Investigating the Relationship Between Therapist Competence and Patient Outcome in Adult Psychological Interventions

Niall Power

A thesis submitted in partial fulfilment of the requirements for the award of Doctor of Clinical Psychology at the University of Sheffield

The University of Sheffield

Faculty of Science

Clinical Psychology Unit, Department of Psychology

May 2021

Declaration

I declare that this thesis has been submitted for the award of Doctorate in Clinical Psychology at the University of Sheffield. It has not been submitted for any other qualification or to any other academic institution.

Word Count

Section one: Literature Review

Excluding tables, references and appendices……7,821

Including tables and references, excluding appendices……..12,677

Section two: Empirical Research Report

Excluding tables, references and appendices……7,640

Including tables and references, excluding appendices……..11,434

Total

Excluding tables, references and appendices……..15,461

Including tables and references, excluding appendices……..24,111

Acknowledgements

First, I would like to thank the participants who shared their outcome data and made this research possible. Thank you to my supervisors; Dr Stephen Kellet for providing strong support with rapidly developing a new project after the COVID-19 pandemic obliterated my original study and your consistent availability thereafter, and Dr Melanie Simmonds-Buckley for your patience and invaluable expertise with some of the more technical aspects of the projects. Thank you also to Jessica Smithies who helped make some data management tasks that could have been quite complex, quite straightforward.

A huge thank you to my family; particularly my mum, Gaz, Cal and Ell. You have seen me through my rollercoaster of a journey and have helped me remember who I am through all of it. Thank you to my close mates, both old and new, who have always been a reliable source of laughter and encouragement. Finally, a wholehearted thank you to my girlfriend Jess who was confined in a small apartment with me for 12 months while I was writing this thesis. I am deeply grateful for your patience, resilience and unrelenting support.

Summary

Despite several decades of research suggesting that psychological interventions are effective in reducing psychological distress, *how* these interventions lead to change remains unclear. Some factors that are considered important when investigating how interventions lead to change are the levels of adherence, competence and integrity displayed by the therapist when delivering interventions. This thesis aimed to investigate the relationship between therapist competence and patient outcome and involved: 1) an updated and extended meta-analytic review investigating the association between adherence/competence/integrity and outcome, and 2) an empirical study investigating whether practitioner competence is associated with clinical outcomes in guided self-help (GSH) for adults with anxiety disorders.

Section one of this thesis reports a systematic review of 62 studies that investigated the association between therapist adherence/competence/integrity and patient outcome. Meta-analyses were conducted to estimate the overall association between adherence, competence and integrity and outcome across both non-hierarchical (e.g., correlation) and hierarchical (e.g., multilevel modelling) statistical methods. The findings suggest that there is a significant association between therapist integrity and outcome, some association between competence and outcome but no association between adherence and outcome. The findings offered mixed support for the findings of the previous meta-analytic review conducted over a decade ago. This may be the first meta-analysis investigating the integrity-outcome association and thus this needs further investigation. The level of heterogeneity across included studies was high and thus findings should be interpreted with caution.

Section two of this thesis reports an empirical study which was a secondary analysis of outcome data collected during a randomised controlled patient preference trial. This study set out to investigate the association between practitioner competence and patient and service outcomes in the delivery of GSH across and between a new intervention protocol (informed by cognitive analytic principles, n=60) and the standard protocol (informed by cognitive behavioural principles, n=20). A fully crossed design was used to confirm very high inter-rater reliability of competence ratings across competence-levels and treatment conditions. No significant associations were found between practitioner competence, patient outcome, treatment engagement nor need for further intervention. Competence did not predict rate of change in patient self-reported anxiety symptoms. The findings of the study suggest that practitioners of low intensity psychological interventions can competently deliver two distinct versions of GSH with brief training and under regular supervision. Analyses were under powered and thus the association between competence and outcome in GSH needs further exploration.

Taken as a whole, the two studies provide evidence that a practitioner's adherence to an intervention's techniques and the skill with which these techniques are applied (in combination) may be associated with the outcome of patients though this association may vary depending on the particular intervention used and the symptoms/diagnoses being treated. Measurement of treatment integrity needs to be integrated into ongoing clinical supervision of therapists/practitioners.

Contents

Section Two: Empirical Research Report

The Relationship Between Practitioner Competence and Outcome During Low Intensity Guided Self-Help for Anxiety

PART ONE: Literature Review

A Meta-Analytic Review of the Relationship Between Therapist Adherence/Competence/Integrity and Treatment Outcome During Adult Psychotherapy: An Update and Extension of Webb et al. (2010)

Abstract

Objectives

Whilst it is widely accepted that psychological therapies can be beneficial, the therapist behaviours which contribute to positive outcomes remain under-investigated. This paper sought to provide a contemporary review of the association between therapist adherence/competence/integrity and patient outcome and extend the scope of previous reviews in this area.

Design

A pre-registered (CRD42020193889) systematic review and meta-analysis.

Method

Studies that assessed the relationship between adherence/competence/integrity and clinical outcomes were identified though systematic searches and grouped according to whether they employed non-hierarchical or hierarchical analyses. Random effects meta-analyses were conducted on 1) adherence-outcome, 2) competence-outcome, and 3) integrity-outcome relationships. Moderator analyses included diagnosis, treatment modality, year of publication, percentage coverage ratings and risk of bias.

Results

The review identified N=62 studies suitable for inclusion across all meta-analyses. Most studies were rated as moderate-low risk of bias and non-hierarchical meta-analyses were graded as most reliable. There was a small but significant positive association between competence and outcome for non-hierarchical studies (r=0.16). The association between adherence and outcome was non-significant for both non-hierarchical (r=0.08) and hierarchical (r = 0.04) studies. The association between integrity and outcome was significant across both non-hierarchical (r=0.11)

and hierarchical (r=0.23) studies. Diagnosis, treatment modality, year of publication and percentage coverage significantly influenced the strength of process-outcome associations.

Conclusions

This evidence review suggests a significant association between integrity and outcome, some association between competence and outcome, and no association between adherence and outcome. Limitations of the review in addition to clinical and research implications are discussed.

Practitioner Points and Limitations

- Therapist intervention integrity (i.e., when there is evidence of both adherence and competence) should be a feature of the assessment of therapists in training.
- Services offering interventions for anxiety and depression using cognitive and behavioural interventions should consider routinely monitoring treatment integrity in order to optimise patient outcomes.
- Practitioners should routinely use clinical supervision to assess, maintain and improve treatment integrity.
- The risk of bias tool used in the study was amended which may have reduced the reliability and/or validity of the tool.

Keywords

Therapist Adherence; Competence; Integrity; Fidelity; Outcome; Psychotherapy.

Introduction

A vast body of evidence attests to the effectiveness of psychotherapy (Lambert, 2013) and drives calls to establish empirically supported therapies (EST) demonstrated to be efficacious and effective for specific problems (Chambless & Hollon, 1998). Evidence based practice refers to the delivery of these ESTs in routine practice (Cook et al., 2017). However, there are still large gaps in understanding how psychotherapy leads to therapeutic change (Kazdin, 2007). In addition, so called 'negative effects' during or after psychotherapy, such as deterioration of symptoms, do occur although current understanding of these negative effects is poor (Rozental et al., 2018; Cuijpers et al., 2018).

Defining and Measuring Adherence, Competence and Integrity

Psychotherapy process research focuses on phenomena that occur between and within therapists and clients during therapy (Orlinsky & Howard, 1986), and may contribute to understanding and explaining variations in outcome. These phenomena include the degree to which therapists demonstrate 'adherence' (i.e., the extent to which they are delivering techniques/methods that are congruent with the therapy model and/or protocol) and 'competence' (i.e., the skill and appropriateness with which the techniques/methods are being delivered; Waltz et al., 1993). Additionally, therapist or treatment 'integrity' (also known as 'fidelity'; Miller & Binder, 2002) measures a composite of adherence and competence (Leichsenring et al., 2011). Definitions of integrity do vary across studies; though they often include treatment 'differentiation' in addition to the adherence-competence composite, where differentiation refers to whether treatments can be shown to differ from one another in the clinical manner intended (Kazdin, 1986; Gearing, et al., 2011). Although adherence, competence and integrity (ACI) are conceptually similar constructs, they have been deemed to be distinct

enough to be considered separately in process-outcome meta-analyses (e.g., Collyer et al., 2019). In addition to conceptual clarification, an important consideration is that of measurement, i.e., how are these constructs best measured?

Hogue et al. (1996) described three distinct procedures for assessing process measures such as ACI. These were: 1) quality control procedures carried out prior to and throughout treatment (e.g., regular supervision with an 'expert'), 2) monitoring therapist 'in-session' behaviour for model adherence (e.g., therapist self-report of adherent activities), and 3) observational review of therapy sessions (e.g., audio/video recordings of sessions that are independently rated according to adherence criteria). The third procedure (observational review) is seen as the gold standard method for measuring ACI, albeit the most resource-intensive (Hogue et al., 1996).

Attempts to measure and investigate ACI in psychotherapy have varied widely. In addition to the aforementioned range of procedures used to measure ACI, studies have differed with respect to whether sessions were rated from audio or video recordings, the experience and independence of raters, and the timing and proportion of sessions rated (Hogue et al., 1996). Such methodological and procedural heterogeneity has made it difficult to synthesise studies in order to develop overall conclusions regarding psychotherapy ACI. Despite a lack of overall understanding of ACI in psychotherapy, these constructs still seem important areas to measure and/or monitor.

Why are Adherence, Competence and Integrity Considered Important?

Monitoring and evidencing therapist ACI during psychological interventions is considered to be crucial in the contexts of both clinical trials and routine clinical practice. During clinical trials, for example, an intervention needs to be shown to be delivered as intended (or,

adherently) in order to draw valid conclusions concerning an intervention's efficacy (Perepletchikova et al., 2007). Further, Kazdin (1986) highlighted the need to evidence that therapists were delivering an intervention with sufficient levels of competence in clinical trials, in order to reduce the likelihood of any difference in outcome between interventions being re-interpreted as differences between the competence of therapists. In terms of everyday clinical practice, the importance of therapist ACI has been highlighted in relation to the clinical governance of routine services and design of clinical training courses. This has included recommendations that services ensure resources are available to monitor intervention ACI, including adequate resources to train and supervise clinicians and raters (Gearing et al., 2011). There have also been efforts towards developing empirically derived 'competency frameworks' to support the training and ongoing development of psychological therapists (Roth & Pilling, 2008).

Competency frameworks have been developed by systematically examining the treatment principles and techniques used in efficacy trials, in order to define and extract the key competencies needed to deliver a high quality intervention in routine practice (Roth & Pilling, 2008). An initial competency framework based on cognitive behaviour therapy (CBT) was developed (Roth & Pilling, 2008) as a precursor for other competency frameworks based on other therapy modalities (e.g., psychodynamic psychotherapy; Lemma et al., 2008; cognitive analytic therapy; Parry et al., 2021). These competency frameworks have partly been based on the premise that interventions that are adherent to an 'empirically supported' definition of the key competencies are more likely to lead to a successful therapeutic outcome for the patient. Although the association between ACI and patient outcome is intuitively appealing, studies exploring this association have been inconsistent.

Relationship to Treatment Outcome

In the first meta-analytic review of adherence/competence-outcome conducted, Webb et al. (2010) identified 36 eligible studies. Overall, Webb et al. found that neither the mean-weighted effect size for adherence-outcome (r=0.02), nor competence-outcome (r=0.07) were significantly different from zero. Sub-group analyses indicated that studies controlling for therapeutic alliance had a significantly smaller competence-outcome association, though this was not the case for adherence. The primary methodological limitations of Webb et al.'s review included (a) the search of only a single bibliographic database (PsycINFO), so limiting the range and pool of studies screened for inclusion; (b) only published papers were included which may have over-inflated effect size estimates; (c) no quality appraisal of included studies was carried out and (d) there was no meta-analysis conducted of the association between integrity and outcome.

The Current Review

As a decade has now passed since Webb et al.'s (2010) review, and with the aforementioned limitations in mind, an update and extension is now warranted. Throughout the last decade, more sophisticated statistical approaches to analysing outcomes such as multilevel modelling have grown in popularity. Such methods enable nested hierarchical data structures (such as patients within therapists) to be represented, with higher-level units and variables appropriately modelled (Greenland, 2000). Failure to appropriately model such effects (i.e., nesting of patients within therapists using multilevel models) may lead to confounding and incorrect inferences (Soldz, 2006). In addition, variance structures differ between hierarchical and non-hierarchical models and thus hierarchical and non-hierarchical analyses should be considered separately. An updated review would enable ACI-outcome research applying these

more complex analyses to be meta-analysed alongside the more traditional approaches; ultimately advancing the overall understanding of the association between therapist ACI and patient outcome. A risk of bias assessment of original studies would also enhance understanding, as this was absent in Webb et al. (2020).

To the authors' knowledge, no previous meta-analyses have been conducted on the association between therapist integrity and patient outcome though this appears to be a highly related construct and a recent meta-analysis of studies on children included integrity-outcome studies alongside competence-outcome and adherence-outcome studies (Collyer et al., 2019). Moreover, variations in the ACI-outcome relationship may be in part explained by differences between studies related to patient, treatment and methodological characteristics as shown by the moderator analyses in Webb et al.'s (2010) review. However, moderator analyses are limited by the number of studies available, and interpretation is often hampered by low power (Deeks et al., 2019). An updated and extended review would allow a replication of Webb et al.'s (2010) moderator analyses in sufficiently powered analyses (e.g., diagnosis, treatment modality), as well as investigation of additional variables such as the influence of treatment format and methodological quality that were not previously examined. Finally, unlike Webb et al. (2010) the current study also undertook an evaluation of the quality of the meta-analyses performed in order to provide an indication of the reliability of the findings using the GRADE approach (Grades of Recommendation, Assessment, Development, and Evaluation) system (Atkins et al., 2005).

Aims of the Review

The specific aims of the current review were to: 1) quantify the strength of the relationship between therapist adherence, competence and integrity and treatment outcome by conducting three separate meta-analyses, and 2) assess whether the strength of these

relationships have changed in comparison to Webb et al. (2010), (3) conduct a range of moderator analyses within each meta-analysis.

Method

Pre-Registration

The current review was pre-registered with PROSPERO (ID CRD42020193889).

Literature Search Strategy

Bibliographic Database Searching

Scopus, PsycINFO and MEDLINE were searched in line with pre-defined inclusion/exclusion criteria. No publication date limits were used. The specific search terms and Boolean logic used are displayed in Table 1. Searches were conducted between 18th and 22nd August 2020.

Table 1

Search Terms Used in Bibliographic Database Searching

	Filter 1: Psychotherapy	**Filter 2: Competence**	**Filter 3: Skill**
Search Fields	**Title, abstract or key words**	**Title, abstract or key words**	**Title only**
Combined using OR ↑	"psycho* therap*" "psychotherap*" "CBT" "cognitive therap*" "behavio* therap*" "counselling" "psychodynamic"	"competen*" "adher*" "fidelity" "integrity"	"skill*" "capab*" "abilit*" "competen*" "adher*" "fidelity" "integrity" "capacity" "proficiency" "techni*" "process" "therapist success" "treatment integrity"
Combined using AND →			

Hand Searching/Unpublished Literature Strategy

In addition to bibliographic database searches, forward and reverse citation searches were carried out on all included studies. Unpublished literature was searched by contacting primary authors of all included studies (if correspondence email addresses were given) in addition to searching the Grey Matters, Ethos and ProQuest: Dissertations and Theses databases. Authors were given a period of two weeks to reply for their unpublished manuscripts to be included in the review. Reference lists of previous, similar systematic literature reviews were also hand-searched for studies meeting the criteria for eligibility.

Inclusion and Exclusion Criteria

Study Inclusion Criteria

1. Included an adult (18+ years old) clinical population
2. Investigated face-to-face individual or group psychotherapy
3. Utilised trained researcher/expert-rated quantifiable measures of therapist adherence/competence/integrity based on videotaped, audiotaped, or transcribed therapy sessions
4. Used quantitative measures of treatment outcome
5. Measured treatment outcome after psychotherapy has ceased
6. Conducted statistical analyses on the relationship between adherence/competence/integrity and treatment outcome
7. Were published in English

Exclusion Criteria

1. Case studies
2. Self-rated or patient-rated therapist competence
3. Internet or telephone-based (non-face-to-face) psychotherapy
4. Competence rating not based on a patient therapy session (e.g., competence established via role-plays or other proxy evaluations)

Independent Screening

The titles and abstracts of a sub-sample (20%) of papers were screened by an independent reviewer (MSB). Similarly, the full texts of a further sub-sample (20%) were screened by the

same independent reviewer (MSB). The levels of agreement were very high for both title/abstract (98.66%) and full text (100%) screening.

Data Extraction

Data from the included studies was extracted by the primary author using a bespoke data extraction tool that was piloted and found to be appropriate. Data extracted included the primary author, year of publication, relevant effect size statistics and whether competence, adherence, integrity or fidelity was measured (according to the authors' description). Additional extracted information included the type of therapy investigated, problem targeted, patient and therapist sample size and demographic information (percentage female and mean age), and the name of the outcome measure used. In terms of the ACI rating procedures, the percentage of sessions rated was extracted along with whether the whole or a subscale of the measure was used.

In order to check the reliability of extraction procedures and the extraction tool, a sub-sample (10%) of included studies were extracted by a second independent Clinical Psychologist reviewer (CS) who was blind to the primary author's extractions. This confirmed reliability of the procedures and tool.

Risk of Bias

An adapted version of the Newcastle-Ottawa Scale (NOS): cohort studies (Wells et al., 2000) was used to assess risk of bias (RoB) in the included studies. A more recently developed RoB tool (Risk Of Bias In Non-randomized Studies of Interventions tool; ROBINS-I) has been developed by the Cochrane group (Sterne et al., 2016). The ROBINS-I, however, was deemed less appropriate for use in the current review particularly given the review's focus on within-group analyses as opposed to comparing two or more intervention groups, which is the focus of

the ROBINS-I (Sterne et al., 2016). Before the ROBINS-I was developed, the Cochrane handbook recommended the NOS as the preferred RoB tool for non-randomised studies (Higgins & Green, 2011).

The original 8-item NOS (Appendix A) assesses three features of methodological quality and thus RoB: 'selection' (4 items), 'comparability' (1 item), and 'outcome' (3 items). In the modified version, the Comparability item ('comparability of cohorts on the basis of the design or analysis') and question 2 of the Selection section of the scale ('selection of the non-exposed cohort') were removed as they were concerned with control conditions which were not analysed by any of the studies. Furthermore, question 3 ('adequacy of follow-up') in the Outcome section was modified so that the follow-up adequacy referred to only the sample for which there were ACI-ratings. These changes were made to enhance compatibility between the tool and the types of studies being assessed. The RoB scoring system was also amended accordingly (see Appendix B). Studies were assessed on a scale of 0-6 (higher scores suggested lower RoB); and categorised as low risk of bias (score of 3 for Selection and 2-3 for Outcome), moderate risk of bias (2 for Selection and 2-3 for Outcome) or high risk of bias (0-1 for Selection and 0-1 for Outcome). A randomly selected subset (30%) of included studies were chosen for RoB assessment by a second rater (CS) blind to the primary rater's ratings. The level of agreement was 'substantial' (Kappa = 0.65) according to Landis and Koch (1977).

GRADE

The quality of evidence for each meta-analytic comparison was also assessed using the GRADE approach (Atkins et al., 2005). Three reviewers (NP, SK and MSB) assessed each meta-analysis against five criteria; limitations of included studies, level of imprecision in estimates, amount of unexplained heterogeneity, indirectness of the evidence, and publication bias,

downgrading or upgrading the level of evidence quality accordingly and by consensus (either high, moderate, low or very low). The GRADE procedure therefore provided an indication of the degree of confidence that can be placed in the meta-analytic results.

Effect Size Extraction and Calculation

For studies that reported a bivariate correlation coefficient to represent the association between ACI and treatment outcome, this effect size was directly extracted and used in the meta-analysis. In cases where a study reported an alternative effect size statistic, a correlation coefficient (*r*) was calculated using the Lenhard and Lenhard (2016) effect size conversion calculator. In cases where a correlation coefficient could not be calculated, a partial correlation was calculated using the relevant standardised regression coefficient (Beta), standard error and number of observations from a regression model. The Meta-Essentials Workbook 6 version 1.5 (Suurmond et al., 2017) was used to calculate partial correlations. If the standard error was not reported, 95% confidence intervals of Beta were used to calculate the standard error (see Appendix C for equation used). If the statistics reported in the article were not sufficient for a correlation nor a partial correlation coefficient to be calculated, the author was contacted to request sufficient data. If the author did not provide a correspondence email address in the article or they did not respond within two weeks, the study was excluded on the basis of insufficient statistical information. Where necessary, the direction of the effect size was switched so that a positive correlation coefficient represented a positive association between therapist ACI and a therapeutic client outcome/clinical improvement.

In cases where multiple effect sizes were reported, a preference hierarchy was used so that each sample only contributed one effect size. If several effect sizes representing ACI-outcome associations were reported due to several different outcome measures being used, the

outcome measure hierarchy was used. The outcome hierarchy was as follows: 1) independent/clinician-rated outcome measures were given preference over self-report outcome measures, 2) problem-specific outcome measures relevant to the target problem (e.g. a depression measure for a depressed patient sample) were given preference over more generic outcome measures (e.g. psychological distress), 3) self-report outcome measures were given preference over 'behavioural' outcome data (e.g. number of self-harm incidents over the past week).

If several effect sizes were reported due to ACI being measured at numerous time-points, the time-point closest to the cessation of therapy was given preference. In cases where outcomes were measured at numerous time-points, the outcome taken closest to the cessation of therapy was given preference (e.g., one-month follow-up outcomes would be given preference over 6-month follow-up outcomes). If several effect sizes were reported due to several sub-scales of the ACI measure being reported, the total ACI association was extracted. If no total was reported, the sub-scale most relevant to the type of therapy being delivered was extracted (e.g., a cognitive-behavioural subscale would be used for CBT). Finally, if no sub-scale had higher relevance to the type of therapy, the mean correlation coefficient of the numerous sub-scale effect sizes was calculated to represent the overall association between ACI and outcome.

Meta-Analytic Strategy

Effect sizes extracted from original studies were first grouped according to whether the ACI-outcome analysis was based on a 'hierarchical' analysis (e.g., patients nested within therapists) or 'non-hierarchical' analysis in the original studies. Second, effect sizes were grouped according to whether they represented an adherence-outcome, competence-outcome, or integrity-outcome association. Random-effects models were used due to the methodological

differences between the studies (Borenstein et al., 2011). Consistent with previous guidance, meta-analyses were only carried out on groupings of at least two studies (Valentine et al., 2010). All analyses were carried out using Meta-Essentials Workbook 5 (correlational data), version 1.5 (Suurmond et al., 2017). Forest plots were produced in R (version 3.6.1) using the package *forestplot*.

The extent of heterogeneity across studies was estimated using both the '*Q*' and 'I^2' statistics. A statistically significant '*Q*' value suggests that inconsistency across studies exceeds what would be expected by sampling error alone (Higgins et al., 2003). As it has been previously highlighted that '*Q'* may be susceptible to the number of studies in an analysis (Higgins et al., 2003), I^2 was also calculated to indicate the extent of variability across studies that is due to heterogeneity rather than chance (Higgins et al., 2003). I^2 is represented by a percentage with 0% indicating no observed heterogeneity and higher percentages indicating higher heterogeneity, with 25%, 50% and 75% being suggested to indicate low, medium and high heterogeneity, respectively (Higgins et al., 2003).

Moderator Analyses

Moderator analyses were carried out to explore heterogeneity across studies (Borenstein et al., 2011). Categorical and continuous study characteristics were analysed. For categorical moderator analyses, a subgroup analysis was conducted when at least 10 studies were eligible for inclusion and there were at least two studies in each subgroup (Deeks et al., 2019). For moderator analyses where the moderator was continuous, a meta-regression was conducted when at least six studies were eligible for inclusion in the specific meta-regression.

The review largely replicated Webb et al.'s (2010) moderator analyses, with some adjustments. The subgroup analyses specified a priori were 1) the primary problem targeted, 2)

the treatment modality, 3) treatment format (i.e., individual or group), 4) timing of the ACI measure in therapy (i.e., early, middle, or late), and 5) alliance-confound (i.e., whether or not the study statistically controlled for therapeutic alliance). In the 'timing of the ACI subgroup analysis, if various/random sessions were rated, the study was excluded from the analysis.

The meta-regression analyses specified a priori tested whether the effect sizes in each meta-analysis varied as a function of 1) the year of publication, and 2) percentage coverage of ACI ratings. A post-hoc meta-regression was also conducted to test whether effect sizes varied as a function of 3) risk of bias assessment. In order to extract the relevant data to calculate the percentage coverage of ACI ratings, a preference hierarchy was used. In studies where a standardised number of therapy sessions were delivered, the percentage was simply the proportion of sessions across which the ACI-ratings were taken. If the therapy durations varied across patients, the mean number of sessions delivered to patients in the study was used. If therapy durations varied and no mean therapy-duration was reported, the maximum number of therapy sessions provided in the study was used to calculate the percentage coverage. If no data was reported regarding therapy duration, the study was excluded from the 'percentage coverage' meta-regression.

As the number of sub-group analyses increases, the likelihood of finding false-positive significance tests increases (Deeks et al., 2019). The level of significance used in moderator analyses was therefore adjusted using Bonferroni corrections for the subgroup and meta-regression analyses separately. False positives were also protected against through the moderator analyses being pre-specified in the pre-registered protocol which was recommended by Thompson and Higgins (2002).

Publication Bias

Publication bias was explored via visual symmetry analysis of funnel plots (Light & Pillemer, 1984) and examined statistically using Egger's regression test (Egger et al., 1997).

Results

Study Selection

A summary of the search strategy and flow of studies is displayed in Figure 1 (adapted from Moher et al., 2009). Electronic database searches identified 7,098 records, after duplicates were removed. A further 31 records were found through hand searching and unpublished literature searching procedures. A total of 7,129 records were therefore screened by title and abstract. Of the 7,129 records, 6,850 were removed (96.51%). The remaining 279 full texts were screened against eligibility criteria and a further 217 records were removed (see Appendix D for titles of studies excluded at full text stage). In total, 62 individual articles were eligible for inclusion in the review which included a total of 90 effect size statistics (i.e., 39 competence-outcome, 43 adherence-outcome, and 8 integrity-outcome effect sizes). Figure 2 summarises the four effect size groupings used in the analysis.

Summary Study Characteristics

Characteristics of included studies are outlined in Table 2 (i.e., non-hierarchical studies) and Table 3 (i.e., hierarchical studies). Characteristics of studies were summarised due to the relatively high number of individual studies and effect sizes included in the meta-analyses. Individual study characteristics can be viewed in Appendix E.

In terms of patient samples included in the review overall; 50% (45 samples) were from the United States of America (USA). The remaining samples were from countries including Germany (12, 13.33%), United Kingdom (11, 12.22%) and Canada (9, 10%). Across all samples,

the mean patient sample size was 95.97 (*SD*=162.53) and the mean therapist sample size was 16.56 (SD=28.38). The most common therapy type delivered to the samples was CBT (27, 30%) with depression (24, 26.67%) being the most treated patient problem. In terms of the proportion of therapy sessions rated, mean coverage was 35.11% (SD=35.82) of all sessions. Similar proportions of ratings were based on audio (42, 46.67%) and video (43, 47.78%) recordings of sessions. The few remaining ratings were based on coded transcripts (2, 2.22%) or did not report the procedure (3, 3.33%). A minority of studies controlled for therapeutic alliance (11, 12.22%).

Risk of Bias Assessment

The risk of bias (RoB) assessment summaries are reported in Tables 2 and 3. For non-hierarchical samples, the majority were rated as low RoB (66.67%, 57.69% and 100.00% for adherence, competence and integrity samples, respectively). Studies were generally of lower RoB across adherence-outcome samples (66.67%-85.71% rated low RoB) compared to competence-outcome samples (50.00%-57.69% rated low RoB). All non-hierarchical integrity-outcome studies were rated low RoB, though only four studies were in this group. For hierarchical samples, the same proportion of competence-outcome studies scored low RoB as scored high RoB (four studies each). Mean (*SD*) RoB assessments for non-hierarchical studies were 4.81 (0.79), 4.65 (0.63) and 5.25 (0.5) for adherence, competence and integrity studies respectively. Hierarchical studies had ratings of 4.86 (0.38), 4.25 (1.16) and 4.67 (0.58) for adherence, competence and integrity studies respectively. Consistent methodological strengths across studies included the representativeness of samples and adequacy of follow-up outcome measures. A consistent weakness was that assessment of outcome was mostly self-report as opposed to independent/blind assessment. RoB assessments of individual studies can be viewed in Appendix F.

Figure 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Diagram Summarising Screening Procedure

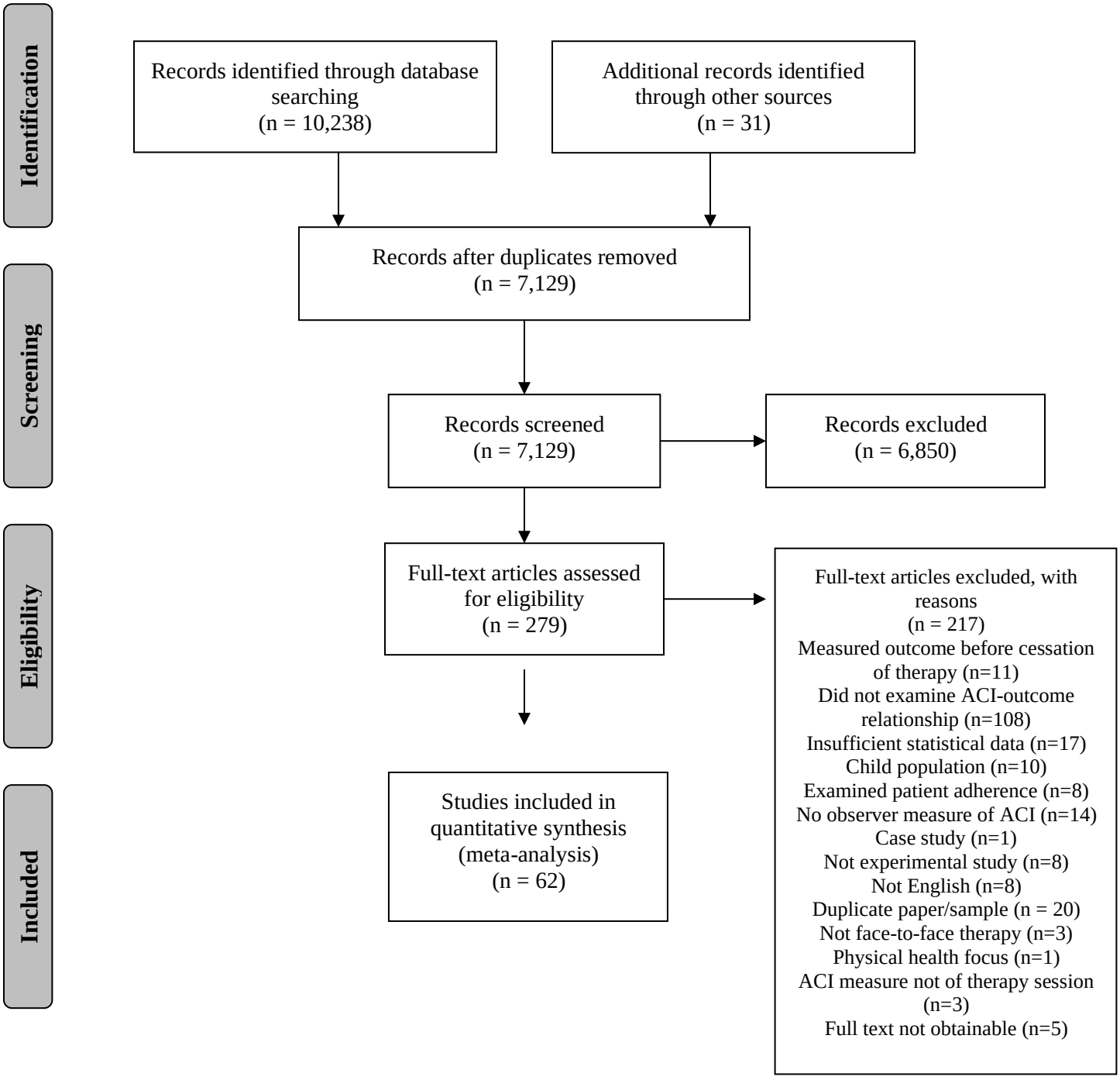


Figure 2
Breakdown of Effect Size Groups

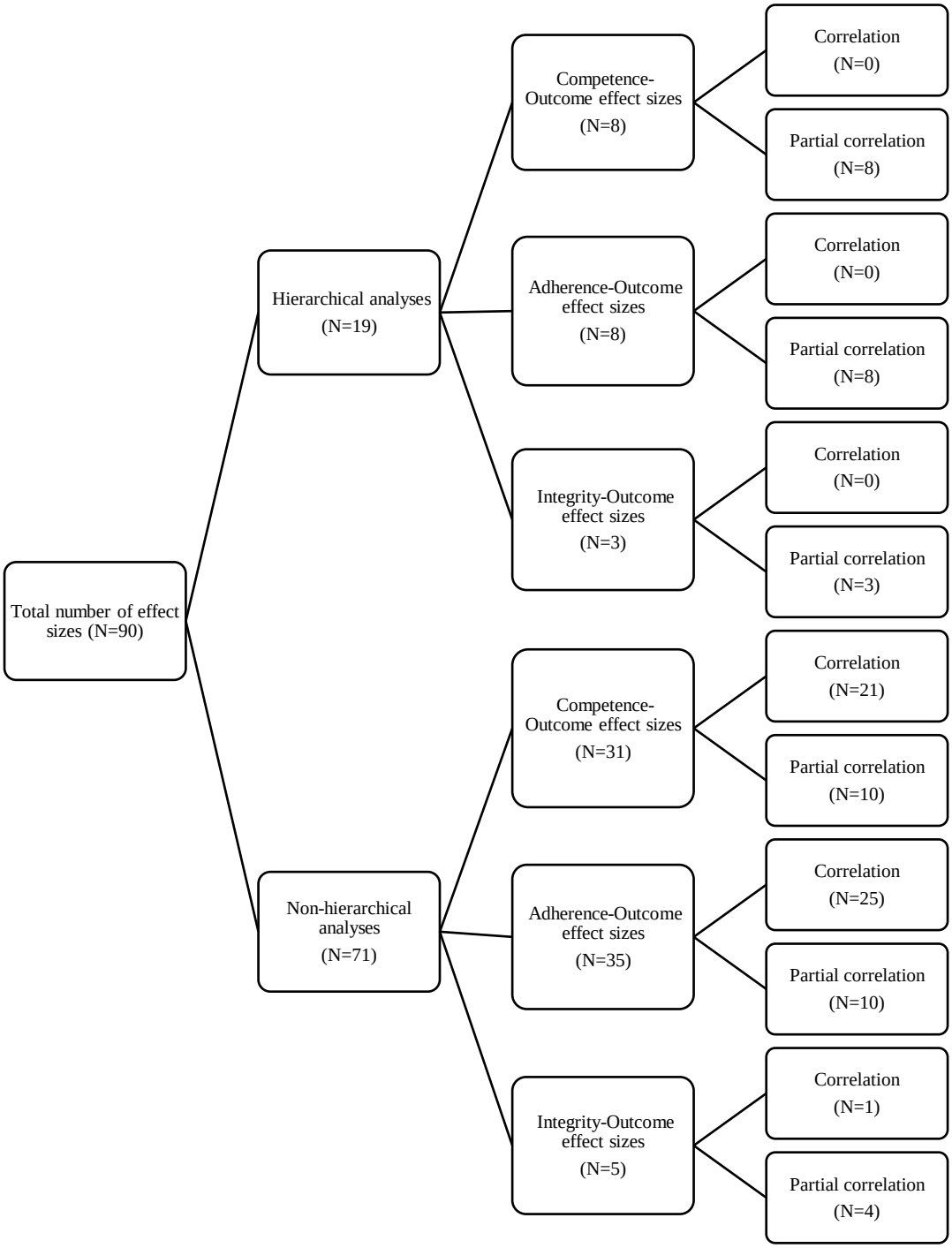


Table 2

A Summary of Study Characteristics for Studies Using Non-Hierarchical Analyses to Investigate Process-Outcome Associations

Meta-Analysis	Patient N	Therapist N	Treatment Modality	Problem Targeted/Diagnosis	Patient Demographics	Therapist Demographics	% coverage (per patient)	Aspect of measure used	Type of session data rated	Alliance controlled	Country	RoB assessment
Non-Hierarchical Adherence-Outcome (*k*=27, samples *n*=35)	57.14 (77.49)	12.09 (8.01)	Psychodynamic (17.14%), SEDP (5.14%), CBT (20.00%), CT (11.43%), MI/MET (8.57%), Mixed (5.71%), EFT (8.57%), IPT (5.71%). All remaining were 2.86%: STIP, STSIP, BAP, IET, BRT, ET.	Anxiety Disorders (20.00%), Substance Use/Addiction (17.14%), Depression (25.71%), Mixed (25.71%), Trauma/PTSD (8.57%), BPD (2.86%)	58.78% female, mean age 36.33 (5.67)	62.66% female, mean age 36.39 (5.17)	25.98%	Sub-scale (51.43%), Whole scale (48.57%)	Transcripts (2.85%), Audio recording (45.71%), Video recording (51.43%)	Yes (5.71%), No (94.29%)	USA (68.57%), Germany (17.14%), Canada (14.29%),	Low (66.67%), Moderate (29.63%), High (3.70)
Non-Hierarchical Competence-Outcome (*k*=26, samples *n*=31)	129.45 (247.46)	21.94 (44.49)	Psychodynamic (9.68%), CBT (38.71%), CT (12.90%), MI/MET (6.45%), EFT (9.68%). All remaining were 3.23%: CPT, Mixed, Experiential/client-centred, IPT, STAPP, ET, Schema-Focused Therapy	Anxiety Disorders (16.13%), Substance Use/Addiction (12.90%), Depression (25.81%), Mixed (32.26%), Trauma/PTSD (12.90%)	52.07% female, mean age 40.25 (5.94)	72.11% female, mean age 35.81 (5.17)	31.92%	Sub-scale (25.81%), Whole scale (74.19%)	Transcripts (3.23%), Audio recording (48.39%), Video recording (45.16%), NR (3.23%)	Yes (12.90%), No (87.10%)	USA (35.48%), Germany (16.13%), Canada (12.90%), Norway (6.45%), UK (25.81%), Australia (3.23%)	Low (57.69%), Moderate (42.3%), High (0.00%)
Non-Hierarchical Integrity-Outcome (*k*=4, samples *n*=5)	199.8 (156.86)	21.33 (19.01)	MI/MET (60.00%), SBNT (20.00%), CBT (20.00%)	Substance Use/Addiction (60.00%), Obesity (20.00%), Anxiety Disorders (20.00%)	42.02% female, mean age 42.66 (10.44)	Mean demographics across studies not calculated as only one study reported	52.50%	Sub-scale (80.00%), Whole scale (20.00%)	Audio recording (40.00%), Video recording (60.00%)	Yes (0.00%), No (100%)	USA (40%), UK (40%), Norway (20%)	Low (100%), Moderate (0.00%), High (0.00%)

Notes. The above table is of effect size samples, not studies. One study/sample may be included across both adherence-outcome and competence-outcome study type as separate meta-analyses were conducted. No participants were included in the same meta-analysis more than once. Some studies did not report particular details and thus the summary characteristics could only be calculated from the studies that did report the relevant information. Risk of bias assessments are the proportions of the individual papers that scored low/moderate/high out of the total number of individual papers in that category (e.g., out of all adherence, non-hierarchical papers). *k* = total number of individual papers in that category, samples *n* = the number of individual samples in a particular category (this may be more than the number of individual studies due to some studies having process-outcome effect sizes for numerous sub-samples (e.g., samples receiving different interventions). SEDP = Supportive Expressive Dynamic Psychotherapy, CBT = Cognitive Behaviour Therapy, CT = Cognitive Therapy, IPT = Interpersonal Psychotherapy, MI/MET = Motivational Interviewing/Motivational Enhancement Therapy, EFT = Emotion Focused Therapy, STAGE-12 = Stimulant Abuser Groups to Engage in 12-Step intervention, STDPP = Short Term Psychodynamic Psychotherapy, STIP = Short-Term Interpretative Psychotherapy, STSIP = Short-Term Supportive Individual Psychotherapy, BAP = Brief-Adaptive Psychotherapy, IET = Interpersonal-Experiential Therapy, MBCT = Mindfulness Based Cognitive Therapy. STAPP = Short-Term Anxiety Provoking Psychotherapy, SBNT = Social Behaviour and Network Therapy, UK = United Kingdom, PTSD = Post Traumatic Stress Disorder, ED = Eating Disorders, BPD = Borderline Personality Disorder, USA = United Stated of America.

