Improving Access to Psychological Therapies (IAPT) in the United Kingdom: A systematic review and meta-analysis of 10-years of practice-based evidence

Article *in* British Journal of Clinical Psychology · June 2020

DOI: 10.1111/bjc.12259

CITATIONS
2

READS
407

6 authors, including:

Sarah Wakefield
The University of Sheffield
23 PUBLICATIONS **237** CITATIONS
SEE PROFILE

Stephen Kellett
The University of Sheffield
130 PUBLICATIONS **2,353** CITATIONS
SEE PROFILE

Mel Simmonds-Buckley
The University of Sheffield
13 PUBLICATIONS **28** CITATIONS
SEE PROFILE

Daniel Cawley Stockton
The University of Sheffield
2 PUBLICATIONS **20** CITATIONS
SEE PROFILE

the british psychological society
promoting excellence in psychology

British Journal of Clinical Psychology (2020)

www.wileyonlinelibrary.com

Improving Access to Psychological Therapies (IAPT) in the United Kingdom: A systematic review and meta-analysis of 10-years of practice-based evidence

Sarah Wakefield[1], Stephen Kellett[2]*, Melanie Simmonds-Buckley[1], Daniel Stockton[3], Abigail Bradbury[1] and Jaime Delgadillo[1]

[1]Clinical Psychology Unit, Department of Psychology, University of Sheffield, UK
[2]University of Sheffield and Sheffield Health and Social Care NHS Foundation Trust, UK
[3]The Mid-Yorkshire Hospitals NHS Foundation Trust, UK

Objectives. Improving Access to Psychological Therapies (IAPT) is a national-level dissemination programme for provision of evidence-based psychological treatments for anxiety and depression in the United Kingdom. This paper sought to review and meta-analyse practice-based evidence arising from the programme.

Design. A pre-registered (CRD42018114796) systematic review and meta-analysis.

Methods. A random effects meta-analysis was performed only on the practice-based IAPT studies (i.e. excluding the clinical trials). Subgroup analyses examined the potential influence of particular methodologies, treatments, populations, and target conditions. Sensitivity analyses investigated potential sources of heterogeneity and bias.

Results. The systematic review identified $N = 60$ studies, with $N = 47$ studies suitable for meta-analysis. The primary meta-analysis showed large pre-post treatment effect sizes for depression (d = 0.87, 95% CI [0.78–0.96], $p < .0001$) and anxiety (d = 0.88, 95% CI [0.79–0.97], $p < .0001$), and a moderate effect on functional impairment (d = 0.55, 95% CI [0.48–0.61], $p < .0001$). The methodological features of studies influenced ESs (e.g., such as whether intention-to-treat or completer analyses were employed).

Conclusions. Current evidence suggests that IAPT enables access to broadly effective evidence-based psychological therapies for large numbers of patients. The limitations of the review and the clinical and methodological implications are discussed.

**Correspondence should be addressed to Stephen Kellett, Clinical Psychology Unit, University of Sheffield, Cathedral Court, Floor F, 1 Vicar Lane, Sheffield S1 2LT, UK (email: s.kellett@sheffield.ac.uk).*

DOI:10.1111/bjc.12259

Practitioner points

- IAPT interventions are associated with large pre-post treatment effect sizes in depression and anxiety measures.
- IAPT interventions are associated with moderate treatment effect sizes with regards to work and social adjustment.
- A reduction in dropout and also the prevention of post-treatment relapse via the offer of follow-up support are important areas for future development.

In the United Kingdom, the National Institute for Health and Care Excellence (NICE) guidelines recommend evidence-based psychological interventions for common mental health problems organized in a stepped care model (NICE, 2011). These guidelines were implemented at a national level in 2008 in England through the *Improving Access to Psychological Therapies* (IAPT) programme. Historically, IAPT was founded on the premise that many patients receiving an evidenced-based psychological therapy would likely recover and return to work, therefore reducing the welfare benefit cost burden (Clark, 2011). This national implementation was supported by positive results from two initial IAPT 'demonstration' sites which provided evidence of the feasibility and effectiveness of the IAPT model (Clark *et al*., 2009). Ten years later, there are over 200 IAPT services across England, which is the largest publicly funded and systematic implementation of evidence-based psychological care in the world. The IAPT programme has subsequently served as a model for the development of similar systems in other countries such as Australia (Cromarty, Drummond, Francis, Watson, & Battersby, 2016), Canada (Naeem, Pikard, Rao, Ayub, & Munshi, 2017), Norway (Knapstad, Nordgreen, & Smith, 2018), and Japan (Kobori *et al*., 2014). IAPT services have three distinctive features: a stepped care model of service provision, the implementation of evidence-based and highly standardized and protocol-driven treatments, and also the systematic use of routine outcome monitoring.

To date, approximately 7.5 million referrals have been received by IAPT services since national statistics were introduced in 2012, of whom approximately 4.9 million received psychological treatment. National statistical reports indicate that the IAPT programme now receives around 1.25 million annual referrals. IAPT services deliver psychological treatments following stepped care principles (Bower & Gilbody, 2005), which is an organizational model supported by evidence from controlled trials (Firth, Barkham, & Kellett, 2015) in which progressively intensive psychological treatments are made available to patients according to need. Patients are initially offered brief (≤8 sessions), low-cost, and low-intensity guided self-help (GSH) based on principles of cognitive behavioural therapy. GSH is psychoeducational in nature and can be delivered over the telephone, via computerized CBT, in large groups or in a one-to-one format. GSH in IAPT services is delivered by psychological well-being practitioners (PWPs), who are trained and supervised to deliver highly standardized, evidence-based interventions guided by a national competency framework and associated assessment and treatment competency measures (Kellett *et al*., 2020). Patients who have not benefited from GSH are stepped up to high-intensity psychological therapies, which involve formal CBT and other therapies such as person-centred experiential counselling, interpersonal psychotherapy (IPT), dynamic interpersonal therapy (DIT), eye movement desensitization and reprocessing (EMDR) and couples counselling for depression. High-intensity interventions are delivered following evidence-based treatment protocols, are lengthier (i.e. typically around 16–20 sessions), and are mostly delivered one-to-one, in person. These

interventions are delivered by qualified therapists , under weekly clinical supervision to ensure fidelity to associated competency frameworks (e.g., Roth & Fonagy, 2005).

IAPT services operate a routine outcome monitoring system in which patients complete a series of standardized questionnaires on a session-to-session basis, including self-reported measures of depression (Patient Health Questionnaire-9; PHQ-9; Kroenke, Spitzer, & Williams, 2001), anxiety (Generalized Anxiety Disorder Scale-7; GAD-7; Spitzer, Kroenke, Williams, & Lowe, 2006), and functional impairment (Work and Social Adjustment Scale; WSAS; Mundt, Marks, Shear, & Greist, 2002). Other disorder-specific questionnaires are also applied when relevant to the patient's problems (Mental Health Policy Team, 2018). This routine outcome monitoring system has enabled the large-scale evaluation of IAPT services around the country, yielding insights into the factors that distinguish more and less effective services (e.g., see Clark *et al*., 2018; Gyani, Shafran, Layard, & Clark, 2013). Furthermore, numerous studies have emerged from IAPT services, supported by practice research networks of IAPT therapists and researchers (e.g., see Lucock *et al*., 2017). The IAPT programme is also remarkable for its transparent and open-access reporting of clinical performance data at a national scale (Clark *et al*., 2018).

The present study is the first systematic review of practice-based studies arising from the first 10 years since the implementation of the IAPT programme. Its primary objective was to quantify the effectiveness of IAPT interventions delivered during routine practice. As such, this review focused specifically on quantitative, practice-based outcome research, excluding randomized controlled trials (RCTs). The rationale for excluding RCTs conducted within IAPT services (e.g., Richards *et al*., 2016) is that these studies often apply strict inclusion/exclusion criteria which render samples that are not typical of routine IAPT populations (e.g., excluding cases with comorbid disorders; Westen & Morrison, 2001). Furthermore, effects from RCT samples may not be realistic reflections of the effects of routine service delivery (e.g., see Baker, McFall, & Shoham, 2009). Because the IAPT programme has expanded to also include assessment and treatment of patients with psychological distress associated with physical health problems (IAPT, 2018), and in order to provide a more comprehensive evaluation of the effectiveness of the programme, studies including patients with long-term physical health conditions were included in this review. A secondary aim was to narratively synthesize the characteristics of the practice-based studies that constitute the IAPT evidence base.

Methods

Inclusion and exclusion criteria

Study inclusion criteria were as follows: (1) an outcome study with an adult clinical population (i.e., 18+ years); (2) quantitatively analysed standardized outcome measures and had at least two points of outcome data collection; (3) published in a peer-reviewed journal and written in English; and (4) conducted in UK-based IAPT service delivering group or individual interventions. Study exclusion criteria were as follows: (1) the focus of the study was on children/adolescent populations; (2) only assessment scores were reported on the outcome measures; (3) the methodology was an RCT design; and (4) qualitative studies/opinion pieces/editorials.

Literature search strategy

The study protocol was prospectively registered (PROSPERO ref: CRD42018114796).

Three databases were searched – Scopus, PsycINFO, and MEDLINE – up until the date of 13-08-2018. The search terms utilized were as follows: 'Improving Access to Psychological Therapies' AND/OR IAPT OR 'stepped care' NOT 'International association for plant taxonomy'. As the IAPT initiative commenced in 2008, the search years were inclusive of 2007 to the current date. The process for capturing all relevant studies followed several components: (1) a systematic search of the three databases using the pre-determined search strings which were operationalized to capture all relevant articles; (2) hand-searching, which involved searching the reference lists of those articles that met inclusion criteria; and (3) of those articles meeting inclusion criteria from steps 1 and 2, a backward/reverse citation search was completed.

Eligibility of relevant articles and data extraction

Sixty studies met the inclusion criteria, with $n = 29$ reporting sufficient statistical information to calculate effect sizes (ESs). For those studies that did not report statistics that were eligible for the meta-analysis (n = 31), we contacted the corresponding author of the article by email and requested 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. Figure 1 is a PRISMA diagram (Moher, Liberati, Tetzlaff, & Altman, 2009) detailing the process of study selection. This process followed two stages and was completed by one author in the first instance (SW). Queries about eligibility were discussed and ratified at subsequent research meetings, including three members of the research team (JD, SK, and 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 [ITT] or completer analysis), intervention, main findings, and outcome measures. Any issues likely to introduce bias were also noted in the data extraction tool.

