
Article:
Saxon, D. orcid.org/0000-0002-9753-8477, Broglia, E., Duncan, C. et al. (1 more author) (2024) Variability in treatment effects in an English national dataset of psychological therapies: the relationships between severity, treatment duration, and therapy type. Journal of Affective Disorders, 362. pp. 244-255. ISSN 0165-0327

https://doi.org/10.1016/j.jad.2024.06.115

Journal of Affective Disorders 362 (2024) 244–255

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Variability in treatment effects in an English national dataset of psychological therapies: The relationships between severity, treatment duration, and therapy type

David Saxon [a,*], Emma Broglia [b], Charlie Duncan [c], Michael Barkham [a]

[a] Clinical and Applied Psychology Unit, Department of Psychology, University of Sheffield, Sheffield S1 2LT, United Kingdom of Great Britain and Northern Ireland
[b] Department of Psychology, University of Sheffield, Sheffield S1 2LT, United Kingdom of Great Britain and Northern Ireland
[c] Research Unit, British Association for Counselling and Psychotherapy, BACP House, 15 St John's Business Park, Lutterworth, Leicestershire LE17 4HB, United Kingdom of Great Britain and Northern Ireland

ARTICLE INFO

Keywords:
(3–6) small effects
Outcomes
Depression
Cognitive behavioral therapy
Person-centered experiential therapy

ABSTRACT

Background: Comparisons between bona fide psychological therapies generally report no effects or small effects favoring cognitive behavioral therapy (CBT), suggesting that differences between therapy modalities are of little importance.
Methods: An observational, cohort study of patients (N = 11,116), drawn from the English Talking Therapies program. Patients presented with depression and were treated by CBT or person-centered experiential therapy (PCET), 67 % were female and the age range was 15–94. Multilevel modeling was used to identify variables associated with outcomes and CBT and PCET outcomes were compared dependent on the number of sessions patients attended and the severity of their depression at intake.
Results: Although overall a small effect size of 0.14 (0.10, 0.18) favored CBT, we found differing effect sizes depending on patient severity and the number of sessions they had attended at outcome. For non-clinical and moderately depressed patients no significant differences between therapies were found. For moderately-severe and severe patients there was a crossover in effectiveness with PCET more effective up to 5 or 6 sessions and CBT more effective thereafter. However, small effects in favour of PCET were unreliable. The largest effects were found for CBT for patients who were moderately-severe and had >10 sessions (d = 0.30, 0.12, 0.48) or were severe and had >12 sessions (d = 0.26 (0.02, 0.49).
Limitations: No data was available on therapy allocation decisions or therapists and the results may not be generalizable to non-NHS Talking Therapies services.
Conclusions: Small effects can mask important contextual differences between therapies and their study can inform allocation decisions.

1. Introduction

Substantial reviews of the psychological therapies outcome literature have concluded that, in general, bona fide therapies are effective and that for the majority of presenting problems, there are either no or relatively small differences between the efficacy or effectiveness of differing therapy models (see Barkham and Lambert, 2021; Lambert, 2013). The evidence derives not only from meta-analyses of randomized controlled trials (e.g., Cuijpers et al., 2020) but also from practice-based research utilizing data drawn from the English National Health Service (NHS) Talking Therapies for Anxiety and Depression program, previously known as the Improving Access to Psychological Therapies (IAPT) program (Clark, 2018). In the Talking Therapies program, the two most often delivered models of individual psychological therapies are cognitive behavioral therapy (CBT) and person-centered experiential therapy (PCET), previously known as Counselling for Depression or Person-centered experiential-counselling for depression (Murphy, 2019). Evidence from practice-based datasets (Saxon et al., 2017), national audit data (Barkham and Saxon, 2018; Pybis et al., 2017), and randomized controlled trials (Barkham et al., 2021) has shown a similar pattern of

* Corresponding author at: Clinical and Applied Psychology Unit (CAPU), Department of Psychology, University of Sheffield, 1 Vicar Lane, Sheffield S1 2LT, United Kingdom of Great Britain and Northern Ireland.
E-mail address: d.saxon@sheffield.ac.uk (D. Saxon).

https://doi.org/10.1016/j.jad.2024.06.115
Received 16 August 2023; Received in revised form 22 March 2024; Accepted 27 June 2024
Available online 28 June 2024

results, namely either negligible or relatively small differences between these two therapies although, in the latter case, results favored CBT at 12-months post-randomization.

The general similarities in efficacy and effectiveness of different psychological therapy types have led various commentators from Rosenzweig (1936) onwards (e.g., Luborsky et al., 2002; Wampold et al., 1997) to label this finding as the Dodo Bird verdict, derived from Lewis Carroll's fairy tale *Alice in Wonderland*, captured in the phrase "Everybody has won and all must have prizes" (Carroll, 1865/2015). Synonymous with such a portrayal of results is the line that even if there are differences, they are small and as such, are not important in the context of other factors that are deemed to carry greater impact, such as therapist effects (Wampold et al., 1997). Notwithstanding these views, other authors have argued that interpreting the literature as showing broad equivalence does not hold in the treatment of depression and anxiety and that, in fact, CBT, is indeed superior to other modalities (e.g., Tolin, 2010).

Contrary to these interpretations of the literature, a further view is that, while at the overall level of group means there is relatively little, if any, difference in the outcomes between differing modalities of therapy, the effects of averaging patient outcome data masks detectable differences between therapies in relation to such features as treatment duration. Although acceptance of the 'Dodo bird' verdict may render the comparative study of modalities redundant, the study of the *variability* between modalities, particularly in their relationships with other factors, can inform patient allocation to the most suitable therapy and improve outcomes. This 'knowledge gap' in how different therapy modalities interact with other factors was recognised in the recent National Institute for Health and Care Excellence (NICE) guidelines for depression (National Institute for Health and Care Excellence, 2022).

To address this 'knowledge gap', current developments in adopting precision methods in the psychological therapies aim to predict patient outcomes based on multiple factors in order to 'match' patients' profiles to the most suitable treatment (Cohen et al., 2021). Individually, these factors may have very small effects on patient outcomes but may interact or accumulate in some patients to have a more significant effect on their outcomes. For example, a study using machine learning methods found subgroups of patients in routine practice settings responded differently to PCET and CBT (Delgadillo and Gonzalez Salas Duhne, 2020). However, this and other research (e.g., Serbanescu et al., 2020) have only focused on the available baseline variables as possible contributing factors.

While precision methods and machine learning approaches aim to include all important factors, some, like the number of sessions the patient will attend before ending therapy (or whether they will leave prematurely or not), are mostly unknown at the start of therapy and are, therefore, difficult to include in 'a priori' outcome predictions. However, patient outcomes have often been modeled on the dose-effect curve in which number of sessions determines outcomes but with diminishing incremental returns over time (Howard et al., 1996). An alternative model suggests that there are individual patient differences and therapy ends when the patient has achieved a 'good enough' level of improvement and hence a function of the rate of individual change for a patient rather than the number of sessions received (Barkham et al., 2006).

In terms of guidance on the number of sessions administered, the 2009 NICE clinical guideline for depression recommended up to 16 sessions of PCET for non-severe depression and up to 16 sessions of CBT for all depression including severe and moderately–severe (National Institute for Health and Care Excellence, 2009). However, the updated 2022 guidelines listed PCET as a treatment for all depression and also recommended >16 sessions of either therapy where patients presented with comorbid health, or physical, or social problems (National Institute for Health and Care Excellence, 2022). However, in practice, a large proportion of patients complete therapy or drop-out after receiving fewer sessions, with a review reporting that the average number of sessions attended in NHS Talking Therapies services was 7 sessions (Wakefield et al., 2021).

The PRaCTICED trial, conducted in a single Talking Therapies service, compared PCET and CBT for moderate or severe depression (Barkham et al., 2021) and found a small non-significant advantage to PCET on most measures at 6 months post randomization, with a small but significant effect in the region of $g = 0.3$ in favour of CBT at 12 months post randomization. Also, at 6 months only the per-protocol analysis, where patient who had attended <4 sessions were excluded, showed a small non-significant advantage to CBT. These findings suggest that CBT had better outcomes when patients had more sessions and PCET had better outcomes when patients had fewer sessions. Furthermore, an audit of Talking Therapies data indicated a different pattern of outcomes between the two therapy modalities across sessions with PCET showing greater change than CBT when patients had attended up to 8 sessions and less change than CBT when patients had attended >8 sessions, thereby indicating a 'crossover' in effectiveness dependent on the number of sessions a patient had attended at the point of their final session (Pybis et al., 2017).

Against this background, the current study utilized a large, English national NHS Talking Therapies dataset comprising outcome data for CBT and PCET, to firstly, test the findings obtained from both the PRaCTICED trial and Pybis et al. (2017) audit with regard to the relationships between therapy modality, sessions attended at discharge, and outcome. Secondly, the aim was to extend this analysis to assess how patient severity level at intake may affect these relationships.