Table 3

A Summary of Study Characteristics for Studies Using Hierarchical Analyses to Investigate Process-Outcome Associations

Meta-Analysis	Patient N	Therapist N	Treatment Modality	Problem Targeted/Diagnosis	Patient Demographics	Therapist Demographics	% coverage (per patient)	Aspect of measure used	Type of session data rated	Alliance controlled	Country	RoB assessment
Hierarchical Adherence-Outcome (*k*=7, samples *n*=8)	81.63 (68.77)	22.43 (18.12)	CBT (37.50%) All remaining were 12.50%: Mixed, STAGE-12, STDPP, CPT, MBCT.	Anxiety Disorders (12.50%), Substance Use/Addiction (12.50%), Depression (37.50%), Mixed (12.50%), Trauma/PTSD (12.50%), ED (12.50%)	64.70% female, mean age 33.50 (7.03)	62.30% female, mean age 48.05 (3.32)	36.42%	Sub-scale (25.00%), Whole scale (75.00%)	Audio recording (37.50%), Video recording (50.00%), NR (12.50%)	Yes (25.00%), No (75.00%)	USA (37.50%), Germany (12.50%), The Netherlands (25.00%), Denmark (12.50%), Switzerland (12.50%)	Low (85.71%), Moderate (14.29%), High (0.00%)
Hierarchical Competence-Outcome (*k*=8, samples *n*=8)	70.88 (36.89)	10.71 (6.24)	Psychodynamic (25.00%), CBT (37.50%) All remaining were 12.50%: STAGE-12, CPT, MBCT	Substance Use/Addiction (12.50%), Depression (50.00%), Mixed (25.00%), Trauma/PTSD (12.50%)	69.40% female, mean age 40.07 (6.67)	89.63% female, mean age 44.67 (6.49)	63.54%	Sub-scale (25.00%), Whole scale (75.00%)	Audio recording (50.00%), Video recording (37.50%), NR (12.50%)	Yes (37.50%), No (62.50%)	USA (50.00%), The Netherlands (12.50%), Switzerland (12.50%), UK (12.50%), New Zealand (12.50%),	Low (50%), Moderate (0.00%), High (50.00%)
Hierarchical Integrity-Outcome (*k*=3, samples *n*=3)	135.00 (89.94)	5.33 (1.53)	MI/MET (66.67%), CPT (33.30%)	Substance Use/Addiction (66.67%), Trauma/PTSD (33.33%)	46.87% female, mean age 51.57 (19.18)	Mean demographics across studies not calculated as only one study reported	56.00%	Sub-scale (66.67%), Whole scale (33.33%)	Audio recording (66.67%), Video recording (33.33%)	Yes (0.00%), No (100.00%)	USA (33.33%), Switzerland (33.33%), Multi-National (33.33%)	Low (66.67%), Moderate (33.33%), High (0.00%)

Notes. The above table is of effect size samples, not studies. One study/sample may be included across both adherence-outcome and competence-outcome study type for instance as separate meta-analyses were conducted. No participants were included in the same meta-analysis more than once. Some studies did not report particular details and thus the summary characteristics could only be calculated from the studies that did report the relevant information. Risk of bias assessments are the proportions of the individual papers that scored low/moderate/high out of the total number of individual papers in that category (e.g., out of all adherence, non-hierarchical papers). *k* = total number of individual papers in that category, samples *n* = the number of individual samples in a particular category (this may be more than the number of individual studies due to some studies having numerous process-outcome effect sizes for different sub-samples). CBT = Cognitive Behaviour Therapy, MI/MET = Motivational Interviewing/Motivational Enhancement Therapy, STAGE-12 = Stimulant Abuser Groups to Engage in 12-Step intervention, STDPP = Short Term Psychodynamic Psychotherapy, MBCT = Mindfulness Based Cognitive Therapy, UK = United Kingdom, PTSD = Post Traumatic Stress Disorder, USA = United Stated of America.

GRADE Assessments

Six meta-analytic comparisons are reported across non-hierarchical and hierarchical analyses for the relationship between adherence, competence and integrity and treatment outcome. GRADE assessments indicating the quality of evidence are provided for each individual comparison. The initial quality of evidence was set as 'high' quality as most included studies involved rigorous assessments of ACI (e.g., were observer-rated using validated tools and involved inter-rater reliability checks) and thus there were no significant limitations that would give a reason to downgrade. A range of gradings were made (from 'very low' to 'high' quality). No comparisons were downgraded on limitations of included studies or indirectness of evidence criteria. The remaining three criteria (unexplained heterogeneity, imprecision in estimates and publication bias) had a mixture of gradings.

Non-Hierarchical Meta-Analyses

Adherence-Outcome Meta-Analysis

The non-hierarchical adherence-outcome meta-analysis from 2,000 patients across 27 studies (k = 35 independent samples) is displayed in Figure 3. The adherence-outcome association was small (r = 0.08, 95% CI [-0.01-0.17], p = 0.069, GRADE = moderate), indicating no significant association between adherence and outcome. A down-grading of quality was made for unexplained heterogeneity. There was evidence of substantial heterogeneity: I^2 = 62.07%; Q(df = 34) = 89.65, p < 0.001. Visual inspection of the funnel plot (see Figure 4a) suggested some asymmetry and thus risk of publication bias, but Egger's regression was non-significant (p = 0.131).

Figure 3

Forest Plot for Non-Hierarchical Adherence-Outcome Meta-Analysis

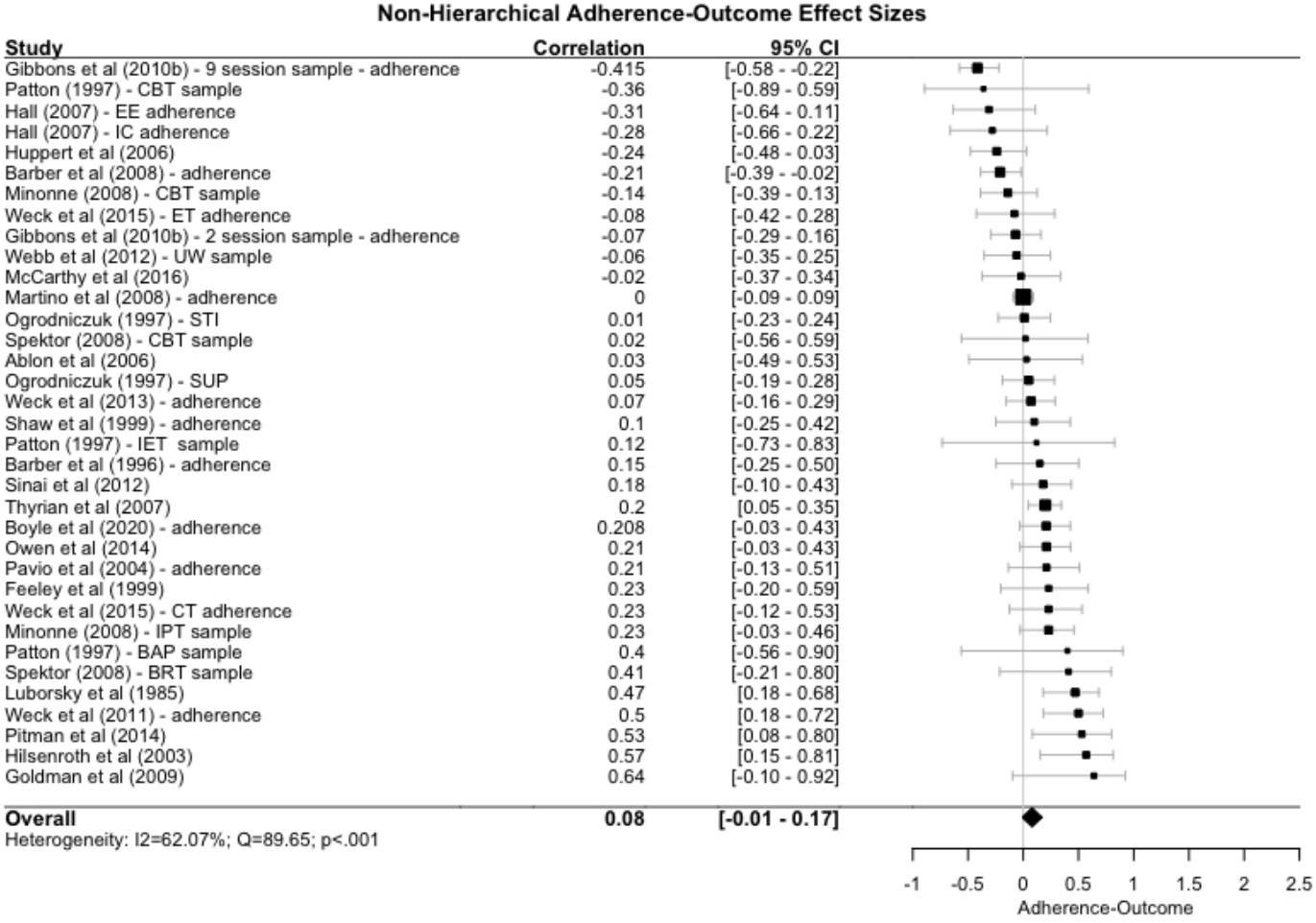
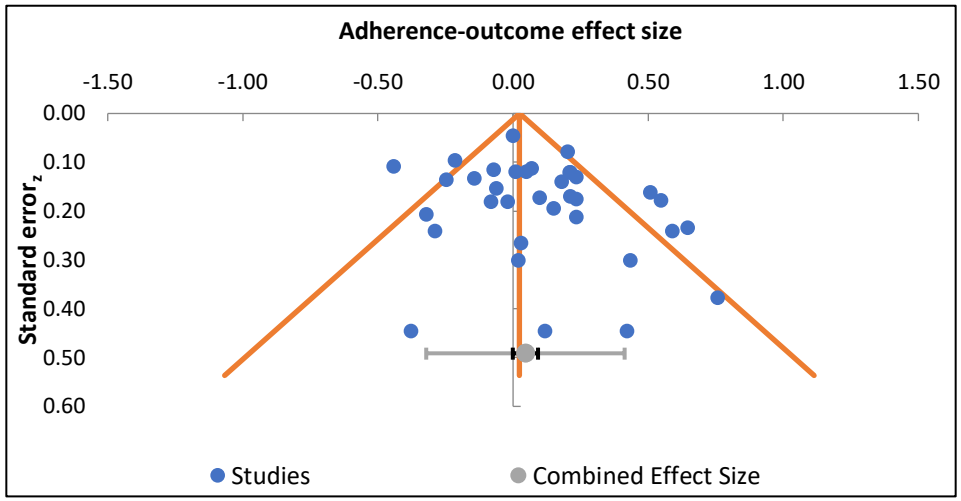


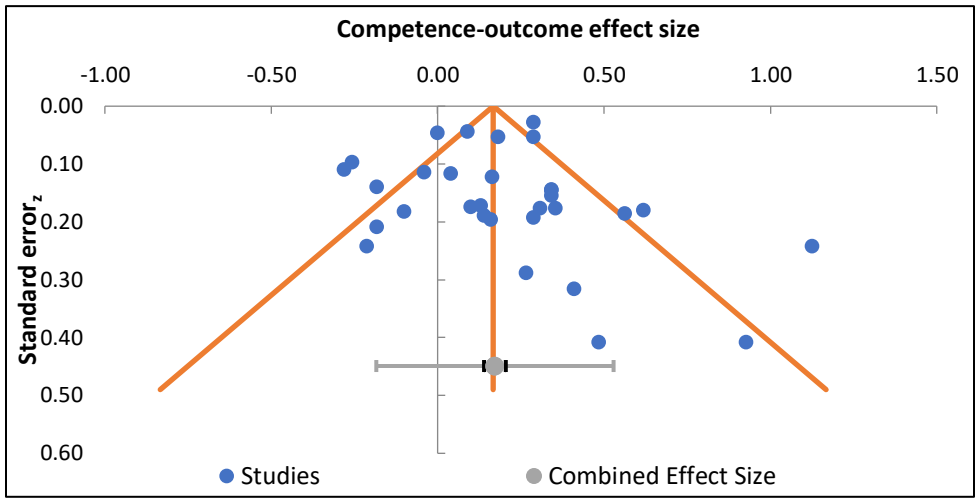
Figure 4

Funnel Plots for a) Adherence-Outcome, b) Competence-Outcome and c) Integrity-Outcome Non-Hierarchical Meta-Analyses

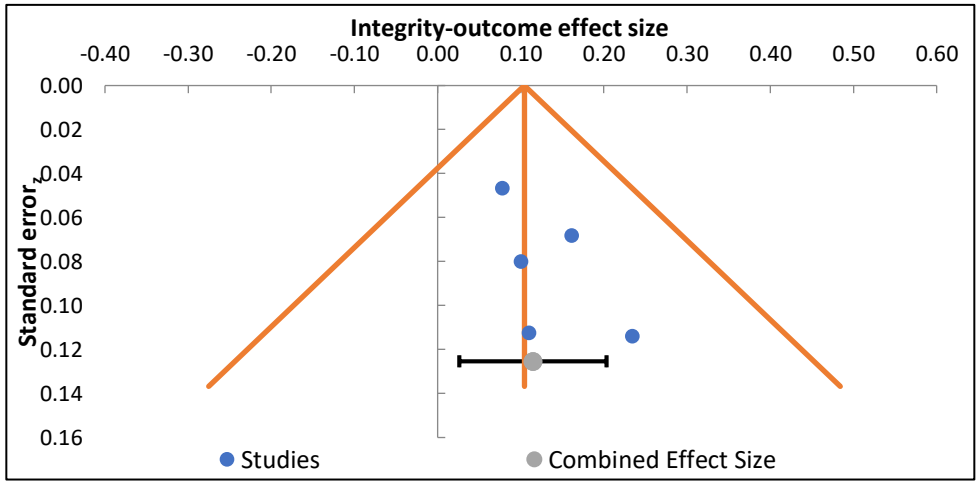
(a)



(b)



(c)

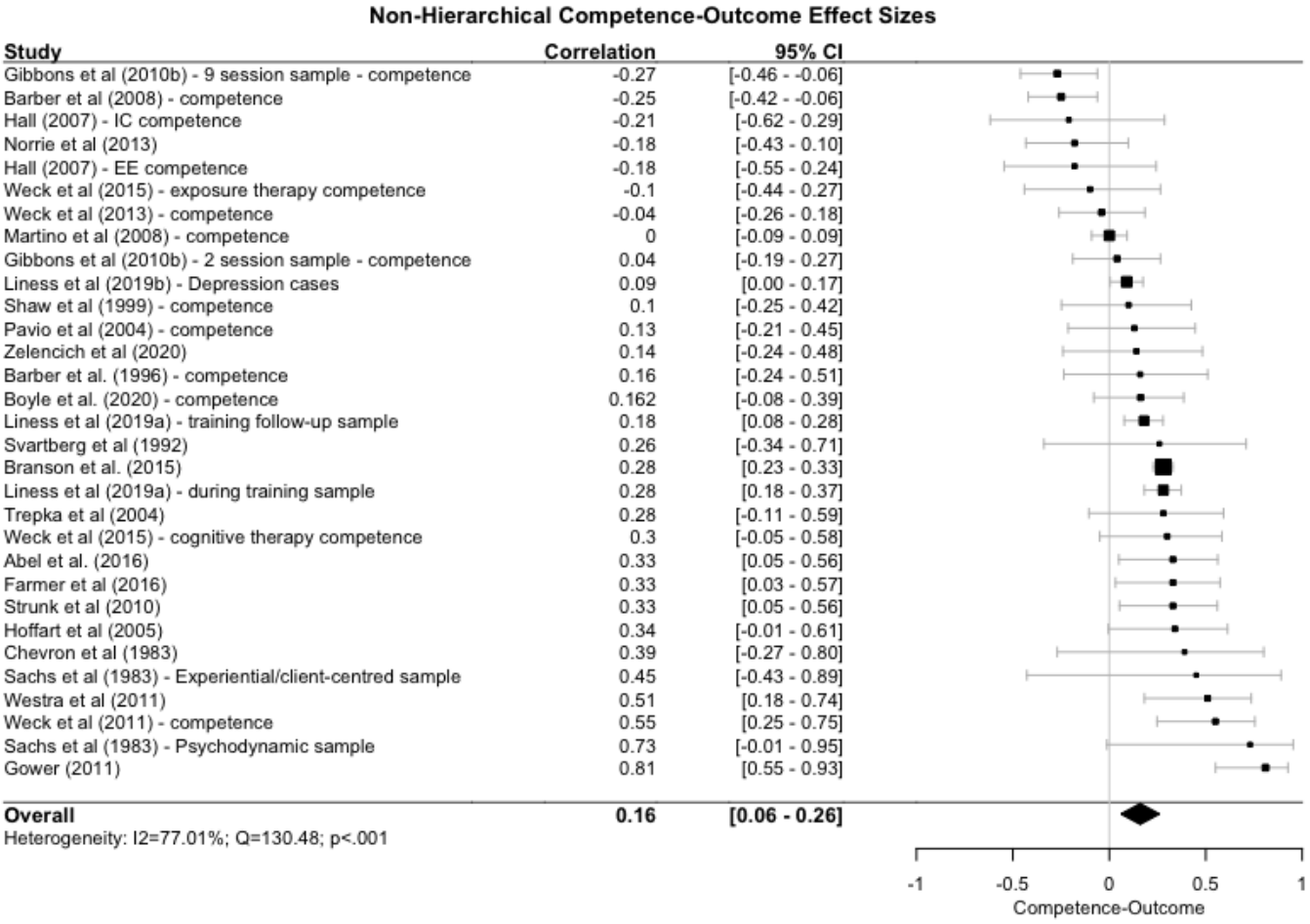


Competence-Outcome Meta-Analysis

The non-hierarchical competence-outcome meta-analysis from 4,013 patients across 26 studies (k = 31 independent samples) is displayed in Figure 5. The overall association was r = 0.16 (95% CI [0.06-0.26], p = 0.001, GRADE = moderate), indicating a small but significant association between competence and outcome. A down-grading of quality was given for unexplained heterogeneity. There was evidence of substantial heterogeneity: I^2 = 77.01%; Q(df = 30) = 130.48, $p < 0.001$. Visual inspection of the funnel plot (see Figure 4b) suggested some asymmetry and thus risk of publication bias. Egger's regression, however, was non-significant (p = 0.917).

Figure 5

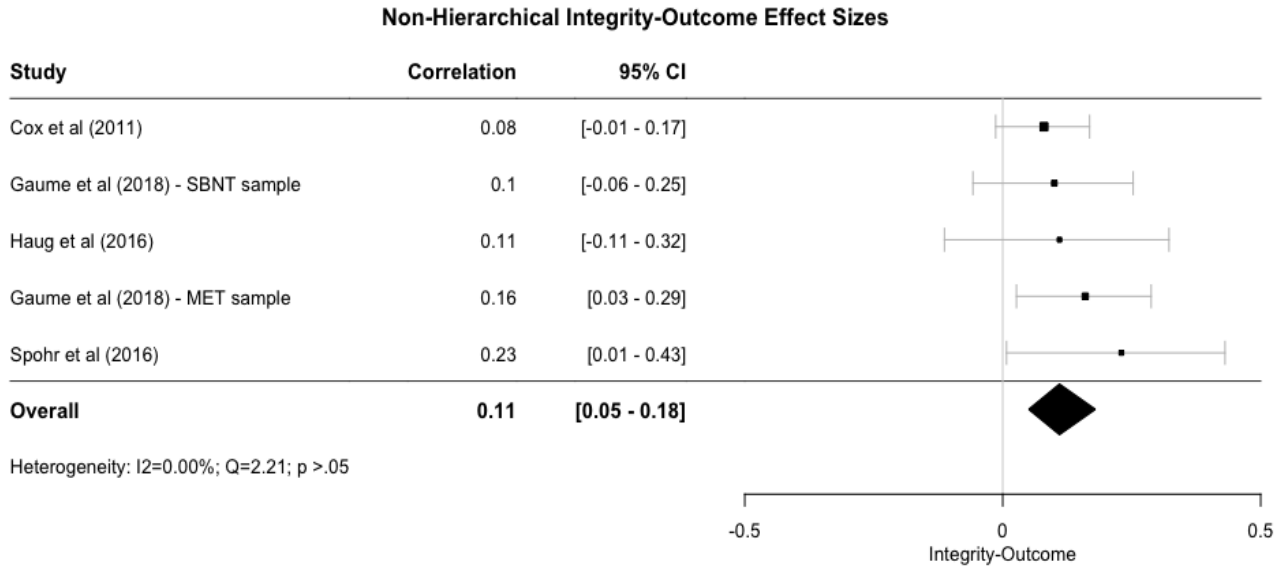
Funnel Plot for Non-Hierarchical Competence-Outcome Meta-Analysis



Integrity-Outcome Meta-Analysis

The meta-analysis summarising integrity-outcome associations from 999 patients across 4 studies (k = 5 independent samples) is displayed in Figure 6. The overall association was small, but significant (r = 0.11, 95% CI [0.05 – 0.18], $p < 0.001$, GRADE = high). There were no down-grading regarding quality. There was little evidence of heterogeneity: I^2 = 0%; Q(df = 34) = 2.21, p = 0.697. Visual inspection of the funnel plot (see Figure 4c) suggested some asymmetry and thus risk of publication bias. Egger's regression, however, was non-significant (p = 0.214).

Figure 6.
Forest Plot for Non-Hierarchical Integrity-Outcome Meta-Analysis



Hierarchical Meta-Analyses

Adherence-Outcome Meta-Analysis

The hierarchical adherence-outcome meta-analysis from 653 patients across 7 studies (*k* = 8 independent samples) is displayed in Figure 7. The adherence-outcome association was small (*r* = 0.04, 95% CI [-0.18-0.25], *p* = 0.679, GRADE = very low), indicating no significant association between adherence and outcome. Quality down-gradings were given for unexplained heterogeneity, imprecision in estimates, and publication bias. There was evidence of substantial heterogeneity: I^2 = 76.04%; *Q*(df = 7) = 29.22, *p* < 0.001. Visual inspection of the funnel plot (see Figure 8a) suggested some asymmetry and thus some risk of publication bias. Egger's regression was also significant (*p* = 0.046).

Figure 7

Forest Plot for Hierarchical Adherence-Outcome Meta-Analysis

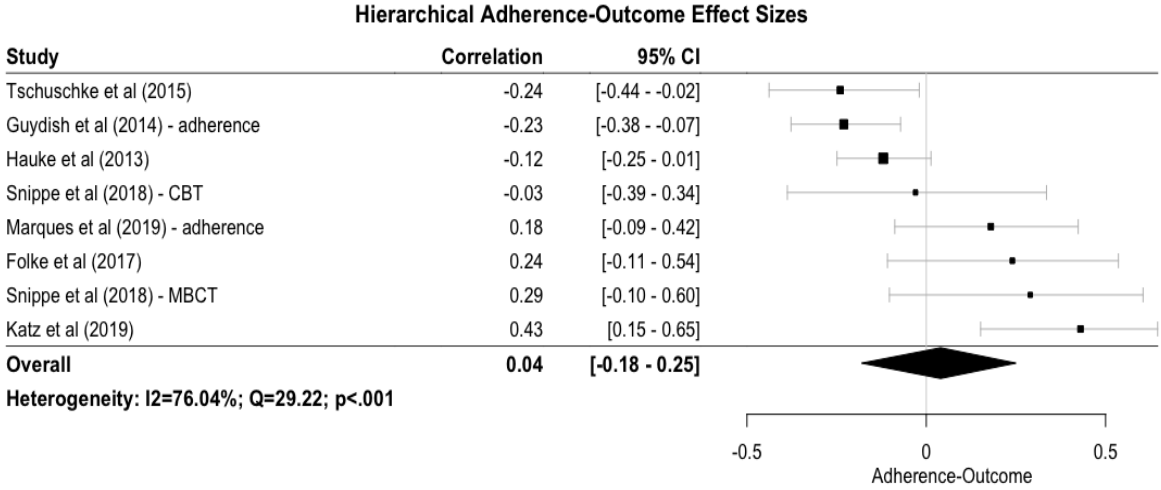
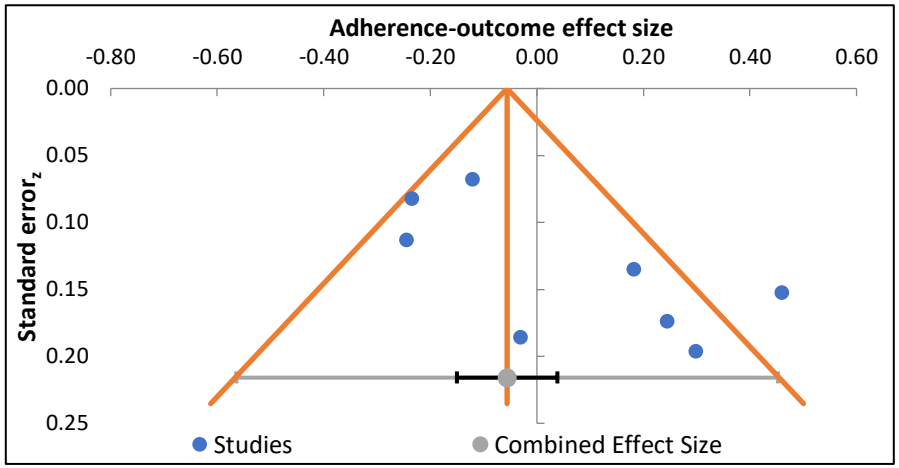


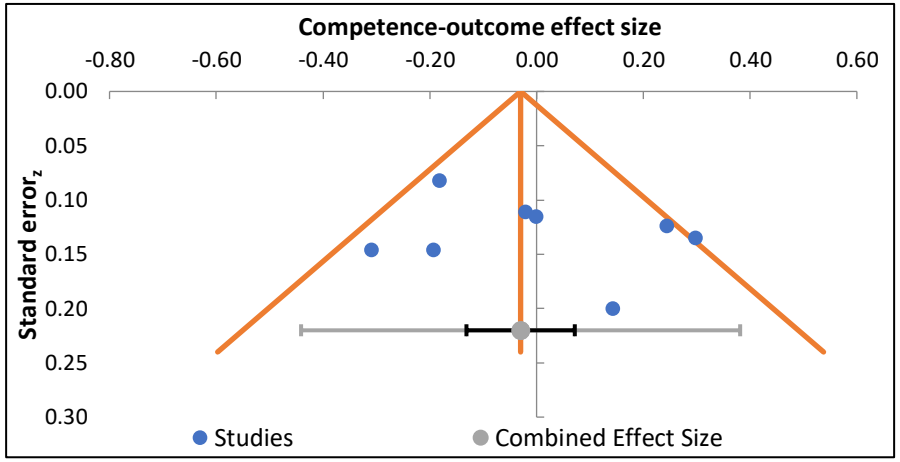
Figure 8

Funnel Plots for a) Adherence-Outcome, b) Competence-Outcome and c) Integrity-Outcome Hierarchical Meta-Analyses

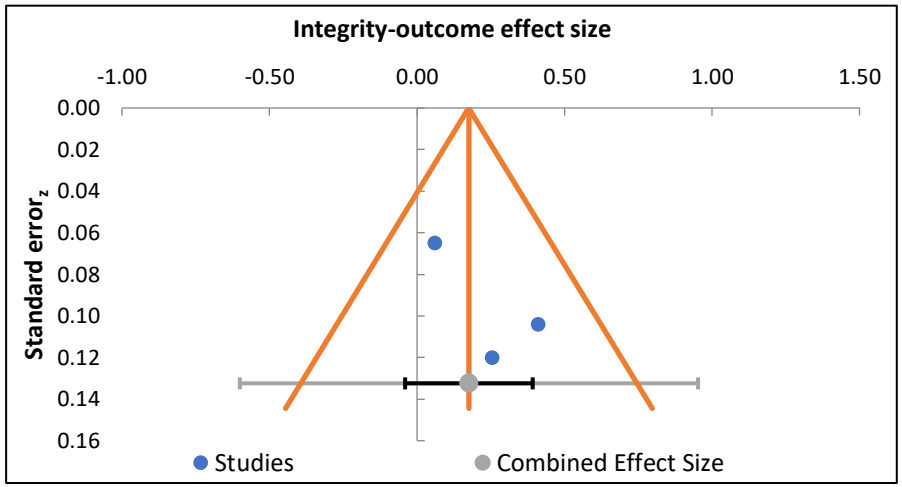
(a)



(b)



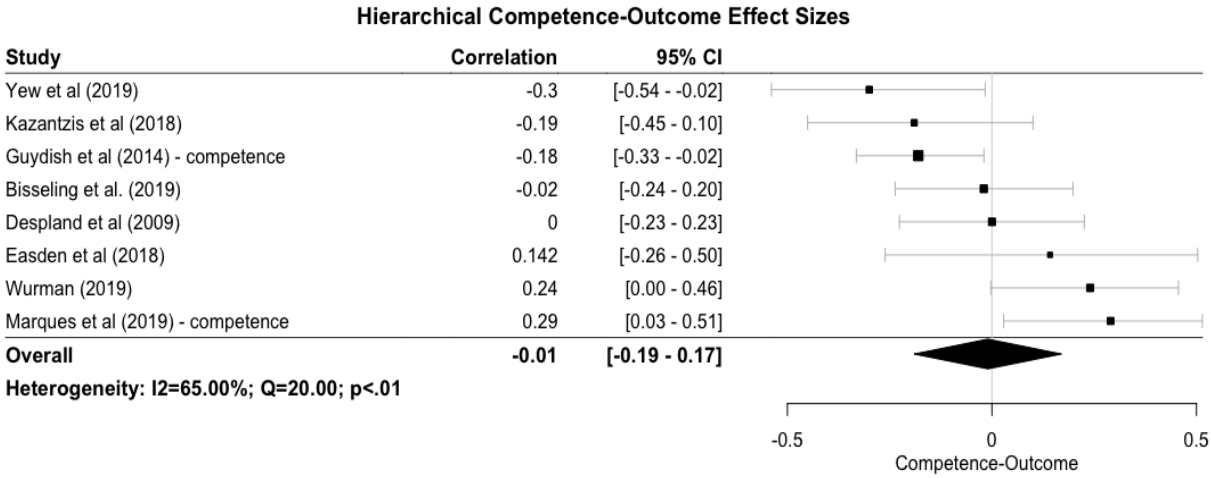
(c)



Competence-Outcome Meta-Analysis

The meta-analysis summarising competence-outcome associations from 567 patients across 8 studies (k = 8 independent samples) is displayed in Figure 9. The overall association was small (r = -0.01, 95% CI [-0.19 - 0.17], p = 0.901, GRADE = low), indicating no significant association between competence and outcome. Quality down-gradings were given for unexplained heterogeneity and imprecision in estimates. There was evidence of considerable heterogeneity: I^2 = 65%; Q(df = 7) = 20, p < 0.006. Visual inspection of the funnel plot (see Figure 8b) suggested some asymmetry and thus risk of publication bias. Egger's regression, however, was non-significant (p = 0.430).

Figure 9
Forest Plot for Hierarchical Competence-Outcome Meta-Analysis

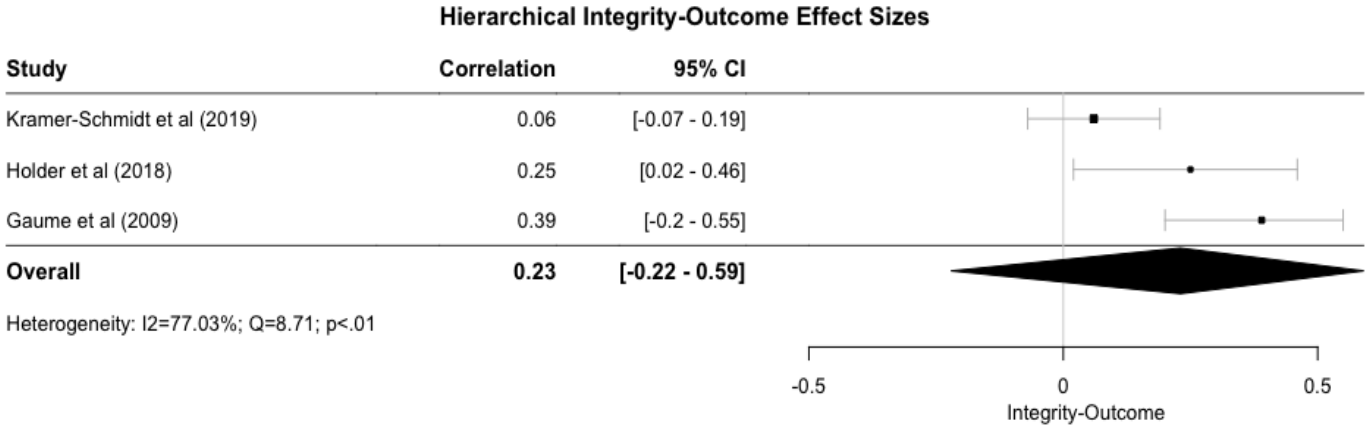


Integrity-Outcome Meta-Analysis

The meta-analysis summarising integrity-outcome associations from 405 patients across 3 studies (k = 3 independent samples) is displayed in Figure 10. The overall association was small, but significant (r = 0.23, 95% CI [-0.22-0.59], p < 0.027, GRADE = low). Quality down-

gradings were given for unexplained heterogeneity and imprecision in estimates. There was evidence of considerable heterogeneity: I^2 = 77.03%; Q(df = 2) = 8.71, p < 0.013. Visual inspection of the funnel plot (see Figure 8c) suggested no asymmetry and thus little risk of publication bias and Egger's regression was also non-significant (p = 0.357).

Figure 10
Forest Plot for Hierarchical Integrity-Outcome Meta-Analysis



Moderator Analyses

Breakdown of Moderator Analyses

In total, 48 individual moderator analyses were planned (based on eight categorical and continuous variables) and studies were again grouped according to whether the effect size was extracted from a hierarchical or non-hierarchical analysis. A total of 18 moderator analyses (6 categorical and 12 continuous) were viable due to sufficient effect sizes being available within the ACI groupings (see Appendix G for a summary of viable/non-viable moderator analyses). Only moderators in the adherence-outcome and competence-outcome effect sizes were possible, with integrity-outcome effect sizes being insufficient in number. No subgroup analyses were viable for any hierarchical studies. A total of 30 moderator analyses were not viable due to

insufficient numbers of studies with data relevant to moderator variables. Within the non-hierarchical adherence-outcome and competence-outcome samples, only three subgroup analyses were viable (problem targeted, treatment modality and alliance confound) out of the five planned, due to insufficient numbers of samples in the other two planned subgroup analyses (treatment format and timing of rating). In the case of the 'timing of rating' moderator, this was not viable due to the analysis planned on the *timing* of the session rated (e.g., early, middle, or late in the intervention). Studies tended to either select sessions randomly or systematically select sessions to rate (e.g., one session from early, middle, and late intervention), and thus there were insufficient numbers of studies only rating early, middle or late sessions.