Quality assessment and risk of bias

The Critical Appraisal Skills Programme (CASP) tool was used to assess the quality of studies (see Table 1). One researcher completed quality assessments for all studies (SW), followed by blind rating by two other raters (rater 1 = accredited IAPT CBT therapist; rater 2 = clinical psychologist). Rater 1 rated 12 papers (which represented 20% of the studies), and rater 2 rated six papers that overlapped with rater 1 (which represented 10% of the studies). Second (blind) ratings were achieved by splitting the 60 included papers into study quality quartiles and then randomly selecting from each quartile (i.e., 15 papers per quartile) to ensure coverage across all study quality levels. Once completed, the ratings were compared and any discrepancies 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 and SK). Inter-rater reliability was calculated using the Kappa statistic (Cohen, 1960); the level of agreement was 'moderate' both 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).

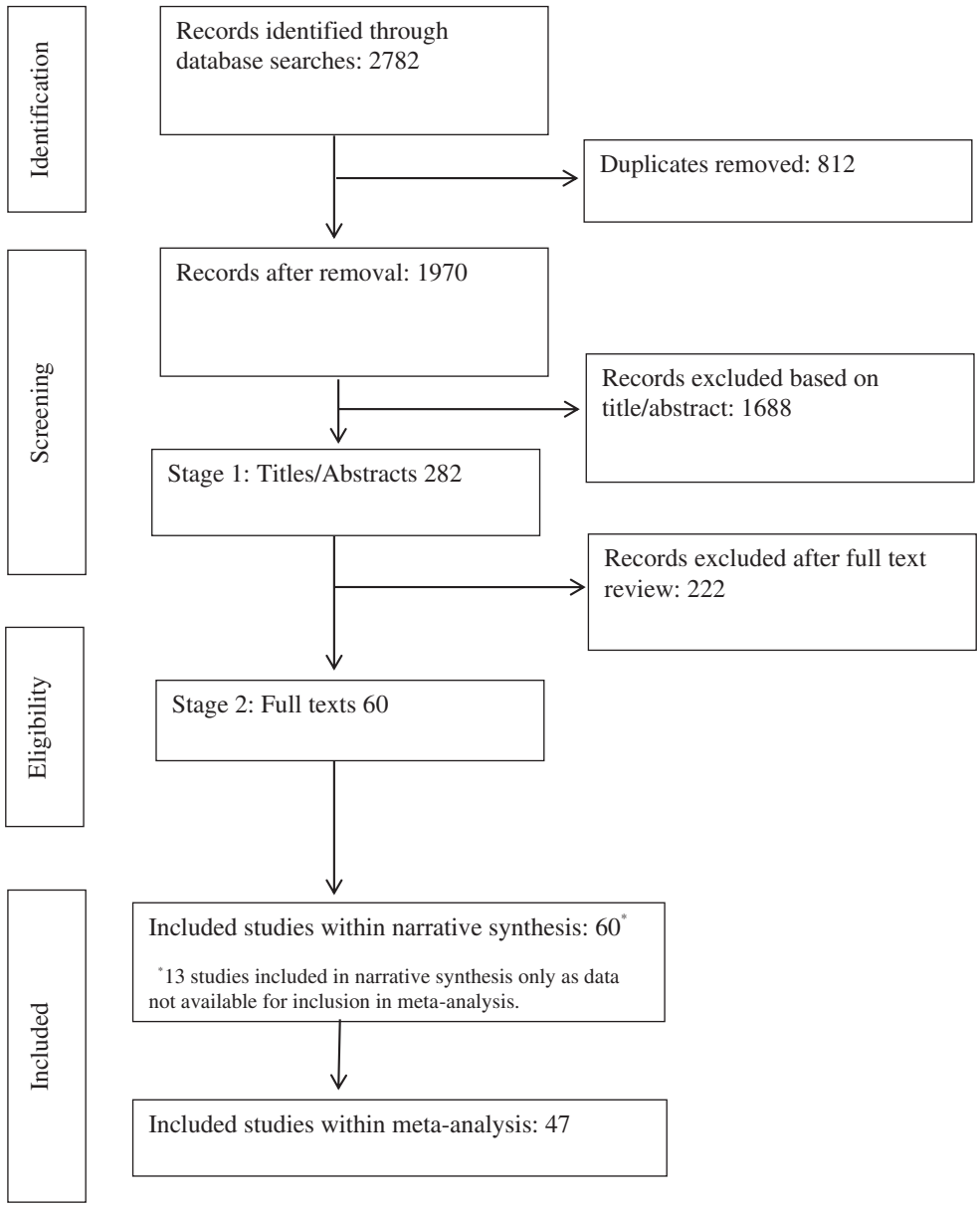

Figure 1. PRISMA summary of included studies

Narrative review and meta-analysis

A narrative synthesis aimed to summarize key study characteristics. A random effects meta-analysis aimed to synthesize 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 as follows: (1) reporting pre- and post-means and *SD*s convertible into an ES (Cohen's *d*; Cohen, 1988), (2) reporting Cohen's *d* ES, (3)

Table 1. Overview of papers in the systematic review

First author and year	Service(s)	Mental health condition(s)	Analysed *N*	Dropout *N*	Analysis (ITT or completers)	Intervention (step of care)	Main outcome measure(s)	Risk of bias:; CASP rating = low, medium, high
Adamson *et al.* (2015)[b]	Lincolnshire IAPT for male offenders (IAPT-O), category B prison	Depression and anxiety-based disorders	627	93	ITT	Step 2 or Step 3 (2 and 3)	PHQ-9 GAD-7	Low
Ali *et al.* (2014)[b]	Single north of England IAPT service	Mild-to-moderate MH symptoms or functional impairment	1,376	*Not specified*	Completers	Low intensity (2)	PHQ-9 GAD-7	Low
Ali *et al.* (2017)	Single IAPT service	Common MH problems	439	165	Completers[c]	Previous course of low-intensity CBT (2)	PHQ-9 GAD-7 WSAS	Low
Baucom *et al.* (2018)[b]	London IAPT services	Depression and relationship distress	63 clients[d] (with 63 partners)	*Not specified*	ITT	High intensity – BCT-D (3)	PHQ-9 GAD-7 CSI-4	Low
Binnie and Boden (2016)[b]	Single outer London borough IAPT service	*Not reported*	140	61	Completers[e]	CBT (3)	PHQ-9 GAD-7	Low
Branson, Myles, Mahdi, and Shafran (2015)	University of Reading and five participating IAPT services	Anxiety and/or depression	1,247	*Not specified*	ITT	CBT (3)	*Client:* PHQ-9 GAD-7 *Therapist:* CTS-R	Medium
Branson *et al.* (2015)	University of Reading and five participating IAPT services (Thames Valley LETB)	Mild-to-moderate anxiety and/or depression	3,688	*Not reported*	ITT	Low intensity (2)	*Client:* PHQ-9 GAD-7 *Therapist:* ReachOut scales	Medium

Continued

Table 1. (*Continued*)

First author and year	Service(s)	Mental health condition(s)	Analysed *N*	Dropout *N*	Analysis (ITT or completers)	Intervention (step of care)	Main outcome measure(s)	Risk of bias:; CASP rating = low, medium, high
Buckman *et al.* (2018)[b]	Single London IAPT service	Problematic alcohol use; common MH problems	In audit = 3,643 Not in audit = 1,687	642	ITT	*Not specified*	AUDIT-C PHQ-9 GAD-7 IAPT Phobias Scale WSAS	Low
Burns *et al.* (2015)[b]	Single north of England IAPT service	Common MH problems	801	261	Completers	Step 2 'Stress Control' group or 'Stress Control+' group (2)	PHQ-9 GAD-7	Low
Chan and Adams (2014)[b]	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[f]	Low and high intensity (50:50) (2 and 3)	PHQ-9 GAD-7	Medium
Cheston and Howells (2016)	Single south-west of England IAPT service	Diagnosis of dementia; carers	4	1	ITT	LivDem group (2)	QoL-AD *Carer-related outcomes:* QoL perception	High
Clark *et al.* (2009)[b]	Two IAPT demonstration sites – Doncaster and Newham IAPT services[g]	Depression and/or anxiety	*Newham*: 221 (follow-up sample = 60)	*Not reported*	Completers	Low and high intensity (2 and 3)	PHQ-9 GAD-7 CORE-OM Employment status *Follow-up*: PHQ-9 GAD-7 Employment status	Low

Continued

Table 1. (*Continued*)

First author and year	Service(s)	Mental health condition(s)	Analysed *N*	Dropout *N*	Analysis (ITT or completers)	Intervention (step of care)	Main outcome measure(s)	Risk of bias:; CASP rating = low, medium, high
Clark *et al.* (2018)[b]	NHS Digital and Public Health England data	Depression and/or anxiety	2014/15: 221 CCG 2015/16: 209 CCG (487,523 used in meta-analysis)	*Not reported*	Completers	*Not specified*	PHQ-9 GAD-7 WSAS	Medium
Clarkson *et al.* (2016)[b]	Military Veterans' IAPT service (North-West)	Mild-to-moderate MH difficulties	505	170	ITT	Low and high intensity (2 and 3)	PHQ-9 GAD-7 WSAS	Low
Delgadillo, McMillan, Leach, *et al.* (2014)[b]	Single north of England IAPT service	Common MH problems	2,891	*Not specified*	ITT	Step 2 (low intensity) and Step 3 (high intensity) (2 and 3)	PHQ-9 GAD-7	Low
Delgadillo, McMillan, Lucock, *et al.* (2014)[b]	Single north of England IAPT service	Common MH problems	1,850	511 (35.1%)	ITT	Low intensity (2)	PHQ-9 GAD-7	Low
Delgadillo, Asaria, Ali, and Gilbody (2016)	211 identifiable CCG areas across England	Common MH problems	110,415	*Not specified*	ITT	*Not specified*	PHQ-9 GAD-7	Low
Delgadillo, Kellett, *et al.* (2016)[b]	Five northern IAPT services	Depression and/or anxiety	4,451	1,359	ITT	Step 2 (low intensity) 'Stress Control' group (2)	PHQ-9 GAD-7 WSAS	Low
Delgadillo, Moreea, *et al.* (2016)[b]	Single north of England IAPT service	Depression and/or anxiety	1,347	*Not specified*	ITT	Step 2 (low intensity) and Step 3 (high intensity) (2 and 3)	PHQ-9 GAD-7 WSAS	Low