2. Method

The design was an observational cohort study of patient depression outcomes in NHS Talking Therapies services in England. Publicly available routinely collected data from all patients who started and ended treatment between 1st April 2016 and 31st March 2017 was provided by NHS Digital, the national provider of data for commissioners, analysts, and clinicians in NHS England. The data was obtained following a formal application and the signing of a Data Sharing Agreement (DSA) between NHS Digital and the University of Sheffield, which included a Data Protection Agreement ensuring the confidentiality of the patients and security of the data (DSA ID: DARS-NIC-85465-H1W9F-v0.8). The write-up followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines (von Elm et al., 2007).

2.1. Ethical statement

All procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The need for informed consent was waived as data has been collected from NHS Digital and is publicly available. All data provided to the research team was anonymised by the NHS Digital.

2.2. Setting

Data were provided from 201 Clinical Commissioning Groups (CCG) covering most of England and each representing a geographical area containing a number of GP clinics. CCGs commission the therapy provided by Talking Therapies services, which are attached to GP clinics. Talking Therapies operates a 'stepped-care' model with monitoring by GP at step 1, a short-term, low-intensity treatment by Psychological Wellbeing Practitioners (PWP) within Talking Therapies services at step 2 and a longer term, high-intensity therapy by a trained therapist, usually CBT or PCET, at step 3. Patients who have not improved following the low-intensity step 2 treatment are stepped-up to a high-intensity treatment by the PWP who assesses the patient and records the patient's main problem or concern and the high intensity therapy they are being stepped up to into an electronic system. The NICE

depression guideline may inform the PWP allocation decision, but patient choice, the availability of therapy, and avoidance of long waiting times may also influence therapy allocation.

2.3. Data and study sample

From a data sample of 635,618 therapy sessions, 113,245 high intensity treatment episodes, defined as a series of sessions of the same therapy type less than six weeks apart from each other, were identified. Single session episodes ($N = 37{,}736$, (33.3 %)) were excluded.

In order to maintain the independence of patients in the analysis, where patients had more than one episode of treatment, only the most recent episode was included ($N = 56{,}098$). A further 1048 were excluded as they did not have a pre and post therapy PHQ-9 score, and 40,584 were excluded as the primary problem was not described as depression. For the majority of these ($N = 27{,}866$) the primary problem was described as anxiety and for 7679 no primary problem was recorded. Of the 14,466 patients identified with depression, 11,904 (82.3 %) were treated by CBT therapists or PCET counselors. Of those excluded, 754 had episodes of treatment that contained more than one therapy modality, where the patient switched therapy or had concurrent therapies. We took an upper limit of 20 sessions as being consistent with the limit prescribed by the PRaCTICED trial. and therefore excluded patients who received >20 sessions ($N = 392$) as being atypical and likely to have more complex problems, and a further 396 were excluded as their CCG and GP identifiers were missing or conflicting, possibly as a result of commissioning region boundary changes. The numbers of patients excluded according to specific criteria are shown in a flowchart in Supplemental files. These inclusion criteria are likely to produce outcomes that cannot be directly compared to those published by Talking Therapies, where change across the care pathway rather than for specific therapy modalities are reported.

The study sample ($N = 11{,}116$) therefore comprised patients whose primary problem at assessment was described as depression, who received between 2 and 20 sessions in an episode of high-intensity treatment from either a CBT therapist or PCET counselor (identified by the 'primary role of the practitioner' variable) and had a self-report standardized outcome measure (PHQ-9) completed at the start and end of therapy.

2.4. Missing data

Because of the complexity of analysis, multiple imputations of missing data were not used. However, with a focus on sessions attended, therapy modality, severity, and outcome, a particular effort was made to minimize exclusions on these variables. Consistent criteria were used to identify an episode of treatment and the number of sessions attended and where the Primary Role of the therapist was missing, it was imputed based on a consistent description of the therapy type provided in each session. Also, in line with Talking Therapies practice, where either of the PHQ-9, GAD-7, or WSAS score at the last session was missing, the score at the previous session was used. Similarly, if the score at the first session was missing, the score at the second session was used as the baseline score. If this resulted in the same session score being used for the first and last scores, the patient was excluded.

2.5. Therapies

The two high-intensity psychological therapy modalities comprised cognitive behavioral therapy (CBT) and person-centered experiential therapy (PCET). The former comprises the mainstay of psychological therapy delivered within the Talking Therapies program and is based on a rich tradition of conceptual, clinical, and empirical research (for a summary, see Newman et al., 2021). The primary mechanisms focus on changing behaviors and maladaptive thoughts (i.e., cognitions) and schema (e.g., Beck, 2011).

The latter is the second most delivered high-intensity psychological therapy within the Talking Therapies program and was initially labeled 'counselling for depression' (Sanders and Hill, 2014) but has subsequently been renamed as person-centered experiential therapy (Murphy, 2019). This combination of terms captures the main theoretical components of the model, namely the person-centered tradition and experiential tradition (see Duffy et al., 2024; Elliott et al., 2021). The person-centered component focuses on the provision of a non-judgemental relationship while the experiential component promotes a style of process guiding by the therapist, thereby making it more action-oriented than traditional person-centered therapy with the focus on changing perceptions of self and situation as well as processing emotional distress.

2.6. Outcomes

The primary outcome was severity of depression, as measured by the Patient Health Questionnaire-9 (PHQ-9: Kroenke et al., 2001) at the last therapy session. PHQ-9 is a nine-item measure of depression severity, with each item scored 0–3 producing a total between 0 and 27. The PHQ-9 has good internal consistency (0.89) and test re-test (0.86) reliability (Kroenke et al., 2001), and is used routinely, at every therapy session in NHS Talking Therapies. PHQ-9 scores indicate the level of depression severity: none/minimal depression, scores 0–4; Mild depression, scores 5–9; moderate depression, scores 10–14; moderately severe depression, scores 15–19 and severe depression, scores 20–27.

To make comparisons of outcomes, effect sizes (Cohen's *d*) with 95 % confidence intervals (CI) were used and are reported with positive effect sizes indicating an advantage to CBT and negative effect sizes indicating an advantage to PCET. In addition, two outcome metrics are reported: (1) the percentage of patients making reliable change, based on the established cut-offs used by Talking Therapies (Talking Therapies Manual, 2018) (i.e., a pre-post change of 6 or more points on the PHQ-9); and (2) the percentage of patients meeting criteria for 'recovery' defined as making both reliable change and clinically significant improvement where a pre-therapy score of 10 or higher falls to 9 or less (i.e., a pre-post change from the clinical range [10–27] to the non-clinical range [(0–9]) (see Jacobson and Truax, 1991).

2.7. Covariates

The variables of interest were (1) therapy modality (CBT or PCET), (2) the number of sessions the patient had attended at the end of therapy, and (3) the severity as indicated by the PHQ-9 score at intake. Other variables were available as potential 'controls' in the modeling of these variables and included two other measures of severity at intake: the Generalised Anxiety Disorder-7 (GAD-7; Spitzer et al., 2006), and the Work and Social Adjustment Scale (WSAS; Mundt et al., 2002), which measure anxiety, and functional impairment, respectively. The GAD-7 is a seven-item measure of anxiety, scored 0–3, with a total of 0–21. It has good internal consistency (0.92) and test re-test (0.83) reliability (Kroenke et al., 2007). WSAS measures functioning across five domains: work, home management, social life, private leisure activities, and family relationships with adequate internal consistency (0.70) (Mundt et al., 2002).

In addition to the measures of patient severity, other available patient variables were: age, gender (male or female), ethnicity (White, Black, (African and Caribbean), Asian (Indian, Pakistani, and Bangladeshi), Mixed Ethnicity and Other), employment status (employed, unemployed: seeking work, student, long-term sickness, homemaker, not working: not seeking work or receiving benefits, unpaid voluntary work, or retired), Index of Multiple Deprivation quintile (IMD: Ministry of Housing, Communities, and Local Government, 2019) and referral source (self, general practitioner or other agency). The number of sessions (low intensity treatment and/or other IAPT treatment) the patient had in the six months prior to start of the current treatment episode was also available. CCGs, ($N = 201$), and GP practices ($N = 4141$) were also

available to control for 'organizational effects' in multilevel analysis.

Table 1 presents descriptive data for patients in the study sample. Although broadly similar on most variables, patients receiving CBT appear to be slightly more severe on assessment measures at intake and had a larger proportion of patients with moderate/severe and severe levels of PHQ-9 scores than PCET. More CBT patients self-referred and lived in less deprived areas than PCET and also tended to be younger, students, male, and from majority ethnic groups. Patients who were homemakers or retired were more likely to have received PCET, which probably reflects the patient gender and age differences between therapies. CBT patients generally attended more sessions and had a wider range (IQR) of sessions attended than PCET.

Fig. 1 shows the frequencies of the number of sessions attended for CBT and PCET. Generally, it shows smaller frequencies as the number of sessions increased although there are differences between treatments, most notably two sessions being the modal number of sessions for CBT while the mode for PCET was six sessions.

2.8. Analyses

In order to address the stated aims, we sought to develop regression models to assess firstly the significance of the interaction between the number of sessions the patient had attended at outcome and therapy modality in its association with PHQ-9 outcome score, and secondly the interactions between sessions, therapy modality, and intake depression severity. The final model was derived from the full sample ($N = 11{,}116$), however initially the full sample was split into two subsamples with a model developed on one half, then 'tested' on the other half (see Supplemental Files). This would assess the reliability of the included explanatory variables and inform model development for the full sample.