Non-Hierarchical Moderator Analyses

Subgroup Analyses. Significant heterogeneity between studies was explored using subgroup analyses to investigate three categorical moderators of adherence-outcome and competence-outcome associations for non-hierarchical samples (Table 4). For adherence-outcome samples, no significant variation in effects was found across any of the subgroups investigated. There was a noteworthy difference between the number of studies which controlled for alliance (k=2) and those that did not (k=33). In terms of the competence-outcome samples, after controlling for multiple testing, significant variations in effect sizes were found in the problem targeted and treatment modality subgroups. The strongest positive association between competence and outcome was found when anxiety disorders were being treated, with substance-use/addiction interventions showing a negative competence-outcome association. In terms of treatment modality, cognitive therapy had the strongest positive competence-outcome association, whereas emotion focused therapy had a negative association. No significant variations in effect sizes were found in the alliance confound subgroup, though a similar pattern

to the adherence-outcome analyses was found in that a much higher number of studies did not control for alliance (*k*=27) than those that did (*k*=4).

Table 4
Subgroup Analyses for Non-Hierarchical Adherence-Outcome and Competence-Outcome Effect Sizes

Group	Variable	Subgroup	*k*	Effect Size (*r*)	95% CI	*Q*	I^2 (%)	Diff between subgroups (*p*)
Adherence - outcome								
	Problem Targeted	Anxiety Disorders	7	0.16	-0.12-0.42	18.44**	67.46	0.568
		Depression/ Mood Disorders	9	0.10	-0.06-0.25	11.68	31.53	
		Mixed Problems	9	0.12	0.02- 0.23	5.21	0.00	
		Substance Use/Addiction	6	-0.02	-0.33-0.30	37.52***	86.68	
		Trauma/PTSD	3	-0.11	-0.70-0.58	4.98	59.80	
	Treatment Modality	CBT	7	-0.01	-0.18-0.16	8.70	31.02	0.198
		CT	4	0.23	-0.18-0.56	6.66	54.97	
		EFT	3	-0.11	-0.7-0.58	4.98	59.80	
		IPT	2	0.21	-0.12-0.49	0.07	0.00	
		MI/MET	3	0.05	-0.28-0.37	5.81	65.60	
		Mixed	2	0.03	-1.0-1.0	23.61***	95.76	
		Psychodynamic	6	0.30	-0.02-0.56	9.52	47.49	
		SEDP	2	-0.07	-0.98-0.98	2.77	63.85	
	Alliance Confound	Controlled	2	0.02	-0.64-0.67	0.46	0.00	0.533
		Not Controlled	33	0.09	-0.01-0.18	89.13***	64.10	
Competence - outcome								
	Problem Targeted	Anxiety Disorders	5	0.34	0.00-0.60	9.68*	58.69	**0.002****
		Depression/ Mood Disorders	9	0.25	0.02-0.45	25.68**	68.79	
		Mixed Problems	9	0.21	0.09-0.33	17.32**	53.81	
		Substance Use/Addiction	4	-0.11	-0.36-0.14	10.40*	71.15	
		Trauma/PTSD	4	0.06	-0.35-0.45	6.00	50.02	
	Treatment Modality							

	CBT	12	0.20	0.05-0.35	48.94***	77.52	**<0.001*** **
	CT	4	0.37	0.15-0.55	2.21	0.00	
	EFT	3	-0.05	-0.49-0.42	1.96	0.00	
	MI/MET	2	0.01	-0.17-0.18	0.10	0.00	
	Psychodynamic	3	0.16	-0.84-0.91	10.58**	81.10	
Alliance Confound	Not Controlled	27	0.19	0.08-0.29	106.48***	75.58	0.109
	Controlled	4	-0.01	-0.36-0.35	8.37*	64.15	

Notes. CI = Confidence Interval. CBT = cognitive behaviour therapy, CT = cognitive therapy, EFT = emotion focused therapy, IPT = interpersonal psychotherapy, MI/MET = motivational interviewing/motivational enhancement therapy, SEDP = supportive expressive dynamic psychotherapy. *significant at $p < .05$ threshold, **significant at $p < .01$ threshold, ***significant at $p < .001$ threshold, between subgroup differences significant at Bonferroni adjusted $p < .02$ threshold for multiple testing (in bold).

Meta-Regression Analyses. Significant heterogeneity between studies was also explored using meta-regressions to investigate three continuous moderators for adherence-outcome and competence-outcome associations in the non-hierarchical samples (Table 5). Adherence-outcome associations were significantly more negative as the percentage coverage of ratings increased. There was no significant relationship between year of publication or RoB rating and adherence-outcome associations. In terms of competence-outcome samples; neither year of publication, percentage coverage nor RoB ratings were significantly related to competence-outcome associations.

Table 5
Meta-Regression Analyses for Non-Hierarchical Adherence-Outcome and Competence-Outcome Effect Sizes

Group	Moderator	*k*	B-coefficient	95% CI	SE	*p*
Adherence – outcome						
	Year of Publication	35	0.00	-0.02 to 0.01	0.01	0.440
	Percentage Coverage	31	0.00	-0.01 to 0.00	0.00	**0.019***
	Risk of bias	35	0.05	-0.08 to 0.17	0.06	0.434

Competence – outcome						
	Year of Publication	31	0.00	-0.01 to 0.01	0.01	0.546
	Percentage Coverage	21	0.00	-0.01 to 0.00	0.00	0.271
	Risk of bias	31	0.02	-0.13 to 0.17	0.08	0.787

Notes. CI = Confidence Interval, SE = Standard Error. *significant at $p < .05$ threshold, **significant at $p < .01$ threshold, ***significant at $p < .001$ threshold, moderators significant at Bonferroni adjusted $p < .02$ threshold for multiple testing (in bold).

Hierarchical Moderator Analyses

Meta-Regression Analyses. Meta-regressions to investigate continuous moderators for adherence-outcome and competence-outcome associations in the hierarchical samples are shown in Table 6. In the adherence-outcome samples, more recently published studies were significantly associated with a larger positive adherence-outcome association (after controlling for multiple testing). Percentage coverage of ratings and RoB ratings were not significantly associated with degree of adherence-outcome associations. In terms of competence-outcome samples; neither year of publication, percentage coverage nor RoB ratings were significantly related to competence-outcome associations.

Table 6

Meta-Regression Analyses for Hierarchical Adherence-Outcome and Competence-Outcome Effect Sizes

Group	Moderator	*k*	B-coefficient	95% CI	SE	*p*
Adherence – outcome						
	Year of Publication	8	0.08	0.03 to 0.14	0.03	**0.001****
	Percentage Coverage	6	0.00	-0.01 to 0.00	0.00	0.098
	Risk of bias	8	0.20	-0.49 to 0.88	0.29	0.496

Competence – outcome						
	Year of Publication	8	0.01	-0.04 to 0.06	0.02	0.633
	Percentage Coverage	6	0.00	-0.01 to 0.01	0.00	0.977
	Risk of bias	8	0.10	-0.05 to 0.26	0.07	0.121

Notes. CI = Confidence Interval, SE = Standard Error. *significant at $p < .05$ threshold, **significant at $p < .01$ threshold, ***significant at $p < .001$ threshold, moderators significant at Bonferroni adjusted $p < .02$ threshold for multiple testing (in bold).

Discussion

This systematic review has identified and synthesised published and non-published studies investigating associations between therapist adherence, competence and integrity and patient clinical outcomes. The meta-analysis sought to update and improve upon prior reviews (Webb et al., 2010) by conducting a more thorough search, conducting a risk of bias assessment of the original studies, grouping studies according to the statistical approach (hierarchical versus non-hierarchical), analysing the treatment integrity-outcome relationship and completing an assessment of the quality of synthesised evidence.

In contrast to Webb et al. (2010), the current review found a significant association between competence and outcome in non-hierarchical, but not hierarchical studies. Adherence was not found to be significantly associated with outcome in the current review, consistent with Webb et al.'s findings. As Webb et al. did not explore integrity-outcome associations specifically, these cannot be compared. Mixed support was therefore found for Webb et al.'s previous meta-analytic findings.

Evidence Synthesis

The main results from the primary meta-analyses suggested a small, but significant overall competence-outcome and integrity-outcome association across studies using non-hierarchical analyses. For studies using hierarchical analyses that controlled for between-therapist variability in treatment outcomes, the small association with outcome was only significant for treatment integrity. The lack of any significant association between adherence and outcome is consistent with the review by Webb et al., (2010). However, the evidence of a small significant association between competence and outcome contradicts earlier evidence

synthesised by Webb et al. (2010). When between-therapist variables were controlled for (using hierarchical analyses only) this association was again non-significant.

Integrity has not previously been included in process-outcome meta-analyses; therefore, this review is the first demonstration of a significant small, positive relationship between treatment integrity and outcome. This finding therefore represents an important step in exploring the associations between specific therapist factors and patient clinical outcomes. This finding is particularly robust since (1) it was replicated in both hierarchical and non-hierarchical studies and (2) it remained significant in hierarchical studies that adjust for expected outcome variability between therapists (i.e., therapist effects). Hence, rather than focusing on adherence or competence in isolation, the most consistent impact on treatment outcome appears to come from taking both concepts into consideration and ensuring treatment models are (a) adhered to (b) in a skilful manner. An alternative explanation may be that the integrity-outcome relationship is primarily driven by the influence of competence, which was significant in the meta-analysis of non-hierarchical studies, whereas adherence was not significant. However, this methodological explanation is less plausible in light of the pattern of results in the hierarchical samples that controlled for therapist effects, and where only the integrity-outcome association was significant.

The average RoB assessments across groupings suggested low RoB. In terms of quality of the meta-analytic comparisons, GRADE assessments indicated that three of the meta-analyses had moderate to high reliability (non-hierarchical adherence, competence and integrity), whereas three had very low to low reliability (hierarchical adherence, competence and integrity). These gradings suggested that meta-analyses of the non-hierarchical studies were most reliable. All meta-analyses except non-hierarchical integrity-outcome, however, found significant heterogeneity between studies. Additionally, there was some visual evidence of the influence of

publication bias across all meta-analyses except hierarchical adherence and integrity and therefore results should be interpreted with caution.

Moderator Analyses

Consistent with the review's third aim, a number of moderator analyses were conducted within each meta-analysis. In terms of competence-outcome studies using non-hierarchical analyses; significant variations in effect sizes were found in the problem targeted subgroup and the treatment modality subgroup. The strongest positive associations between competence and outcome were found when anxiety disorders and depression were the problems targeted. The finding that problem targeted is a significant moderator of the competence-outcome association mirrors Webb et al.'s (2010) findings where major depressive disorder had the greatest positive association. It may be that symptoms of both anxiety disorders and depression are particularly responsive to a therapist's level of competence, which may partly contribute to the highly comparable anxiety and depression outcomes when the same treatment modality is delivered to people with anxiety disorders and depression (e.g., Wakefield et al., 2021). In terms of treatment modality, CT and CBT had the strongest positive competence-outcome association with EFT having a negative association. A similar pattern was found in Webb et al.'s (2010) review although this was non-significant in their review. It may be that most CT and CBT studies included in the current review tended to treat different problem sets (e.g., mostly anxiety disorders and depression) compared to those treated with EFT (mostly psychological trauma). Presenting problems therefore may have partly confounded the moderation analysis between treatment modality and competence-outcome association. It is possible, however, that cognitively-focused therapies such as CT and CBT competence may be treatments whose therapeutic benefits are more sensitive to therapist competence compared to other treatments

such as MI/MET. This assertion is supported by Zarafonitis- Müller et al. (2014) who conducted a meta-analysis investigating the relationship between adherence/competence and outcome in CBT specifically. They found that therapist competence had a small but significant association with outcome and that this association was strongest when depression was treated. Similar to the current review, Zarafonitis- Müller et al. (2014) found that therapist adherence was not significantly associated with post-treatment outcome.

Percentage coverage of ratings was a significant moderator of non-hierarchical adherence-outcome associations, suggesting that as the proportion of adherence-rated sessions increased, the relationship between adherence and outcome became weaker. Although this association was non-significant in studies using hierarchical analyses, these studies were markedly lower in number (k=6) compared to non-hierarchical studies (k=31), which likely resulted in less statistical power to detect an association. One explanation for this finding may be that therapist adherence is unstable over the course of a therapy. Strunk, Brotman and DeRubeis (2010) investigated the association between therapist session-to-session adherence and patient session-to-session symptom change in CT for depression. They found that adherence to some aspects of CT (e.g., to negotiating the content of sessions) predicted symptom improvement, whilst adherence to other components (e.g., behavioural methods) did not predict symptom change. This highlights the potential importance of adherence sub-scales and within-patient variation in process variables, neither of which were explored in the current review due to a primary focus on overall ACI-outcome correlations. A surprising finding was that more recently published studies were significantly associated with a stronger positive adherence-outcome association (i.e., greater adherence related to better outcome) in studies using hierarchical analyses. Closer examination of the pool of studies in this moderator analysis

suggests that a confounding variable was patient sample size. For instance, the three studies conducted between 2013 and 2015 had a cumulative sample size of 452 whereas the five remaining studies published in subsequent years had a cumulative sample size of 201. It is likely, therefore, that more recently published studies in the hierarchical adherence-outcome pool of studies had smaller sample sizes and thus may have used less representative samples compared to the slightly older studies. In terms of risk of bias assessments, analyses indicated that the RoB of included studies was not associated with the strength of ACI-outcome associations found.

Limitations and Methodological Improvements

In terms of limitations of the current review; firstly, a large number of moderator analyses (subgroup and meta-regression analyses) were conducted, increasing the risk of type 1 errors (Mascha, 2015). Secondly, studies were classified as adherence/competence/integrity according to the original authors' description of what was being measured in the primary studies. No verification was done to check consistency of definitions used in the primary studies included in the review. It is therefore possible that there was some variation in classification of ACI due to inconsistent definitions being used. If so, this would have reduced validity of the meta-analyses. Thirdly, the risk of bias tool used (NOS) needed to be amended in order to be fit-for-purpose for assessing risk of bias in the original studies. Although this is not unusual in systematic reviews, the reliability and validity of the NOS is therefore open to question. A more overarching limitation of the process-outcome literature being examined in the current review is that many of the studies had rigorous methodological procedures which involved selecting, training and monitoring therapists' delivery of interventions (e.g., analysing data from randomised controlled trials or therapy training courses). This may have reduced the generalisability of the findings to routine clinical practice settings. Further, a limited range of ACI scores may result in a lower

estimated magnitude of process-outcome associations than a more varied and generalisable sample (Bisseling et al., 2019; Liness et al., 2019).

In terms of future studies investigating ACI-outcome associations, the confound of therapeutic alliance should be routinely controlled for. Studies should also attempt to systematically investigate the potential moderating effect of the timing of the measurement of competency/adherence/integrity (e.g., early, middle, late therapy) across treatment in addition to measuring overall intervention competency/adherence/integrity. More studies of ACI and its relationship to outcome based in routine practice would be welcome, particularly in those systems that are supposedly delivering evidence-based practice. This review also highlights the need for studies to employ independent assessments of both ACI and clinical outcome. The particular presenting problems/diagnoses of patients appears to be important when considering the impact of 'therapist effects' (e.g., competence), mirroring previous findings (Johns et al., 2019). Studies should therefore continue to attempt to investigate process-outcome associations with particular client groups (e.g., patients with specific anxiety disorders) as opposed to highly clinically heterogeneous groups. Finally, the findings in the current review were obtained from studies that were primarily conducted in Europe and the USA. How generalisable these findings are to clinical contexts in other Westernised and non-Westernised countries and cultures is currently unknown. Therefore, many more studies of the relationship between ACI are particularly needed in non-Westernised countries.

Clinical Implications

This review highlights the need for: 1) practice-based studies to more consistently measure ACI rather than these solely being a feature of clinical trials, 2) measurement of treatment integrity needs to be integrated in the training of therapists, 3) measurement of

treatment integrity needs to be integrated into ongoing clinical supervision of qualified therapists, 4) commissioning of services needs to be based both on integrity evidence as well as outcome evidence and 5) services offering treatment for anxiety and depression using cognitive and behavioural interventions in particular should consider routinely monitoring therapist integrity to optimise patient outcomes.

Conclusions

A comprehensive search and analysis of the literature examining the association between therapist ACI and patient outcome provides support for the notion that treatment integrity (which combines adherence and competence) is robustly associated with treatment outcomes, particularly for the treatment of depression and anxiety problems. This review highlights the importance of ensuring both adherence and competence are maintained when therapists are delivering interventions in both research and routine practice contexts.

References

References with an asterisk were included in the meta-analyses.

*Abel, A., Hayes, A. M., Henley, W., & Kuyken, W. (2016). Sudden Gains in Cognitive-Behavior Therapy for Treatment-Resistant Depression: Processes of Change. *Journal of Consulting and Clinical Psychology*, *84*(8), 726–737. https://doi.org/10.1037/ccp0000101

Atkins, D., Briss, P. A., Eccles, M., Flottorp, S., Guyatt, G. H., Harbour, R. T., ... & Williams, J. W. (2005). Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. *BMC Health Services Research*, *5*(1), 1-12. https://doi.org/10.1186/1472-6963-5-25

*Ablon, J. S., Levy, R. A., & Katzenstein, T. (2006). Beyond brand names of psychotherapy: Identifying empirically supported change processes. *Psychotherapy*, *43*(2), 216–231. https://doi.org/10.1037/0033-3204.43.2.216

*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*(3), 619–622. https://doi.org/10.1037/0022-006X.64.3.619

*Barber, Jacques P., Gallop, R., Crits-Christoph, P., Barrett, M. S., Klostermann, S., McCarthy, K. S., & Sharpless, B.A. (2008). The role of the alliance and techniques in predicting outcome of supportive-expressive dynamic therapy for cocaine dependence. *Psychoanalytic Psychology*, *25*(3), 461–482. https://doi.org/10.1037/0736-9735.25.3.461

*Bisseling, E. M., Schellekens, M. P. J., Spinhoven, P., Compen, F. R., Speckens, A. E. M., & Lee, M. L. van der. (2019). Therapeutic alliance—not therapist competence or group cohesion—contributes to reduction of psychological distress in group-based mindfulness-based cognitive therapy for cancer patients. *Clinical Psychology and Psychotherapy*, *26*(3), 309–318. https://doi.org/10.1002/cpp.2352

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

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to Meta-Analysis*. John Wiley & Sons.

*Boyle, K., Deisenhofer, A.-K., Rubel, J. A., Bennemann, B., Weinmann-Lutz, B., & Lutz, W. (2020). Assessing treatment integrity in personalized cbt: The inventory of therapeutic interventions and skills. *Cognitive Behaviour Therapy*, *49*(3), 210-227. https://doi.org/10.1080/16506073.2019.1625945

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

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66*(1), 7-18. https://doi.org/10.1037/0022-006X.66.1.7

*Chevron, E. S., & Rounsaville, B. J. (1983). Evaluating the clinical skills of psychotherapists. A comparison of techniques. *Archives of General Psychiatry, 40*(10), 1129–1132. https://doi.org/10.1001/archpsyc.1983.01790090091014

Collyer, H., Eisler, I., & Woolgar, M. (2019). Systematic literature review and meta-analysis of the relationship between adherence, competence and outcome in psychotherapy for children and adolescents. *European Child & Adolescent Psychiatry, 29.* 417-431. https://doi.org/10.1007/s00787-018-1265-2

Cook, S. C., Schwartz, A. C., & Kaslow, N. J. (2017). Evidence-based psychotherapy: Advantages and challenges. *Neurotherapeutics, 14*(3), 537-545. https://doi.org/10.1007/s13311-017-0549-4

*Cox, M. E., Yancy, W. S., Coffman, C. J., Østbye, T., Tulsky, J. A., Alexander, S. C., Brouwer, R. J. N., Dolor, R. J., & Pollak, K. I. (2011). Effects of counseling techniques on patients' weight-related attitudes and behaviors in a primary care clinic. *Patient Education and Counseling*, *85*(3), 363–368. https://doi.org/10.1016/j.pec.2011.01.024

Cuijpers, P., Reijnders, M., Karyotaki, E., de Wit, L., & Ebert, D. D. (2018). Negative effects of psychotherapies for adult depression: a meta-analysis of deterioration rates. *Journal of Affective Disorders*, *239*, 138-145. https://doi.org/10.1016/j.jad.2018.05.050

Deeks, J. J., Higgins, J. P., Altman, D. G. (2019). Analysing data and undertaking meta-analyses. In J.P. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M.J. Page, V.A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (pp. 241-284). John Wiley & Sons. https://doi.org/10.1002/9781119536604.ch10

*Despland, J.-N., Roten, Y. de, Drapeau, M., Currat, T., Beretta, V., & Kramer, U. (2009). The role of alliance in the relationship between therapist competence and outcome in brief psychodynamic psychotherapy. *The Journal of Nervous and Mental Disease*, *197*(5), 362–367. https://doi.org/10.1097/NMD.0b013e3181a20849

*Easden, M. H., & Fletcher, R. B. (2018). Therapist competence in case conceptualization and outcome in cbt for depression. *Psychotherapy Research, 30*(2), 151-169. https://doi.org/10.1080/10503307.2018.1540895

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629-634. https://doi.org/10.1136/bmj.315.7109.629

*Farmer, C. C., Mitchell, K. S., Parker-Guilbert, K., & Galovski, T. E. (2016). Fidelity to the cognitive processing therapy protocol: Evaluation of critical elements. *Behavior Therapy, 48*(2), 195-206. https://doi.org/10.1016/j.beth.2016.02.009

*Feeley, M., DeRubeis, R. J., & Gelfand, L. A. (1999). The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, *67*(4), 578–582. https://doi.org/10.1037/0022-006X.67.4.578

*Folke, S., Daniel, S. I. F., Gondan, M., Lunn, S., Takker, L., & Poulsen, S. (2017). Therapist adherence is associated with outcome in cognitive-behavioral therapy for bulimia nervosa. *Psychotherapy*, *54*(2), 195–200. https://doi.org/10.1037/pst0000107

*Gaume, J., Gmel, G., Faouzi, M., & Daeppen, J.-B. (2009). Counselor skill influences outcomes of brief motivational interventions. *Journal of Substance Abuse Treatment*, *37*(2), 151–159. https://doi.org/10.1016/j.jsat.2008.12.001

*Gaume, J., Heather, N., Tober, G., & McCambridge, J. (2018a). A mediation analysis of treatment processes in the UK alcohol treatment trial. *Journal of Consulting and Clinical Psychology*, *86*(4), 321–329. https://doi.org/10.1037/ccp0000287

Gearing, R. E., El-Bassel, N., Ghesquiere, A., Baldwin, S., Gillies, J., & Ngeow, E. (2011). Major ingredients of fidelity: a review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review*, *31*(1), 79-88. https://doi.org/10.1016/j.cpr.2010.09.007

*Gibbons, C. J., Nich, C., Steinberg, K., Roffman, R. A., Corvino, J., Babor, T. F., & Carroll, K. M. (2010). Treatment process, alliance and outcome in brief versus extended treatments for marijuana dependence. *Addiction*, *105*(10), 1799–1808. https://doi.org/10.1111/j.1360-0443.2010.03047.x

Ginzburg, D. M., Bohn, C., Höfling, V., Weck, F., Clark, D. M., & Stangier, U. (2012). Treatment specific competence predicts outcome in cognitive therapy for social anxiety disorder. *Behaviour Research and Therapy*, *50*(12), 747-752. https://doi.org/10.1016/j.brat.2012.09.001

*Goldman, G. A., & Gregory, R. J. (2009). Preliminary relationships between adherence and outcome in dynamic deconstructive psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, *46*(4), 480–485. https://doi.org/10.1037/a0017947

*Gower, P.J. (2011). *Therapist competence, case conceptualisation and therapy outcome in cognitive behavioural therapy* [Unpublished doctoral thesis]. College of Life and Environmental Sciences, University of Exeter.

Greenland, S. (2000). Principles of multilevel modelling. *International Journal of Epidemiology*, *29*(1), 158-167. https://doi.org/10.1093/ije/29.1.158

*Guydish, J., Campbell, B. K., Manuel, J. K., Delucchi, K. L., Le, T., Peavy, K. M., & McCarty, J. (2014). Does treatment fidelity predict client outcomes in 12-Step Facilitation for stimulant abuse? *Drug and Alcohol Dependence*, *134*, 330–336. https://doi.org/10.1016/j.drugalcdep.2013.10.020

*Hall, I.E. (2007). *Therapist relationship and technical skills in two versions of emotion focused trauma therapy* [Unpublished doctoral thesis]. Faculty of Graduate Studies, University of Windsor.

*Haug, T., Nordgreen, T., Ost, L.G., Tangen, T., Kvale, G., Hovland, O. J., Heiervang, E. R., & Havik, O. E. (2016). Working alliance and competence as predictors of outcome in cognitive behavioral therapy for social anxiety and panic disorder in adults. *Behaviour Research and Therapy*, *77*, 40–51. https://doi.org/10.1016/j.brat.2015.12.004

*Hauke, C., Gloster, A. T., Gerlach, A. L., Richter, J., Kircher, T., Fehm, L., Stoy, M., Lang, T., Klotsche, J., Einsle, F., Deckert, J., & Wittchen, H.-U. (2014). Standardized treatment manuals: Does adherence matter? *Sensoria: A Journal of Mind, Brain & Culture, 10*(2), 1–13. https://doi.org/10.7790/sa.v0i0.362

Higgins, J. P. T., & Green, S. (2011). Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. https://handbook-5-1.cochrane.org/

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*(7414), 557-560. https://doi.org/10.1136/bmj.327.7414.557

*Hilsenroth, M. J., Ackerman, S. J., Blagys, M. D., Baity, M. R., & Mooney, M. A. (2003). Short-term psychodynamic psychotherapy for depression: An examination of statistical, clinically significant, and technique-specific change. *The Journal of Nervous and Mental Disease, 191*(6), 349–357. http://doi.org/10.1097/01.NMD.0000071582.11781.67

*Hoffart, A., Sexton, H., Nordahl, H. M., & Stiles, T. C. (2005). Connection between patient and therapist and therapist's competence in schema-focused therapy of personality problems. *Psychotherapy Research, 15*(4), 409–419. https://doi.org/10.1080/10503300500091702

Hogue, A., Liddle, H. A., & Rowe, C. (1996). Treatment adherence process research in family therapy: A rationale and some practical guidelines. *Psychotherapy: Theory, Research, Practice, Training, 33*(2), 332-345. https://doi.org/10.1037/0033-3204.33.2.332

*Holder, N., Holliday, R., Williams, R., Mullen, K., & Surís, A. (2018). A preliminary examination of the role of psychotherapist fidelity on outcomes of cognitive processing therapy during an RCT for military sexual trauma-related PTSD. *Cognitive Behaviour Therapy*, *47*(1), 76–89. https://doi.org/10.1080/16506073.2017.1357750

*Huppert, J. D., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S.W. (2006). The Interaction of Motivation and Therapist Adherence Predicts Outcome in Cognitive Behavioral Therapy for Panic Disorder: Preliminary Findings. *Cognitive and Behavioral Practice*, *13*(3), 198–204. https://doi.org/10.1016/j.cbpra.2005.10.001

Johns, R. G., Barkham, M., Kellett, S., & Saxon, D. (2019). A systematic review of therapist effects: A critical narrative update and refinement to review. *Clinical Psychology Review*, *67*, 78-93. https://doi.org/10.1016/j.cpr.2018.08.004

*Katz, M., Hilsenroth, M. J., Gold, J. R., Moore, M., Pitman, S. R., Levy, S. R., & Owen, J. (2019). Adherence, flexibility, and outcome in psychodynamic treatment of depression. *Journal of Counseling Psychology*, *66*(1), 94–103. https://doi.org/10.1037/cou0000299

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

Kazdin, A. E. (1986). Comparative outcome studies of psychotherapy: methodological issues and strategies. *Journal of Consulting and Clinical Psychology*, *54*(1), 95-105. https://doi.org/10.1037/0022-006X.54.1.95

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, *3*, 1-27. https://doi.org/10.1146/annurev.clinpsy.3.022806.091432

*Kramer Schmidt, L., Moyers, T. B., Nielsen, A. S., & Andersen, K. (2019). Is fidelity to motivational interviewing associated with alcohol outcomes in treatment-seeking 60+ year-old citizens? *Journal of Substance Abuse Treatment*, *101*, 1–11. https://doi.org/10.1016/j.jsat.2019.03.004

Lambert, M. J. (2013). Outcome in psychotherapy: The past and important advances. *Psychotherapy*, *50*(1), 42–51. https://doi.org/10.1037/a0030682

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, *33*(2), 363-374. https://doi.org/10.2307/2529786

Leichsenring, F., Salzer, S., J Hilsenroth, M., Leibing, E., Leweke, F., & Rabung, S. (2011). Treatment integrity: An unresolved issue in psychotherapy research. *Current Psychiatry Reviews*, *7*(4), 313-321. https://doi.org/10.2174/157340011797928259

Lemma, A., Roth, A.D., & Pilling, S. (2008). The competences required to deliver effective psychoanalytic/ psychodynamic therapy. University College London. https://www.ucl.ac.uk/drupal/site_pals/sites/pals/files/migrated-files/PPC_Clinicians_Background_Paper.pdf

Lenhard, W. & Lenhard, A. (2016). *Computation of Effect Sizes*. Psychometrica. https://doi.org/10.13140/RG.2.2.17823.92329

Light, R. J., & Pillemer, D. B. (1984). *Summing up: the science of reviewing research*. Harvard University Press.

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

*Liness, S., Beale, S., Lea, S., Byrne, S., Hirsch, C. R., & Clark, D. M. (2019b). Multi-professional IAPT CBT training: Clinical competence and patient outcomes. *Behavioural and Cognitive Psychotherapy, 47*(6), 672–685. https://doi.org/10.1017/S1352465819000201

*Luborsky, L., McLellan, A. T., Woody, G. E., O'Brien, C. P., & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry, 42*(6), 602–611. http://doi.org/10.1001/archpsyc.1985.01790290084010

*Marques, L., Valentine, S. E., Kaysen, D., Mackintosh, M. A., Dixon De Silva, L. E., Ahles, E. M., Youn, S. J., Shtasel, D. L., Simon, N. M., & Wiltsey-Stirman, S. (2019). Provider fidelity and modifications to cognitive processing therapy in a diverse community health clinic: Associations with clinical change. *Journal of Consulting and Clinical Psychology, 87*(4), 357–369. https://doi.org/10.1037/ccp0000384

*Martino, S., Ball, S. A., Nich, C., Frankforter, T. L., & Carroll, K. M. (2008). Community program therapist adherence and competence in motivational enhancement therapy. *Drug and Alcohol Dependence, 96*(1-2), 37–48. https://doi.org/10.1016/j.drugalcdep.2008.01.020

Mascha, E. J. (2015). Alpha, beta, meta: guidelines for assessing power and type I error in meta-analyses. *Anesthesia & Analgesia, 121*(6), 1430-1433. https://doi.org/10.1213/ANE.0000000000000993

*McCarthy, K. S., Keefe, J. R., & Barber, J. P. (2016). Goldilocks on the couch: Moderate levels of psychodynamic and process-experiential technique predict outcome in psychodynamic therapy. *Psychotherapy Research*, *26*(3), 307–317. https://doi.org/10.1080/10503307.2014.973921

Miller, S. J., & Binder, J. L. (2002). The effects of manual-based training on treatment fidelity and outcome: A review of the literature on adult individual psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, *39*(2), 184-198. https://doi.org/10.1037/0033-3204.39.2.184

*Minonne, G. A. (2009). *Therapist adherence, patient alliance, and depression change in the NIMH Treatment for Depression Collaborative Research Program* [Unpublished doctoral thesis]. The University of Michigan.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

*Norrie, J., Davidson, K., Tata, P., & Gumley, A. (2013). Influence of therapist competence and quantity of cognitive behavioural therapy on suicidal behaviour and inpatient hospitalisation in a randomised controlled trial in borderline personality disorder: Further

analyses of treatment effects in the BOSCOT study. *Psychology and Psychotherapy*, *86*(3), 280–293. https://doi.org/10.1111/papt.12004

*Ogrodniczuk, J.S. (1997). *Therapist adherence to treatment manuals and its relation to the therapeutic alliance and therapy outcome: scale development and validation* [Unpublished doctoral thesis]. Faculty of Graduate Studies and Research: Medical Sciences, University of Alberta.

Orlinsky, D. E., & Howard, K. I. (1986). The psychological interior of psychotherapy: Explorations with the Therapy Session Reports. In L. S. Greenberg & W. M. Pinsof (Eds.), *Guilford clinical psychology and psychotherapy series. The psychotherapeutic process: A research handbook* (p. 477–501). Guilford Press.

*Owen, J., & Hilsenroth, M. J. (2014). Treatment adherence: The importance of therapist flexibility in relation to therapy outcomes. *Journal of Counseling Psychology*, *61*(2), 280–288. https://doi.org/10.1037/a0035753

*Paivio, S. C., Holowaty, K. A. M., & Hall, I. E. (2004). The influence of therapist adherence and competence on client reprocessing of child abuse memories. *Psychotherapy*, *41*(1), 56–68. https://doi.org/10.1037/0033-3204.41.1.56

*Patton, J.R. (1997). *A study in reliability and validation: what is the nature of the unique and joint contributions of therapist adherence and therapeutic alliance to treatment outcome?* [Unpublished doctoral thesis]. Graduate Faculty in Psychology, The City University of New York.

Parry, G., Bennett, D., Roth, A. D., & Kellett, S. (2021). Developing a competence framework for cognitive analytic therapy. *Psychology and Psychotherapy: Theory, Research and Practice*, *94*(1), 151-170. https://doi.org/10.1111/papt.12306

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, *75*(6), 829. https://doi.org/10.1037/0022-006X.75.6.829

*Pitman, S., Slavin-Mulford, J., & Hilsenroth, M. (2014). Psychodynamic techniques related to outcome for anxiety disorder patients at different points in treatment. *The Journal of Nervous and Mental Disease*, *202*(5), 391–396. https://doi.org/10.1097/NMD.0000000000000137

Roth, A. D., & Pilling, S. (2008). Using an evidence-based methodology to identify the competences required to deliver effective cognitive and behavioural therapy for depression and anxiety disorders. *Behavioural and Cognitive Psychotherapy*, *36*(2), 129. https://doi.org/10.1017/S1352465808004141

Rozental, A., Castonguay, L., Dimidjian, S., Lambert, M., Shafran, R., Andersson, G., & Carlbring, P. (2018). Negative effects in psychotherapy: commentary and recommendations for future research and clinical practice. *BJPsych Open*, *4*(4), 307-312. https://doi.org/10.1192/bjo.2018.42

*Sachs, J. S. (1983). Negative factors in brief psychotherapy: An empirical assessment. *Journal of Consulting and Clinical Psychology*, 51(4), 557. https://doi.org/10.1037/0022-006X.51.4.557

*Shaw, B. F., Elkin, I., Yamaguchi, J., Olmsted, M., Vallis, T. M., Dobson, K. S., Lowery, A., Sotsky, S. M., Watkins, J. T., & Imber, S. D. (1999). Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting and Clinical Psychology*, *67*(6), 837–846. https://doi.org/10.1037/0022-006X.67.6.837

*Sinai, D., Gur, M., & Lipsitz, J. D. (2012). Therapist adherence to interpersonal vs. Supportive therapy for social anxiety disorder. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *22*(4), 381–388. https://doi.org/10.1080/10503307.2012.658452

*Snippe, E., Schroevers, M. J., Tovote, K. A., Sanderman, R., Emmelkamp, P. M. G., & Fleer, J. (2019). Explaining variability in therapist adherence and patient depressive symptom improvement: The role of therapist interpersonal skills and patient engagement. *Clinical Psychology and Psychotherapy*, *26*(1), 84–93. https://doi.org/10.1002/cpp.2332

Soldz, S. (2006). Models and meanings: Therapist effects and the stories we tell. *Psychotherapy Research*, *16*(2), 173-177. https://doi.org/10.1080/10503300500264937

*Spektor, D. (2008). *Therapists' adherence to manualized treatments in the context of ruptures* [Unpublished doctoral thesis]. New School for Social Research, The New School.