Continued

Table 1. (*Continued*)

First author and year	Service(s)	Mental health condition(s)	Analysed *N*	Dropout *N*	Analysis (ITT or completers)	Intervention (step of care)	Main outcome measure(s)	Risk of bias:; CASP rating = low, medium, high
Delgadillo, Dawson, *et al.* (2017)[b]	Single north of England IAPT service	Depression and anxiety-related problems with or without LTCs	28,498	*Not reported*	ITT	Low and high intensity (2 and 3)	PHQ-9 GAD-7 WSAS	Low
Delgadillo, Huey, *et al.* (2017)[b]	Single northern England IAPT service	Depression, anxiety, or other MH problems	1,512	31.3% (low intensity = 32.2%; high intensity = 28.5)	ITT	Low and high intensity (2 and 3)	PHQ-9 GAD-7 WSAS SAPAS	Low
Delgadillo, Overend, *et al.* (2017)[b]	Single north of England IAPT service	Depression and anxiety problems	594	*Not specified*	ITT	Low and high intensity (2 and 3)	PHQ-9 GAD-7	Low
Elison *et al.* (2017)[b]	Single Greater Manchester IAPT service	Range of MH issues	1,068	216	ITT	Low intensity (e-Therapy self-help) (2)	PHQ-9 GAD-7 WSAS	Low
Firth *et al.* (2015)[b]	Single citywide IAPT service	*Not specified*	6,111	1,553	ITT	Step 2 (low intensity) (2)	PHQ-9 GAD-7 WSAS	Low
Giebel *et al.* (2014)[b]	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	Low
Goddard, Wingrove, and Moran (2015)	Southwark Psychological Therapies Service (IAPT)	Comorbid personality disorder with depression and/or anxiety	1,005	35%	ITT	Low and high intensity (2 and 3)	PHQ-9 GAD-7 WSAS SAPAS	Low
		Not reported	1,122	0	ITT	Step 2	PHQ-9	Low

Continued

Table 1. (*Continued*)

First author and year	Service(s)	Mental health condition(s)	Analysed *N*	Dropout *N*	Analysis (ITT or completers)	Intervention (step of care)	Main outcome measure(s)	Risk of bias:; CASP rating = low, medium, high
Green, Barkham, Kellett, and Saxon (2014)[b]	Six IAPT services located within the north of England					(2)	GAD-7	
Griffiths and Griffiths (2015)[b]	Four IAPT services (three midlands; one south-east)	Those scoring 'severe' on outcome measures (depression, anxiety, and functioning)	25,034 (severe anxiety sample *n* = 14,612 used in meta-analysis)	0	ITT	*Not specified*	PHQ-9 GAD-7 WSAS	Medium
Gyani *et al.* (2013)[b]	*N* = 24 year one IAPT services	Depression and/or anxiety	19,395 (11,535 used in meta-analysis)	*Not specified*	Completers	Low and high intensity (2 and 3)	PHQ-9 GAD-7	Low
Hammond *et al.* (2012)[b]	*N* = 7 IAPT services in east of England region	*Not reported*	4,106	0	ITT	Low intensity – OTT or FTF (2)	PHQ-9 GAD-7 WSAS	Low
Highfield *et al.* (2016)[b]	Coventry and 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) (2 and 3)	PHQ-9 GAD-7 SEMCD scale (step 2 only)	High

Continued

Table 1. (*Continued*)

First author and year	Service(s)	Mental health condition(s)	Analysed *N*	Dropout *N*	Analysis (ITT or completers)	Intervention (step of care)	Main outcome measure(s)	Risk of bias:; CASP rating = low, medium, high
Jolley *et al.* (2015)[b]	SLaM – IAPT-SMI demonstration site	Service users with psychosis experience	54	11	Completers	CBT-p (16–30 sessions) (3)	*Clinical outcomes:* CHOICE WEMWBS WSAS PSYRATS *Other outcomes:* service user experience, satisfactions, and feedback questionnaires Friends and Family Test EQ5D	Medium
Kellett *et al.* (2020)[b]	Single northern England LTC/MUS Pathfinder site	Depression and/or anxiety alongside LTCs or MUS.	1,016	130	ITT	Step 2 (low intensity) and Step 3 (high intensity) (2 and 3)	PHQ-9 GAD-7	Low
Kellett *et al.* (2017)[b]	Single IAPT service	Depressive symptoms	26	1	ITT	Behavioural activation group (3)	PHQ-9 GAD-7 WSAS	Low
Kenwright *et al.* (2017)[b]	North Midlands IAPT service	Anxiety disorders and comorbid IBS	104	23	ITT	Step 2 and Step 3 (2 and 3)	PHQ-9 GAD-7 WSAS IBS-specific measures	Low

Continued

Table 1. (*Continued*)

First author and year	Service(s)	Mental health condition(s)	Analysed *N*	Dropout *N*	Analysis (ITT or completers)	Intervention (step of care)	Main outcome measure(s)	Risk of bias:; CASP rating = low, medium, high
Kuhn (2011)[b]	Newham Primary Care Psychological Services	Common MH problems	65	7	ITT	Systemic therapy (3)	PHQ-9 GAD-7 CORE-OM WHO DAS II CSQ-8 Client satisfaction questionnaire Employment questionnaire	High
Lucock *et al.* (2018)[b]	Single north of England IAPT service	Remission of symptoms following psychological intervention for depression	11	4	ITT	Low intensity – 'SMArT' intervention (2)	PHQ-9 GAD-7	Medium
Luik *et al.* (2017)[b]	NHS-funded charity in Manchester IAPT service	Insomnia-related depression and/or anxiety	72	26	Completers	Digital CBT (dCBT) (*Not specified*)	PHQ-9 GAD-7 ISI	Medium
Prina *et al.* (2014)	Six IAPT services in the east of England region	Depression and/or anxiety	16,236	4,931	ITT	Step 2 and Step 3 (2 and 3)	PHQ-9 GAD-7	Low
McDevitt-Petrovic *et al.* (2018)[b]	Northern Ireland IAPT service	Common MH difficulties	163	*Not specified*	ITT	Low-intensity CBT (2)	PHQ-9 GAD-7	Medium
Meadows and Kellett (2017)[b]	Single IAPT service	Depression and/or anxiety	10	7	Completers	Step 2 – CAT-SH (2)	PHQ-9 GAD-7 WSAS	Low
	BTSS	PTSD	6	0	Completers			PHQ-9

Continued

Table 1. (*Continued*)

First author and year	Service(s)	Mental health condition(s)	Analysed *N*	Dropout *N*	Analysis (ITT or completers)	Intervention (step of care)	Main outcome measure(s)	Risk of bias:; CASP rating = low, medium, high
Methley, Woodruff, Sayer, and Nevin (2016) Medium						Step 4 waiting list – psychoeducation PTSD group (4)	GAD-7 IES-R SCS-SF ERQ	
Mofrad and Webster (2012)	Single north-east of England IAPT service	Depression and simple phobia	1	0	ITT	Behavioural activation (3)	PHQ-9 GAD-7	High
Morrison, Walker, Ruggeri, and Hacker Hughes (2014)[b]	Single east of England IAPT service	Depression (with little or no comorbid anxiety)	12	5	Completers	Low-intensity 'MindBalance' intervention (2)	PHQ-9 WSAS BDI	Medium
Murray (2017)[b]	Single east of England IAPT services	PTSD	57 (PHQ-9 and GAD-7), 21 (IES-R)	*Not reported*	Completers	Step 3 – TR-CBT (3)	PHQ-9 GAD-7 IES-R	High
Pack and Condren (2014)[b]	Single IAPT service	Low self-esteem	50	39[h]	Completers	CBT group *(Not specified)*	PHQ-9 GAD-7 RSES	Medium
Pereira, Barkham, Kellett, and Saxon (2016)[b]	One IAPT service	Depression and/or anxiety	4,980	*Not reported*	ITT	Low and high intensity (2 and 3)	PHQ-9 WSAS	Medium
Pettit *et al.* (2017)	South-west of England IAPT services	*Not specified*	'Attenders' = 54,328 'Completers' = 22,858	*Not reported*	Completers	*Not reported*	PHQ-9 GAD-7	Medium
Poots *et al.* (2014)[b]	Single (Westminster) IAPT service	Depression	1,426	3,208	Completers	*Not reported*	PHQ-9	High

Continued

Table 1. (*Continued*)

First author and year	Service(s)	Mental health condition(s)	Analysed *N*	Dropout *N*	Analysis (ITT or completers)	Intervention (step of care)	Main outcome measure(s)	Risk of bias:; CASP rating = low, medium, high
Pybis *et al.* (2017)[b]	(Up to) *N* = 121 IAPT services involved in the 2nd NAPT	Depression and/or anxiety, or other common MH problems	33,243 (CBT *n* = 23,595; Counselling *n* = 9,648)	9,262	ITT	Step 3 CBT and Step 3 counselling (3)	PHQ-9 GAD-7	Low
Radhakrishnan *et al.* (2013)[b]	*N* = 5 PCT IAPT services, east of England	*Not specified*	8,464	1,961	ITT	Low and high intensity (2 and 3)	PHQ-9 GAD-7	Low
Richards and Borglin (2011)[b]	Single north of England IAPT service	Common mental health difficulties	4,183	969	ITT	Low and high intensity (2 and 3)	PHQ-9 GAD-7	Low
Rimes *et al.* (2017)[b]	*N* = 4 London borough IAPT service(s)	Common MH difficulties within 6 different sexual orientation groups	1) 182 2) 213 3) 6,382 4) 619 5) 72 6) 2,901	*Not reported*	ITT[i]	Low and high intensity (2 and 3)	PHQ-9 GAD-7 WSAS	Low
Saunders, Cape, Fearon, and Pilling (2016)[b]	Two London services	Depression and anxiety disorders	16,636 (split into two samples): *n* = 8,321; *n* = 8,315	*Not specified*	ITT	Step 1 ('brief interventions') and Step 2 ('formal interventions') (2)	PHQ-9 GAD-7 WSAS Phobia Scale – self-rating	Low
Saxon *et al.* (2016)[b]	*Not specified*	Common MH problems	4,034	*Not reported*	Completers	Step 3 counselling or CBT	PHQ-9	Low