To split the full sample, it was stratified by intake severity level (defined by PHQ-9 cut-offs) within therapy modality, within each CCG and patients were randomized within each severity level to subsamples A or B. The two samples (described in Supplemental files), were broadly similar and each is similar to the full sample (Table 1).

Because of the nested structure of the data, with patients at level 1, clustered in GP practices at level 2 and GP practices clustered in CCGs at level 3, multilevel modeling (MLM) was used, with GP practice and CCG tested for significance in the model as random effects at level 2 and level 3 respectively (Snijders and Bosker, 2012). The effect each level had was calculated as the proportion of the total variance at that level, the intra-class correlation coefficient (ICC), and reported as a percentage. Models and the plots of predicted outcomes were produced using MlwiN software v 3.05 (Charlton et al., 2020) and used Iterative Generalised Least Squares (IGLS) procedures.

The development of the multilevel models was similar. Patient variables (intake severity and demographic factors) were entered first and retained if significant, based on z-scores. The number of sessions that each patient had attended at outcome was then included, initially as a linear term. Higher terms (quadratic, cubic etc) were assessed in turn and the reduction in −2*loglikelihood values compared against the chi-squared statistic for the additional degrees of freedom. The relationship between sessions and outcome was established when the additional term made no significant contribution (at 0.05 level) to model fit. Therapy modality and its interaction with sessions attended at outcome were then included followed by interactions between PHQ-9 intake severity and sessions attended.

Interactions with non-linear terms were included and tested sequentially, beginning with the linear term, then the cubic term, etc. Again, improvements in model fit were tested by comparing the reductions in the −2*loglikelihood values against the chi-squared statistic for the additional degrees of freedom. Finally, higher level (CCG and GP) effects were included and tested for significance and improvements in model fit.

Additional analysis considered intake depression severity as a

Table 1
Descriptives of study sample at baseline ($N = 11{,}116$).

Variable	All	CBT	PCET
N (%)	1116	8572 (77.1)	2544 (22.9)
PHQ-9 at first session: mean (sd)	15.5 (6.03)	15.7 (5.99)	14.8 (6.12)
PHQ-9 Severity n(%)			
None: score 0–4	488	352 (4.1)	136 (5.3)
Mild: score 5–9	1435	1043 (12.2)	392 (15.4)
Moderate: score 10–14	2783	2116 (24.7)	667 (26.2)
Moderate/severe: score 15–19	3220	2520 (29.4)	700 (27.5)
Severe: score 20–27	3190	2541 (29.6)	649 (25.5)
GAD-7: mean (sd)	12.8 (5.28)	13.0 (5.22)	12.4 (5.47)
Missing	2	2	0
WSAS: mean (sd)	20.8 (9.45)	21.4 (9.26)	18.9 (9.85)
Missing	520	340	180
Age mean (sd)	41.5 (13.68)	40.8 (13.66)	43.9 (13.46)
Sessions attended: Mean (sd)	8.6 (4.96)	9.1 (5.15)	6.9 (3.83)
Median	8	9	6
IQR	4–12	5–13	4–9
Number of contacts in the 6 months prior to start of treatment episode: mean (sd)	2.7 (3.01)	2.9 (3.12)	2.2 (2.53)
Sex n(%):			
Male	3655	2891 (33.8)	764 (30.1)
Female	7448	5670 (66.2)	1778 (69.9)
Missing	13	11	2
Ethnicity n(%):			
White	9077	7030 (84.1)	2047 (83.3)
Mixed Ethnicity	283	215 (2.6)	68 (2.8)
Asian	543	400 (4.8)	143 (5.8)
Black	361	275 (3.2)	86 (3.5)
Other	195	152 (1.8)	43 (1.7)
Not stated	362	291 (3.5)	71 (2.9)
Missing	295	209	86
IMD quintile n(%)			
1 Most deprived	2634	1876 (21.9)	758 (29.8)
2	2383	1841 (21.5)	542 (21.3)
3	2139	1671 (19.5)	468 (18.4)
4	1974	1580 (18.5)	394 (15.5)
5 Least deprived	1965	1587 (18.6)	378 (14.9)
Missing	21	17	4
Employment status n(%)			
Employed	5528	4337 (53.3)	1191 (51.2)
Unemployed SW RB	1297	1015 (12.5)	282 (12.1)
Student	423	367 (4.5)	56 (2.4)
Long-term sickness	1515	1149 (14.1)	366 (15.7)
Homemaker	607	453 (5.6)	154 (6.6)
Not working, Not SW, Not RB	334	266 (3.3)	68 (2.9)
Unpaid Voluntary work	54	45 (0.6)	9 (0.4)
Retired	701	500 (6.1)	201 (8.6)
Missing	657	440	217
Referral Source n(%):			
Self	5385	4370 (53.8)	1015 (41.3)
GP	4382	3101 (38.2)	1281 (52.1)
Other Agency	812	651 (8.0)	161 (6.6)
Missing	537	450	87

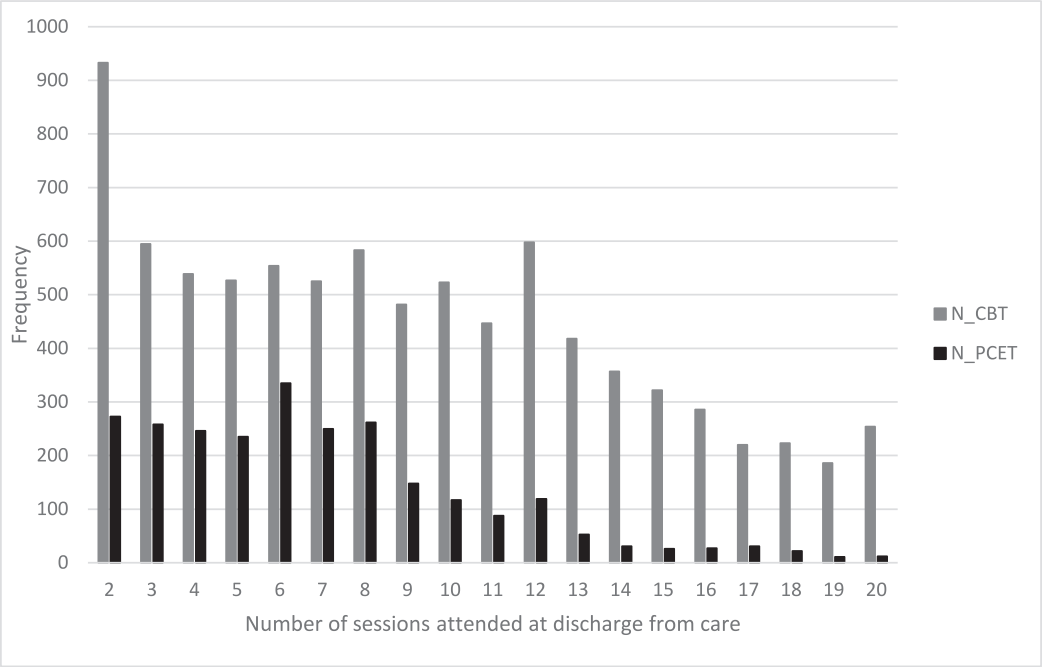

Fig. 1. Frequencies for Number of Sessions Attended for CBT and PCET.

categorical variable in the model based on the PHQ-9 cut-off scores. The two non-clinical categories, scores 0–4 and 5–9, were combined due to smaller sample sizes, other categories were as standard: score 10–14, moderate; 15–19, moderately severe and 20–27, severe. Finally, the full sample was split into severity level subgroups and the model for the full sample was applied to each.

3. Results

We first report briefly on the development of the model and testing using the split sample (presented in Supplemental files), before presenting the model produced by the full sample in more detail. Finally, the results of the models for PHQ-9 severity levels are reported. Plots of predicted outcomes derived from the models are presented where appropriate and comparative outcomes are reported.

3.1. Model development and testing

The model developed on half of the dataset (subsample A, $N = 5558$) found the three measures of intake severity, employment status, referral source, the number of sessions attended by the patient in the six months prior to the current episode, sessions attended at outcome for the current episode and therapy modality to be associated with PHQ-9 outcome. Including a cubic relationship between sessions attended and outcome, reduced the −2*loglikelihood value by 14.24, which compared to the chi squared statistic for the one additional degree of freedom (df) was significant ($p < 0.001$). Extending to the 4th power was not significant and did not improve model fit (−2*loglikelihood change of 0.076, (1df), $p = 0.783$). There were also significant interactions between therapy modality and sessions and intake severity on PHQ-9 and sessions but there was no significant interaction between therapy modality and intake severity. The interactions with sessions were only significant with its linear term.