*Spohr, S. A., Taxman, F. S., Rodriguez, M., & Walters, S. T. (2016). Motivational Interviewing Fidelity in a Community Corrections Setting: Treatment Initiation and Subsequent Drug Use. *Journal of Substance Abuse Treatment*, *65*, 20–25. https://doi.org/10.1016/j.jsat.2015.07.012

Sterne, J.A.C., Higgins, J.P.T., Elbers, R.G., Reeves, B.C., & the development group for ROBINS-I (2016). Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I): detailed guidance. http://www.riskofbias.info

Strunk, D. R., Brotman, M. A., & DeRubeis, R. J. (2010). The process of change in cognitive therapy for depression: Predictors of early inter-session symptom gains. *Behaviour Research and Therapy, 48*(7), 599-606. https://doi.org/10.1016/j.brat.2010.03.011

*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*(3), 429–437. https://doi.org/10.1037/a0019631

Suurmond, R., van Rhee, H., & Hak, T. (2017). Introduction, comparison and validation of Meta-Essentials: A free and simple tool for meta-analysis. *Research Synthesis Methods*, *8*(4), 537–553. https://doi.org/10.1002/jrsm.1260

*Svartberg, M., & Stiles, T. C. (1992). Predicting patient change from therapist competence and patient-therapist complementarity in short-term anxiety-provoking psychotherapy: A pilot study. *Journal of Consulting and Clinical Psychology*, *60*(2), 304–307. https://doi.org/10.1037/0022-006X.60.2.304

Thompson, S. G., & Higgins, J. P. (2002). How should meta-regression analyses be undertaken and interpreted?. *Statistics in Medicine, 21*(11), 1559-1573. https://doi.org/10.1002/sim.1187

*Thyrian, J. R., Freyer-Adam, J., Hannöver, W., Röske, K., Mentzel, F., Kufeld, C., Bischof, G., Rumpf, H.-J., John, U., & Hapke, U. (2007). Adherence to the principles of Motivational Interviewing, clients' characteristics and behavior outcome in a smoking cessation and relapse prevention trial in women postpartum. *Addictive Behaviors*, *32*(10), 2297–2303. https://doi.org/10.1016/j.addbeh.2007.01.024

*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*(2), 143–157. https://doi.org/10.1023/B:COTR.0000021536.39173.66

*Tschuschke, V., Crameri, A., Koehler, M., Berglar, J., Muth, K., Staczan, P., Von Wyl, A., Schulthess, P., & Koemeda-Lutz, M. (2015). The role of therapists' treatment adherence, professional experience, therapeutic alliance, and clients' severity of psychological problems: Prediction of treatment outcome in eight different psychotherapy approaches. Preliminary results of a naturalistic. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *25*(4), 420–434. https://doi.org/10.1080/10503307.2014.896055

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, *35*(2), 215-247. https://doi.org/10.3102/1076998609346961

Wakefield, S., Kellett, S., Simmonds-Buckley, M., Stockton, D., Bradbury, A., & Delgadillo, J. (2021). Improving Access to Psychological Therapies (IAPT) in the United Kingdom: A systematic review and meta-analysis of 10-years of practice-based evidence. *British Journal of Clinical Psychology*, *60*(1), 1-37. https://doi.org/10.1111/bjc.12259

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, *61*(4), 620–630. https://doi.org/10.1037/0022-006X.61.4.620

*Webb, C. A., Derubeis, R. J., Dimidjian, S., Hollon, S. D., Amsterdam, J. D., & Shelton, R. C. (2012). Predictors of patient cognitive therapy skills and symptom change in two

randomized clinical trials: The role of therapist adherence and the therapeutic alliance. *Journal of Consulting and Clinical Psychology, 80*(3), 373–381. https://doi.org/10.1037/a0027663

Wells, G., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., & Tugwell, P. (2000). The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottowa Hospital Research Institute. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

*Weck, F., Bohn, C., Ginzburg, D. M., & Stangier, U. (2011). Assessment of adherence and competence in cognitive therapy: Comparing session segments with entire sessions. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *21*(6), 658–669. https://doi.org/10.1080/10503307.2011.602751

*Weck, F., Richtberg, S., Jakob, M., Neng, J. M. B., & Hofling, V. (2015). Therapist competence and therapeutic alliance are important in the treatment of health anxiety (hypochondriasis). *Psychiatry Research*, *228*(1), 53–58. https://doi.org/10.1016/j.psychres.2015.03.042

*Weck, F., Rudari, V., Hilling, C., Hautzinger, M., Heidenreich, T., Schermelleh-Engel, K., & Stangier, U. (2013). Relapses in recurrent depression 1 year after maintenance cognitive-behavioral therapy: The role of therapist adherence, competence, and the therapeutic alliance. *Psychiatry Research*, *210*(1), 140–145. https://doi.org/10.1016/j.psychres.2013.05.036

*Westra, H. A., Constantino, M. J., Arkowitz, H., & Dozois, D. J. (2011). Therapist differences in cognitive–behavioral psychotherapy for generalized anxiety disorder: A pilot study. *Psychotherapy*, *48*(3), 283-292. https://doi.org/10.1037/a0022011

*Wurman, T.V. (2019). *Therapist competence in dynamic interpersonal therapy and its association with treatment outcome* [Unpublished doctoral thesis]. University College London.

*Yew, R. Y., Dobson, K. S., Zyphur, M., & Kazantzis, N. (2019). Mediators and moderators of homework-outcome relations in cbt for depression: A study of engagement, therapist skill, and client factors. *Cognitive Therapy and Research, 45,* 209-223. https://doi.org/10.1007/s10608-019-10059-2

Zarafonitis-Müller, S., Kuhr, K., & Bechdolf, A. (2014). The relationship between therapist's competence and adherence to outcome in cognitive-behavioural therapy-results of a metaanalysis. *Fortschritte der Neurologie-psychiatrie*, *82*(9), 502-510. https://doi.org/10.1055/s-0034-1366798

*Zelencich, L. M., Wong, D., Kazantzis, N., McKenzie, D. P., Downing, M., & Ponsford, J. L. (2020). Predictors of Anxiety and Depression Symptom Improvement in CBT Adapted for Traumatic Brain Injury: Pre/Post-Injury and Therapy Process Factors. *Journal of the International Neuropsychological Society*, *26*(1), 97–107. https://doi.org/10.1017/S1355617719000791

Appendix A
Newcastle-Ottawa Quality Assessment Scale (Blank Copy)

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
COHORT STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability

Selection

1) Representativeness of the exposed cohort
   a) truly representative of the average _____ (describe) in the community ✸
   b) somewhat representative of the average _____ in the community ✸
   c) selected group of users eg nurses, volunteers
   d) no description of the derivation of the cohort

2) Selection of the non exposed cohort
   a) drawn from the same community as the exposed cohort ✸
   b) drawn from a different source
   c) no description of the derivation of the non exposed cohort

3) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview ✸
   c) written self report
   d) no description

4) Demonstration that outcome of interest was not present at start of study
   a) yes ✸
   b) no

Comparability

1) Comparability of cohorts on the basis of the design or analysis
   a) study controls for _____ (select the most important factor) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Outcome

1) Assessment of outcome
   a) independent blind assessment ✸
   b) record linkage ✸
   c) self report
   d) no description

2) Was follow-up long enough for outcomes to occur
   a) yes (select an adequate follow up period for outcome of interest) ✸
   b) no

3) Adequacy of follow up of cohorts
   a) complete follow up - all subjects accounted for ✸
   b) subjects lost to follow up unlikely to introduce bias - small number lost - > ____ % (select an adequate %) follow up, or description provided of those lost) ✸
   c) follow up rate < ____% (select an adequate %) and no description of those lost
   d) no statement

Appendix B

Adapted Newcastle-Ottawa Scale and Scoring System

Amended scoring criteria

A modified version of the Newcastle-Ottawa scale (Wells et al., 2000) was used to quality assess the final sample of papers (the 'Cohort Studies' version was used). In the modified version the Comparability portion of the scale and question 2 of the Selection portion of the scale were removed as they pertained to control conditions which were not present in or analysed by any of the studies. The scoring system was also adjusted accordingly (see below).

Item 5: Follow-up long enough

As the extraction hierarchy for effect sizes in the current meta-analysis was end of treatment (or if only follow-up data points were available, closest to end of treatment was given preference), a 'long enough follow-up' was defined as any outcome data point taken at the end of treatment.

Item 6: Adequacy of follow up of cohorts

The adequacy of follow-up item was interpreted as the proportion of the sample from which a competence-outcome effect size was calculated for. As stated above, the 'follow-up' for some studies may have been the end of treatment data point. The following scoring criteria were used for item 6:

a) complete follow up - all subjects with a competency/adherence rating during treatment accounted for (STAR)
b) subjects with a competency/adherence rating lost to follow up unlikely to introduce bias - small number lost (< 50 % follow up), or description provided of those lost) STAR
c) subjects with a competency/adherence rating follow up rate > 50 % and no description of those lost
d) no statement

For all other items in the NOS, the manual was followed.

Scoring Requirements for Different Ratings

Low risk of bias: 3 Stars in Selection section and 2 or 3 in Outcome section
Moderate risk of bias: 2 Stars in Selection section and 2 or 3 in Outcome section
High risk of bias: 0 or 1 in Selection section and 0 or 1 in Outcome section

If a study met the criteria for one rating in one section (i.e. Low risk of bias) but met the criteria for a lower rating in another section (i.e. Moderate risk of bias) they would be awarded the lower of the two ratings.

Note – the original NOS used 'poor', 'fair' and 'good' overall quality ratings. This was changed to 'high', 'moderate' and 'low' risk of bias in the current review, respectively, to reduce confusion with GRADE quality assessments.

Appendix C

Equations Used to Calculate Standard Error

$SE\ of\ \beta y = ln(upperCIy) - ln(lowerCIy)$ / 3.92

SE of B = B / t value

Where:

SE = standard error

B = unstandardised beta coefficient

βy = standardised beta coefficient

ln = natural logarithm

*upper*CIy = upper 95% Confidence Interval

*lower*CIy = lower 95% Confidence Interval

Appendix D

Studies Excluded at Full Text Stage

Stage 1 (Titles/Abstracts)	**Number**
Total papers after duplicates removed	7,098
Stage 1 excluded	6,850
Stage 1 included	248
Stage 2 (Full Texts)	
Full texts screened from database searching	248
Reasons for full-text exclusion	
Measured outcome before cessation of therapy	10
Didn't examine ACI-outcome relationship	102
Insufficient statistical data	12
Child population	9
Examined patient adherence	8
No observer measure of therapist ACI	13
Case study	1
Not an experimental study	8
Non-English	8
Duplicate	12
Not face-to-face psychotherapy	3
Physical health focused	1
ACI measure not of therapy session	3
Could not obtain full text	5
Total Excluded at Stage 2 (titles and exclusion reasons below)	195
Total included	53
Total screened	**248**
Hand searched/screened	31
Hand searched included	9
Grand total papers included in review/meta-analysis	**62**

Exclusion Reasons with Excluded Study Titles (Bibliographic Databases)

Measured outcome before cessation of therapy (*k*=10)
Development of a therapist adherence/competence rating scale for supportive-expressive dynamic psychotherapy: A preliminary report.
Factor structure of therapist fidelity to individual resiliency training in the Recovery After an Initial Schizophrenia Episode Early Treatment Program
A re-examination of process-outcome relations in cognitive therapy for depression: Disaggregating within-patient and between-patient effects.
Identifying moderators of the adherence-outcome relation in cognitive therapy for depression.
The process of change in cognitive therapy for depression: predictors of early inter-session symptom gains.
Therapist use of specific and nonspecific strategies across two affect-focused psychotherapies for depression: Role of adherence monitoring
Patient characteristics and variability in adherence and competence in cognitive-behavioral therapy for panic disorder.
How Do Supportive Techniques Bring About Therapeutic Change: The Role of Therapeutic Alliance as a Potential Mediator
Interpersonal Factors Are Associated with Lower Therapist Adherence in Cognitive–Behavioural Therapy for Panic Disorder
What process works for whom: Individual differences and the impact of therapy techniques and treatment mechanisms.

Didn't examine ACI-outcome relationship (*k*=102)
Clinical supervision in cognitive behavior therapy improves therapists' competence: a single-case experimental pilot study
Promoting Counseling Competence using Silf-Reflection
Developing a measure of fidelity for an ecological approach to family therapy.
Enhancing treatment fidelity in psychotherapy research: novel approach to measure the components of cognitive behavioural therapy for relapse prevention in first-episode psychosis.
The influence of supervision on manual adherence and therapeutic processes.
Therapist adherence in the strong without anorexia nervosa (SWAN) study: A randomized controlled trial of three treatments for adults with anorexia nervosa.
The treatment of postnatal depression by health visitors: impact of brief training on skills and clinical practice.
An evaluation of the effect of an educational intervention for Australian social workers on competence in delivering brief cognitive behavioural strategies: a randomised controlled trial.
Scaling up the evaluation of psychotherapy: evaluating motivational interviewing fidelity via statistical text classification.

Test Interpretation Competence: A Comparison of Microskills and Mental Practice Training
Effects of Separate and Combined Overt and Covert Practice Modes on Counseling Trainee Competence and Motivation
Microskills practice versus mental practice training for competence in decision-making counseling
Implementation of a fidelity monitoring process to assess delivery of an evidence-based adherence counseling intervention in a multi-site biomedical HIV prevention study
Development of an adherence/competence rating scale for individual drug counseling.
Therapists' adherence and competence and treatment discrimination in the NIDA Collaborative Cocaine Treatment Study.
Development of the Cognitive Therapy Adherence and Competence scale.
Assessing the development of competence during postgraduate cognitive-behavioral therapy training.
Manual-based cognitive behaviour therapy for chronic fatigue syndrome: therapists' adherence and perceptions.
Trainee self-assessment of cognitive behaviour therapy competence during and after training
Building an Evidence Base for Effective Supervision Practices: An Analogue Experiment of Supervision to Increase EBT Fidelity
Assessing Adherence, Competence and Differentiation in a Stepped-Wedge Randomised Clinical Trial of a Complex Behaviour Change Intervention
Fidelity considerations in translational research: Eating As Treatment - a stepped wedge, randomised controlled trial of a dietitian delivered behaviour change counselling intervention for head and neck cancer patients undergoing radiotherapy.
Developing a measure of provider adherence to improve the implementation of behavioral health services in primary care: a Delphi study.
Therapy contamination as a measure of therapist treatment adherence in a trial of cognitive behaviour therapy versus befriending for psychosis.
Strengthening motivational interviewing skills following initial training: a randomised trial of workplace-based reflective practice.
Implementing an evidence-based psychological intervention for suicidal thoughts and behaviors on an inpatient unit: Process, challenges, and initial findings
Process evaluation of Internet-based cognitive behavioural therapy for adults with tinnitus in the context of a randomised control trial
Assessing the strength and integrity of an intervention.
Is the process of delivery of an individually tailored lifestyle intervention associated with improvements in LDL cholesterol and multiple lifestyle behaviours in people with familial hypercholesterolemia?.
The RoadMAP relapse prevention group counseling toolkitTM: Counselor adherence and competence outcomes.
Factors associated with competence in cognitive therapists.
Can physical therapists deliver a pain coping skills program? An examination of training processes and outcomes.
Therapist skill and patient variables in homework compliance: Controlling an uncontrolled variable in cognitive therapy outcome research.

Measuring competence in systemic practice: development of the ‘Systemic Family Practice – Systemic Competency Scale’ (SPS)
Measuring adherence in time-limited dynamic psychotherapy.
Assessing fidelity of treatment delivery in group and individual 12-step facilitation.
Etherapy: A training program for development of clinical skills in distance psychotherapy.
Internal validity of Project MATCH treatments: discriminability and integrity.
A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions.
Adherence to principles of motivational interviewing and client within-session behavior.
The Mindfulness-Based Relapse Prevention Adherence and Competence Scale: Development, interrater reliability, and validity.
Implementation of a smoking cessation treatment integrity protocol: treatment discriminability, potency and manual adherence.
The therapeutic alliance and therapist adherence as predictors of dropout from cognitive therapy for depression when combined with antidepressant medication.
Characterizing the integration of CBT and psychodynamic techniques in interpersonal reconstructive therapy for patients with severe and comorbid personality pathology.
Do supervisors and independent judges agree on evaluations of therapist adherence and competence in the treatment of cocaine dependence?
Disorder- and Treatment-Specific Therapeutic Competence Scales for Posttraumatic Stress Disorder Intervention: Development and Psychometric Properties
Treatment integrity and members' change in group counseling: A pilot study on counselor's mentalizing interventions.
The design of the MBT-G adherence and quality scale
Person-Centred and Experiential Psychotherapy Scale: Development and reliability of an adherence/competence measure for person-centred and experiential psychotherapies.
Therapist adherence and competence with manualized cognitive-behavioral therapy for PTSD delivered via videoconferencing technology.
A nine session manual of motivational enhancement therapy for methamphetamine dependence: adherence and efficacy.
Assessing treatment integrity in alcohol behavioral couple therapy.
Applying the Collaborative Study Psychotherapy Rating Scale to rate therapist adherence in cognitive-behavior therapy, interpersonal therapy, and clinical management.
The Relationship Between the Level of Program Integrity and Pre- and Post-Test Changes of Responsive-Aggression Regulation Therapy (Re-ART) Outpatient: A Pilot Study.
Mutual influence in therapist competence and adherence to motivational enhancement therapy.
Evaluating therapist adherence in motivational interviewing by comparing performance with standardized and real patients.
Moderators of trainee therapists' competence in cognitive therapy.

Multifamily Group Psychoeducation in New York State: Implementation and Fidelity Outcomes.
Assessing competence in cognitive-behavioural therapy.
Cognitive analytic therapy for borderline personality disorder: therapist competence and therapeutic effectiveness in routine practice.
An assessment of physiotherapist's delivery of behaviour change techniques within the SOLAS feasibility trial
What can we expect from trainee therapists? A study of acquisition of competence in dynamic psychotherapy.
Counselor trainee achievement goal orientation and the acquisition of time-limited dynamic psychotherapy skills.
Demonstrating functional analytic psychotherapy as an independent variable in efficacy research: A new measure of treatment fidelity.
A sequential analysis of motivational interviewing technical skills and client responses
Evaluating CBT clinical competence with standardised role plays and patient therapy sessions.
Therapist fidelity with an exposure-based treatment of PTSD in adults with schizophrenia or schizoaffective disorder.
A method for assessing fidelity of delivery of telephone behavioral support for smoking cessation.
Assessing fidelity of delivery of smoking cessation behavioural support in practice.
Adherence has been defined as 'the extent to which a therapist used interventions and approaches prescribed by the treatment manual and avoided the use of interventions proscribed by the manual' (Waltz et al., 1993)
Use of fidelity assessments to train clinicians in the CBT for PTSD program for clients with serious mental illness.
Organizational readiness for change in community-based addiction treatment programs and adherence in implementing evidence-based practices: A national study
Assessing treatment fidelity and contamination in a cluster randomised controlled trial of motivational interviewing and cognitive behavioural therapy skills in type 2 diabetes
Nurse-led psychological interventions to improve diabetes control: assessing competencies.
Are all cognitive therapies alike? A comparison of cognitive and noncognitive therapy process and implications for the application of empirically supported treatments.
Psychotherapy adherence of therapists treating HIV-positive patients with depressive symptoms.
Measuring therapist adherence in psychotherapy for anorexia nervosa: Scale adaptation, psychometric properties, and distinguishing psychotherapies.
"Mentalization-based therapy adherence and competence stimulates insession mentalization in psychotherapy for borderline personality disorder with co-morbid substance dependence": Corrigendum.
Therapist adherence to manualized cognitive-behavioral therapy for anger management delivered to veterans with PTSD via videoconferencing.
Effects of motivational interviewing fidelity on substance use treatment engagement in primary care

Treatment adherence and competency ratings among therapists, supervisors, study-related raters and external raters in a clinical trial of a 12-step facilitation for stimulant users
A prenatal coparenting intervention with unmarried father-mother dyads: Fidelity of intervention delivery by male-female community mentor teams
Characteristics of trainees' early sessions: A naturalistic process-outcome study tribute to Jeremy D. Safran.
Training and dissemination of cognitive behavior therapy for depression in adults: a preliminary examination of therapist competence and client outcomes.
Therapist treatment fidelity in prescriptive vs. exploratory psychotherapy.
Therapist competence: its temporal course, temporal stability, and determinants in short-term anxiety-provoking psychotherapy.
An evaluation of treatment integrity in a randomized trial of behavioural therapy for low mood in stroke patients with aphasia.
The relationship between therapist competency in cognitive therapy and general therapy skill.
The relationship between therapist competence and homework compliance in maintenance cognitive therapy for recurrent depression: secondary analysis of a randomized trial.
Evaluating therapist competency and adherence to behavioral family management with bipolar patients.
Associations between therapy skills and patient experiences of change processes in cognitive behavioral therapy for psychosis.
Clinical supervision in cognitive behavior therapy improves therapists' competence: a single-case experimental pilot study
Measuring adherence and competence of dynamic therapists in the treatment of cocaine dependence.
Improvement in therapist skills over sessions in brief motivational interventions predicts client language and alcohol use outcomes
Therapist adherence in individual cognitive-behavioral therapy for binge-eating disorder: assessment, course, and predictors.
Comparing the Treatment Process in Successful and Unsuccessful Cases in Two Forms of Psychotherapy for Cluster C Personality Disorders
A pilot evaluation of a brief CBT training course: Impact on trainees' satisfaction, clinical skills and patient outcomes
The relation between specific and general dimensions of the psychotherapy process in interpersonal psychotherapy of depression.
Differential effects of alliance and techniques on Panic-Specific Reflective Function and misinterpretation of bodily sensations in two treatments for panic
Focal adherence in brief dynamic psychotherapy: a comparison from two independent studies
Psychotherapy process in the National Institute of Mental Health Treatment of Depression Collaborative Research Program.
Intervention and impact: An examination of treatment adherence, therapeutic alliance, and outcome in cognitive therapy.

Insufficient statistical data (*k*=12)

Community Program Therapist Adherence and Competence in a Motivational Interviewing Assessment Intake Session (adherence and competence)
Therapeutic Alliance and Treatment Adherence in Two Interventions for Bulimia Nervosa: A Study of Process and Outcome
The role of therapist adherence, therapist competence, and alliance in predicting outcome of individual drug counseling: Results from the National Institute Drug Abuse Collaborative Cocaine Treatment Study (adherence and competence)
CBT competence in novice therapists improves anxiety outcomes (adherence and competence)
Adherence and competence in two manual-guided therapies for co-occurring substance use and posttraumatic stress disorders: clinician factors and patient outcomes (ICBT and IAC groups - adherence and competence also)
Therapist competence and clinical outcome in the Prevention of Parasuicide by Manual Assisted Cognitive Behaviour Therapy Trial: the POPMACT study
Treatment specific competence predicts outcome in cognitive therapy for social anxiety disorder
Teacher Competence in Mindfulness-Based Cognitive Therapy for Depression and Its Relation to Treatment Outcome
Therapist Competenceand PatientOutcomein Interpersonal Psychotherapyof Depression
THE EFFECT OF COGNITIVE-BEHAVIORAL THERAPY, THERAPIST COMPETENCY, AND GROUP PROCESS ON DEPRESSION AMONG THE ELDERLY
Clinician Integrity in Multiple Family Groups: Psychometric Properties and Relationship with Schizophrenia Client and Caregiver Outcomes
Clinicians' Fidelity to a Manual-Based Family Treatment as a Predictor of the One-Year Course of Bipolar Disorder

Child population (*k*=9)
A descriptive evaluation of long-term treatment integrity.
Adherence to and competence in cognitive behavioral therapy for youth anxiety: Psychometric evaluation.
From counselor skill to decreased marijuana use: Does change talk matter?
Exploring the link among behavior intervention plans, treatment integrity, and student outcomes under natural educational conditions.
The influence of parental factors on therapist adherence in multi-systemic therapy.
Transporting efficacious treatments to field settings: the link between supervisory practices and therapist fidelity in MST programs.
Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning
Investigating the 'active ingredients' of cognitive behaviour therapy and counselling for patients with chronic fatigue in primary care: Developing a new process measure to assess treatment fidelity and predict outcome

The Relationship Between the Level of Program Integrity and Pre- and Post-Test Changes of Responsive-Aggression Regulation Therapy (Re-ART) Outpatient: A Pilot Study.

Examined patient adherence (*k*=8)
Predictors of symptomatic change and adherence in Internet-based cognitive behaviour therapy for social anxiety disorder in routine psychiatric care.
Outpatient and self-referred participants: Adherence to treatment components and outcome in an internet intervention targeting anxiety disorders
Patient recall of specific cognitive therapy contents predicts adherence and outcome in adults with major depressive disorder
Weekly brief phone support in self-help cognitive behavioral therapy for insomnia disorder: Relevance to adherence and efficacy.
Program Evaluation of Group-based Cognitive Behavioral Therapy for Insomnia: a Focus on Treatment Adherence and Outcomes in Older Adults with Co-morbidities
Treatment modality preferences and adherence to group treatment for panic disorder with agoraphobia.
Process and treatment adherence factors in group cognitive-behavioral therapy for partner violent men.
Adherence to Internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis.

No observer measure of therapist ACI (*k*=13)
A Prospective Study of Therapist Facilitative Interpersonal Skills as a Predictor of Treatment Outcome
Therapist facilitative interpersonal skills and training status: A randomized clinical trial on alliance and outcome
Psychometric assessment of the Primary Care Behavioral Health Provider Adherence Questionnaire (PPAQ)
Using the Primary Care Behavioral Health Provider Adherence Questionnaire (PPAQ) to identify practice patterns.
Four process studies in the behavioral treatment of chronic headache.
Impact of an educational intervention on general practitioners' skills in cognitive behavioural strategies: a randomised controlled trial.
Counseling and Psychotherapy Skills Training for Family Physicians
Implementation of evidence-based rehabilitation for non-specific back pain and common mental health problems: a process evaluation of a nationwide initiative.
Drug treatment process indicators for probationers and prediction of recidivism.
An Investigation of the Relationship Between the Alliance Negotiation Scale and Psychotherapy Process and Outcome

Therapist Adherence to Good Psychiatric Practice in a Short-Term Treatment for Borderline Personality Disorder.
Therapist competence, comorbidity and cognitive-behavioral therapy for depression.
Traveling the road to healing: An examination of the relationship between psychotherapeutic technique and cognitive, emotional and physiological outcomes in women exposed to violence.

Case study (*k*=1)
The process of change in brief psychotherapy: effects of psychodynamic and cognitive-behavioral prototypes.

Not an experimental study (*k*=8)
Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium.
An interpersonal perspective on therapy alliances and techniques.
Treatment integrity of studies that compare short-term psychodynamic psychotherapy with cognitive-behavior therapy.
Does therapists' competence matter in delivering psychological therapy?
Experiential therapy in practice: The process-experiential approach.
Humanistic-existential psychotherapy competencies and the supervisory process.
The adherence/resource priming paradigm - a randomised clinical trial conducting a bonafide psychotherapy protocol for generalised anxiety disorder
The role of adherence in mediating the relationship between depression and health outcomes

Non-English (*k*=8)
Development of interpersonal problem solving competences during psychotherapy training in psychodrama.
Therapist adherence in cognitive-behavioral therapy for binge-eating disorder.
Process analysis of behavioral family management with schizophrenic patients

Effects of skill level on individual behavioral counseling for smoking cessation
Behavior Therapy Competency Checklist. Competency evaluation, quality assurance and supervision.
Definition, operationalization and quality assurance of psychotherapy. An investigation with the behavior therapy-competence-checklist (btcc).
Validation of the Therapist Competence Scale with clients with generalized anxiety disorder.
Treatment technique in guided imagery psychotherapy, hypnopsychotherapy, and autogenic psychotherapy: The therapist version of the "Process Scale for Imagery- and Trance-Based Psychodynamic Therapies" (PIT).

Duplicate (sample or paper) (*k*=12)
Empirical validation of treatment models: An examination of psychotherapy process and outcome.
Strengthening competence of therapists-in-training in the treatment of health anxiety (hypochondriasis): Validation of the Assessment of Core CBT Skills (ACCS)
"Mentalization-based therapy adherence and competence stimulates insession mentalization in psychotherapy for borderline personality disorder with co-morbid substance dependence": Corrigendum.
Fidelity to the cognitive processing therapy protocol: Further evaluation of critical elements.
Therapist adherence to treatment manuals and its relation to the therapeutic alliance and therapy outcome: Scale development and validation.
A study in reliability and validation: What is the nature of the unique and joint contributions of therapist adherence and therapeutic alliance to treatment outcome?
Therapists' adherence to manualized treatments in the context of ruptures.
Psychotherapy technique related to changes in anxiety symptoms with a transdiagnostic sample.
Weck et al (2014) - Assessing Treatment Integrity in Cognitive-Behavioral Therapy: Comparing Session Segments With Entire Sessions
Weck et al (2015a) - Treatment failure in cognitive-behavioural therapy: Therapeutic alliance as a precondition for an adherent and competent implementation of techniques
Svartberg et al (1994) - Therapeutic Alliance, Therapist Competence, and Client Change in Short-Term Anxiety-Provoking Psychotherapy
Thyrian et al (2010) - Population-based smoking cessation in women post partum: adherence to motivational interviewing in relation to client characteristics and behavioural outcomes

Not face-to-face psychotherapy (*k*=3)
Do adherence variables predict outcome in an online program for the prevention of eating disorders?.
The effects of treatment adherence and treatment-specific therapeutic competencies on outcome and goal attainment in telephone-based therapy with caregivers of people with dementia.
Development and evaluation of a scale assessing therapist fidelity to guidelines for delivering therapist-assisted Internet-delivered cognitive behaviour therapy
Physical health focus (*k*=1)
Is a motivational interviewing based lifestyle intervention for obese pregnant women across Europe implemented as planned? Process evaluation of the DALI study

ACI measure not of therapy session (*k*=3)
The Relationship between Competence and Patient Outcome with Low-Intensity Cognitive Behavioural Interventions
Effects of trained health professionals' behavioral counseling skills on smoking cessation outcomes
Influence of supervision, therapist's competence, and patient's ego level on the effects of time-limited psychotherapy.

Could not obtain full text (*k*=5)
Psychoanalytic psychotherapy outcome: Relation to technical competence, training and clinical experience
The impact of therapeutic alliance and other therapy process variables as predictors of outcome in cognitive therapy with older adults.
Therapeutic alliance and adherence in cognitive therapy for depression.
The role of therapist actions in process-experiential therapy. Davis
Processes of symptom change in psychotherapy: Investigating the role of therapist adherence, competence and the therapeutic alliance.

Hand-Searched Studies

The below table is a summary of the hand-searched papers screened by full text (with exclusion reasons). This is a separate list because the tables above only include the texts that were derived from bibliographic database searching. The PRISMA diagram in the main text includes both bibliographic and hand-searched papers.

Hand-searched papers screened (full texts) - e.g. grey literature, other reviews (*k*=31 screened, *k*=9 included)	**Included/ Excluded**	**Reason for exclusion**
Carroll et al (1997) - Contribution of the Therapeutic Alliance to Outcome in Active Versus Control Psychotherapies	Excluded	Insufficient statistical data
Crits-Christoph et al (1988) - The Accuracy of Therapists' Interpretations and the Outcome of Dynamic Psychotherapy	Excluded	Child population (age range started at 15)
DeRubeis (1990) - Determinants of Change in Cognitive Therapy for Depression	Excluded	Measured outcome before cessation of therapy
Farmer et al (2016) - Fidelity to the Cognitive Processing Therapy Protocol: Evaluation of Critical Elements	Included	
Feeley et al (1999) - The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression	Excluded	Duplicated (already included on bibliographic list)
Fischer et al (2015) - Motivational Interviewing Treatment Integrity and Client Change: Using ROC Analysis to Explore the Relationship Between MI Fidelity Level and Drinking Outcome.	Excluded	Insufficient statistical data
Gaston et al (1992) - preliminary results on the inventory of therapeutic strategies	Excluded	Insufficient statistical data
Gaston et al (1994) - Alliance and Technique for Predicting Outcome in Short-and Long-Term Analytic Psychotherapy	Excluded	Didn't examine adherece/competence-outcome relationship
Gaston et al (1998) - Alliance, Technique, and their Interactions in Predicting Outcome of Behavioral, Cognitive, and Brief Dynamic Therapy	Excluded	Didn't examine adherece/competence-outcome relationship
Gower (2011) - Therapist competence, case conceptualisation and therapy outcome in cognitive behavioural therapy	Included	
Hall (2007) - Therapist relationship and technical skills in two versions of emotion focused trauma therapy	Included	
Horowitz et al (1984) - Brief psychotherapy of bereavement reactions	Excluded	Didn't examine adherece/competence-outcome relationship
Jasin (1986) - THE EFFECT OF COGNITIVE-BEHAVIORAL THERAPY, THERAPIST COMPETENCY, AND GROUP PROCESS ON DEPRESSION AMONG THE ELDERLY	Excluded	Insufficient statistical data
Katz et al (2019) - Adherence, Flexibility, and Outcome in Psychodynamic Treatment of Depression	Excluded	Duplicated (already included on bibliographic list)
Katz et al (2018) - Psychodynamic technique early in treatment related to outcome for depressed patients	Excluded	Duplicated sample

Markowitz et al (2000) - Psychotherapy Adherence of Therapists Treating HIV-Positive Patients With Depressive Symptoms	Excluded	Insufficient statistical data
Marziali (1984) - prediction of Outcome of Brief Psychotherapy From Therapist Interpretive Interventions	Excluded	Didn't examine adherece/competence-outcome relationship
McDonell (2004) - TREATMENT INTEGRITY AND CLIENT OUTCOMES IN MULTIPLE FAMILY GROUP TREATMENT	Excluded	Duplicated sample
Minonne (2008) - Therapist Adherence, Patient Alliance, and Depression Change in the NIMH Treatment for Depression Collaborative Research Program	Excluded	Duplicated sample
Ogrodniczuk (1997) - therapist adherence to treatment manuals and its relation to the therapeutic alliance and therapy outcome: scale development and validation	Included	
Ogrodniczuk et al (1999) - Measuring Therapist Technique in Psychodynamic Psychotherapies: Development and Use of a New Scale	Excluded	Duplicated sample
Patton (1997) - STUDY IN RELIABILITY AND VALIDATION: WHAT IS THE NATURE THE UNIQUE AND JOINT CONTRIBUTIONS OF THERAPIST ADHERENCE AND THERAPEUTIC ALLIANCE TO TREATMENT OUTCOME?	Included	
Piper et al (1986) - Relationships between the object focus of therapist interpretations and outcome in short-term individual psychotherapy	Excluded	Didn't examine adherece/competence-outcome relationship
Piper et al (1991) - Transference Interpretations, Therapeutic Alliance, and Outcome in Short-term Individual Psychotherapy	Excluded	Didn't examine adherece/competence-outcome relationship
Pitman et al (2017) - Psychotherapy Technique Related to Changes in Anxiety Symptoms With a Transdiagnostic Sample	Excluded	Duplicated sample
Sachs (1983) - Negative factors in brief psychotherapy: an empirical assessment	Included	
Slavin-Mulford et al (2011) - Therapeutic Interventions Related to Outcome in Psychodynamic Psychotherapy for Anxiety Disorder Patients	Excluded	Duplicated sample - Same sample as Pitman et al (2014)
Spektor et al (2008) - therapists' Adherence To Manualized Treatments In The Context of Ruptures	Included	
Taylor (2010) - Traveling the Road to Healing: An Examination of the Relationship between Psychotherapeutic Technique and Cognitive, Emotional and Physiological Outcomes in Women Exposed to Violence	Excluded	Adherence/competence was self-rated by therapist
Wurman (2019) - THERAPIST COMPETENCE IN DYNAMIC INTERPERSONAL THERAPY AND ITS ASSOCIATION WITH TREATMENT OUTCOME	Included	

Westra et ak (2011) - Therapist Differences in Cognitive–Behavioral Psychotherapy for Generalized Anxiety Disorder: A Pilot Study	Included	

Appendix E

Individual Study Characteristics

Table 1a

Non-Hierarchical Adherence-Outcome Study Characteristics

Study	Patient N	Therapist N	Treatment Group	Problem Targeted	Patient Demographics	Therapist Demographics	Outcome Measure	% coverage (per patient)	Aspect of measure used	Type of session data rated	Alliance controlled	Country	RoB assessment
Ablon et al. (2006)	17	7	Psychodynamic	Anxiety Disorders	88.2% female, mean age 35	28.57% female	Panic Disorder Severity Scale (PDSS)	4.17	Sub-scale	Transcripts	No	US	6/6 (low)
Barber et al (2008) - adherence	108	13	Supportive Expressive Dynamic Psychotherapy	Substance Use/Addiction	19% female, mean age 33 (6.1)	69.23% female	Addiction Severity Index (ASI) the higher score, higher addiction severity	2.78	Sub-scale	Audio	No	US	5/6 (low)
Barber et al. (1996) - adherence	29	4	Supportive Expressive Dynamic Psychotherapy	Depression	72% female, mean age 38.4 (10.36)	NR	BDI	6.25	Sub-scale	Audio	No	US	5/6 (low)
Boyle et al. (2020) - adherence	70	30	CBT	Mixed	64.3% female, mean age 36.8 (12.4)	86.7% female, 31.2 mean age	Bern post-session report - patient (BPSR-P)	10	Whole	Video	No	Germany	5/6 (low)
Feeley et al (1999)	25	4	CT	Depression	88% female, mean age 32.9 (11.2)	25% female	BDI	20.55	Sub-scale	Audio	No	US	4/6 (moderate)
Gibbons et al (2010b) - 2 session sample - adherence	77	11	MI/MET	Substance Use/Addiction	NR	55% female, mean age 44	Percentage of days of marijuana use	100	Sub-scale	Video	No	US	5/6 (low)
Gibbons et al (2010b) - 9 session sample - adherence	86	11	Mixed	Substance Use/Addiction	NR	55% female, mean age 45	Percentage of days of marijuana use	100	Sub-scale	Video	No	US	
Goldman et al (2009)	10	5	Psychodynamic	BPD	90% female, mean age 27.40 (6.85)	NR	Borderline Evaluation of Severity Over Time (BEST)	9.62	Whole	Video	No	US	4/6 (moderate)
Hall (2007) - EE adherence	26	11	EFT	Trauma/PTSD	57.69% female, mean age 45.62 (11.57)	63.64% female, age range 25-57	Resolution Scale (RS)	15	Whole	Video	No	Canada	4/6 (moderate)
Hall (2007) - IC adherence	20	11	EFT	Trauma/PTSD	50% female, mean age 45.7 (14.96)	63.64% female, age range 25-55	Resolution Scale (RS)	15	Whole	Video	No	Canada	

Hilsenroth et al (2003)	21	10	Psychodynamic	Depression	47.62% female, mean age 34.43 (12.7)	50% female, age not reported	Clinician-rated DSM-IV Rating scale of major depressive episode symptoms	23.33	Sub-scale	Video	No	US	6/6 (low)
Huppert et al (2006)	56	13	CBT	Anxiety Disorders	62% female, mean age 36	NR	Panic Disorder Severity Scale (PDSS)	52.73	Whole	Audio	No	US	5/6 (low)
Luborsky et al (1985)	41	9	Mixed	Substance Use/Addiction	100% male, no mean age reported (18-55)	NR	Addiction Severity Index: Drug use subscale		Whole	Audio	No	US	5/6 (low)
Martino et al (2008) - adherence	461	35	MI/MET	Substance Use/Addiction	29% female, mean age 35.3 (9.7)	60% female, mean age 38.9 (11.8)	Timeline follow-back method (days primary percent drug abstinence)	100	Sub-scale	Audio	No	US	5/6 (low)
McCarthy et al (2016)	33		Psychodynamic	Depression	61% female, mean age 35.5	NR	HRSD	5.88%	sub-scale	audio	no	US	6/6 (low)
Minonne (2008) - CBT sample	59	8	CBT	Depression	NR	NR	BDI	24.69	Sub-scale	Video	No	US	4/6 (moderate)
Minonne (2008) - IPT sample	61	10	IPT	Depression	NR	NR	BDI	24.69	Sub-scale	Video	No	US	
Ogrodniczuk (1997) - STI	72	18	Short-term interpretative psychotherapy	Mixed	full sample (across two treatment conditions) = 63% female, mean age 34.3 (9.6)	55.56% female, mean age 42.6 (7.7)	Various outcomes collected from patient, therapist and external assessor - pre/post scores used to calculate residual gain scores	45	Whole	Audio	No	Canada	5/6 (low)
Ogrodniczuk (1997) - SUP	72	18	Short-term supportive individual psychotherapy	Mixed	full sample (across two treatment conditions) = 63% female, mean age 34.3 (9.6)	55.56% female, mean age 42.6 (7.7)	Various outcomes collected from patient, therapist and external assessor - pre/post scores used to calculate residual gain scores	45	Whole	Audio	No	Canada	
Owen et al (2014)	70	28	Psychodynamic	Mixed	75% female, mean age 29.8 (11)	53.57% female, age not reported	Resolution Scale (RS)	14.29	Sub-scale	Video	No	US	5/6 (low)
Patton (1997) - BAP sample	8	4	Brief-adaptive psychotherapy (BAP)	Mixed	NR	NR	Symptom Checklist-90-R: global severity index; inventory of interpersonal		Sub-scale	Audio	No	US	4/6 (moderate)

							problems - 64 combined						
Patton (1997) - CBT sample	8	4	CBT	Mixed	NR	NR	Symptom Checklist-90-R: global severity index; inventory of interpersonal problems - 64 combined		Sub-scale	Audio	No	US	
Patton (1997) - IET sample	8	4	Interpersonal-experiential therapy (IET)	Mixed	NR	NR	Symptom Checklist-90-R: global severity index; inventory of interpersonal problems - 64 combined		Sub-scale	Audio	No	US	
Pavio et al (2004) - adherence	37	10	EFT	Trauma/PTSD	78% female, mean age 38 (11.32)	70% female, mean age 34.2 (7.41)	Resolution Scale (RS)	15.79	Whole	Audio	No	Canada	4/6 (moderate)
Pitman et al (2014)	20	14	Psychodynamic	Anxiety Disorders	75% female, mean age 27.5 (10.25)	57.14% female	The Brief Symptom Inventory (BSI): anxiety subscale (ANX)	7.14	Sub-scale	Video	No	US	5/6 (low)
Shaw et al (1999) - adherence	36	8	CBT	Depression	not reported	75% female	Hamilton Rating Scale for Depression (HRSD)	42.86	Whole	Video	No	US	6/6 (low)
Sinari et al (2012)	53	3	IPT	Anxiety Disorders	patients mean age 35 (9.5) sex not reported	NR	Liebowitz Social Anxiety Scale (LSAS)	21.43	Sub-scale	Video	No	US	5/6 (low)
Spektor (2008) - BRT sample	14	13	Brief Relational Therapy	Mixed	78.57% female, mean age 40.93 (14.74)	69.23% female, mean age 34.5 (5.5)	Symptom Checklist-90-Revised (SCL-90-R)	6.67	Sub-scale	Audio	No	US	3/6 (high)
Spektor (2008) - CBT sample	14	10	CBT	Mixed	28.57% female, mean age 41.14 (10.53)	60% female, mean age 37.54 (5.5)	Symptom Checklist-90-Revised (SCL-90-R)	6.67	Sub-scale	Audio	No	US	
Thyrian et al (2007)	161	NR	MI/MET	Substance Use/Addiction	100% female, mean age 26.28 (5.73)	NR	Smoking status (smoking/not smoking)	33.33	Sub-scale	Audio	No	Germany	5/6 (low)
Webb et al (2012) - UW sample	45	3	CT	Depression	NR	NR	BDI-II	12.5	Sub-scale	Video	Yes	US	4/6 (moderate)
Weck et al (2011) - adherence	34	10	CT	Anxiety Disorders	41.2% female, mean age 34.8 (9.32)	80% female, mean age 33.7 (7)	Clinical Global Impression Scale-Improvement (CGI-I)	6.25	Whole	Video	No	Germany	6/6 (low)
Weck et al (2013) - adherence	80	26	CBT	Depression	68.6% female, mean age 48.3 (11.6)	80.8% female, mean age 33.8 (8.7)	Time to the first relapse	6.25	Whole	Video	Yes	Germany	5/6 (low)

Weck et al (2015) - cognitive therapy adherence	35	17	CT	Anxiety Disorders	57.14% female, mean age 38.14 (10.54)	76.47% female, mean age 30.94 (4.01)	Yale–Brown Obsessive Compulsive Scale for Hypochondriasis (H-YBOCS)	16.67	Whole	Video	No	Germany	4/6 (moderate)
Weck et al (2015) - exposure therapy adherence	33	17	Exposure Therapy	Anxiety Disorders	54.55% female, mean age 41.67 (12.52)	82.35% female, mean age 30.35 (3.87)	Yale–Brown Obsessive Compulsive Scale for Hypochondriasis (H-YBOCS)	16.67	Whole	Video	No	Germany	

Notes. NR = Not Reported, CT = Cognitive Therapy, CBT = Cognitive Behaviour Therapy, MBCT = Mindfulness Based Cognitive Therapy, IPT = Interpersonal Psychotherapy, MI/MET = Motivational Interviewing/Motivational Enhancement Therapy, BDI = Beck Depression Inventory, HRSD = Hamilton Rating Scale for Depression.

