The University Of Sheffield.

Exploring new developments within, and the evidence base for, the Improving Access to Psychological Therapies programme in the United Kingdom

A thesis submitted in partial fulfilment of the requirements for the Doctorate in Clinical Psychology

The University of Sheffield

May 2020

Sarah Wakefield

Declaration

This thesis has not been submitted to any other institution, or for the purpose of obtaining any other qualification.

Word Counts

Literature review:

Excluding references and tables – 7944

Including references and tables – 14976

Research report:

Excluding references and tables – 7040

Including references and tables – 8821

Total word count (combining overall abstract, literature review and research report):

Excluding references and tables – 14980

Excluding references and tables – 23797

Overall Abstract

The Improving Access to Psychological Therapies (IAPT) initiative was launched in 2008 in the United Kingdom (UK) to offer evidence-based psychotherapies to patients with anxiety and depression in a stepped care service delivery model. The programme reports evaluation data in the form of monthly and annual reports of recovery rates by service and care commissioning group. In recent years the plurality of intervention in this programme has also been expanded

A systematic review and meta-analysis was conducted to review the published practice-based studies arising from the first 10-years of the English IAPT programme, focussing on the effectiveness of IAPT interventions delivered in routine practice. A total of 60 studies were included overall, with n=29 of those making up the meta-analysis. Results found large pre-post treatment effect sizes for depression and anxiety, and a medium pre-post treatment effect size for functional impairments. Implications for future work include exploring how IAPT can become more effective for people with comorbid long-term conditions and medically unexplained symptoms and improving the designs of practice-based studies.

The research report investigated the effectiveness of cognitive analytic therapy (CAT) within the IAPT programme at step 3 (high intensity). Outcome comparisons with the 'treatment as usual' intervention offered at step 3 (i.e. cognitive behavioural therapy; CBT) were completed. Longitudinal multilevel modelling (MLM) was used to investigate symptomatic changes over time and between groups. Results suggest equivalence between CAT and CBT in an IAPT setting. CAT and CBT trajectories of symptom change do not appear to differ over time during the therapies. Implications for CAT being offered as a first-line treatment within the IAPT programme are discussed.

Acknowledgements

Firstly, thank you to my supervisors: Dr Stephen Kellett, for his unwavering positive attitude throughout the many changing faces of the project(s); Dr Jaime Delgadillo, for his ability to make the most complicated topics understandable and the most hopeless situations hopeful; and Dr Daniel Stockton, for his support, encouragement and kindness. Each one has taught me so much through their guidance and expertise.

I would like to thank research collaborators, including Stephen White and Liz Fawkes for their encouragement of the project and for the opportunity to be involved. Thank you to Abi Bradbury for her invaluable input to the work. Many thanks to those in the wider team who have offered advice and support in the collaborative process of shaping the literature review into a published paper.

I have met so many amazing individuals along the way, and I would like to acknowledge the support given by each of my placement supervisors. I feel privileged to have had the opportunity to work alongside you all.

Thank you to my fellow trainees for your encouragement throughout this journey even when all of yours had ended. A particular thanks to Dr Alexandra Leedham who has been relentless in her championing of me – I must repay you one day! I could not have kept going without my friends, who are my family, so thank you for never giving up hope.

Last but not least, a big thanks to my family unit – Neil, Leo, Voltron and Columbo – for your endurance, sympathy, encouragement and company. And for giving me the motivation to keep going!

Contents

Part 1: Literature Review

Ten years of the Improving Access to Psychological Therapies programme in the United Kingdom: A review and meta-analysis of outcomes to date

*This paper has been accepted by the British Journal of Clinical Psychology.

Abstract

Objectives: To review the first 10-years' worth of practice-based studies arising from the English Improving Access to Psychological Therapies (IAPT) programme.

Methods: A systematic review and meta-analysis was conducted. Three databases (SCOPUS, PsycINFO, MedLine) were searched for relevant articles. Inclusion criteria were utilised and included studies focusing on working age adults, quantitative methodology, and the inclusion of validated outcome measures for at least two time points. A narrative overview of the studies is included, describing the characteristics of the studies (e.g., outcomes analysed, step of treatment, demographic variables). A meta-analysis is included from those studies reporting appropriate metric pre-post treatment outcomes: (means and standard deviations, or Cohen's *d* effect sizes). Subgroup analyses examine the potential influence of particular methodologies, treatments, populations or target conditions. Sensitivity checks investigated influential heterogeneity and bias, and articles are quality assessed.

Results: N=60 studies were included in the systematic review, with n=47 reporting metrics that could be used in the meta-analysis. The studies include a range of treatments delivered at both step 2 (low intensity) and step 3 (high intensity). The primary meta-analysis indicated a large pre-post effect size for reductions in depression ($d = 0.87$) and anxiety ($d = 0.88$), and a medium effect regarding reductions to impairment ($d = 0.55$). Differing features of the studies influenced the size of the effect found, such as whether intention-to-treat or completer analyses were used.

Conclusions: IAPT interventions produce large effect sizes in routine practice, with the programme increasing access for large populations. Limitations of the review and future clinical implications are discussed.

Practitioner points:

- Group interventions appear similarly effective as individual sessions and could be implemented more often within services which would increase number of patients seen as well as be an efficient use of staff resources without reducing clinical outcomes
- Disorder specific outcome measures are very limited in use within the IAPT literature and thus this questions whether therapists are adequately capturing change. Practitioners should adopt measures well attuned to the presenting problem
- The use of adherence, competency and treatment integrity measures are lacking within the literature, which raises uncertainty about whether this is reflective of what is happening within clinical practice. The use of such measures within IAPT should be integrated into everyday practice and reported in the research literature
- Future service needs include further expansion of the programme (e.g., continuing the understanding and specialised treatment in those with co-morbid long-term conditions/medically unexplained symptoms)
- IAPT has integrated large-scale and transparent outcome monitoring practices throughout the programme which is consequently allowing for the continued evaluation of practice-based evidence

Keywords: Improving Access to Psychological Therapies, IAPT, meta-analysis, outcome measures, depression, anxiety.

1. Introduction

In the United Kingdom, the National Institute for Health and Care Excellence (NICE) guidelines support the implementation of evidence-based psychological interventions for common mental health problems (e.g. depression, anxiety disorders, obsessive-compulsive disorder and post-traumatic stress disorder; NICE CG90; NICE CG113). These guidelines were implemented at a population level in 2008 in England through a national programme called Improving Access to Psychological Therapies (IAPT). IAPT is based on the funding assumption that patients receiving an evidenced-based psychological therapy would likely recover, return to work and so reduce welfare benefit costs (Clark et al., 2018). Full implementation (via initial pathfinder sites) was based on the development and evaluation of two 'demonstration sites' in Doncaster and Newham (Clark et al., 2009). IAPT addresses the increased demand for psychological therapies in services organised via the principles of 'stepped-care'. It has continued to grow and develop since its inception and is now inclusive of all age groups (e.g., children and older adults) and patient groups (e.g., military veterans and people with long-term health conditions). Patient choice has expanded to include offering therapies other than cognitive behavioural therapy (CBT; e.g. interpersonal psychotherapy, dynamic interpersonal psychotherapy, counselling for depression and couple work focal to depression).

IAPT implementation was based on recruiting and training a new psychological workforce. A significant proportion of this workforce were low intensity practitioners (called psychological wellbeing practitioners; PWPs) working with mild-to-moderate anxiety and depression at the lower step of the model. Traditional (mainly CBT) therapists working with moderate-to-severe anxiety and depression provide therapy at the higher step (The Depression Report, 2006; Clark et al., 2009). National training curricula for both low and high intensity

IAPT therapists mean that services deliver the same protocol-driven psychological interventions (Health Education England). A competencies framework was developed to support the delivery of the CBT treatment protocols (IAPT Programme 'competency framework', 2007). Implementation of clinical and case-management supervision guidelines for IAPT support the stepped-care approach, ensure fidelity to treatment protocols and reduce therapeutic drift (IAPT Supervision Guidance, 2011).

The IAPT stepped care approach offers an effective, brief and less intensive intervention as a first step, before reviewing and either 'stepping up' or 'stepping down' should it be necessary. Low intensity interventions are based on cognitive-behavioural theory, are brief (6-8, 35 minute sessions) and can be delivered over the telephone, in large groups or in a one-to-one format. Low intensity treatment uses a guided and psychoeducational self-help approach with PWPs acting as coaches rather than traditional therapists (Turpin, 2010). Step 3 interventions are based in a formulation-driven approach, are lengthier (typically the protocols define 16 sessions) and typically delivered one-to-one (Roth & Fonagy, 2005). Brief and less intensive interventions are cost effective with a low intensity treatment costing £1,043 per patient, in comparison to high intensity costs of £2,895 per patient (Radhakrishnan et al., 2013).

The collection of sessional outcome measures is a key feature of the IAPT programme. This meant creating a standardised set of generic outcome measures matched to the presenting problems commonly encountered in this setting – known as the 'minimum data set' (MDS). The MDS includes a measure of depression symptomatology (Patient Health Questionnaire (PHQ-9)) (Kroenke, Spitzer & Williams, 2001), a measure of anxiety symptomatology (Generalised anxiety disorder (GAD-7)) (Spitzer, Kroenke, Williams, & Lowe, 2006), and a measure of functional disability (Work and Adjustment Scale (WSAS)) (Mundt, Marks, Shear, & Greist,

2002). ‘Disorder-specific’ measures are also used pre-post when relevant to the individual client and their presenting difficulty (National IAPT guidance, 2010), such as the Obsessive Compulsive Inventory (OCI; Foa, Kozak, Salkovskis, Coles, & Amir, 1998). The MDS is used to calculate the numbers of patients who ‘reliably recover’ following an IAPT intervention (i.e. the case moves below the caseness threshold on all measures at the end of treatment).

PWPs, therapists and services expected to attain a 50% recovery rate (this counts the number of patients who were above the clinical threshold on depression and/or anxiety at pre-treatment, with recovery occurring if a patient subsequently scores below the clinical threshold on depression and anxiety at the end of treatment). This 50% figure is drawn from the results of the clinical trials that make up the evidence base that forms the NICE guidelines. IAPT service outcomes are presented monthly as well as bi-annual full programme reports in aggregated forms and IAPT has chosen to use the ‘recovery’ indicator to assess effectiveness. Approximately 7.5 million referrals have been received by IAPT services since 2012/13 (when the national data outputs are available from) to the latest data output (2017/18), with an average of 1.25 million referrals. From this same data, around 4.9 million patients started treatment with just over 2.6 million patients being recorded as having completed treatment. Treatment duration across IAPT services is typically seven sessions (data from December 2018: Health and Social Care Information Centre, 2019).

In its first 3 years, IAPT treated one million individuals and recovery rates were slightly below the 50% target (NHS England, 2015). For example, in 2015, 44% of treatment finishers in the IAPT service made a ‘reliable recovery’ (Briefing Paper, 2015). In the latest national statistical report this target has been met as recorded by the Health and Social Care Information Centre and is at 51.5% (Health and Social Care Information Centre, March 2019). Clearly, a

huge amount of investment has occurred to enable and then maintain the IAPT programme and it has transformed the landscape of psychological services for people with anxiety and depression in the UK (Firth et al., 2019) and the programme has also served as a model for other national implementation programmes (e.g. Australia, Germany and Canada).

It has been 10 years since the implementation of the IAPT programme, and first demonstration site evaluations (Clark et al., 2009). The implementation of evidence-based practice, through the use of the NICE guidelines and stepped-care, has shaped IAPT as a programme to deliver highly standardised psychological interventions. However, research that makes up the greatest influence on national guidelines such as those of NICE is based on highly rigorous clinical trial studies that utilise randomised controlled trials (RCT) methodology. Whilst this type of research is undeniably scientifically useful and in studies which are high in internal validity, it does this at the expense of external validity. Instead, practice-based evidence studies (i.e., lower internal reliability, but high external reliability) of real-world clinical services, also serve a function, particularly in terms of the generalisability of the clinical trial evidence. IAPT can be defined as a national practice-research network in that it is implementing evidenced-based practice at a national level, but in a practice-based context. The establishment and growth of IAPT over the last 10 years has seen the collection of treatment outcomes from large samples. Thus, IAPT offers a unique insight into the pros and cons of attempts to scale up any psychological service delivery model to a national level. Over the last decade, a range of studies have been published, but to date no systematic review or quality appraisal and meta-analysis of the IAPT evidence base has been completed.

1.1 Aims

The aim of this systematic review and meta-analysis was to review the published practice-based studies arising from the English IAPT programme conducted since its inception. The study sought to exclude clinical trials and to focus on the effectiveness of IAPT interventions delivered in routine practice. This review aimed to supplement nationally available metrics of improvement with conventional effect sizes to enable comparison with the wider psychological therapies for depression and anxiety literature. The review is timely, after ten years of implementation work. This is the first known example of a meta-analysis of studies underpinning a national-level psychological services implementation programme. As this is the first review of its kind, a secondary aim was to also narratively detail the characteristics of outcome studies of the IAPT initiative over the last 10-years.

2. Methods

2.1 Study protocol registration

The study protocol for this meta-analysis was prospectively registered in the PROSPERO website: CRD42018114796.

2.2 Inclusion and exclusion criteria

Articles were included in the review if they met the following inclusion criteria: i) the main focus population was working age adults (i.e., 18 years and over); ii) the study design included the analysis of validated outcome measures and included at least two points of outcome data collection; iii) the studies were published in a peer-reviewed journal; iv) the article was written (or translated) into English; v) the IAPT service was UK-based; vi) quantitative methodology was employed. Table 1 sets out the operationalised inclusion criteria. Exclusion criteria were: i) the focus of the study was on children/adolescent populations; ii) outcome

measures were collected at one time-point only; iii) the study design was experimental in nature (e.g., RCT); iv) the article was not a peer-reviewed publication; v) the article was not in English (or not translated into English); vi) the IAPT service was not UK-based; vii) purely qualitative methodology, opinion pieces or editorials.

Table 1: Operationalised inclusion criteria

Detail	Operationalised inclusion criteria
Participants	Adults who have received treatment from an IAPT service (18 years and over, no upper age limit). Those studies that focus on adult populations, but which include some individuals younger than 18 years old will be included.
Intervention	All psychological treatments received by individuals in the context of the IAPT service (UK), such as cognitive behavioural therapy (CBT), interpersonal therapy (IPT) and counselling. All formats will also be captured, e.g., face-to face 1:1, group-based and computer-based.
Comparator	None.
Outcomes	Primary outcomes: those that measure the range of mental health difficulties that individuals experience treated in IAPT services, such as measures of depression (PHQ-9) and anxiety (GAD-7), collected at a pre- and post- treatment time point. Secondary outcomes: any other measures that are captured during routine clinical practice by IAPT services.
Setting	UK-based IAPT services.
Study design	Pre-post designed practice-based evidence studies.

2.3 Literature search strategy

Three databases were searched for appropriate articles – SCOPUS, PsycInfo, Medline – up until the date of 13-08-2018. The search terms utilised were: “Improving Access to Psychological Therapies” AND/OR IAPT OR “stepped care” NOT “International association for plant taxonomy”. As the IAPT initiative first commenced in 2008, the search years were inclusive of 2007-current date.

The process for capturing all relevant peer-reviewed articles relating to the review question followed several components: i) a systematic search of the three databases using the pre-determined search strings which were operationalised to capture all relevant articles; ii) hand

searching which involved searching the reference lists of those articles that met inclusion criteria; iii) of those articles meeting inclusion criteria from steps 1) and 2) a backward/reverse citation search was completed to capture any other articles relevant to the review.

2.4 Eligibility of relevant articles

Sixty studies met the inclusion criteria, with *n*=29 reporting sufficient statistical information to enable a meta-analysis of their results. For those studies that did not report statistics that were extractable for the meta-analysis (*n*=31), then the corresponding author of the article was contacted by email and requested to supply the relevant study statistics. This resulted in accessing data from *n*=18 additional studies and enabled these studies to be included in the meta- analysis. A narrative synthesis was also carried out including all eligible studies. The PRISMA diagram (Moher, Liberati, Tetzlaff, & Altman, 2009) details the process of study selection (Figure 1). This process followed two stages which was completed by one author in the first instance (SW), and any queries about eligibility were discussed and ratified at subsequent research meetings including three members of the research team (JD, SK, SW). The eligibility process initially reviewed and removed inappropriate articles (i.e., duplicates), followed by the reviewing of the title and abstract, and finally by accessing and reviewing the full-text. A bespoke data extraction tool was used and contained the following items: author/year, service, mental health condition, analysed N, dropout N, analysis (intention-to-treat, or completers analysis), intervention, main findings and outcome measures. Any issues likely to create bias were also noted on the data extraction tool.

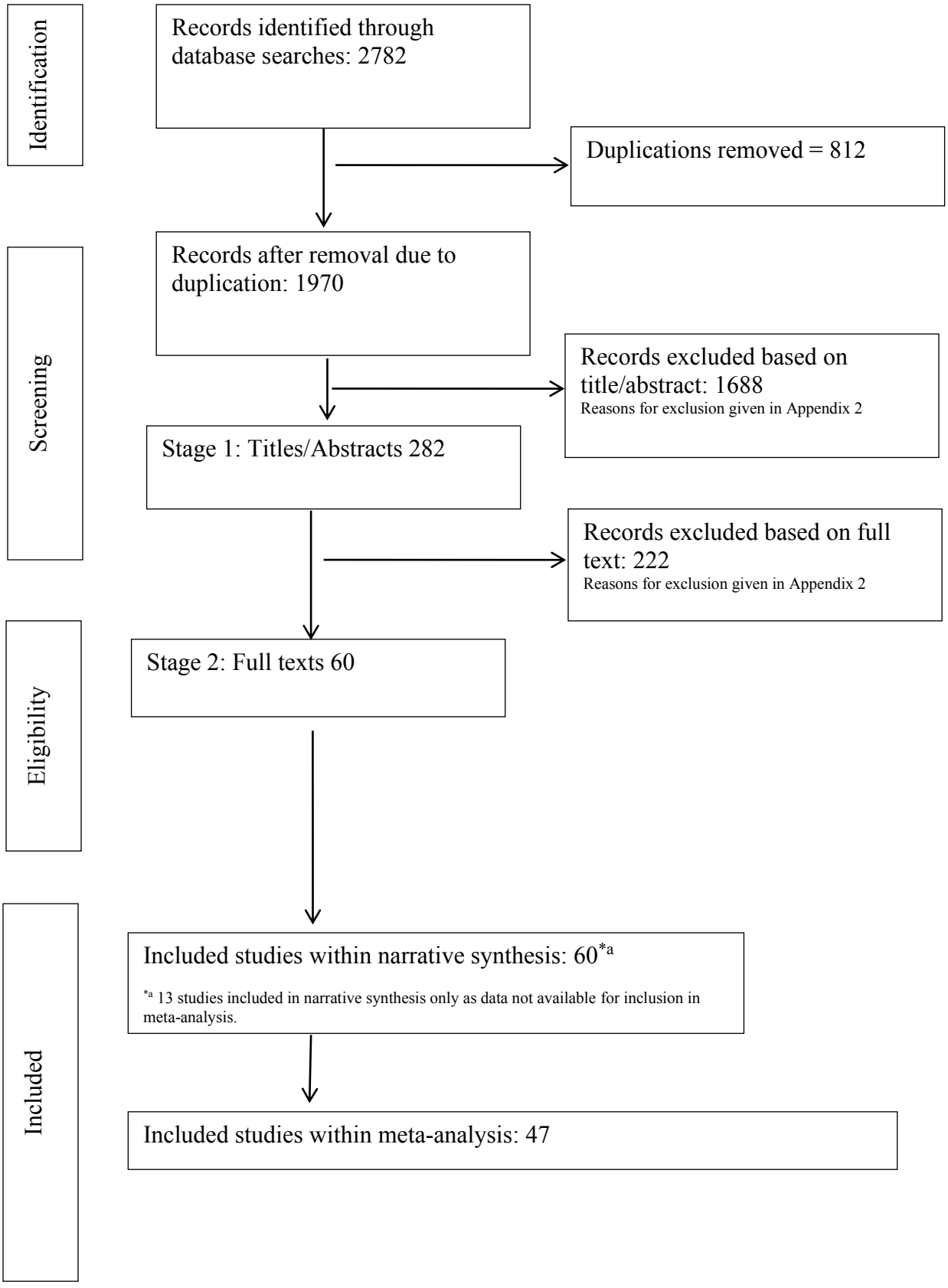

Figure 1: PRISMA diagram of included papers throughout

2.4 Quality assessment and risk of bias

The Critical Appraisal Skills Programme (CASP) tool was used to assess the quality of included studies (Downloaded from: https://casp-uk.net/casp-tools-checklists/). A quality assessment is a crucial element of the research process due to the large variations between the rigor of individual studies. Certain elements of the study, which may be due to methodological flaws, may influence important variables and thus lead to confounding effects. Whether or not these biases are captured and taken into account during the analysis process will be important in understanding whether the treatment effect is biased by these confounding factors and thus to what extent the reported results are truly due to the manipulation of the independent variable (IV) and thus how robust a study's findings are. In addition, not only does the quality assessment capture the potential biases or limitations of a study, it is also used to capture the strengths of a particular study to further understand the robustness of the results.

One researcher completed the quality assessment for all studies (SW) followed by blind rating by two other raters (rater 1 = CBT therapist; rater 2 = clinical psychologist). The process for secondary ratings involved the raters assessing a random selection of papers. Rater 1 rated twelve papers (which represented 20% of the studies), and rater 2 rated six papers that overlapped with rater 1 (which represented 10% of the studies). The process for this consisted of splitting the 60 included papers into study quality groups and taking a random selection to cover all quality levels. The raters were blind to initial quality scores. Once completed, the ratings were compared, and any discrepancies were discussed. An overall agreement consensus for the rating of each paper was completed where possible. Where this was not possible, other members of the research team not involved in quality rating were consulted (JD, SK). Interrater reliability was calculated using Cohen's kappa value (Cohen, 1960) and was calculated from initial ratings.

Based on Cohen's kappa interpretations, the agreement was deemed to be 'moderate' between the original rater and rater 1 (k = 0.526 95% CI 0.430-0.662), and between the original rater and rater 2 (k = 0.546 95% CI 0.369-0.683).

The overall score for each paper can be found in Table 3 (the full quality assessment for each paper can be found in Appendix 3).

2.5 Data synthesis: narrative review and meta-analysis

A narrative synthesis aimed to summarise key study characteristics. A random effects meta-analysis aimed to synthesise the available outcome data (i.e. pre-post treatment, within-group effect sizes derived from available statistics). Analyses were conducted using R packages *metafor* via *MAVIS: Meta-analysis via Shiny* and *forestplot* (R version 3.6.3) (Gordon & Lumley, 2019; Hamilton, Aydin & Mizumoto, 2016; Viechtbauer, 2010). Inclusion criteria for meta-analysis were: (1) reporting pre and post means and SDs convertible into an effect size (ES; Cohen's *d*; Cohen, 1988), (2) reporting Cohen's *d* effect size or, (3) reporting other ESs, but with sufficient additional information (i.e., means/SDs) to enable Cohen's *d* to be calculated or (3) reporting the mean pre-post change and SD. The calculation for Cohen's d was:

$$d = (M_1 - M_1) / SD\ pooled$$

$$\text{where, } SD\ pooled = \sqrt{((SD_1^2 + SD_2^2) / 2)}$$

Cohen's power primer definitions (Cohen, 1988) were used to interpret ESs: 'small' (d = 0.2), 'medium' (d = 0.5) or 'large' (d = 0.8), with anything <0.2 classified as 'negligible'. Forest plots summarize the ES for each study, as well as the pooled (combined) depression, anxiety and functioning ESs across studies. Numbers needed-to-treat (NNT) results are provided for each of the outcome measures to increase the clinical significance of the meta-analysis results. Publication bias was assessed using funnel plots (Egger, Davey, Smith, Schneider & Minder,

1997) and by using the fail-safe N (Orwin, 1983) and rank correlation tests (Begg & Mazumdar, 1994). Heterogeneity was examined using the I^2 statistic and Cochrane's Q test. Moderator analyses examined potential sources of heterogeneity in between-study ES. Subgroup analysis investigated five categorical variables; methodological design (intention-to-treat/completer), step of care (step two/ step three/ step two & three), primary condition (mental health only/comorbid physical health), format (individual/group) and risk of bias (low/medium/high). Meta-regression investigated four continuous variables; gender, age, mean baseline score and treatment duration. The alpha threshold for significance was adjusted to $p < 0.01$ for subgroup and meta-regression analyses to account for multiple testing.

3. Results

3.1 Characteristics of included studies

Table 2 describes individual characteristics of the included studies (n=60). Table 3 outlines the main findings from each study and the associated quality assessment ratings (CASP). Tables 4 and 5 provide a summary of the moderator analyses performed on studies included in the meta-analysis.

Table 2: Overview of papers

First Author & Year	Service(s)	Mental health condition(s)	Analysed *N*	Drop out *N*	Analysis (ITT or completers)	Intervention	Main outcome measure(s)
Adamson et al (2015)†	Lincolnshire IAPT for male offenders (IAPT-O), category B prison	Depression and anxiety-based disorders	627	93	ITT	Step 2 or Step 3	PHQ-9 GAD-7
Ali et al (2014)†	Single North of England IAPT service	Mild-to-moderate MH symptoms or functional impairment	1376	*Not specified*	Completers	Low intensity	PHQ-9 GAD-7
Ali et al (2017)	Single IAPT service	Common MH problems	439	165	Completers[*1]	Previous course of low intensity CBT	PHQ-9 GAD-7 WSAS
Baucom et al (2018)†	London IAPT services	Depression, relationship distress	63 clients[*2] (with 63 partners)	*Not specified*	ITT	High intensity - BCT-D	PHQ-9 GAD-7 CSI-4
Binnie & Boden (2016)	Single outer London Borough IAPT service	*Not reported*	140	61	Completers[*3]	CBT	PHQ-9 GAD-7
Branson et al (2015)	University of Reading and five participating IAPT services	Anxiety and/or depression	1247	*Not specified*	ITT	CBT	*Client:* PHQ-9 GAD-7 *Therapist:* CTS-R
Branson et al (2018)	University of Reading and five participating IAPT services (Thames Valley LETB)	Mild to moderate anxiety and/or depression	3688	*Not reported*	ITT	Low intensity	*Client:* PHQ-9 GAD-7 *Therapist:* ReachOut

							scales
Buckman et al (2018)	Single London IAPT service	Problematic alcohol use; common MH problems	3643	642	ITT	*Not specified*	AUDIT-C PHQ-9 GAD-7 IAPT Phobias Scale WSAS
Burns et al (2016)†	Single North of England IAPT service	Common MH problems	801	261	Completers	Step 2 'Stress Control' group or 'Stress Control+' group	PHQ-9 GAD-7
Chan et al (2014)†	Single Suffolk IAPT service	Mild to moderate depression and/or anxiety	100 (randomly selected from overall *N*)	12 (3 from low intensity; 9 from high intensity)	ITT[*4]	Low and high intensity (50:50)	PHQ-9 GAD-7
Cheston et al (2016)	Single South-West of England IAPT service	Diagnosis of dementia; carers	4	1	ITT	LivDem group	QoL-AD *Carer-related outcomes:* QoL perception
Clark et al (2009)	Two IAPT demonstration sites – Doncaster and Newham IAPT services[*5]	Depression and/or anxiety	*Newham*: 221 (follow-up sample = 60)	*Not reported*	Completers	Low and high intensity	PHQ-9 GAD-7 CORE-OM Employment status *Follow-up*: PHQ-9 GAD-7 Employment status

Clark et al (2018)	NHS Digital and Public Health England data	Depression and/or anxiety	2014/15: 221 CCG 2015/16: 209 CCG	*Not reported*	Completers	*Not specified*	PHQ-9 GAD-7
Clarkson et al (2016)†	Military Veterans IAPT service (North-West)	Mild to moderate MH difficulties	505	170	ITT	Low and high intensity	PHQ-9 GAD-7 WSAS
Delgadillo et al (2014)†	Single North of England IAPT service	Common MH problems	2891	*Not specified*	ITT	Step 2 (low intensity) and Step 3 (high intensity)	PHQ-9 GAD-7
Delgadillo et al (2014)	Single North of England IAPT service	Common MH problems	1850	511 (35.1%)	ITT	Low intensity	PHQ-9 GAD-7
Delgadillo et al (2016)	211 identifiable CCG areas across England	Common MH problems	110415	*Not specified*	ITT	*Not specified*	PHQ-9 GAD-7
Delgadillo et al (2016)†	Five Northern IAPT services	Depression and/or anxiety	4451	1359	ITT	Step 2 (low intensity) 'Stress Control' group	PHQ-9 GAD-7 WSAS
Delgadillo et al (2016)	Single North of England IAPT service	Depression and/or anxiety	1347	*Not specified*	ITT	Step 2 (low intensity) and Step 3 (high intensity)	PHQ-9 GAD-7 WSAS
Delgadillo et al (2017)†	Single North of England IAPT service	Depression and anxiety-related problems with or without LTCs	28498	*Not reported*	ITT	Low and high intensity	PHQ-9 GAD-7 WSAS

Delgadillo et al (2017)	Single Northern England IAPT service	Depression, anxiety or other MH problems	1512	31.3% (Low intensity = 32.2%; High intensity = 28.5)	ITT	Low and high intensity	PHQ-9 GAD-7 WSAS SAPAS
Delgadillo et al (2017)	Single North of England IAPT service	Depression and anxiety problems	594	*Not specified*	ITT	Low and high intensity	PHQ-9 GAD-7
Elison et al (2017)	Single Greater Manchester IAPT service	Range of MH issues	1068	216	ITT	Low intensity (e-Therapy self-help)	PHQ-9 GAD-7 WSAS
Firth et al (2015)†	Single citywide IAPT service	*Not specified*	6111	1553	ITT	Step 2 (low intensity)	PHQ-9 GAD-7 WSAS
Giebel et al (2014)†	North-west veteran-specific IAPT service	Clinical and social problems, including physical disability	366	289 (40.1%)	ITT	*Not reported*	PHQ-9 GAD-7 WSAS
Goddard et al (2015)	Southwark Psychological Therapies Service (IAPT)	Comorbid personality disorder with depression and/or anxiety	1005	35%	ITT	Low and high intensity	PHQ-9 GAD-7 WSAS SAPAS
Green et al (2014)†	Six IAPT services located within the North of England	*Not reported*	1122	0	ITT	Step 2	PHQ-9 GAD-7
Griffiths et al (2014)	Four IAPT services (three Midlands; one Southeast)	Those scoring ‘severe’ on outcome measures (depression, anxiety, functioning)	25034	0	ITT	*Not specified*	PHQ-9 GAD-7 WSAS