Continued

Table 1. (*Continued*)

First author and year	Service(s)	Mental health condition(s)	Analysed *N*	Dropout *N*	Analysis (ITT or completers)	Intervention (step of care)	Main outcome measure(s)	Risk of bias:; CASP rating = low, medium, high
						(3)		
Scott (2018)	North-west of England IAPT services	Various MH difficulties	29	*Not reported*	Completers	*Not reported*	PHQ-9 GAD-7	High
Vaillancourt, Manley, and McNulty (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 (2 and 3)	PHQ-9 GAD-7	Medium
Wright and Abrahams (2015)[b]	Single inner London borough IAPT service	Anxiety and/or depression or other common MH difficulties	24	0	ITT	DIT (3)	PHQ-9 GAD-7	Medium
Wroe, Rennie, Gibbons, Hassy, and Chapman (2014)	*Not reported*	Low mood and worry alongside T2DM	*Variable depending on phase of service development*	*Not specified*	Completers	Step 2 'Wellbeing Group' (2)	PHQ-9 GAD-7 DHP SDSCA Physiological measures: HbA_{1c}	Medium

Continued

Table 1. (*Continued*)

First author and year	Service(s)	Mental health condition(s)	Analysed *N*	Dropout *N*	Analysis (ITT or completers)	Intervention (step of care)	Main outcome measure(s)	Risk of bias:; CASP rating = low, medium, high
Young *et al.* (2017)[b]	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 (2 and 3)	PHQ-9 GAD-7	Medium

AUDIT-C = Alcohol Use Disorders Identification Test – Consumption; BAG = behavioural activation in groups; BCT-D = Behavioural Couple Therapy for Depression; BDI = Beck Depression Inventory; BSL = British Sign Language; BTSS = Berkshire Traumatic Stress Service; CCG = Clinical Commissioning Groups; CHOICE = Choice of Outcome in Cognitive therapy for psychosEs; CORE-OM = Clinical Outcomes in Routine Evaluation – Outcome Measure; CSI-4 = Couples Satisfaction Index (4-item); CTS-R = Cognitive Therapy Scale – Revised; DHP = Diabetes Health Profile; DIT = dynamic interpersonal therapy; EQ5D = EuroQol Group (Quality of Life questionnaire); ERQ = Emotion Regulation Questionnaire; FTF = face-to-face; HbA_{1c} = glycosylated haemoglobin; IAPT-SMI = Improving Access to Psychological Therapies for people with severe mental illness; IBS = irritable bowel syndrome; IES-R = Impact of Events Scale – Revised; ISI = Insomnia Severity Index; LivDem = living well with dementia; LTC = long-term conditions; MH = mental health; MUS = medically unexplained symptoms; NAPT = National Audit of Psychological Therapies; OTT = over the telephone; PCTs = primary care trusts; PSYRATS = Psychotic Symptom Rating Scales; PTSD = post-traumatic stress disorder; PWP = psychological well-being practitioners; QoL-AD = Quality of Life in Alzheimer's Disease; RSES = Rosenberg's Self-Esteem Scale; SAPAS = Standardised Assessment of Personality – Abbreviated Scale; SCS-SF = Self-Compassion Scale – Short Form; SDSCA = Summary of Diabetes Self-Care Activities questionnaire; SEMCD scale = Self-Efficacy for Managing Chronic Disease Scale; SLaM = South London and Maudsley NHS Foundation Trust; SMArT = Self-Management After Therapy; T2DM = type 2 diabetes mellitus; Thames Valley LETB = Thames Valley Local Education and Training Board; TR-CBT = trauma-focused CBT; WEMWBS = Warwick-Edinburgh Mental Wellbeing Scale.

[a]Risk of bias – more information can be found in the Table S1; [b]Those studies included in the meta-analyses; [c]Those who completed treatment were recruited, and following this stage, the data were analysed using ITT (survival analysis) of all participants, even those lost to follow-up; [d]Clients data only reported within this review; [e]Completer 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; [f]Some missing data and not used, but analysis included dropouts; [g]Doncaster outcomes are reported in full in another paper (Richards & Borglin, 2011), and therefore, only Newham data from the Clark *et al.* (2009) paper was used; [h]'Non-completers' used – no information about whether this includes only those who dropped out or others also. Therefore, this figure is an approximate; [i]Estimate based on information given within the paper, as unsure that enough data are available to determine.

reporting other ESs, but with sufficient additional information (i.e., means/*SD*s) to enable Cohen's *d* to be calculated, or (4) reporting the mean pre-post change and *SD*. The calculation for Cohen's *d* was $d = (M_1 - M_1)/SD$ pooled, where *SD* pooled = $\sqrt{((SD_1^2 + SD_2^2)/2)}$. Cohen's power primer definitions (Cohen, 1992) 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 (ITT/completer), step of care (step two/step three/steps two and 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 $p < .01$ for subgroup and meta-regression analyses to account for multiple testing.

Results

Section one of the results presents the narrative synthesis and section two the meta-analysis. Table 1 describes the characteristics and risk of bias assessment of all included studies ($n = 60$). Tables 2 and 3 provide a summary of the moderator analyses performed on studies included in the meta-analysis (see Tables S1–S3 for summaries of the main findings from all included studies).

Demographics

Sample sizes ranged from a single-case study ($n = 1$; Mofrad & Webster, 2012) to data from 209 clinical commissioning groups ($n = 537{,}131$; Clark *et al.*, 2018). One study included only male patients (Adamson, Gibbs, & McLaughlin, 2015), and 17 studies did not report the gender distribution of the patients. Of those studies that reported gender, the average percentage of females was 60.2%. Twenty-seven studies did not report ethnicity data; those studies that did on ethnicity tended to vary in the depth of detail provided. With the exception of three studies, the category of 'White'/'White British'/'Caucasian' was the largest ethnic group. North of England services contributed the largest number of studies ($n = 17$), and London-based services contributed $N = 11$ studies.

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; 21/60 (35%) reported impairment outcomes. Thirty-two other outcome measures were used across 18/60 (30%) studies; only two studies reported patient satisfaction (Jolley *et al.*, 2015; Kuhn, 2011).

Table 2. 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	0.67–0.90	99	5701.68***	**.001****
		COM	22	1.04	0.97–1.12	98	3128.89***	
	Study bias	Low	44	0.82	0.71–0.93	99	5576.75***	.016*
		Medium	17	0.95	0.84–1.06	99	4043.44***	
		High	4	1.26	0.96–1.56	77	13.16**	
	Primary condition	Mental health	52	0.94	0.87–1.01	99	9080.97***	**.001****
		Physical health	10	0.43	0.13–0.74	94	170.82***	
	Step of care	Step 2 only	15	0.80	0.68–0.93	97	686.96***	.038*
		Step 3 only	9	1.09	0.85–1.33	93	120.46***	
	Format	Individual	38	0.77	0.65–0.89	99	5194.26***	.500
		Group	6	0.88	0.59–1.16	92	69.92***	
GAD-7	Methodology	ITT	41	0.80	0.68–0.91	99	5484.73***	<**.001****
		COM	19	1.06	0.98–1.14	99	3967.79***	
	Study bias	Low	41	0.83	0.72–0.94	98	3655.71***	**.001****
		Medium	16	0.97	0.84–1.09	99	5223.23***	
		High	3	1.36	1.10–1.62	24	2.63	
	Primary condition	Mental health	47	0.96	0.88–1.04	99	10813.46***	**.006****
		Physical health	10	0.50	0.19–0.82	94	175.50***	
	Step of care	Step 2 only	14	0.88	0.74–1.03	98	776.13***	.182
		Step 3 only	6	1.16	0.77–1.56	84	32.46*	
	Format	Individual	33	0.76	0.62–0.89	99	4782.75***	.291
		Group	6	0.91	0.66–1.16	93	73.78**	
WSAS[a]	Methodology	ITT	21	0.54	0.48–0.61	98	1236.70***	.154
		COM	3	0.44	0.32–0.57	0	0.74	
	Study bias	Low	19	0.55	0.48–0.62	97	780.55**	.389
		Medium	5	0.48	0.34–0.62	99	810.32***	
	Step of care	Step 2 only	7	0.52	0.43–0.61	98	432.12***	.239
		Step 3 only	2	0.44	0.33–0.55	0	0.18	
	Format	Individual	12	0.48	0.41–0.55	98	916.96	.930
		Group	2	0.48	0.45–0.51	0	0.87	

Note. CI = confidence interval; COM = completer; GAD-7 = Generalized Anxiety Disorder Scale-7; ITT = intention to treat; *k* = number of comparisons per subgroup; PHQ-9 = Patient Health Questionnaire-9; WSAS = Work and Social Adjustment Scale.

[a]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; *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 (in bold).

Mental health conditions and populations

The majority of studies investigated outcomes for depression and anxiety. Six studies (9.8%) investigated outcomes for physical health conditions, with one study investigating outcomes for dementia (Cheston & Howells, 2016). Other target conditions included psychosis, relationship distress, and problematic alcohol use (one study each; 4.9% overall). One study (1.6%) was set in a prison for male offenders (Adamson, Gibbs, & McLaughlin, 2015), whilst two papers (3.3%) studied outcomes with veterans (Clarkson *et al.*, 2016; Giebel, Clarkson, & Challis, 2014). One study explored the effectiveness of an

Table 3. 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	−.00	−0.01 to 0.00	.00	.034*
	Mean age	(31–49 years; *M* = 39.8)	45	−.01	−0.03 to 0.00	.01	.131
	Mean intake score	(7.9–18.8; *M* = 15.0)	58	.02	0.00 to 0.04	.01	.015*
	Mean number of sessions	(3–16 sessions; *M* = 6.7)	42	.03	0.01 to 0.05	.01	**.001****
GAD-7	Gender (% female)	(0–100%; *M* = 59.1)	49	.00	0.00 to 0.00	.00	.079
	Mean age	(31–49 years; *M* = 39.7)	43	−.01	−0.03 to 0.00	.01	.061
	Mean intake score	(3.7–18.3; *M* = 13.5)	52	.07	0.04 to 0.09	.01	<**.001*****
	Mean number of sessions	(3–16 sessions: *M* = 6.6)	38	.03	0.01 to 0.05	.01	.015*
WSAS	Gender (% female)	(0–100%; *M* = 54.9)	22	.00	0.00 to 0.00	.00	.283
	Mean age	(31–49 years; *M* = 39.3)	20	.00	−0.02 to 0.02	.01	.689
	Mean intake score	(14.8–24.5; *M* = 19.3)	22	−.01	−0.05 to 0.04	.02	.751
	Mean number of sessions	(4–16 sessions; *M* = 6.7)	18	.02	−0.01 to 0.05	.01	.163

Note. CI = confidence interval; GAD-7 = Generalized Anxiety Disorder Scale-7; *k* = number of comparisons; *M* = mean; PHQ-9 = Patient Health Questionnaire-9; *SE* = standard error; WSAS = Work and Social Adjustment Scale.