Applying the model to sub-sample B ($N = 5558$) found the same variables to be associated with outcome with similar effects but with some differences. For example, student outcomes were not significantly different to employed patients' outcomes in sub-sample A ($p = 0.373$), but were significantly different in sub-sample B ($p = 0.041$). Importantly the interaction between therapy modality and sessions was significant in both models. A notable difference was that in sub-sample B, there was no significant cubic term for sessions attended at outcome (change in −2*loglikelihood = 0.282, (1df), $p = 0.595$). These differences may be due to the data splitting procedure and small numbers in some categories (i.e., where patients had attended a large number of sessions).

With regard to random effects, the GP practice was not significantly associated with outcome in either model (i.e., the variability between GP practices within CCGs had little effect on patient outcomes) and was not included. However, the CCG had a small but significant effect of 1.51 % for subsample A and 1.87 % for subsample B, indicating that the variability between CCGs had a small but significant effect on patient outcomes in both samples.

3.2. Full sample model

Informed by the sub-sample analysis, a model was developed using the full sample ($N = 11{,}116$). The variables included in the model are presented in Table 2.

The same variables were associated with outcome as in the subsamples but there were differences. Patients who were more severe on all three outcome measures, and were referred by an outside agency, or who had had more contacts with the service in the previous 6 months generally had poorer outcomes in all models. Employed patients had significantly better outcomes than all other categories except 'retired' who were not significantly different. 'Homemaker' was significant in this larger sample but non-significant in both subsamples while 'student' was significant here, and subsample B but not subsample A.

In the full sample, the cubic relationship between sessions and outcome was significant (as in subsample A) and its inclusion improved model fit (reducing the −2*loglikelihood by 5.90, (1 df), $p = 0.015$). Also, there was a significant interaction between intake score on PHQ-9 and the quadratic term for sessions. The model indicated greater intake severity of depression was associated with a poorer outcome but this was moderated by sessions. Having more sessions tended to reduce the negative effect of intake severity to some extent but this benefit was countered slightly at the extremes of sessions attended.

As with both subsamples, GP practice was not significant in the model. The CCG effect was 2.2 %, which was slightly larger than in either subsample.

Table 2
Full sample model: variables associated with PHQ-9 outcome.

Fixed effects	$N = 9581$		
	B	S.E	*p*-Value
cons	9.04	0.135	
First PHQ9-gm	0.39	0.017	<0.001
First GAD7-gm	0.12	0.015	<0.001
First WSAS-gm	0.10	0.008	<0.001
Referral source: Self			
General Practitioner	0.23	0.128	0.073
Outside Agency	0.99	0.219	<0.001
Employment status: Employed			
Unemployed: Seeking Work	0.92	0.176	<0.001
Student	0.62	0.277	0.036
Long-Term sick	1.53	0.169	<0.001
Homemaker	0.57	0.237	0.017
Not Working_not SW or RB	1.41	0.318	<0.001
Unpaid Voluntary Work	1.53	0.766	0.046
Retired	0.13	0.225	0.560
Number of contacts in the 6 months prior to start of treatment episode -gm	0.08	0.011	<0.001
N Sessions-gm	−0.31	0.025	<0.001
(N Sessions-gm)^2	0.04	0.003	<0.001
(N Sessions-gm)^3	−0.001	0.0005	0.001
Therapy: CBT			
PCET	0.53	0.152	<0.001
Therapy:PCET* N Sessions-gm	0.142	0.033	<0.001
First PHQ9-gm* N Sessions-gm	−0.028	0.002	<0.001
First PHQ9-gm* (N Sessions-gm)^2	0.001	0.0004	0.03

Fig. 2 plots the predicted outcomes for CBT and PCET patients derived from the full sample model. It shows PCET had better outcomes where patients had received 2–4 sessions although the differences were small and the 95 % CIs indicate no reliable difference. For patients who had received 5 sessions, the outcomes were identical while for 6–20 sessions, CBT had superior outcomes although the 95 % CIs indicate that the difference was only reliable from 10 to 20 sessions.

Plots of predicted outcomes for the subsamples are reported in Supplemental files. They too show the small superiority of PCET for patients who had 2–4 sessions, with a crossover in effectiveness at session 5 and a superiority of CBT from 6 sessions onwards. In subsample A, CBT was reliably more effective from 12 sessions onwards while for subsample B it was from 11 sessions. However, in the larger sample, the narrower confidence intervals resulted in CBT being reliably more effective where patients had received 10 or more sessions.

Considering these three treatment durations of therapy (i.e., 2–4 sessions, 5–9 sessions and 10–20 sessions), 2067 (24.1 %) of CBT patients had 2–4 sessions, 2671 (31.2 %) had 5–9 sessions, while 3834 (44.7 %) had 10–20 sessions. The respective frequencies for PCET were 777 (30.5 %), 1230 (48.3 %) and 537 (21.4 %), indicating that the percentage of patients in CBT who had the largest dose (i.e., 10–20 sessions) was over twice that of PCET.

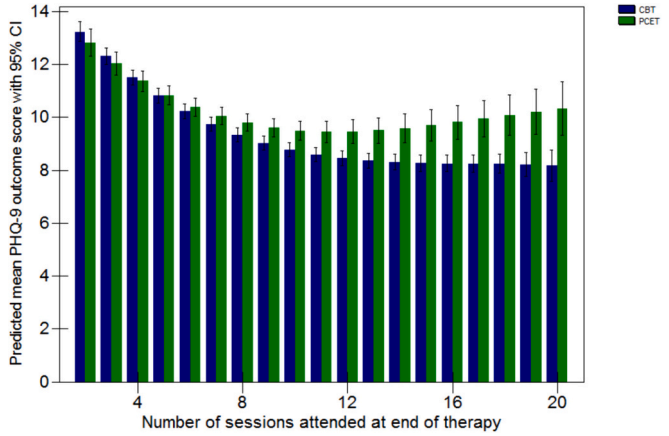

Fig. 2. Plot of predicted PHQ-9 outcomes derived from the full sample model.

Overall, the mean (SD) pre-post change on PHQ-9 for CBT was 5.2 (6.40) and 4.3 (6.01) giving a small effect in favour of CBT ($d = 0.14$, 0.10, 0.18) (which was similar to the effect sizes for GAD-7 and WSAS, see Supplemental files). For patients who had received 2–4 sessions, the mean (SD) pre-post change on PHQ-9 was 1.86 (4.71) for CBT and 2.31 (5.36) for PCET, an effect size (95 % CI) in favour of PCET of $d = -0.10$ (−0.02, −0.18). For patients who had 5–9 sessions the difference between therapies was negligible with changes of 4.89 (6.21) for CBT and 4.99 (6.01) for PCET, an effect size of $d = -0.02$ (−0.08, 0.05) For 10–20 sessions the change was 7.27 (6.53) for CBT and 5.45 (6.23) for PCET, an effect size in favour of CBT of $d = 0.28$ (0.19, 0.37).

Although there is a sizable amount of missing data (25 %) describing the end of therapy, there is an indication that CBT patients were less likely to complete therapy to an agreed ending and more likely to drop-out than PCET patients. Overall the therapy completion and drop-out rates for CBT 78.6 % and 15.6 % respectively compared with 82.0 % and 14.0 % for PCET. The remainder, 5.7 % for CBT and 4.0 % for PCET, were patients considered unsuitable for Talking Therapies who were referred onto other services. Considering the three therapy dose categories, where patients had received 2–4 sessions, the completion, drop-out, and referred-on rates were 59.5 %, 30.9 %, and 9.6 % for CBT patients and 66.5 %, 29.4 %, and 4.1 % for PCET patients, respectively. This suggests more patients experienced PCET as being a suitable treatment compared with CBT. For patients receiving 5–9 sessions, the rates were 76.8 %, 17.8 %, and 5.4 % for CBT compared with 86.9 %, 9.4 %, and 3.7 % for PCET, indicating that a smaller percentage had completed therapy and a larger percentage had dropped-out for CBT compared to PCET. Where patients had received 10–20 sessions, the figures were more similar: 86.7 %, 8.8 %, and 4.5 % for CBT and 88.3 %, 7.0 %, and 4.7 % for PCET.

The full sample model was based on 9581 of the 11,116 patients due to missing data for GAD-7, WSAS, employment status and referral source. In order to test the cross-over finding, these variables were excluded from the model. The resulting model and plot of predicted outcomes in presented in Supplemental files and they show the cross-over to be present, albeit from 7 sessions, and CBT to be reliably more effective for 10 sessions or more.

3.3. *Severity of depression at baseline*

The models reported above show that intake severity of depression, as a continuous measure (PHQ-9 score), was associated with outcome. Before producing models for each level of baseline depression severity subgroup, a model was developed replacing PHQ-9 as a continuous variable with the categorical severity variable (non-clinical, moderate, moderately severe, and severe) in the full sample (see Supplemental files). The same patient variables were included and it showed that compared to non-clinical scores, the three severity groups labeled moderate, moderately severe, and severe added 1.1, 2.7 and 5.2 PHQ-9 points, respectively, to outcome scores. However, these effects were reduced by 0.24, 0.38 and 0.49 of a PHQ-9 point, respectively, for each session above the average (8.6 sessions) that a patient had attended and increased by the same amounts where patients attended fewer than average sessions. The number of sessions attended was only significant as a linear term where it interacted with therapy modality and severity level. The quadratic and cubic terms for sessions were significant and the quadratic term interacted with severity level, thereby resulting in adjustments to the effect of severity level, particularly for a larger number of sessions attended. As in the model including the PHQ-9 as a continuous score, the interaction between therapy modality and depression severity was not significant.