Gyani et al (2013)	*N*=24 Year One IAPT services	Depression and/or anxiety	19395	*Not specified*	Completers	Low and high intensity	PHQ-9 GAD-7
Hammond et al (2012)	*N*=7 IAPT services in East of England region	*Not reported*	4106	0	ITT	Low intensity – OTT or FTF	PHQ-9 GAD-7 WSAS
Highfield et al (2016)†	Coventry & Warwickshire IAPT service	Depression and/or anxiety alongside LTCs or MUS	Step 2 = 28 Step 3 = 28	*Not specified*	Completers	Step 2 ("Mind and Body" CBT-based group); Step 3 (individual adapted CBT)	PHQ-9 GAD-7 SEMCD scale (step 2 only)
Jolley et al (2015)†	SLaM – IAPT-SMI demonstration site	Service users with psychosis experience	54	11	Completers	CBT-p (16-30 sessions)	*Clinical outcomes:* CHOICE WEMWBS WSAS PSYRATS *Other outcomes:* Service user experience, satisfactions and feedback questionnaires Friends and Family Test EQ5D
Kellett et al (2016)†	Single Northern England LTC/MUS Pathfinder site	Depression and/or anxiety alongside LTCs or MUS.	1016	130	ITT	Step 2 (low intensity) and Step 3 (high intensity)	PHQ-9 GAD-7
Kellett et al (2017)†	Single IAPT service	Depressive symptoms	26	1	ITT	Step 3 (high intensity) 'BAG'	PHQ-9 GAD-7 WSAS

Kenwright et al (2017)†	North Midlands IAPT service	Anxiety disorders and co-morbid IBS	104	23	ITT	Step 2 and Step 3	PHQ-9 GAD-7 WSAS IBS-specific measures
Kuhn (2011)†	Newham Primary Care Psychological Services	Common MH problems	65	7	ITT	Systemic therapy	PHQ-9 GAD-7 CORE-OM WHO DAS II CSQ-8 Client satisfaction questionnaire Employment questionnaire
Lucock et al (2018)†	Single North of England IAPT service	Remission of symptoms following psychological intervention for depression	11	4	ITT	Low intensity – 'SMArT' intervention	PHQ-9 GAD-7
Luik et al (2017)†	NHS-funded charity in Manchester IAPT service	Insomnia-related depression and/or anxiety	72	26	Completers	Digital CBT (dCBT)	PHQ-9 GAD-7 ISI
Matthew Prina et al (2014)	Six IAPT services in the East of England region	Depression and/or anxiety	16236	4931	ITT	Step 2 and Step 3	PHQ-9 GAD-7
McDevitt-Petrovic et al (2018)	Northern Ireland IAPT service	Common MH difficulties	163	*Not specified*	ITT	Low intensity CBT	PHQ-9 GAD-7
Meadows et al (2017)†	Single IAPT service	Depression and/or anxiety	10	7	Completers	Step 2 – CAT-SH	PHQ-9 GAD-7 WSAS

Methley et al (2016)	BTSS	PTSD	6	0	Completers	Step 4 waiting list – psychoeducation PTSD group	PHQ-9 GAD-7 IES-R SCS-SF ERQ
Mofrad et al (2014)	Single North East of England IAPT service	Depression and simple phobia	1	0	ITT	Behavioural activation	PHQ-9 GAD-7
Morrison et al (2014)†	Single East of England IAPT service	Depression (with little or no comorbid anxiety)	12	5	Completers	Low intensity 'MindBalance' intervention	PHQ-9 WSAS BDI
Murray (2017)	Single East of England IAPT services	PTSD	57 (PHQ-9 & GAD-7), 21 (IES-R)	*Not reported*	Completers	Step 3 – TR-CBT	PHQ-9 GAD-7 IES-R
Pack et al (2014)†	Single IAPT service	Low self-esteem	50	39[*6]	Completers	CBT group	PHQ-9 GAD-7 RSES
Pereira et al (2016)	One IAPT service	Depression and/or anxiety	4980	*Not reported*	ITT	Low and high intensity	PHQ-9 WSAS
Pettit et al (2017)	South West of England IAPT services	*Not specified*	'Attenders' = 54328 'Completers' = 22858	*Not reported*	Completers	*Not reported*	PHQ-9 GAD-7
Poots et al (2014)	Single (Westminster) IAPT service	Depression	1426	3208	Completers	*Not reported*	PHQ-9

Pybis et al (2017)†	(Up to) *N*=121 IAPT services involved in the 2nd NAPT	Depression and/or anxiety, or other common MH problems	33243 (CBT *n* = 23595; Counselling *n* = 9648)	9262	ITT	Step 3 CBT and Step 3 Counselling	PHQ-9 GAD-7
Radhakrishnan et al (2013)†	*N*=5 PCT IAPT services, East of England	*Not specified*	8464	1961	ITT	Low and high intensity	PHQ-9 GAD-7
Richards et al (2011)†	Single North of England IAPT service	Common mental health difficulties	4183	969	ITT	Low and high intensity	PHQ-9 GAD-7
Rimes et al (2018)	*N*=4 London borough IAPT service(s)	Common MH difficulties within different sexual orientation groups	*Variable depending on outcome measure*	*Not reported*	ITT*[7]	Low and high intensity	PHQ-9 GAD-7 WSAS
Saunders et al (2016)	Two London services	Depression and anxiety disorders	16636 (split into two samples): n = 8321; n = 8315	*Not specified*	ITT	Step 1 ('brief interventions') and Step 2 ('formal interventions')	PHQ-9 GAD-7 WSAS Phobia scale – self-rating
Saxon et al (2017)	*Not specified*	Common MH problems	4034	*Not reported*	Completers	Step 3 Counselling or CBT	PHQ-9
Scott (2018)	North West of England IAPT services	Various MH difficulties	29	*Not reported*	Completers	*Not reported*	PHQ-9 GAD-7
Vaillancourt et al (2015)	Single South London IAPT service	Common MH problems	Time 1 = 454 Time 2 = 534	Step 2: Time 1 = 29%; Time 2 = 22% Step 3: Time 1 = 17%; Time 2 = 19%	Completers	Low and High intensity	PHQ-9 GAD-7

Wright et al (2015)†	Single inner London borough IAPT service	Anxiety and/or depression or other common MH difficulties	24	0	ITT	DIT	PHQ-9 GAD-7
Wroe et al (2015)	*Not reported*	Low mood and worry alongside T2DM	*Variable depending on phase of service development*	*Not specified*	Completers	Step 2 'Wellbeing Group'	PHQ-9 GAD-7 DHP SDSCA Physiological measures: HbA_{1c}
Young et al (2017)†	BSL-IAPT and standard IAPT services	Anxiety and/or depression in Deaf BSL clients	Standard IAPT: 116 (pre) and 98 (post) BSL-IAPT: 429 (pre) and 366 (post)	*Not specified*	Completers	Step 2 or Step 3	PHQ-9 GAD-7

† indicates those studies included in the meta-analyses.

MH = mental health; BCT-D = Behavioural Couple Therapy for Depression; CSI-4 = Couples Satisfaction Index (4-item); CTS-R = Cognitive Therapy Scale-Revised; AUDIT-C = Alcohol Use Disorders Identification Test-Consumption; QoL-AD = Quality of Life in Alzheimer's Disease; CORE-OM = Clinical Outcomes in Routine Evaluation-Outcome Measure; SAPAS = Standardised Assessment of Personality – abbreviated scale; SEMCD scale = Self-efficacy for Managing Chronic Disease Scale; CHOICE = Choice of outcome in cognitive therapy for psychoses; WEMWBS = Warwick-Edinburgh Mental Wellbeing Scale; PSYRATS = Psychotic Symptom Rating Scales ; EQ5D = EuroQol Group (Quality of Life questionnaire); ISI = Insomnia Severity Index; IES-R = Impact of Events Scale – Revised; SCS-SF = Self-Compassion Scale Short-Form; ERQ = Emotion Regulation Questionnaire; RSES = Rosenberg's Self-Esteem Scale; DHP = Diabetes Health Profile; SDSCA = Summary of Diabetes Self-Care Activities' questionnaire; HbA_{1c} = glycosylated haemoglobin; BDI = Beck Depression Inventory;

LivDem = Living Well with Dementia; OTT = Over the telephone; FTF = face-to-face; BAG = Behavioural Activation in Groups; DIT = Dynamic Interpersonal Therapy; SMArT = Self-Management After Therapy; TR-CBT = Trauma-focussed CBT.

Thames Valley LETB = Thames Valley Local Education and Training Board; CCG = Clinical Commissioning Groups; PWP = psychological wellbeing practitioners; SLaM = South London and Maudsley NHS Foundation Trust; IAPT-SMI = Improving Access to Psychological Therapies for people with Severe Mental Illness; PCTs = Primary Care Trusts; BTSS = Berkshire Traumatic Stress Service; NAPT = National Audit of Psychological Therapies; BSL = British Sign Language.

LTC = long-term conditions; MUS = medically unexplained symptoms; IBS = Irritable Bowel Syndrome; PTSD = Post-traumatic stress disorder; T2DM = Type 2 Diabetes Mellitus.

*1 Those who completed treatment were recruited and following this stage the data was analysed using ITT (survival analysis) of all participants, even those lost to follow-up.

*2 Clients data only reported within this review

*3 Completers analysis used for the outcomes from CBT intervention. However, this study does compare those who dropped out with the rest of the sample on other variables, such as demographics.

*4 Some missing data and not used, but analysis included dropouts.

*5 Doncaster outcomes are reported in full in another paper (Richards et al., 2011) and therefore, only Newham data from the Clark et al (2009) paper will be used

*6 'Non-completers' used – no information about whether this includes only those who dropped out or others also. Therefore, this figure is an approximate.

*7 Estimate based on information given within the paper, as unsure that enough data is available to determine.

3.2 Narrative synthesis

3.2.1 Demographics

There were a number of sources of heterogeneity regarding the characteristics of the studies. The included *n* in each study ranged from a case study (*n* = 1; Mofrad & Webster, 2014), to data from 209 clinical commissioning groups (CCGs) (*n* = 537,131; Clark et al., 2018) (median = 997; average = 16,025). One study included only male patients (Adamson et al., 2015) and 17 studies that did not report/specify gender demographic data. Of those studies that reported gender, the percentage of females reported ranged from 7.7% (Clarkson et al., 2016; military sample) to 100% (case study; Mofrad & Webster, 2014) (median = 64.3%; average = 60.2%). There were 27 studies which did not report any ethnicity data. Of those that did report ethnicity, it varied in the depth of detail given and with the exception of three studies, the category of 'White'/'White British'/'Caucasian' was the largest ethnic group. The exceptions here included Mofrad & Webster (2014) who reported on a case study of a woman with Middle Eastern ethnicity, Kuhn (2011) who reported 'Asian' ethnicity accounted for 34% of the included sample ('White British' accounted for 33%, and 'White other' accounted for 13%), and Jolley et al (2015) who reported 'BME' as representing 58% of the sample (with 'Non-BME' accounting for the rest of the sample). From the geographical region reported, the North of England represented the largest number of studies (with n=17 studies noting North of England sites; and n=5 noting North-West sites), whilst London IAPT services were the next largest group (n=11 studies). Other areas included IAPT services in the South of England (n=5 studies), the East and Midland regions (n=4 and 2, respectively), Northern Ireland where n=1 study was completed and n=4 studies which utilised data from throughout England.

3.2.2 Outcome measures

Only two studies did not include an analysis of PHQ-9 outcomes. GAD-7 outcomes were reported in 54/60 studies (90%). The WSAS outcomes were reported much less frequently with only 21/60 studies (35%) reporting impairment outcomes. Thirty-two other outcome measures were used across 18/60 (30%) studies (e.g., SDSCA, a measure of diabetes self-care was utilised when the target condition was low mood and worry alongside T2DM (type II diabetes mellitus; Wroe et al., 2015). Only two studies reported on patient satisfaction (Kuhn, 2011; Jolley et al., 2015). Due to a lack of consistent reporting of disorder-specific outcome measures across the studies, only the MDS measures were investigated within the meta-analysis.

3.2.3 Mental health conditions and populations

The majority of studies investigated IAPT services for the conditions it was originally commissioned for (i.e., depression, anxiety, or related/common mental health conditions). In six studies (9.8%), outcomes with physical health conditions were investigated (including long-term conditions, medically unexplained symptoms), whilst in a further one study (1.6%) dementia was the clinical context. Other target conditions included psychosis, relationship distress and problematic alcohol use (one study each; 4.9% overall). One paper (1.6%) investigated the effectiveness of an IAPT service provision within a prison for male offenders (Adamson, Gibbs & McLaughlin, 2016), whilst two papers (3.3%) studied outcomes for military veteran-based services (Clarkson et al., 2016; Giebel, Clarkson & Challis, 2014). One study explored the effectiveness of both a generic IAPT service and one specifically aimed at British Sign Language (BSL) users for a deaf population (Young et al., 2017), whilst another examined any differences in outcomes based on reported sexual orientation of the client (Rimes et al., 2018).

3.2.4 Interventions and stepped care

The specific treatment protocols used to treat patients tended not to be reported in the studies, as studies tended to simply state either generic step or step 3 interventions were delivered. Overall, *n* = 21 studies reported on interventions at either Step 2 or Step 3, whilst any 'Step 2' intervention was reported in *n* = 11 studies, and any 'Step 3' intervention was reported in *n* = 1 study. Where there was a specific intervention discussed and evaluated, this ranged in the intensity and type. CBT was specified as the intervention in *n* = 6 studies (Binnie & Boden, 2016; Branson, Shafran & Miles, 2015; Highfield et al., 2016; McDevitt-Petrovic et al., 2018; Pybis, Saxon, Hill & Barkham, 2017; Saxon, Firth & Barkham, 2017), with another study looking at those clients who had received a previous course of CBT (Ali et al., 2017), and two others using manuals based on CBT – one for psychosis (CBT-p) (Jolley et al., 2015) and one trauma-informed CBT intervention (Murray, 2017). Five studies investigated group interventions (*n* = 6 studies) including two using 'Stress Control' group intervention at step 2 (Burns, Kellett & Donohoe, 2016; Delgadillo et al., 2016b), one high intensity behavioural activation group ('BAG') at step 3 (Kellett, Simmonds-Buckley, Bliss, & Waller, 2017), and another step 2 intervention for dementia patients and their carers (Cheston & Howells, 2016). Other single studies analysed outcomes for systemic therapy (Kuhn, 2011), dynamic interpersonal therapy (DIT) (Wright & Abrahams, 2015), couples' therapy (BCT-D) (Baucom et al., 2018) and a guided self-help CAT intervention (CAT-SH) (Meadows & Kellett, 2017). The full range of interventions can be found in Table 2.

Table 3: Main findings and quality assessment ratings

Author & Year	Main Findings	Quality score* (CASP) (bias range – low, medium, high)
Adamson et al (2015)	Clinical contacts: 82% Low-intensity; average sessions 4.5 (SD 3.06, range 2-8); 18% High-intensity, average sessions 7.5 (SD 6.11, range 2-20). Pre-treatment: 91% clinical range (PHQ) and 94% (GAD); post treatment: 55% recovered (PHQ) and 52% (GAD).	Low
Ali et al (2014)	Clinical contacts: average sessions 5.2 (SD 2.2). Therapist effects minimal effect on outcomes; between-patient variability has larger impact.	Low
Ali et al (2017)	Following low-intensity CBT treatment, 52.8% of patients experienced relapse (deterioration of symptoms to clinically significant levels), with 79% of those occurring within the first 6 months post-treatment. Patients with residual depression symptoms at the end of treatment (PHQ-9 scores 5-9) were more likely to relapse and have shorter remission times than those completing treatment without residual depression symptoms.	Low
Baucom et al (2018)	Clinical contacts: average 10.85 (SD 6.07; range 2-26). Recovery (both PHQ-9/GAD-7 below caseness): full sample of clients = 57.1% recovered; in subsample (those with both client and partner in clinical range pre-treatment), clients = 45.2% recovered and partners = 48.4% recovered. Client effect sizes: PHQ-9 *d* = 2.23; GAD-7 *d* = 1.80; CSI-4 *d* = 0.43.	Low
Binnie & Boden (2016)	Higher levels of depression significantly predicted drop-out. Reasons stated by patients for non-attendance included: feeling unwell (both in relation to physical and mental health symptoms), other priorities, relationship (with the therapist or therapy modality).	Low
Branson et al (2015)	CBT competence, as measured by therapy session video recordings and rated using the CTS-R, was generally not associated with patient outcomes except in the extreme ends of competence.	Medium
Branson et al (2018)	Patients treated by qualified PWPs showed higher rates of improvement and lower rates of deterioration than those treated by trainees. Clinical competence, as measured by the OSCE, was generally not associated with patient outcomes except in the extreme ends of competence.	Medium

Buckman et al (2018)	High alcohol consumption (as indicated by >=6 or >=8 on AUDIT-C) was not associated with treatment outcomes (recovery or RCSC measures). Scores >=8 on the AUDIT-C was associated with greater odds of dropping out of treatment.	Low
Burns et al (2016)	Stress Control group: Clinical contacts = 6 total (2 hours each), 2x PWPs to run. Attendees = average 74 people (range n=23-106). Recovery = 47.1% recovered (when all 6 sessions attended, recovery was 59.2%) – there was a significant association between number of sessions attended and recovery. Those with depression or anxiety more likely to recover than those with comorbidity. Caseness: n=387 pre-treatment on GAD-7) - 58.9% reliably improved and 2.8% reliably deteriorated post-treatment. Caseness: n=302 pre-treatment on PHQ-9 - 45.4% reliably improved and 2% reliably deteriorated post-treatment.	Low
Chan et al (2014)	Clinical contacts: Low-intensity, average sessions 6.61 (SD 3.99); High-intensity, average session 15.98 (SD 8.95). Recovery = 50% High-intensity; 55.3% Low-intensity. PHQ-9 post-treatment effect sizes: High-intensity *d* = 0.86, Low-intensity *d* = 0.66. GAD-7 post-treatment effect sizes: High-intensity *d* = 0.99, Low-intensity *d* = 0.85.	Medium
Cheston et al (2016)	Four patients and their carers took part. Some improvements in Quality of Life outcome measures - 2/3 patients scores increased; 3/3 carers scores increased. Qualitative interviews - confirm quantitative results (in terms of the group being helpful or not).	High
Clark et al (2009)	Recovery post-treatment = 55%.	Low
Clark et al (2018)	Service organisation features predicting clinical outcomes – Positive associations with outcome: i) % of cases with a problem descriptor; ii) number of treatment sessions; iii) % of referrals. Negative associations with outcome: i) time waited to start treatment; ii) % of appointments missed. Negative associations with outcome: i) social deprivation of a CCG.	Medium

Clarkson et al (2016)	Post-treatment effect sizes (full sample, n=505): PHQ-9 *d* = 0.62, GAD-7 *d* = 0.63; WSAS *d* = 0.41 Post-treatment effect sizes (completed treatment, n=156) PHQ-9 *d* = 1.07, GAD-7 *d* = 1.03, WSAS *d* = 0.75 Of those at caseness pre-treatment, 33% and 30% recovered at post-treatment (PHQ-9 & GAD-7, respectively).	Low
Delgadillo et al (2014a)	Post-treatment recovery rates: depression = 42.2%, anxiety = 43.5%. Post-treatment effect sizes: depression *d* = 0.81, anxiety *d* = 0.90.	Low
Delgadillo et al (2014b)	Early improvements can predict end of treatment RCSI (even after confounds taken into account). Early change in treatment was a better predictor of recovery than baseline severity and pre-treatment symptom improvement. It takes at least 4 sessions to achieve >50% RCSI rates for patients accessing low-intensity interventions. Highest attrition rates occur in sessions 1-4.	Low
Delgadillo et al (2016a)	Larger number of referrals were moderately associated with areas of greater deprivation. There was no association between deprivation and case-load size. Significant associations between deprivation and outcomes were found - lower recovery rates were associated with areas of greater deprivation. Using the 50% recovery target - a large proportion of services can be deemed 'underperforming' (72.5%). Whereas using an IMD-adjusted benchmark indicates a lower proportion are 'underperforming' (43.1%).	Low
Delgadillo et al (2016b)	Recovery rates of those meeting caseness pre-treatment: depression = 41.0%, anxiety = 42.2%, overall = 41.6%. Post-treatment effect sizes: depression *d* = 0.59, anxiety *d* = 0.70, functioning *d* = 0.47.	Low
Delgadillo et al (2016c)	PHQ-9 (depression) model classified 63.4% of sample on pre-treatment predictors of poor depression outcomes: - younger age, unemployment, having a self-reported disability, higher baseline WSAS scores. GAD-7 (anxiety) model classified 62.1% of sample on pre-treatment predictors of poor anxiety outcomes: - higher baseline depression scores (PHQ-9), low treatment outcomes expectations. Leeds Risk Index (LRI) developed based on predictors of poorer outcomes.	Low
Delgadillo et al (2017a)	Demographics and outcome measures: age - weak association; gender - no association; unemployment - higher distress at post-treatment correlated with unemployment; ethnicity - higher post-treatment outcomes in non-White patients; deprivation (living area) - higher post-treatment scores associated with more deprived socioeconomic area. Long-term conditions (LTCs): Five LTCs were associated with higher post-treatment scores (controlling for demographics and clinically relevant information). People with LTCs are more likely to be offered High-intensity interventions.	Low

Delgadillo et al (2017b)	Depression prognostic model: Demographic characteristics explained more variance in post-treatment RSCI outcomes (22.5%) relative to clinical characteristics (15%) and personality features (14.7%). Anxiety prognostic model: Clinical characteristics explained more variance in post-treatment RCSI outcomes (55.9%) relative to personality features (23.9%) and demographic characteristics (15.2%). Complex cases were more likely to obtain poorer outcomes after treatment. Those complex cases assigned to High-intensity treatment initially (rather than assigned to Low-intensity and then stepped up to High-intensity) were more likely to show RCSI (PHQ-9 only; GAD-7 approaching significance).	Low
Delgadillo et al (2017c)	Patients in the 'OF' cohort had shorter average duration of treatment and lower average cost of treatment episode. Patients in 'controls' cohort more likely to be described as 'not-on-track' (NOT). No significant differences in outcomes by both cohorts ('controls' and 'OF') - pre-post IAPT MDS scores. Qualitative interviews - OF influenced therapists use of outcome data in several ways (e.g.,): - discussed outcome data more consistently with patients; - prioritised 'NOT' patients in supervision.	Low
Elison et al (2017)	Post-treatment effect sizes (Pearsons r): Breaking Free Online (substance misuse treatment) - *d* = 0.63 (depression); *d* = 0.60 (anxiety); *d* = 0.51 (functioning). Living Life to the Full Interactive (low mood, stress, anxiety treatment) – *d* =0.73 (depression); *d* = 0.73 (anxiety); *d* = 0.47 (functioning). Sleepio (insomnia treatment) – *d* = 0.78 (depression); *d* = 0.69 (anxiety); *d* = 0.54 (functioning). Significant reductions in those at clinical thresholds on the outcome measures at post-treatment in all eTherapy programmes.	Low
Firth et al (2015)	Post-treatment effect sizes: PHQ-9 *d* = 0.82; GAD-7 *d* = 0.90; WSAS *d* = 0.60 RCSI post-treatment: PHQ-9 = 32%; GAD-7 = 36%. Reliable deterioration post-treatment: PHQ-9 = 3%; GAD-7 = 5%	Low
Giebel et al (2014)	Post-treatment effect sizes: PHQ-9 *d* = 0.63; GAD-7 *d* = 0.63; WSAS *d* = 0.40.	Low
Goddard et al (2015)	Patients with higher SAPAS scores (indicating the likely presence of 'personality disorder') were more likely to have poorer outcome measure scores at post-intervention for depression, anxiety and functioning. Recovery rates were also lower in the group with higher SAPAS scores.	Low

Green et al (2014)	Therapist effects: PHQ-9 outcomes = 8.7%; GAD-7 outcomes = 8.8% - i.e., variability in patient outcomes is due to variability between PWPs. Post-treatment effect sizes: PHQ-9 *d* = 0.52; GAD-7 *d* = 0.55 Comparisons of effect sizes from lowest ranked and highest ranked PWPs: Lowest ranked - PHQ *d* = 0.20, GAD *d* = 0.22 Highest ranked - PHQ *d* = 0.92; GAD *d* = 0.95.	Low
Griffiths et al (2014)	Post-treatment effect sizes on those who: i) scored severe on the PHQ-9 at screening: PHQ-9 *d* = 0.52, GAD-7 *d* = 0.40, WSAS *d* = 0.30. ii) scored severe on the GAD-7 at screening: PHQ-9 *d* = 0.45, GAD-7 *d* = 0.55, WSAS *d* = 0.30. iii) scored severe on the WSAS at screening: PHQ-9 *d* = 0.37, GAD-7 *d* = 0.32, WSAS *d* = 0.49. iv) scored severe on all measures at screening: PHQ-9 *d* = 0.46, GAD-7 *d* = 0.42, WSAS *d* = 0.44.	Medium
Gyani et al (2013)	40.3% of sample showed reliable recovery (range between services: 23.9% to 56.5%). 6.6% of sample showed reliable deterioration (range between services: 2.1% to 11.4%). Compliance with NICE guidelines (recommendation of CBT for depression and GAD; counselling for depression) resulted in higher recovery rates (high intensity treatment). Those receiving guided self-help (compared with those receiving non-guided self-help) had higher recovery rates (low intensity treatment). Factors predicting reliable recovery included: patient-level (e.g., initial severity) and service-level (e.g., number of sessions, size of service).	Low
Hammond et al (2012)	In general, both OTT and F2F intervention conditions are effective for those attending low-intensity interventions within IAPT services - on measures of depression, anxiety and functioning. Only those with in an older age bracket and those with more severe symptoms were more likely to benefit from F2F. Costings: F2F mean session cost = £119; OTT mean session cost = £79.	Low
Highfield et al (2016)	Step 2: Significant improvement from pre to post treatment on PHQ-9, GAD-7 and self-efficacy scale. Post-treatment recovery: 35.71%. Step 3: Post-treatment recovery - non-trained workers: depression = 58%; anxiety = 54%. Post-treatment recovery - trained workers: depression = 79%; anxiety = 90%.	High
Jolley et al (2015)	Clinical contacts: completers = 15.6 (SD 7.4); drop-out = 2.8 (SD 1.1). Post-treatment effect sizes: CHOICE: *d* = 0.7 (completers), *d* = 0.2 (drop-out). WSAS *d* = 0.4. Satisfaction ratings - mean score 8.2 (range 6-17, lower scores more favourable).	Medium

Kellett et al (2016)	Post-treatment effect sizes (Partial eta squared): Long-term condition – PHQ-9 *d* = 0.35, GAD-7 *d* = 0.32 Medically-unexplained symptoms – PHQ-9 *d* = 0.25, GAD-7 *d* = 0.31	Low
Kellett et al (2017)	Post-treatment effect sizes: PHQ-9 *d* = 0.73; GAD-7 *d* = 0.49; WSAS *d* = 0.52	Low
Kenwright et al (2017)	Clinical contacts: BCA - 7.2 hours (SD 5.8); Non-BCA - 6.9 hours (SD 5.5). Step: BCA - 79% received step 3/high-intensity treatment (CBT therapist); Non-BCA - 63% received step 3/high-intensity treatment (CBT therapist) (non-significant difference). Pre-treatment to 6-month follow-up effect sizes: BCA – PHQ-9 *d* = 2.0, GAD-7 *d* = 2.2. Non-BCA - PHQ-9 *d* = 1.8, GAD-7 *d* = 2.4. IBS scale - total: *d* = 2.0; BCA showed greater improvement than non-BCA at 6-month follow-up.	Low
Kuhn (2011)	Clinical contacts: average sessions = 8 (SD 4.9). Post-treatment effect sizes: PHQ-9 *d* = 1.28, GAD-7 *d* = 1.15, CORE-OM (distress) *d* = 1.29. Client satisfaction: Therapy clients were 97% mostly/very satisfied with the service; and 92% satisfied with the choice about treatment.	High
Lucock et al (2018)	Remission (<10 on PHQ-9): 86% of patients were in remission by the third contact (telephone call; final point) vs. 60% in remission at the first session.	Medium
Luik et al (2017)	Post-treatment recovery: n=48 (68%). Insomnia symptoms significantly decreased.	Medium

Matthew Prina et al (2014)	Over 65-year olds accounted for 4% of referrals in the 6 PCTs in the East of England during the study period. Older age (over 65-years) was associated with a shorter wait from referral to assessment and from referral to first treatment session. In general, older age was associated with higher recovery rates however this was largely non-significant.	Low
McDevitt-Petrovic et al (2018)	Depression: Reliable improvement = 59.5% of patients; Reliable deterioration = 3%. Anxiety: Reliable improvement = 70.6% of patients; Reliable deterioration = 5.5%. Overall: Reliable recovery: 47.9%; Reliable improvement = 76.7% of patients; Reliable deterioration = 6.1%.	Medium
Meadows et al (2017)	Post-treatment effect size: PHQ-9 *d* = 1.27, GAD-7 *d* = 1.66, WSAS *d* = 1.28.	Low
Methley et al (2016)	Some changes on outcome measures were reported post-intervention in relation to depression, anxiety and PTSD although in general participation in the treatment group on these symptoms was deemed to be inconclusive.	Medium
Mofrad et al (2014)	Subjective reports - positive outcomes reported. IAPT MDS - mixed sessional outcomes over the treatment, with a general decline (although no means/SDs given).	High
Morrison et al (2014)	Recovery: PHQ-9 – 70% (*n* = 10).	Medium
Murray (2017)	Following the training programme for therapists, significant differences in the outcomes on the IES-R were reported but not on the PHQ-9 or GAD-7. Those patients treated after the training programme had significantly lower scores on the IES-R than those patients treated before the training programme.	High
Pack et al (2014)	RSES (self-esteem measure) - clinical and significant change pre-post treatment only. PHQ-9 - clinical and significant change pre-post treatment; clinical and significant change at pre- to 3-month follow-up. GAD-7 - clinical and significant change pre-post treatment; clinical and significant change at pre- to 3-month follow-up.	Medium