*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 (in bold).

IAPT for deaf patients (Young *et al.*, 2017), whilst another explored differences in outcomes based on sexual orientation (Rimes *et al.*, 2018).

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 2 (low-intensity GSH) or step 3 (high-intensity) interventions were delivered. Overall, $n = 21$ studies reported on interventions delivered at either step 2 or step 3, whilst only 'step 2' interventions were evaluated in $n = 17$ studies, and only 'step 3' interventions were reported in $n = 17$ studies (step of care was not specified in $n = 10$ studies). At step 2, the mean reported intake scores were PHQ-9 = 13.48, GAD-7 = 12.06, and WSAS = 16.87. At step 3, the mean reported intake scores were PHQ-9 = 15.26, GAD-7 = 13.04, and WSAS = 18.40. Pre-treatment symptom severity was therefore similar for step 2 and step 3, PHQ-9: $t(14) = -1.388$, $p = .187$; GAD-7: $t(11) = -0.529$, $p = .607$; WSAS: $t(3) = -0.777$, $p = .494$. Where specific interventions were named, this ranged in terms of intensity and type. CBT was specified in six studies (Binnie & Boden, 2016; Branson, Shafran, & Myles, 2015; Highfield *et al.*, 2016; McDevitt-Petrovic *et al.*, 2018; Pybis, Saxon, Hill, & Barkham, 2017; Saxon, Firth, & Barkham, 2017). One study described using a treatment manual for CBT with psychosis (Jolley *et al.,* 2015), and another a treatment manual for trauma-informed CBT (Murray, 2017). Five studies investigated group interventions (n = 6 studies) including two analysing outcomes for the 'Stress Control' psychoeducational group intervention at step 2 (Burns, Kellett, & Donohoe, 2015; Delgadillo, Kellett, et al., 2016), one high-intensity behavioural activation group (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 (Wright & Abrahams, 2015), couples' therapy (Baucom *et al.*, 2018), and a GSH version of cognitive analytic therapy delivered at step 2 (Meadows & Kellett, 2017).

Follow-up

There were four studies that had a post-treatment follow-up period, and this ranged from 4 to 52 weeks.

Risk of bias assessment

Overall, the majority of studies (58%) were rated as having low risk of bias, 30% had medium risk (30%), and 12 % had high risk. Study quality was particularly affected by the lack of follow-up data.

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 and reported across the studies, this resulted in different numbers of studies 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; and 19 used the WSAS as an outcome measure. Some of the included studies reported

more than one 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 ESs were included as independent samples. This was consistently implemented across the whole meta-analysis and subgroup analyses. For example, in the paper by Delgadillo, Dawson, et al. (2017) a separate ES is 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 limited number of studies reporting pre-post outcomes also included 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.

Primary meta-analysis

Results for the PHQ-9 summarizing outcomes from 636,734 patients (mean n = 9,796; median n = 619) across 46 studies (n = 65 independent samples) are reported in Figure 2. The overall combined pre-post treatment PHQ-9 ES was large (d = 0.87, 95% CI [0.78-0.96], p < .0001, 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 plot asymmetry (p = .083). The fail-safe N analysis indicating the number of non-significant studies needed to be published to overturn the findings to a small clinically non-significant effect was 97. Results for the GAD-7 included outcomes from 598,166 patients (mean n = 9,969; median n = 541) across 41 studies (n = 60 independent samples) and are reported in Figure 4. The overall combined pre-post treatment GAD-7 ES 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.The overall combined pre-post treatment GAD-7 ES 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 some evidence for 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 = 1,351) from 19 studies (n = 24 independent samples) and are summarized in Figure 6. The overall combined WSAS ES 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 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 ES to a small clinically non-

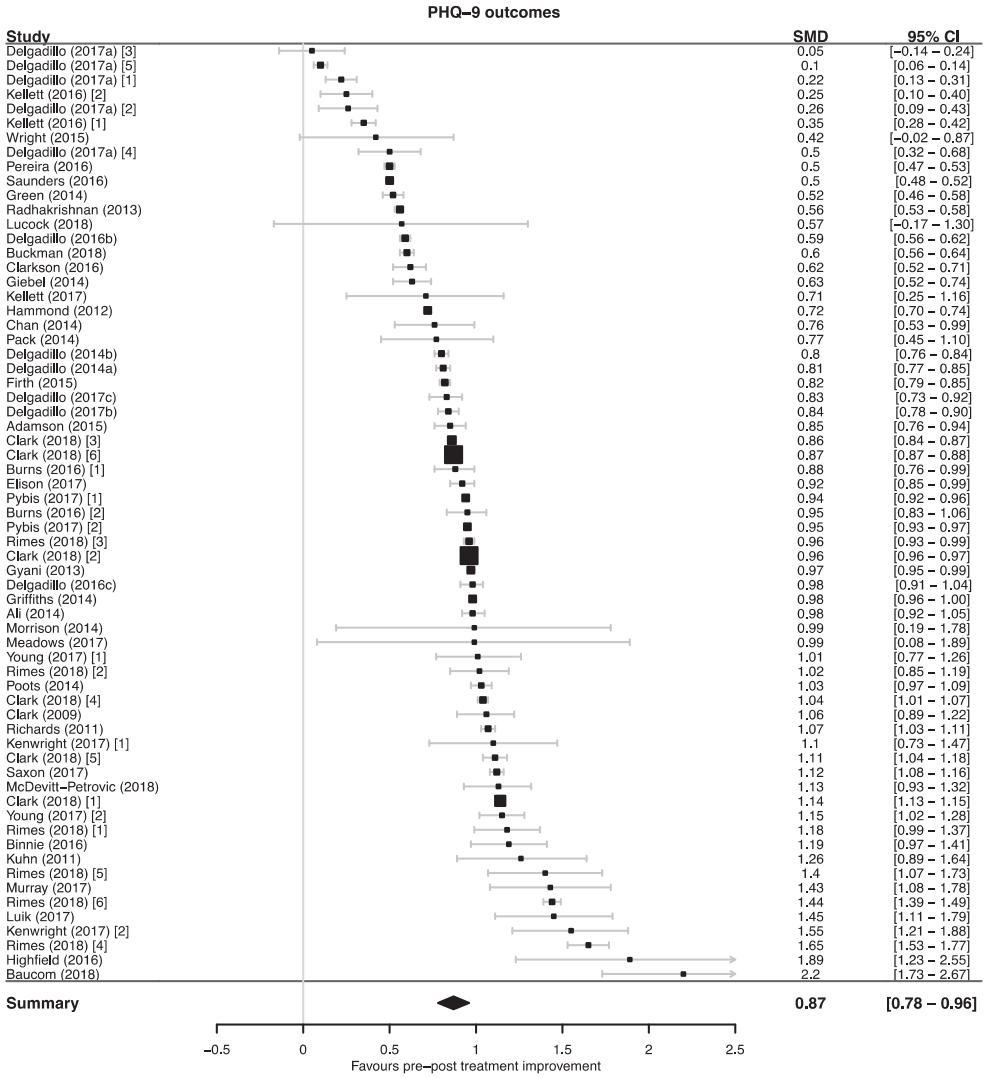

Figure 2. Forest plot of pre-post PHQ-9 independent samples' effect sizes and the pooled treatment effect.

significant pre-/post-improvement (d = 0.35); however, the rank correlation test was not significant (p = .572).

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 2). For PHQ-9 outcomes, significant variations in ES by subgroups were evident for type of methodology used, primary condition, step of care, and level of study bias. Completer analyses produced significantly larger ES than ITT analyses. Studies of primary

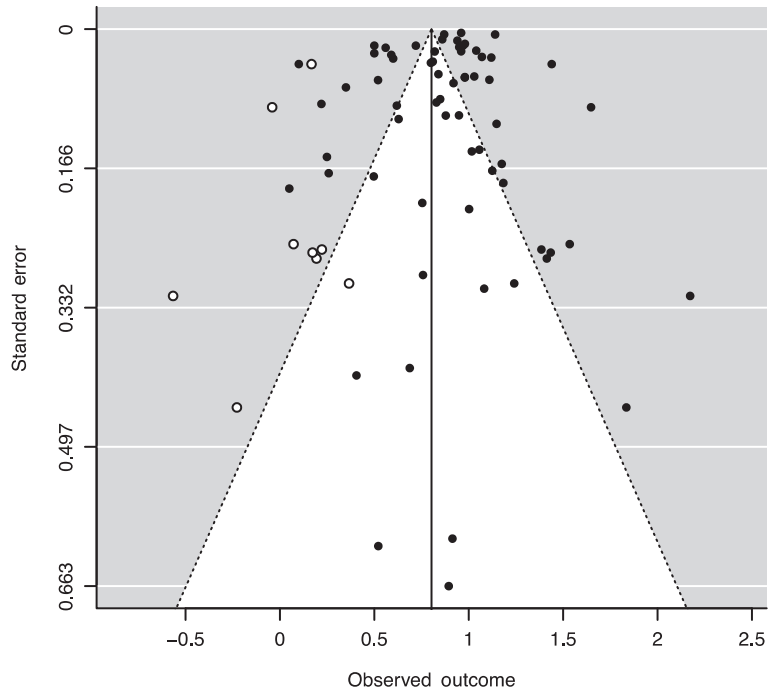

Figure 3. Funnel plot of the distribution of studies reporting pre-post PHQ-9 outcomes.

mental health conditions produced significantly larger effects than studies which included patients with a chronic physical illness as the primary condition. Studies with increased risk of bias produced larger treatment effects than studies with low risk of bias. Samples reporting outcomes for step 3 (high-intensity) interventions produced larger effects than those reporting outcomes for step 2 (low-intensity) interventions. However, the subgroup differences in the latter two comparisons were no longer significant after accounting for multiple testing. For GAD-7 outcomes, significant variations in ESs by subgroups were evident for type of methodology used (completer vs. ITT analysis), primary condition (mental health vs. physical illness), and risk of study bias, showing the same pattern as in the PHQ-9 outcomes. 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 was found for WSAS outcomes.

