to Psychological Therapies programme. Limited support was found for a general association but more patients than expected reliably improved when treated by therapists who showed the highest competence. These findings have potential importance for understanding the complex relationship between competence, patient outcome, and the implementation of CBT.

Acknowledgements

References

American Psychiatric Association. (1994). Diagnostic and Statistical Manual of Mental Disorders IV. Washington DC: Author.

Barber, J., P, Sharpless, B., A, Klostermann, S., & McCarthy, K., S. (2007). Assessing Intervention Competence and Its Relation to Therapy Outcome: A selected review derived from the outcome literature. Professional psychology, research and practice, *38*, 493-500.

Barber, J. P., Crits-Christoph, P., & Luborsky, L. (1996). Effects of Therapist Adherence and Competence on Patient Outcome in Brief Dynamic Therapy. Journal of consulting and clinical psychology, *64*, 619.

Barnfield, T. V., Mathieson, F. M., & Beaumont, G. R. (2007). Assessing the Development of Competence During Postgraduate Cognitive-Behavioral Therapy Training. Journal of Cognitive Psychotherapy, *21*, 140-147.

Bein, E., Anderson, T., Strupp, H., Henry, W., Schacht, T., Binder, J., & Butler, S. (2000). The Effects of Training in Time-Limited Dynamic Psychotherapy: Changes in Therapeutic Outcome. Psychotherapy Research, *10*, 119-132.

Blackburn, I., James, I., Milne, D., Baker, C., Standart, S., Garland, A., & Reichelt, F. K. (2001). The Revised Cognitive Therapy Scale (CTS-R): Psychometric Properties. Behavoural and Cognitive Psychotherapy, *29*, 431-446.

Blackburn, I. M., James, I. A., Milne, D. L., & Reichelt, F. K. (1999). CTS-R: Revised scale for assessing competence in cognitive therapy (1st ed.). In. Newcastle: Newcastle Cognitive and Behavioural Therapies Centre.

Branson, A., & Shafran, R. (2015). Therapist Characteristics and their Effect on Training Outcomes: What Counts? Behavioural and Cognitive Psychotherapy, *FirstView*, 1-7.

Brown, L. A., Craske, M. G., Glenn, D. E., Stein, M. B., Sullivan, G., Sherbourne, C., Bystritsky, A., Welch, S. S., Campbell-Sills, L., Lang, A., Roy-Byrne, P., & Rose, R. D. (2013). CBT Competence in Novice Therapists Improves Anxiety Outcomes. Depression and Anxiety, *30*, 97-115.

Clark, D. M., Gyani, A., Layard, R., & Shafran, R. (2013). Enhancing recovery rates: lessons from year one of the English 'Improving access to psychological therapies' programme. CEP Discussion Papers, CEPDP1227. In. Centre for Economic Performance, London School of Economics and Political Science, London, UK.

Clark, D. M., von Consbruch, K., Hinrichs, J., & Stangier, U. (2007). Cognitive Therapy of Social Phobia: checklist of therapist competency. In. London: Kins College London.

Costa, P. T., & McCrae, R. R. (2006). NEO Revised. NEO Personality Inventory (NEO PI-R) Manual (UK Edition ed.). Oxford: Hogrefe.

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., Gallop, R., McGlinchey, J. B., Markley, D. K., Gollan, J. K., Atkins, D. C., Dunner, D. L., & Jacobson, N. S. (2006). Randomized Trial of Behavioral Activation, Cognitive Therapy, and Antidepressant Medication in the Acute Treatment of Adults With Major Depression. Journal of consulting and clinical psychology, *74*, 658-670.

Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence Based Mental Health, 1*, 70-72.

Fairburn, C. G., & Cooper, Z. (2011). Therapist Competence, Therapy Quality, and Therapist Training. *Behaviour Research and Therapy, 49*, 373-378.

Ginzburg, D. M., Bohn, C., Höfling, V., & Weck, F. (2012). Treatment specific competence predicts outcome in cognitive therapy for social anxiety disorder. *Behaviour Research and Therapy, 50*, 747-752.

Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2011). Enhancing Recovery Rates in IAPT Services: Lessons from analysis of the Year One data. In (Vol. 2012): IAPT.

Heinonen, E., Lindfors, O., Laaksonen, M. A., & Knekt, P. (2012). Therapists' professional and personal characteristics as predictors of outcome in short- and long-term psychotherapy. *Journal of affective disorders, 138*, 301-312.

Huppert, J. D., Bufka, L. F., Barlow, D. H., Gorman, J. M., & Shear, M. K. (2001). Therapists, Therapist Variables, and Cognitive-Behavioral Therapy Outcome in a Multicentre Trial for Panic Disorder. *Journal of consulting and clinical psychology, 69*, 747-755.

IAPT. (2011a). The IAPT Data Handbook Appendices: Including the IAPT Data Standard. In NHS (Ed.). London.

IAPT. (2011b). The IAPT Data Handbook: Guidance on recording and monitoring outcomes to support local evidence-based practice. In NHS (Ed.). London.

Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., Gortner, E. T., & Prince, S. E. (1996). A Component Analysis of Cognitive-Behavioral Treatment for Depression. *Journal of consulting and clinical psychology, 64*.

Jacobson, N. S., & Truax, P. A. (1991). Clinical Significance: A Statistical Approach to Defining Meaningful Change in Psychotherapy Research. *Journal of consulting and clinical psychology, 59*, 12-19.

James, I. A., Blackburn, I. M., Milne, D. L., & Reichfelt, F. K. (2001). Moderators of Trainee Therapists' Competence in Cognitive Therapy. *British Journal of Clinical Psychology, 40*, 131-141.

Keen, A. J. A., & Freeston, M. H. (2008). Assessing Competence in Cognitive-Behavioural Therapy. *British Journal of Psychiatry, 193*, 60-64.

Kolb, D. A. (1984). Experiential Learning: Experience as the Source of Learning and Development. Englewood Cliffs: Prentice-Hall.

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.

Manring, J., Beitman, B. D., & Dewan, M. J. (2003). Evaluating Competence in Psychotherapy. *Academic Psychiatry, 27*, 136-144.

McManus, F., Westbrook, D., Vazquez-Montes, M., Fennell, M., & Kennerley, H. (2010). An Evaluation of the Effectiveness of Diploma-Level Training in

Cognitive Behaviour Therapy. *Behaviour Research and Therapy*, *48*, 1123-1132.

Moore, R. G., & Garland, A. (2003). Cognitive Therapy for Chronic and Persistent Depression. West Sussex: Wiley.

Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review*, *33*, 484-499.

NICE. (2009). Depression. The treatment and management of depression in adults. This is a partial update of NICE clinical guideline 23. Clinical guideline 90. . In NICE (Ed.). London: National Collaborating Centre for Mental Health.

NICE. (2011). Generalised Anxiety Disorder in Adults; Management in Primary, Secondary and Community Care. National Clinical Guideline Number 113. In NICE (Ed.). London: The British Psychological Society and The Royal College of Psychiatrists.

Okiishi, J. C., Lambert, M. J., Eggett, D., Nielsen, L., & Dayton, D. D. (2006). An Analysis of Therapist Treatment Effects: toward Providing Feedback to Individual Therapists on Their Clients' Psychotherapy Outcome. *Journal of Clinical Psychology*, *62*, 1157-1172.

Okiishi, J. C., Lambert, M. J., Nielsen, S., & Olgles, b. (2003). Waiting for Supershrink: An Empirical analysis of Therapist Effects. *Clinical Psychology & Psychotherapy*, *10*, 361-373.

Riso, L. P., & Newman, C. F. (2003). Cognitive therapy for chronic depression. Journal of Clinical Psychology, *59*, 817-831.

Shrout, P. E. (1998). Measurement reliability and agreement in psychiatry. Statistical Methods in Medical Research, *7*, 301-317.

Simons, A. D., Padesky, C. A., Montemarano, J., Lewis, C. C., Murakami, J., Lamb, K., DeVinney, S., Reid, M., Smith, D. A., & Beck, A. T. (2010). Training and Dissemination of Cognitive Behavior Therapy for Depression in Adults: A Preliminary Examination of Therapist Competence and Client Outcomes. Journal of consulting and clinical psychology, *78*, 751-756.

Singer, A. R., Dobson, K. S., & Dozois, D. J. A. (2008). Generalized anxiety disorder, obsessive-compulsive disorder, and posttraumatic stress disorder. In M. Whisman (Ed.), Adapting cognitive therapy for depression: Managing complexity and comorbidity (pp. 88-109). New York Guilford Press.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. Archives of Internal Medicine, *166*, 1092-1097.

Strunk, D. R., Brotman, M. A., DeRubeis, R. J., & Hollon, S. D. (2010). Therapist Competence in Cognitive Therapy for Depression: Predicting Subsequent Symptom Change. Journal of consulting and clinical psychology, *78*, 429-437.

Trepka, C., Rees, A., Shapiro, D. A., Hardy, G. E., & Barkham, M. (2004). Therapist Competence and Outcome of Cognitive Therapy for Depression. Cognitive Therapy and Research, *28*, 143-157.

Vallis, M. T., Shaw, B. F., & Dobson, K. S. (1986). The Cognitive Therapy Scale. *Journal of consulting and clinical psychology, 54*, 381-385.

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: a meta-analytic review. *Journal of consulting and clinical psychology, 78*.

Wilborg, J. F., Knoop, H., Wensing, M., & Bleijenberg, G. (2012). Therapist effects and the dissemination of cognitive behaviour therapy for chronic fatigue syndrome in community-based mental health care. *Behaviour Research and Therapy, 50*, 393-396.

Yager, J., & Kay, J. (2003). Assessing Psychotherapy Competence in Psychiatric Residents: Getting Real. *Harvard Review of Psychiatry, 11*, 109-112.

Young, J. E., & Beck, A. T. (1980). Cognitive Therapy Scale: Rating manual. Unpublished Manuscript. In. Philadelphia, PA.: University of Pennsylvania.