Pereira et al (2016)	Therapist effect of 6.7% based on patient depression scores. Higher levels of mindfulness, resilience and the two combined were found in those deemed more effective therapists than those deemed less effective therapists.	Medium
Pettit et al (2017)	Those in younger age bands (20-24 years, 25-29 years) were more likely to be referred to IAPT services for common mental health difficulties. Lowest attendance rates were found in younger age bands (20-24 years, 57.34% attendance) compared to older age bands (60-64 years, 76.97% attendance). MCID or reliable improvement increases throughout the age bands (from 20-24years through to 65-69 years), with the lowest rates being in those 70-74 years.	Medium
Poots et al (2014)	Those from higher areas of deprivation were significantly more likely to have higher baseline depression scores. There was no significant difference in the change scores on the depression measure between levels of deprivation. Inequity of service by deprivation level is not supported.	High
Pybis et al (2017)	Clinical contacts: Overall average sessions = 8.5 (SD 6.18); CBT sessions = 8.9 (SD 6.34); Counselling sessions = 7.5 (SD 5.54). Post-treatment effect size: CBT *d* = 0.94 (95% CI 0.92-0.95). Counselling *d* = 0.95 (95% CI 0.92-0.98).	Low
Radhakrishnan et al (2013)	Clinical contacts: Of those patients who attended 2+ sessions, 44.9% completed treatment, 18.2% dropped out. Costings: Across all sites, average cost per session of Low-intensity = £98.59 (range £78.31 - £150.17); average cost of High-intensity = £176.97 (range £140 - £270.41). Cost per recovered client: Across all sites, average cost of any step £1766 (with assessment). Low-intensity, average cost of recovery £1043 (with assessment). High-intensity, average cost of recovery £2895 (with assessment). Cost per completed treatment: Across all sites, average cost of completed treatment £877.	Low
Richards et al (2011)	Clinical contacts: Average number of sessions = 5.49 (SD 4.31) Average number of face-to-face contacts = 2.33 (SD 2.96) Average number of telephone contacts = 3.17 (SD 3.01). Post-treatment effect size: PHQ-9 *d* = 1.07 (0.88-1.29); GAD-7 *d* = 1.04 (0.88-1.23).	Low

Rimes et al (2018)	Changes in symptomatology outcomes (when controlling for potential confounding factors): Females - Smaller change scores for depression (PHQ-9) and functioning (WSAS) were reported in those who identified as lesbian or bisexual compared to those who identified as heterosexual; smaller change scores for anxiety (GAD-7) was reported in those who identified as bisexual compared to those who identified as heterosexual. Males - no significant differences across sexuality groups. 'Failure to recover' (when controlling for potential confounding factors): Females - 64.9% (lesbian women), 75.2% (bisexual women), 61.3% (heterosexual women) 'failed to meet recovery'. This same pattern, albeit slightly higher, was found in the analyses investigating 'failure to meet reliable recovery'. Significant differences were reported between bisexual women and heterosexual women only. Males - 58.1% (gay men), 64.0% (bisexual men), 59.5% (heterosexual men) 'failed to meet recovery'. This same pattern, albeit slightly higher, was found in the analyses investigating 'failure to meet reliable recovery'. Logistic regression analyses were non-significant.	Low
Saunders et al (2016)	Latent profiles for groups of patients based on similar characteristics were generated. Examples include: - those with low symptomatology baseline scores (anxiety & depression) plus high functioning in everyday life are high likelihood of achieving recovery (74% - brief intervention) in comparison to the full dataset (40%). - those with higher symptomatology baseline scores (anxiety & depression) plus low functioning in everyday life have a lower likelihood of achieving recovery (11% - brief intervention) in comparison to the full dataset.	Low
Saxon et al (2017)	A therapist effect of 5.8% was reported, with those therapists rated 'more effective' showing higher recovery rates in the patients that they treated (compared to 'less effective' therapists). Patient outcomes improved with additional sessions, in general.	Low
Scott (2018)	Post-treatment recovery rates: Depression (n = 26) – recovery = 21.3%, reliable improvement = 11.5%. Anxiety (n = 26) - recovery = 46.0%, reliable improvement = 11.5%. Combined (n = 25) - recovery = 24.0%, reliable improvement = 0%.	High
Vaillancourt et al (2015)	Factors compared at time 1 ("high recovery") and time 2 ("low recovery"): - wait to triage - on average longer wait at time 2 (11 days at step 2; 9 days at step 3). Recovery rates and associated factors: - baseline symptom severity was associated with recovery (i.e., those who recovered were more likely to have lower starting BL severity). - number of treatment sessions, i.e., more sessions were completed by those who recovered at time 1 & 2 (an average of >=2 sessions at step 2/low-intensity with a similar pattern found at step 3/high-intensity).	Medium
Wright et al (2015)	Post-treatment effect size: PHQ-9 d = 0.46. GAD-7 d = 0.52.	Medium

Wroe et al (2015)	Adaptations can be made to Step 2/low-intensity interventions within IAPT to support those with comorbid LTCs (specifically, T2DM).	Medium
Young et al (2017)	BSL cut-offs used for recovery for IAPT-BSL; standard cut-offs used for IAPT generic BSL clients. BSL cut-offs: PHQ = 8 (standard = 10), GAD = 6 (standard = 8). Post-treatment recovery rates: Recovered – generic IAPT = 41.2%, BSL-IAPT = 43.0% Not recovered – generic IAPT = 58.8%, BSL-IAPT = 57.0% Post-treatment reliable improvement rates: Reliably improved - generic IAPT = 63.5%, BSL-IAPT = 66.8 No reliable change - generic IAPT = 32.9%, BSL-IAPT = 27.2% Reliably deteriorated - generic IAPT = 3.5%, BSL-IAPT = 6.0	Medium

*Quality score – more information can be found in the appendix.
Reported effect sizes are Cohen's *d* unless otherwise stated.

Key: CTS-R = Cognitive Therapy Scale Revised; PWPs = Psychological Wellbeing Practitioners; OSCE = Observed Standardised Clinical Examination; CSI-4 = Couples Satisfaction Index (4-item); AUDIT-C = Alcohol Use Disorders Identification Test-Consumption; CORE-OM = Clinical Outcomes in Routine Evaluation-Outcome Measure; SAPAS = Standardised Assessment of Personality – abbreviated scale; CHOICE = Choice of outcome in cognitive therapy for psychoses; IES-R = Impact of Events Scale – Revised; RSES = Rosenberg's Self-Esteem Scale; OTT = Over the telephone; FTF = face-to-face; CCG = Clinical Commissioning Groups; PCTs = Primary Care Trusts; BSL = British Sign Language; PTSD = Post-traumatic stress disorder; T2DM = Type 2 Diabetes Mellitus; GAD = Generalised Anxiety Disorder; BCA = Bowel Control Anxiety; IAPT MDS = Improving Access to Psychological Treatment Minimum Dataset; RSCS = Reliable and Clinically Significant Change; RCSI = Reliable and Clinically Significant Improvement; MCID = Minimal Clinical Importance Difference; IMD = Index of Multiple Deprivation; OF = Outcome Feedback.

3.3 Meta-analysis

Overall, *n*=47 studies were included in the meta-analysis. The analyses were organised according to the outcome measures routinely used within IAPT services. Due to discrepancies with which measures were used within each study, this created different numbers in each analysis. Within the studies included here, 46 used the PHQ-9 as an outcome measure; 41 used the GAD-7 as an outcome measure; 19 used the WSAS as an outcome measure. Some of the included studies reported more than one effect size (ES) for independent samples contained within their original research (*n* = 8 studies). Where this occurred and the separate ES reported did not contain overlapping patient data, the effect sizes were included as independent samples. This was consistently implemented across the whole meta-analysis and subgroup analyses. For example, in the paper by Delgadillo et al. (2017a) separate ES are reported for different patient groups and therefore each group is represented by the individually reported ES. This means that whilst the number of studies is given in each description below, this does not always match the actual number in the ES calculations included in the meta-analysis. The number of studies and number of ES reported in each analysis will be reported for clarity. A very limited number of studies also include follow-up data (*n* = 4; Clark et al, 2009; Kenwright, McDonald, Talbot & Janjua, 2017; Meadows & Kellett, 2017; Pack & Condren, 2014). Due to the small number of these studies, follow-up outcomes have not been included within this meta-analysis.

3.3.1 Primary meta-analysis

Results for the PHQ-9 summarizing outcomes from 636,734 patients (mean *n* = 9796; median *n* = 619) across 46 studies (*n* = 65 independent samples) are reported in Figure 2. The overall combined pre-post treatment PHQ-9 effect size was large (*d* = 0.87, 95% CI [0.78-0.96],

p<.0001, number needed to treat [NNT]=2.17), indicating a statistically significant and large reduction in depression severity.

There was evidence of considerable heterogeneity across PHQ-9 outcome studies: I^2 = 98%; Q(df = 64) = 3600.47, p<.0001. Funnel plot asymmetry (see Figure 3) suggested the presence of publication bias. However, there was a non-significant rank correlation test (p=.196) and non-significant regression test for funnel asymmetry (p=.083). The fail-safe N analysis indicating the number of non-significant studies likely needed to be published to reduce the findings to a small clinically non-significant effect (d = 0.35) was 97.

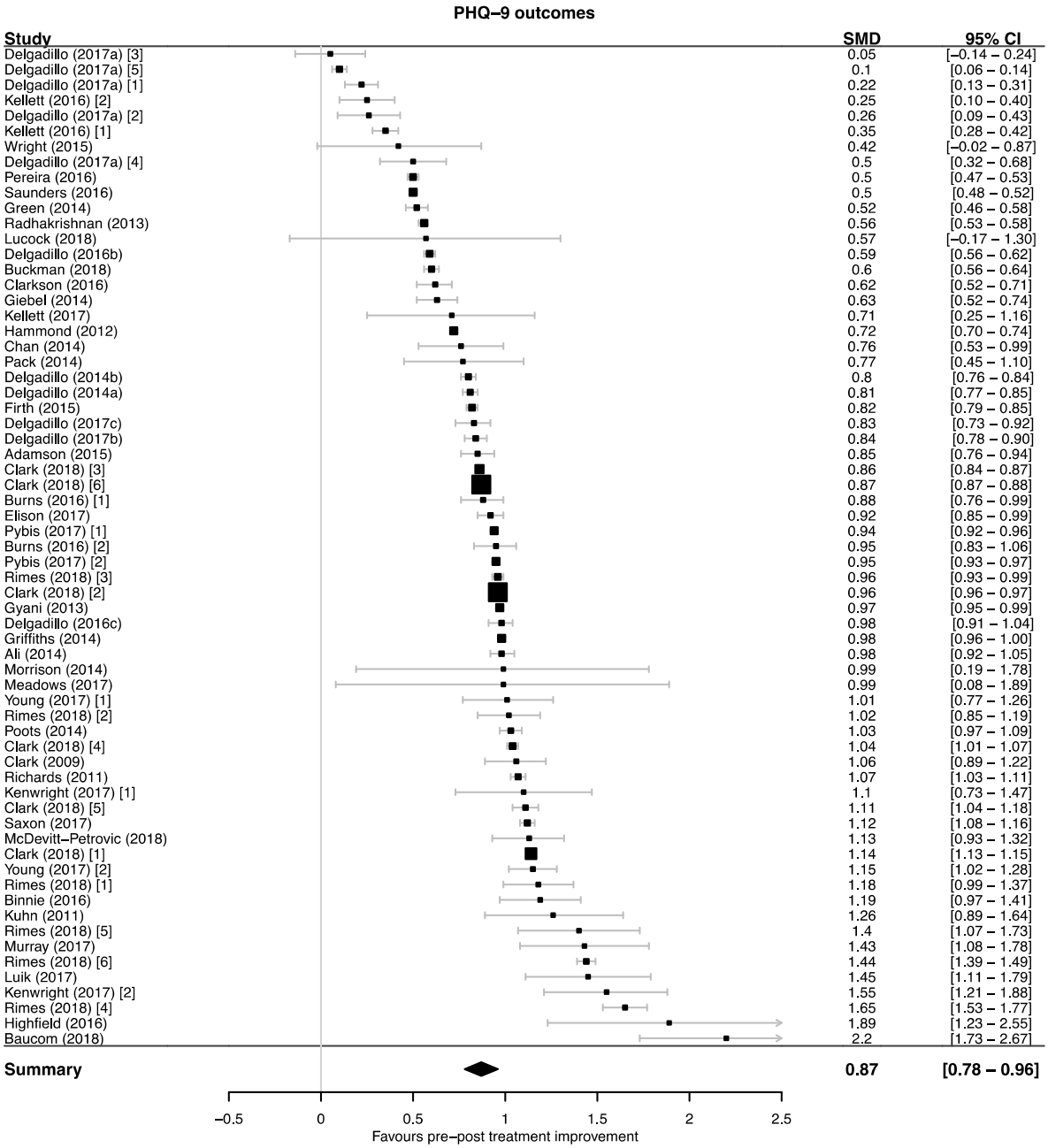

Figure 2: Forest plot of pre-post PHQ-9 independent samples effect sizes and the pooled treatment effect.

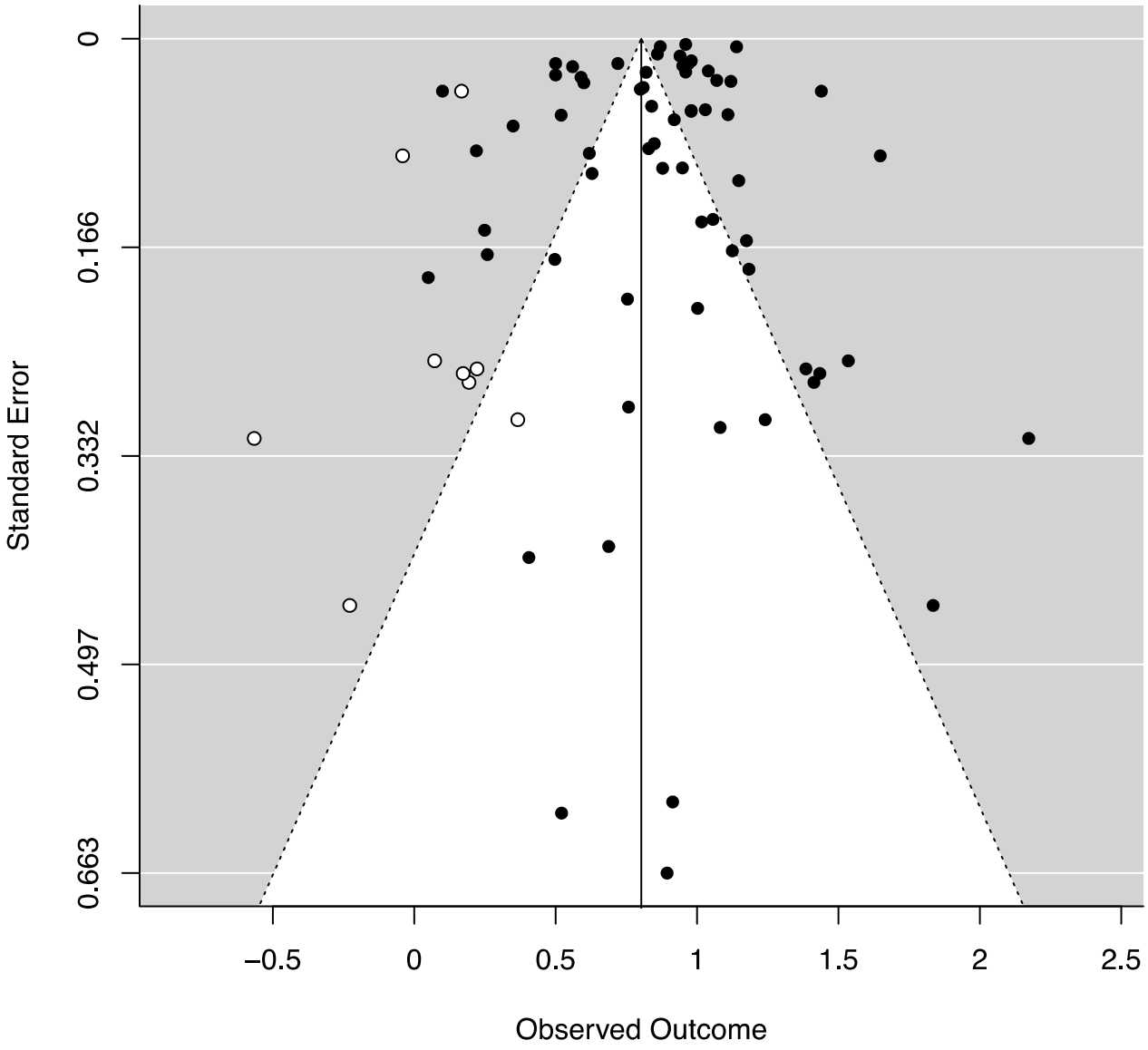

Figure 3: Funnel plot of the distribution of studies reporting pre-post PHQ-9 outcomes

Results for the GAD-7 included outcomes from 598,166 patients (mean *n* = 9969; median *n* = 541) across 41 studies (*n* = 60 independent samples) are reported in Figure 4. The overall combined pre-post treatment GAD-7 effect size was large (*d* = 0.88, 95% CI [0.79-0.97], p<.0001, NNT=2.15), indicating a statistically significant and large reduction in anxiety severity.

There was evidence of considerable heterogeneity across studies (I^2 = 98%; Q(df = 59) = 4239.30, p<.0001). There was some evidence of funnel plot asymmetry (see Figure 5); the funnel plot asymmetry regression test (p=.014) and the rank correlation test (p=.008) were significant indicating there may be some publication bias. However, the fail-safe N analysis indicated that 92 studies with null findings would be necessary to reduce the results to clinically non-significant. The results for the WSAS included data from 478,693 patients (mean *n* = 19,946; median *n* = 1351) from 19 studies (*n* = 24 independent samples) and are summarized in Figure 6.

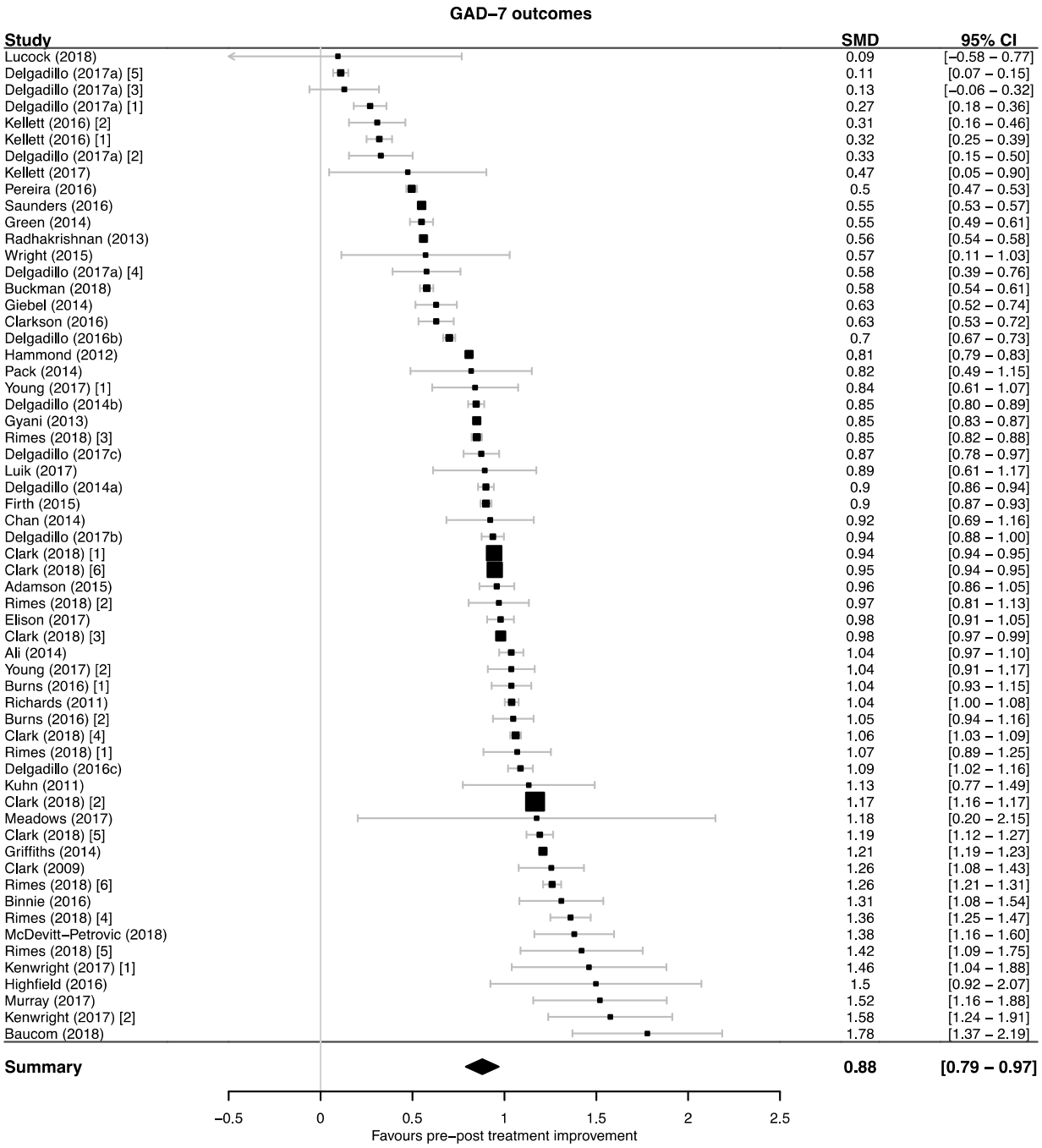

Figure 4: Forest plot of pre-post GAD-7 independent samples effect sizes and the pooled treatment effect.

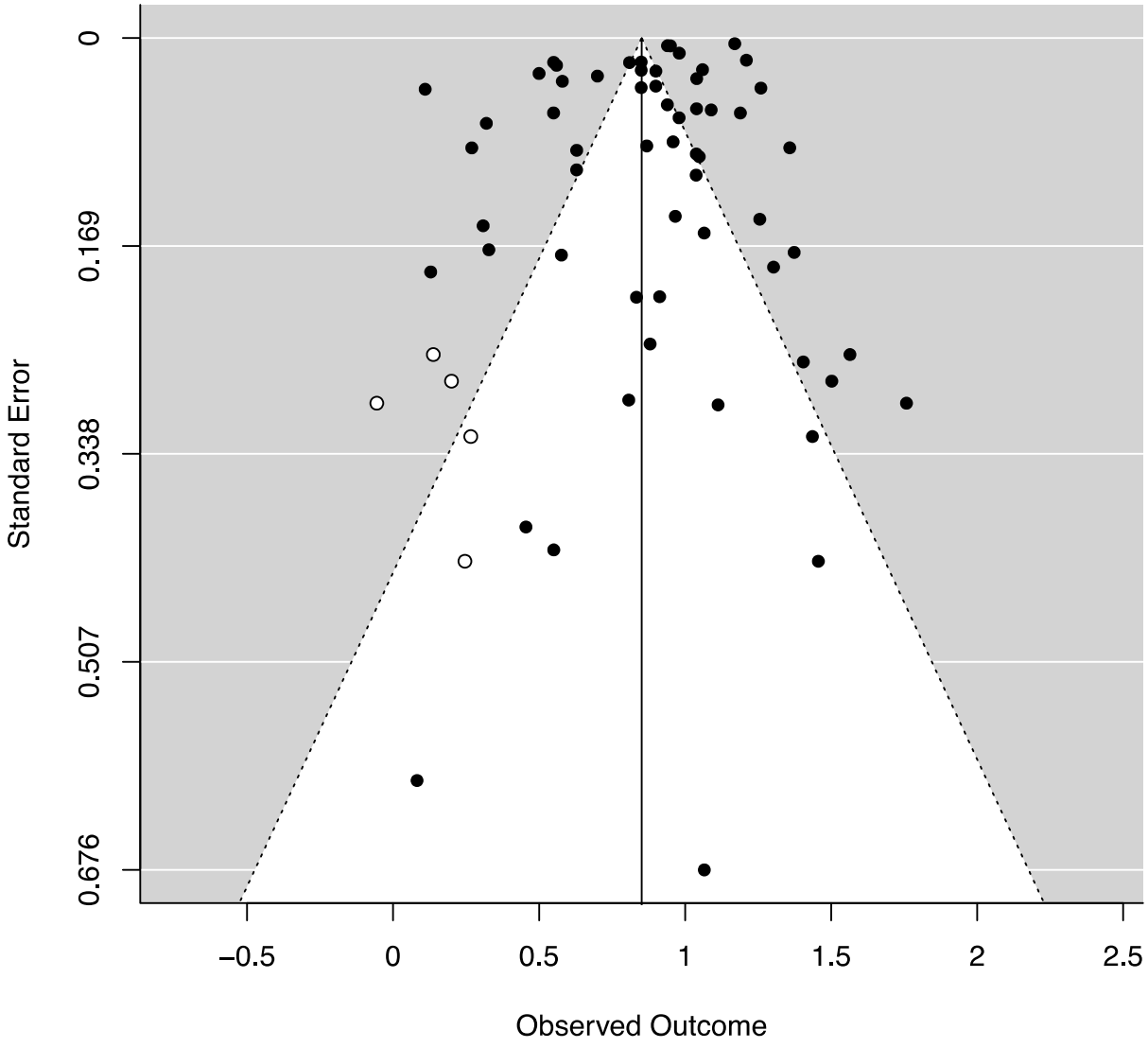

Figure 5: Funnel plot of the distribution of studies reporting pre-post GAD-7 outcomes.

The overall combined WSAS effect size was moderate (d = 0.55, 95% CI [0.48-0.61], p<.0001, NNT=3.30), indexing a statistically significant treatment effect on work and social adjustment.

There was evidence of considerable significant heterogeneity across studies (I^2 = 95%; Q(df = 23) = 524.11, p<.0001). Funnel plots were visually inspected and suggested some asymmetry with missing studies demonstrating larger effects (see Figure 7). The statistical tests showed mixed evidence of publication bias; the funnel plot asymmetry regression suggested significant asymmetry (p=.027) and the fail-safe N indicated 13 null-finding studies would reduce the average effect size to a small clinically non-significant pre-post improvement (d = 0.35), however the rank correlation test was not significant (p=.572).

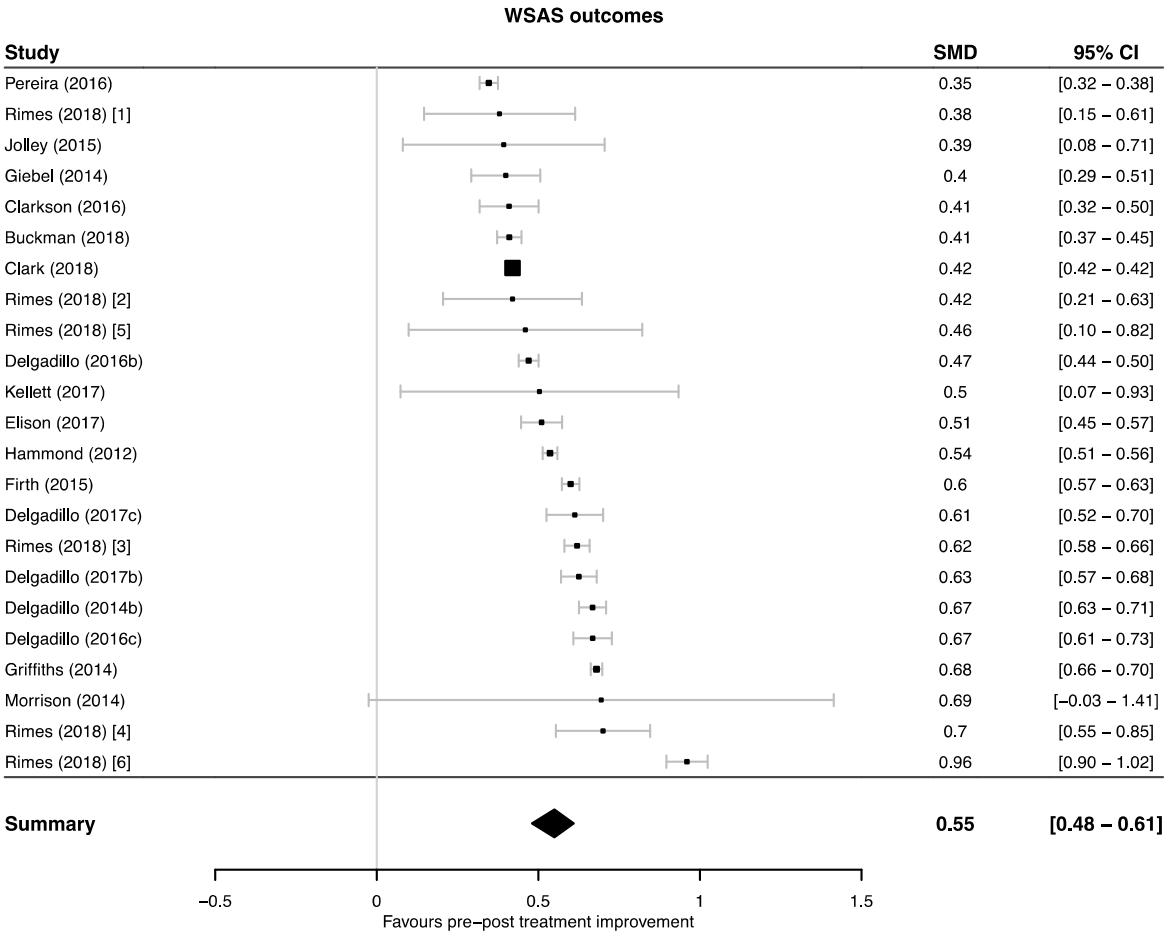

Figure 6: Forest plot of pre-post WSAS independent samples effect sizes and the pooled treatment effect.