Table 1b

Non-Hierarchical Competence-Outcome Study Characteristics

Study name	**Patient N**	**Therapist N**	**Treatment Group**	**Problem Targeted**	**Patients Demographics**	**Therapists Demographics**	**Outcome Measure**	**% coverage** (per patient)	**Aspect of measure used**	**Type of session data rated**	**Alliance controlled**	**Country of study**	**RoB assessment**
Abel et al. (2016)	50	11	CBT	Depression	73.1% female, 49.6 (11.6)	90.91% female	BDI-II	NR	Whole	Audio	No	UK	5/6 (low)
Barber et al (2008) - competence	108	13	Psychodynamic	Substance Use/Addiction	19% female, M age 33 (6.1)	69.23% female	Addiction Severity Index (ASI) the higher score, higher addiction severity	2.78	Sub-scale	Audio	Yes	US	5/6 (low)
Barber et al. (1996) - competence	29	4	Psychodynamic	Depression	72% female, age 38.4 (10.36)	NR	BDI	6.25	Sub-scale	Audio	No	US	5/6 (low)
Boyle et al. (2020) - competence	70	30	CBT	Mixed	64.3% female, M age 36.8 (12.4)	86.7% female, 31.2 mean age	Bern post-session report - patient (BPSR-P)	10	Whole	Video	No	Germany	5/6 (low)
Branson et al. (2015)	1247	43	CBT	Mixed	Not reported (only reported for therapists)	67.44% female, mean age 38.6 (8.29)	PHQ-9	NR	Whole	Audio	No	UK	5/6 (low)
Chevron et al (1983)	13	9	IPT	Depression	NR	NR	Hamilton Depression Scale	18.75	Whole	Video	No	US	4/6 (moderate)
Farmer et al (2016)	45	8	CPT	Trauma/PTSD	sex not reported, mean age 37.75 (11.59)	NR	Posttraumatic Stress Diagnostic Scale (PDS)	100	Whole	Video	No	US	4/6 (moderate)
Gibbons et al (2010b) - 2 session sample - competence	77	11	MI/MET	Substance Use/Addiction	NR	55% female, mean age 44	percentage of days of marijuana use	100	Sub-scale	Video	No	US	5/6 (low)
Gibbons et al (2010b) - 9 session sample - competence	86	11	Mixed	Substance Use/Addiction	NR	55% female, mean age 45	percentage of days of marijuana use	100	Sub-scale	Video	No	US	

Gower (2011)	20	9	CBT	Depression	60% female, mean age 49.9 (10.5)	88.89% female, mean age 37 (5.6)	BDI-II	11.11	Whole	Audio	No	UK	5/6 (low)
Hall (2007) - EE competence	26	11	EFT	Trauma/PTSD	57.69% female, mean age 45.62 (11.57)	63.64% female, age range 25-58	Resolution Scale (RS)	15	Whole	Video	No	Canada	4/6 (moderate)
Hall (2007) - IC competence	20	11	EFT	Trauma/PTSD	50% female, mean age 45.7 (14.96)	63.64% female, age range 25-56	Resolution Scale (RS)	15	Whole	Video	No	Canada	
Hoffart et al (2005)	35	2	Schema-focused therapy	Anxiety Disorders	80% female, mean age 40.1 (9.5)	NR	Structured Clinical Interview for DSM□ IV - axis II (SCID II): PD cluster C index (baseline evaluation and 1 year follow-up)	10	Whole	Video	No	Norway	5/6 (low)
Liness et al (2019a) - during training sample	360	45	CBT	Mixed	NR	73% female, median age 33	PHQ-9	NR	Whole	Audio	No	UK	5/6 (low)
Liness et al (2019a) - training follow-up sample	360	45	CBT	Mixed	NR	73% female, median age 34	PHQ-9	NR	Whole	Audio	No	UK	
Liness et al (2019b) - Depression cases	520	252	CBT	Mixed	NR	79.76% female, mean age 34.23	PHQ-9	NR	Whole	Audio	No	UK	5/6 (low)
Martino et al (2008) - competence	461	35	MI/MET	Substance Use/Addiction	29% female, M age 35.3 (9.7)	60% female, mean age 38.9 (11.8)	Timeline followback method (days primary percent drug abstinence)	100	Sub-scale	Audio	No	US	5/6 (low)
Norrie et al (2013)	54	5	CBT	Mixed	NR	NR	Number of suicidal acts in past 12 months	NR	Whole	Audio	No	UK	4/6 (moderate)
Pavio et al (2004) - competence	37	10	EFT	Trauma/PTSD	78% female, mean age 38 (11.32)	70% female, mean age 34.2 (7.41)	Resolution Scale (RS)	NR	Whole	Audio	No	Canada	4/6 (moderate)
Sachs et al (1983) - Experiential/client-centred sample	9	4	Experiential/client-centred	Mixed	100% male, mean age not reported	NR	Overall Improvement (Assessments of the patient's global improvement made by the patient, the therapist, and the independent clinician were summed)	NR	Sub-scale	Audio	No	US	4/6 (moderate)
Sachs et al (1983) - Psychodynamic sample	9	3	Psychodynamic	Mixed	100% male, mean age not reported	NR	Overall Improvement (Assessments of the patient's global improvement made by the patient, the therapist, and the independent clinician were summed)	NR	Sub-scale	Audio	No	US	
Shaw et al (1999) - competence	36	8	CBT	Depression	NR	75% female	Hamilton Rating Scalefor Depression (HRSD)	45	Whole	Video	No	US	6/6 (low)

Strunk et al (2010)	51	6	CT	Depression	58% female, mean age 40 (12)	33.33% female, mean age 45 (4)	Hamilton Rating Scalefor Depression (HRSD)	NR	Whole	Video	No	US	4/6 (moderate)
Svartberg et al (1992)	15	8	Short term anxiety provoking psychotherapy (STAPP)	Mixed	60% female, mean age 30	NR	Symptom Check List-90-R: Global Severity Index	5	Whole	Transcripts	No	Norway	4/6 (moderate)
Trepka et al (2004)	30	6	CT	Depression	70% female, mean age 34.3 (9.36)	100% female	BDI	5	Whole	Not Reported	Yes	UK	5/6 (low)
Weck et al (2011) - competence	34	10	CT	Anxiety Disorders	41.2% female, mean age 34.8 (9.32)	80% female, mean age 33.7 (7)	Clinical Global Impression Scale-Improvement (CGI-I)	6.25	Whole	Video	No	Germany	6/6 (low)
Weck et al (2013) - competence	80	26	CBT	Depression	68.6% female, mean age 48.3 (11.6)	80.8% female, mean age 33.8 (8.7)	time to the first relapse	6.25	Whole	Video	Yes	Germany	5/6 (low)
Weck et al (2015) - cognitive therapy competence	35	17	CT	Anxiety Disorders	57.14% female, mean age 38.14 (10.54)	76.47% female, mean age 30.94 (4.01)	Yale–Brown Obsessive Compulsive Scale for Hypochondriasis (H-YBOCS)	16.67	Whole	Video	No	Germany	4/6 (moderate)
Weck et al (2015) - exposure therapy competence	33	17	Exposure Therapy	Anxiety Disorders	54.55% female, mean age 41.67 (12.52)	82.35% female, mean age 30.35 (3.87)	Yale–Brown Obsessive Compulsive Scale for Hypochondriasis (H-YBOCS)	16.67	Whole	Video	No	Germany	
Westra et al (2011)	32	4	CBT	Anxiety Disorders	71.88% female	50% female, mean age 29	Penn State Worry Questionnaire (PSWQ)	25	Whole	Audio	No	Canada	4/6 (moderate)
Zelencich et al (2020)	31	6	CBT	Mixed	29% female, mean age 47.32 (15.26)	NR	Hospital Anxiety and Depression Scale (HADS) - anxiety subscale / Depression Anxiety Stress Scales (DASS)-depression subscale	55.56	Sub-scale	Audio	Yes	Australia	4/6 (moderate)

Notes. NR = Not Reported, CT = Cognitive Therapy, CBT = Cognitive Behaviour Therapy, MBCT = Mindfulness Based Cognitive Therapy, IPT = Interpersonal Psychotherapy, MI/MET = Motivational Interviewing/Motivational Enhancement Therapy, BDI = Beck Depression Inventory, HRSD = Hamilton Rating Scale for Depression, STAGE-12 = Stimulant Abuser Groups to Engage in 12-Step (STAGE-12) intervention.

Table 1c

Non-Hierarchical Integrity-Outcome Study Characteristics

Study / Sample	Patient N	Therapist N	Treatment Group	Problem Targeted	Patient Demographics	Therapists Demographics	Outcome Measure	% coverage (per patient)	Aspect of measure used	Type of session data rated	Alliance controlled	Country	RoB assessment
Cox et al (2011)	461	40	MI/MET	Obesity	66% female, mean age 59.8(13.9)	NR	Likert scale: patient motivation to lose weight	100%	Sub-scale	Audio	No	US	5/6 (low)
Gaume et al (2018) - MET sample	217	not reported	MI/MET	Substance Use/Addiction	26.7% female, mean age 42.5 (10.2)	NR	Form 90	33.33	Sub-scale	Video	No	UK	5/6 (low)
Gaume et al (2018) - SBNT sample	159	not reported	Social Behaviour and Network Therapy	Substance Use/Addiction	23.9% female, mean age 42.5 (9.6)	NR	Form 90	12.5	Sub-scale	Video	No	UK	
Haug et al (2016)	82	22	CBT	Anxiety Disorders	58.5% female, mean age 33.6 (10.3)	63.6% female, mean age 43.5 (11.3)	Clinician Severity rating	16.67%	Whole	Video	No	Norway	6/6 (low)
Spohr et al (2016)	80	2	MI/MET	Substance Use/Addiction	35% female, mean age 34.9 (12)	NR	Treatment initiation (behavioural outcome)	100%	Sub-scale	Audio	No	US	5/6 (low)

Notes. NR = Not Reported, CBT = Cognitive Behaviour Therapy, CPT = Cognitive Processing Therapy, MI/MET = Motivational Interviewing/Motivational Enhancement Therapy.

Table 1d

Hierarchical Adherence-Outcome Study Characteristics

Study	Patient N	Therapist N	Treatment Group	Problem Targeted	Patient Demographics	Therapist Demographics	Hierarchical Analysis	Outcome Measure	% coverage (per patient)	Aspect of measure used	Type of session data rated	Alliance controlled	Country	RoB assessment
Folke et al (2017)	36	4	CBT	Eating Disorders	97.22% female, mean age 25.7 (5.3)	75% female	Yes	Frequency of binge eating	5	Whole	Audio	No	Denmark	5/6 (low)
Guydish et al (2014) - adherence	151	NR	STAGE-12	Substance Use/Addiction	60% female, mean age 39.1 (10.4)	NR	Yes	Anxiety sensitivity index (ASI)-lite: drug composite	100	Sub-scale	not reported	Yes	US	5/6 (low)

Hauke et al (2013)	220	58	CBT	Anxiety Disorders	72.7% female, mean age 34.9 (10.6)	NR	Yes	Panic and Agoraphobia Scale (PAS)		Whole	Video	No	Germany	4/6 (moderate)
Katz et al (2019)	46	25	Short term psychodynamic psychotherapy (STDPP)	Depression	65% female, mean age 31.8 (11.2)	48% female, mean age not reported	Yes	the Brief Symptom Inventory Depression subscale (BSI-DEP)	6.35	Sub-scale	Video	No	US	5/6 (low)
Marques et al (2019) - adherence	58	19	CPT	Trauma/PTSD	69% female, mean age 39.9 (13.39)	78.9%, mean age 45.7 (13.8)	Yes	Posttraumatic stress disorder symptoms (PCL-S)	100	Whole	Audio	No	US	5/6 (low)
Snippe et al (2018) - CBT	32	12	CBT	Depression	NR	NR	Yes	BDI-II	25	Whole	Video	No	The Netherlands	5/6 (low)
Snippe et al (2018) - MBCT	29	9	MBCT	Depression	NR	NR	Yes	BDI-II	25	Whole	Video	No	The Netherlands	
Tschuschke et al (2015)	81	30	Mixed	Mixed	56.79% female, mean age 39.6 (11.8)	60% female, mean age 50.4	Yes	Brief Symptom Inventory: Global severity index (BSI-GSI), outcome questionniare-45 (OQ-45) and global assessment of functioning scale (GAF) - pre and post were all combined to result in a single 'outcome T-score' for each patient		Whole	Audio	Yes	Switzerland	5/6 (low)

Notes. NR = Not Reported, CT = Cognitive Therapy, CBT = Cognitive Behaviour Therapy, MBCT = Mindfulness Based Cognitive Therapy, IPT = Interpersonal Psychotherapy, MI/MET = Motivational Interviewing/Motivational Enhancement Therapy, BDI = Beck Depression Inventory, HRSD = Hamilton Rating Scale for Depression.

Table 1e

Hierarchical Competence-Outcome Study Characteristics

Study name	Patient N	Therapist N	Treatment Group	Problem Targeted	Patients Demographics	Therapists Demographics	Outcome Measure	% coverage (per patient)	Aspect of measure used	Type of session data rated	Alliance controlled	Country of study	RoB assessment
Bisseling et al. (2019)	84	9	MBCT	Mixed	83.3% female, M age 52.6 (10.7)	NR	Hospital Anxiety and Depression Scale (HADS)	25	Whole	Video	No	The Netherlands	5/6 (low)
Despland et al (2009)	78	15	Psychodynamic	Mixed	58.97% female, mean age 29.2(8.5)	NR	Symptom Check List-90-R	25	Whole	Video	Yes	Switzerland	4/6 (high)
Easden et al (2018)	28	7	CBT	Depression	64.29% female, mean age 44.75 (11.51)	100% female, mean age 36.86 (11.60)	BDI-II	100	Whole	Video	No	New Zealand	3/6 (high)
Guydish et al (2014) - competence	151	NR	STAGE-12	Substance Use/Addiction	60% female, mean age 39.1 (10.4)	NR	Anxiety sensitivity index (ASI)-lite: drug composite/subscale	100	Sub-scale	not reported	Yes	US	5/6 (low)
Kazantzis et al (2018)	50	4	CBT	Depression	76% female, mean age 39.2	sex not reported, mean age 43.5	BDI	NR	Whole	Audio	Yes	US	3/6 (high)
Marques et al (2019) - competence	58	19	CPT	Trauma/PTSD	69% female, mean age 39.9 (13.39)	78.9%, mean age 45.7 (13.8)	Posttraumatic stress disorder symptoms (PCL-S)	100	Whole	Audio	No	US	5/6 (low)
Wurman (2019)	68	17	Psychodynamic	Depression	67.6% female, mean age 36.6 (11.3)	Only 10/17 gave personal information - 90% female, mean age 52.6 (5.4)	Hamilton Rating Scalefor Depression (HRSD)	31.25	Sub-scale	Audio	No	UK	6/6 (low)
Yew et al (2019)	50	4	CBT	Depression	76% female, mean age 39.22 (8.78)	NR	BDI	NR	Whole	Audio	No	US	3/6 (high)

Notes. NR = Not Reported, CT = Cognitive Therapy, CBT = Cognitive Behaviour Therapy, MBCT = Mindfulness Based Cognitive Therapy, IPT = Interpersonal Psychotherapy, MI/MET = Motivational Interviewing/Motivational Enhancement Therapy, BDI = Beck Depression Inventory, HRSD = Hamilton Rating Scale for Depression, STAGE-12 = Stimulant Abuser Groups to Engage in 12-Step (STAGE-12) intervention.

Table 1f

Hierarchical Integrity-Outcome Study Characteristics

Study / Sample	Patient N	Therapist N	Treatment Group	Problem Targeted	Patient Demographics	Therapists Demographics	Outcome Measure	% coverage (per patient)	Aspect of measure used	Type of session data rated	Alliance controlled	Country	RoB assessment
Gaume et al (2009)	95	5	MI/MET	Substance Use/Addiction	22% female, mean age 38	NR	Mean alcohol consumption	100%	Sub-scale	Audio	No	Switzerland	5/6 (low)
Holder et al (2018)	72	4	CPT	Trauma/PTSD	83.33% female, mean age not reported	100% female	The PTSD checklist (PCL)	12	Whole	Video	No	US	4/6 (moderate)
Kramer-Schmidt et al (2019)	238	7	MI/MET	Substance Use/Addiction	35.29% female, mean age 65.13 (4.01)	NR	The Drinkers Inventory of Consequences (DrInC): abstain (categorical outcome)	NR	Sub-scale	Audio	No	International (Denmark, Germany and USA)	5/6 (low)

Notes. NR = Not Reported, CBT = Cognitive Behaviour Therapy, CPT = Cognitive Processing Therapy, MI/MET = Motivational Interviewing/Motivational Enhancement Therapy.

Appendix F

Risk of Bias Assessment Summary Table

Study	Representativeness of the exposed cohort	Ascertainment of exposure	Outcome not present at start of study	Assessment of outcome	Long enough follow-up	Adequacy of follow-up	Overall RoB rating
Abel2016.pdf	truly representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	5/6 (low)
Ablon2006.pdf	selected/self-selected	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	4/6 (moderate)
Barber1996.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	5/6 (low)
Barber2008.pdf	selected/self-selected	secure record (*)	yes (*)	independent blind assessment (*)	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	5/6 (low)
Bisseling 2019.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	5/6 (low)
Boyle 2020.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or	5/6 (low)

						description of those lost is provided (*)	
Branson2015.pdf	truly representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	5/6 (low)
Chevron1983.htm	selected/self-selected	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	4/6 (moderate)
Cox2011.pdf	truly representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	5/6 (low)
Despland2009.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	not reported	4/6 (high)
Easden2018.pdf	selected/self-selected	secure record (*)	yes (*)	self-report	yes (*)	not reported	3/6 (high)
Feeley1999.pdf	no description of selection	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	4/6 (moderate)
Folke2017.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	5/6 (low)
Gaume2009.pdf	truly representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	5/6 (low)

Gaume2018.pdf	no description of selection	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	4/6 (moderate)
Gibbons2010b.pdf	truly representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	5/6 (low)
Goldman2009.pdf	no description of selection	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	4/6 (moderate)
Guydish2014.pdf	truly representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	5/6 (low)
Haug2016.pdf	truly representative of the population (*)	secure record (*)	yes (*)	independent blind assessment (*)	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	6/6 (low)
Hauke2013.pdf	selected/self-selected	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	4/6 (moderate)
Hilsenroth2003.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	independent blind assessment (*)	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	6/6 (low)
Hoffart2005.pdf	selected/self-selected	secure record (*)	yes (*)	independent blind assessment (*)	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	5/6 (low)

Holder2018.pdf	selected/self-selected	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	4/6 (moderate)
Huppert2006.pdf	selected/self-selected	secure record (*)	yes (*)	independent blind assessment (*)	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	5/6 (low)
Katz2019.pdf	truly representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	5/6 (low)
Kazantzis2018.pdf	no description of selection	secure record (*)	yes (*)	self-report	yes (*)	not reported	3/6 (high)
Kramer-Schmidt2019.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	5/6 (low)
Liness2019a.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	5/6 (low)
Liness2019b.pdf	truly representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	5/6 (low)
Luborsky1985.pdf	truly representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	5/6 (low)
Marques2019.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp	5/6 (low)

						rating accounted for (*)	
Martino2008.pdf	truly representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	5/6 (low)
McCarthy2016.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	independent blind assessment (*)	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	6/6 (low)
Norrie2013.pdf	no description of selection	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	4/6 (moderate)
Owen2014.pdf	truly representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	5/6 (low)
Paivio2004.pdf	selected/self-selected	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	4/6 (moderate)
Pitman2014.pdf	truly representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	5/6 (low)
Shaw1999.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	independent blind assessment (*)	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	6/6 (low)
Sinai2012.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or	5/6 (low)

						description of those lost is provided (*)	
Snippe2018.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	5/6 (low)
Spohr2016.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	5/6 (low)
Strunk2010.pdf	no description of selection	secure record (*)	yes (*)	independent blind assessment (*)	yes (*)	not reported	4/6 (moderate)
Svartberg1992.pdf	selected/self-selected	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	4/6 (moderate)
Thyrian2007.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	5/6 (low)
Trepka2004.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	5/6 (low)
Tschuschke2015.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	5/6 (low)
Webb2012.pdf	no description of selection	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or	4/6 (moderate)

						description of those lost is provided (*)	
Weck2011.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	independent blind assessment (*)	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	6/6 (low)
Weck2013.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	independent blind assessment (*)	yes (*)	not reported	5/6 (low)
Weck2015.pdf	no description of selection	secure record (*)	yes (*)	independent blind assessment (*)	yes (*)	not reported	4/6 (moderate)
Yew2019.pdf	no description of selection	secure record (*)	yes (*)	self-report	yes (*)	not reported	3/6 (high)
Zelencich2020.pdf	no description of selection	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	4/6 (moderate)
Farmer2016.pdf	no description of selection	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	4/6 (moderate)
Gower2011.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	5/6 (low)
Hall2007.PDF	selected/self-selected	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	4/6 (moderate)

Minonne_2008.pdf	no description of selection	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	4/6 (moderate)
Ogrodniczuk 1997.pdf	somewhat representative of the population (*)	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	5/6 (low)
Patton1997.pdf	no description of selection	secure record (*)	yes (*)	self-report	yes (*)	complete follow-up: all pts with comp rating accounted for (*)	4/6 (moderate)
Sachs1983.pdf	no description of selection	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	4/6 (moderate)
Spektor2008.pdf	selected/self-selected	secure record (*)	yes (*)	self-report	yes (*)	not reported	3/6 (high)
Westra2011.pdf	selected/self-selected	secure record (*)	yes (*)	self-report	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	4/6 (moderate)
Wurman 2019.pdf	truly representative of the population (*)	secure record (*)	yes (*)	independent blind assessment (*)	yes (*)	subjects lost to follow-up <50% or description of those lost is provided (*)	6/6 (low)

Appendix G

Summary of Moderator Analysis Viability

Moderator	Competence		Adherence		Integrity	
Categorical	HA	Non-HA	HA	Non-HA	HA	Non-HA
1. Problem targeted	N	Y	N	Y	N	N
2. Treatment modality	N	Y	N	Y	N	N
3. Alliance confound	N	Y	N	Y	N	N
4. Treatment format	N	N	N	N	N	N
5. Timing of rating	N	N	N	N	N	N
Continuous						
1. Year of publication	Y	Y	Y	Y	N	N
2. Percentage coverage	Y	Y	Y	Y	N	N
3. Quality rating	Y	Y	Y	Y	N	N

Notes. HA = Hierarchical Analysis, Non-HA = Non-Hierarchical Analysis. Y = yes/viable, N = no/not viable

This page is intentionally blank

PART TWO: Research Report

The Relationship Between Practitioner Competence and Outcome During Low Intensity Guided Self-Help for Anxiety

Abstract

Objectives

There is scarce research on the relationship between practitioner competence in low intensity psychological interventions and treatment outcome. Therefore, this study investigated the association between practitioner competence in delivering guided self-help (GSH) and clinical outcome, rate of clinical change, need for further intervention, attendance and drop out.

Design

Data from a randomised controlled patient preference trial of GSH for anxiety disorders in an Improving Access to Psychological Therapies service were used. The trial compared GSH based on cognitive analytic therapy (CAT-GSH) with standard GSH based on cognitive behaviour therapy (CBT-GSH).

Method

Sessions were randomly sampled from the 6-8 session manualised treatments (CAT-GSH, n=60 and CBT-GSH, n=20) and rated using a validated competence rating tool. A 'fully crossed' independent rating design was used with two independent rating groups. Competence ratings were then used to explore associations with anxiety, depression and functioning, treatment engagement (attendance and drop-out), and need for further intervention.

Results

Competence ratings had very high inter-rater reliability across all competence levels and in both treatments (intraclass correlation coefficients between 0.85 and 0.99). No significant associations were found between competence and clinical outcome, treatment engagement or need for further intervention. Competence did not predict rate of change in anxiety symptoms.

Conclusions

Practitioners were competent in the delivery of either version of GSH, but competency did not predict clinical or service outcomes. Analyses were under-powered and exploratory and directions for future LI competency-outcome research are provided.

Practitioner Points and Limitations

- The competency of GSH delivery is an important consideration in the supervision of Psychological Wellbeing Practitioners (PWP).
- PWPs appear able (under supervision) to deliver two distinct versions of GSH.
- How competency relates to outcome needs further investigation for low intensity psychological interventions.
- The main limitation in this study was the unequal distribution of treatments, created by the marked patient preference for the CAT-GSH approach.

Keywords

Therapist Competence; Clinical Outcome; IAPT; Process-Outcome; Guided Self-Help.

Introduction

Background

Common mental disorders (CMDs) include depression and various forms of anxiety disorder (e.g., panic disorder, phobias, etc.; Baker 2020). These disorders not only have a large detrimental impact on individuals but also on their families and society as a whole. Internationally, the lifetime prevalence of CMD is estimated to be around 29.2% (Steel et al., 2014). Standardised clinical interviewing suggests that around 15.7% (roughly one in six) adults experience symptoms of CMD, with around 8.1% experiencing severe symptoms (Stansfeld et al., 2014). It is clear, therefore, that CMD represent a major health problem globally and so developing and delivering effective interventions is a key priority. Evolution in the scope and scale of the evidence base for CMD interventions has resulted in a number of effective psychological interventions being recommended by the National Institute for Health and Clinical Excellence (NICE; NICE, 2011).

IAPT

In order to provide evidence-based psychological therapies identified in the CMD NICE guidelines, the Improving Access to Psychological Therapies (IAPT) initiative was funded and implemented in England (Layard & Clark, 2014). IAPT services are designed around a 'stepped-care model' whereby early and least intensive interventions occur before the patient "steps up" to more intensive interventions as necessary (Scogin et al., 2003). Stepped care has been found to have significantly better treatment outcomes for anxiety and depression when compared to care as usual (Firth et al., 2015; Ho et al., 2016). IAPT services therefore provide 'low-intensity' interventions (LI) at step 2 and traditional 'high-intensity'

interventions (HI) at step 3 (National Collaborating Centre for Mental Health; NCCMH, 2020).

In an effort to clarify the distinction between high and low intensity CBT, Shafran et al. (2021) defined low intensity CBT as interventions that: 1) utilised self-help materials, 2) involved six hours or less of contact time with each contact being typically 30 minutes or less, and 3) provided by trained practitioners or supporters. During step 2 interventions, the patient works through the GSH materials with the support of a 'Psychological Wellbeing Practitioner' (PWP; Cujipers et al., 2010). PWPs are practitioners who are specifically trained and qualified to deliver low-intensity interventions. Cuijpers et al. (2010) conducted a meta-analysis of randomised controlled trials (RCT) and found that GSH for anxiety disorders showed moderate-large effect sizes when compared to placebo and small effect size compared to face-to-face psychotherapy.

Although GSH treatments in IAPT have traditionally been based on CBT principles (Baguley et al., 2010), an alternative GSH treatment has recently been developed, based on cognitive analytic therapy (CAT) principles (Meadows & Kellett, 2017). Whereas CBT-GSH emphasises the relationship between a patient's physical, behavioural and cognitive symptoms using cognitive and behavioural change methods (Baguley et al., 2010), CAT-GSH emphasises reformulating, recognising and then revising the relational patterns in which distress originates and is maintained (Meadows & Kellett, 2017). CAT-GSH has demonstrated fidelity to GSH principles and promising feasibility and effectiveness evidence in a small-scale case-series design (Meadows & Kellett, 2017).

Therapist Competence and Clinical Outcome

One of the key constructs measured which gives an indication to the 'quality' of an intervention being delivered is the competence of the therapist/practitioner delivering the intervention. Along with therapist adherence, therapist competence can be considered as one of the 'specific factors' responsible for clinical improvements of patients, as opposed to 'common factors' which apply across therapies (e.g., therapeutic alliance; Castonguay, 1993). According to Sharpless and Barber (2009), 'adherence' refers to the extent to which a therapist delivers theory-specific techniques/interventions, whereas 'competence' refers to the skilfulness with which an intervention is implemented. Various studies to date have investigated the construct of therapist competence and its associations with patient clinical outcomes, but these have tended to have been of psychological therapy.

Webb et al. (2010) for example conducted a meta-analytic review of 36 studies to establish the relationships between therapist intervention-adherence, therapist competence and clinical outcomes of patients. Combining all studies included in the review, Webb et al. found that neither therapist adherence nor competence were significantly associated with patient outcome. They concluded that adherence and competence may not play a significant role in determining patient outcome in the modalities and presentations reviewed. Since 2010, various primary studies have been conducted which appear to continue the equivocal pattern of findings.

Studies conducted after the Webb et al. (2010) meta-analysis have shown higher therapist competence being associated with improved patient outcomes across various interventions and presenting problems, including cognitive therapy (CT) for depression (Strunk et al., 2010), social anxiety disorder (Ginzburg et al., 2012) and hypochondriasis (Weck et al., 2015), CBT for patients diagnosed with Borderline Personality Disorder (Norrie et al., 2013),

a 12-step intervention for patients with problematic substance use (Guydish et al., 2014) and improved post-traumatic stress symptoms in patients treated with cognitive processing therapy (CPT; Marques et al., 2019). It does, therefore, appear that therapist competence may have some role to play in determining client outcomes across therapy models and clinical presentations, but the current evidence base is solely grounded in studies of formal psychological therapy.

In addition to understanding whether there is an association between therapist competence and overall patient outcome, a valuable area of investigation has been exploring the rate at which clients experience clinical changes. The hypothesis being that more competent therapists enable more rapid change. Strunk et al. (2010) investigated CT for depression and found competence ratings significantly predicted early-treatment session-to-session symptom change. More recently, Wurman (2019) investigated therapist competence in high-intensity dynamic interpersonal therapy (DIT) and found that higher levels of competence were associated with a faster rate of clinical improvement.

Competence and Outcome in IAPT

There have been various studies investigating the competence of therapists delivering step 2 and step 3 interventions in IAPT services. Branson et al. (2015) found that a higher proportion of patients at step 3 of the most competent trainee high-intensity CBT therapists demonstrated a reliable improvement in anxiety. Additionally, a significantly higher proportion of patients seen by the least competent therapists experienced a reliable deterioration in anxiety. In analyses of competence and outcome data from trainee and qualified high-intensity CBT therapists, there have been non-significant positive trends between competence and depression and anxiety outcomes (Liness et al., 2019a), and

competence predicting depression, but not anxiety outcomes (Liness et al., 2019b). It seems, therefore, that there may be an association between CBT therapist competence and patient outcome in a high-intensity IAPT service context, although this appears to be relatively inconsistent and the evidence is based on trainee and not qualified therapists. It is also largely unclear whether these associations are present in an LI context where interventions tend to be characteristically different (e.g., shorter and more structured) compared to formal psychological therapy. This is problematic due to the British Government's national guidance suggesting that roughly 40% of a core IAPT service should be PWPs and therefore many patients receive an LI intervention annually (NCCMH, 2020).

Branson et al. (2018) assessed the competence of trainee and qualified PWPs by evaluating performance in Observed Structured Clinical Examinations (OSCE) of LI CBT and linked this to the clinical outcomes of their clients. They found more patients of the highest-competence PWPs showed reliable symptom improvement in anxiety and depression than would be expected by chance. Additionally, fewer patients of the most competent PWPs deteriorated when compared to those seen by the least competent PWPs. In an exploratory study of the associations between therapist effects and patient outcome, Delgadillo et al. (2020) used a multilevel modelling approach with the same dataset. This study found that PWP personality factors (such as 'agreeableness') were related to outcome, although clinical competence (measured by the OSCE) was not associated with patient outcome when various PWP factors (e.g., age, sex, years of experience) were controlled for.

Key limitations of these LI competence-outcome studies, however, were that the operationalisation of competence (i.e., OSCE scores) did not have robust psychometric

properties, may have had low ecological validity due to ratings being based on 'actor-patient' sessions and also may have measured adherence as opposed to competency.

Furthermore, although 'therapist factors' such as skills, training and experience have been associated with psychotherapy drop-out (Roos & Werbart, 2013), most studies have focused primarily on the links between client characteristics and drop-out (Villeneuve et al., 2010). Thus, the association between therapist competence and treatment drop-out is under-investigated, particularly in an LI context. In one of the few studies to investigate this, Haug et al. (2016) found that lower therapist competence was associated with higher drop-out.