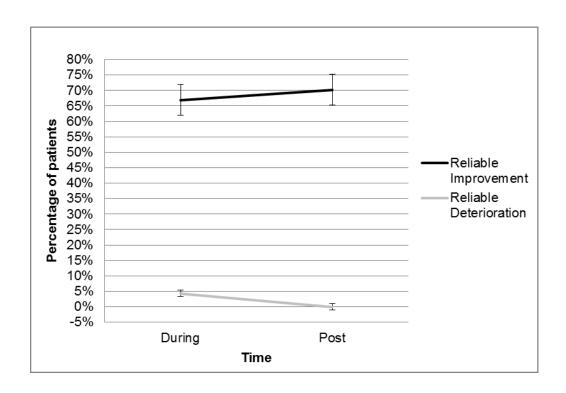*Figure 1 Graph showing median percentage of patients reliably improving, and deteriorating by timeframe*

	n	Minimum	Maximum	Mean	Std. Deviation
CTS-R 1	*43*	19	49	35.72	7.53
CTS-R 2	*42*	22	57	38.76	6.28
CTS-R 3	*40*	28	55	37.48	5.95
Mean CTS-R	*43*	28.56	50.16	37.25	5.36

Table 1 Mean CTS-R scores

Table 2 Correlations (Spearman's Rho) between mean CTS-R, therapist characteristics, mean reliable change score for PHQ-9 and GAD-7, proportion of patients reliably improving and proportion of patients reliably deteriorating during and post training

a)

DURING (n=37)	CTS-R	Age	Sex	Degree	Experience	PHQ-9 change[a]	GAD-7 change[a]	Reliable improvement[b]	Reliable deterioration[b]
CTS-R	1	-.266	.008	-.235	-.074	-.146	-.257	.283	-.135
Age		1	-.162	.248	.523*	.024	.136	-.103	.29
Sex			1	-.164	.2	-.019	-.005	.135	-.065
Degree				1	.198	-.106	.145	.007	.173
Experience					1	-.262	-.037	.213	.118
PHQ-9 change						1	.794**	-.672**	.146
GAD-7 change							1	-.760**	.377
Reliable improvement								1	-.006
Reliable deterioration									1

b)

POST (n=25)	CTS-R	Age	Sex	Degree	Experience	PHQ-9 change[a]	GAD-7 change[a]	Reliable improvement[b]	Reliable deterioration[b]
CTS-R	1	-.479	-.069	-.176	-.04	-.218	-.225	.07	.027
Age		1	.116	.105	.344	-.213	-.012	.303	-.481
Sex			1	.043	.237	-.048	-.109	.403	.103
Degree				1	.143	.211	.172	-.033	-.044
Experience					1	-.079	-.199	.015	.13
PHQ-9 change						1	.584*	-.705**	.413
GAD-7 change							1	-.479*	.328
Reliable improvement								1	-.504
Reliable deterioration									1

* p<0.01 (2tailed), ** p<0.001 (2 tailed)
a. Mean reliable PHQ-9/GAD-7 change scores calculated by applying Jacobson and Truax's (1991) formula for reliable change to the difference between patients' first and last PHQ-9 scores.
b. Proportion of patients showing reliable improvement/deterioration on the PHQ-9 and/or GAD-7, with no deterioration on either measure.

Table 3 *Contingency table showing the frequency and expected frequency of patients reliably improving, deteriorating and showing no reliable change on the PHQ-9 by CTS-R performance group*

| | | | Reliable change (PHQ-9) | | | |
			Reliable Improvement	Reliable Deterioration	No Reliable Change	*Total*
CTS-R	Top 10%	Count	162	7	81	*250*
		Expected Count	146.7	8.7	94.6	*250.0*
		Std. Residual	1.3	-.6	-1.4	
	Mid-range	Count	590	37	417	*1044*
		Expected Count	612.8	36.2	395.0	*1044.0*
		Std. Residual	-.9	.1	1.1	
	Bottom 10%	Count	78	5	37	*120*
		Expected Count	70.4	4.2	45.4	*120.0*
		Std. Residual	.9	.4	-1.2	
	Total	Count	*830*	*49*	*535*	*1414*

Table 4 Contingency table showing the frequency and expected frequency of patients reliably improving, deteriorating and showing no reliable change on the GAD-7 by CTS-R performance group

			Reliable change (GAD-7)			
			Reliable Improvement	Reliable Deterioration	No Reliable Change	*Total*
CTS-R	Top 10%	Count	**164**	6	**80**	*250*
		Expected Count	**143.2**	8.1	**98.7**	*250.0*
		Std. Residual	**1.7**	-.7	**-1.9***	
	Mid-range	Count	577	32	435	*1044*
		Expected Count	598.0	34.0	412.0	*1044.0*
		Std. Residual	-.9	-.3	1.1	
	Bottom 10%	Count	69	**8**	43	*120*
		Expected Count	68.7	**3.9**	47.4	*120.0*
		Std. Residual	.0	**2.1***	-.6	
	Total	*Count*	*810*	*46*	*558*	*1414*