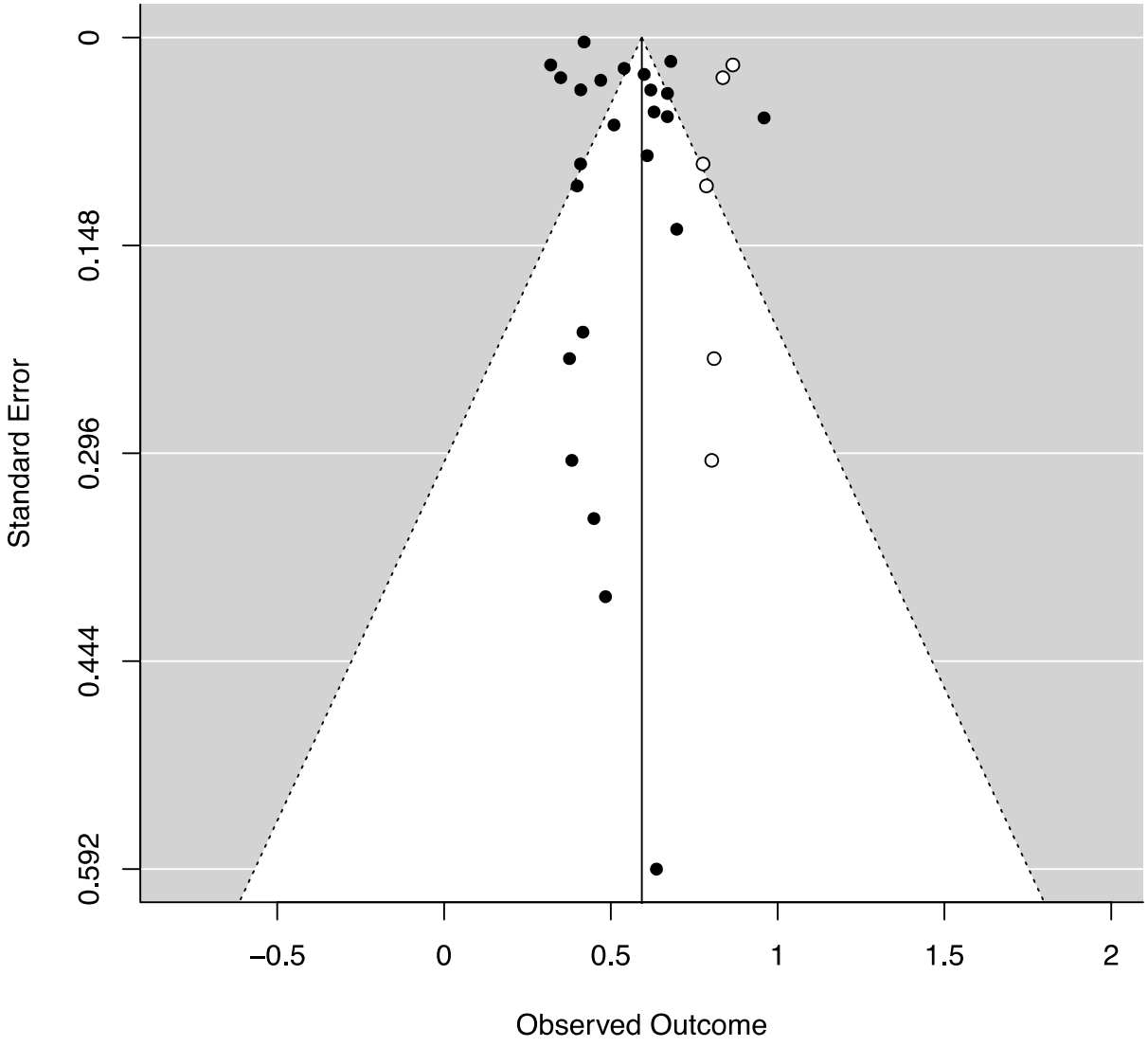

Figure 7: Funnel plot of the distribution of studies reporting pre-post WSAS outcomes.

Moderator and sensitivity analyses

Subgroup analyses of categorical variables.

Significant between-study heterogeneity was explored using subgroup analyses to investigate five categorical moderators of treatment effects across the three outcomes (Table 4). For PHQ-9 outcomes, significant variations in effect sizes by subgroups were evident for type of methodology used, primary condition, step of care and level of study bias. Completer analysis produced significantly larger effect sizes than ITT and studies of primary mental health conditions produced significantly larger effects than studies with physical health as the primary condition. Studies with increased levels of bias produced larger treatment effects than studies with less influence of bias and step 3 produced larger effects than step 2, however the subgroup differences in both comparisons were no longer significant after accounting for multiple testing. For GAD-7 outcomes, significant variations in effect sizes by subgroups were evident for type of methodology used , primary condition and level of study bias, showing the same pattern as in the PHQ-9 outcomes; significantly larger effects for completer analysis versus ITT, for treatment of primary mental health conditions versus physical health conditions and in studies with higher levels of bias. Effects for step of care were not significantly different for GAD-7 outcomes. The format of treatment did not explain variations in treatment effects for either PHQ-9 or GAD-7 outcomes and no significant variation in effects across subgroups were found for WSAS outcomes.

Table 4: Subgroup analysis of pre-post treatment effects

Outcome	Variable	Subgroup	*k*	Effect size	95% CI	I^2(%)	Q	Diff between subgroups (*p*)
PHQ-9	Methodology	ITT	43	0.78	0.67 to 0.90	99%	5701.68***	**.001****
		COM	22	1.04	0.97 to 1.12	98%	3128.89***	
	Study bias	Low	44	0.82	0.71 to 0.93	99%	5576.75***	.016*
		Medium	17	0.95	0.84 to 1.06	99%	4043.44***	
		High	4	1.26	0.96 to 1.56	77%	13.16**	
	Primary condition	Mental health	52	0.94	0.87 to 1.01	99%	9080.97***	**.001****
		Physical health	10	0.43	0.13 to 0.74	94%	170.82***	
	Step of care	Step 2 only	15	0.80	0.68 to 0.93	97%	686.96***	.038*
		Step 3 only	9	1.09	0.85 to 1.33	93%	120.46***	
	Format	Individual	38	0.77	0.65 to 0.89	99%	5194.26***	.500
		Group	6	0.88	0.59 to 1.16	92%	69.92***	
GAD-7	Methodology	ITT	41	0.80	0.68 to 0.91	99%	5484.73***	**<.001****
		COM	19	1.06	0.98 to 1.14	99%	3967.79***	
	Study bias	Low	41	0.83	0.72 to 0.94	98%	3655.71***	**.001****
		Medium	16	0.97	0.84 to 1.09	99%	5223.23***	
		High	3	1.36	1.10 to 1.62	24%	2.63	
	Primary condition	Mental health	47	0.96	0.88 to 1.04	99%	10813.46***	**.006****
		Physical health	10	0.50	0.19 to 0.82	94%	175.50***	

	Step of care	Step 2 only	14	0.88	0.74 to 1.03	98%	776.13***	.182
		Step 3 only	6	1.16	0.77 to 1.56	84%	32.46*	
	Format	Individual	33	0.76	0.62 to 0.89	99%	4782.75***	.291
		Group	6	0.91	0.66 to 1.16	93%	73.78**	
WSAS[1]	Methodology	ITT	21	0.54	0.48 to 0.61	98%	1236.70***	.154
		COM	3	0.44	0.32 to 0.57	0%	0.74	
	Study bias	Low	19	0.55	0.48 to 0.62	97%	780.55**	.389
		Medium	5	0.48	0.34 to 0.62	99%	810.32***	
	Step of care	Step 2 only	7	0.52	0.43 to 0.61	98%	432.12***	.239
		Step 3 only	2	0.44	0.33 to 0.55	0%	0.18	
	Format	Individual	12	0.48	0.41 to 0.55	98%	916.96	.930
		Group	2	0.48	0.45 to 0.51	0%	0.87	

*significant at $p < .05$ threshold; **significant at $p < .01$ threshold, ***significant at $p < .0001$ threshold, **between subgroup differences significant at Bonferroni adjusted $p<.01$ threshold for multiple testing**; [*1] Moderator analysis for 'primary condition' was not undertaken for the WSAS outcome measure as all studies included were deemed to be investigating mental health with none focusing purely on physical health. Abbreviations: *k:* number of comparisons per subgroup, CI: confidence interval, PHQ-9: Patient Health Questionnaire-9, GAD-7: Generalized Anxiety Disorder scale-7; WSAS; Work and social adjustment scale; ITT: Intention to treat; COM; completer.

Meta-regression analyses of continuous variables.

Significant between-study heterogeneity was explored using meta-regressions to investigate four continuous moderators of treatment effects across the three outcomes (Table 5). For GAD-7 and WSAS outcomes, between-study variations in effect sizes were not related to differences in the mean age or gender proportions of the study samples. PHQ-9 outcomes did show larger treatment effects when proportions of males increased, however the effect did not remain significant after adjusting for multiple testing. Mean treatment duration was significantly associated with between-study effect size variations for both PHQ-9 and GAD-7 outcomes, with larger effects evident when there was a greater mean number of sessions attended. Larger effects were also associated with higher baseline severity scores for PHQ-9 and GAD-7 outcomes (although the PHQ-9 effect did not remain significant after accounting for multiple testing. There was no association between intake score or treatment duration and variation in treatment effects for WSAS outcomes.

Table 5: Meta-regression analysis of pre-post treatment effects

Outcome	Variable	Range and mean	*k*	B-coefficient	95% CI	SE	*p*
PHQ-9	Gender (% female)	(0 – 100%; M=59.5)	52	-0.00	-0.01 to 0.00	0.00	.034*
	Mean age	(31 – 49 years; M=39.8)	45	-0.01	-0.03 to 0.00	0.01	.131
	Mean intake score	(7.9 – 18.8; M=15.0)	58	0.02	0.00 to 0.04	0.01	.015*
	Mean number of sessions	(3 – 16 sessions; M=6.7)	42	0.03	0.01 to 0.05	0.01	**.001****
GAD-7	Gender (% female)	(0 – 100%; M=59.1)	49	0.00	0.00 to 0.00	0.00	.079
	Mean age	(31 – 49 years; M=39.7)	43	-0.01	-0.03 to 0.00	0.01	.061
	Mean intake score	(3.7 – 18.3; M=13.5)	52	0.07	0.04 to 0.09	0.01	**<.001*****
	Mean number of sessions	(3 – 16 sessions: M=6.6)	38	0.03	0.01 to 0.05	0.01	.015*
WSAS	Gender (% female)	(0 – 100%; M=54.9)	22	0.00	0.00 to 0.00	0.00	.283
	Mean age	(31 – 49 years; M=39.3)	20	0.00	-0.02 to 0.02	0.01	.689
	Mean intake score	(14.8 – 24.5; M=19.3)	22	-0.01	-0.05 to 0.04	0.02	.751
	Mean number of sessions	(4 - 16 sessions; M=6.7)	18	0.02	-0.01 to 0.05	0.01	.163

Note: *significant at $p < .05$ threshold; **significant at $p < .01$ threshold, ***significant at $p < .0001$ threshold, **significant at Bonferroni adjusted *p*<.01 threshold for multiple testing**; Abbreviations: *k:* number of comparisons, CI: confidence interval; SE: standard error; M: mean; PHQ-9: Patient Health Questionnaire-9, GAD-7: Generalized Anxiety Disorder scale-7; WSAS; Work and social adjustment scale.

Sensitivity analysis excluding atypical studies.

Sensitivity analyses investigated the aggregated effect size for those studies that were more similar to each other, through excluding studies deemed to be atypical of routine IAPT services in terms of their population, target condition or treatment type. There were 8 studies excluded on this basis. The excluded studies focused on samples of: male offenders (Adamson et al, 2015), two studies of veterans (Clarkson et al, 2016; Giebel et al, 2014), deaf patients (Young et al, 2017), two studies of systemic therapy (Kuhn, 2011; DIT, Wright et al, 2015) and two studies due to both the population and treatment delivered (couples and BCT-D, Baucom et al, 2018; psychosis and CBT-p, Jolley et al, 2015). Meta-analyses for each outcome were re-run with the atypical studies excluded. Overall, and in comparison, to the primary meta-analysis of all studies, there was minimal difference in the ES from the sensitivity analysis. With regards to the PHQ-9, 57 separate comparisons contributed to the analysis producing a moderate to large ES of d = 0.85 (95% CI [0.80 to 0.90]; p<.0001, NNT=2.19). There was evidence of considerable heterogeneity across studies (I^2 = 98%; Q(df = 56) = 3557.37, p<.0001). The GAD-7 pooled ES was calculated from 52 typical studies and still indicated a large effect (d = 0.87, 95% CI [0.81 to 0.94]; p<.0001, NNT=2.17) with large between-study heterogeneity (I^2 = 98%; Q(df = 51) = 4201.15, p<.0001). Twenty-one comparisons contributed to the WSAS pooled ES, producing a moderate effect (d = 0.56, 95% CI [0.48 to 0.62]; p<.0001, NNT=3.25) with considerable heterogeneity still evident between studies (I^2 = 96%; Q(df = 20) =523.88, p<.0001). The moderator analyses were repeated in the typical study sample finding similar effects to the main analysis (see Appendix 5).

3.3 Quality assessment

The CASP tool was used to assess the quality of included studies. The summary results for methodological quality can be found in Table 3 (full results can be found in Appendix 3). Overall, more than half of studies within this review had low risk of bias (58%) and a smaller percentage indicated medium (30%) or high (12%) levels of bias. Just under half of studies either failed or could not be identified as having identified confounding factors within the design/analysis and did not take such factors into account. Whether the article findings fitted with other available evidence was difficult to assess or inconclusive in almost half of studies.

The studies rated as having high risk of bias tended to be studies focused on areas outside of mainstream IAPT (e.g., Cheston & Howells 2016). The quality of the evidence base was affected by the dearth of studies capturing follow-up outcomes for IAPT services. The most common sources of bias in the reviewed studies related to whether confounding factors within the study was not only reported, but also taken into account during analysis, as well as whether the study results were subject to follow-up. Around one-third of the included studies either did not discuss factors that could have influenced the results or did not take them into account within the analysis. This was particularly evident in those studies using completer analysis as differences between profiles of those participants who completed the therapy and those who were classed as dropping out were not adequately compared or discussed. In some studies, there were issues with confounding variables and difficulties ascertaining whether or not these had been taken into account in the analysis. Furthermore, only a minority of studies completed follow-up analysis. In four studies follow-ups were achieved, although in one of those studies it was deemed that the follow-up period was not adequately long enough. Further work in this area

would identify whether the outcomes achieved within the literature was maintained once therapy had ended, and for how long.

4. Discussion

This systematic review has identified and synthesized all available, peer reviewed, practice-based evidence generated within the UK's IAPT programme – an initiative originally designed to increase rapid access to evidence-based psychological treatments for those experiencing common mental disorders (Clarke et al, 2009; Clarke, 2011). The narrative review summarised *n*=60 studies that varied markedly in terms of the methods used, samples studied and outcomes analyzed. The meta-analysis aimed to quantify the overall impact of IAPT interventions standardized outcome measures, including data from over 600,000 patients. The exclusion of RCTs was based on a rationale of wanting to focus solely on outcomes achieved in routine practice, due to the common issues regarding generalizing from experimental studies to routine service delivery contexts (Lorenzo-Luaces, Johns, & Keefe, 2018).

The main findings from the primary meta-analysis found a large pre-post ES for both reductions in depression and anxiety, with a medium ES regarding improvements in work and social adjustment. The GAD-7 ES mirrors the results of the Stewart and Chambless (2009) meta-analysis of the effectiveness of CBT for adult anxiety disorders delivered in routine practice, which illustrated that pre-post outcomes on disorder-specific measures were large, and when benchmarked against the outcomes achieved in RCTs were equivalent. The PHQ-9 ES results mirror the Thimm and Antonsen (2014) meta-analysis of the treatment of depression in routine practice in that the ES at post-treatment was large (d = .97), and at post-treatment 44% demonstrated a significant improvement in depression. The tests of heterogeneity throughout the current meta-analyses indicated high levels of variability across studies and there was some

evidence of publication bias (for GAD-7 outcomes) so results should be interpreted cautiously. The ES reported here therefore compliment the recovery rates that are routinely submitted by IAPT services (Clark et al. 2019) to assess the effectiveness of the IAPT programme (alongside other targets related to wait-times for assessment, entry into treatment, return to work rates, etc).

4.1 Subgroup analyses

Studies using intention-to-treat (ITT) analyses were compared with completer analyses (COM). This is an important methodological distinction, due to resultant differences in those included participants whose data is analyzed (Kyrios, Hordern, & Fassnacht, 2015). ITT methods are recommended to minimise bias (Ranganathan et al, 2016), whereas COM tends to increase the rate of Type I errors (Fergusson, 2002). The ES for COM studies were larger than those using ITT across both anxiety and depression outcomes, and this is further compelling evidence that study designs which employ COM approaches for routinely delivered psychological interventions risk yielding overoptimistic and biased results.

Significant differences were found in the magnitude of ES observed at low and high intensity treatments for depression (albeit no longer significant after accounting for multiple testing). Although differences between low and high intensity were not significant for anxiety outcomes and functional impairment, there was a pattern of larger ES for high intensity interventions. This may have been due to the fact that when intake scores were assessed, there were no differences in symptom levels between step 2 and step 3 studies. PWPs are trained to post graduate certificate level via a national curriculum to work with mild-to-moderate anxiety and depression, with the psychoeducational approaches used being originally designed for such presentations (Kellett et al. 2020). Therefore, ES may be being suppressed at step 2 because PWPs are working with presentations that are too complex for the skill level of the practitioner

or the content of the intervention. This finding is a challenge to stepped care principles, as low intensity interventions are not assumed to be less effective, just less intense in format and more flexible in terms of service delivery method (Firth et al. 2015). Patients deemed 'complex cases' and accessing high intensity interventions directly show better outcomes than those accessing the low intensity treatments and then being stepped-up to high intensity interventions (Delgadillo, Morea & Lutz, 2016c; Delgadillo, Huey, Bennett & McMillan, 2017b). Other research has also investigated the use of predictive models to identify factors that may impact on outcomes at the various steps of IAPT – both at patient (e.g., demographic and clinical factors) and therapist levels (e.g., Delgadillo, Morea & Lutz, 2016c; Firth et al, 2015). The average duration of IAPT treatments (mean = 6.7) was associated with larger treatment effects for depression and anxiety outcomes (although anxiety effects were not significant after controlling for multiple testing). This finding is in line with national evidence that suggests the average length of an IAPT treatment is 7 sessions and that patients that move to recovery attend 8 sessions on average (NHS Digital, 2019).

4.2 Study limitations

There are some limitations within this systematic review and meta-analysis that should be considered when interpreting the above results. The absence of any control comparators across studies means that the reported pre-post effect sizes are likely to be influenced by statistical artefacts such as regression to the mean and also natural recovery (Posternak & Miller, 2001; Whiteford et al., 2013). Therefore, the magnitude of effects reported in the meta-analysis could be overestimating the "true effect" of routinely delivered psychological interventions. Furthermore, the exclusion criteria imposed in the systematic searches could have further restricted the potential for studies to be included. For example, there are many examples of

evaluation work completed that does not get reported in peer-reviewed scientific journals, such as in the 'grey' literature including clinical audits and service evaluation projects. However, the funnel plots produced using meta-analysis did not provide evidence of publication bias. As IAPT produces monthly and bi-annually reports with outcome measure data, it is possible that there was some overlap between studies regarding patient outcome data. This was minimised as much as possible, such as through the exclusion of studies where the data is presented in an earlier publication. However, this can only be minimised and not completely excluded as some studies used data from across CCG areas (e.g., Delgadillo et al, 2016a), which may have overlapped with data contained within other individual IAPT service studies.

A consistent study limitation concerned the lack of any indices of treatment fidelity, integrity or competency (i.e., the studies cannot be sure that what was said to be delivered, was actually delivered). IAPT strongly emphasises adherence and fidelity to NICE guidance, however, without evidenced fidelity to the particular delivered intervention, this can appear meaningless. This may be the mechanism driving the evidence concerning widespread differences between therapists (Green et al., 2014), teams and services (Gyani et al., 2013) that exists. A lack of fidelity has also implications for internal and external validity, statistical power and ethical considerations (Moncher & Prinz, 1991). For example, without treatment adherence and fidelity measures it is difficult to truly ascertain that any symptomatic changes following treatment is due to that particular treatment. Furthermore, any deterioration in a patients' progress within therapy where fidelity to the model has been less than adequate (or not recorded) could lead to arguments that non-adherence of protocols can lead to 'harm'. Treatment adherence, fidelity and integrity can be assessed in various ways, such as through the auditing of the use of treatment protocols, matching in-session change methods to the appropriate stage of

the treatment protocol, within live supervision, and through the reports provided by patients. Evidence has shown a differing uptake of the use of fidelity measures, with one study indicating an increase in treatment fidelity reporting within research outputs (Moncher & Prinz, 1991) whilst another study did not find this trend (Borrelli, Sepinwall, Ernst, Bellg, & Czajkowski et al., 2005).

There was evidence of high study heterogeneity which is interesting given the aim of IAPT being to introduce homogeneity of service delivery. Therefore, whilst there were high levels of heterogeneity in the primary meta-analysis, some of this could be accounted for by study features, such as those using a completers-type methodology were more likely to be similar to each other than those studies using ITT methods. The lack of follow-up studies means that relatively little is known regarding the durability of the effects of IAPT.

As previously reported, from the available data from the last 6 years (since 2012/2013), approximately 4.9 million patients started treatment with just over 2.6 million patients being recorded as having completed treatment (i.e., receiving 2 or more sessions within the service). From the present systematic review, we have found that a total of just under 1 million patients' data has been included from 60 studies, which indicates that there are substantially more patients entering the IAPT programme (even just within the last 6 years of available data), than are being reported within published studies from the last 10-years. This is a major mismatch between delivery and evaluation. This limits the potential for learning from the IAPT programme and potentially improving systems and outcomes. It also highlights just how little research is being completed and published in relation to the large amounts of available evidence.

4.3 Future work and clinical implications

Whilst IAPT continues to expand and evolve, the present review has highlighted implications for practitioners working with IAPT services, and also wider service implications. Taking into account the small number of studies looking at group interventions, the review indicated that group interventions have a similar effectiveness in regard to outcome measure compared to individual interventions. This therefore suggests that services could look into implementing group interventions more often. This may increase the efficiency of services at no loss of effectiveness. Group interventions still require highly skilled delivery and close supervision. In the review, group interventions were utilised for individuals with several different presenting problems and occurred in both mental health and long-term physical health populations – for example, in those experiencing common mental health difficulties, long-term conditions (e.g., diabetes) and more enduring conditions (e.g., PTSD) as well as within both low and high intensity therapies. This again is encouraging and suggests that group interventions can be effective with a range of presenting conditions and at different steps of the IAPT model.

One of the IAPT programmes' unique features is that it produces transparent monitoring of outcome measures on a large, national scale. As seen within this review, the minimum dataset of outcome measures within IAPT includes a measure of depression, anxiety and functional disability. Whilst other disorder-specific measures can be used, this review highlighted that practitioners are seldom using these to monitor outcome changes. This brings into question whether change is being captured accurately. There are no reports of idiographic outcomes in IAPT patients and this would supplement nomothetic outcome analysis well. Supplementing the minimum dataset with other disorder-specific measures by practitioners would further expand the understanding of the IAPT programmes' impact. Alongside this is the issue of reliable competency assessment of interventions at step 2 and step 3 and whether the treatment protocols

are being adhered to. Whilst IAPT practitioners have competency assessments integrated into their training and post qualification clinical supervision (and they form part of BABCP accreditation for CBT therapists), the measures are not reported in the outcome evidence base. Future research would therefore benefit from integrating such assessments into their methods.

In addition to the practitioner and service implications, the current review has also shed light onto areas of potential useful future research. The following are indicated (a) studies analyzing outcomes on other disorder-specific measures; (b) studies describing the interventions in greater detail; (c) consistent use of measures of treatment fidelity and competency; (d) studies investigating moderators and mediators of depression and anxiety outcomes; (e) studies collecting longer-term follow-up outcomes data; (f) more consistent reporting of drop-out rates, and (g) studies modelling and exploring variability between therapists/services/regions and using site as a moderator in any future meta-analysis. Any future IAPT outcome paper should as matter of course report whether they are taking an IIT versus completer approach, the percentage of patients treated at each step, the stepping up rate, the dropout rate, pre and post-treatment means (SDs) and effect sizes on the IAPT MDS measures as well the disorder-specific outcome measures. In terms of the policy implications, the following are of note; (a) the commissioning of routine follow-up data collection, (b) identifying numbers of patients that are re-referred for IAPT treatment; (c) open access to routinely collected patient-level IAPT datasets, to enable research to keep pace with the often rapidly shifting IAPT policy context. Cross and Hickie (2017) have argued that stepped-care models fail to deal with the high levels of comorbidity and/or complexity seen in routine services, suggesting that that technically integrated transdiagnostic models are more fit for purpose in matching presentation to timely delivery of the right intensity and dose of intervention to ensure rapid access. The main clinical implications of

the research are; (a) the expansion of choice that has occurred at high intensity (e.g. provision of interpersonal psychotherapy, dynamic interpersonal psychotherapy, counselling for depression and couple work for depression) has not been mirrored at low intensity and needs to take place (e.g. Meadows & Kellett, 2017); (b) there is increasing evidence to support stratified models of treatment-matching for more complex cases; (c) solely reporting recovery rates are a limited index of the effectiveness of any therapist and/or service and (d) combining the PHQ-9, GAD-7 and WSAS numbers needed to treat (NNT) results suggests that 2.54 patients need to be seen for one patient to gain benefit from an IAPT intervention. The clinical significance of the results overall is that original aim of the IAPT programme was to increase access to evidence-based talking treatments and large numbers are being treated annually, with evidence that recovery rates are slowly increasing and achieving the 50% target (IAPT, 2019). This however leaves open the possibility of a creating a 'forgotten-fifty' of referrals that do not experience any benefit and also the so far unrecorded rates of patients that seek further IAPT interventions, thus creating a 'revolving door' scenario (Cotton, 2019). There is evidence to suggest that patient complexity exerts a negative influence on outcomes, particularly when depression is inappropriately treated at step 2 initially in IAPT services (Delgadillo, Huey, Bennett, & McMillan, 2017).

Another area of inquiry could be to continue exploring how IAPT can become more effective for people with comorbid long-term conditions and medically unexplained symptoms. Within this present review, the extent to which these more physical health conditions are being researched, which at the current time is quite limited. In terms of disorder specific outcome measures, then more research needs to be produced of effect sizes for IAPT interventions on measures well attuned to the presenting problem. There is a clear need for studies to describe the interventions used at each step in more detail, to expand the use of measures of fidelity and

competency and also to gather short and long-term follow-up outcomes. Within the studies quantified using meta-analytic methods (n=29), there were only four studies that included a follow-up period, which ranged in length from 4 weeks (Meadows & Kellett, 2017) to an average of 42 weeks (Clark et al., 2009). Some recent research has been investigating the use of 'top-up' sessions following IAPT treatment for those individuals who do relapse which has shown positive results (Lucock et al., 2018). More studies need to understand what moderates and mediates depression and anxiety outcomes at both step 2 and step 3 of IAPT services. More research needs to explore why some individuals drop out of or do not benefit from IAPT intervention, including areas of harm or deterioration following IAPT intervention. Using mixed methods to investigate this particular field of research would open up further areas of exploration to continue to understand the impact of the IAPT programme. More studies need to be conducted that explore what creates the variability between IAPT therapists, teams and services, rather than just recording that variability exists.

A growing area of research has further explored therapist effects, i.e., the degree to which patients treated by different therapists have similar treatment outcomes. There has been variation between studies reporting therapist effects, and a recent systematic review indicated a therapist effects range of 0.2-29% from different study designs (the majority being practice-based) and further supported the notion that it matters most who the therapist is to those who enter treatment with higher baseline symptom severity (Johns, Barkham, Kellett & Saxon, 2019). Despite the positive and significant results arising from the meta-analysis, there was a lack of longer-term follow-up from the majority of the studies. Within the studies quantified using meta-analytic methods (n=47), there were only four studies that included a follow-up period, which ranged in length from 4 weeks (Meadows & Kellett, 2017) to an average of 42 weeks (Clark et al., 2009).

It is important to understand whether the positive and significant outcomes following IAPT treatment is maintained long-term, for how long and also for whom. It is expected that relapse will occur in some of those clients who made positive progress as has been shown in the literature (Ali et al., 2017). Some recent research has been investigating the use of 'top-up' sessions following IAPT treatment for those individuals who do relapse which has shown positive results (Lucock et al., 2018). Other areas of growing research have also been looking into those individuals more at risk of both dropping out and of experiencing relapse, with certain clinical, demographic and therapist factors contributing to these (Marshall et al., 2016; Ali et al., 2017). Therefore, if more was known about those who may be at a higher risk of relapse, which may be further evidenced by following-up clients going through the IAPT treatment once it has finished, there may be potential for working with those individuals to support them in their recovery and help to prevent some individuals from relapsing altogether.

4.4 Conclusion

The IAPT programme is an example of public healthcare transformation informed by scientific evidence (Clarke, Thomas & James, 2013). A huge amount of investment has occurred to enable and then maintain the IAPT programme and it has transformed the landscape of psychological services for people with anxiety and depression in the United Kingdom (Firth et al., 2019). It has served as a model for other national implementation programmes (e.g., Australia, Germany and Canada). It will be interesting to learn reciprocally from these programmes (and conduct cross-national meta-analyses) to continue to enhance and improve the services for those in receipt of IAPT care. From the results of the present meta-analysis, further expansion of the IAPT programme is warranted in the area of physical health, and this has begun already albeit it is in the early stage. Further, the results indicate that group interventions, at

different steps of the service, appear to show similar effectiveness compared to individual interventions. Whilst this should be interpreted with some caution due to smaller study numbers in the group sub-group analysis, this is a promising finding which could have encouraging implications at a client and service level. The transparency of outcome measurement and the sheer scale of data completion is quite unique and thus the learning potential that programmes such as IAPT afford are huge (Clark et al., 2018). Regular meta-analyses should supplement the reporting of national recovery rates each year. The IAPT programme is yet to achieve its scientific potential in offering a unique infrastructure from which further evidence-based practice can be developed, at both the national and international level.