Table 3 presents models for the four PHQ-9 severity levels, applying the variables significant in the full sample model (Table 2). Significant

Table 3
Models for each level of depression severity based on PHQ-9 cut-offs.

	None/Mild (0–9) ***N*** = 1923		Moderate (10–14) ***N*** = 2783		Moderately severe (15–19) ***N*** = 3220		Severe (20–27) ***N*** = 3190	
	B (S.E.)	*p*-Value	B (S.E.)	*p*-Value	B (S.E.)	*p*-Value	B (S.E.)	*p*-Value
Fixed effects								
cons	4.66		7.124		9.373		12.894	
	(0.203)		(0.195)		(0.224)		(0.283)	
First PHQ9-gm	0.24	<0.001	0.357	<0.001	0.335	0.001	0.535	<0.001
	(0.06)		(0.095)		(0.101)		(0.087)	
First GAD7-gm	0.087	0.003	0.085	<0.001	0.085	0.001	0.184	<0.001
	(0.029)		(0.024)		(0.026)		(0.038)	
First WSAS-gm	0.06	<0.001	0.101	<0.001	0.113	<0.001	0.112	<0.001
	(0.015)		(0.013)		(0.014)		(0.016)	
Referral source: Self								
General Practitioner	0.415	0.051	0.172	0.396	0.243	0.288	0.093	0.732
	(0.213)		(0.203)		(0.229)		(0.271)	
Outside Agency	0.902	0.029	0.558	0.159	0.784	0.061	1.142	0.009
	(0.414)		(0.396)		(0.418)		(0.435)	
Employment status: Employed								
Unemployed: Seeking Work	0.233	0.568	0.758	0.02	1.29	<0.001	0.957	0.007
	(0.409)		(0.326)		(0.317)		(0.353)	
Student	1.042	0.056	0.552	0.227	0.564	0.252	0.584	0.376
	(0.544)		(0.457)		(0.492)		(0.66)	
Long-Term sick	1.349	0.002	1.03	0.002	1.159	<0.001	1.813	<0.001
	(0.445)		(0.33)		(0.304)		(0.327)	
Homemaker	0.686	0.133	0.635	0.101	0.197	0.653	0.901	0.1
	(0.456)		(0.387)		(0.439)		(0.547)	
Not Working_not SW or RB	−0.079	0.907	0.514	0.436	1.308	0.023	2.061	0.001
	(0.675)		(0.66)		(0.575)		(0.609)	
Unpaid Voluntary Work	−0.942	0.426	3.348	0.027	1.965	0.19	2.063	0.225
	(1.183)		(1.51)		(1.50)		(1.70)	
Retired	0.192	0.59	0.048	0.891	0.842	0.058	−0.831	0.17
	(0.357)		(0.352)		(0.445)		(0.606)	
Number of contacts in the 6 months prior to start of treatment episode -gm	0.025	0.184	0.088	<0.001	0.109	<0.001	0.1	<0.001
	(0.019		(0.017)		(0.021)		(0.027)	
N Sessions-gm	−0.062	0.151	−0.255	<0.001	−0.325	<0.001	−0.469	<0.001
			(0.041)		(0.046)		(0.057)	
(N Sessions-gm)^2	0.025	0.003	0.038	<0.001	0.045	<0.001	0.032	<0.001
	(0.009)		(0.006)		(0.006)		(0.006)	
(N Sessions-gm)^3	−0.002	0.066	−0.001	0.16	−0.002	0.003	−0.001	0.323
	(0.001)		(0.001)		(0.001)		(0.001)	
Therapy: CBT								
PCET	−0.019	0.937	0.518	0.039	0.538	0.056	0.878	0.011
	(0.242)		(0.252)		(0.281)		(0.346)	
Therapy:PCET* N Sessions-gm	0.068	0.262	0.036	0.541	0.233	<0.001	0.222	0.003
	(0.061)		(0.058)		(0.063)		(0.076)	
First PHQ9-gm* N Sessions-gm	−0.044	<0.001	−0.016	0.314	−0.055	0.001	−0.041	<0.001
	(0.011)		(0.015)		(0.016)		(0.012)	
First PHQ9-gm* (N Sessions-gm)^2	0.004	0.037	−0.00004	0.987	0.002	0.411	−0.001	0.742
	(0.002)		(0.003)		(0.003)		(0.002)	
CCG effect	0 (NS)		1.7 % (NS)		1.9 %		1.4 % (NS)	

variables are in bold to aid comparisons. In general, fewer variables were associated with outcome for non-clinical patients while most were significant for severe patients. Higher severity scores on each of the three outcome measures were associated with a poorer outcome for all levels of depression severity, but with considerable differences in effects between non-clinical and severe categories, with higher severity scores having a greater effect in the severe category. For the non-clinical category, the interaction between PHQ-9 score at intake and sessions attended was significant for both the linear and quadratic term for sessions, while for moderately severe and severe only the linear term was significant. For the moderate category there was no significant interaction between PHQ-9 score and sessions attended.

Compared to self-referral, referral by the GP was non-significant, while referral by an outside agency was associated with outcome for non-clinical and severe patients but not for moderate or moderately severe, although for the latter it is borderline. With regard to employment status, long term sick was associated with a poorer outcome for all severity levels, while for other categories of employment status the associations differed between severity levels. Number of sessions in the previous six months was associated with outcome for all severity levels except non-clinical.

The relationship between the number of sessions attended at outcome and PHQ-9 outcome score differed between severity levels. For non-clinical, only the quadratic term for sessions was significant, while for moderately severe a cubic relationship was also significant. For moderate and severe categories, the linear and quadratic terms were significant.

PCET had slightly better outcomes than CBT for non-clinical patients,

but the effect was non-significant. For moderate, moderately-severe, and severe categories, PCET had poorer outcomes, but the difference was non-significant ($p = 0.056$) for the moderately-severe group.

Table 3 also shows that the CCG effect was only significant for the moderately severe category (1.9 %) and for each level of severity the effect it was smaller than in the full sample model. Notably for non-clinical severity, there was no CCG effect.

Fig. 3 plots the predicted outcomes for each severity level derived from their models. It shows that for non-clinical and moderate, where there was no interaction between therapy modality and sessions and no crossover and the outcomes for both modalities run parallel with an advantage in favour of PCET for non-clinical patients and in favour of CBT for moderate patients. However, the differences in both were small and the 95 % CI overlap across all sessions attended.

Comparing outcomes for CBT and PCETS for non-clinical patients, the mean (SD) pre-post change for CBT was 0.78 (4.43) while for PCET it was 0.81 (4.54) ($d = -0.01, -0.11, 0.09$). For moderate patients, change was 3.79 (4.86) for CBT and 3.18 (4.86) for PCET ($d = 0.12, -0.04, 0.21$). With moderate scores, recovery rate and reliable improvement rate are the same and were 39 % for CBT and 31.8 % for PCET.

The plots for moderately-severe and severe depression show the crossover of effectiveness. For moderately-severe patients, PCET had better outcomes for patients who attended 2–6 sessions (although 95 % CIs overlap) with a mean (SD) pre-post PHQ-9 change for CBT of 3.34 (5.51), compared with 4.83 (5.62) for PCET, giving an effect size favoring PCET ($d = -0.27, -0.40, 0.14$). The reliable improvement rate and recovery rate for these patients were 31.6 % and 25.0 %, respectively, for CBT and 44.5 % and 35.2 %, respectively, for PCET.