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 outcome measures (Table 3). For GAD-7 and WSAS outcomes, between-study variations in ESs 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 ES variations for both PHQ-9 and GAD-7 outcomes, with larger effects evident when there were a greater mean

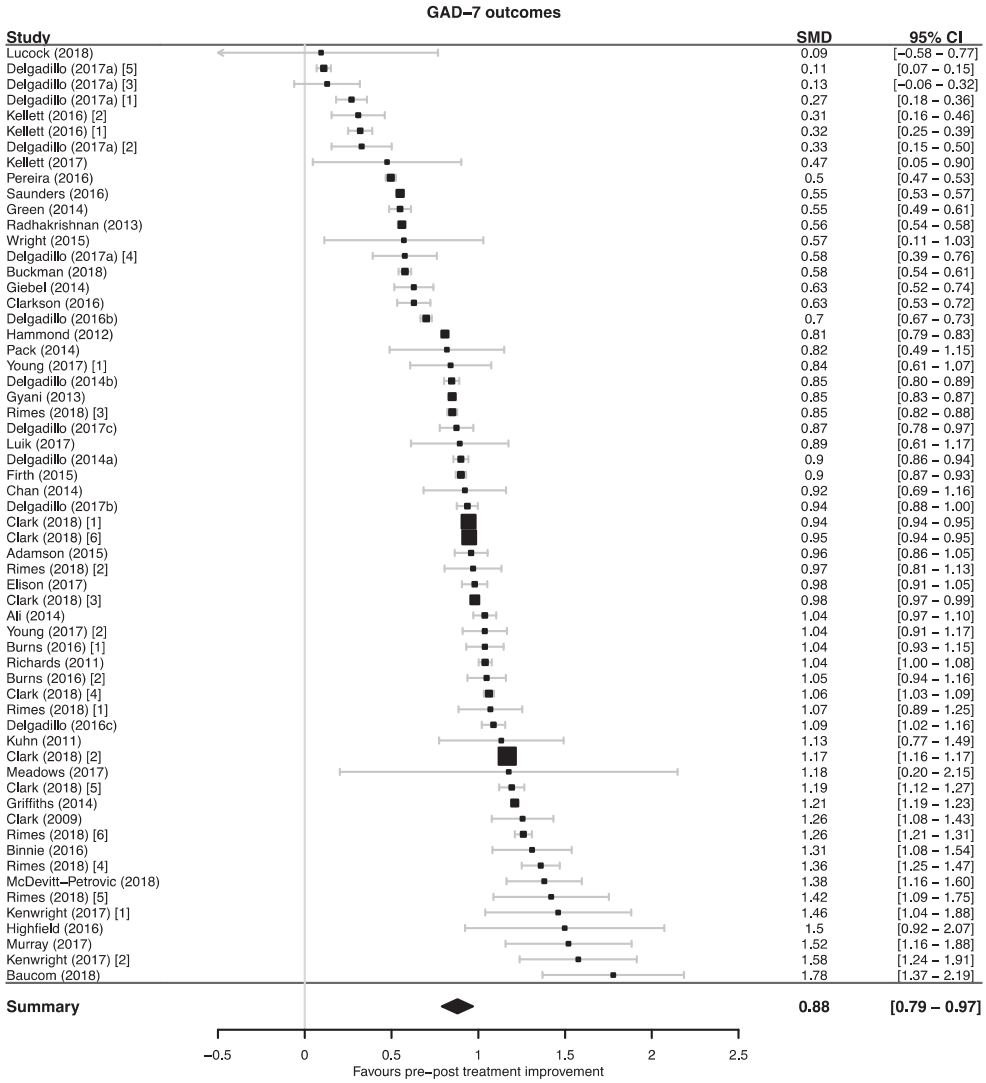

Figure 4. Forest plot of pre-post GAD-7 independent samples effect sizes and the pooled treatment effect.

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.

Sensitivity analysis excluding atypical studies

Sensitivity analyses investigated the aggregated ES 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 eight studies excluded on this basis. The excluded studies focused on samples of male offenders

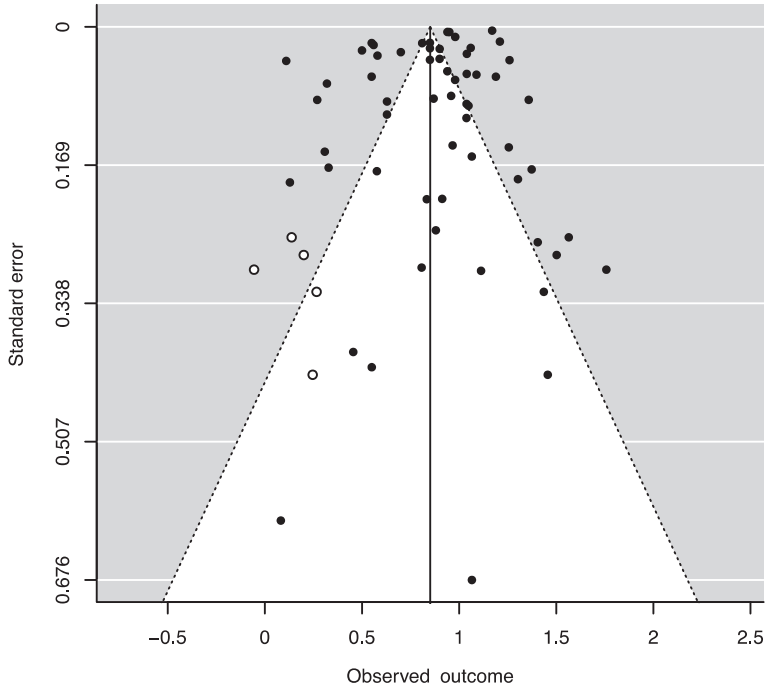

Figure 5. Funnel plot of the distribution of studies reporting pre-post GAD-7 outcomes.

(Adamson, Gibbs, & McLaughlin, 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 & Abrahams, 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 completed with the atypical studies excluded. Overall, and in comparison with the primary meta-analysis, there was only a minimal difference in the ES found in the sensitivity analyses. With regard to the PHQ-9, 57 separate comparisons contributed to the analysis producing a moderate-to-large ES of $d = 0.85$ (95% CI [0.80–0.90]; $p < .0001$, NNT = 2.19). There was still 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 [0.81–0.94]; $p < .0001$, 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 [0.48–0.62]; $p < .0001$, 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 (reported in the Tables S1–S3). Overall, this indicates that findings from the primary meta-analyses were stable and robust to sample selection across sub-group analyses.

Discussion

This systematic review has identified and synthesized all available, peer-reviewed, practice-based evidence generated by the IAPT programme – an initiative originally

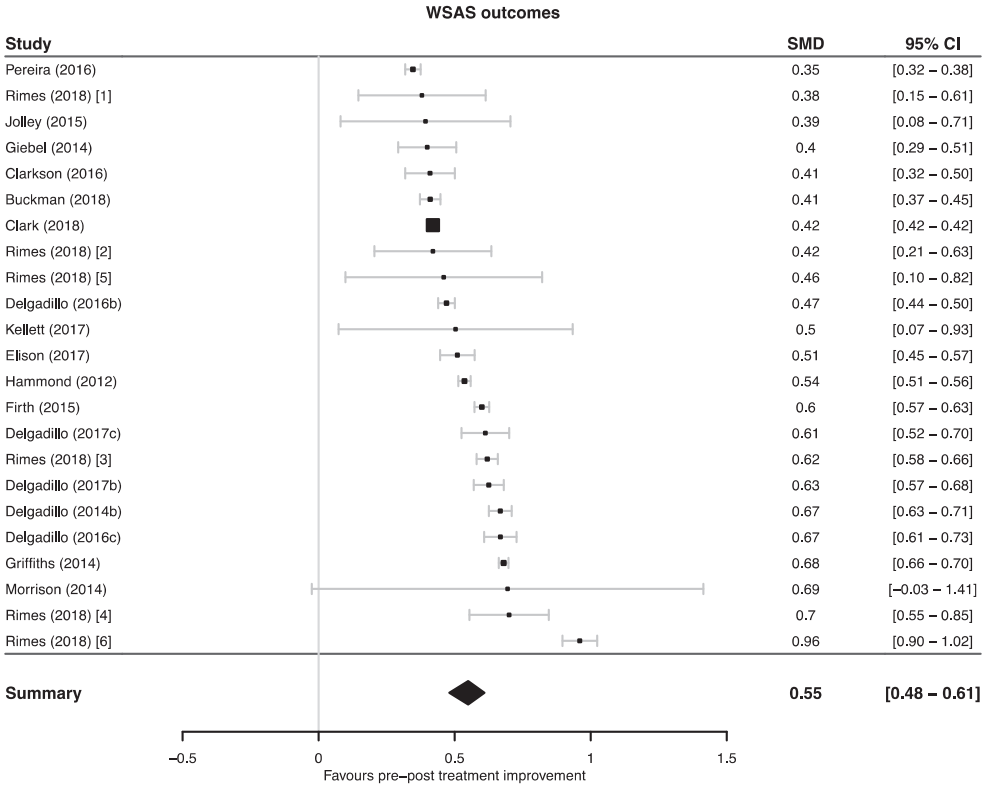

Figure 6. Forest plot of pre-post-WSAS independent samples effect sizes and the pooled treatment effect.