References

* indicates included studies

*Adamson, V., Gibbs, M., & McLaughlin, D. (2014). Evaluation of the improving access to psychological therapies for offenders programme at HMP Lincoln: a three year prospective cohort study. *The Journal of Forensic Psychiatry & Psychology, 26*(2), 185–201. doi:10.1080/14789949.2014.985695

*Ali, S., Littlewood, E., McMillan, D., Delgadillo, J., Miranda, A., Croudace, T., & Gilbody, S. (2014). Heterogeneity in patient-reported outcomes following low-intensity mental health interventions: A multilevel analysis. *PLoS ONE, 9*(9), e99658. doi:10.1371/journal.pone.0099658

*Ali, S., Rhodes, L., Moreea, O., McMillan, D., Gilbody, S., Leach, C., … Delgadillo, J. (2017). How durable is the effect of low intensity CBT for depression and anxiety? Remission and relapse in a longitudinal cohort study. *Behaviour Research and Therapy, 94*, 1–8. doi:10.1016/j.brat.2017.04.006

*Baucom, D. H., Fischer, M. S., Worrell, M., Corrie, S., Belus, J. M., Molyva, E., & Boeding, S. E. (2017). Couple-based intervention for depression: An effectiveness study in the national health service in England. *Family Process, 57*(2), 275–292. doi:10.1111/famp.12332

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1088-1101. doi: 10.2307/253344

*Binnie, J., & Boden, Z. (2016). Non-attendance at psychological therapy appointments. *Mental Health Review Journal, 21*(3), 231–248. doi:10.1108/mhrj-12-2015-0038

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

*Branson, A., Myles, P., Mahdi, M., & Shafran, R. (2017). The relationship between competence and patient outcome with low-intensity cognitive behavioural interventions. *Behavioural and Cognitive Psychotherapy, 46*(1), 101–114. doi:10.1017/s1352465817000522

*Buckman, J. E. J., Naismith, I., Saunders, R., Morrison, T., Linke, S., Leibowitz, J., & Pilling, S. (2018). The impact of alcohol use on drop-out and psychological treatment outcomes in improving access to psychological therapies services: An audit. *Behavioural and Cognitive Psychotherapy, 46*(5), 513–527. doi:10.1017/s1352465817000819

*Burns, P., Kellett, S., & Donohoe, G. (2015). “Stress Control” as a large group psychoeducational intervention at step 2 of IAPT services: Acceptability of the approach and moderators of effectiveness. *Behavioural and Cognitive Psychotherapy, 44*(4), 431–443. doi:10.1017/s1352465815000491

*Chan, S. W. Y., & Adams, M. (2014). Service use, drop-out rate and clinical outcomes: A comparison between high and low intensity treatments in an IAPT service. *Behavioural and Cognitive Psychotherapy, 42*(6), 747–759. doi:10.1017/s1352465813000544

*Cheston, R., & Howells, L. (2015). A feasibility study of translating “Living Well with Dementia” groups into a primary care improving access to psychological therapy service (innovative practice). *Dementia, 15*(2), 273–278. doi:10.1177/1471301215582104

*Clark, D. M., Layard, R., Smithies, R., Richards, D. A., Suckling, R., & Wright, B. (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behaviour Research and Therapy, 47*(11), 910–920. doi:10.1016/j.brat.2009.07.010

*Clark, D. M., Canvin, L., Green, J., Layard, R., Pilling, S., & Janecka, M. (2018). Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. *The Lancet, 391*, 679–686. doi:10.1016/s0140-6736(17)32133-5

*Clarkson, P., Giebel, C. M., Challis, D., Duthie, P., Barrett, A., & Lambert, H. (2016). Outcomes from a pilot psychological therapies service for UK military veterans. *Nursing Open, 3*(4), 227–235. doi:10.1002/nop2.57

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

Cotton, E. (2019). IAPT revolving door data casts doubt on recovery rate. Downloaded from: http://www.cbtwatch.com/iapt-revolving-door-data-casts-doubt-on-recovery-rate/

Critical Appraisal Skills Programme (CASP). (2014). CASP Cohort Studies Checklist. Oxford: CASP. Retrieved from: https://casp-uk.net/casp-tools-checklists/

Cross, S.P.M., & Hickie, I. (2017). Transdiagnostic stepped care in mental health. *Public Health Research Practice, 27*(2):e2721712. doi: org/10.17061/phrp2721712

*Delgadillo, J., Asaria, M., Ali, S., & Gilbody, S. (2016). On poverty, politics and psychology: the socioeconomic gradient of mental healthcare utilisation and outcomes. *British Journal of Psychiatry, 209*(5), 429–430. doi:10.1192/bjp.bp.115.171017

*Delgadillo, J., Dawson, A., Gilbody, S., & Böhnke, J. R. (2017). Impact of long-term medical conditions on the outcomes of psychological therapy for depression and anxiety. *British Journal of Psychiatry, 210*(1), 47–53. doi:10.1192/bjp.bp.116.189027

*Delgadillo, J., Huey, D., Bennett, H., & McMillan, D. (2017). Case complexity as a guide for psychological treatment selection. *Journal of Consulting and Clinical Psychology, 85*(9), 835–853. doi:10.1037/ccp0000231

*Delgadillo, J., Kellett, S., Ali, S., McMillan, D., Barkham, M., Saxon, D., … Lucock, M. (2016). A multi-service practice research network study of large group psychoeducational cognitive behavioural therapy. *Behaviour Research and Therapy, 87*, 155–161. doi:10.1016/j.brat.2016.09.010

*Delgadillo, J., McMillan, D., Leach, C., Lucock, M., Gilbody, S., & Wood, N. (2014). Benchmarking routine psychological services: A discussion of challenges and methods. *Behavioural and Cognitive Psychotherapy, 42*(1), 16–30. doi:10.1017/s135246581200080x

*Delgadillo, J., McMillan, D., Lucock, M., Leach, C., Ali, S., & Gilbody, S. (2014). Early changes, attrition, and dose-response in low intensity psychological interventions. *British Journal of Clinical Psychology, 53*(1), 114–130. doi:10.1111/bjc.12031

*Delgadillo, J., Moreea, O., & Lutz, W. (2016). Different people respond differently to therapy: A demonstration using patient profiling and risk stratification. *Behaviour Research and Therapy, 79*, 15–22. doi:10.1016/j.brat.2016.02.003

*Delgadillo, J., Overend, K., Lucock, M., Groom, M., Kirby, N., McMillan, D., … de Jong, K. (2017). Improving the efficiency of psychological treatment using outcome feedback technology. *Behaviour Research and Therapy, 99*, 89–97. doi:10.1016/j.brat.2017.09.011

Department of Health. (2012). IAPT three-year report. The first million patients. Retrieved from: http://www.iapt.nhs.uk/silo/files/iapt-3-year-report.pdf

Egger, M., Davey., Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629 634. doi:10.1136/bmj.315.7109.629

*Elison, S., Ward, J., Williams, C., Espie, C., Davies, G., Dugdale, S., … Smith, K. (2017). Feasibility of a UK community-based, eTherapy mental health service in Greater Manchester: repeated-measures and between-groups study of "Living Life to the Full Interactive", "Sleepio" and "Breaking Free Online" at "Self Help Services." *BMJ Open, 7*(7), e016392. doi:10.1136/bmjopen-2017-016392

Fergusson, D. (2002). Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ, 325*(7365), 652–654. doi:10.1136/bmj.325.7365.652

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

Firth, N., Delgadillo, J., Kellett, S., & Lucock, M. (2019). The influence of socio-demographic similarity and difference on adequate attendance of group psychoeducational cognitive behavioural therapy. *Psychotherapy Research*, *30*(3), 362-374. doi:10.1080/10503307.2019.1589652

Foa, E.B., Kozak, M.J., Salkovskis, P.M., Coles, M.E., & Amir, N. (1998). The validation of a new obsessive-compulsive disorder scale: The Obsessive-Compulsive Inventory. *Psychological Assessment, 10*(3), 206-214.

*Giebel, C. M., Clarkson, P., & Challis, D. (2014). Demographic and clinical characteristics of UK military veterans attending a psychological therapies service. *The Psychiatric Bulletin, 38*(6), 270–275. doi:10.1192/pb.bp.113.046474

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

Gordon, M. & Lumley, T. (2019). 'forestplot'. R package version 1.9.

*Green, H., Barkham, M., Kellett, S., & Saxon, D. (2014). Therapist effects and IAPT psychological wellbeing [ractitioners (PWPs): A multilevel modelling and mixed methods analysis. *Behaviour Research and Therapy, 63*, 43–54. doi:10.1016/j.brat.2014.08.009

*Griffiths, C. A., & Griffiths, L. J. (2015). Recovery and reliable change rates for patients scoring severe on depression, anxiety or impaired functioning in a psychological therapies service: IAPT. *Mental Health Review Journal, 20*(1), 28–35. doi:10.1108/mhrj-06-2014-0022

*Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: Lessons from year one of IAPT. *Behaviour Research and Therapy, 51*(9), 597–606. doi:10.1016/j.brat.2013.06.004

Hamilton, W. K., Aydin, B., & Mizumoto, A. (2016). 'MAVIS: Meta Analysis via Shiny'. R package version 1.1.2.

*Hammond, G. C., Croudace, T. J., Radhakrishnan, M., Lafortune, L., Watson, A., McMillan-Shields, F., & Jones, P. B. (2012). Comparative effectiveness of cognitive therapies delivered face-to-face or over the telephone: An observational study using propensity methods. *PLoS ONE, 7*(9), e42916. doi:10.1371/journal.pone.0042916

Health Education England. Retrieved from: https://www.hee.nhs.uk/our-work/mental-health/improving-access-psychological-therapies

Health and Social Care Information Centre (HSCIC). (2015). Improving Access to Psychological Therapies (IAPT) Report. Retrieved from: http://www.hscic.gov.uk/catalogue/PUB17755

Health and Social Care Information Centre (HSCIC). (2019). Improving Access to Psychological Therapies (IAPT) Executive Summary. Retrieved from: https://files.digital.nhs.uk/B8/5E5409/iapt-month-dec-2018-exec-sum.pdf

*Highfield J.A., Lowe K., Lewis E., Warren R., Martin K., Walkett E. (2016). Clinical psychologists training and supervising IAPT therapists to work with long-term conditions and medically unexplained symptoms: A service development project. *Clinical Psychology Forum,* (280), 34-39.

Humphreys, M., de la Sierra, R. S., & van der Windt, P. (2013). Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration. *Political Analysis, 21,* 1–20. doi:10.1093/pan/mps021

Johns, R.G., Barkham, M., Kellett, S., & Saxon, D. (2018) A systematic review of therapist effects: A critical narrative update and refinement to Baldwin and Imel's (2013) review. *Clinical Psychology Review, 67*, 78-93. doi: 10.1016/j.cpr.2018.08.004.

*Jolley, S., Garety, P., Peters, E., Fornells-Ambrojo, M., Onwumere, J., Harris, V., … Johns, L. (2015). Opportunities and challenges in improving access to psychological therapies for people with severe mental illness (IAPT-SMI): Evaluating the first operational year of the South London and Maudsley (SLaM) demonstration site for psychosis. *Behaviour Research and Therapy, 64*, 24–30. doi:10.1016/j.brat.2014.11.006

*Kellett, S., Simmonds-Buckley, M., Bliss, P., & Waller, G. (2017). Effectiveness of group behavioural activation for depression: A pilot study. *Behavioural and Cognitive Psychotherapy, 45*(4), 401–418. doi:10.1017/s1352465816000540

*Kellett, S., Webb, K., Wilkinson, N., Bliss, P., Ayers, T., & Hardy, G. (2016). Developing services for patients with depression or anxiety in the context of long-term physical health conditions and medically unexplained symptoms: Evaluation of an IAPT pathfinder site. *Behavioural and Cognitive Psychotherapy, 44*(5), 553–567. doi:10.1017/s1352465816000114

*Kenwright, M., McDonald, J., Talbot, J., & Janjua, K. (2017). Do symptoms of irritable bowel syndrome improve when patients receive cognitive behavioural therapy for co-morbid anxiety disorders in a primary care psychological therapy (IAPT) service? *Behavioural and Cognitive Psychotherapy, 45*(5), 433–447. doi:10.1017/s1352465817000200

*Kuhn, P. (2011). Improving access to psychological therapies: systemic therapy in the Newham pilot site. *Journal of Family Therapy, 33*(4), 400–414. doi:10.1111/j.1467-6427.2011.00545.x

Layard, R., & CEP Mental Health Policy Group. (2006). The depression report: A new deal for depression and anxiety disorders. CEP Special Papers 15, Centre for Economic Performance, LSE.

Lorenzo-Luaces, L., Johns, E., & Keefe, J.R. (2018). The generalizability of randomized controlled trials of self-guided internet-based cognitive behavioral therapy for depressive symptoms: systematic review and meta-regression analysis. *Journal of Medical Internet Research*, 2018;20(11):e10113. doi:10.2196/10113

*Lucock, M., Bartys, S., Cupac, J., Delgadillo, J., Denton, C., Gaines, S., … Stebbings, R. (2018). Using implementation intentions to prevent relapse after psychological treatment for depression – the SMArT intervention. *Behavioural and Cognitive Psychotherapy, 46*(5), 626–632. doi:10.1017/s1352465818000255

*Luik, A. I., Bostock, S., Chisnall, L., Kyle, S. D., Lidbetter, N., Baldwin, N., & Espie, C. A. (2017). Treating depression and anxiety with digital cognitive behavioural therapy for insomnia: A real world NHS evaluation using stadardised outcome measures. *Behavioural and cognitive psychotherapy, 45,* 91-96. doi:10.1017/s1352465816000369

Marshall, D., Quinn, C., Child, S., Shenton, D., Pooler, J., Forber, S & Byng, R. (2016). What IAPT services can learn from those who do not attend. *Journal of Mental Health, 25*(5). 410-415, doi: 10.3109/09638237.2015.1101057

*Matthew Prina, A., Marioni, R. E., Hammond, G. C., Jones, P. B., Brayne, C., & Dening, T. (2014). Improving access to psychological therapies and older people: Findings from the Eastern region. *Behaviour Research and Therapy, 56*, 75–81. doi:10.1016/j.brat.2014.03.008

*McDevitt-Petrovic, O., Kirby, K., McBride, O., Shevlin, M., McAteer, D., Gorman, C., & Murphy, J. (2018). Preliminary findings of a new primary and community care psychological service in Northern Ireland: Low-intensity cognitive behavioural therapy for common mental health difficulties. *Behavioural and Cognitive Psychotherapy, 46*(6), 761–767. doi:10.1017/s1352465818000322

*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. doi:10.1017/s1352465816000485

*Methley, A., Woodruff, P., Sayer, N., & Nevin, J. (2016). Readiness to engage in therapy for PTSD: A service evaluation of a pilot pre-therapy group within an IAPT service. *Clinical Psychology Forum*, (288), 23-27.

*Mofrad, L., & Webster, L. A. D. (2012). The treatment of depression and simple phobia through an interpreter in the North East of England: a case study. *The Cognitive Behaviour Therapist, 5*(4), 102–111. doi:10.1017/s1754470x13000044

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med 6*(7), e1000097. doi:10.1371/journal.pmed1000097

*Morrison, C., Walker, G., Ruggeri, K., & Hacker Hughes, J. (2014). An implementation pilot of the MindBalance web-based intervention for depression in three IAPT services. *The Cognitive Behaviour Therapist, 7*. doi:10.1017/s1754470x14000221

*Murray, H. (2017). Evaluation of a trauma-focused CBT training programme for IAPT services. *Behavioural and Cognitive Psychotherapy, 45*(5), 467–482. doi:10.1017/s1352465816000606

National institute for Health and Care Excellence. (2009). *Depression: the treatment and management of depression in adults (update).* NICE clinical guideline 90 (CG90). Retrieved from: https://www.nice.org.uk/guidance/cg90

National institute for Health and Care Excellence. (2011). *Common mental health problems: identification and pathways to care* (CG90). Retrieved from: https://www.nice.org.uk/guidance/cg123/chapter/1-guidance

NHS England. (2018). *NHS England Funding and Resource 2018/19: Supporting 'Next Steps for the NHS Five Year Forward View'.* Retrieved from: https://www.england.nhs.uk/wp-content/uploads/2018/03/Business-plan-2018-19-Annex-Update.pdf

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8,* 157-159.

*Pack, S., & Condren, E. (2014). An evaluation of group cognitive behaviour therapy for low self-esteem in primary care. *The Cognitive Behaviour Therapist, 7*. doi:10.1017/s1754470x14000051

*Pereira, J.-A., Barkham, M., Kellett, S., & Saxon, D. (2016). The role of practitioner resilience and mindfulness in effective practice: A practice-based feasibility study. *Administration and Policy in Mental Health and Mental Health Services Research, 44*(5), 691–704. doi:10.1007/s10488-016-0747-0

Perfect, D., Jackson, C., Pybis, J., & Hill, A. (2016). Choices of therapies in IAPT: An overview of the availability and client profile of step 3 therapies. *British Association for counselling and Psychotherapy.* Retrieved from: http://www.bacp.co.uk//docs/pdf/15641_napt_v6.pdf

*Pettit, S., Qureshi, A., Lee, W., Stirzaker, A., Gibson, A., Henley, W., & Byng, R. (2017). Variation in referral and access to new psychological therapy services by age: an empirical quantitative study. *British Journal of General Practice, 67*(660), e453–e459. doi:10.3399/bjgp17x691361

*Poots, A. J., Green, S. A., Honeybourne, E., Green, J., Woodcock, T., Barnes, R., & Bell, D. (2014). Improving mental health outcomes: achieving equity through quality improvement. *International Journal for Quality in Health Care, 26*(2), 198–204. doi:10.1093/intqhc/mzu005

Posternak, M. A., & Miller, I. (2001). Untreated short-term course of major depression: a meta-analysis of outcomes from studies using wait-list control groups. *Journal of Affective Disorders, 66*(2), 139-146.

*Pybis, J., Saxon, D., Hill, A., & Barkham, M. (2017). The comparative effectiveness and efficiency of cognitive behaviour therapy and generic counselling in the treatment of depression: evidence from the 2nd UK national audit of psychological therapies. *BMC Psychiatry, 17*(1). doi:10.1186/s12888-017-1370-7

*Radhakrishnan, M., Hammond, G., Jones, P. B., Watson, A., McMillan-Shields, F., & Lafortune, L. (2013). Cost of improving access to psychological therapies (IAPT) programme: An analysis of cost of session, treatment and recovery in selected primary care trusts in the East of England region. *Behaviour Research and Therapy, 51*(1)*,* 37-45. doi: 10.1016/j.brat.2012.10.001

Ranganathan, P., Pramesh, C.S., & Aggarwal, R. (2016). Common pitfalls in statistical analysis: Intention-to-treat versus per-protocol analysis. *Perspectives in Clinical Research, 7*(3). 144–146. doi: 10.4103/2229-3485.184823

*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. doi:10.1016/j.jad.2011.03.024

Richards, D. A., & Suckling, R. (2009). Improving access to psychological therapies: Phase IV prospective cohort study. *British Journal of Clinical Psychology, 48*(4), 377–396. doi:10.1348/014466509x405178

*Rimes, K. A., Broadbent, M., Holden, R., Rahman, Q., Hambrook, D., Hatch, S. L., & Wingrove, J. (2017). Comparison of treatment outcomes between lesbian, gay, bisexual

and heterosexual individuals receiving a primary care psychological intervention. *Behavioural and Cognitive Psychotherapy, 46*(3), 332–349. doi:10.1017/s1352465817000583

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638—64. doi:10.1037/0033-2909.86

Roth, A. D. & Fonagy, P. (2005) *What works for whom: A critical review of psychotherapy researc*h, New York: Guilford Press.

*Saunders, R., Cape, J., Fearon, P., & Pilling, S. (2016). Predicting treatment outcome in psychological treatment services by identifying latent profiles of patients. *Journal of Affective Disorders, 197*, 107–115. doi:10.1016/j.jad.2016.03.011

*Saxon, D., Firth, N., & Barkham, M. (2016). The relationship between therapist effects and therapy delivery factors: Therapy modality, dosage, and non-completion. *Administration and Policy in Mental Health and Mental Health Services Research, 44*(5), 705–715. doi:10.1007/s10488-016-0750-5

*Scott, M. J. (2018). Improving access to psychological therapies (IAPT) - The need for radical reform. *Journal of Health Psychology, 23*(9), 1136–1147. doi:10.1177/1359105318755264

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

Stewart, R. E., & Chambless, D. L. (2009). Cognitive–behavioral therapy for adult anxiety disorders in clinical practice: A meta-analysis of effectiveness studies. *Journal of Consulting and Clinical Psychology, 74*, 595-606. doi: org/10.1037/a0016032

The Mandate: A mandate from the Government to NHS England: April 2015 to March 2016. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/386221/NHS_England_Mandate.pdf

Thimm, J.C., & Antonsen, L. (2014). Effectiveness of cognitive behavioral group therapy for depression in routine practice. *BMC Psychiatry 14,* 292. doi: org/10.1186/s12888-014-0292-x

Turpin, G. (Ed.). (2010). IAPT good practice guide to using self-help materials. NMHDU/IAPT, 1-40.

Turpin, G., & Wheeler, S. (2011). *IAPT Supervision Guidance*. Retrieved from: https://www.uea.ac.uk/documents/246046/11919343/IAPT+Supervision+Guidance+2011.pdf/3a8f6c76-cea0-4f76-af78-e2f0a1f64c5e

*Vaillancourt, K., Manley, J., & McNulty, N. (2015). Why has our recovery rate dropped? An audit examining waiting times, starting scores and length of treatment in relation to recovery within an IAPT service. *The Cognitive Behaviour Therapist, 8,* 1-14. doi:10.1017/s1754470x15000148

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 48. https://www.jstatsoft.org/v36/i03/

Whiteford, H. A., Harris, M. G., McKeon, G., Baxter, A., Pennell, C., Barendregt, J. J., & Wang, J. (2013). Estimating remission from untreated major depression: a systematic review and meta-analysis. *Psychological Medicine, 43*(8), 1569-1585.

*Wright, D., & Abrahams, D. (2015). An investigation into the effectiveness of dynamic interpersonal therapy (DIT) as a treatment for depression and anxiety in IAPT. *Psychoanalytic Psychotherapy, 29*(2), 160–170. doi:10.1080/02668734.2015.1035740

*Wroe, A. L., Rennie, E. W., Gibbons, S., Hassy, A., & Chapman, J. E. (2014). IAPT and long term medical conditions: What can we offer? *Behavioural and Cognitive Psychotherapy, 43(*4), 412–425. doi:10.1017/s1352465813001227

*Young, A., Rogers, K., Davies, L., Pilling, M., Lovell, K., Pilling, S., … Oram, R. (2017). Evaluating the effectiveness and cost-effectiveness of British sign language improving access to psychological therapies: An exploratory study. *Health Services and Delivery Research, 5*(24), 1–196. doi:10.3310/hsdr05240

Appendices

1: Stepped care model from NICE CG123

Figure 1: Stepped-care model: a combined summary for common mental health disorders

Focus of the intervention	Nature of the intervention
Step 3: Persistent subthreshold depressive symptoms or mild to moderate depression that has not responded to a low-intensity intervention; initial presentation of moderate or severe depression; GAD with marked functional impairment or that has not responded to a low-intensity intervention; moderate to severe panic disorder; OCD with moderate or severe functional impairment; PTSD.	**Depression:** CBT, IPT, behavioural activation, behavioural couples therapy, counselling*, short-term psychodynamic psychotherapy*, antidepressants, combined interventions, collaborative care**, self-help groups. **GAD:** CBT, applied relaxation, drug treatment, combined interventions, self-help groups. **Panic disorder:** CBT, antidepressants, self-help groups. **OCD:** CBT (including ERP), antidepressants, combined interventions and case management, self-help groups. **PTSD:** Trauma-focused CBT, EMDR, drug treatment. All disorders: Support groups, befriending, rehabilitation programmes, educational and employment support services; referral for further assessment and interventions.
Step 2: Persistent subthreshold depressive symptoms or mild to moderate depression; GAD; mild to moderate panic disorder; mild to	**Depression:** Individual facilitated self-help, computerised CBT, structured physical activity, group-based peer support (self-

moderate OCD; PTSD (including people with mild to moderate PTSD).	help) programmes**, non-directive counselling delivered at home†, antidepressants, self-help groups. **GAD and panic disorder:** Individual non-facilitated and facilitated self-help, psychoeducational groups, self-help groups. **OCD:** Individual or group CBT (including ERP), self-help groups. **PTSD:** Trauma-focused CBT or EMDR. **All disorders:** Support groups, educational and employment support services; referral for further assessment and interventions.
Step 1: All disorders – known and suspected presentations of common mental health disorders.	**All disorders:** Identification, assessment, psychoeducation, active monitoring; referral for further assessment and interventions.

* Discuss with the person the uncertainty of the effectiveness of counselling and psychodynamic psychotherapy in treating depression.

** For people with depression and a chronic physical health problem.

† For women during pregnancy or the postnatal period.

CBT, cognitive behavioural therapy; ERP, exposure and response prevention; EMDR, eye movement desensitisation and reprocessing; GAD, generalised anxiety disorder; OCD, obsessive compulsive disorder; IPT, interpersonal therapy; PTSD, post-traumatic stress disorder.

2: Reasons for paper exclusion, PRISMA diagram

Titles/Abstracts	
Reasons for exclusion	*n* of studies excluded
Age of client group	94
Book	181
Book review	12
Comment/reply to comment	27
Commentary/Discussion/Position paper	29
Conference proceeding	1
Correction	5
Editorial	15
Guidelines	4
Interview	1
Language	4
Mix of IAPT and non-IAPT data	0
Not pre-post design	12
Not IAPT	807
Not IAPT UK	31
Other 'IAPT' research field	15
Protocol	63
Qualitative study	102
RCT/experiemental design	117

Review	168
Republished article	1
More comprehensive dataset	0

Full text

Reasons for exclusion	*n* of studies excluded
Age of client group	0
Book	0
Book review	0
Comment/reply to comment	1
Commentary/Discussion/Position paper	8
Conference proceeding	1
Correction	0
Editorial	2
Guidelines	1
Interview	0
Language	1
Mix of IAPT and non-IAPT data	4
Not pre-post design	35
Not IAPT	145
Not IAPT UK	14
Other 'IAPT' research field	1

Protocol	1
Qualitative study	0
RCT/experiemental design	5
Review	1
Republished article	0
More comprehensive dataset	1

3: CASP quality review table

Table A1: Overview of CASP quality ratings

	CASP questions													
Paper	Q1	Q2	Q3	Q4	Q5a	Q5b	Q6a	Q6b	Q7[1]	Q8[2]	Q9	Q10	Q11	Q12
Adamson et al (2015)	Y	Y	Y	Y	Y	Y	N	N			Y	C/T	C/T	Y
Ali et al (2014)	Y	Y	Y	Y	Y	N	N	N			Y	Y	Y	Y
Ali et al (2017)	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y	Y	Y
Baucom et al (2018)	Y	Y	Y	Y	Y	Y	N	N			Y	C/T	Y	Y
Binnie & Boden (2016)	Y	Y	Y	Y	Y	Y	N	N			Y	Y	Y	Y
Branson et al (2015)	Y	Y	Y	Y	C/T	N	N	N			Y	C/T	C/T	C/T
Branson et al (2018)	Y	Y	Y	Y	C/T	C/T	N	N			Y	C/T	C/T	C/T
Buckman et al (2018)	Y	Y	Y	Y	Y	Y	N	N			Y	C/T	Y	Y
Burns et al (2016)	Y	Y	Y	Y	Y	N	N	N			Y	Y	Y	Y
Chan et al (2014)	Y	C/T	N	N	C/T	Y	N	N			C/T	Y	Y	C/T
Cheston et al (2016)	Y	C/T	Y	N	N	N	N	N			N	N	C/T	N
Clark et al (2009)	Y	Y	Y	Y	Y	N	Y	Y			Y	Y	Y	Y
Clark et al (2018)	Y	Y	C/T	Y	C/T	Y	N	N			Y	C/T	C/T	Y
Clarkson et al (2016)	Y	Y	Y	Y	Y	Y	N	N			Y	C/T	Y	Y
Delgadillo et al (2014a)	Y	Y	Y	Y	C/T	Y	N	N			Y	Y	Y	Y

Delgadillo et al (2014b)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y
Delgadillo et al (2016a)	Y	Y	Y	C/T	C/T	Y	N	N	Y	Y	Y	Y
Delgadillo et al (2016b)	Y	Y	N	Y	Y	Y	N	N	Y	Y	C/T	Y
Delgadillo et al (2016c)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y
Delgadillo et al (2017a)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y
Delgadillo et al (2017b)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y
Delgadillo et al (2017c)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y
Elison et al (2017)	Y	Y	N	C/T	Y	Y	N	N	Y	Y	Y	Y
Firth et al (2015)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y
Giebel et al (2014)	Y	Y	Y	Y	Y	N	N	N	Y	C/T	Y	Y
Goddard et al (2015)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y
Green et al (2014)	Y	Y	Y	Y	C/T	Y	N	N	Y	Y	Y	C/T
Griffiths et al (2014)	Y	Y	Y	Y	N	N	N	N	C/T	Y	C/T	Y
Gyani et al (2013)	Y	Y	C/T	Y	Y	Y	N	N	Y	Y	Y	Y
Hammond et al (2012)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	C/T	Y
Highfield et al (2016)	Y	C/T	N	Y	N	N	N	N	N	C/T	C/T	N
Jolley et al (2015)	Y	C/T	Y	Y	Y	N	N	N	Y	C/T	C/T	C/T
Kellett et al (2016)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	C/T	Y
Kellett et al (2017)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y

Kenwright et al (2017)	Y	Y	Y	C/T	C/T	C/T	C/T	Y	Y	Y	Y	Y
Kuhn (2011)	Y	N	Y	Y	N	N	N	N	N	N	C/T	C/T
Lucock et al (2018)	Y	C/T	Y	Y	C/T	C/T	N	N	C/T	C/T	Y	Y
Luik et al (2017)	Y	Y	Y	Y	C/T	N	N	N	C/T	Y	C/T	Y
Matthew Prina et al (2014)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	C/T	Y
McDevitt-Petrovic et al (2018)	Y	Y	Y	Y	N	N	N	N	N	C/T	C/T	Y
Meadows et al (2017)	Y	Y	Y	Y	Y	C/T	Y	N	Y	Y	Y	Y
Methley et al (2016)	Y	Y	Y	Y	N	N	N	N	N	C/T	C/T	C/T
Mofrad et al (2014)	Y	C/T	Y	C/T	N	N	N	N	N	N	C/T	C/T
Morrison et al (2014)	Y	C/T	Y	Y	Y	N	N	N	Y	Y	Y	C/T
Murray (2017)	Y	C/T	Y	Y	N	N	N	N	C/T	C/T	C/T	C/T
Pack et al (2014)	Y	Y	Y	Y	C/T	N	N	C/T	C/T	Y	Y	C/T
Pereira et al (2016)	Y	C/T	Y	Y	Y	Y	N	N	Y	Y	C/T	C/T
Pettit et al (2017)	Y	Y	C/T	Y	C/T	N	N	N	Y	Y	Y	Y
Poots et al (2014)	Y	Y	C/T	C/T	C/T	N	C/T	C/T	N	C/T	C/T	C/T
Pybis et al (2017)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	C/T	Y
Radhakrishnan et al (2013)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y
Richards et al (2011)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	C/T
Rimes et al (2018)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	C/T

Saunders et al (2016)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	C/T	Y
Saxon et al (2017)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y
Scott (2018)	Y	C/T	C/T	Y	N	N	N	N	N	C/T	N	C/T
Vaillancourt et al (2015)	Y	Y	Y	Y	C/T	N	N	N	Y	C/T	Y	Y
Wright et al (2015)	Y	Y	Y	Y	C/T	C/T	N	N	C/T	Y	C/T	Y
Wroe et al (2015)	Y	Y	Y	Y	C/T	N	N	N	C/T	Y	Y	Y
Young et al (2017)	Y	Y	Y	Y	Y	N	N	N	C/T	C/T	C/T	Y

[1:] Please see summary table (Table 3) for overview of each article's reported findings.
[2:] Non-categorical score.