The Current Study

The current study aimed to further investigate the impact of PWP competence on patient outcome and engagement in LI for anxiety disorders. This was done through use of data collected during the first phase of a pragmatic patient preference randomized controlled trial (see Kellett et al., 2021 for the published protocol) and with qualified PWPs. Firstly, the study will extend previous investigations by examining the association between PWP competence and patient outcome in step 2 GSH interventions using an empirically developed measure of practitioner competence. Secondly, in addition to establishing the association between competence and outcome in GSH, the study will enable preliminary comparisons of this association between CAT-GSH and CBT-GSH. Thirdly, the study will be the first study to explore the association between PWP competence and treatment engagement (session attendance and drop-out) and the need for further intervention (i.e., the 'step-up' rate to step 3) across and within two GSH treatments. Finally, the study will provide a preliminary investigation of the association between PWP competence and the rate of patient clinical

change, as this has not been attempted previously with LI interventions, despite the need for rapid change given the brevity of the treatment contract.

Aims

1. Investigate the association between PWP competence and patient clinical outcome across and between CBT-GSH and CAT-GSH for anxiety disorders.
2. Investigate the association between PWP competence and secondary outcomes including treatment engagement (attendance and drop-out) and stepping up rates across and between CBT-GSH and CAT-GSH for anxiety disorders.
3. Investigate the association between PWP competence and rate of patient anxiety outcome change across CBT-GSH and CAT-GSH for anxiety disorders.

Hypotheses

1. Higher PWP competence will be significantly associated with improved patient outcomes (anxiety, depression and functioning), although there will be no differences between CBT-GSH and CAT-GSH groups.
2. Higher PWP competence will be significantly associated with lower patient dropout and increased treatment session attendance across the overall sample, although there will be no differences between CBT-GSH and CAT-GSH groups.
3. PWP competence level will be significantly associated with step-up rates within and across treatment conditions although there will be no differences between CBT-GSH and CAT-GSH groups (exploratory).
4. PWP competence will predict patient rate of change in sessional anxiety outcomes across the overall sample (exploratory).

Method

Design

The current study was a quantitative analysis of data obtained from a patient preference randomised controlled trial (see Kellett et al., 2021, for full protocol). The trial set out to investigate whether the CAT-GSH intervention was equally efficacious when compared to treatment as usual (CBT-GSH). After engaging in the informed consent process of the trial (see Appendix A for the participant-facing documents), participants chose or were randomised to either receive CBT-GSH or CAT-GSH. All participants were patients referred to the service due to anxiety. Anxiety disorders were defined and measured through conducting the Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1997) at the research screening interview. The primary outcome was the Beck Anxiety Inventory (BAI; Beck et al., 1988). Secondary outcomes included: number of sessions attended, drop-out rate, stepping up rate and the IAPT minimum dataset (MDS) measures: the Generalised Anxiety Disorder-7 (GAD-7; Spitzer et al., 2006), Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001), Work and Social Adjustment Scale (WSAS; Mundt et al., 2002). These outcome measures were collected at baseline and every treatment session attended. The secondary outcomes of the trial (session attendance, drop-out rate, step-up rate and MDS outcome measures; see Appendix B) formed the primary outcomes of the current study. During the trial, each participant had one treatment session from the 6-8 session treatment contract randomly selected and audio recorded. The recordings were then assessed for competency.

Ethical Considerations

The main trial was ethically reviewed and approved (see Appendix C): IRAS reference number, 240751. The study was also pre- registered: ClinicalTrials.gov identifier,

NCT03730532. Patient participants in the trial consented to one of their treatment sessions being recorded and then competency rated by the research team (the research team included the raters in the current study). The trial moved to telephone sessions due to the COVID-19 pandemic (ethics approval amendment was made). The current study was also pre-registered: https://aspredicted.org/blind.php?x=xa7ns9.

Setting and Participants

The trial was based in an NHS IAPT service in the North of England where the Local Authority has been ranked in the 20 most deprived of 317 Local Authorities in the country (Local Government Association, 2019). The IAPT service offers evidence-based psychological interventions for individuals aged 16+ years with common mental health problems. Patients were typically seen at community-based clinics when face-to-face LI work was undertaken. Step 2 interventions in the service previously consisted of 6-8 sessions (30-35 minutes in duration) of guided self-help cognitive behaviour therapy (CBT-GSH) with a PWP.

The patient participant eligibility criteria used in the main trial were as follows:

Inclusion criteria:

- Referred by their GP or other health or social care professional or self-referred.
- Aged 18 years or older.
- Met criteria for an anxiety disorder on the MINI screening interview assessment and scored >10 on the BAI at assessment.
- An anxiety disorder was the patient's primary disorder.
- Motivated to engage in GSH.
- Motivated to engage in treatment and could attend at least six sessions of face-to-face GSH.

Exclusion criteria:

- Engaged in another IAPT step 2 intervention at time of recruitment.
- Did not meet criteria for an anxiety disorder as defined by the MINI and the BAI score.
- Met criteria for depression and a comorbid anxiety disorder where the depression was more severe and was the patient's main concern.
- Were involved in psychiatric or secondary care mental health services.
- Had a diagnosis of social phobia or PTSD (IAPT guidelines indicate that these disorders are treated at step 3).
- The GSH sessions needed an interpreter.
- Unable to read and write.

Sample

Patient Participants

Data from a sub-sample of patients in the main trial were used in the current study. The sample size was constrained by two factors: 1) the main trial was still collecting data when the current study commenced, and 2) audio-recording of sessions was randomised and thus some sessions were not captured before the patient dropped out or were discharged early due to recovery.

A total of 96 recordings were screened for use in the current study. After excluding unrateable (e.g., corrupt/inaudible) and training recordings, the number of recordings available was *n*=80. The CAT-GSH group had a substantially larger sample of participants (*n*=60) compared to the CBT-GSH group (*n*=20). This mirrored the larger pattern in the trial, where *n*=180 patients chose to receive CAT-GSH compared to *n*=73 who chose to receive CBT-GSH (see Appendix D for trial newsletter). The reason for this appeared to be that patients were

choosing CAT-GSH because of returning the service and not wanting the intervention they had previously received (i.e., always a CBT-informed intervention). Of the CAT-GSH group, 50% had previously received psychological intervention compared to 25% of the CBT-GSH group (see Table 1 for summarised patient characteristics).

PWP Participants

All PWPs in the IAPT service took part in the trial and delivered both GSH intervention conditions. All *n*=19 PWPs (one male) were qualified (i.e., had all completed a 1-year Post Graduate Certificate course in Low-Intensity Interventions accredited by the British Psychological Society). PWPs had a mean age of 31.31 (*SD* = 5.07) years. Mean years of post-qualification clinical experience was 4.75 (*SD* = 4.43). All PWPs were in two clinical skills supervision groups (one for each intervention) and in receipt of weekly case management supervision.

Independent Raters

Six expert independent raters (four females) were recruited to provide independent ratings of a sub-sample of recordings to determine inter-rater reliability of competence. All independent raters were all employed as senior PWPs in the NHS and as teaching staff at a University accredited by the British Psychological Society to deliver clinical training to PWPs. Independent raters had between two and six years of experience routinely rating trainee PWP audio recordings using a competence rating tool as part of course requirements.

CAT Expert Raters

Three qualified CAT practitioners (all female and with at least 10 years post-certification experience) were recruited to provide independent verification of a randomly selected CAT-GSH session. The aim of the CAT expert rating process was to 1) verify that CAT theory and

techniques were being adhered to in CAT-GSH sessions, and 2) to further assess inter-rater reliability.

Measures

Therapist Competence

Therapist competence for both interventions was rated using the Low Intensity Treatment Competency scale (LITC; Kellett, Simmonds-Buckley et al., 2021), a six-item measure of LI competence. The LITC has been shown to have good internal consistency (α=.84) and moderate-good interrater reliability (.74) and have adequate construct and predictive validity. A total score of 18 on the LITC differentiates competent and incompetent LI-CBT treatment sessions. This cut off was based on the scoring methods of extant high-intensity competency measures (requiring that each item scores 3 or more to be classed as 'competent'). The specific prompts on the LITC were amended slightly for the description of CAT-GSH change method specification, although the wording of each of the six items rated were identical. See Appendix E for copies LITC measures used for CBT-GSH and CAT-GSH sessions.

Anxiety

Anxiety was measured using the GAD-7. The GAD-7 is a seven-item, self-report measure based on the Diagnostic and Statistical Manual (DSM-IV, American Psychiatric Association; APA, 2000) diagnostic criteria for generalised anxiety. Scores range from 0-21 with higher scores suggesting a greater severity of anxiety and scores of 5, 10, and 15 represent clinical cut-offs for mild, moderate and severe anxiety respectively. The GAD-7 was found to have good internal reliability (α = .92; Spitzer et al., 2006). A change score of >=5 points was deemed to represent a reliable change on the GAD-7 (Richards & Borglin, 2011). Caseness was defined as a score of >8 points on the GAD-7 (NCCMH, 2020).

Depression

Depression was measured using the PHQ-9. The PHQ-9 is a nine-item self-report measure of depression with items being derived from DSM-IV (APA, 2000) depression symptom criteria. Total scores range from 0-27 and scores of 5, 10, 15, and 20 represent clinical cut-off for mild, moderate, moderate severe and severe depression respectively. The scale has good internal reliability (α = .87; Kroenke, et al. 2001). A change score of >=6 points was deemed to represent a reliable change on the PHQ-9 (Richards & Borglin, 2011). Caseness was defined as a score of >10 points on the PHQ-9 (NCCMH, 2020).

Functional Impairment

Functional impairment was measured using the WSAS. The WSAS is an eight-item self-report measure of disability or functional impairment, attributable to a specific problem. Total scores range from 0-40 and scores >10 are associated with significant functional impairment. The WSAS has good internal consistency (α = .87) and test-retest reliability (r = 0.73; Mundt et al., 2002). No reliable change nor caseness thresholds were used for the WSAS due these statistics not being available at the time of the current study.

Session Attendance, Drop-Out and Stepping-Up Rate

A treatment completer was defined as a participant that attended at least six sessions or that reached recovery before session six. Session attendance was measured and categorised as: minimal attendance (1-2 sessions), moderate attendance (3-5 sessions) or full attendance (6-8 sessions). Patient drop-out was defined as a patient that did not complete treatment and were discharged from the service. These participants were not discharged early due to clinical recovery, referred for an alternative treatment, or ‘stepped up’ to high-intensity treatment.

'Stepping-up rate' referred to the percentage of patients that were 'stepped up' (i.e., referred for a step 3/high intensity intervention) following step 2/GSH treatment.

Clinical Outcomes

For the three outcome measures which comprised the primary outcomes (GAD-7, PHQ-9 and WSAS), a pre-post change score was calculated for each patient by subtracting a patient's baseline score from their final score. Measures of reliable change (RC) and reliable and clinically significant improvement (RCSI) were also calculated for GAD-7 and PHQ-9. The RCSI highlights whether a patient has made a therapeutic change from having symptoms within the range of a clinical population to being within a 'non-clinical range' (Jacobson et al., 1984). The RC highlights whether a patient's score has changed more than would be expected by measurement error (Jacobson & Truax, 1991). Patients that did not have RC were classed as 'no change'. It was not possible to carry out WSAS RC/RCSI calculations due to the lack of established cut-off values.

Procedure

Treatment

Both treatments were 6-8 sessions and used highly structured workbooks with homework exercises. During appointments, the PWPs therefore supported the patient to work through a structured workbook. The structure of treatment sessions in both treatment groups was similar with the exception of using different self-help materials. All PWPs attended a one-day training on introducing CAT and using the CAT-GSH protocol. Further details of the interventions used in the trial are reported elsewhere (Kellett et al., 2021).

Data Handling

Data were fully anonymised before the analysis was conducted by the primary researcher and stored on a secure encrypted electronic server.

Competence Rating and Training

The research co-ordinator of the main trial (JS) assigned the appropriate participant ID to each audio recording (anonymised) before uploading the recordings to a secure university drive. A third-year Trainee Clinical Psychologist with a Post Graduate Certificate in CBT (NP) attended a three-hour training session on use of the LITC tool from two of the developers of the measure (SK and MSB) to ensure reliable rating. NP was the first rater and listened to and rated all the eligible audio-recordings.

The six independent raters attended one of two, three-hour LITC training sessions. As the raters were all familiar with the LITC, the primary objectives of the training sessions were to 1) introduce the CAT-GSH protocol, and 2) calibrate ratings of two CAT-GSH session recordings. The CAT expert group did not receive training on the LITC tool, due to this process being primarily to verify CAT adherence of CAT-GSH.

Competence Rating Reliability

Independent raters were allocated to two rating groups (Group A and Group B). Each rating group rated the same six sessions (two rated as low competence, two rated medium, two rated as high by the primary rater) which were randomly chosen from the full sample. Half of the rating group's six sessions to be rated were from each treatment condition. This rating matrix was designed in order to create a 'fully crossed' design where multiple subjects are rated by the same set of coders. Although fully crossed designs require a higher number of ratings, this enables systematic bias between raters to be controlled for in an inter-rater reliability estimate such as an intraclass correlation (Hallgren, 2012). Two rating groups were created in order to

rate a higher number of sessions than possible with one group of raters. As each rating group rated six recordings, the total number of recordings that was rated with a fully crossed design was 12.

Recordings to be rated by the independent raters were chosen by collating participant IDs in the order of their attained competence rating. A range of competency ratings were selected for independent rating to assess inter-rater reliability across competency levels as previously conducted (Haddock et al., 2001). Sub-samples of low, medium and high therapist competence were created, and each recording was given a 'sub-sample' ID. A random number generator (using the 'RAND' function in Excel) was then used to select a sub-sample ID at random from each competence-level. To control for order-effects, the sequence in which raters were asked to rate the recordings was counter-balanced by generating a random rating sequence for raters. After independent ratings were conducted, interrater agreement was measured via two-way (target x judges) random-effects intra-class correlation coefficients (ICC; Shrout & Fleiss, 1979) using 'absolute agreement'.

Data Analysis

All data analyses were conducted using Microsoft Excel and SPSS Statistics v26.0. Data analyses were categorised as primary analyses (to test hypotheses 1 and 2) and exploratory analyses (to test hypothesis 3 and 4). Hypotheses were classed as 'exploratory' if they were not pre-specified (hypothesis 3) or were under-powered (hypothesis 4). A series of chi-square, correlation and regression analyses were conducted to test hypotheses. G-Power (version 3.1.9.6) was used to conduct a-priori power calculations for the correlation and multiple linear regression analyses which indicated samples sizes of n=67 and n=85 were required, respectively. The total sample based on the available eligible recordings was n=80. See Appendix F for further details

regarding power calculations. Prior to conducting correlation and regression analyses, the relevant assumptions were tested (see Appendix G). All chi-square and correlation analyses were adjusted for multiple tests.

Descriptive statistics were initially calculated for sample characteristics, PWP competence ratings and outcomes. In terms of hypothesis 1, Pearson Product-Moment correlation coefficients were calculated to establish the association between overall PWP competence and end of treatment clinical outcomes (end of treatment score and pre-post change for GAD-7, PHQ-9 and WSAS) for each treatment condition and the overall sample. Correlations for the overall sample were sufficiently powered, though the within-condition correlations were not. Cohen's definitions were used to interpret correlation coefficients: 'small' (r = .10), 'medium' (r = .30) and 'large' (r = .50) associations (Cohen, 1992). In order to establish whether different competence 'levels' were differently associated with clinical recovery, 'high', 'medium' and 'low' competence levels were created as was a nominal variable of RCSI for anxiety and depression. Competence levels were created using 'competence quartiles' (low = <25%, medium = 25-75%, high = >75%) of overall competence scores (Branson et al., 2018). Chi-squared tests were then performed between competence level and RCSI status for anxiety and depression outcomes, for treatment conditions and the overall sample. To further explore the association between PWP competence and anxiety outcome, a standard multiple regression was conducted. The model used the outcome variable 'end of treatment GAD-7 score', predictor variables 'overall competence' and 'treatment condition', and 'baseline GAD-7 score' and 'session number rated' as covariates. The multiple regression analysis did not have sufficient power (due to at least n=85 being required).

In terms of hypothesis 2, chi-square tests were conducted to investigate the association between ‘competence level’ and ‘treatment drop-out’ (a nominal yes/no variable). To further investigate the relationship between PWP competence and treatment drop-out, a binomial logistic regression was conducted where the outcome variable was ‘treatment drop-out’ (yes/no), predictor variables were ‘overall competence’ and ‘treatment condition’ and covariates were ‘baseline GAD-7 score’ and ‘session number rated’. To test whether there was a significant association between competence level and session attendance (minimal, moderate or full attendance), chi-square tests were conducted. A sensitivity analysis was conducted in order to establish the effect of the timing of competence-rating on the overall competence-outcome and competence-drop-out associations. This was carried out by re-running the multiple regression and logistic regression analyses including only cases that had a competence-rating based on an ‘early’ treatment session (session 1 or 2) and comparing this model to the original models.

For hypothesis 3, chi-squared tests were conducted to investigate the association between ‘competence level’ and ‘stepping up’ (a nominal yes/no variable). In terms of hypothesis 4, a number of exploratory longitudinal Linear Mixed Model (LMM) analyses were developed in order to explore the rate of change in sessional anxiety outcomes and the impact of PWP competence on this rate of change. LMM were conducted in SPSS with Maximum Likelihood (ML) estimation. A two-level hierarchical structure was created where sessional GAD-7 scores (level 1) were nested within the patient (level 2). The area of interest was whether interindividual differences in growth trajectories (e.g., initial anxiety status and pattern of clinical change) were systematically related to PWP competence rating, with the session number rated and treatment group being controlled for. LMM analyses were conducted on all cases except those with only one or two data points available (in order to explore curvilinear change patterns). LMM were

carried out with all cases which had a maximum of 9 data points (baseline + 8 treatment sessions), due to the protocol for the main trial stipulating that treatment was 6-8 sessions. The LMM were developed following published guidance (Singer &Willett, 2003; Shek & Ma, 2011). The model-fit statistic used was -2 Log-Likelihood (-2LL). Chi-square analyses were used to establish whether the -2LL were significantly different.

Firstly, an unconditional model was run to provide a baseline model. Next, a number of individual growth trajectories were tested (linear, log-linear, quadratic and cubic models) as fixed-effects to establish the best fitting time trend. The log-linear model was the simplest (4 parameters) model with the best fitting -2LL statistic and thus this was taken forward (consistent with the rule of parsimony). After unconditional fixed-effects growth models, an unconditional model with log-linear (time trend) as random effects was run in order to test whether allowing intercepts and slopes to vary would improve the model. Then the best-fitting within-individual error covariance structure was selected (after variance components, AR1H and unstructured structures were run). The best fitting covariance structure was an AR1: Heterogenous structure (modelling for auto-correlation in the longitudinal data). After unconditional models were run, the conditional models were run with overall competence (continuous) and time (continuous) entered into the LMM analyses as predictors and session number rated and treatment group entered as covariates. Before continuous covariates/predictors were entered, they were mean-centred. A patient-level interaction term (competence*time) was then added to the model in order to test whether the between-patient variation in change in GAD-7 scores over time was explained by the competence of their PWP, with session number rated and treatment condition being controlled for.

It was expected that the outcome data contained missing data points due to, for instance, non-attendance and treatment drop-out. For all analyses except LMM, missing outcome data was handled by the Last Observation Carried Forward (LOCF) procedure. In terms of LMM models, these procedures are sufficiently flexible to allow for unbalanced data (e.g., unequal sample sizes between groups and missing data) through using maximum likelihood estimation (Rabe-Hesketh & Skrondal, 2008; Shek & Ma, 2011).

Results

Sample Description

Participant characteristics are summarised in Table 1 for each treatment and the overall sample. Characteristics were comparable between treatments with no significant differences in any of the demographics. Similarly, there were no significant differences between groups in sessions attended, step-up rate, drop-out rate, previous psychological intervention (either according to clinical records or self-report), nor treatment allocation method.

Table 1

Summary of Patient Sample Characteristics Overall and per Treatment Condition

			Overall (*n*=80)	CBT-GSH (*n*=20)	CAT-GSH (*n*=60)	Test of Difference
Previous psychological intervention (%)	Yes		35 (43.8)	5 (25)	30 (50)	χ^2 (1) = 3.81, p = 0.051
	No		45 (56.3)	15 (75)	30 (50)	
Treatment allocation method (%)	Patient choice		74 (92.5)	18 (90)	56 (93.3)	χ^2 (1) = 0.24, p = 0.624
	Randomised		6 (7.5)	2 (10)	4 (6.7)	
Mean age (SD)			36.51 (13.83)	38.5 (15.22)	35.85 (13.41)	t(78) = 0.74, p = 0.462
Gender (%)		Male	19 (23.8)	5 (25)	14 (23.3)	χ^2 (1) = 0.02, p = 0.879
		Female	61 (76.3)	15 (75)	46 (76.7)	
Psychotropic medication (%)		Yes	44 (55)	14 (70)	30 (50)	χ^2 (1) = 2.42, p = 0.119
		No	36 (45)	6 (30)	30 (50)	
Ethnicity (%)		White British	73 (91.3)	17 (85)	56 (93.3)	χ^2 (5) = 10.45, p = 0.064
		White Irish	1 (1.3)	1 (5)	0	
		Pakistani	1 (1.3)	1 (5)	0	
		Mixed: White and Black Caribbean	3 (3.8)	0	3 (5)	
		Caribbean	1 (1.3)	0	1 (1.7)	
		Not Stated	1 (1.3)	1 (5)	0	
Marital status (%)		Single	3 (3.8)	0	3 (5)	χ^2 (4) = 4.69, p = 0.320
		Married/Civil Partnership	11 (13.8)	2 (10)	9 (15)	
		Separated	1 (1.3)	0	1 (1.7)	

	Widowed/Surviving Civil Partner	1 (1.3)	1 (5)	0	
	Not Stated	63 (78.8)	16 (80)	47 (78.3)	
Employment status (%)	Full-Time Employment	41 (51.2)	11 (55)	30 (50)	χ^2 (7) = 9.62, p = 0.211
	Part-Time Employment	10 (12.5)	2 (10)	8 (13.3)	
	Unemployed/Benefits	13 (16.3)	1 (5)	12 (20	
	Self-Employed	2 (2.5)	0	2 (3.3)	
	Student	8 (10)	3 (15)	5 (8.3)	
	Retired	4 (5)	3 (15)	1 (1.7)	
	Volunteer	1 (1.3)	0	1 (1.7)	
	Not Stated	1 (1.3)	0	1 (1.7)	
Mean sessions attendance (SD)		6.45 (1.79)	6.4 (1.6)	6.47 (1.85)	t(78)= -0.14, p = 0.886
Session attendance groupings (%)	Minimal (1-2 sessions)	3 (3.38)	0 (0)	3 (5)	χ^2 (2) = 3.64, p = 0.162
	Moderate (3-5 sessions)	14 (17.5)	6 (30)	8 (13.3)	
	Full (6+ sessions)	63 (78.8)	14 (70)	48 (80)	
Stepped-up (%)	Yes	14 (17.5)	2 (10)	12 (20)	χ^2 (1) = 1.04, p = 0.308
	No	66 (82.5)	18 (90)	48 (80)	
Dropped out (%)	Yes	10 (12.5)	1 (5)	9 (15)	χ^2 (1) = 1.37, p = 0.242
	No	70 (87.5)	19 (95)	51 (85)	

Notes. *significant at $p < .05$ threshold, **significant at $p < .01$ threshold, ***significant at $p < .001$ threshold, significant at Bonferroni adjusted threshold for multiple testing in bold (though none are significant at the adjusted threshold).

Competency Ratings

Inter-Rater Reliability of Competence Ratings

Inter-rater reliability statistics were calculated 1) within the rating group only, and 2) across the rating group and the primary rater. As displayed in Table 2, all raters across both rating groups achieved an acceptable level of agreement in ratings of PWP competence, with all ICCs being between 0.85 ('good') and 0.99 ('excellent'; Koo & Li, 2016).

Table 2

A Summary of the ICC for the Individual Rating Groups with and without the Primary Rater Included.

	Rating Group A (6 recordings)		Rating Group B (6 recordings)	
	Rating Group Alone	Rating Group and Primary Rater	Rating Group Alone	Rating Group and Primary Rater
Overall Score only (95% CI)	0.92 (0.68 – 0.99)	0.85 (0.53 – 0.98)	0.86 (0.34 – 0.98)	0.90 (0.6 – 0.98)
Items and Overall Score (95% CI)	0.98 (0.97 – 0.99)	0.98 (0.97 – 0.99)	0.99 (0.98 – 0.99)	0.99 (0.99 – 1.0)

Notes. 'Absolute agreement' was used for ICC. CI = Confidence Intervals.

CAT Expert Group

The CAT expert rating group verified that the main characteristics and foci of CAT were present in the session. The CAT expert group had an ICC of 0.996 (95% CI = 0.98 - 1) alone and 0.99 (95% CI = 0.95 – 1) with the primary rater's ratings included. The ICC calculations indicated that there was 'excellent' agreement between the CAT therapists and between the CAT therapists and the primary rater (Koo & Li, 2016). Taken together the expert PWP and CAT therapist ratings meant that the single rating of the primary researcher were reliable. All following results are based on the single rating of competency by the primary researcher.

PWP Competency

Table 3 displays descriptive statistics of competency-items and overall PWP competency ratings for the total sample and across treatments. Competence was not significantly different between treatment groups (neither by individual item nor overall).

Table 3

Descriptive Statistics of Item and Overall PWP Competency Ratings per Treatment Condition and for the Overall Sample.

Mean Competence Rating (*SD*)	Overall (*n*=80)	CBT-GSH (*n*=20)	CAT-GSH (*n*=60)	Test of Difference
1: Focusing the session	2.55 (0.86)	2.75 (0.79)	2.48 (0.87)	$t(78)=1.21$, $p = 0.230$
2: Continued engagement	3.39 (0.56)	3.38 (0.48)	3.4 (0.59)	$t(78)=-0.17$, $p = 0.864$
3: Interpersonal competencies	3.53 (0.66)	3.73 (0.68)	3.47 (0.64)	$t(78)=1.53$, $p = 0.129$
4: Information gathering	3.15 (0.79)	3.38 (0.72)	3.08 (0.81)	$t(78)=1.48$, $p = 0.144$
5: Within session self-help change method	3.6 (0.6)	3.7 (0.55)	3.57 (0.62)	$t(78)=0.86$, $p = 0.395$
6: Planning and shared decision making	3.26 (0.56)	3.43 (0.63)	3.2 (0.53)	$t(78)=1.56$, $p = 0.122$
Overall Score	*19.48 (2.82)*	*20.35 (2.58)*	*19.19 (2.86)*	$t(78)=1.61$, $p = 0.113$

Notes. *significant at p < .05 threshold, **significant at p < .01 threshold, ***significant at p < .001 threshold, significant at Bonferroni adjusted threshold for multiple testing in bold (though none are significant at the adjusted threshold).

Summary of Clinical Outcomes

A summary of the self-reported clinical outcomes is shown in Table 4. In terms of baseline to end of treatment (pre-post) outcome change, on average all outcomes in both treatments had improved at the end of treatment compared to baseline. Standard deviations (*SD*) indicated that the variance around the mean in the sample was broadly similar for each outcome across treatment conditions. T-tests between groups showed that the average self-reported outcomes were not significantly different between treatment conditions, after adjustment for multiple testing was applied.

Table 4

Descriptive Statistics of Outcome Measure Scores for Baseline, End of Treatment and Pre-Post Change by Treatment Condition and Overall Sample.

	Baseline	End of Treatment	Pre-post change
Anxiety (GAD-7)			
Overall Sample	16.31 (3.9)	8.66 (5.54)	-7.65 (5.98)
CBT-GSH	16.55 (3.27)	7.45 (4.11)	-9.10 (5.6)
CAT-GSH	16.23 (4.11)	9.07 (5.92)	-7.17 (6.07)
Test of Difference	$t(78)=0.31$, $p = 0.755$	$t(47.09)=-1.35$, $p = 0.183$	$t(78)=-1.26$, $p = 0.212$
Depression (PHQ-9)			
Overall Sample	15.43 (4.79)	9.94 (6.12)	-5.49 (6.08)
CBT-GSH	15.95 (3.82)	8.10 (5.15)	-7.85 (5.26)
CAT-GSH	15.25 (5.09)	10.55 (6.34)	-4.70 (6.17)
Test of Difference	$t(78)=0.56$, $p = 0.575$	$t(78)=-1.56$, $p = 0.122$	$t(78)=-2.05$, $p = 0.044$*
Functioning (WSAS)			
Overall Sample	20.36 (7.31)	14.59 (9.25)	-5.78 (8.97)
CBT-GSH	22.75 (7.0)	11.40 (7.99)	-8.00 (7.85)
CAT-GSH	19.57 (7.29)	15.65 (9.45)	-5.03 (9.25)
Test of Difference	$t(78)=1.71$, $p = 0.092$	$t(78)=1.81$, $p = 0.075$	$t(78)=-1.29$, $p = 0.202$

Notes. *significant at $p < .05$ threshold, **significant at $p < .01$ threshold, ***significant at $p < .001$ threshold, significant at Bonferroni adjusted threshold for multiple testing in bold (though none are significant at the adjusted threshold).

In order to explore individual outcomes, reliable and clinical change status was calculated for each patient in the sample. As displayed in Table 5, there were no significant differences in outcome rates between treatment conditions.

Table 5

Summaries of Outcome Change at End of Treatment for the Overall Sample and Each Treatment Condition.

	Overall (*n*=80)	CBT-GSH (*n*=20)	CAT-GSH (*n*=60)	Test of Difference
Anxiety (GAD-7)				
Reliable improvement	56 (70%)	17 (85%)	39 (65%)	χ^2 (1) = 2.86, p = 0.091
Reliable deterioration	1 (1.3%)	0	1 (1.7%)	χ^2 (1) = 0.34, p = 0.561
RCSI	43 (53.8%)	14 (70%)	29 (48.3%)	χ^2 (1) = 2.83, p = 0.092
No change	21 (26.3%)	3 (15%)	18 (30%)	χ^2 (1) = 1.74, p = 0.187
Depression (PHQ-9)				
Reliable improvement	41 (51.2%)	14 (70%)	27 (45%)	χ^2 (1) = 3.75, p = 0.053
Reliable deterioration	2 (2.5%)	0	2 (3.3%)	χ^2 (1) = 0.68, p = 0.408
RCSI	31 (38.8)	11 (55%)	20 (33.3%)	χ^2 (1) = 2.97, p = 0.085
No change	34 (42.5%)	5 (25%)	29 (48.3%)	χ^2 (1) = 3.34, p = 0.068

Notes. *significant at $p < .05$ threshold, **significant at $p < .01$ threshold, ***significant at $p < .001$ threshold, significant at Bonferroni adjusted threshold for multiple testing in bold (though none are significant at the adjusted threshold).

Relationship between Competence and Outcome (Hypothesis 1)

Pre-Post Intervention Symptom Scores

Table 6 reports the correlation coefficients between overall competence and clinical outcome variables. Neither the end of treatment nor the pre-post change anxiety, depression or functioning outcomes were significantly correlated with overall PWP competence.

Table 6

Pearson Product-Moment Correlation Coefficients Between Overall Competence and Clinical Outcome Variables (Post-Treatment Score and Pre-Post Change)

		Overall Sample (*n*=80)	CBT-GSH Condition (*n*=20)	CAT-GSH Condition (*n*=60)
		r *(p)*	*r* *(p)*	*r* *(p)*
Anxiety				
	Post-tx GAD-7	0.15	0.30	0.15
		(0.095)	(0.103)	(0.123)
	GAD-7 Change	0.01	-0.001	0.04
		(0.476)	(0.498)	(0.375)
Depression				
	Post-tx PHQ-9	0.08	-0.04	0.15
		(0.244)	(0.433)	(0.127)
	PHQ-9 Change	-0.04	-0.13	0.03
		(0.349)	(0.292)	(0.418)
Functioning				
	Post-tx WSAS	0.01	0.01	0.05
		(0.469)	(0.478)	(0.340)
	WSAS Change	-0.08	0.01	-0.07
		(0.241)	(0.483)	(0.294)

Notes. Correlations were interpreted using the following thresholds: 'small' (r = .10), 'medium' (r = .30) and 'large' (r = .50) correlations (Cohen, 1992). Tx = Treatment. Significance level one-tailed. *significant at $p < .05$ threshold, **significant at $p < .01$ threshold, ***significant at $p < .001$ threshold, significant at Bonferroni adjusted threshold for multiple testing in bold (though none are significant at the adjusted threshold).

Recovery Rates

As presented in Table 7, there was no association between PWP competence level and patient recovery rates.

Table 7

Chi-Square Analyses for Rates of RCSI Across PWP Competence Levels per Treatment Condition and for the Overall Sample.

		Overall Sample (*n*=80)			CAT-GSH Condition (*n*=60)			CBT-GSH Condition (*n*=20)		
		Low[1]	Medium[2]	High[3]	Low	Medium	High	Low	Medium	High
Anxiety	Total *n*	18	44	18	15	34	11	3	10	7
	RCSI	12	21	10	9	15	5	3	6	5
	No RSCI	6	23	8	6	19	6	0	4	2
	Chi-square	χ^2 (2) = 1.87, p = 0.392			χ^2 (2) = 1.10, p = 0.578			χ^2 (2) = 1.77, p = 0.413		
Depression	RCSI	8	16	7	6	11	3	2	5	4
	No RCSI	10	28	11	9	23	8	1	5	3
	Chi-square	χ^2 (2) = 0.35, p = 0.839			χ^2 (2) = 0.50, p = 0.780			χ^2 (2) = 0.28, p = 0.870		

Notes. RCSI = Reliable and Clinically Significant Improvement. Significance is Asymptotic Significance (two-sided). [1] = low competence was defined as an overall competence score < 17.5, [2] = medium competence was defined as an overall score between 17.5 and 21.5, [3] = high competence was defined as an overall score >21.5. *significant at p < .05 threshold, **significant at p < .01 threshold, ***significant at p < .001 threshold, significant at Bonferroni adjusted threshold for multiple testing in bold (though none are significant at the adjusted threshold).

Competence Predicting Outcome

In terms of the multiple regression model, the overall model explained 14.5% of the variance in post-treatment anxiety symptoms. Table 8 reports the regression model at each step. After initially controlling for the covariates, which explained 13.1% of the variance, PWP competence did not significantly improve the model only accounting for an additional 1.3% of the variance in anxiety outcomes. The final step assessing whether there was an interaction effect between competence and treatment condition was not significant indicating the association between overall competence and end of treatment GAD-7 score did not differ across the CBT-

GSH and CAT-GSH treatment conditions. A sensitivity analysis was also performed to help elucidate whether PWP competence influenced clinical outcome, or clinical outcome influenced PWP competence. In this analysis, only sessions that were rated early in treatment (rated at session 1 or 2) were included in the multiple regression model (see Appendix H for results).

Table 8

Hierarchical Regression Model for Association Between Competence and Anxiety Outcome, Controlling for Baseline Severity, Treatment Condition and Timing of Session Rating.

	B	*B* SE	β	*p*
Step 1 (R^2=0.13)				
Constant	2.19	3.67	-	0.553
Baseline GAD-7	0.35*	0.15	0.25	0.024
Treatment condition	1.78	1.36	0.14	0.195
Session number rated	-0.78*	0.35	-0.24	0.029
Step 2 (R^2=0.14)				
Constant	-2.30	5.60	-	0.683
Baseline GAD-7	0.32	0.16	0.22	0.045
Treatment condition	2.04	1.38	0.16	0.145
Session number rated	-0.76*	0.35	-0.23	0.034
Competence	0.23	0.22	0.12	0.292
Step 3 (R^2=0.15)				
Constant	-7.18	19.90	-	0.719
Baseline GAD-7	0.31	0.16	0.22	0.049
Treatment condition	4.77	10.78	0.38	0.660
Session number rated	-0.76*	0.35	-0.23	0.034
Competence	0.48	0.98	0.24	0.629
Competence*Treatment	-0.14	0.53	-0.23	0.799

Notes. Step 2: ΔR^2 = 0.01. Step 3ΔR^2 = 0.001. * p<.05, ** p<.01, ***p<.001. SE = Standard Error.

Relationship between Competence and Secondary Outcomes (Hypothesis 2)

Treatment Drop-Out and Session Attendance

As is shown in Table 9, there was no association between PWP competence level and treatment drop-out. Similarly, there was no association between PWP competence level and session attendance.

Table 9

Chi-Square Analyses for Drop-Out Status and Session Attendance Across PWP Overall Competence Levels per Treatment Condition and for the Overall Sample.

		Overall Sample (*n*=80)			CAT-GSH Condition (*n*=60)			CBT-GSH Condition (*n*=20)		
Drop-out		Low	Medium	High	Low	Medium	High	Low	Medium	High
	Total *n*	18	44	18	15	34	11	3	10	7
	Dropped out	2	8	0	2	7	0	0	1	0
	Non-drop-out	16	36	18	13	27	11	3	9	7
	Chi-square	χ^2 (2) = 3.90, *p* = 0.142			χ^2 (2) = 2.81, *p* = 0.246			χ^2 (2) = 1.05, *p* = 0.591		
Attendance		Low	Medium	High	Low	Medium	High	Low	Medium	High
	Total *n*	18	44	18	15	34	11	3	10	7
	Minimal attendance	0	3	0	0	3	0	0	0	0
	Moderate attendance	3	8	3	2	5	1	1	3	2
	Full attendance	15	33	15	13	26	10	2	7	5
	Chi-square	χ^2 (4) = 2.66, *p* = 0.617			χ^2 (4) = 2.76, *p* = 0.598			χ^2 (2) = 0.02, *p* = 0.989		

Notes. *significant at $p < .05$ threshold, **significant at $p < .01$ threshold, ***significant at $p < .001$ threshold, significant at Bonferroni adjusted threshold for multiple testing in bold (though none are significant at the adjusted threshold).

Competence predicting drop-out

In terms of the logistic regression analysis, the overall model explained 36% of the variation in drop-out. Table 10 reports the regression model at each step. After initially controlling for the co-variates, which explained 29% of the variance, PWP competence did not significantly improve the model only accounting for an additional 6% of the variance in drop-out. The final step assessing whether there was an interaction effect between competence and treatment condition was not significant indicating the association between overall competence and drop-out did not differ across the CBT-GSH and CAT-GSH groups. Similar to the multiple regression model above, a sensitivity analysis was conducted for the logistic regression, where only early-rated sessions were included in the model (see Appendix H for results).