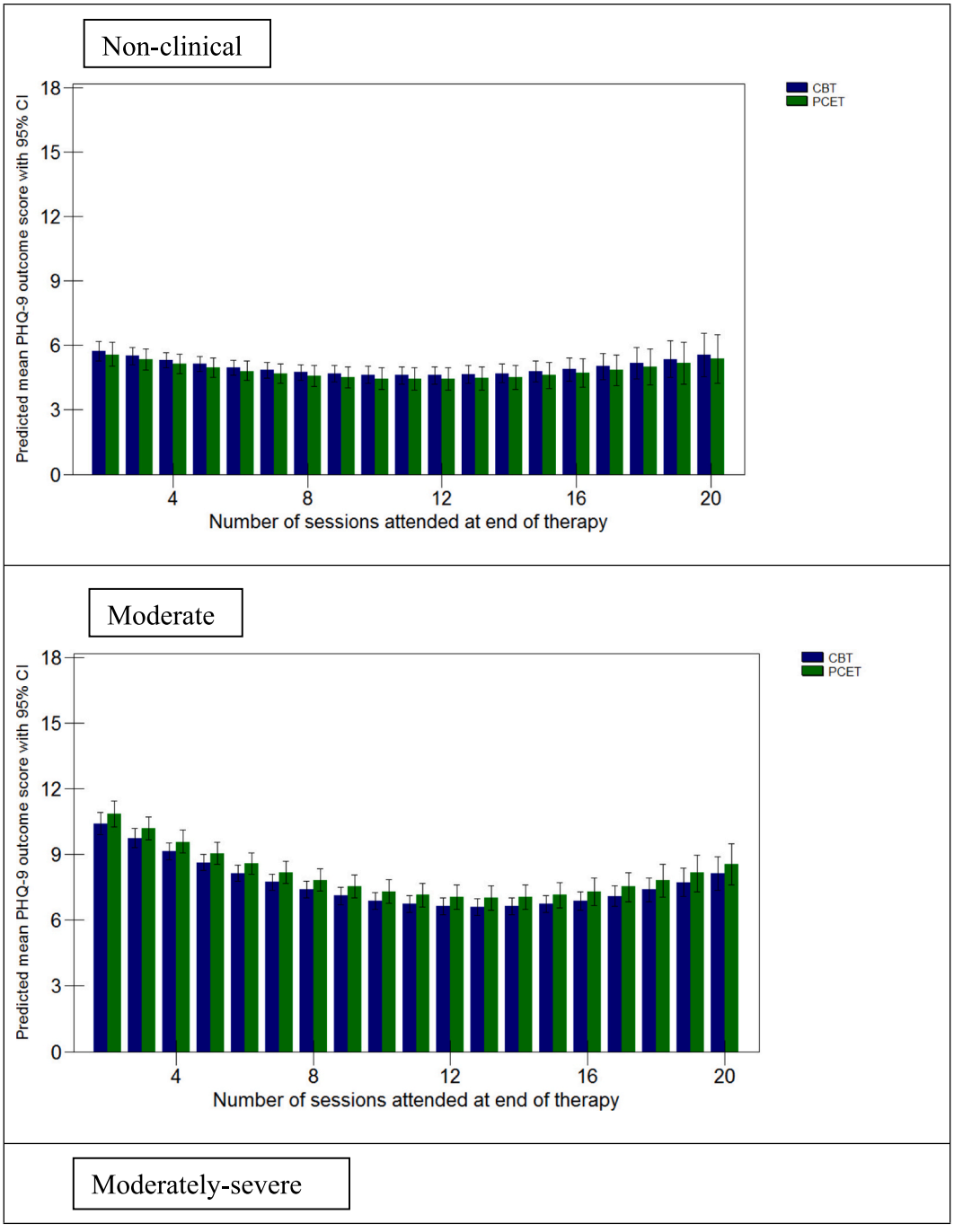

Fig. 3. Plots of predicted PHQ-9 outcomes for each depression severity level.

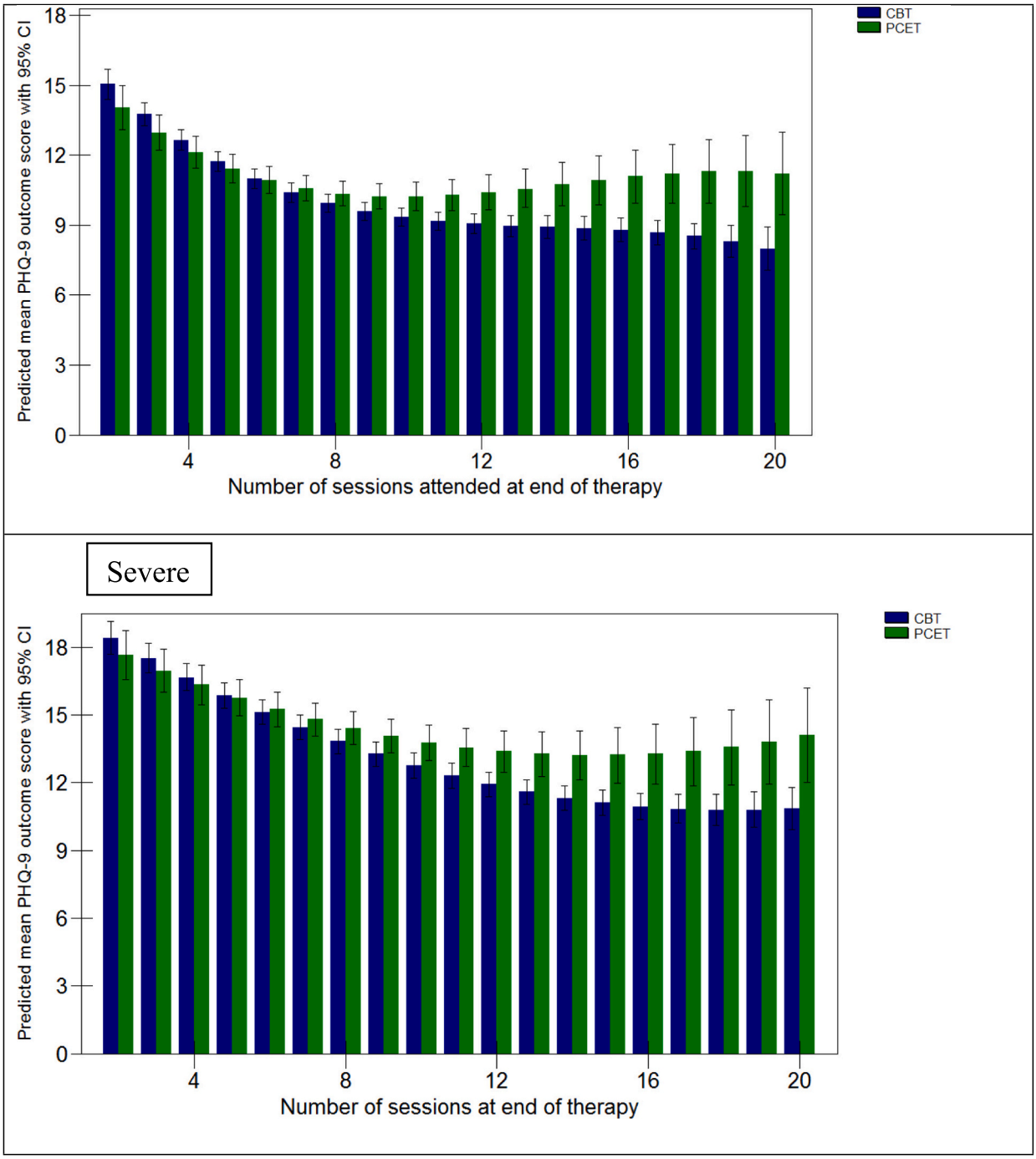

Fig. 3. (*continued*).

For moderately-severe patients who had 7–20 sessions, CBT had better outcome scores with no overlap of 95 % CIs for 11–20 sessions. For 7–10 sessions where the 95 % CI overlap, the pre-post change for CBT was 6.78 (5.98) compared with 6.48 (5.99) for PCET (d = 0.05, −0.10, 0.20). The improvement and recovery rates were similar, 59.6 % and 50.2 %, for CBT and 58.2 % and 49.0 % for PCET. For patients who had received 11–20 sessions, the pre-post change for CBT was 7.75 (5.61) compared with 6.07 (6.11) for PCET, giving an effect size (95 % CI) of d = 0.30 (0.12, 0.48). The improvement and recovery rates for CBT were 70.5 % and 60.4 % respectively while for PCET they were 57.3 % and 50.4 %.

Where patients were severe and had received 2–5 sessions, outcomes favored PCET although again the 95 % CIs overlap for each session. Overall for 2–5 sessions, the mean (SD) pre-post PHQ-9 change for CBT was 3.88 (5.35) and for PCET it was 5.01 (6.12) (d = −0.20, −0.36, 0.06). For CBT, 30.6 % had made reliable improvement while 8.5 % had recovered which compared to 35.0 % and 13.9 % for PCET.

CBT had better outcomes where severe patients had had 6–12 sessions although 95 % CIs overlap. The mean (SD) pre-post change for CBT was 8.59 (7.12) while for PCET it was 7.35 (6.44) (d = 0.18 (0.05, 0.30). Improvement and recovery rates were 60.8 % and 32.9 % for CBT and 57.3 % and 24.6 % for PCET.

Finally, for severe patients who had received 13 to 20 sessions, where CBT had better outcomes and the 95 % CIs did not overlap, the mean (SD) pre-post change for CBT was 10.84 (6.94) while for PCET it was 9.05 (6.47) giving an effect size (95 % CI) of d = 0.26 (0.02, 0.49). In this

sample, 74.8 % of CBT patients had made reliable improvement while 46.6 % had recovered compared to 66.2 % and 36.4 % for PCET.

4. Discussion

In contrast to much of the literature reporting little difference in effectiveness, our results highlight key differences between CBT and PCET according to referrals, patient journey, and outcomes. Most noteworthy, results indicate a cross-over in effectiveness for patients who were moderately severe or severe at intake, with PCET more effective where patients had attended fewer sessions at outcome and CBT more effective where patients had attended more sessions. Where patients were non-clinical or had moderate severity there was no cross-over and little difference between therapy modality outcomes.