4:

Figure 2: Forest plot, PHQ-9 independent samples effect sizes and pooled effect size, Table format

	Study	N1	N2	d	ES	SV
1	Adamson et al 2015	314	313	0.85	0.849	0.007
2	Ali et al 2014	688	688	0.96	0.959	0.003
3	Baucom et al 2018	32	31	2.23	2.202	0.102
4	Burns et al 2016	207	207	0.88	0.878	0.011
5	Burns et al 2016	211	211	0.95	0.948	0.011
6	Chan et al 2014	50	50	0.77	0.764	0.043
7	Clark et al 2009	111	110	1.06	1.056	0.021
8	Clarkson et al 2016	253	252	0.62	0.619	0.008
9	Delgadillo et al 2014	1446	1445	0.81	0.810	0.001
10	Delgadillo et al 2016	2226	2225	0.59	0.590	0.001
11	Delgadillo et al 2017	254	253	0.22	0.220	0.008
12	Delgadillo et al 2017	69	68	0.26	0.259	0.029
13	Delgadillo et al 2017	56	55	0.05	0.050	0.036
14	Delgadillo et al 2017	67	67	0.50	0.497	0.031
15	Delgadillo et al 2017	1158	1158	0.10	0.100	0.002
16	Firth et al 2015	3056	3055	0.82	0.820	0.001
17	Giebel et al 2014	183	183	0.63	0.629	0.011
18	Green et al 2014	561	561	0.52	0.520	0.004
19	Highfield et al 2016	14	14	1.86	1.806	0.201
20	Kellett et al 2016	422	422	0.94	0.939	0.005
21	Kellett et al 2016	87	87	0.74	0.737	0.025
22	Kellett et al 2017	13	13	0.73	0.707	0.163
23	Kenwright et al 2017	24	24	0.98	0.964	0.093
24	Kenwright et al 2017	40	39	1.56	1.545	0.066
25	Kuhn 2011	26	26	1.28	1.261	0.092
26	Lucock et al 2018	6	5	0.50	0.457	0.376
27	Luik et al 2017	36	35	1.33	1.315	0.069
28	Meadows et al 2017	5	5	1.02	0.921	0.442
29	Morrison et al 2014	6	6	1.06	0.978	0.373
30	Pack et al 2014	25	25	0.79	0.778	0.086
31	Pybis et al 2017	11798	11797	0.94	0.940	0.000

32	Pybis et al 2017	4824	4824	0.95	0.950	0.000
33	Radhakrishnan et al 2013	4232	4232	0.72	0.720	0.001
34	Richards et al 2011	2092	2091	1.07	1.070	0.001
35	Wright et al 2015	12	12	0.44	0.425	0.170
36	Young et al 2017	49	49	1.02	1.012	0.046
37	Young et al 2017	183	183	1.15	1.148	0.013

Figure 4: Forest plot, GAD-7 independent samples effect sizes and pooled effect size, Table format

	Study	N1	N2	d	ES	SV
1	Adamson et al 2015	314	313	0.96	0.959	0.007
2	Ali et al 2014	688	688	1.00	0.999	0.003
3	Baucom et al 2018	32	31	1.80	1.778	0.089
4	Burns et al 2016	256	256	1.04	1.038	0.009
5	Burns et al 2016	245	245	1.05	1.048	0.009
6	Chan et al 2014	50	50	0.93	0.923	0.044
7	Clark et al 2009	111	110	1.26	1.256	0.022
8	Clarkson et al 2016	253	252	0.63	0.629	0.008
9	Delgadillo et al 2014	1446	1445	0.90	0.900	0.002
10	Delgadillo et al 2016	2226	2225	0.70	0.700	0.001
11	Delgadillo et al 2017	254	253	0.27	0.270	0.008
12	Delgadillo et al 2017	69	68	0.33	0.328	0.030
13	Delgadillo et al 2017	56	55	0.13	0.129	0.036
14	Delgadillo et al 2017	67	67	0.58	0.577	0.031
15	Delgadillo et al 2017	1158	1158	0.11	0.110	0.002
16	Firth et al 2015	3056	3055	0.90	0.900	0.001
17	Giebel et al 2014	183	183	0.63	0.629	0.011
18	Green et al 2014	561	561	0.55	0.550	0.004
19	Highfield et al 2016	14	14	1.88	1.825	0.202
20	Kellett et al 2016	422	422	0.93	0.929	0.005
21	Kellett et al 2016	87	87	0.86	0.856	0.025
22	Kellett et al 2017	13	13	0.49	0.475	0.158
23	Kenwright et al 2017	24	24	1.32	1.298	0.101
24	Kenwright et al 2017	40	39	1.39	1.376	0.063
25	Kuhn 2011	26	26	1.15	1.133	0.089

```
26         Lucock et al 2018    6    5 0.09 0.082 0.367
27           Luik et al 2017   36   35 0.79 0.781 0.061
28        Meadows et al 2017    5    5 1.33 1.201 0.472
29           Pack et al 2014   25   25 0.82 0.807 0.087
30 Radhakrishnan et al 2013 4232 4232 0.82 0.820 0.001
31       Richards et al 2011 2092 2091 1.04 1.040 0.001
32         Wright et al 2015   12   12 0.63 0.608 0.174
33          Young et al 2017   49   49 0.86 0.853 0.045
34          Young et al 2017  183  183 1.10 1.098 0.013
```

Table S2: Subgroup analysis of pre-post treatment effects in the typical study sample (*n*=8 atypical studies excluded)

Outcome	Variable	Subgroup	*k*	Effect size	95% CI	I^2(%)	Q	Diff between subgroups (*p*)
PHQ-9	Methodology	ITT	37	0.77	0.65 to 0.89	99%	5636.35***	**<.001****
		COM	20	1.04	0.96 to 1.12	99%	3121.97***	
	Study bias	Low	40	0.81	0.70 to 0.93	99%	5518.52***	.061
		Medium	14	0.94	0.83 to 1.06	99%	4029.45***	
		High	3	1.26	0.84 to 1.68	83%	11.95**	
	Primary condition	Mental health	45	0.93	0.86 to 1.01	99%	8985.62***	**.002****
		Physical health	10	0.43	0.13 to 0.74	94%	170.82***	
	Step of care	Step 2 only	15	0.80	0.68 to 0.93	97%	686.96***	.013*
		Step 3 only	6	1.05	0.90 to 1.20	94%	83.76***	
	Format	Individual	32	0.75	0.62 to 0.89	99%	5133.20***	.454
		Group	6	0.88	0.59 to 1.16	92%	69.92***	
GAD-7	Methodology	ITT	35	0.79	0.66 to 0.91	99%	5425.53***	**<.001****
		COM	17	1.07	0.99 to 1.15	99%	3964.94***	
	Study bias	Low	37	0.82	0.71 to 0.94	99%	3600.73***	**<.001****
		Medium	13	0.98	0.84 to 1.13	99%	5215.82***	
		High	2	1.51	1.49 to 1.53	0%	0	
	Primary condition	Mental health	40	0.96	0.88 to 1.04	99%	10732.73***	**.006****
		Physical health	10	0.50	0.19 to 0.82	94%	175.50***	

	Step of care	Step 2 only	14	0.88	0.74 to 1.03	98%	776.13***	.366
		Step 3 only	3	1.15	0.58 to 1.73	87%	16.21*	
	Format	Individual	27	0.75	0.58 to 0.91	99%	4746.63***	.283
		Group	6	0.91	0.66 to 1.16	93%	73.78**	
WSAS[1]	Methodology	ITT	19	0.56	0.49 to 0.63	98%	1224.64***	.363
		COM	2	0.46	0.27 to 0.66	0%	0.71	
	Study bias	Low	17	0.57	0.49 to 0.64	97%	772.84**	.438
		Medium	4	0.49	0.33 to 0.66	99%	810.27***	
	Step of care	Step 2 only	7	0.52	0.43 to 0.61	98%	432.12***	-
		Step 3 only	1	0.50	-	-	-	
	Format	Individual	9	0.51	0.42 to 0.59	99%	908.50	.568
		Group	2	0.48	0.45 to 0.51	0%	0.87	

*significant at $p < .05$ threshold; **significant at $p < .01$ threshold, ***significant at $p < .0001$ threshold, **between subgroup differences significant at Bonferroni adjusted $p<.01$ threshold for multiple testing**; [*1] Moderator analysis for 'primary condition' was not undertaken for the WSAS outcome measure as all studies included were deemed to be investigating mental health with none focusing purely on physical health. Abbreviations: *k:* number of comparisons per subgroup, CI: confidence interval, PHQ-9: Patient Health Questionnaire-9, GAD-7: Generalized Anxiety Disorder scale-7; WSAS; Work and social adjustment scale; ITT: Intention to treat; COM; completer. .

Table S3: Meta-regression analysis of pre-post treatment effects in the typical study sample (*n*=8 atypical studies excluded)

Outcome	Variable	Range and mean	*k*	B-coefficient	95% CI	SE	*p*
PHQ-9	Gender (% female)	(0 – 100%; M=59.5)	44	-0.01	-0.01 to 0.00	0.00	**<.001*** **
	Mean age	(31 – 49 years; M=39.8)	40	-0.01	-0.02 to 0.01	0.01	.483
	Mean intake score	(7.9 – 18.8; M=15.0)	51	0.03	0.01 to 0.05	0.01	.012*
	Mean number of sessions	(3 – 16 sessions; M=6.7)	35	0.04	0.02 to 0.07	0.01	**<.001*** **
GAD-7	Gender (% female)	(0 – 100%; M=59.1)	41	-0.00	-0.00 to 0.00	0.00	**.001** **
	Mean age	(31 – 49 years; M=39.7)	38	-0.01	-0.02 to 0.01	0.01	.481
	Mean intake score	(3.7 – 18.3; M=13.5)	45	0.07	0.04 to 0.09	0.01	**<.001*** **
	Mean number of sessions	(3 – 16 sessions: M=6.6)	31	0.04	0.01 to 0.07	0.01	**.002** **
WSAS	Gender (% female)	(0 – 100%; M=54.9)	19	-0.00	-0.00 to 0.00	0.00	.041*
	Mean age	(31 – 49 years; M=39.3)	17	0.00	-0.02 to 0.03	0.01	.738
	Mean intake score	(14.8 – 24.5; M=19.3)	20	0.02	-0.01 to 0.05	0.01	.127
	Mean number of sessions	(4 - 16 sessions; M=6.7)	16	0.03	-0.02 to 0.07	0.02	.201

Note: *significant at $p < .05$ threshold; **significant at $p < .01$ threshold, ***significant at $p < .0001$ threshold, **significant at Bonferroni adjusted *p*<.01 threshold for multiple testing**; Abbreviations: *k:* number of comparisons, CI: confidence interval; SE: standard error; M: mean; PHQ-9: Patient Health Questionnaire-9, GAD-7: Generalized Anxiety Disorder scale-7; WSAS; Work and social adjustment scale.

Part 2: Research Report

The effectiveness of brief cognitive analytic therapy for clients with depression and anxiety: a pilot study.

Abstract

Objectives: To investigate the effectiveness of brief cognitive analytic therapy (CAT) for anxiety and depression via a comparison with matched cases receiving cognitive behavioural therapy (CBT).

Methods: A pilot treatment comparison analysis of routinely collected sessional outcomes for CAT and CBT. The setting for the study was an Improving Access to Psychological Therapies (IAPT) service. The measures used were the PHQ-9, GAD-7 and WSAS. Propensity score matching (PSM) was used to create equivalent CAT and CBT samples that received up to 8-sessions of treatment. Longitudinal multilevel modelling (LMLM) was then used to investigate rates of symptomatic change between the two therapies.

Results: The primary longitudinal ML model was not significant, indicating few differences between the outcomes achieved by the two therapies (i.e. outcomes did not change as a function of the interaction between Time (i.e., sessional outcomes over time) and Sample (i.e., CAT vs CBT). This suggests that the trajectories of symptomatic and severity-level changes that occur over time were similar in both treatment models. Regression analyses also indicated no significant differences in the rate of change when comparing CAT and CBT. Effect size calculations indicate small between-group post-treatment effects.

Conclusions: The results are discussed with reference to the equivalence paradox for routinely delivered psychological therapies. Limitations of the study are labelled, and the design of future studies discussed. Whilst CAT shows some promise as a potential IAPT high intensity therapy, future research is clearly indicated.

Practitioner points:

- Practitioners under appropriate supervision could consider using a brief 8-session CAT when treating patients with anxiety and depression.

- The brief 8-session version of the model developed for use in Primary Care appears to hold some organisational promise.
- Practitioners need to ensure that brief CAT interventions still retain the theoretical integrity and so adhere to the reformulation, recognition and revision structure of the model.

Keywords: Improving Access to Psychological Therapies, IAPT, cognitive analytic therapy, CAT, cognitive behavioural therapy, CBT, practice-based, outcome measures, depression, anxiety, functioning.

1. Introduction

Mental health care features on the UK Government's agenda consistently (Layard, 2014). For example, in 2012, the Health and Social Care Act set out to achieve a 'parity of esteem' between mental and physical health services (Department of Health, 2012). The system for delivering evidence-based psychological therapies in the UK is informed by the National Institute for Health and Care Excellence (NICE) guidelines for the treatment of anxiety, depression, obsessive compulsive disorder (OCD) and post-traumatic stress disorder (PTSD). NICE guidance is informed by research and is hierarchical in nature (i.e., the most evidence generated by robust randomised controlled trials (RCTs) are given largest weightings, followed by less robust research evidence, such as cohort studies). Cognitive behavioural therapy (CBT) features in the NICE guidelines for the common mental health problems (i.e. anxiety disorders and depression) due to its robust evidence base in comparison to other psychotherapies. In routine services there has been a shift to delivery of CBT-informed therapies being manualised in order to create greater therapist adherence to the evidence base and to ensure evidence-based practice. Through continued rigorous research efforts, CBT as a therapeutic model has developed an extensive evidence base and became the initial cornerstone for the UK's national primary care mental health programme.

1.1 Improving Access to Psychological Therapies

The Improving Access to Psychological Therapies (IAPT) programme was developed in 2008, and its remit was to increase the access to evidenced based psychological therapies for people experiencing common mental health problems. It has continued to develop and expand since its first inception, and now is inclusive of all ages and covers a large range of conditions such as depression, generalised anxiety disorder, post-traumatic stress disorder, social anxiety disorder, obsessive-compulsive disorder, panic disorder, specific phobias, and other problems. It also covers the treatment of psychological distress associated with log-term

physical health conditions (Clark, Canvin, Green, Layard, Pilling & Janecka, 2018). IAPT is based on a stepped-care model (Figure 1) that delivers low and high intensity psychological interventions. As a first-line treatment to those accessing the service, briefer and less intensive therapies can be offered which typically range from 6-8 sessions and formats for the delivery of low-intensity interventions can include telephone, group, 1:1 or computerised treatment (The National Collaborating Centre for Mental Health, 2018). The stepped care aspect of the model means individuals can be 'stepped up' if they do not respond to low-intensity treatments or their risk changes, to more high-intensity therapies which are typically lengthier (up to 20 sessions).

There has been a more recent move in IAPT to increase the plurality of intervention at step 3. However, a recent study reported that of 114 services analysed, only one service offered the full range of the five NICE-recommended therapies for depression (i.e., CBT, IPT, couples therapy, counselling, psychodynamic psychotherapy), whilst 40 services offered four of the five recommended therapies (Perfect, Jackson, Pybis & Hill, 2016). The offer of differing therapies is based on the evidence that few differences are apparent between therapies offered in routine. For example, patient outcomes between CBT and counselling treatments have found the two to be generally comparable (e.g., Pybis et al., 2017); briefer forms of counselling for depression (2-7 sessions) have a larger impact on recovery rates, whilst the benefits of CBT are generally better following longer treatments (above 8 sessions) (Pybis et al., 2017). On average, treatment duration is around seven sessions in IAPT services (December 2018: Health and Social Care Information Centre, 2019).

IAPT services across England collect data from standardised patient-reported outcome measures on a session-by-session basis, which are used to monitor patients' response to treatment and to inform decisions about treatment planning. Typically, patients complete validated measures of depression, anxiety and functional impairment at the start of

each treatment session, and this is reviewed with their therapists as part of the therapy process. This information is then input into computerised data collection systems that enable services to assess their clinical performance. These data are anonymised and made publicly available each month and shared via NHS Digital. To put the scale of the outcome monitoring into context, the most recent monthly data available (November 2019) indicated that over 98,000 individuals began therapy within IAPT and the average number of sessions each individual will receive is 7 (NHS Digital, 2020). The IAPT evidence base consists of the publicly available outcomes, a large pool of evidence-based research studies (Wakefield et al., *in press*) and a small number of RCTs (Richards et al, 2016).

A standardised set of outcome measures, known as the 'minimum data set' (MDS), is utilised which matches the most common presenting problems. The MDS includes a measure of depression symptomatology (Patient Health Questionnaire (PHQ-9): Kroenke, Spitzer & Williams, 2001), a measure of anxiety symptomatology (Generalised anxiety disorder (GAD-7): Spitzer, Kroenke, Williams, & Lowe, 2006), and a measure of functional disability (Work and Adjustment Scale (WSAS): Mundt, Marks, Shear, & Greist, 2002). IAPT uses the 'recovery' indicator to assess the effectiveness of interventions, with 'reliable recovery' indicating that an individual patient moves below the caseness threshold on all measures at the end of treatment (National Collaborating Centre for Mental Health, 2019). A 50% recovery rate is set as a target for all IAPT services. This counts the number of patients who were above the clinical threshold on depression and/or anxiety at pre-treatment, with recovery occurring if a patient subsequently scores below the clinical threshold on depression and anxiety at the end of treatment. This 50% figure is drawn from the results of the clinical trials that make up the evidence base that forms the NICE guidelines. The latest figures show that this recovery rate is being met with the November 2019 figures indicating a 50.6% recovery rate (NHS Digital, 2020). 'Disorder-specific' measures are also intended to be used

when relevant to the individual client and their presenting difficulty (National IAPT guidance, 2010), however these appear to be collected less often than may be necessary (Wakefield, Kellett, Simmonds-Buckley, Stockton, Bradbury & Delgadillo, *in press*).

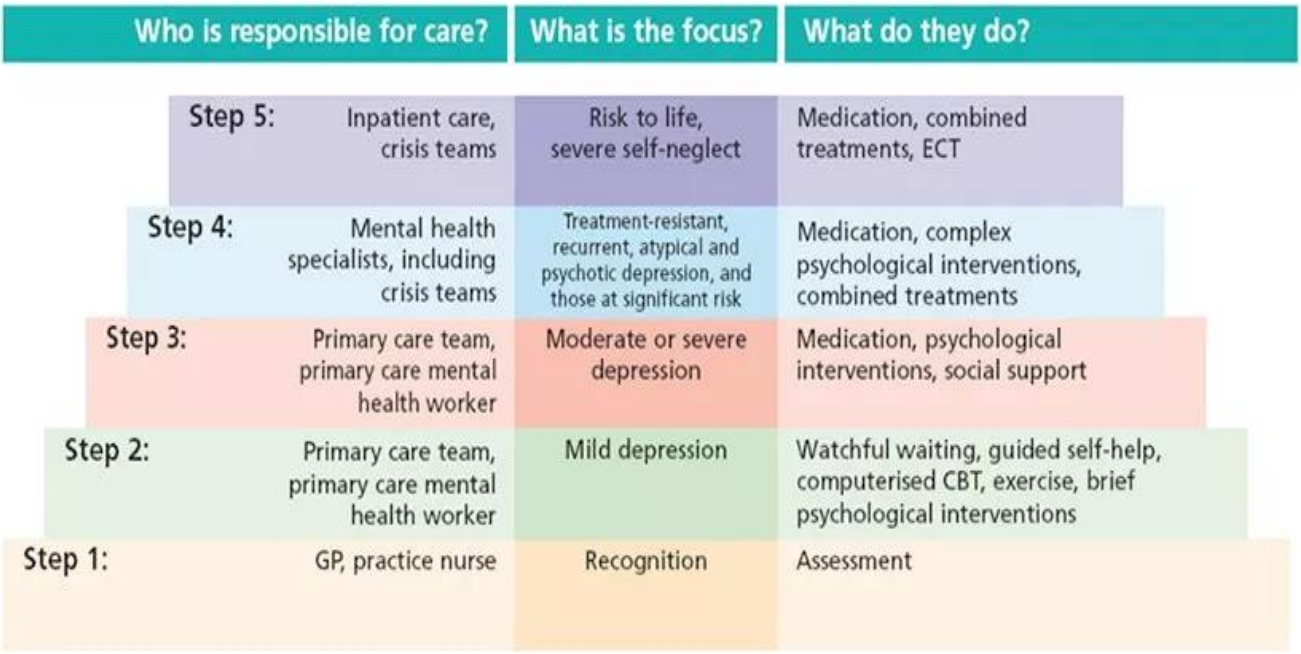

Figure 1: The proposed stepped-care model for IAPT services.

1.2 Cognitive analytic therapy

Cognitive analytic therapy (CAT) was first proposed as a bone fide psychotherapy in 1985 (Ryle & Kerr, 2002). Anthony Ryle pioneered this therapy during a time when the NHS was pressured both for time and finances – making CAT as relevant today in the current NHS climate as it was during its inception. CAT is a time-limited, relational, collaborative and integrative therapy which draws from both cognitive and analytic theories (Ryle, Kellett, Hepple & Calvert 2014). Specifically, it integrates analytic theory relating to object relations (Ogden, 1983) with cognitive theory of personal constructs (Kelly, 1956). Self to self, self to other, and other to self relationship patterns are summarised as 'reciprocal roles' (RRs) and are a reflection of the analytic aspect of the theory, with resultant procedural sequences reflecting the cognitive element (Ryle & Kellett, 2019). Generally, CAT is offered over 8, 16 or 24 sessions plus a follow-up period (Ryle & Kerr, 2002) and follows a reformulation, recognition, and revision structure (Ryle & Kellett, 2019).

CAT has become popular with therapists and has grown and developed notably from its first inception over 30-years ago to now be delivered internationally (Ryle et al, 2014). CAT is largely used with clients with severe and debilitating difficulties, such as those diagnosed with 'personality disorders' (Kellett, Bennett, Ryle & Thake, 2013; Clarke, Thomas & James, 2013). However, more recently CAT has expanded and research studies into the use of CAT with other clinical populations have gained momentum. For example, CAT has shown beneficial effects with clients struggling with eating disorders (Treasure et al., 1995), morbid jealousy (Kellett & Totterdell, 1995), and has also been utilised in physical health settings with clients following acquired brain injuries (Yeates et al, 2008). CAT has been criticised for developing an evidence base with complex psychological disorders at the expense of the common mental health problems (Hammonds, Simmonds-Buckley & Kellett, 2020). There is a small evidence base for the effectiveness of CAT with anxiety and depression, with the studies being limited by the completion of simple pre and pest designs (Calvert & Kellett, 2014). A recent meta-analysis of the CAT evidence base (Kellett et al, 2020) included 28 studies (k=10 RCTS and k=18 PBE) studies showing significant reductions in depression symptoms (ES=1.05, Z=9.17; p<.001).

In clinical practice, whilst the research element of CAT is somewhat lacking, the model is continuing to evolve and increase in its applicability to different situations. For example, a recent dismantling CAT trial investigated a brief 8-session CAT within IAPT for clients with depression (randomised to either reformulation letter or no reformulation letter) and found significant reductions to depression at both end of treatment and follow-up, with associated improvements in levels of functioning (Kellett et al, 2018). Due to the impetus to increase the range of therapies offered at step 3 of IAPT services, it is particularly worthwhile to investigate CAT in relation to 'treatment as usual' (i.e., CBT). This is particularly important given that CAT is a time-limited and structured psychotherapy and that could

potentially fit in well with the IAPT programme. Other treatments that have a psychodynamic orientation origin and a time-limited structure, such as Interpersonal Psychotherapy (IPT), have been implemented into the available treatments for patients and thus the current study can add to this growing list of potential available treatment options for patients accessing IAPT services (e.g., Wright & Abrahams, 2015). The outcomes achieved by the 8-session version of the model are particularly ripe for investigation, as the average number of sessions in IAPT is seven (NHS Digital, 2020) and therefore the design and evaluation of a therapy to fit that evidence is indicated. Therefore, a practice-based evidence pilot study is reported here that investigated the effectiveness of brief 8-session CAT in comparison to well-matched CBT treatment cases.

1.3 Aims

The proposed study aimed to analyse practice-based evidence of CAT and high intensity CBT for well-matched patients attending at an IAPT service for common mental health difficulties. Growth curve models using sessional outcome measures will be defined and compared to investigate the trajectories of change across the two therapies. The study also sought also to define the effect sizes achieved in CAT and CBT. We did not expect any statistically significant differences between treatments, based on the prevalence of the equivalence paradox in routine practice (i.e., bone fide psychotherapies tend to produce similar post-treatment outcomes).

1.4 Hypotheses

i. CAT and CBT have similar change trajectories over sessional time (primary analysis).

ii. Effect sizes will not differ between CAT and CBT.

iii. The acceptability of CAT (i.e., the dropout rate) will be comparable to CBT.

2. Methodology

2.1 Study design

A retrospective, single-site pilot study from a routine IAPT service comparing clinical outcomes on pre-existing datasets. In this study, datasets have been recorded within the IAPT service via routine clinical practice during step 3 interventions (CAT and CBT). The study utilised IAPT sessional outcome data to investigate trajectories of change in this data. As is routine in IAPT, clients attending the services therefore completed relevant sessional measures that allowed to track change for individuals during therapy.

2.2 Setting

The research dataset was drawn from individuals who attended an IAPT service based at a site within the South of England. The IAPT service has stepped-care and so offers evidence-based treatments with qualified professionals trained to deliver step-3 (high intensity) interventions, based on NICE recommendations, for individuals referred with common mental health problems. The service was also piloting the use of an 8-session CAT intervention.

2.3.1 8-session CAT model description

A description and outline of the 8-session CAT model has been detailed elsewhere (White & Hepple, *in submission*). CAT was designed to be offered in the service where the presenting difficulties raised by patients contained ‘complex relational problems, personality disorder traits or histories of adverse childhood experience’. The reformulation letter aspect of CAT was removed due to implications on therapist time and evidence from the Kellett et al (2018) dismantling study showing that narrative reformulations does not improve outcomes. The therapy still adhered to the reformulation, recognition and revision structure of the model (Ryle & Kellett, 2019), as target problem procedures (TPPs) were identified early in

treatment, a sequential diagrammatic reformulation (SDR) was completed by mid-treatment and goodbye letters were shared between patients and therapists at the end of treatment. The tools of CAT (e.g. the psychotherapy file and TPP rating sheets) were also included. All the CAT treatments were supervised by an ACAT accredited CAT psychotherapist and supervisor. A brief description of the stages of the 8-session CAT model is outlined:

Early sessions (1-3): The therapist introduces the premises around CAT, particularly the relational nature of the therapy. Both therapist and patient collaborate to understand relationship patterns from the patients past and collaborate to identify 'target problems' (TPs) that the patient wants to focus on during the therapy and the TPPs that maintain the target problems in the current day and that may disable or interfere with the effectiveness of the therapy. The therapist also draws on their experience of the patient in the session (and through use of supervision) to start to build the case formulation and includes information from the Psychotherapy File. The central aim of the initial phase of treatment is to conceptualize how past relationships may be impacting on present relationships, including the therapeutic relationship via identification and patient monitoring of TPs and TPPs.

Mid-session (4): The introduction of the concept of reciprocal roles is introduced here and mapped out visually by the therapist to the patient in the form of a sequential diagrammatic reformulation (SDR), which visually summarises the key reciprocal roles and the procedures that link them. In collaboration, the therapist and patient map therefore map out the reciprocal roles that summarize self-to-self, self-to-other and other-to-self relating that are based in early experiences and also the current day procedures that maintain the reciprocal roles. The patient and therapist also use the SDR to enable the patient to recognize when reciprocal roles and procedures are being enacted in close relationships, including the therapeutic relationship via enactment analysis and also rupture repair efforts.

Later sessions (5-7): This is the phase of CAT which is focused on facilitating change in reciprocal roles and associated procedures. Analysis of enactments and also ruptures within the therapeutic relationship continue to be part of the change process. Therapists also encourages the patient to create 'exits' on the SDR to change reciprocal roles and disrupt procedures ad these are put into place in the patient's life outside of therapy. 'Exits' in CAT are catholic and as long as they are based in the SDR then the therapist is free to work using any change method, as long as this is within the patient's zone of proximal development (Ryle & Kellett, 2019). The goodbye letter is planned at session 7 and these are shared at session 8. There is a worksheet to provide patients with advice on what might be helpful to reflect on and therefore contains relational prompts.

Final session (8): The therapist and patient their share goodbye letters in this final session, capturing the relational understanding that has been built-up collaboratively during the therapy, labelling what are effective exits and how these can be maintained and also naming relapse prevention strategies (Ryle & Kellett, 2019).

2.3.2 CBT model description

IAPT CBT interventions are highly standardised and delivered following the treatment protocols drawn from randomised controlled trials (RCTs) forming the basis of the depression and anxiety NICE guidelines (CG 90; CG 113). CBT treatments are also underpinned by the competencies framework for the cognitive or behavioural treatment of anxiety and depression (Roth & Pilling, 2008). IAPT CBT interventions are typically 16-20 sessions. There are two treatment protocols for depression, one is cognitive (Beck, Rush, Shaw, & Emery, 1979) and one is behavioural (Martell, Addis, & Jacobsen, 2001). The protocols for the range of anxiety disorders, OCD, BDD and PTSD are all cognitive-

behavioural. All the protocols share the common ingredients of a disorder-specific formulation, within-session change methods, homework and relapse prevention.

2.4 Sample selection and characteristics

Ethical approval was sought to access the dataset (South of England: 19/HRA/0025, 16th May 2018), including comparator CBT treatments (see Appendix 4). The following inclusion and exclusion criteria were used:

Inclusion Criteria

- Adult patients who accessed high intensity CBT or 8-session CAT in the participating IAPT site;
- CAT was in particular offered to patients with difficulties that are recognised as experiencing common mental health problems (i.e., depression and/or anxiety) that reflect complex relational problems, personality disorder traits or histories of adverse childhood experience;
- Patients either completed therapy or dropped out of therapy.

Exclusion Criteria

- Patients who completed a treatment modality other than CAT or high intensity CBT;
- Patients that had a CAT treatment that lasted longer than 8-sessions;
- Patients who did not complete any treatment sessions (i.e., only completed assessment session).

The CONSORT flow diagram in Figure 2 outlines how those who met inclusion into the study were filtered into the final datasets that were analysed for this study.