Table 10

Binomial Logistic Regression Model for Association Between Competence and Drop-Out, Controlling for Baseline Severity, Treatment Condition and Timing of Session Rating.

	B	*B* SE	Exp(B)	*p*
Step 1 (*Nagelkerke R^2=0.29*)				
Constant	-4.55	2.48	0.01	0.067
Baseline GAD-7	0.25	0.13	1.28	0.059
Session number rated	-0.74*	0.32	0.48	0.021
Step 2 (*Nagelkerke R^2=0.35*)				
Constant	-3.17	4.61	0.04	0.492
Baseline GAD-7	0.27*	0.13	1.31	0.044
Session number rated	-0.81*	0.34	0.44	0.016
Treatment condition	1.12	1.17	3.05	0.338
Competence	-0.19	0.16	0.83	0.235
Step 3 (*Nagelkerke R^2=0.36*)				
Constant	-13.85	21.01	0.00	0.510
Baseline GAD-7	0.26	0.13	1.30	0.050

Treatment condition	6.85	10.96	945.28	0.532
Session number rated	-0.82*	0.34	0.44	0.017
Competence	0.34	0.99	1.40	0.732
Competence*Treatment	-0.28	0.52	0.78	0.590

Notes. Step 2: $\Delta R^2 = 0.06$. Step $3\Delta R^2 = 0.01$. * $p<.05$, ** $p<.01$, ***$p<.001$. SE = Standard Error.

Relationship between Competence and Stepping up (Hypothesis 3)

As is shown in Table 11, there was no association between PWP competence level and stepping-up rates.

Table 11

Chi-Square Analyses for Stepping up Rate Across PWP Competence Levels per Treatment Condition and for the Overall Sample.

	Overall Sample (*n*=80)			CAT-GSH Condition (*n*=60)			CBT-GSH Condition (*n*=20)		
	Low	Medium	High	Low	Medium	High	Low	Medium	High
Total *n*	18	44	18	15	34	11	3	10	7
Stepped up	5	7	2	5	5	2	0	2	0
Not stepped up	13	37	16	10	29	9	3	8	7
Chi-square	χ^2 (2) = 1.90, p = 0.386			χ^2 (2) = 2.29, p = 0.319			χ^2 (2) = 2.22, p = 0.329		

Notes. *significant at $p < .05$ threshold, **significant at $p < .01$ threshold, ***significant at $p < .001$ threshold, significant at Bonferroni adjusted threshold for multiple testing in bold (though none are significant at the adjusted threshold).

Relationship between Competence and Rate of Change (Hypothesis 4)

An LMM was built to assess the impact of competence on rate of change in anxiety symptoms over the course of treatment, controlling for treatment and session number rated. The best fitting selected model applied a log-linear growth trend with a AR1 heterogenous covariance structure (see Appendix I for results on model building and selection). The statistics of the final model are displayed in Table 12. The main coefficient of interest was the competence * time

interaction term, which was non-significant, indicating that practitioner competence was not associated with rate of change in self-reported anxiety during treatment.

Table 12

Fixed effects in the Final LMM Model.

	Coefficient	95% CI	SE	*p*
Competence	0.99	-0.01 – 1.99	0.50	0.052
Time log	-3.85***	-4.51 - -3.18	0.33	<0.001
Competence*Time log	-0.13	-0.8 – 0.54	0.34	0.699
Session number rated	-0.11	-0.93 – 0.7	0.41	0.778
Treatment	0.29	-1.60 – 2.17	0.95	0.764

Notes. CI = Confidence Intervals, SE = Standard Error. * $p<.05$, ** $p<.01$, *** $p<.001$.

Discussion

The current study reported secondary analyses of data collected during a pragmatic randomised controlled patient preference trial evaluating the effectiveness of CAT-GSH treatment for anxiety disorders compared to treatment as usual (CBT-GSH). The aim of the current study was to investigate the relationship between PWP competence, patient clinical outcome, treatment engagement and step-up rates during GSH treatment for anxiety disorders in step 2 of an IAPT service. An additional aim was to conduct exploratory investigations into whether PWP competence could predict rate of change in anxiety outcomes. Investigations of the relationship between LI competency and outcome have not been conducted previously on data drawn from an RCT nor with qualified PWPs working in a routine service in an area of high socio-economic deprivation.

In terms of clinical outcomes generally, 53.8% and 38.8% reached RCSI on the GAD-7 and PHQ-9 respectively compared to an overall rate of 51.1% reaching RCSI across anxiety and depression and across diagnoses, nationally (NHS Digital, 2020). RCSI was higher on anxiety outcomes compared to depression, which was perhaps unsurprising given that all patients had a primary diagnosis of an anxiety disorder and the interventions were targeting their anxiety symptoms.

In contrast to the hypotheses, PWP competence was not significantly associated with patient clinical outcome nor was it associated with treatment drop-out, session attendance or step-up neither within nor across treatment conditions. Furthermore, PWP competence did not significantly predict rate of change on anxiety outcomes across the full sample.

The finding that PWP competence was not significantly associated with patient clinical outcome was in contrast to previous studies on high intensity therapies suggesting that higher

therapist competence is significantly associated with therapeutic treatment outcome when delivering CT (Strunk et al., 2010; Weck et al., 2011; Weck et al., 2015), CBT (Norrie et al., 2013), CPT (Marques et al., 2019) and DIT (Wurman, 2019). The findings were, however, not the first to suggest that with some interventions/diagnoses, competence may not have a significant impact on clinical outcomes. Weck et al. (2015), for instance, found that the association between competence and outcome was mediated by therapeutic alliance. One possibility in the current case, therefore, is that a sufficient therapeutic alliance was not able to be developed within 6-8, 30-35 minute sessions of step 2 GSH treatment, and thus no competence-outcome association was shown. It is also possible that there was not an adequate range of competency ratings in the current study to find a significant association with outcome and that all PWPs performed at least at a 'good enough' skill level, as they were all qualified, were taking part in a clinical trial and were closely supervised. This limited range of competency ratings is a common limitation in therapist competency studies (e.g., Bisseling et al, 2019; Wurman, 2019). A lack of competency ranges may be particularly pertinent in cases where recordings took place in the context of a clinical trial, which is a context where therapists' performance is highly scrutinised and may not be representative of routine practice (Branson et al., 2015).

After all, the most recent meta-analysis conducted by Webb et al. (2010) included 17 competence-outcome findings and showed that the mean-weighted effect size was not significantly different from zero ($r = 0.07$). It may be, therefore, that there is only a very slight association between competence and outcome if one exists at all. It is important to note in the current case, however, that most analyses conducted did not have sufficient power and therefore non-significant results may have been due to a lack of power to detect a significant association.

The finding that the rate of patient clinical change was not predicted by PWP competence was also in contrast with previous findings using high intensity therapies (Strunk et al., 2010; Wurman, 2019). It is important to highlight, however, that the LMM analyses were under-powered and therefore non-significant findings cannot firmly suggest evidence of a lack of a significant effect.

PWP competence may not influence patient outcome as much as during traditional high-intensity psychotherapy, because the practitioner is following highly structured workbooks (Shafran et al., 2021). This may suggest that this may require less 'skill' than a high intensity psychological therapy. Therefore, utilising measures of adherence rather than competency may be suited more to the LI context and clinical method. Other factors which may influence step 2 patient outcomes are beginning to be investigated using contemporary statistical techniques such as multilevel modelling. Delgadillo et al. (2020), for instance, found that the level of PWP agreeableness (a personality factor) was significantly associated with patient outcome when controlling for other therapist factors (e.g., age, sex, years of experience). This is a relatively new area of study, however, and thus further investigations into the mediators and moderators of patient outcome in low-intensity interventions will be highly valuable.

Strengths, Limitations and Future Research

The study had a number of key strengths. The main trial had a randomised controlled patient preference design which resulted in a number of factors being controlled for across conditions (e.g., involvement in the trial, specific PWP factors, treatment duration). The competence rating procedures were also a strength of the study and were sampling LI practice in a routine service setting. Firstly, every patient that had a session recorded had an associated PWP competence-rating. This is an improvement on some previous study designs that only rate a

small sub-sample of therapists' recordings. Secondly, PWP competence was rated using a valid and reliable tool designed specifically to assess the competence of PWPs in delivering low intensity interventions at step 2 of IAPT services. Thirdly, the good inter-rater reliability of the tool was further demonstrated in the study through the use of a fully-crossed two-group design finding a very high level of agreement across both rating groups. The use of a CAT expert rating group also confirmed the theoretical integrity of the CAT-GSH.

The study also had a number of methodological limitations. For instance, several analyses conducted did not have sufficient statistical power as mentioned above. In terms of statistical analyses, LOCF procedures were employed to offer protection against attrition bias. The LOCF approach is known, however, to have limitations and may result in biased clinical outcomes (Molnar et al., 2008). Further, as a patient preference design was used, a large proportion of the patients included in the current study chose to engage in CAT-GSH over CBT-GSH potentially due to prior engagement in the service and so prior experience of CBT-GSH. This resulted in significantly unequal sample sizes across treatment conditions and may have introduced greater sampling bias (e.g., patients that chose CAT-GSH may have been more likely to have had an 'unsuccessful' CBT-informed treatment in the past). This sampling bias may be remedied in future by using a more conventional RCT design where all participants are randomised to conditions. The CAT expert rating group could have also rated CBT-GSH sessions to better confirm the theoretical differences between the interventions. Finally, a measure of LI adherence needs to be developed and evaluated due to adherence potentially being more pertinent in LI.

Clinical Implications

In terms of clinical implications, CAT-GSH can be delivered by qualified PWPs with comparable levels of competence to standard GSH, with only brief additional training and

ongoing supervision support. This presents the opportunity for genuinely expanding choice at step 2 beyond the LI interventions currently offered based on CBT. The preference rates suggest that a relationally informed GSH has some attraction to IAPT service users. The study has also further demonstrated that the LITC is a reliable tool to measure competence of PWPs delivering either CBT-GSH or CAT-GSH interventions for anxiety disorders. The use of the measure is therefore indicated during the clinical and case management supervision of PWPs and on training courses, offering a more methodologically robust alternative to current measures for their OSCE assessments and session submissions.

Conclusions

In conclusion, the current study suggests that qualified PWPs can successfully be trained to deliver a CAT-GSH treatment protocol with comparable competence to the delivery of the standard CBT-GSH treatment in IAPT step 2. PWP competence does not appear to be significantly associated with patient clinical outcome nor treatment engagement, though a limited range of competence was observed which may have restricted the likelihood of significant associations being found. Finally, there was no evidence that PWP competence significantly predicted rate of clinical change across GSH treatments, however these analyses were under-powered and thus this does not provide firm evidence for a lack of association between competence and rate of change during GSH treatment. Further investigations of the competency-outcome relationship during LI work are clearly indicated.

References

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., Text Revision). Author.

Baguley, C., Farrand, P., Hope, R., Leibowitz, J., Lovell, K., Lucock, M., O'Neill, C., Paxton, R., Pilling, S., Richards, D., Turpin, G., White, J., & Williams, C. (2010). *Good Practice Guidance on the use of Self-Help Materials within IAPT Services*. IAPT.

Baker, C. (2020, 23 January). *Mental health statistics for England: prevalence, services and funding [Briefing Paper, Number 6988].* House of Commons Library. https://commonslibrary.parliament.uk/research-briefings/sn06988/

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: psychometric properties. *Journal of Consulting and Clinical Psychology, 56*(6), 893-897. https://doi.org/10.1037/0022-006X.56.6.893

Bisseling, E. M., Schellekens, M. P., Spinhoven, P., Compen, F. R., Speckens, A. E., & van der Lee, M. L. (2019). Therapeutic alliance—not therapist competence or group cohesion—contributes to reduction of psychological distress in group-based mindfulness-based cognitive therapy for cancer patients. *Clinical Psychology & Psychotherapy, 26*(3), 309-318. https://doi.org/10.1002/cpp.2352

Branson, A., Myles, P., Mahdi, M., & Shafran, R. (2018). The relationship between competence and patient outcome with low-intensity cognitive behavioural interventions. *Behavioural and Cognitive Psychotherapy, 46*(1), 101-114. https://doi.org/10.1017/S1352465817000522

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

Castonguay, L. G. (1993). " Common factors" and" nonspecific variables": Clarification of the two concepts and recommendations for research. *Journal of Psychotherapy Integration*, *3*(3), 267-286. https://doi.org/10.1037/h0101171

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159. https://doi.org/10.1037/0033-2909.112.1.155

Cuijpers, P., Donker, T., van Straten, A., Li, J., & Andersson, G. (2010). Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. *Psychological Medicine*, *40*(12), 1943-1957. https://doi.org/10.1017/S0033291710000772

Delgadillo, J., Branson, A., Kellett, S., Myles-Hooton, P., Hardy, G. E., & Shafran, R. (2020). Therapist personality traits as predictors of psychological treatment outcomes. *Psychotherapy Research*, *30*(7), 857-870. https://doi.org/10.1080/10503307.2020.1731927

Firth, N., Barkham, M., & Kellett, S. (2015). The clinical effectiveness of stepped care systems for depression in working age adults: a systematic review. Journal of Affective Disorders, 170, 119-130. https://doi.org/10.1016/j.jad.2014.08.030

Ginzburg, D. M., Bohn, C., Höfling, V., Weck, F., Clark, D. M., & Stangier, U. (2012). Treatment specific competence predicts outcome in cognitive therapy for social anxiety disorder. *Behaviour Research and Therapy*, *50*(12), 747-752. https://doi.org/10.1016/j.brat.2012.09.001

Guydish, J., Campbell, B. K., Manuel, J. K., Delucchi, K. L., Le, T., Peavy, K. M., & McCarty, D. (2014). Does treatment fidelity predict client outcomes in 12-Step Facilitation for stimulant abuse?. *Drug and Alcohol Dependence, 134*, 330-336. https://doi.org/10.1016/j.drugalcdep.2013.10.020

Haddock, G., Devane, S., Bradshaw, T., McGovern, J., Tarrier, N., Kinderman, P., ... & Harris, N. (2001). An investigation into the psychometric properties of the cognitive therapy scale for psychosis (CTS-Psy). *Behavioural and Cognitive Psychotherapy, 29*(2), 221-233. http://dx.doi.org/10.1017/S1352465801002089

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23-34. https://doi.org/10.20982/tqmp.08.1.p023

Haug, T., Nordgreen, T., Ost, L.-G., Tangen, T., Kvale, G., Hovland, O. J., Heiervang, E. R., & Havik, O. E. (2016). Working alliance and competence as predictors of outcome in cognitive behavioral therapy for social anxiety and panic disorder in adults. *Behaviour Research and Therapy, 77*, 40–51. https://doi.org/10.1016/j.brat.2015.12.004

Ho, F. Y. Y., Yeung, W. F., Ng, T. H. Y., & Chan, C. S. (2016). The efficacy and cost-effectiveness of stepped care prevention and treatment for depressive and/or anxiety disorders: a systematic review and meta-analysis. Scientific Reports, 6(1), 1-10. https://doi.org/10.1038/srep29281

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15*(4), 336-352. https://doi.org/10.1016/S0005-7894(84)80002-7

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to denning meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12-19. https://doi.org/10.1037//0022-006x.59.1.12

Kellett, S., Bee, C., Aadahl, V., Headley, E., & Delgadillo, J. (2021). A pragmatic patient preference trial of cognitive behavioural versus cognitive analytic guided self-help for anxiety disorders. *Behavioural and Cognitive Psychotherapy, 49*(1), 104-111. https://doi.org/10.1017/S1352465820000442

Kellett, S., Simmonds-Buckley, M., Limon, E., Hague, J., Hughes, L., Stride, C., & Millings, A. (2021). Defining the assessment and treatment competencies to deliver low-intensity cognitive behavior therapy: A multi-center validation study. *Behavior Therapy, 52*(1), 15-27. https://doi.org/10.1016/j.beth.2020.01.006

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

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

Layard, R., & Clark, D. M. (2014). *Thrive: The Power of Evidence-Based Psychological Therapies*. Penguin.

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

Liness, S., Beale, S., Lea, S., Byrne, S., Hirsch, C. R., & Clark, D. M. (2019b). Multi-professional IAPT CBT training: Clinical competence and patient outcomes. *Behavioural and Cognitive Psychotherapy, 47*(6), 672–685. https://doi.org/10.1017/S1352465819000201

Local Government Association, 2019. *Indices of Deprivation 2019.* Local Government Inform. https://lginform.local.gov.uk/reports/view/lga-research/lga-research-report-indices-of-deprivation-2019?mod-area=E08000004&mod-group=AllMetropolitanBoroughLaInCountry_England&mod-type=namedComparisonGroup

Marques, L., Valentine, S. E., Kaysen, D., Mackintosh, M. A., Dixon De Silva, L. E., Ahles, E. M., ... & Wiltsey-Stirman, S. (2019). Provider fidelity and modifications to cognitive processing therapy in a diverse community health clinic: Associations with clinical change. *Journal of Consulting and Clinical Psychology, 87*(4), 357-369. https://doi.org/10.1037/ccp0000384

Meadows, J., & Kellett, S. (2017). Development and evaluation of Cognitive Analytic Guided Self-Help (CAT-SH) for use in IAPT services. *Behavioural and Cognitive Psychotherapy, 45*(3), 266-284. https://doi.org/10.1017/S1352465816000485

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *The British Journal of Psychiatry, 180*(5), 461-464. https://doi.org/10.1192/bjp.180.5.461

Molnar, F. J., Hutton, B., & Fergusson, D. (2008). Does analysis using "last observation carried forward" introduce bias in dementia research?. *Cmaj, 179*(8), 751-753. https://doi.org/10.1503/cmaj.080820

National Collaborating Centre for Mental Health (2020). *The Improving Access to Psychological Therapies Manual [Version 4]*. Author. https://www.england.nhs.uk/wp-content/uploads/2020/05/iapt-manual-v4.pdf

National Institute for Health and Clinical Excellence. (2011, May 25). *Common mental health problems: identification and pathways to care.* NICE Guideline (CG123). https://www.nice.org.uk/guidance/cg123

NHS Digital (2020, July 30). *Annual report on the use of IAPT services, England 2019-2020.* https://files.digital.nhs.uk/B8/F973E1/psych-ther-2019-20-ann-rep.pdf

Norrie, J., Davidson, K., Tata, P., & Gumley, A. (2013). Influence of therapist competence and quantity of cognitive behavioural therapy on suicidal behaviour and inpatient hospitalisation in a randomised controlled trial in borderline personality disorder: Further analyses of treatment effects in the BOSCOT study. *Psychology and Psychotherapy: Theory, Research and Practice*, *86*(3), 280-293. https://doi.org/10.1111/papt.12004

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata*. STATA Press.

Richards, D. A., & Borglin, G. (2011). Implementation of psychological therapies for anxiety and depression in routine practice: two year prospective cohort study. *Journal of Affective Disorders*, *133*(1-2), 51-60. https://doi.org/10.1016/j.jad.2011.03.024

Roos, J., & Werbart, A. (2013). Therapist and relationship factors influencing dropout from individual psychotherapy: A literature review. *Psychotherapy Research*, *23*(4), 394-418. https://doi.org/10.1080/10503307.2013.775528

Scogin, F. R., Hanson, A., & Welsh, D. (2003). Self-administered treatment in stepped-care models of depression treatment. *Journal of Clinical Psychology*, *59*(3), 341-349. https://doi.org/10.1002/jclp.10133

Shafran, R., Myles-Hooton, P., Bennett, S., & Öst, L. G. (2021). The concept and definition of 'low intensity' cognitive behaviour therapy. *Behaviour Research and Therapy*, 103803. https://doi.org/10.1016/j.brat.2021.103803

Sharpless, B. A., & Barber, J. P. (2009). A conceptual and empirical review of the meaning, measurement, development, and teaching of intervention competence in clinical psychology. *Clinical Psychology Review*, *29*(1), 47-56. https://doi.org/10.1016/j.cpr.2008.09.008

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Janavs, J., Weiller, E., Keskiner, A., ... & Dunbar, G. C. (1997). The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *European Psychiatry*, *12*(5), 232-241. https://doi.org/10.1016/S0924-9338(97)83297-X

Shek, D. T., & Ma, C. (2011). Longitudinal data analyses using linear mixed models in SPSS: concepts, procedures and illustrations. *The Scientific World Journal*, *11*, 42-76. https://doi.org/10.1100/tsw.2011.2

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420-428. https://doi.org/10.1037/0033-2909.86.2.420

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence.* Oxford University 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*(10), 1092-1097. https://doi.org/10.1001/archinte.166.10.1092

Stansfeld, S., Clark, C., Bebbington, P., King, M., Jenkins, R., & Hinchliffe, S. (2014). Chapter 2: common mental disorders. In S. McManus, P. Bebbington, R. Jenkins, & T. Brugha (Eds.), *Mental health and wellbeing in England: Adult psychiatric morbidity survey 2014* (pp. 2-32). NHS Digital. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/556596/apms-2014-full-rpt.pdf

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: a systematic review and meta-analysis 1980–2013. *International Journal of Epidemiology*, *43*(2), 476-493. https://doi.org/10.1093/ije/dyu038

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*(3), 429-437. https://doi.org/10.1037/a0019631

Villeneuve, K., Potvin, S., Lesage, A., & Nicole, L. (2010). Meta-analysis of rates of drop-out from psychosocial treatment among persons with schizophrenia spectrum disorder. *Schizophrenia Research, 121*(1-3), 266-270. https://doi.org/10.1016/j.schres.2010.04.003

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78*(2), 200-211. https://doi.org/10.1037/a0018912

Weck, F., Richtberg, S., Jakob, M., Neng, J. M., & Höfling, V. (2015). Therapist competence and therapeutic alliance are important in the treatment of health anxiety (hypochondriasis). *Psychiatry Research, 228*(1), 53-58. https://doi.org/10.1016/j.psychres.2015.03.042

Wurman, T.V. (2019). *Therapist competence in dynamic interpersonal therapy and its association with treatment outcome* [Unpublished doctoral dissertation]. University College London.

Appendix A

Participant-Facing documents

Emma Beattie-Edwards
Trainee Clinical Psychologist
University of Sheffield
Department of Psychology
Floor F, Cathedral Court
1 Vicar Lane
Sheffield S1 2LT
UK

Email: ebeattieedwards1@sheffield.ac.uk

Study Title:
Cognitive-behavioural versus cognitive-analytic guided self-help for anxiety; a patient preference clinical trial (IRAS reference number: 240751 version 8)

Name of Researcher: Emma Beattie Edwards
Study participant ID number:

	Please initial box if happy after reading
I confirm that I have read and understand the information sheet explaining the above research project and I have had the chance to ask questions.	☐
I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without there being any negative consequences. In addition, should I not wish to answer any particular question or questions, I am free to decline.	☐
I understand that if I withdraw during the course of the study, any data I have provided until that point will still be used.	☐
I understand that my responses will be kept confidential. So I understand that I will not be identified or identifiable in the report or reports that result from the research.	☐
I agree for any data collected from me to be stored anonymously.	☐

I understand that I need to choose a treatment and if I am happy to receive either treatment, then the research team will randomly select a treatment for me and that treatment is delivered on the telephone. ☐

I give consent for my GP to be contacted by letter and informed that I am participating in a research study whilst completing my routine treatment for my anxiety in the Improving Access to Psychological Therapies service. Yes ☐ No ☐

I give my consent that one of my therapy sessions will be recorded and this recording will be listened to by a member of the research team to check that the treatment I am receiving is being delivered correctly. ☐

I understand that if I have benefited from the treatment I will be asked to participate in an interview and that this interview is voluntary and will be audio-recorded and then transcribed by a third par that has signed a confidentiality agreement. ☐

I understand that I will be followed-up at 8 and 24 weeks if I have completed the guided self-help, dropped out or been allocated to another intervention in the Trust. ☐

I understand that despite the efforts made to protect my anonymity relating to the interviews I might be able to recognize myself in any written reports in a direct quote by myself. ☐

I understand the limits of confidentiality explained to me in the information sheet, should I suggest that I am a risk to myself, to another person, I am at risk from another person or make a criminal disclosure then this will be shared with the relevant people

I agree to take part in this research project. ☐

_____	_____	_____
Name of Participant	Date	Signature
_____	_____	_____
Lead Researcher	Date	Signature

To be signed and dated in presence of the participant

Copies:

Once this has been signed by all parties the participant should receive a copy of the signed and dated participant consent form and the information sheet. A copy of the signed and dated consent form should be placed in the project's main record (e.g. a site file), which must be kept in a secure location.

Emma Beattie-Edwards
Trainee Clinical Psychologist
University of Sheffield
Department of Psychology
Floor F, Cathedral Court
1 Vicar Lane
Sheffield S1 2LT
UK

Email: ebeattieedwards1@sheffield.ac.uk

Study Title:
Cognitive-behavioural versus cognitive-analytic guided self-help for anxiety; a patient preference clinical trial (IRAS reference number: 240751 version 7)

PARTICPANT INFORMATION SHEET

You are being invited to take part in a research project. Before you decide, it is important to understand why the research is being done and what the research will involve for you. Please read the following information carefully, and please feel free ask any questions you may have.

Who is doing this research?

The University of Sheffield is organizing this research and the project has some funding from a charity. This project has been previously reviewed in terms of its scientific merit by the University of Sheffield Clinical Psychology department and has NHS ethics approval. This research is being undertaken in part fulfillment of an educational qualification.

Why have I been invited?

You have been invited to take part in this research project because you are going to complete a guided self-help intervention in the Improving Access to Psychological Therapies (IAPT) service. This treatment has the aim of helping you with the anxiety that you are currently experiencing. The treatment is delivered by Psychological Wellbeing Practitioners (PWPs) whom are well trained, experienced and supervised in delivering this type of treatment.

Do I have to take part?

It is completely up to you! If you decide to take part, you can keep this information sheet and will be asked to sign a consent form. You can withdraw at any time without giving a reason. Getting help in the IAPT service for your anxiety is not dependent on being part of this research project.

What will happen if I take part?

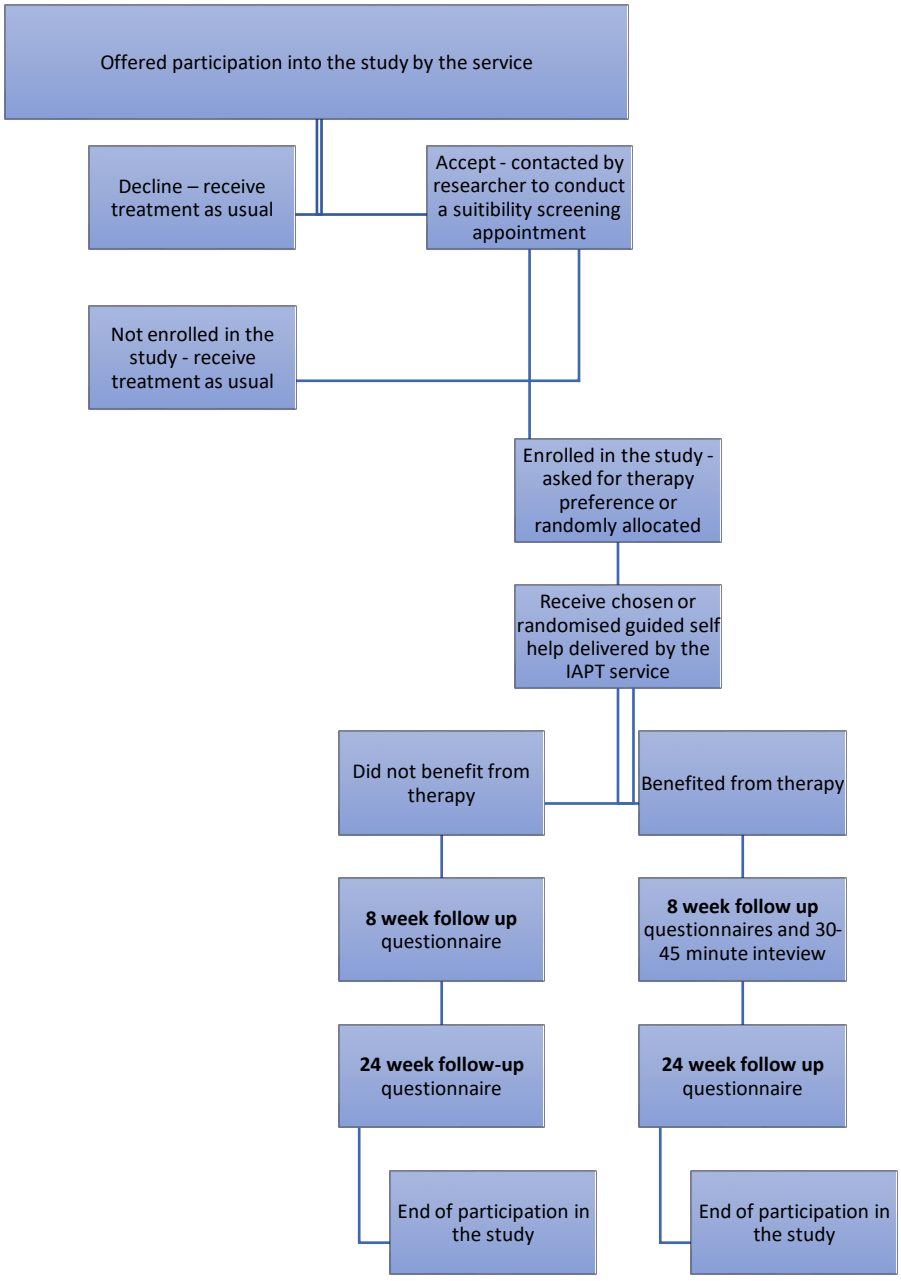
You will be asked to take part in a screening appointment (an interview that asks you about your anxiety and you will be asked to complete a questionnaire that rates the severity of your anxiety; this is called the Beck Anxiety Inventory). The information that you give us in the interview and the total anxiety score from the questionnaire will tell us whether you are suitable for the research study.

If you are suitable for the study, you will be given information that helps you make a choice between the two differing types of guided self-help on offer. If you do not have a strong preference for either one of the treatments, you will be randomly allocated to a treatment by the research team. We really want you to have the opportunity to choose your treatment, but there is no pressure for you to choose. If you are really struggling to decide, the researcher cannot make you decide or make that decision for you. But, we can offer to allocate you to treatment at random, so that decision is taken out of your hands. Please note that due to the COVID-19 health crisis that each of the possible treatments will take place on the telephone. Please do note that guided self-help is often delivered on the telephone, so there is nothing to worry about there.

With your permission, we will inform your GP that you are taking part in a research study. You will also fill in measures of your levels of distress at each of the sessions with the Psychological Wellbeing Practitioner (PWP) who is helping you with your anxiety. This happens with all patients that are seen in the Improving Access to psychological therapies service. The research will also record how many sessions that you attend, whether you dropout and whether you need any more help after this treatment has finished.

One of your guided self-help treatment sessions will be recorded so that we can check that the intervention is being delivered correctly. This session will be selected at random and the recording will have you and the PWP talking on it. It will not have your name attached to it, but rather your study ID and therefore the information is anonymous. The content of the session will be checked for treatment fidelity (i.e. is the PWP doing their job well) by a member of the research team from Sheffield University. Once the session is scored in terms of its fidelity, then it will be deleted. Eight weeks after completing your treatment, or if you have dropped out of treatment or have been allocated to another treatment you will be contacted via the telephone or letter by a researcher. You will be asked to complete the Beck Anxiety Inventory on the telephone. You will also be contacted again at 24-week follow-up by telephone and this is the final follow-up for the study. Again, you will be asked to complete the Beck Anxiety Inventory on the telephone.

If you have really benefitted from the treatment (i.e. you no longer have symptoms that place you in a 'clinical' group) due to effect of the guided self-help, then you would also be asked if you want to take part in a 30-45 minute interview about the treatment you received. This will be conducted either on the telephone or in your home. This will take place at the time of the first follow-up at 8-weeks. In this interview, you will be asked questions about your experience of the guided self-help intervention, including what you found helpful or found unhelpful. These interviews will be audio-taped and then transcribed by a member of the research team at Sheffield University (i.e. they will be typed up to include everything that is said by a professional transcriber bound by a confidentiality agreement). Once transcribed the audio recording will be deleted and the transcribed record stored at Sheffield University. Below is a flow diagram that easily explains the process of participating in the study.



You will be allocated an anonymous study number if you choose to participate in the study and so no information can be attributed to you as a person. Interviews will therefore be attached to a study identification number (i.e. ID number) and not your individual name. Any direct quotes used in summary reports will only have a study ID number attached to them and your anonymity and confidentiality will therefore be protected. The quotes will be kept to a minimum and these direct quotes will be anonymized, but of course there is a small chance that you might recognize yourself in the final write-up if you can remember what you told us during the interview, but no one else really could.

Treatment choices

You are being offered one of two treatments. Both treatments are the same length (6 sessions that last 30-35 minutes each) and both use a guided self-help approach. This means that you will work through a workbook with the support of a PWP. One treatment is called 'cognitive-behavioural guided self-help' and one is called 'cognitive analytic guided self-help'. The key difference is that the cognitive behavioural self-help works in the 'here and now' with your anxiety, whilst the cognitive analytic self-help uses your past and how you grew up, as a way of understanding your anxiety, before moving onto making changes in the present day. You will be given an information sheet that describes the treatment choices and that will help you make the choice that suits you. If you have no strong preference and either treatment appeals to you, then please say and you will be allocated to a treatment by the research team. This allocation is done 'at random.' This means that a computer selects which treatment that you will receive based on a random sequence, to make sure that there is no bias. There is no pressure to choose or to be allocated; either are fine and are both part of the research.

What are the benefits of taking part?

You get the opportunity to share your experience of completing a guided self-help intervention in IAPT. This feedback is helpful for supporting the development and adaptation of guided self-help interventions in the IAPT service. We hope to improve the effectiveness of guided self-help for others through doing this research.

What if there is a problem?

If you feel that there is a problem at any time with participating in the research, you can let the research team know. If you experience any distress whilst sharing your experience, the researcher will be able to discuss this with you, and discuss what further support might be of help.

Will all the information be kept confidential?

Pennine Care NHS Foundation Trust IAPT service will collect information from you and your medical records for this research study in accordance with our instructions.

Pennine Care NHS Foundation Trust will keep your name, NHS number and contact details confidential and will not pass this information to The University of Sheffield. Pennine Care NHS Foundation Trust will use this information as needed, to contact you about the research study, and make sure that relevant information about the study is recorded for your care, and to oversee the quality of the study. Certain individuals from The University of Sheffield and regulatory organizations may look at your medical and research records to check the accuracy of the research study. The University of Sheffield will only receive information without any identifying information. The people who analyze the information will not be able to identify you and certainly will not be able to find out your name, NHS number or contact details.

You will also not be personally identifiable in any reports or publications. As stated, we will only use anonymized short quotes from the interview data. All the number results will be presented as group averages or percentages, and so no single person can be identified.

How long will the data be stored and how will it be handled?

The University of Sheffield is the sponsor for this study based in the United Kingdom. We will be using information from you in order to undertake this study and will act as the data controller for this study. This means that we are responsible for looking after your information and using it properly. The University of Sheffield will keep identifiable information about you for 6 months after the study has finished.

Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible.

You can find out more about how we use your information at https://www.hra.nhs.uk/planning-and-improving-research/policies-standardslegislation/data-protection-and-information-governance/gdpr-guidance or by contacting one of the researchers involved in this study.

The research data will be stored for 5 years. The data will be transferred between the IAPT service and the University. This will be via secure email and the files will also be password protected.

How many times will my data get used?

This information will not identify you and will not be combined with other information in a way that could identify you. The information will only be used for the purpose of health and care research, and cannot be used to contact you or to affect your care. It will not be used to make decisions about future services available to you, such as insurance.

What are the limits of confidentiality?

If during the research screening process or any of the follow-up meetings then you disclose an issue that has implications for your own safety (or the safety of others) or make a disclosure concerning criminal activities, then the research team have a duty of care to pass this information onto the relevant authorities. The researcher conducting the screening or the follow-ups will inform you of this is this is an outcome from that process.

Will I receive any reimbursement of expenses for taking part in this research?

No. There will be no reimbursement of expenses for this research, as we do not anticipate that any will be generated for you. You will not be paid for participating.

What will happen to the results of the study?

The results will be shared at national conferences and also in publications. You can obtain a copy of the results by contacting the researcher on ebeattieedwards1@sheffield.ac.uk or s.kellett@sheffield.ac.uk. Once the study has been published, you will be able to access it on the following University of Sheffield website https://www.sheffield.ac.uk/clinicalpsychology/research/pubs-grants

What if I wish to complain about the way the study has been carried out?

Health and care research should serve the public interest, which means that we have to demonstrate that our research serves the interests of society as a whole. We do this by following the UK Policy Framework for Health and Social Care Research.

If you wish to raise a complaint on how we have handled your personal data, you can contact Andrew Thompson, Director of Research Training at the University of Sheffield who will investigate the matter A.r.thompson@sheffield.ac.uk. If you are not satisfied with our response or believe we are processing your personal data in a way that is not lawful you can complain to the Information Commissioner's Office (ICO).

If you feel that your complaint has not been handled to your satisfaction following this, you can contact the University's Registrar and Secretary Dr Andrew West, Email: registrar@sheffield.ac.uk and Tel *0114 222 1051*

Contact Information

This research is being conducted by **Emma Beattie Edwards** Trainee Clinical Psychologist under the supervision of the Chief Investigator Dr Stephen Kellett. This research will be used to write a thesis which fulfils part of their doctoral training. If you have any questions about the research, you can leave a telephone message with the Research Support Officer on: 0114 222 6650 and he will ask **Emma Beattie Edwards** to contact you.