designed to increase rapid access to evidence-based psychological treatments for those experiencing common mental disorders (Clark, 2011; Clark *et al.*, 2009). The narrative review summarized $n = 60$ studies that varied markedly in terms of the methods used, samples studied, and outcomes analysed. The meta-analysis aimed to quantify the overall impact of IAPT interventions using standardized outcome measures, including data from over 600,000 patients. RCTs were excluded from this review in order to gain a better understanding of 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 results from the primary meta-analysis found large pre-post treatment effect sizes for reductions in depression and anxiety, with a medium effect regarding improvements in work and social adjustment. The GAD-7 effect 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-treatment outcomes on disorder-specific measures were large and, when benchmarked against the outcomes achieved in RCTs, were equivalent. The PHQ-9 effect mirrors the Thimm and Antonsen (2014) meta-analysis of the treatment of depression in routine practice, in that the ES at post-treatment was large ($d = 0.97$), and 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

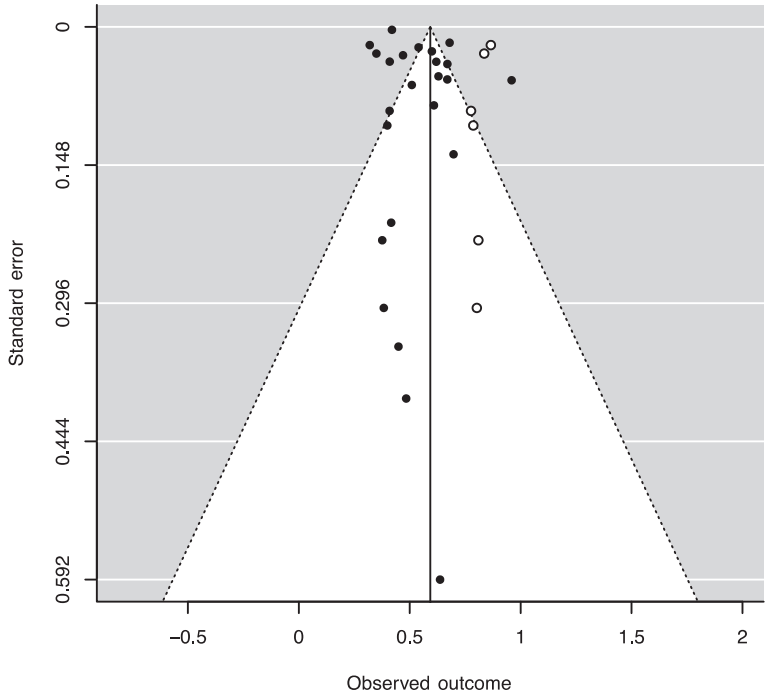

Figure 7. Funnel plot of the distribution of studies reporting pre-post WSAS outcomes.

publication bias (for GAD-7 outcomes) so results should be interpreted cautiously. The ES reported here therefore complements the recovery rates that are routinely reported by services (Clark, 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.

Moderator analyses

Studies using ITT analyses were compared with completer analyses (COM), which is an important and well-known methodological distinction (Kyrios, Hordern, & Fassnacht, 2015). ITT methods are recommended to minimize bias (Ranganathan, Pramesh, & Aggarwal, 2016), whereas COM tends to increase the rate of Type I errors (Fergusson, 2002). The ESs in COM studies were larger than those using ITT analysis across both anxiety and depression outcomes, and this is further 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 effect sizes observed for low and high intensity interventions for depression; however, these were no longer significant after accounting for multiple testing. Although differences between low and high intensity interventions were not significant for anxiety outcomes and functional impairment, there was a pattern of larger effects for high intensity interventions. This may have been due to the fact that when intake scores were assessed, there were no differences in initial assessment scores between the step 2 and step 3 studies. Psychological well-being practitioners delivering low-intensity interventions in IAPT 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, ESs may possibly be attenuated in some patients with more complex problems, where the skill level of the practitioner or the content of the intervention may be insufficient. 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, Barkham, Kellett, & Saxon, 2015). Recent studies suggest that 'complex cases' tend to have poor treatment outcomes when they are initially allocated to and receive low-intensity therapies, compared to high-intensity interventions (Delgadillo, Huey, *et al.*, 2017; Delgadillo, Moreea, & Lutz, 2016). 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, Moreea, *et al.*, 2016; Firth, Barkham, *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 seven sessions and that patients that move to recovery attend eight sessions on average (NHS England, 2019).

Study limitations

The absence of any control comparators means that the observed effects may be confounded by statistical phenomena such as regression to the mean and/or a possible natural recovery phenomenon (Posternak & Miller, 2001; Whiteford *et al.*, 2013). The lack of any indices of treatment fidelity, integrity, or competency in the studies raises uncertainty as to whether the interventions described were actually delivered as intended. The moderate rate of agreement concerning risk of bias ratings could have created unreliable treatment effect estimates in the meta-analysis (Armijo-Olivo *et al.*, 2014). The lack of precision in the studies regarding the specificity of low- and high-intensity interventions means that there was insufficient granularity in the descriptions of the interventions. There were relatively fewer purely low-intensity or high-intensity outcome studies for inclusion, and this weakened the specificity of the moderator analyses conducted. The lack of studies with adequate post-treatment follow-up data means that the durability of IAPT interventions is still open to question.

Research, policy, and clinical implications

In order to continue to improve our understanding of the effects of routinely delivered interventions, there is a need for the following: (1) studies analysing outcomes on other disorder-specific measures; (2) studies describing the interventions in greater detail; (3) consistent use of measures of treatment fidelity and competency; (4) studies investigating moderators and mediators of depression and anxiety outcomes; (5) studies collecting longer-term follow-up outcome data; (6) more consistent reporting of dropout rates; and (7) studies modelling and exploring variability between therapists/services/regions. Future IAPT studies should apply ITT analyses and report the percentage of patients treated at each step, the stepping up rate, the dropout rate, pre- and post-treatment means (*SD*s), and ESs on the standard IAPT outcome measures as well as the disorder-specific outcome measures used in routine care.

In terms of the policy implications, the following are of note: (1) the commissioning of routine follow-up support post-treatment, (2) identifying numbers of patients that are re-referred for IAPT treatment; and (3) open access to routinely collected patient-level IAPT data sets, to enable research to keep pace with the rapidly shifting IAPT policy context. National performance reports could be improved through the commissioning of rigorous meta-analytic evaluations, as exemplified in this study. In addition, it is clear that clinical outcomes are attenuated in populations with chronic and long-term illnesses and multidisciplinary care is advisable for this population based on the wider evidence base (e.g., see Delgadillo, Dawson, *et al.*, 2017). Furthermore, the extent to which the effects of IAPT interventions endure over time is largely unknown, and the little available data on this topic indicate that relapse after low-intensity interventions is likely to be very common (Ali *et al.*, 2017). A major area for improvement is the consistent implementation of evidence-based relapse prevention support, such as booster sessions (Gearing, Schwalbe, Lee, & Hoagwood, 2013) or mindfulness-based relapse prevention (Kuyken *et al.*, 2016). A promising development in this regard concerns telephone-delivered relapse prevention support which could be implemented at low cost to support IAPT patients to maintain their improvement after the acute phase of therapy (Lucock *et al.*, 2018) and during the first 6 months after therapy which is known to be the time of highest risk of relapse (Ali *et al.*, 2017).

This broad review of routinely delivered IAPT interventions has some implications for clinical practice. First, the expansion of high-intensity treatment options (e.g., provision of interpersonal psychotherapy, dynamic interpersonal psychotherapy, person-centered experiential counselling, and couples therapy for depression) has not been mirrored for low-intensity interventions which are mainly based on CBT principles. An expansion of other evidence-based low-intensity treatment options could provide greater choice for the highly heterogeneous clinical populations treated by IAPT services (Meadows & Kellett, 2017). There is increasing evidence to support stratified models of treatment matching for more complex cases, who evidently have higher dropout rates and poorer outcomes when offered very brief interventions. The original aim of the IAPT programme was to increase access to evidence-based talking treatments and there is evidence that large numbers are being treated annually, and that recovery rates are slowly increasing and achieving the 50% target (IAPT, 2019). There is, however, considerable room for improvement, particularly for patients who do not attain clinically significant improvement and who may find themselves in a 'revolving door' scenario of repeated treatment episodes (Cotton, 2019). There is also evidence to suggest that a considerable proportion (~30%) of IAPT patients have complex presentations (e.g., severe symptoms, comorbidity, socioeconomic adversity, and personality disorder traits), and they derive less benefit from routinely delivered interventions (Delgadillo, Huey, *et al.*, 2017). It is also evident that some complex cases do not benefit from low-intensity interventions, and therefore identifying complex cases early and signposting to high-intensity interventions is an important area for future development.

Conclusion

The IAPT programme is a notable example of psychological public health care transformation informed by scientific evidence (Clark *et al.*, 2018). Analysis of the evidence accumulated over the last 10 years supports the effectiveness of the IAPT programme and also demonstrates that innovative research and practice development have flourished within this context. A huge amount of investment has occurred to enable

and to maintain the IAPT programme and this has been achieved via mental health service infrastructure change, human resource investment in recruiting a new therapies workforce and overall organisational culture development/change. This transformation of the landscape of psychological services for people with anxiety and depression in the United Kingdom has served as a model for similar developments in other countries. This review has demonstrated that the systematic routine outcome monitoring implemented at scale in the IAPT programme also has huge scientific potential (Clark *et al.*, 2018).

Conflicts of interest

All authors declare no conflict of interest.

Author contributions

Stephen Christopher Kellett, D Clin Psy (Conceptualization; Methodology; Supervision; Writing – original draft; Writing – review and editing); Sarah Wakefield (Data curation; Formal analysis; Writing – original draft; Writing – review and editing); Jaime Delgadillo (Conceptualization; Methodology; Supervision); Mel Simmonds-Buckley (Formal analysis; Writing – review and editing); Dan Stockton (Formal analysis); Abigail Bradbury (Formal analysis).