Patients assigned to each therapy type were broadly similar on most characteristics, but with some notable differences. Older and retired patients received PCET, which is supported by an apparent resistance by GPs to refer retired patients to CBT (Collins and Corna, 2018). Also, a larger proportion of patients assigned to CBT had self-referred, were from less deprived neighborhoods, or tended to have more severe symptoms: 59 % of patients in receipt of CBT with moderately-severe or severe levels of depression compared with 53 % for PCET. CCGs had a small but significant effect on patient outcomes, which may be due in part to differences in the availability of CBT and PCET and how each is utilized at different CCGs.

Patients receiving CBT had more sessions on average, nine compared with six for PCET, as previously reported (Barkham et al., 2021; Pybis et al., 2017). There may be historical reasons for this difference as before the introduction of NHS Talking Therapies/IAPT services, counselors usually offered brief therapy of 6–8 sessions (e.g., Shepherd et al., 2005). We found that the difference in number of sessions attended was not due to a larger proportion of PCET patients ending therapy prematurely, as CBT had the higher drop-out rate and was more likely to consider patient unsuitable for Talking Therapies treatment.

For the full sample, and the two randomized sub-samples, greater baseline severity and attending fewer sessions were both strongly associated with a poorer outcome. Given that more sessions, generally, improved outcomes (Wakefield et al., 2021) and PCET generally provided fewer sessions, the adjustment for sessions attended improved the outcomes of PCET relative to CBT. Patients who had had more contacts with the service in the previous 6 months, perhaps a proxy measure of severity, had poorer outcomes as did those who were referred to the service by an outside agency (such as legal, education and social services) and those in some employment status categories, particularly long-term sickness and unemployed.

Supporting previous findings (Pybis et al., 2017), the relationship between outcome and the number of sessions attended differed for the two therapies. Where patients had had fewer sessions PCET was more effective, while CBT was more effective where patients had more sessions, leading to a cross-over in effectiveness. This enhanced early impact of PCET has been noted previously (Duffy et al., 2022). One possible explanation for the differential effect of treatment 'dose' might be that change and treatment ending in CBT follows the classic dose-response model (Howard et al., 1996) while in PCET change follows the good enough level (GEL) model (Baldwin et al., 2009; Barkham et al., 2006). Such an account would suggest that both models of change are valid, and determining the optimal approach may rely on matching the patient's needs and expectations with the appropriate therapy (see Moggia et al., 2023). Future research should consider the components of each therapy modality that contribute to these differential outcomes.

We found the cross-over to occur where patients had attended >5 sessions while previously it was found to be around 8 sessions (Pybis et al., 2017). This is likely to be due to factors such as the population, the sample sizes for each dose of therapy and variables included in the model, however the general finding is that attending more sessions has greater benefit for CBT patients.

The intake severity levels of the patient population may also influence the cross-over point. Our analysis of depression severity levels found a cross-over for moderately-severe and severe patients (a population more similar to the PRaCTICED trial) but not for non-clinical or moderate patients. Also, for moderately-severe we found the cross-over to be at 7 sessions while for severe it was 6 sessions. Therefore, the composition of the population in terms of severity levels is likely to influence where the cross-over occurs. This, along with sample sizes for each level of sessions attended (and therefore the 95 % CI) may also determine where differences between therapy outcomes are reliable.

PCET was not reliably more effective than CBT for any dose of therapy. For non-clinical patients, although PCET was superior across the number of sessions attended the effect size was only −0.01 (−0.11, 0.09). For moderately-severe patients who had attended up to 6 sessions it was −0.27 (−0.40, 0.14) and for severe patients who attended up to 5 sessions it was −0.20 (−0.36, 0.06).

CBT on the other hand was found to be superior to PCET for moderately-severe and severe patients who attended 11–20 and 13–20 sessions respectively. The effect sizes were 0.30 (0.12, 0.48) for the former and 0.26 (0.02, 0.49) for the latter. Although these effect sizes may be considered small to medium, in terms of recovery rate there was almost a 10 % difference between therapy modalities, which if scaled-up in a national service over time may represent many thousands of individual patients.

Our results go beyond the Dodo bird verdict whereby, in acknowledging the value of smaller effects, we establish a more nuanced differential effect arising due to the relationships between the therapy type and other factors. A small effect may vary such that it favors one therapy type in some circumstances and another therapy in different circumstances. Small effects can therefore yield important findings for therapy when delivered at scale and also inform the assessment and allocation of patients to therapy types (see Barkham, 2023).

Clinical and service implications.

The combination of the crossover effect and the differential assignment as well as dropout rate suggests that both therapy modalities have areas to address in terms of enhancing their impact: CBT to retain more patients in therapy and for PCET to provide improved outcomes for patients who require >5 sessions. Clinical services can improve assessment and allocation when stepping-up patients and assessors should consider patient preferences and motivation for each therapy and the likelihood of drop-out and also the severity level of the patient. Our results suggest that for patients who are non-clinical or moderately depressed either step 3 therapy will be similarly effective. However, for patients who are more severe if they are informed and prepared to attend at least 10 sessions, then CBT would be more effective. Patients wanting fewer sessions and/or who are more likely to drop-out of treatment, might be better suited to PCET.

4.1. Limitations

No data were available on therapy allocation decisions within services, including the influence of patient preference, PWP clinical decisions, and high intensity therapy type availability. Furthermore, the main problem of the patient (i.e. depression) was determined by the PWP, who was unlikely to be qualified in making a formal diagnosis, therefore it can only be seen as an informed opinion. Also, therapist effects could not be controlled for as data were unavailable. Although it is likely that there would be a significant therapist effect (e.g. Saxon et al., 2017), there is little evidence that therapist variability would significantly alter the main findings (e.g., Pybis et al., 2017). Also, the findings may be limited to NHS Talking Therapies services as the delivery of therapy for depression may vary considerably between countries, particularly regarding the number of sessions patients are offered (Flückiger et al., 2020).

5. Conclusions

Small differential effect sizes overall may imply that therapy type is of little importance. However, they can mask important differences between therapies for different patient subgroups. Notably, it does matter which therapy a patient attends based on their depression severity and number of sessions they will attend. We found that although the number of sessions the patient attended moderates the effect of intake severity on outcome for all patients, for those who are moderately-severe or severe, therapy modality also moderates the effect of sessions on outcome. The study of small effects can therefore inform therapy assignment and how different therapies can be best utilized to improve overall effectiveness and can contribute to the field of precision medicine and matching patients to therapies. It should also be recognised that small effect sizes, when scaled up to national routine services, can represent large numbers of patients.

Funding

This research was partially funded by the British Association for Counselling & Psychotherapy (BACP).

CRediT authorship contribution statement

David Saxon: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Software, Writing – original draft, Writing – review & editing. **Emma Broglia:** Conceptualization, Investigation, Writing – review & editing. **Charlie Duncan:** Conceptualization, Writing – review & editing. **Michael Barkham:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The contribution of David Saxon was partially funded by BACP.

Emma Broglia has previously been employed by BACP.

Charlie Duncan is employed by BACP.

Michael Barkham has previously received funding from BACP and previously been a member of various BACP research and scientific committees.

No authors are members of BACP.

Acknowledgments

We would like to acknowledge the assistance and support of NHS Digital, NHS England, in the initial preparation, anonymization and provision of the data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2024.06.115.

References

Baldwin, S.A., Berkeljon, A., Atkins, D.C., Olsen, J.A., Nielsen, S.L., 2009. Rates of change in naturalistic psychotherapy: contrasting dose-effect and good-enough level models of change. J. Consult. Clin. Psychol. 77 (2), 203–211. https://doi.org/10.1037/a0015235.

Barkham, M., 2023. Smaller effects matter in the psychological therapies: 25 years on from Wampold et al. (1997). Psychother. Res. 33 (4), 530–532. https://doi.org/10.1080/10503307.2022.2141589.

Barkham, M., Lambert, M.J., 2021. The efficacy and effectiveness of psychological therapies. In: Barkham, M., Lutz, W., Castonguay, L.G. (Eds.), Bergin and Garfield's Handbook of Psychotherapy and Behavior Change, 7th ed. Wiley, pp. 135–189.

Barkham, M., Saxon, D., 2018. The effectiveness of high-intensity CBT and counselling alone and following low-intensity CBT: a reanalysis of the 2nd UK National Audit of Psychological Therapies data. BMC Psychiat. 18, 321.

Barkham, M., Connell, J., Stiles, W.B., Miles, J.N.V., Margison, J., Evans, C., Mellor-Clark, J., 2006. Dose-effect relations and responsive regulation of treatment duration: the good enough level. J. Consult. Clin. Psychol. 74, 160–167.

Barkham, M., Saxon, D., Hardy, G.E., Bradburn, M., Galloway, D., Wickramasekera, N., Keetharuth, A.D., Bower, P., King, M., Elliott, R., Gabriel, L., Kellett, S., Shaw, S., Wilkinson, T., Connell, J., Harrison, P., Ardern, K., Bishop-Edwards, L., Ashley, K., Brazier, J.E., 2021. Clinical and cost-effectiveness of person-centred experiential therapy vs. cognitive behavioural therapy for moderate and severe depression as delivered in the English Improving Access to Psychological Therapies (IAPT) national programme: a pragmatic randomised non-inferiority trial [PRaCTICED]. The Lancet Psychiat. 8, 487–499.