Appendix B

Outcome Measures [REMOVED]

Appendix C

Ethical Approval Documents

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

NHS
Health Research Authority

Miss Emma Beattie-Edwards
Trainee Clinical Psychologist
Sheffield Health and Social Care NHS Foundation Trust
University of Sheffield, Clinical Psychology Unit
Floor F, Cathedral Court, 1 Vicar Lane
Sheffield
S1 2LT

Email: hra.approval@nhs.net
Research-permissions@wales.nhs.uk

16 October 2018

Dear Miss Beattie-Edwards

HRA and Health and Care Research Wales (HCRW) Approval Letter

Study title:	**Cognitive-behavioural versus cognitive-analytic guided self-help for anxiety; a patient preference clinical trial**
IRAS project ID:	**240751**
Protocol number:	**URMS 156722**
REC reference:	**18/EM/0240**
Sponsor	**University of Sheffield**

I am pleased to confirm that **HRA and Health and Care Research Wales (HCRW) Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

How should I continue to work with participating NHS organisations in England and Wales?
You should now provide a copy of this letter to all participating NHS organisations in England and Wales, as well as any documentation that has been updated as a result of the assessment.

Following the arranging of capacity and capability, participating NHS organisations should **formally confirm** their capacity and capability to undertake the study. How this will be confirmed is detailed in the "*summary of assessment*" section towards the end of this letter.

You should provide, if you have not already done so, detailed instructions to each organisation as to how you will notify them that research activities may commence at site following their confirmation of capacity and capability (e.g. provision by you of a 'green light' email, formal notification following a site initiation visit, activities may commence immediately following confirmation by participating organisation, etc.).

IRAS project ID	240751

It is important that you involve both the research management function (e.g. R&D office) supporting each organisation and the local research team (where there is one) in setting up your study. Contact details of the research management function for each organisation can be accessed here.

How should I work with participating NHS/HSC organisations in Northern Ireland and Scotland?
HRA and HCRW Approval does not apply to NHS/HSC organisations within the devolved administrations of Northern Ireland and Scotland.

If you indicated in your IRAS form that you do have participating organisations in either of these devolved administrations, the final document set and the study wide governance report (including this letter) has been sent to the coordinating centre of each participating nation. You should work with the relevant national coordinating functions to ensure any nation specific checks are complete, and with each site so that they are able to give management permission for the study to begin.

Please see IRAS Help for information on working with NHS/HSC organisations in Northern Ireland and Scotland.

How should I work with participating non-NHS organisations?
HRA and HCRW Approval does not apply to non-NHS organisations. You should work with your non-NHS organisations to obtain local agreement in accordance with their procedures.

What are my notification responsibilities during the study?
The document "*After Ethical Review – guidance for sponsors and investigators*", issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

I am a participating NHS organisation in England or Wales. What should I do once I receive this letter?
You should work with the applicant and sponsor to complete any outstanding arrangements so you are able to confirm capacity and capability in line with the information provided in this letter.

The sponsor contact for this application is as follows:

Name: Andrew Thompson
Tel: 01142226637
Email: a.r.thompson@sheffield.ac.uk

Who should I contact for further information?
Please do not hesitate to contact me for assistance with this application. My contact details are below.

Your IRAS project ID is **240751**. Please quote this on all correspondence.

IRAS project ID	240751

Yours sincerely

Chris Kitchen
Assessor

Email: hra.approval@nhs.net

Copy to: *Dr Andrew Thompson, University of Sheffield (Sponsor Contact)*
Ms Brenda Pimlott , Pennine Care NHS Trust (R&D Contact)

NHS
Health Research Authority

East Midlands - Nottingham 1 Research Ethics Committee
The Old Chapel
Royal Standard Place
Nottingham
NG1 6FS

Please note: This is the favourable opinion of the REC only and does not allow you to start your study at NHS sites in England until you receive HRA Approval

28 September 2018

Miss Emma Beattie-Edwards
Sheffield Health and Social Care NHS Foundation Trust
University of Sheffield, Clinical Psychology Unit
Floor F, Cathedral Court, 1 Vicar Lane
Sheffield
S1 2LT

Dear Miss Beattie-Edwards,

Study title:	**Cognitive-behavioural versus cognitive-analytic guided self-help for anxiety; a patient preference clinical trial**
REC reference:	**18/EM/0240**
Protocol number:	**URMS 156722**
IRAS project ID:	**240751**

Thank you for your letter of 27 September 2018, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair and two other members of the Committee.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this opinion letter. Should you wish to provide a substitute contact point, require further information, or wish to make a request to postpone publication, please contact hra.studyregistration@nhs.net outlining the reasons for your request.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA and HCRW Approval (England and Wales)/ NHS permission for research is available in the Integrated Research Application System, at www.hra.nhs.uk or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database within 6 weeks of recruitment of the first participant (for medical device studies, within the timeline determined by the current registration and publication trees).

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from the HRA. Guidance on where to register is provided on the HRA website.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites

NHS sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Evidence of indemnity]	V.1	05 September 2017
GP/consultant information sheets or letters [GP info letter]	I	06 July 2018
Interview schedules or topic guides for participants [Interview schedule]	2	02 June 2018
IRAS Application Form [IRAS_Form_16072018]		16 July 2018
Letter from funder [Letter confirmation]		
Letter from statistician [Scientific approval letter]	II	02 July 2018
Non-validated questionnaire [Secondary outcome measure]		
Other [Unfavourable opinion letter Rec Review]	I	24 May 2018
Other [Letter addressing REC review recommendations]	I	06 July 2018
Other [CAT-GSH workbook]		
Other [MINI]		
Other [preference support]	II	06 July 2018
Other [peer review feedback]	I	06 July 2018
Other [peer reivew letter]	I	06 July 2018
Other [Trust lone working policy]	I	06 July 2018
Other [PWP study invitation script]	V.1	28 August 2018
Other [REC Provisional opinion letter]	1	22 August 2018
Other [Letter - response to provisional opinion]	1	03 September 2018
Other [Letter to REC 27.09.18]	I	27 September 2018
Participant consent form [consent form]	IV	27 September 2018
Participant information sheet (PIS) [participant info sheet]	IV	27 September 2018
Research protocol or project proposal [Research Protocol]	II	06 July 2018
Summary CV for Chief Investigator (CI) [Emma CV]	2	07 June 2018
Summary CV for Chief Investigator (CI) [CI CV]	1	03 September 2018
Summary CV for student [Student CV]	II	06 July 2018
Summary CV for supervisor (student research)		
Validated questionnaire [Beck Anxiety Inventory examplar]	1	06 January 2018

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document *"After ethical review – guidance for researchers"* gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Training

We are pleased to welcome researchers and R&D staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

18/EM/0240	**Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project.

Yours sincerely,

Mr Murthy Nyasavajjala
Chair

Email:NRESCommittee.EastMidlands-Nottingham1@nhs.net

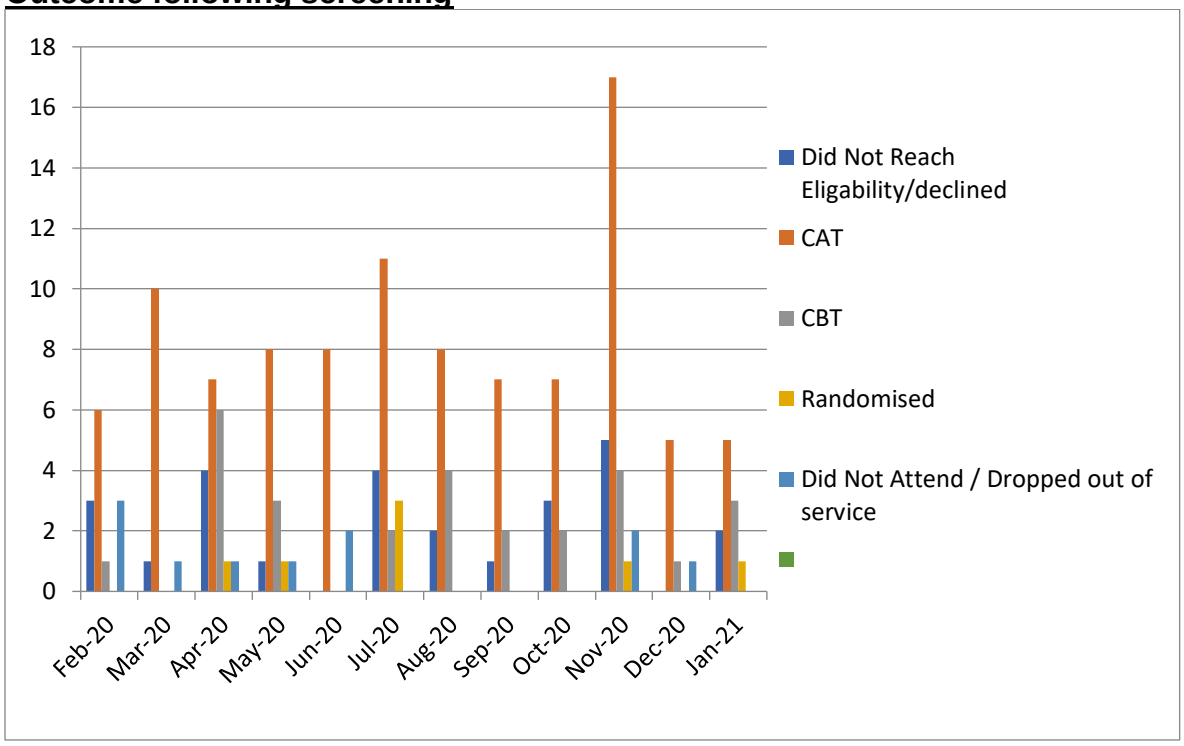
Appendix D

Monthly Trial Newsletter

March 2021

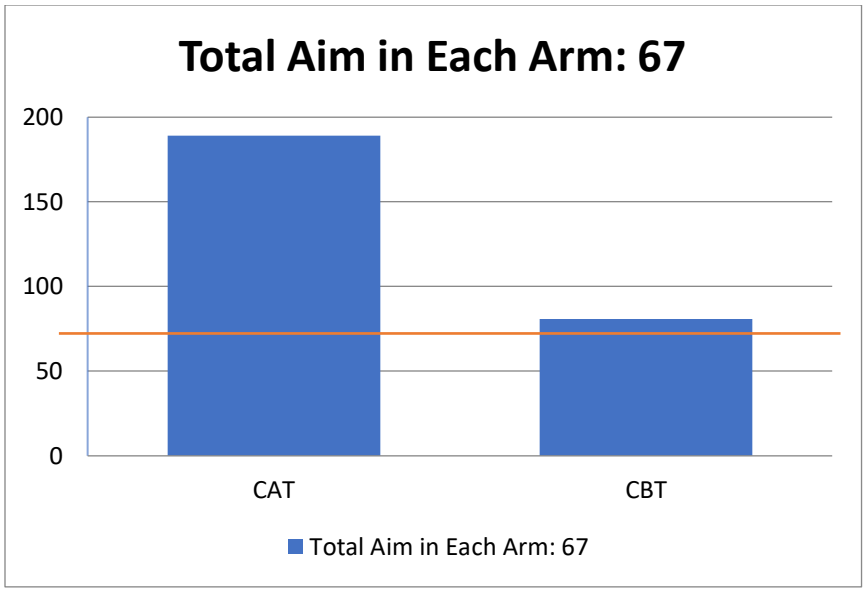
Cognitive-Behavioural Versus Cognitive-Analytic Guided Self-help for Anxiety;
A patient preference clinical trial

Outcome following screening



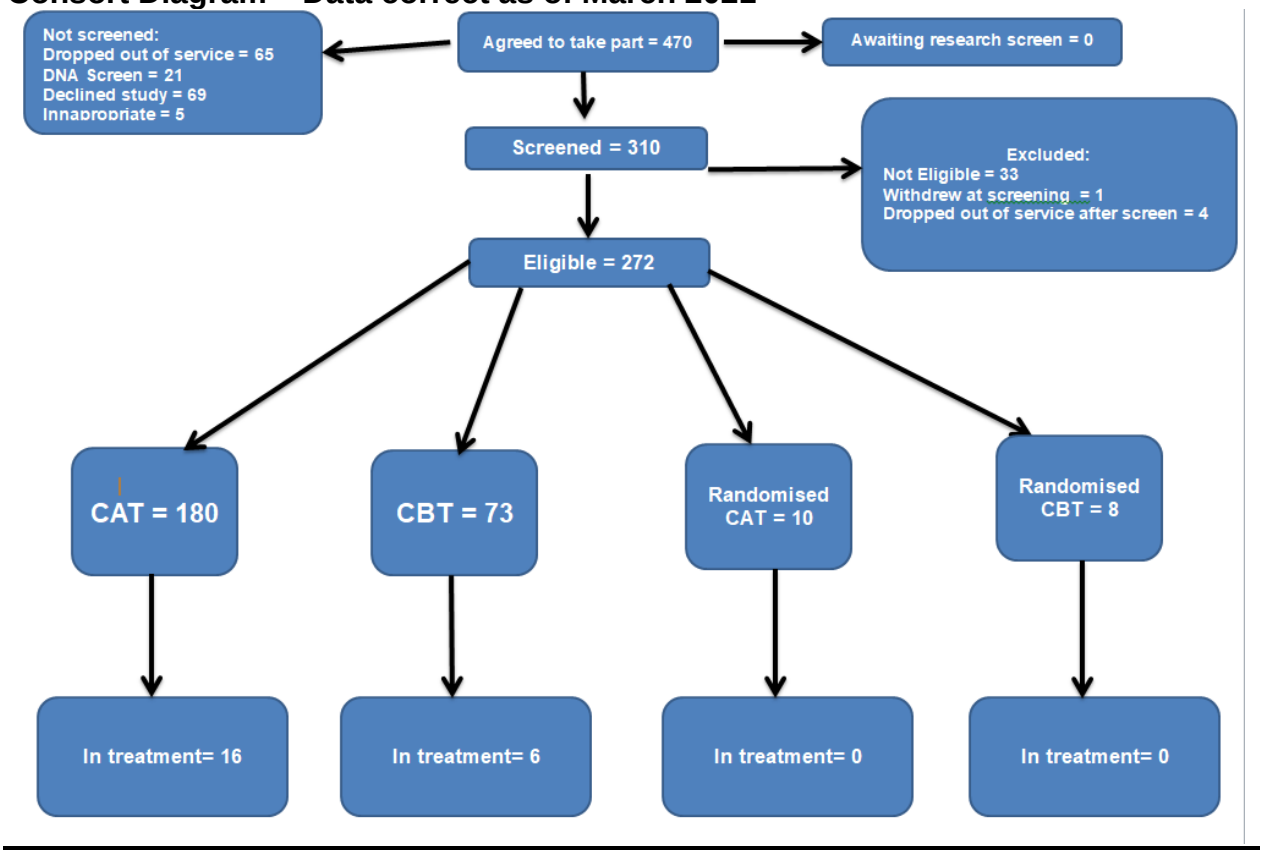
The trial is based on a power analysis of N=67 participants in each arm (CBT-GSH versus CBT-GSH) at least starting their treatment.

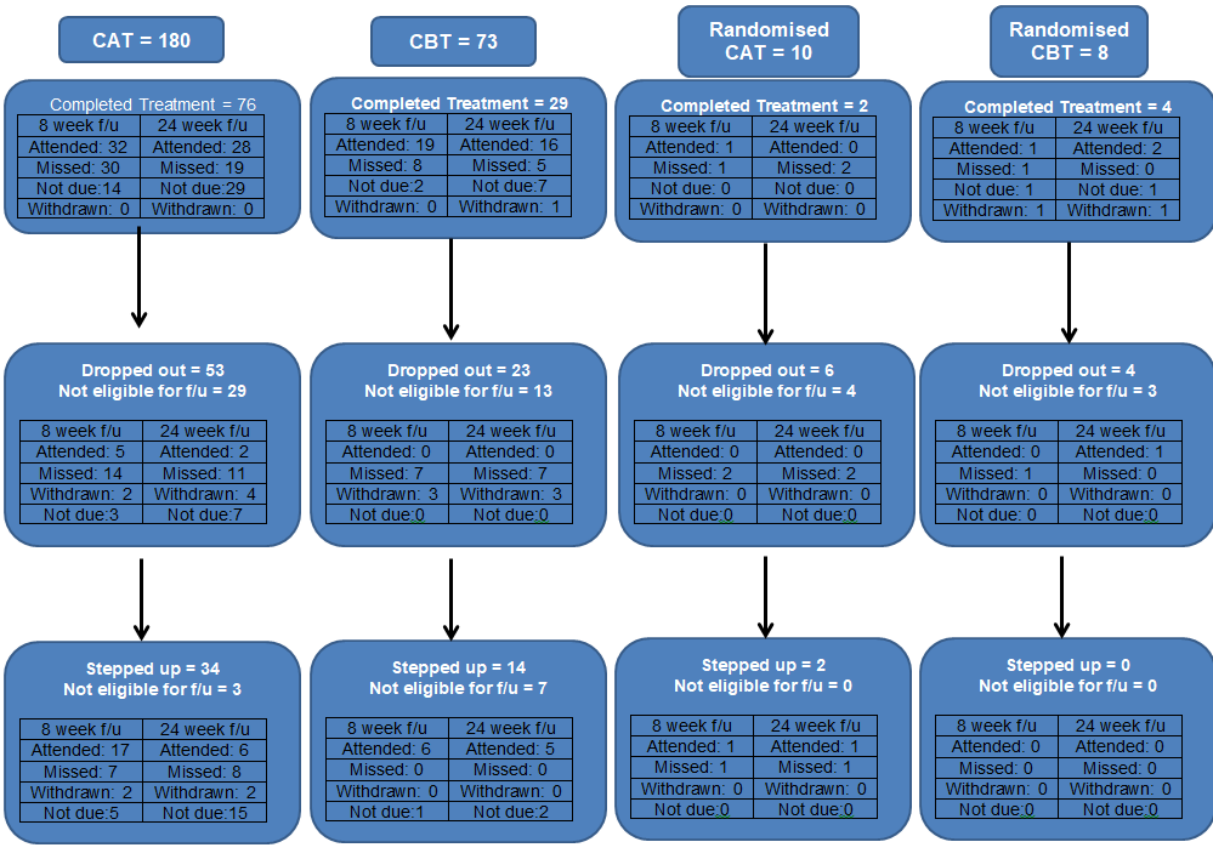
This is what we look like to date:



CAT= 190
CBT= 81

Consort Diagram – Data correct as of March 2021





Now we have completed recruitment to the study we are continuing to focus on contacting patients following discharge from the service to collect their follow-up data.

There are currently a total of 23 active study cases still receiving treatment.

Adherence checking

There are currently N= 80 sessions recording that have been listen to and rated. Three teams (1 x 3 CAT therapists and 2 x 3 senior PWPs) have

participated in a fully crossed reliability study. A report will be available for the service in July 2021 and the remaining N= 50 tapes then rated.

Supporting CAT-GSH in the service after the trial

We have agreed to provide a feedback report on the IAPT MDS outcomes, uptake, dropout and attendance in the four arms in October 2021 in order that they continue to support delivery (i.e. if there is equivalence of outcomes). The management and PWPs are strongly supporting continuing at present. The final PWP CAT-GSH supervision group will take place May 2021 and this is being diligently planned for.

Associated Projects Update

Experience of providing CAT –GSH

12 PWP's have now been interviewed and the interviews are in the process of being transcribed. A report will be available for the service in October 2021.

Analysis of change during GSH

There have been N=20 interviews conducted with participants that have experienced reliable change on the GAD-7 during treatment. This has been analysed and a report will be available for the service in October 2021.

Conversation analysis

There are currently N=80 sessions being analysed in order that a computer be training to identify the key phases used by PWPs to enable change. The initial coding has been completed and then machine learning analysis will be shortly completed.

Appendix E

Blank Competence Rating Tools

LITC (CAT-GSH)

Focusing the CAT-GSH session						
	INCOMPETENT	**NOVICE**	**ADVANCED BEGINNER**	**COMPETENT**	**PROFICIENT**	**EXPERT**
Agrees collaborative agenda with patient						
Subsequent adherence to that agenda						
Overall Section Competency Rating	**0**	**1**	**2**	**3**	**4**	**5** **6**

Continued Engagement in CAT-GSH						
	INCOMPETENT	**NOVICE**	**ADVANCED BEGINNER**	**COMPETENT**	**PROFICIENT**	**EXPERT**
Collaborative approach concerning change or difficulties						
Acknowledges progress or difficulties by use of simple reflections						
Acknowledges progress or difficulties by use of complex reflections						
Use of capsule summaries regarding progress or difficulties						
Use of major section summaries						

Ratio of questions to feedback to facilitate change						
Overall Section Competency Rating	**0**	**1**	**2**	**3**	**4**	**5**	**6**
Interpersonal Competencies in the session						
	INCOMPETENT	**NOVICE**	**ADVANCED BEGINNER**	**COMPETENT**	**PROFICIENT**	**EXPERT**
Empathises through verbal communication						
Non-verbal behaviour ***(please note do not score if the session is audio and do not let this alter the main rating in this section)***						
Encouraging and reinforcing						
Warmth and compassion						

Pacing of the session						
Overall Section Competency Rating	**0**	**1**	**2**	**3**	**4**	**5** **6**

Information Gathering Competencies Specific to Change

COM-B		INCOMPETENT	NOVICE	ADVANCED BEGINNER	COMPETENT	PROFICIENT	EXPERT
CAPABILITY MOTIVATION OPPORTUNITY	**Questioning skills**						
CAPABILITY OPPORTUNITY MOTIVATION	**Problem statement review** ***This only gets completed at session 3 of CAT-GSH and is not then present in subsequent sessions, so do not let this alter the main rating in this section.***						

CAPABILITY OPPORTUNITY MOTIVATION	**Review of goal progress** ***(please note may be absent and do not let this alter the main rating in this section)***						
CAPABILITY OPPORTUNITY MOTIVATION	**Homework review**						
	Review of medication ***(please note can be a very quick check in for a 3 score)***						
CAPABILITY MOTIVATION OPPORTUNITY	**Risk review** ***(please note can be a very quick check in for a 3 score)***						

	Outcome monitoring						
Overall Section Competency Rating	**0**	**1**	**2**	**3**	**4**	**5**	**6**

Within Session Self-Help Change Method Competencies							
COM -B		**INCOMPETENT**	**NOVICE**	**ADVANCED BEGINNER**	**COMPETENT**	**PROFICIENT**	**EXPERT**
CAPABILITY MOTIVATION OPPORTUNITY	**Rationale for treatment** (e.g. introduced or reiterated) ↓						
CAPABILITY OPPORTUNITY MOTIVATION	**Adherence to principles of PWP intervention** Displays fidelity to low intensity treatment (eg. using psychoeducational materials in session) ↓						

CAPABILITY MOTIVATION OPPORTUNITY	**Appropriateness of PWP intervention** Relevant PWP intervention appropriate for patient, stage of intervention and presenting problem ↓						
CAPABILITY MOTIVATION OPPORTUNITY	**Change method** (e.g. use of diaries, ABC or 5-areas conceptualisation to drive the low intensity change methods such as BA, CT ect)						
Overall Section Competency Rating		**0**	**1**	**2**	**3**	**4**	**5** **6**

Planning and Shared Decision Making Competencies							
COM -B		**INCOMPETENT**	**NOVICE**	**ADVANCED BEGINNER**	**COMPETENT**	**PROFICIENT**	**EXPERT**
CAPABILITY OPPORTUNITY MOTIVATION	**Agrees next steps of treatment and the between session work** ↓						

CAPABILITY MOTIVATION OPPORTUNITY	**Defines and agrees the implementation plan for the between session work** ↓						
	Session review and ending						
Overall Section Competency Rating	**0**	**1**	**2**	**3**	**4**	**5**	**6**

LITC (CBT-GSH)

LOW INTENSITY TREATMENT SESSION **COMPETENCY RATING:_____/36**

Worker: ID100 ..... Date: .....

Focusing the Treatment Session Competency						
	INCOMPETENT	**NOVICE**	**ADVANCED BEGINNER**	**COMPETENT**	**PROFICIENT**	**EXPERT**
Agrees collaborative agenda with patient **Standing items;** - review of progress - risk - measures - homework ↓			Text			
Subsequent adherence to agenda			Text			
Overall Section Competency Rating	**0**	**1**	**2**	**3**	**4**	**5** **6**

Interpersonal Competencies						
	INCOMPETENT	NOVICE	ADVANCED BEGINNER	COMPETENT	PROFICIENT	EXPERT
Empathises through verbal communication			Text			
Non-verbal behaviour			Text			
Encouraging and reinforcing			Text			
Warmth and compassion			Text			
Pacing			Text			
Overall Section Competency Rating	0	1	2	3	4	5 6

Information Gathering Competencies Specific to Change							
COM-B		INCOMPETENT	NOVICE	ADVANCED BEGINNER	COMPETENT	PROFICIENT	EXPERT
CAPABILITY MOTIVATION OPPORTUNITY	Questioning skills			Text			
CAPABILITY OPPORTUNITY MOTIVATION	Problem statement review			Text			
CAPABILITY OPPORTUNITY MOTIVATION	Review of goal progress			Text			
	Review of medication			Text			

Establishing and Maintaining Engagement Competencies						
	INCOMPETENT	NOVICE	ADVANCED BEGINNER	COMPETENT	PROFICIENT	EXPERT
Collaborative approach concerning change or difficulties			Text			
Acknowledges progress or difficulties by use of simple and complex reflections			Text			
Use of capsule summaries regarding progress or difficulties			Text			
Use of major section summaries			Text			
Ratio of questions to feedback to facilitate change			Text			
Overall Section Competency Rating	0	1	2	3	4	5 6

CAPABILITY MOTIVATION OPPORTUNITY	**Risk review**			Text			
	Outcome monitoring			Text			
CAPABILITY OPPORTUNITY MOTIVATION	**Homework review**			Text			
Overall Section Competency Rating		**0**	**1**	**2**	**3**	**4**	**5** **6**

Within Session Self-Help Change Method Competencies							
COM-B		INCOMPETENT	NOVICE	ADVANCED BEGINNER	COMPETENT	PROFICIENT	EXPERT
	Rationale for treatment (e.g. introduced or reiterated) ↓			Text			
CAPABILITY OPPORTUNITY MOTIVATION	**Adherence to principles of PWP intervention** Displays fidelity to low intensity treatment (e.g. using psychoeducational materials in session) ↓			Text			
CAPABILITY MOTIVATION OPPORTUNITY	**Appropriateness of PWP intervention** Relevant PWP intervention appropriate for patient, stage of intervention and presenting problem ↓			Text			
	Change method (e.g. use of a stage of a low intensity treatment protocol)			Text			
Overall Section Competency Rating		**0**	**1**	**2**	**3**	**4**	**5** **6**

Planning and Shared Decision Making Competencies							
COM -B		INCOMPETENT	NOVICE	ADVANCED BEGINNER	COMPETENT	PROFICIENT	EXPERT
CAPABILITY OPPORTUNITY MOTIVATION	Agrees next steps of treatment and the between session work ↓			Text			
CAPABILITY MOTIVATION OPPORTUNITY	Defines and agrees the implementation plan for the between session work ↓			Text			
	Session review and ending			Text			
Overall Section Competency Rating		0	1	2	3	4	5 6

Appendix F

Power Calculation Details

G-Power (version 3.1.9.6) was used to conduct a-priori statistical power calculations for the planned correlational and multiple linear regression analyses. In terms of the correlation analyses, Branson et al. (2017) found a moderate positive correlation between PWP competence and reliable symptom improvement for anxiety and depression (r=.405, p=.01). For a one-tailed calculation with correlation H1 set at 0.3, alpha level at 0.05, power set at 0.80 and correlation H0 at 0, a sample size of n=67 was required.

The following determinants were used for multiple regression analyses: F-test, linear multiple regression: fixed model, R^2 deviation from zero and effect size set at a medium level (0.15; Cohen, 1988). Alpha error probability was set at 0.05 and Power set as 0.80 with 4 predictors (two predictor variables: competence rating, treatment condition; two covariates: baseline GAD-7 score, session number rated). The total sample size suggested to achieve sufficient statistical power for the multiple regression analysis was n=85.

Appendix G

Assumption-Testing for Correlations and Regressions

Pearson Correlation Assumptions (for competence-outcome correlations)

A number of assumptions were tested before conducting Pearson Product-Moment correlation analyses.

Linear relationship

Through examination of simple scatter plots, the assumption of linearity was met for the full sample between the outcome variables (LOCF and change scores).

Specifically, there was no non-linear relationships between neither the GAD-7 change nor LOCF GAD-7 and overall competence. Similarly, the assumption was met for PHQ-9 LOCF and PHQ-9 change scores with overall competence. This was the same case for the WSAS LOCF and change scores.

Outliers

No outliers were identified through visual inspection of scatterplots.

Normality Assumption

The assumption of normality was tested by exploring the distribution of the individual variables to be included in the correlation analyses. This was initially tested statistically before the distributions of the variables were explored through graphical interpretation. Statistically, the Shapiro-Wilk test indicated that GAD-7 LOCF, PHQ-9 LOCF and WSAS LOCF variables were not normally distributed ($p < 0.05$). The change variable of GAD-7, PHQ-9 and WSAS and the overall competence variable were all normally distributed as indicated by the Shapiro-Wilk test.

Similarly, examination of histograms suggested that the change variable of the GAD-7, PHQ-9 and WSAS was normally distributed. The overall competence variable was also normally distributed. The LOCF variable of the GAD-7, PHQ-9 and WSAS were not normally distributed according to histograms. Visual inspection of Normal Q-Q Plots indicated that all variables were normally distributed.

Multiple Regression Assumptions (for competence-outcome multiple regression)

Before conducting the planned multiple regression analysis, six assumptions of this procedure were tested.

Independence of observations

There was independence of residuals, as assessed by a Durbin-Watson statistic of 2.02. This assumption was, therefore, passed.

Linear relationship between the outcome variable and continuous predictor variables

A simple scatterplot was generated between the studentised residual and the unstandardised predicted values to establish whether there was a linear relationship between the outcome (GAD-7 final score) and continuous predictor (overall competence, session number rated and baseline GAD-7 score) variables collectively. Visual inspection of the scatterplot suggested a linear relationship existed between the predictor and outcome variables.

Partial regression plots were also generated to establish whether a linear relationship existed between the outcome variable and each of the predictor variables. Visual inspection of the plots suggested that a linear relationship did exist between the outcome and predictor variables individually.

The Assumption of Homoscedasticity

There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values.

No multicollinearity

Through inspection of correlation coefficients, no two predictor variables correlated with one another above a coefficient of 0.7. Calculation of 'Tolerance' and 'VIF' statistics indicated that there were no issues of collinearity (the lowest Tolerance value was 0.93).

Assumption of no significant unusual points

Any potential outliers, high leverage points and highly influential points were examined. Through use of casewise diagnostics, with outliers defined as greater than ±3 standard deviations (as commonly used), no outliers were identified in the data set. There were also no problematic leverage values by use of Huber's (1981) leverage suggestions (<0.2 = safe; 0.2 – 0.5 = risky; >0.5 = dangerous) – suggesting no high leverage points in the data set. Finally, as there were no Cook's Distance values above 1 (Cook and Weisberg, 1982), it was concluded there were no highly influential points and therefore no significant unusual points in the data set.

Assumption of normality

Through visual inspection of a histogram of the outcome variable (GAD-7 final score), it was established that the variable was normally distributed. This finding was confirmed through visual inspection of a Normal P-P Plot which indicated the residuals were close enough to normal for the assumption of normality to be passed.

Binominal Logistic Regression Assumptions (for competence-drop-out regression)

Linear relationship between the continuous predictor variable and the logit transformation of the dependent variable

First, whether the continuous predictor/covariate variables (overall competence, baseline GAD-7 score and session number rated) were linearly related to the logit of the outcome variable (drop-out) was investigated using the Box-Tidwell (1962) procedure. A Bonferroni correction was applied based on all eight terms (including the intercept) in the model when assessing the assumption of linearity (Tabachnick & Fidell, 2014) resulting in an alpha level of $p < 0.006$. Based on the adjusted alpha level, all continuous predictor variables were linearly related to the logit of the dependent variable.

Multicollinearity

Similar to the assumption-testing for the aforementioned multiple regression analysis, through inspection of correlation coefficients, no two predictor variables correlated with one another above a coefficient of 0.7. Calculation of 'Tolerance' and 'VIF' statistics indicated that there were no issues of collinearity (the lowest Tolerance value was 0.93).

Assumption of no significantly unusual data points

Any potential outliers, high leverage points and highly influential points were examined. Through use of casewise diagnostics, with outliers defined as greater than ±3 standard deviations (as commonly used), no outliers were identified in the data set. There were also no problematic leverage values by use of Huber's (1981) leverage suggestions (<0.2 = safe; 0.2 – 0.5 = risky; >0.5 = dangerous) – suggesting no high leverage points in the data set. Finally, as there were no Cook's Distance values above 1 (Cook and Weisberg, 1982), it was concluded there were no highly influential points and therefore no significant unusual points in the data set.

Appendix H

Sensitivity Analyses from Regressions

Sensitivity Analysis for Multiple Regression

In the multiple regression model, the 'session number rated' covariate had a significant main effect on end of treatment GAD-7 score when all other variables in the model were controlled for, B = -0.76 (95% CI = -1.45 – -0.06), SE = 0.35, p = 0.03. To help elucidate whether PWP competence influenced clinical outcome, or clinical outcome influenced PWP competence, a sensitivity analysis was conducted with only sessions that were rated early in treatment (rated at session 1 or 2) being included in the multiple regression model (n=39). This indicated that the main effect of overall competence on end of treatment GAD-7 score was still non-significant with treatment, baseline GAD-7 score and session number rated being controlled for, B = 0.16 (95% CI = -0.63 – 0.95), SE = 0.39, p = 0.68.

Sensitivity Analysis for Logistic Regression

A sensitivity analysis was also conducted for the logistic regression, where the same early-rated sample were included in the logistic regression model (n=39). This indicated that the main effect of overall competence on patient drop-out was still non-significant with baseline GAD-7 score, session number rated, treatment condition and competence by treatment interaction being controlled for, B (1) = -0.27, SE = 0.21, p = 0.20.

Appendix I

LMM Model-Building Procedure

Unconditional Models

Individual Growth Trajectories

After an unconditional model with a simple variance components structure was run, a number of different individual growth trajectories were tested to establish the best fitting time trend. After testing linear, log-linear, quadratic and cubic models (all with time as fixed effects); the log-linear model was selected. Chi-square analyses were used to establish whether the -2 Log-Likelihood were significantly different. Consistent with the rule of parsimony, the simplest model with the best fit statistics was selected. The log-linear model was the simplest (4 parameters) model with the best fitting -2 Log-Likelihood statistic which was less than the linear model statistic. A chi-squared test was unnecessary due to the log-linear model having the same number of parameters as the linear model. The more complex structures (i.e. quadratic and cubic) did not fit the data significantly better than log-linear and so the log-linear structure was taken forward.

After an unconditional fixed growth model with log-linear trend was run, an unconditional model with log-linear (time trend) as random effects was run. This was in order to test whether allowing intercepts and slopes to vary would improve the model. This resulted in substantial improvement in model fit which explained an additional 13% of the residual variance.

Covariance Structure

Next, the best-fitting within-individual error covariance structure was selected before entering any predictors. The best fitting covariance structure was a AR1: Heterogenous structure (modelling for auto-correlation in the longitudinal data).

Conditional Models (Adding Covariates and Predictor)

After unconditional models were run, the conditional models were run with overall competence (continuous) and time (continuous) entered into the LMM analyses as predictors and session number rated and treatment group entered as covariates. Before either covariates or the predictor were entered, they were mean-centred, as is recommended in LMM (Shek & Ma, 2011).

The covariates were entered first in order to see their individual main effects. Adding session number rated (fixed effect) significantly improved the model fit (-2LL was significantly reduced), however no additional residual variance was explained. Adding treatment (fixed effect) to the model did not significantly improve model fit, nor did it explain any additional residual variance.

Adding overall competence as a fixed effects predictor significantly improved the model fit. This model only explained an additional 0.1% of residual variance. A patient-level interaction term (competence * time) was then added to the model in order to test whether the between-patient variation in change in GAD-7 scores over time was explained by the competence of their PWP, with session number rated and treatment condition being controlled for. The model fit was not significantly improved compared to the model with only covariates included. The level of PWP competence, therefore, did not appear to have a significant impact on the rate of change in self-reported anxiety of patients through GSH intervention for the overall sample.

Model	**Deviance (-2LL), df**	**Change in Deviance, df**	**Residual variance (% additional variance explained)**	**Subject level intercept variance**	**Subject level slope variance for time (2,2)**	**Intercept-Slope Covariances (2,1)**
1. Unconditional model	3300, 3	---	18.20	12.42	---	---
2. Unconditional fixed linear growth model (time only as fixed effect) – VC	3082, 4	355*, 1	11.53 (37%)	12.90	---	---
3. Unconditional fixed growth model with high-order trends of time						
Log-linear	**3071, 4**	**11, 0**	**11.29 (1%)**	**12.76**	**---**	**---**
Quadratic	3069, 5	2, 1	11.24 (0.3%)	12.75	---	---
Cubic	3071, 5	2, 0	11.28 (-0.2%)	12.73	---	---
4. Unconditional random growth model (best fitting time trend as random effects) (VC)	**3029, 5**	**42, 0**	**9.00 (13%)**	**10.33**	**4.16**	**---**
Testing different covariance structures						
AR1 Heterogenous	**3025, 6**	**4*, 1**	**8.69 (2%)**	**13.32**	**5.50**	**-0.36**
Unstructured	3025, 6	0, 0	8.69 (0%)	13.32	5.50	-3.12
5. Adding main effects of covariates (session rated and treatment)						
Session number rated as covariate	3005, 7	20*, 1	8.74 (-0.3%)	13.56	5.55	-0.38
Treatment group as covariate	3005, 8	0, 1	8.74 (0%)	13.30	5.55	-0.37
6. Adding main effect of competence	3001, 9	4*, 1	8.73 (0.1%)	12.47	5.57	-0.37
7. Adding competence*time interaction	3000, 10	1, 1	8.74 (-0.1%)	12.47	5.55	-0.37

Table displaying statistics at each stage of the model-building process:

Notes. * = chi-square is significant at $p < 0.05$ level. The model taken forward to the subsequent phases is in bold. VC = variance components (covariance structure).