South of England dataset

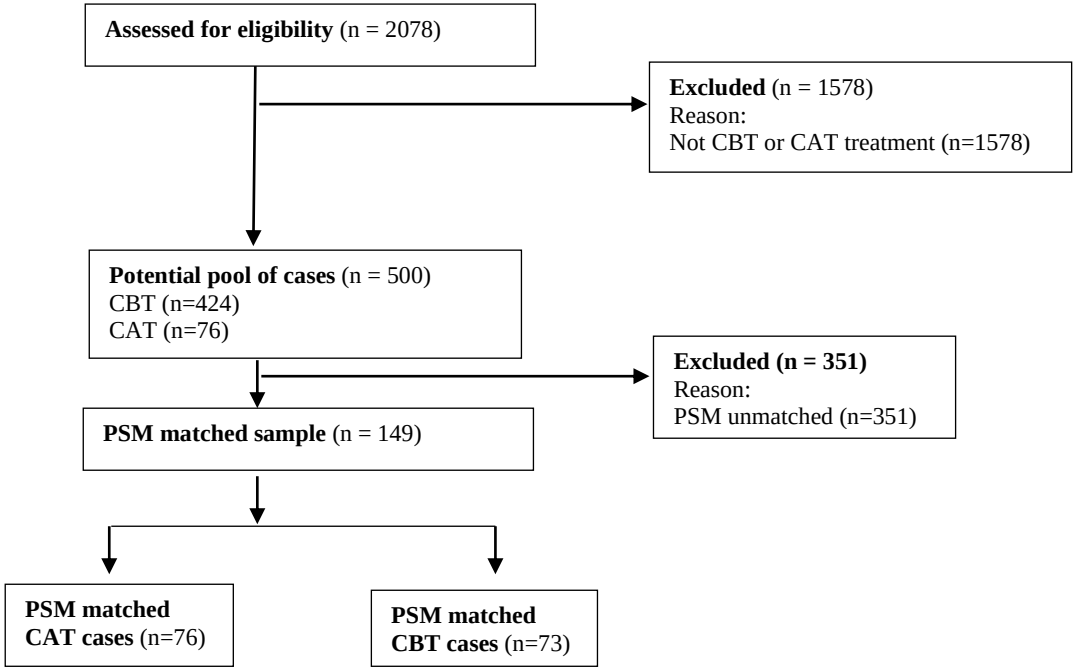

Figure 2: Sample selection diagram

2.5 Outcome measures:

Outcome measures were completed at an initial appointment, as well as at each treatment session. Measures are collected routinely within all IAPT services and therefore the measures are pre-determined (IAPT minimum dataset described below). Details, including psychometric evidence, are outlined below:

Generalised Anxiety Disorder (GAD-7) (Spitzer, Kroenke, Williams, & Lowe, 2006) - The GAD-7 measures anxiety symptom severity. The 7-items are scored on a Likert scale of 0-3 (score range 0-21; higher scores indicating worse symptoms). Psychometric properties have reported good validity and reliability using a cut-off ≥8 to identify clinically significant anxiety disorders, with sensitivity and specificity values of 89% and 82%, respectively (Spitzer et al., 2006).

Patient Health Questionnaire (PHQ-9) (Kroenke, Spitzer & Williams, 2001) - The PHQ-9 measures depression symptom severity. The 9-items are scored on a Likert scale of 0-3 (score range is 0-27; higher scores indicating worse symptoms). Psychometric properties have been reported and research has indicated excellent validity and reliability (Kroenke et al., 2001) using a clinical cut-off ≥10, with sensitivity and specificity values of .77 and .94, respectively (Wittkampf et al., 2007).

Work and Social Adjustment Scale (WSAS) (Mundt, Marks, Shear, & Greist, 2002) - The WSAS measures the impact of MH on a person's ability to function in the areas of work, home management, social leisure activities, private leisure activities and close relationships. The 5-items are scored on a Likert scale of 0-8 (score range is 0-40; higher scores indicating worse symptoms). Psychometric properties have shown the WSAS to have good internal, temporal reliability and is sensitive to differences in disorder severity and responsivity to treatment (Purdie, Kellett, & Bickerstaffe, 2012).

2.6 Data preparation

Data was prepared using the Statistical Package for Social Sciences (SPSS) v.25 The data cleaning process entailed several steps. The initial cleaning required identification of patients who had accessed only those included treatment modalities (i.e., CAT or CBT delivered at step 3). New variables into the data were created, which could identify first and last treatment session, and a new 'Sample' variable was added to classify which treatment each individual patient had accessed for the upcoming analysis.

Following data cleaning, propensity score matching (PSM) was applied. This is a method that enables case-control matching of participants within each treatment based on available characteristics. The PSM method is a valid statistical tool that mimics the randomisation procedure used in RCTs to match groups as equivalent as possible on important factors (Beal & Kupzyk, 2014). Using the CAT group as the reference group, the PSM procedure was set-up to match with CBT cases based on age, gender and baseline scores (PHQ-9, GAD-7, WSAS).

2.6.1 Missing Data

An intention to treat analysis was completed therefore all data from participants who commenced therapy was utilised. As the intention was to utilise longitudinal multilevel modelling analyses, missing data did not need to be imputed, as this approach is capable of modelling growth curves over missing data-points (Singer & Willett, 2003).

2.7 Statistical Analyses

2.7.1 Baseline comparisons of sample characteristics

Data was analysed using the Statistical Package for Social Science (SPSS) v.25. Analysis of the data firstly investigated baseline characteristics of each of the samples prior to

the PSM procedure and then following matching of samples. Here, non-parametric tests (Mann Whitney-U, Chi-Square) were employed to compare any differences between groups, including those who have been included and excluded from the final sample.

2.7.2 Primary analysis

The primary analysis consisted of analysing the research dataset (n=149) with the final PSM treatment modality samples (CAT n=76; CBT n=73). PSM is a statistical method that enables case-control matching in such a way that balances important baseline (pre-treatment) characteristics, as a way to artificially mimic the balancing of covariates that is achieved in randomised controlled trials (Beal & Kupzyk, 2014; Rosenbaum & Rubin, 1983). This matching procedure was based on a logistic regression predicting CAT group membership, entering all available demographic and pre-treatment clinical measures as predictors (age, gender, PHQ-9, GAD-7, WSAS), using a one-to-one nearest neighbours approach with a conservative tolerance level (caliper = 0.2) specified a priori, and allowing replacement to maximize matching precision. This one-to-one matching process should –in theory– result in two balanced treatment samples (CBT; CAT) with identical sample sizes. However, obtaining identical sample sizes depends on the extent to which an exact match (based on the available baseline characteristics) is found for every single case in the sample. In practice, sometimes this procedure yields close but not identical sample sizes in the matched samples (Rosenbaum & Rubin, 1983).

The primary analysis was undertaken using longitudinal multilevel modelling (MLM) to investigate any differences in the treatment trajectories using the sessional data from the outcome measures (PHQ-9, GAD-7, WSAS). Any patient that had received treatment within the IAPT service that lasted more than 8-sessions were analysed in a way to ensure only the initial 8-sessions of CAT were evaluated. The CBT comparison group contained treatment

contacts of up to sixteen sessions. Therefore, the CBT comparator group contained therapies that had been longer than 8-sessions, but the first 8-sessions were used to create the comparisons with CAT in the subsequent growth curve modelling. Both and CAT and the CBT comparators groups therefore could contain therapies of less than 8-sessions, due to patient drop-out. This meant that a direct comparison of the effectiveness of the two therapies could be undertaken, with this method also eliminating any potential differences between the two treatment groups due to factors regarding treatment length. Each outcome measure was analysed in separate longitudinal MLM models. A two-level model was created which included session-by-session outcome measure scores (level 1) nested within cases (level 2). The model had both fixed and random effects (random intercepts and random slopes for time), with an unstructured covariance structure. The analysis followed conventional guidelines for multilevel modelling, which build regression models in a series of steps, progressing from unconditional to fully adjusted (conditional) models that are optimised for goodness-of-fit (Singer & Willett, 2003). In the first part of the analysis the initial task was to develop unconditional models which fit different time trends across the data to find the closest fit. The trends used were: linear, quadratic, cubic and log-linear. The model with the best goodness-of-fit was determined by the -2*loglikelihood ratio test, and the equation for this is as follows:

$$\mathbf{X^2\ Change} = (-2*\text{Log Likelihood}_{old}) \mid (-2\text{Log Likelihood}_{new})$$

$$\mathbf{Df\ Change} = \text{Number of parameters}_{old} - \text{Number of parameters}_{new}$$

(Field, 2009)

Following this, a conditional growth curve model was built, adding ‘Sample’ (i.e., treatment modality group; CAT or CBT) as a predictor to test how outcomes changed between groups over time: Time*Sample interaction.

Growth curve analyses (or, growth curve modelling: GCM) were used to compare patterns of mean-level changes longitudinally. Advantages of using GCM include that it is a robust enough method to allow for missing data, unequal sample sizes and differing time intervals between data points (e.g., Shek & Ma, 2011). GCM analyses data at both a within-subjects (level 1) and between-subjects level (level 2). In the current data set, the within-subjects level included sessional data (level 1) nested within each individual patient (level 2); therefore the 'Sample' variable was conceptualised as a level-2 (between-subjects) predictor.

2.7.3 Secondary analyses

Between-groups effect sizes were calculated for all outcome measures comparing post-treatment means between the CAT and CBT treatments. This was completed within the main PSM sample. Effect sizes are reported using Cohen's *d* (Cohen, 1988), with the calculation for this being:

$$d = (M_1 - M_2) / SD_{pooled}$$

$$\text{where, } SD_{pooled} = \sqrt{((SD_1^2 + SD_2^2) / 2)}$$

Cohen's power primer definitions were used to interpret the effects sizes found in the between group differences: 'small' ($d = 0.2$), 'medium' ($d = 0.5$) or 'large' ($d = 0.8$) (Cohen, 1988).

Secondary analyses also investigated any differences in the number of treatment sessions received in each of the treatment groups. Furthermore, measures of clinically significant change (CSC) and reliable change (RC) was completed to highlight any changes within groups of change from pre- to post-treatment. The CSC would indicate whether an individual had made meaningful and significant change from a point of being within a 'clinical population range' (i.e., dysfunctional) to being within a 'non-clinical range' (i.e., functional) (Jacobson, Follette, & Ravenstorf, 1984). The RC would indicate whether an

individual's scores had changed more than would be determined by that of error inherent within the measure (Jacobson & Truax, 1991). These comparisons were completed on the PHQ-9 and GAD-7 measure only.

3. Results

3.1 Sample inclusion

Datasets were prepared and PSM methodology applied to generate the final sample for the analysis as previously outlined. The route to sample inclusion is shown diagrammatically in Figure 2.

3.2 Sample characteristics

3.2.1 Baseline measures and demographics

Table 1 summarises the demographics, including baseline outcome measures, that were eligible for inclusion before PSM procedures (CAT n = 76; CBT n = 424). Independent samples Mann Whitney-U and chi-square tests were used to compare all those who had completed CBT or CAT treatments on these characteristics. There were no significant age differences between those who were treated using CBT and those who were treated using CAT ($p=0.119$; CBT M = 40.45 (15.08), CAT M = 42.58 (11.96), and gender was equally distributed between the therapies ($p=0.879$; CBT 64% = female, CAT 63% = female). Comparing the number of treatment sessions attended between the therapies resulted in a non-significant finding ($p=0.071$; CBT M = 9.19 (6.90), CAT M = 9.80 (5.59). Baseline outcome measures (compared using Mann Whitney-U tests), indicated significant differences between the distribution of scores on each measure ($p<0.05$) with the CAT cases tending to report greater levels of depression (CAT M = 18.12 (5.08); CBT M = 16.42 (5.92)), anxiety (CAT M = 14.85 (4.39); CBT M = 13.34 (5.37)) and functional impairment (CAT M = 24.52 (8.72); CBT M = 21.51 (9.32)).

Table 1: Characteristics of the sample within the dataset

Characteristics	All cases	All CAT cases	All CBT cases
n	2078	76	424
Age: ***mean (SD; range)***	41.77 (15.07; 18-93)	42.58 (11.96; 18-75)	40.45 (15.08; 18-82)
Gender: ***n*** **(%)**	Male: 685 (33) Female: 1392 (67) *n=1 missing*	Male: 28 (37) Female: 48 (63)	Male: 152 (36) Female: 271 (64) *n=1 missing*
PHQ-9_session 1: ***mean (SD; range)***	*n=465* 16.69 (5.82; 0-27)	*n=73* 18.12 (5.08; 6-27)	*n=392* 16.42 (5.92; 0.27)
GAD-7_session 1: ***mean (SD; range)***	*n=465* 13.57 (5.25; 0-21)	*n=73* 14.85 (4.39; 0-21)	*n=392* 13.34 (5.37; 0.21)
WSAS_session 1: ***mean (SD; range)***	*n=454* 22.00 (0-40)	*n=73* 24.52 (8.72; 1-40)	*n=381* 21.51 (9.32; 0-40)
No. of sessions: ***mean (SD; range)***	6.86 (5.17; 2-40)	9.80 (5.59; 2-28)	9.19 (6.90; 2-40)

3.2.2 Propensity score matching samples

To assess the accuracy of the propensity score matching (PSM) procedure, the baseline characteristics of the filtered PSM sample was compared using Mann Whitney-U and chi-square tests. Table 2 summarises the demographics, including baseline outcome measures, for the PSM sample.

3.2.3.1 PSM dataset

Baseline demographic factors, age and gender, showed no significant differences between those who received CAT or CBT treatment in this PSM subsample. The first session PHQ-9 and GAD-7 outcome measures were also matched sufficiently. However, significant minimal differences were evident following the matching in the baseline WSAS scores between groups ($t(144) = 2.082$, $p<0.05$; CBT M = 21.63 (8.04), CAT M = 24.52 (8.72)). 'Caseness', i.e., whether an individual scored above the clinical cut-off range on outcome measures is also highlighted for the PHQ-9 (clinical range = ≥ 10) and GAD-7 (clinical range = ≥ 8) measures ($p=0.092$).

Table 2: Characteristics of the PSM samples

Characteristics		PSM CAT	PSM CBT	*p-level*[*]
n		76	73	
Age: *mean (SD; range)*		42.58 (11.96; 18-75)	39.88 (13.68; 19-71)	*p=.188*
Gender: *n*		Male: 28 Female: 78	Male: 28 Female: 45	*p=.849*
PHQ-9_session 1: *mean (SD; range)*		18.12 (5.08; 6-27)	16.70 (5.50; 0-27)	*p=.151*
GAD-7_session 1: *mean (SD; range)*		14.85 (4.39; 0-21)	14.03 (5.15; 0-23)	*p=.403*
WSAS_session 1: *mean (SD; range)*		24.52 (8.72; 1-40)	21.63 (8.04; 0-40)	***p<0.05***
PHQ-9_severity ranges: *n (%)*	Minimal-none (0-4) Mild (5-9) Moderate (10-14) Moderate severe (15-19) Severe (20-27)	0 (0) 4 (5) 11 (15) 26 (36) 32 (44)	1 (1) 8 (11) 11 (15) 28 (38) 24 (33)	
GAD-7_severity ranges: *n (%)*	Minimal-none (0-4) Mild (5-9) Moderate (10-14) Severe (15+)	1 (1) 7 (10) 25 (34) 40 (55)	1 (1) 14 (19) 19 (26) 39 (53)	
PHQ-9_casesness start: *no (%)*		69 (95)	63 (86)	*p=.0.93*
GAD-7_casesness start: *no (%)*		70 (99)	62 (85)	***p<0.05***

Note: * Mann Whitney-U conducted except for Gender and PHQ-9/GAD-7 caseness data where Chi square conducted; p-level compares PSM CAT cases with PSM CBT cases.

3.4 Primary analysis

The primary analysis was conducted on the PSM sample (n=149; CAT=76, CBT=73) and was separated across outcome measures (Table 2). Four unconditional models based on different trends (linear, quadratic, cubic, loglinear) were initially built and -2*loglikelihood (-2LL) change was compared to find the model with the best fit. Following this, growth curve modelling was used, and a conditional model was built including the Sample variable (i.e., which treatment modality was received; CAT or CBT) as a predictor.

From the PHQ-9 primary analysis, the Sample*Time interaction term (primary hypothesis test) was not statistically significant (B = 0.33, SE = 0.22, p=0.138) indicating that trajectories of change over time in depression symptoms between treatment modality groups (CAT, CBT) did not significantly differ (i.e., the different treatments did not have differential effects on symptoms over time). Thus, patients receiving either CAT or CBT had on average similar PHQ-9 scores and change profiles over treatment time. The effect of Time was significant (B = 0.065, SE = 0.028, p<0.019) indicating that change does occur over sessions. Baseline severity was not significantly different between the two samples (B = -1.70, SE = 0.90, p=0.061). Symptomatic change on the PHQ-9 followed a quadratic trend (see Figure 3).

From the GAD-7 primary analysis, the Sample*Time interaction term was not significant (B = 0.11, SE = 0.19, p=0.566) indicating that trajectories of change over time on anxiety symptomatology between therapies were not significantly different. Thus, the patients receiving either CAT or CBT had similar GAD-7 change scores and outcome trajectories over time. There was a significant effect of a linear trend for Time (p<0.0001) but not of Sample (p=0.537), indicating that whilst GAD-7 outcome scores did significantly change as a function of Time, there was no significant differences in initial severity on the GAD-7 between CAT and CBT.

Symptomatic change on this measure of anxiety followed a linear trend and this can be viewed visually (Figure 4). From the visual representation, the two treatments appear to follow a similar linear trajectory from session 1 to session 8.

From the WSAS primary analysis, the Sample*Time interaction term was again not significant (B = 0.48, SE = 0.31, p=0.121) indicating that trajectories of change over time on levels of functioning between CAT ad CBT were not significantly different (i.e., functioning changes occurred similarly over time between the therapies). Thus, the patients receiving either CAT or CBT had on average similar WSAS change scores over time. There was a significant effect of a linear Time trend ($p<0.0001$) but not of Sample ($p=0.076$), indicating that whilst WSAS scores did significantly change as a function of Time, there was no significant differences in initial severity on this measure of functioning between groups. Symptomatic change on the WSAS followed a linear trend and this can be viewed visually in Figure 5. From the visual representation, the linear model appears similar to that of the GAD-7 analysis in that the two treatments appear to follow a similar trajectory from session 1 to session 8.

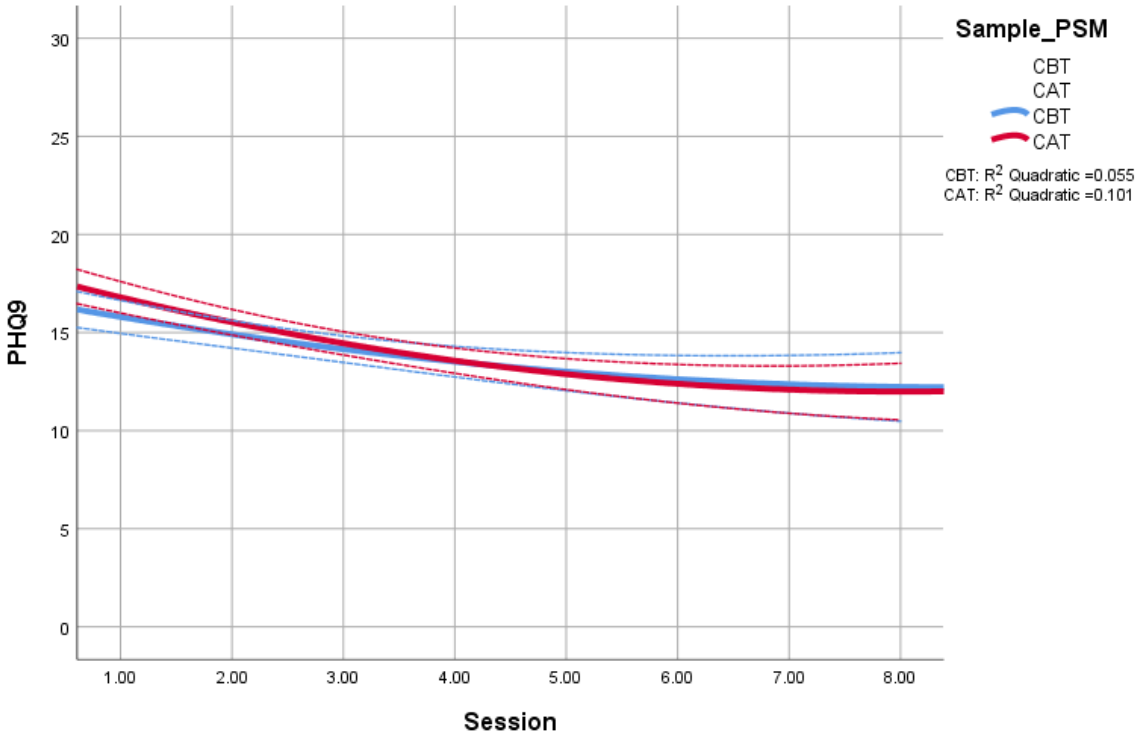

Figure 3: Sample*Time growth curve model showing a quadratic trend in the PHQ-9 data

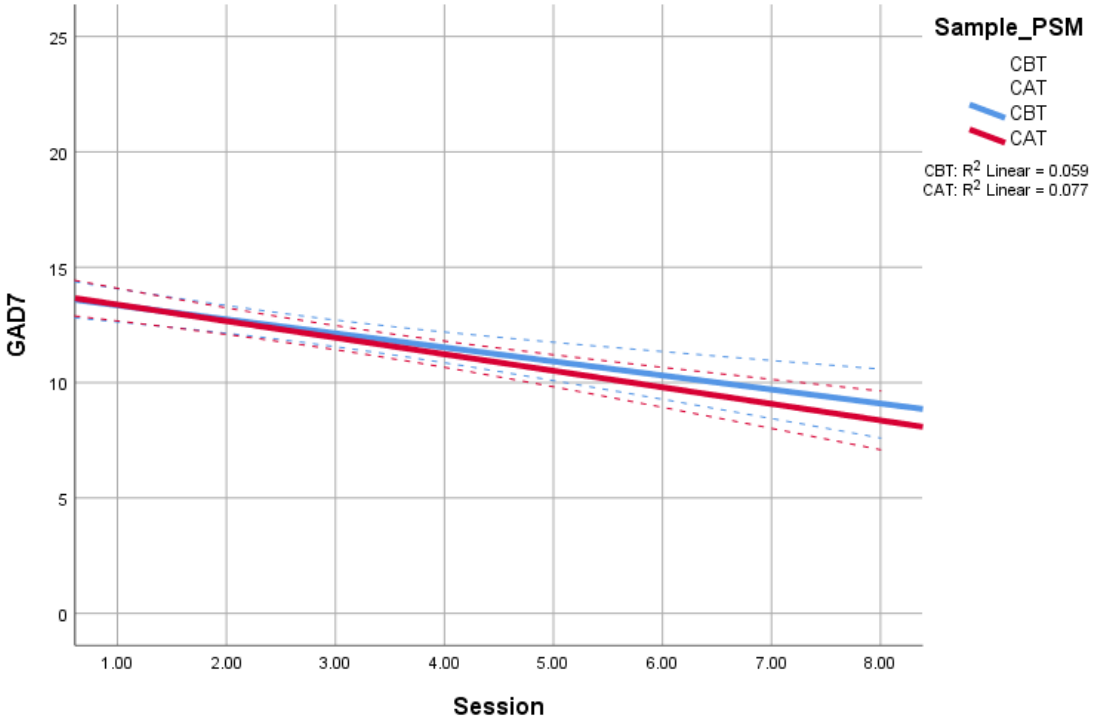

Figure 4: Sample*Time growth curve model showing a linear trend in the GAD-7 data

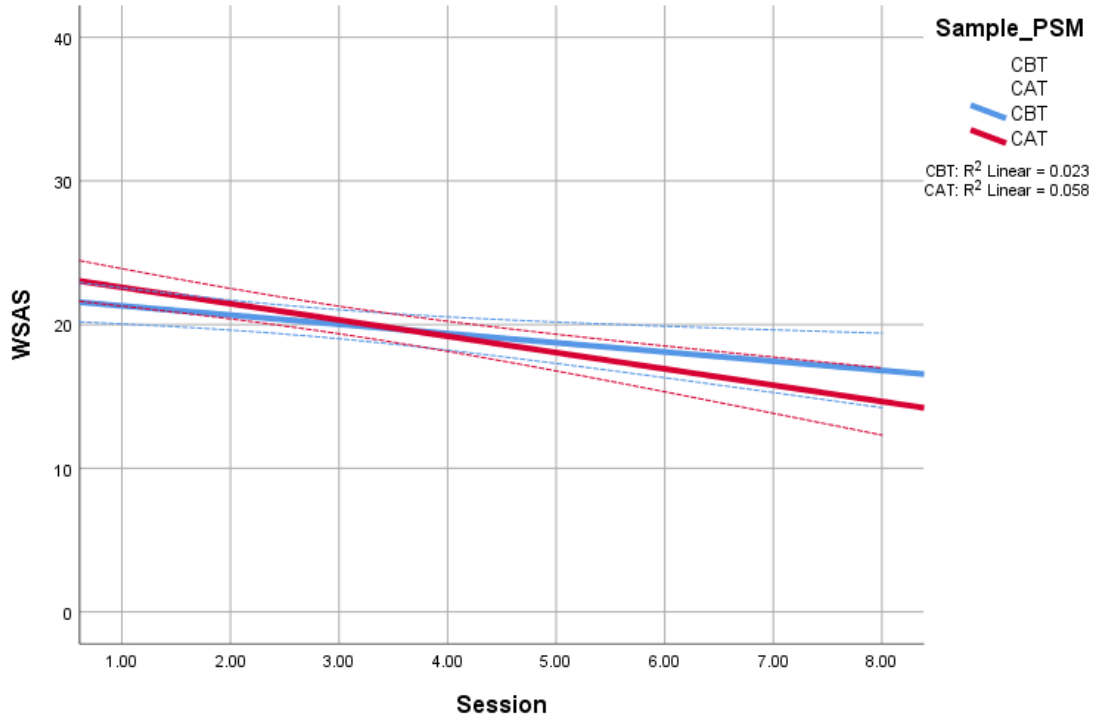

Figure 5: Sample*Time growth curve model showing a linear trend in the WSAS data

3.5 Secondary analyses

3.5.1 Sensitivity analyses: effect size comparisons, treatment sessions, drop-out and change

Between-group post-treatment effect sizes were compared for each outcome measure using Cohen's *d* (Table 3). In the combined dataset of all PSM cases, the outcome measures showed between-group effect sizes in the small to minimal range. When the number of treatment sessions attended was compared this indicated a significant difference in the average number of CAT and CBT treatment sessions ($t(144) = 2.890$, $p<0.004$). This showed that CAT had a significantly higher number of treatment sessions completed (M = 6.01 (2.54)) compared to the CBT treatment group (M = 4.74 (2.78)) in the PSM samples.

Table 3: Effect size data from the PSM dataset

		Post-treatment: Mean (SD)		**Between group effect size**
		Group		
		CAT	CBT	
	Outcome measure	*mean (SD)*	*mean (SD)*	
PSM dataset	PHQ-9	11.49 (7.12)	12.95 (7.61)	0.20
	GAD-7	9.62 (6.33)	10.63 (6.27)	0.16
	WSAS	16.49 (10.43)	17.51 (10.90)	0.10

In regards to drop-out rates (indicated by those ceasing therapy before session 3), within the CBT group, *n=21* (28.8%) patients left therapy =<session 2, whilst within the CAT group *n=11* (15.1%) patients left therapy =<session 2. In addition, clinically significant change and reliable change was examined to highlight any changes between pre- and post-intervention in each treatment group (Table 4). The results indicate that both treatment groups had a similar number of individuals meeting 'caseness' (and thus, similar number not meeting 'caseness') following treatment, with 60% of individuals still meeting 'caseness' on the PHQ-9 measure, and 59-62% on the GAD-7 measure.

Table 4: Clinically significant change and reliable change from the PSM datasets

		Group	
		CAT	CBT
	Outcome measure	*number (%)*	*number (%)*
Caseness_end*	PHQ-9	44 (60)	44 (60)
	GAD-7	43 (59)	45 (62)
Clinically significant change	PHQ-9	27 (37)	19 (26)
	GAD-7	28 (38)	18 (25)
Reliable change	PHQ-9	40 (55)	21 (29)
	GAD-7	38 (52)	27 (37)
Reliable and clinically significant change	PHQ-9	26 (36)	17 (23)
	GAD-7	28 (38)	17 (23)

Note: 'Caseness' refers to the number (and percentage) of individuals who continued to score above clinical cut-offs on the outcome measures at the end of treatment.

4. Discussion

The present pilot study aimed to analyse practice-based outcomes data from patients who accessed CAT or CBT for common mental health problems in an IAPT service. The focus on outcomes for brief 8-session CAT in IAPT contributes to the CAT evidence base which has few studies of outcome for anxiety and depression (Calvert & Kellett, 2014). This analysis was innovative as it was completed using longitudinal multilevel modelling techniques to investigate the trajectory of change over time in treatment groups matched on baseline characteristics through propensity score matching methods. Secondary analyses examined between-group effect sizes, treatment length and rates of reliable and clinically significant change.

The main findings from the primary analysis indicate no significant differences between treatment modalities were evident on symptom change over time. In other words, patients accessing either CAT or CBT experienced similar patterns of change in depression, anxiety, and functioning over sessional time. The propensity score matching procedure used to enable this

comparison derived comparable samples, which strengthens the credibility of the main findings of little differences being apparent in terms of outcome. From the unmatched sample, all baseline symptom measures significantly differed between groups. This may have been due to the CAT sample being offered to patients with identified interpersonal difficulties and complex life histories. This was minimised in the PSM sample as any differences on baseline depression or baseline anxiety scores were minimal (not significant), and only differences on baseline WSAS scores remained albeit minimally.

The findings indicate that patients accessing either therapy tended to experience a decrease (i.e., improvement) in symptom severity over time. This outcome confirms the hypothesis that the treatment outcomes of CBT and CAT would not be significantly different. The findings here compliment an extensive literature of work identifying similarities in treatment outcomes across therapeutic models. In 1977, Smith and Glass reported similar results to the findings in the current study with no significant differences being found when comparing the treatment effectiveness of behavioural and non-behavioural therapies. In subsequent research outputs, authors have further investigated this using sophisticated meta-analyses techniques and have added further support to this finding. For example, comparisons between CBT and other psychotherapies (e.g., IPT) have found no significant differences in regard to treatment outcomes (Cuijpers et al, 2013; Cuijpers et al, 2017). Furthermore, meta-analyses techniques have been utilised to investigate treatment formats in relation to treatment outcomes, and findings have indicated that in treatments which have a human element (i.e., where the patient is supported by a therapist) each is as effective as the other (Cuijpers et al, 2019). The treatment formats include individualised therapy, group work, telephone intervention and guided self-help. In addition, this work found that where a human element was not present, i.e., in unguided self-help, treatment

outcomes were significantly lower than the other formats (Cuijpers et al, 2019). In terms of implications for services, this body of evidence adds support for alternative formats to individual therapy being offered, which is consistent with the IAPT model of working. This also fits with a recent finding reported in a meta-analysis review of evidence-based research studies (Wakefield et al., *in press*). This study indicated that group-based formats delivered within the IAPT model were comparatively as effective as individual formats. With more formats being offered which have been shown to be similarly effective, services have the potential to be able to make therapeutic interventions more accessible to a wider population.