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Adamson, V., Gibbs, M., & McLaughlin, D. (2015). 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*, 185–201. https://doi.org/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*, e99658. https://doi.org/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. https://doi.org/10.1016/j.brat.2017.04.006

Armijo-Olivo, S., Ospina, M., da Costa, B. R., Egger, M., Saltaji, H., Fuentes, J., . . . Cummings, G. G. (2014). Poor reliability between Cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. *PLoS One*, *9*, e96920. https://doi.org/10.1371/journal.pone.0096920

Baker, T. B., McFall, R. M., & Shoham, V. (2009). Current status and future prospects of clinical psychology: Toward a scientifically principled approach to mental and behavioral health. *Psychological Science in the Public Interest*, *9*, 67–103. https://doi.org/10.1111/j.1539-6053.2009.01036.x

Baucom, D. H., Fischer, M. S., Worrell, M., Corrie, S., Belus, J. M., Molyva, E., & Boeding, S. E. (2018). Couple-based intervention for depression: An effectiveness study in the National Health Service in England. *Family Process*, *57*, 275–292. https://doi.org/10.1111/famp.12332

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

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

Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: Access, effectiveness and efficiency. Narrative literature review. *British Journal of Psychiatry*, *186*, 11–17. https://doi.org/10.1192/bjp.186.1.11

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*, 101–114. https://doi.org/10.1017/s1352465817000522

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

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*, 513–527. https://doi.org/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*, 431–443. https://doi.org/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*, 747–759. https://doi.org/10.1017/s1352465813000544

Cheston, R., & Howells, L. (2016). A feasibility study of translating "Living Well with Dementia" groups into a Primary Care Improving Access to Psychological Therapy service. *Dementia*, *15*, 273–278. https://doi.org/10.1177/1471301215582104

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry*, *23*, 375–384. https://doi.org/10.3109/09540261.2011.606803

Clark, D. M. (2019). *IAPT at 10: achievements and challenges*, England: NHS. https://www.england.nhs.uk/blog/iapt-at-10-achievements-and-challenges

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. *Lancet*, *391*, 679–686. https://doi.org/10.1016/S0140-6736(17)32133-5

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*, 910–920. https://doi.org/10.1016/j.brat.2009.07.010

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*, 227–235. https://doi.org/10.1002/nop2.57

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

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

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

Cromarty, P., Drummond, A., Francis, T., Watson, J., & Battersby, M. (2016). NewAccess for depression and anxiety: Adapting the UK Improving Access to Psychological Therapies Program

across Australia. *Australasian Psychiatry*, *24*, 489–492. https://doi.org/10.1177/1039856216641310

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*, 429–430. https://doi.org/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. https://doi.org/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*, 835–853. https://doi.org/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. https://doi.org/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*, 16–30. https://doi.org/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*, 114–130. https://doi.org/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. https://doi.org/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. https://doi.org/10.1016/j.brat.2017.09.011

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. https://doi.org/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". *British Medical Journal Open*, *7*, e016392. https://doi.org/10.1136/bmjopen-2017-016392

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

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

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. https://doi.org/10.1016/j.brat.2015.04.001

Gearing, R. E., Schwalbe, C. S. J., Lee, R. H., & Hoagwood, K. E. (2013). The effectiveness of booster sessions in CBT treatment for child and adolescent mood and anxiety disorders. *Depression and Anxiety*, *30*, 800–808. https://doi.org/10.1002/da.22118

Giebel, C. M., Clarkson, P., & Challis, D. (2014). Demographic and clinical characteristics of UK military veterans attending a psychological therapies service. *Psychiatric Bulletin*, *38*, 270–275. https://doi.org/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. https://doi.org/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 Practitioners (PWPs): A multilevel modelling and mixed methods analysis. *Behaviour Research and Therapy*, *63*, 43–54. https://doi.org/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*, 28–35. https://doi.org/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*, 597–606. https://doi.org/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*, e42916. https://doi.org/10.1371/journal.pone.0042916

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.

Improving Access to Psychological Therapies. (2018). *Pathway for people with long-term physical health conditions and medically unexplained symptoms*. Mental Health Policy and Strategy Team. Retrieved from https://www.england.nhs.uk/mental-health/resources/

Improving Access to Psychological Therapies. (2019). *Talking therapies: New statistics show an increase in referrals, numbers starting treatment and recovery rates during 2018–19*. NHS Digital. Retrieved from https://digital.nhs.uk/news-and-events/latest-news/iapt-2018-19

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. https://doi.org/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*, 401–418. https://doi.org/10.1017/s1352465816000540

Kellett, S., Simmonds-Buckley, M., Limon, E., Hague, J., Hughes, L., Stride, C., & Millings, A. (2020). Defining the assessment and treatment competencies to deliver low intensity cognitive behavior therapy: A multi-centre validation study. *Behavior Therapy* (in press). https://doi.org/10.1016/j.beth.2020.01.006

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*, 433–447. https://doi.org/10.1017/s1352465817000200

Knapstad, M., Nordgreen, T., & Smith, O. R. (2018). Prompt mental health care, the Norwegian version of IAPT: Clinical outcomes and predictors of change in a multicenter cohort study. *BMC Psychiatry*, *18*(1), 260. https://doi.org/10.1186/s12888-018-1838-0

Kobori, O., Nakazato, M., Yoshinaga, N., Shiraishi, T., Takaoka, K., Nakagawa, A., . . . Shimizu, E. (2014). Transporting Cognitive Behavioral Therapy (CBT) and the Improving Access to Psychological Therapies (IAPT) project to Japan: Preliminary observations and service evaluation in Chiba. *The Journal of Mental Health Training, Education and Practice*, *9*, 155–166. https://doi.org/10.1108/JMHTEP-10-2013-0033

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

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

Kuyken, W., Warren, F. C., Taylor, R. S., Whalley, B., Crane, C., Bondolfi, G., . . . Dalgleish, T. (2016). Efficacy of mindfulness-based cognitive therapy in prevention of depressive relapse: An individual patient data meta-analysis from randomized trials. *JAMA Psychiatry*, *73*, 565–574. https://doi.org/10.1001/jamapsychiatry.2016.0076

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*, *3*, 181–190. https://doi.org/10.1016/j/ijchp.2015.07.003

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*, *20*, e10113. https://doi.org/10.2196/10113

Lucock, M., Barkham, M., Donohoe, G., Kellett, S., McMillan, D., Mullaney, S., . . . Delgadillo, J. (2017). The role of Practice Research Networks (PRN) in the development and implementation of evidence: The Northern improving access to psychological therapies PRN case study. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*, 919–931. https://doi.org/10.1007/s10488-017-0810-5

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*, 626–632. https://doi.org/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 standardized outcome measures. *Behavioural and Cognitive Psychotherapy*, *45*, 91–96. https://doi.org/10.1017/S1352465816000369

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*, 761–767. https://doi.org/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*, 266–284. https://doi.org/10.1017/s1352465816000485

Mental Health Policy Team (2018). *The improving access to psychological therapies manual*, London: National Collaborating Centre for Mental Health.

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*, 102–111. https://doi.org/10.1017/s1754470x13000044

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*, e1000097. https://doi.org/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*, e15. https://doi.org/10.1017/s1754470x14000221

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. *British Journal of Psychiatry*, *180*, 461–464. https://doi.org/10.1192/bjp.180.5.461

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

Naeem, F., Pikard, J., Rao, S., Ayub, M., & Munshi, T. (2017). Is it possible to provide low-intensity cognitive behavioral treatment (CBT Lite) in Canada without additional costs to the health system? First-year evaluation of a pilot CBT Lite program. *International Journal of Mental Health*, *46*, 253–268. https://doi.org/10.1080/00207411.2017.1345039

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. (2019). *Talking therapies: New statistics show an increase in referrals, numbers starting treatment and recovery rates in 2018–209*. Retrieved from https://digital.nhs.uk/news-and-events/latest-news/iapt-2018-19

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*, 157–159. https://doi.org/10.2307/1164923

Pack, S., & Condren, E. (2014). An evaluation of group cognitive behaviour therapy for low self-esteem in primary care. *The Cognitive Behaviour Therapist*, *7*, e7. https://doi.org/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*, 691–704. https://doi.org/10.1007/s10488-016-0747-0

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*, e453–e459. https://doi.org/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*, 198–204. https://doi.org/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*, 139–146. https://doi.org/10.1016/S0165-0327(00)00304-9

Prina, A. M., 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. https://doi.org/10.1016/j.brat.2014.03.008

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), 215. https://doi.org/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*, 37–45. https://doi.org/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. *Perspective Clinical Research*, *7*, 144–146. https://doi.org/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*, 51–60. https://doi.org/10.1016/j.jad.2011.03.024

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. https://doi.org/10.1016/S0140-6736(16)31140-0

Rimes, K. A., Broadbent, M., Holden, R., Rahman, Q., Hambrook, D., Hatch, S. L., & Wingrove, J. (2018). Comparison of treatment outcomes between lesbian, gay, bisexual and heterosexual individuals receiving a primary care psychological intervention. *Behavioural and Cognitive Psychotherapy*, *46*, 332–349. https://doi.org/10.1017/s1352465817000583

Roth, A. D., & Fonagy, P. (2005). *What works for whom: A critical review of psychotherapy research*. New York, NY: 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. https://doi.org/10.1016/j.jad.2016.03.011

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*, 705–715. https://doi.org/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*, 1136–1147. https://doi.org/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*, 1092–1097. https://doi.org/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. https://doi.org/10.1037/a0016032

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

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. https://doi.org/10.1017/s1754470x15000148

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 48.

Westen, D., & Morrison, K. (2001). A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *69*, 875–899. https://doi.org/10.1037/0022-006X.69.6.875

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*, 1569–1585. https://doi.org/10.1017/S0033291712001717

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*, 160–170. https://doi.org/10.1080/02668734.2015.1035740

Wroe, A. L., Rennie, E. W., Gibbons, S., Hassy, A., & Chapman, J. E. (2015). IAPT and long term medical conditions: What can we offer? *Behavioural and Cognitive Psychotherapy*, *43*, 412–425. https://doi.org/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*, 1–196. https://doi.org/10.3310/hsdr05240

Received 17 November 2019; revised version received 14 May 2020

Supporting Information

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

Table S1. Main findings of each study and quality assessment ratings.
Table S2. Subgroup analysis of pre-post treatment effects in the typical study sample (n = 8 atypical studies excluded).
Table S3. Meta-regression analysis of pre-post treatment effects in the typical study sample (n = 8 atypical studies excluded).