Beck, J.S., 2011. Cognitive Behavior Therapy: Basics and beyond, 2nd ed. Guilford Press.

Carroll, L., 1865/2015. In: Burstein, M. (Ed.), Alice's Adventures in Wonderland. Princeton University Press.

Charlton, C., Rasbash, J., Browne, W.J., Healy, M., Cameron, B., 2020. MLwiN Version 3.05. Centre for Multilevel Modelling, University of Bristol.

Clark, D.M., 2018. Realizing the mass public benefit of evidence-based psychological therapies: the IAPT Program. Annu. Rev. Clin. Psychol. 14, 159–183. https://doi.org/10.1146/annurev-clinpsy-050817-084833.

Cohen, Z.D., Delgadillo, J., DeRubeis, R.J., 2021. Personalized treatment approaches. In: Barkham, M., Lutz, W., Castonguay, L.G. (Eds.), Bergin and Garfield's Handbook of Psychotherapy and Behavior Change, 7th ed. Wiley, pp. 667–700.

Collins, N., Corna, L., 2018. General practitioner referral of older patients to improving access to psychological therapies (IAPT): an exploratory qualitative study. Brit. J. Psych. Bull. 42 (3), 115–118. https://doi.org/10.1192/bjb.2018.10.

Cuijpers, P., Karyotaki, E., de Wit, L., Ebert, D.D., 2020. The effects of fifteen evidence-supported therapies for adult depression: a meta-analytic review. Psychother. Res. 30 (3), 279–293. https://doi.org/10.1080/10503307.2019.1649732.

Delgadillo, J., Gonzalez Salas Duhne, P., 2020. Targeted prescription of cognitive-behavioral therapy verses person-centered counseling for depression using a machine learning approach. J. Consult. Clin. Psychol. 88 (1), 14–24. https://doi.org/10.1037/ccp0000476.

Duffy, K.E.M., Simmonds-Buckley, M., Saxon, D., Delgadillo, J., Barkham, M., 2022. Early response as a prognostic indicator in person-centered experiential therapy for depression. J. Counsel. Psych. 69 (6), 803–811. https://doi.org/10.1037/cou0000633.

Duffy, K.E.M., Simmonds-Buckley, M., Haake, R., Delgadillo, J., Barkham, M., 2024. The efficacy of individual humanistic-experiential therapies for the treatment of depression: a systematic review and meta-analysis of randomized controlled trials. Psychother. Res. 34 (3), 323–338. https://doi.org/10.1080/10503307.2023.2227757.

Elliott, R., Watson, J., Timulak, L., & Sharbanee, J. (2021). Research on humanistic-experiential psychotherapies: updated review. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), Bergin and Garfield's Handbook of Psychotherapy and Behavior Change. (7th ed., pp. 421–467). John Wiley & Sons, Inc.

von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C., Vandenbroucke, J.P., 2007. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann. Intern. Med. 147 (8), 573–577. https://doi.org/10.7326/0003-4819-147-8-200710160-00010.

Flückiger, C., Wampold, B.E., Delgadillo, J., Rubel, J., Vîslă, A., Lutz, W., 2020. Is there an evidence-dased number of sessions in outpatient psychotherapy? - a comparison of naturalistic conditions across countries. Psychother. Psychosom. 89 (5), 333–335. https://doi.org/10.1159/000507793.

Howard, K.I., Moras, K., Brill, P.L., Martinovich, Z., Lutz, W., 1996. Evaluation of psychotherapy. Efficacy, effectiveness, and patient progress. Am. Psychol. 51, 1059–1064. https://doi.org/10.1037//0003-066x.51.10.1059.

Jacobson, N.S., Truax, P., 1991. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J. Consult. Clin. Psychol. 59 (1), 12–19. https://doi.org/10.1037//0022-006x.59.1.12.

Kroenke, K., Spitzer, R.L., Williams, J.B., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16 (9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., Monahan, P.O., Löwe, B., 2007. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann. Intern. Med. 146, 317–325. https://doi.org/10.7326/0003-4819-146-5-200703060-00004.

Lambert, M.J., 2013. The efficacy and effectiveness of psychotherapy. In: Lambert, M.J. (Ed.), Bergin and Garfield's Handbook of Psychotherapy and Behavior Change, 6th edition. Wiley.

Luborsky, L., Rosenthal, R., Diguer, L., Andrusyna, T.P., Berman, J.S., Levitt, J.T., Seligman, D.A., Krause, E.D., 2002. The dodo bird verdict is alive and well—mostly. Clin. Psychol–Sci. Pr. 9 (1), 2–12. https://doi.org/10.1093/clipsy.9.1.2.

Ministry of Housing, Communities and Local Government, 2019. The English Indices of Deprivation 2019 (IoD2019). Available at: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019.

Moggia, D., Saxon, D., Lutz, W., Hardy, G.E., Barkham, M., 2023. Applying precision methods to treatment selection for moderate/severe depression in person-centered experiential therapy or cognitive behavioral therapy. Psychother. Res. 1–16. Advance online publication. https://doi.org/10.1080/10503307.2023.2269297.

Mundt, J.C., Marks, I.M., Shear, M.K., Greist, J.H., 2002. The work and social adjustment scale: a simple measure of impairment in functioning. Brit. J. Psychiat. 180, 461–464. https://doi.org/10.1192/bjp.180.5.461.

Murphy, D., 2019. Person-Centred Experiential Counselling for Depression. Sage.

National Institute for Health and Care Excellence. (2009). Depression in adults: recognition and management. [NICE Guideline 90]. https://www.nice.org.uk/guidance/cg90.

National Institute for Health and Care Excellence. (2022). Depression in adults: treatment and management. [NICE Guideline 222]. https://www.nice.org.uk/guidance/ng222.

Newman, M.G., Agras, W.S., Haaga, D.A.F., Jarrett, R.B., 2021. Cognitive, behavioral, and cognitive behavioral therapy. In: Barkham, M., Lutz, W., Castonguay, L.G. (Eds.), Bergin and Garfield's Handbook of Psychotherapy and Behavior Change. John Wiley & Sons, Inc, pp. 469–505.

Pybis, J., Saxon, D., Hill, A., Barkham, M., 2017. The comparative effectiveness and efficiency of cognitive behaviour therapy and counselling in the treatment of depression: evidence from the 2nd UK National Audit of Psychological Therapies. BMC Psychiat. 17, 215.

Rosenzweig, S., 1936. Some implicit common factors in diverse methods of psychotherapy. Am. J. Orthop. 6 (3), 412–415. https://doi.org/10.1111/j.1939-0025.1936.tb05248.x.

Sanders, P., Hill, A., 2014. Counselling for Depression: A Person-centred and Experiential Approach to Practice. SAGE Publishing, London, UK.

Saxon, D., Firth, N., Barkham, M., 2017. The relationship between therapist effects and therapy delivery factors: therapy modality, dosage, and non-completion. Adm. Policy Ment Hlth. 44, 705–715. https://doi.org/10.1007/s10488-016- 0750-5.

Serbanescu, I., Backenstrass, M., Drost, S., Weber, B., Walter, H., Kleine, J.P., Zobel, I., Hautzinger, M., Meister, R., Härter, M., Schramm, E., Schoepf, D., 2020. Impact of baseline characteristics on the effectiveness of disorder-specific cognitive Behavioural Analysis System of Psychotherapy (CBASP) and Supportive Psychotherapy in outpatient treatment for persistent depressive disorder. Front. Psychiatry 11, 607300. https://doi.org/10.3389/fpsyt.2020.607300.

Shepherd, M., Ashworth, M., Evans, C., Robinson, S.I., Rendall, M., Ward, S., 2005. What factors are associated with improvement after brief psychological interventions in primary care? Issues arising from using routine outcome measurement to inform clinical practice. Couns. Psychother. Res. 5 (4), 273–280. https://doi.org/10.1080/14733140600571326.

Snijders, A.B., Bosker, R.J., 2012. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling, 2nd Ed. Sage.

Spitzer, R.L., Kroenke, K., Williams, J.B., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch. Intern. Med. 166, 1092–1097 (doi: archinte.166.10.1092).

Tolin, D.F., 2010. Is cognitive-behavioral therapy more effective than other therapies? A meta-analytic review. Clin. Psychol. Rev. 30 (6), 710–720. https://doi.org/10.1016/j.cpr.2010.05.003.

Wakefield, S., Kellett, S., Simmonds-Buckley, M., Stockton, D., Bradbury, A., Delgadillo, J., 2021. Improving access to psychological therapies (IAPT) in the United Kingdom: a systematic review and meta-analysis of 10-years of practice-based evidence. Brit. J. Clin. Psych. 60 (1), 1–37. https://doi.org/10.1111/bjc.12259.

Wampold, B.E., Mondin, G.W., Moody, M., Stich, F., Benson, K., Ahn, H.-n., 1997. A meta-analysis of outcome studies comparing bona fide psychotherapies: empirically, "all must have prizes". Psychol. Bull. 122 (3), 203–215. https://doi.org/10.1037/0033-2909.122.3.203.