Despite the clear theoretical differences between the treatment studies here, there was a non-significant difference of outcome, suggesting a common pan-model mechanism of action. The equivalence paradox also forwards the idea the common therapeutic actions shared by psychotherapies. This is despite the case that both interventions have differing model assumptions and approaches, such as with CBT and CAT. CBT is a model of therapy based on the inter-related relationship between a person's own thoughts, feelings, behaviour and physiology, and uses a range of both cognitive and behavioural techniques to challenge unhelpful behaviour or thinking patterns. In contrast, CAT has a key therapeutic aim of building a therapeutic relationship to allow re-enactments of reciprocal roles to be explored in and out of the room in the effort to build a greater range and flexibility of roles and procedures.

The current sample within this research project, both within CAT and CBT treatments, either entered treatment within the IAPT service at step 3 (high intensity) or were 'stepped up' to step 3, which would indicate that their initial symptom severity was severe enough to warrant a high intensity intervention. A current growing body of research has highlighted other sources of heterogeneity that could have an impact on treatment outcomes. These include patient

characteristics (e.g., deprivation level; Delgadillo et al., 2016) and service-level factors (e.g., number of sessions, size of service; Gyani et al., 2013).

Another area of research investigating factors related to outcome change is therapist effects, i.e., the degree to which patients treated by different therapists have similar treatment outcomes. There has been variation between studies reporting therapist effects, and a recent systematic review indicated a therapist effects range of 0.2-29% from different study designs (the majority being practice-based) and further supported the notion that it matters most who the therapist is to those who enter treatment with higher baseline symptom severity (Johns et al., 2019). IAPT-specific therapist effects have been researched and evidence has indicated a 6.7% effect in combined low and high intensity interventions (Pereira et al., 2016), and a 5.8% effect on treatment outcomes in high intensity interventions undertaken within the IAPT programme (Saxon et al., 2016). Further work is currently being investigated and that potentially links in with the IAPT model well is around outcome feedback (OF). The IAPT model monitors and collects outcome sessional data and this is one of the key components of the model. It may be that OF is a potential next step in the IAPT evolution as OF technology takes those outcomes and encourages therapists to spend time at each session going through changes with the individual client. It can also be utilised within supervision to track clients progress and identify clients deemed ‘not on track’. Research outcomes into this area have shown that in comparisons between those clients treated by therapists using OF and those clients by therapists not using OF, treatment outcomes did not show significant differences however the average length of treatment was shorter in the OF condition and thus the cost of treatment was less (Delgadillo, Overend, Lucock, Groom, Kirby, et al., 2017). Again, this research is mostly focussed in CBT

interventions although could be extended to further interventions and psychological models within the IAPT services.

Time did indicate a significant effect, i.e., change on outcome measures occurs over sessions, and did so in a decreasing manner (i.e., improvement). Whilst this is not a surprising finding, it did indicate that both sample groups, on average, showed decreases in symptom severity pre-post treatment with a trajectory that generally showed a steeper reduction in outcome scores in the earlier sessions of the intervention. One line of research has proposed that, specifically in CBT interventions, those with depression or anxiety symptomatology and who show early gains are more likely to have better outcomes post-treatments than those who do not report early gains (Aderka, Nickerson, Boe & Hofmann, 2012) and this finding may be extended to CAT interventions at least in the briefer form reported here in the current study.

As a psychological treatment, CAT was presented as a time-limited therapy offering 16-24 sessions for individuals experiencing a range of mental health difficulties (Ryle & Kerr, 2002). Whilst this treatment length can be considered a short-term therapy next to the more traditional psychotherapies offered that were delivered, the findings from the present study indicate that CAT can have positive effects delivered in a much shorter timeframe. CAT has expanded over the past decade and research studies into the use of CAT with other populations are gaining momentum including beneficial effects being found with those struggling with eating disorders (Treasure et al., 1995), morbid jealousy (Kellett & Totterdell, 1995), and following acquired brain injuries (Yeates et al., 2008). Furthermore, CAT has also been shown to be successful when delivered as a group intervention (Calvert et al., 2015) and more recently has been used within a randomised dismantling trial for depression (Kellett et al., 2018). Positive outcomes following therapy with those struggling with depression and anxiety have also been

reported (Hamill & Mahoney, 2011). A recent meta-analysis indicated significant reductions in depression symptomatology following CAT intervention (Hallam, Simmonds-Buckley & Kellett, 2020). Whilst CBT is the NICE recommended treatment of choice for both these latter problems (CG90; CG113), CAT can be recommended via clinical judgement when CBT is not warranted (either by the therapist or client themselves), however it has not been until more recently that CAT has been offered as a therapeutic option within the IAPT programme. A low dropout rate for CAT has been previously identified (Calvert & Kellett, 2014). In the current study, 15% of CAT patients left therapy prior to session 3 which fits with the current literature around drop-out rates within CAT. A recent meta-analysis reported that drop-out rates ranged between 0-38% (average = 23%) from those accessing CAT (Hallam et al, 2020). The evidence presented here in the present study further contributes to the benefits of CAT and without it being inferior to CBT.

Whilst the IAPT programme does have a reputation for relying heavily on cognitive-behavioural treatments, more recently researchers have begun to question the differential effect of CBT as a NICE-recommended treatment for depression, as outcomes between CBT and counselling treatments have found the two to be generally comparable (e.g., Pybis et al., 2017). Both are step-3 or 'high-intensity' treatment and studies have shown the benefits of CBT following more intensive (longer) treatments (above 8 sessions) whilst counselling for depression has been reported as showing more improvement in recovery rates in briefer forms (2-7 sessions). In the current study, when investigating CAT and CBT with similar average treatment lengths, they showed similar change trajectories over time.

4.1 Critique of the present study

The current study is limited by the lack of randomisation to treatment groups and also lack of any follow-up data. This is particularly a limitation, as the CAT model stresses the

importance of endings and structured follow-up. The current study was also unable to investigate adherence to the model or competence of the therapists. There is an extant measure of CAT competency (e.g., Bennett & Parry, 2004) and the Revised Cognitive Therapy Scale (CTS-R; James, Blackburn & Reichelt, 2001) is a well-established CBT competency tool. Some argue that adhering to the recommended protocol is key to treatment outcomes (e.g., Waller, 2009) and that trained practitioners are aware of the evidence-based therapeutic protocols, yet evidence has shown that they are used sparingly in practice (e.g., Waller, Stringer & Mayer, 2012). Thus, having ways to monitor fidelity to the treatment offered can ensure that what is reported as being the therapeutic intervention is in fact being delivered. Within IAPT services, including the ones within this current study, manualised protocols are employed to those attending treatment sessions. The IAPT programme has also recognised that supervision, both clinical and case-management, are integral to delivering the stepped-care model and thus guidelines have been produced to support this process, which can be used as one way to ensure fidelity to treatment protocols and reduce therapeutic drift (IAPT Supervision Guidance, 2011). On the contrary, it is also important to recognise that others have reported that more experienced therapists may be able to understand when it is appropriate to 'drift' (known as 'therapist drift') from the protocol and still maintain positive treatment outcomes (Tschuschke et al., 2015). Without being able to monitor the fidelity or competency within the current study, we are unable to comment here on whether these factors had any influence on the findings, yet this is an area of research that could be progressed in future studies.

The analysis used here, namely, longitudinal MLM, allows for the modelling of trajectories over time, which is particularly systematic and robust when dealing with large amounts of data from a programme such as IAPT. The data presented here consisted of a

hierarchical structure whereby scores on outcome measures were nested within patients, and the MLM procedure is able to handle issues surrounding missing data and correlated variables (i.e., the assumption of independence). Being able to input random effects into MLM makes the model more accurate when analysing data from clinical practice, rather than assuming effects (e.g., baseline scores) are fixed across groups. Non-parametric tests were utilised to compare baseline characteristics between groups, which takes into account skewness in the data. The PSM method is a valid statistical tool that mimics the randomisation procedure used in RCTs to match groups as equivalent as possible on important factors and is recommended for such a task (Beal & Kupzyk, 2014). The matching within the present study was done to an acceptable level, with minimal differences between groups following the procedure, augmenting the strength of the findings.

Despite this, there are still areas of potential limitation within the present study that should be addressed. One of the limitations of this study design was that the propensity score matching procedure yielded balanced samples on most baseline characteristics, but some imbalance remained in some features (WSAS) and in sample sizes (CAT n=76; CBT n=73). Although the PSM procedure is a methodologically robust approach in contexts where randomisation is not feasible or ethical (Rosenbaum & Rubin, 1983), it is nevertheless possible that the imperfect balancing of covariates may have introduced some bias into the analysis. Whilst the number of treatment sessions received was compared, other areas of interest that the study was unable to investigate include the dropout rate between the two treatments which may have further enhanced the understandings of the two offered therapies. Another drawback of the current study is the lack of any follow-up. Therefore, whilst this study provides preliminary information about the comparison between two high intensity interventions delivered within the

UK IAPT programme, future research should look to investigate whether the findings are maintained over time.

The advantages and disadvantages of using pre-existing data sets have been discussed within the literature. Regarding the data set in question here, there are specific advantages to conducting this research study. Firstly, the data has been collected as part of routine clinical practice in two working IAPT services. In the dataset, CAT as a treatment was being piloted and therefore, since there became a sufficient amount of completed data to analyse, it is ethically compelling to investigate whether this specific type of therapy is of benefit to this population. Furthermore, time and effort on behalf of the patients receiving this treatment in completing sessional outcome measures has been spent, and whilst they may be utilised in individual sessions, an ethical case may be made for utilising the data set as a whole to investigate whether this type of therapy is appropriate, and the results may be used for service development (e.g., by way of rolling it to other IAPT services if benefits are found). Secondly, the use of a separate researcher/team to carry out an analysis of the data, as opposed to it being carried out by those involved in data collection and potentially being the ones giving the treatment, reduces bias and means that conflicts of interest are not present and thus bias in the interpretation of the results is minimised. Clearly, capturing follow-up data of CAT outcomes in IAPT services is at a premium.

4.2 Clinical, service and research implications

In terms of clinical and service implications, the results of this pilot practice-based study add some support for the effectiveness of analytically informed therapies for those accessing IAPT services. The study has been novel as it has been the first to study brief high intensity CAT in IAPT and so supplements the evidence concerning low intensity CAT in IAPT (Meadows &

Kellett, 2015). Patients with anxiety and depression deserve access to a range of psychological therapies and services need to make more use of offering informed choice. Whilst the preference for psychological therapy over pharmacological treatment is well established (McHugh, Whitton, Peckham, Welge, & Otto, 2013), less is known about patient preferences for psychological therapies and also the role of matching patients to therapies. The results presented here indicate no significant difference between the two treatments which suggests that CAT has a potential role to play in IAPT not only for those individuals who decline or do not benefit from CBT, but also as a first-line treatment. The development of the brief 8-session model of CAT seems to be well suited to IAPT particularly in which the average session attendance is 7 sessions.

In terms of research implications, a further way to investigate whether CAT is acceptable in this population accessing psychological treatments is to conduct patient-preference trials whereby CAT is offered to patients alongside other therapies with one outcome indicating the potential uptake of treatments alongside those already offered. The future evaluation of the brief 8-session CAT model would benefit from the addition of follow-ups to the methodology to assess the durability of the intervention. It would also be interesting to compare the outcomes for PIT and DIT with CAT in IAPT, although again, these are delivered in longer formats. The use of disorder-specific IAPT outcome measurement would also be a valuable way of evaluating effectiveness of the 8-sesion model. Finally, including treatment integrity checks would be a valuable addition to the methods of any future CAT in IAPT outcome research.

As the current study investigated average differences, what it was unable to answer was whether some patients may respond better and thus benefit more from one type of therapy over the other. This should be investigated in other work by observing the individual personalised

differences to treatments and whether patient characteristics could influence preferential treatment offered.

4.3 Conclusions

The UK Government set the agenda to achieve a ‘parity of esteem’ between mental and physical health services (Department of Health, 2012) and one important step in this has been the IAPT programme. Following the national roll-out of the programme in 2008, IAPT has been developing and evolving in many ways, with one of these being in widening the range of high intensity therapies that the programme offers. This study has compared outcomes on well matched cases of CAT and CBT to find that few differences between the therapies emerged in the longitudinal MLM analyses. Therefore, the study adds to the voluminous evidence base of the equivalence of outcomes between psychotherapies when they are delivered in routine practice settings. The 8-session CAT model appears to have good acceptability with patients and is matched in terms of its effectiveness with CBT; the advantage of the model is its brevity and the fact that it is delivered in 8-sessions whilst retaining fidelity to the reformulation, recognition and revision structure of the model. Further, more controlled investigations of the effectiveness and efficacy of the brief 8-sesion model in IAPT appear therefore indicated.

References

Aderka, I. M., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80*(1), 93–101. doi:10.1037/a0026455

Beal, S.J., & Kupzyk, K.A. (2013). An introduction to propensity scores. *The Journal of Early Adolescence, 34*(1), 66–92. doi:10.1177/0272431613503215

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford Press.

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

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

Calvert, R., Kellett, S., & Hagan, T. (2015). Group cognitive analytic therapy for female survivors of childhood sexual abuse. *British Journal of Clinical Psychology*, *54*(4), 391–413. doi:10.1111/bjc.12085

Clark, D. M., Canvin, L., Green, J., Layard, R., Pilling, S., & Janecka, M. (2018). Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. *The Lancet, 391*, 679–686. doi:10.1016/s0140-6736(17)32133-5

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

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kloeiboer, A., & Dobson, K. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *The Canadian Journal of Psychiatry, 58*(7), 376-385.

Cuijpers, P., Noma, H., Karyotaki, E., Cipriani, A., & Furukawa, T. A. (2019). Effectiveness and acceptability of cognitive behavioural therapy delivery formats in adults with depression. A network meta-analysis. *JAMA Psychiatry, 76*(7), 700-707. doi: 10.1001/jamapsychiatry.2019.0268

Delgadillo, J., Asaria, M., Ali, S., & Gilbody, S. (2016). On poverty, politics and psychology: the socioeconomic gradient of mental healthcare utilisation and outcomes. *British Journal of Psychiatry, 209*(5), 429–430. doi:10.1192/bjp.bp.115.171017

Delgadillo, J., Huey, D., Bennett, H., & McMillan, D. (2017). Case complexity as a guide for psychological treatment selection. *Journal of Consulting and Clinical Psychology, 85*(9), 835–853. doi:10.1037/ccp0000231

Delgadillo, J., Moreea, O., & Lutz, W. (2016). Different people respond differently to therapy: A demonstration using patient profiling and risk stratification. *Behaviour Research and Therapy, 79*, 15–22. doi:10.1016/j.brat.2016.02.003

Delgadillo, J., Overend, K., Lucock, M., Groom, M., Kirby, N., McMillan, D., … de Jong, K. (2017). Improving the efficiency of psychological treatment using outcome feedback technology. *Behaviour Research and Therapy, 99*, 89–97. doi:10.1016/j.brat.2017.09.011

Department of Health. (2012). IAPT three-year report. The first million patients. Retrieved from: http://www.iapt.nhs.uk/silo/files/iapt-3-year-report.pdf

Digital NHS. (2020). Psychological therapies: reports on the use of IAPT services, England November 2019. Final including reports on the IAPT pilots/ Retrieved from: https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-report-on-the-use-of-iapt-services/november-2019-final-including-reports-on-the-iapt-pilots

Field, A. (2009). *Discovering Statistics Using SPSS.* (3rd Ed.). London, England: Sage Publications.

Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: Lessons from year one of IAPT. *Behaviour Research and Therapy, 51*(9), 597–606. doi:10.1016/j.brat.2013.06.004

Hallam,C.,Simmonds-Buckley, M., & Kellett, S. (2020). The acceptability, efficacy, effectiveness and durability of cognitive analytic therapy: systematic review and meta-analysis. *Psychology and Psychotherapy, Theory, Research and Practice* (*in print*).

Hamill, M., & Mahony, K. (2011). 'The long goodbye': Cognitive analytic therapy with carers of people with dementia. *British Journal of Psychotherapy, 27*, 292–304. doi: 10.1111/j. 1752-0118.2011.01243

Health and Social Care Information Centre (HSCIC). (2019). Improving Access to Psychological Therapies (IAPT) Executive Summary. Retrieved from: https://files.digital.nhs.uk/B8/5E5409/iapt-month-dec-2018-exec-sum.pdf

IBM Corp. Released. (2017). IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.

James, I. A., Blackbur, I-M., & Reichelt, F. K. (2001). Manual of the revised cognitive therapy scale (CTS-R). Retrieved from: https://www.ed.ac.uk/files/atoms/files/ctsrmanual.pdf

Johns, R.G., Barkham, M., Kellett, S., & Saxon, D. (2018) A systematic review of therapist effects: A critical narrative update and refinement to Baldwin and Imel's (2013) review. *Clinical Psychology Review, 67,* 78-93. doi: 10.1016/j.cpr.2018.08.004

Kellett, S., Bennett, D., Ryle, T., & Thake, A. (2013). Cognitive analytic therapy for borderline personality disorder: Therapist competence and therapeutic effectiveness in routine practice. *Clinical Psychology and Psychotherapy, 20,* 216–225. doi: 10.1002/cpp.796

Kellett, S., & Totterdell, P. (2013). Taming the green-eyed monster: Temporal responsivity to cognitive behavioural and cognitive analytic therapy for morbid jealousy. *Psychology and Psychotherapy: Theory, Research and Practice, 86,* 52–69. doi:10.1111/j.2044-8341.2011. 02045.x

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

Kelly, G. (1956). The psychology of personal constructs. New York, NY: Norton.

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,* 606–13. doi: 10.1046/j.1525-1497.2001.016009606

Kyrios, M., Hordern, C., & Fassnacht, D. B. (2015). Predictors of response to cognitive behaviour therapy for obsessive-compulsive disorder. *International Journal of Clinical and Health Psychology, 15*(3), 181–190. doi:10.1016/j.ijchp.2015.07.003

Layard, R., Clark, A. E., Cornaglia, F., Powdthavee, N., & Vernoit, J. (2014). What predicts a successful life? A life-course model well-being. *Economic Journal, 124,* F720-F738. doi: 10.1111/ecoj.12170

Martell, C. R., Addis, M. E, & Jacobsen, N. S. (2001). *Depression in context: strategies for guided action*. New York: WW Norton.

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,* 461-464. doi: 10.1192/bjp.180.5.461.

National institute for Health and Care Excellence. (2009). *Depression: the treatment and management of depression in adults (update).* NICE clinical guideline 90 (CG90). Retrieved from: https://www.nice.org.uk/guidance/cg90

National institute for Health and Care Excellence. (2011). *Generalised anxiety disorder and panic disorder in adults: management*. NICE clinical guideline 113 (CG113). Retrieved from: https://www.nice.org.uk/guidance/cg113

Ogden, T. (1983). The concept of internal object relations. *International Journal of Psychanalysis, 64,* 227–241.

Pereira, J-A., Barkham, M., Kellett, S., & Saxon, D. (2016). The role of practitioner resilience and mindfulness in effective practice: A practice-based feasibility study. *Administration and Policy in Mental Health and Mental Health Services Research, 44*(5), 691–704. doi:10.1007/s10488-016-0747-0

Perfect, D., Jackson, C., Pybis, J., & Hill, A. (2016). Choices of therapies in IAPT: An overview of the availability and client profile of step 3 therapies. *British Association for counselling and Psychotherapy*. Retrieved from: http://www.bacp.co.uk//docs/pdf/15641_napt_v6.pdf

Purdie, F., Kellett, S., & Bickerstaffe, D. (2012). Predictors of functional disability in disability welfare claimants. *Journal of Occupational Rehabilitation, 22,* 447-455. doi: 10.1007/s10926-012-9368-y

Richards, D. A., Ekers, D., McMillan, D., Taylor, R. S., Byford, S., Warren, F. C., … & Finning, K. (2016). Cost and outcome of behavioural activation versus cognitive behavioural therapy for depression (COBRA): a randomised, controlled, non-inferiority trial. *Lancet, 388,* 871-880. doi:10.1016/ S0140-6736(16)31140-0

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41–55. doi: http://dx.doi.org/10.1093/biomet/70.1.41

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), 129–147. doi: 10.1017/S1352465808004141

Ryle, A., & Kerr, I. B. (2002). *Introducing Cognitive Analytic Therapy: Principles and Practice*. Chichester: John Wiley & Sons.

Ryle, A., Kellett, S., Hepple, J., & Calvert, R. (2014). Cognitive analytic therapy at 30. *Advances in Psychiatric Treatment, 20*(4), 258–268. doi:10.1192/apt.bp.113.011817

Saxon, D., Firth, N., & Barkham, M. (2017). The relationship between therapist effects and therapy delivery factors: Therapy modality, dosage, and non-completion. *Administration and Policy in Mental Health and Mental Health Services Research, 44*(5), 705–715. doi:10.1007/s10488-016-0750-5

Shek, D. T. L., & Ma, C. M. S. (2011). Longitudinal data analyses using linear mixed models in SPSS: Concepts, procedures and illustrations. *Child Health and Human Development, 11,* 42-76. doi: 10.1100/tsw.2011.2

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

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist,* 752-760.

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

The Mandate: A mandate from the Government to NHS England: April 2015 to March 2016. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/386221/NHS_England_Mandate.pdf

The National Collaborating Centre for Mental Health. (2018). The Improving Access to Psychological Therapies Manual. Version 2. Retrieved from: https://www.england.nhs.uk/wp-content/uploads/2019/02/improving-access-to-psychological-therapies-manual.pdf

Treasure, J., Todd, G., Brolly, M., Tiller, J., Nehmed, A., & Denman, F. (1995). A pilot study of randomised trial of cognitive analytical therapy vs educational behavioral therapy for adult anorexia nervosa. *Behaviour Research and Therapy, 33*, 363–367. doi:10.1016/0005-7967(94) 00070-Z

Tschuschke, V., Crameri, A., Koemeda, M., Köhler., M., Berglar, J., Muth, K., ... 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 study. *Psychotherapy Research, 25*, 420-434. http://dx.doi.org/10.1080/10503307.2014.896055

Turpin, G., & Wheeler, S. (2011). IAPT Supervision Guidance. Retrieved from: https://www.uea.ac.uk/documents/246046/11919343/IAPT+Supervision+Guidance+2011.pdf/3a8f6c76-cea0-4f76-af78-e2f0a1f64c5e

Wakefield, S., Kellett, S., Simmonds-Buckley, M., Stockton, D., Bradbury, A., & Delgadillo, J. (*2020*). 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. (in press).*

Waller, G. (2009). Evidence-based treatment and therapist drift. *Behaviour Research and Therapy, 47,* 119-127. doi:10.1016/j.brat.2008.10.018

Waller, G., Stringer, H., & Meyer, C. (2012). What cognitive behavioral techniques do therapists report using when delivering cognitive behavioral therapy for the eating disorders? *Journal of Consulting and Clinical Psychology, 80*(1), 171–175. doi:10.1037/a0026559

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

Wright, D., & Abrahams, D. (2017). An investigation into the effectiveness of Dynamic Interpersonal Therapy (DIT) as a treatment for depression and anxiety in IAPT. *Psychoanalytic Psychotherapy, 29(2),* 160-170. doi: 10.1080/02668734.2015.1035740

Yeates, G., Hamill, M., Sutton, L., Psaila, K., Gracey, F., Mohamed, S., & O'Dell, J. (2008). Dysexecutive problems and interpersonal relating following frontal brain injury: Reformulation and compensation in cognitive analytic therapy (CAT). *Neuro-Psychoanalysis, 10*, 43–58. doi: 10.1080/15294145.2008.10773571

Appendices

1. PHQ-9 outcome measure scale

*Copyright material removed.

2. GAD-7 outcome measure scale

GAD-7

Over the last 2 weeks, how often have you been bothered by the following problems? *(Use "✔" to indicate your answer)*	Not at all	Several days	More than half the days	Nearly every day
1. Feeling nervous, anxious or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it is hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid as if something awful might happen	0	1	2	3

(For office coding: Total Score T____ = ____ + ____ + ____)

3. WSAS outcome measure scale

Work and Social Adjustment Scale (WSAS)

Identifier _____ Date _____

People's problems sometimes affect their ability to do certain day-to-day tasks in their lives. To rate your problems look at each section and determine on the scale provided how much your problem impairs your ability to carry out the activity. This assessment is not intended to be a diagnosis. If you are concerned about your results in any way, please speak with a qualified health professional.

If you're retired or choose not to have a job for reasons unrelated to your problem, tick here ☐

0	1	2	3	4	5	6	7	8
Not at all		Slightly		Definitely		Markedly		Very severely

1	Because of my [problem] my **ability to work** is impaired. '0' means 'not at all impaired' and '8' means very severely impaired to the point I can't work.	
2	Because of my [problem] my **home management** (cleaning, tidying, shopping, cooking, looking after home or children, paying bills) is impaired.	
3	Because of my [problem] my **social leisure activities** (with other people e.g. parties, bars, clubs, outings, visits, dating, home entertaining) are impaired.	
4	Because of my [problem], my **private leisure activities** (done alone, such as reading, gardening, collecting, sewing, walking alone) are impaired.	
5	Because of my [problem], my ability to form and maintain **close relationships** with others, including those I live with, is impaired.	

Print Form Clear Form Total WSAS score =

4. Ethical approval letter

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

NHS
Health Research Authority

Dr Sarah Wakefield
Trainee Clinical Psychologist
Sheffield Health and Social Care NHS Foundation Trust
Clinical Psychology Unit, Floor F
Cathedral Court, 1 Vicar Lane
Sheffield
S1 1HD

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

16 May 2018

Dear Dr Wakefield,

HRA and Health and Care

Study title:	**The effectiveness of 8-session cognitive analytic therapy (CAT) for clients with depression and anxiety presenting in IAPT services; benchmarking outcomes with treatment comparators.**
IRAS project ID:	**241809**
Protocol number:	**156722**
REC reference:	**19/HRA/0025**
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.

*'In flight studies' which have already started an SSI (Site Specific Information) application for NHS organisations in Wales will continue to use this route. Until 10 June 2018, applications on either documentation will be accepted in Wales, but after this date all local information packs should be shared with NHS organisations in Wales using the Statement of Activities/Schedule of Events for non-commercial studies and template agreement/ Industry costing template for commercial studies.

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.).

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 nonNHS organisations to obtain local agreement in accordance with their procedures.

What are my notification responsibilities during the study?

The attached document *"After HRA Approval – guidance for sponsors and investigators"* gives detailed guidance on reporting expectations for studies with HRA and HCRW Approval, 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: Mr Amrit Sinha

Tel: 01142226650

Email: a.sinha@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 **241809**. Please quote this on all correspondence.

Yours sincerely,

Emma Stoica

Senior Assessor

Email: hra.approval@nhs.net

Copy to:

Mr Amrit Sinha, University of Sheffield

List of Documents

The final document set assessed and approved by HRA and HCRW Approval is listed below.

Document	*Version*	*Date*
IRAS Application Form [IRAS_Form_03042018]		03 April 2018
Letter from sponsor [Letter from sponsor]	1	22 March 2018
Other [CV]	1	27 March 2018
Research protocol or project proposal [Research Protocol]	2.4	15 March 2018
Summary CV for Chief Investigator (CI) [CV]	1	23 March 2018
Summary CV for student [CV]	1	23 March 2018
Summary CV for supervisor (student research) [CV]	1	23 March 2018

Summary of assessment

The following information provides assurance to you, the sponsor and the NHS in England and Wales that the study, as assessed for HRA and HCRW Approval, is compliant with relevant standards. It also provides information and clarification, where appropriate, to participating NHS organisations in England and Wales to assist in assessing, arranging and confirming capacity and capability.

Assessment criteria

Section	Assessment Criteria	Compliant with Standards	Comments
1.1	IRAS application completed correctly	Yes	No comments
2.1	Participant information/consent documents and process	Yes	No comments
3.1	Protocol assessment	Yes	No comments

4.1	Allocation of responsibilities and rights are agreed and documented	Yes	As a student study which is taking place at a single NHS site and which does not require review by an NHS REC, a Statement of Activities is not required. An agreement is also not expected.
4.2	Insurance/indemnity arrangements assessed	Yes	No comments
Section	**Assessment Criteria**	**Compliant with Standards**	**Comments**
4.3	Financial arrangements assessed	Yes	No application for external funding has been made.
5.1	Compliance with the Data Protection Act and data security issues assessed	Yes	No comments
5.2	CTIMPS – Arrangements for compliance with the Clinical Trials Regulations assessed	Not Applicable	No comments
5.3	Compliance with any applicable laws or regulations	Yes	No comments
6.1	NHS Research Ethics Committee favourable opinion received for applicable studies	Not Applicable	No comments
6.2	CTIMPS – Clinical Trials Authorisation letter received	Not Applicable	No comments
6.3	Devices – MHRA notice of no objection received	Not Applicable	No comments
6.4	Other regulatory approvals and authorisations received	Not Applicable	No comments

Participating NHS Organisations in England and Wales

This provides detail on the types of participating NHS organisations in the study and a statement as to whether the activities at all organisations are the same or different.

There is only one NHS organisation participating in the study, where research activities are undertaken as described in the application.

The Chief Investigator or sponsor should share relevant study documents with participating NHS organisations in England and Wales in order to put arrangements in place to deliver the study. The documents should be sent to both the local study team, where applicable, and the office providing the research management function at the participating organisation. Where applicable, the local LCRN contact should also be copied into this correspondence.

If chief investigators, sponsors or principal investigators are asked to complete site level forms for participating NHS organisations in England and Wales which are not provided in IRAS, the HRA or HCRW websites, the chief investigator, sponsor or principal investigator should notify the HRA immediately at hra.approval@nhs.net or HCRW at Research-permissions@wales.nhs.uk. We will work with these organisations to achieve a consistent approach to information provision.

Principal Investigator Suitability

This confirms whether the sponsor position on whether a PI, LC or neither should be in place is correct for each type of participating NHS organisation in England and Wales, and the minimum expectations for education, training and experience that PIs should meet (where applicable).
A Local Principal Investigator should be in place at the NHS site for this type of study. GCP training is not a generic training expectation, in line with the HRA/HCRW/MHRA statement on training expectations.

HR Good Practice Resource Pack Expectations

This confirms the HR Good Practice Resource Pack expectations for the study and the pre-engagement checks that should and should not be undertaken
No additional HR arrangements are required as the research activities at the host NHS site will be undertaken by local staff.

Other Information to Aid Study Set-up

This details any other information that may be helpful to sponsors and participating NHS organisations in England and Wales to aid study set-up.
The applicant has indicated that they do not intend to apply for inclusion on the NIHR CRN Portfolio.