Long-term psychotherapy in tertiary care: A practice-based benchmarking study

Some of the authors of this publication are also working on these related projects:

Project   An Investigation into E-therapies for Older Adults   View project

Project   Routine outcome monitoring and feedback   View project

Long-Term Psychotherapy in Tertiary Care: A Practice-Based Benchmarking Study

Chris Gaskell[1]*, Stephen Kellett[2,1], Melanie Simmonds-Buckley[1,2], Joe Curran[3], Jack Hetherington[3], & Jaime Delgadillo[1,2]

1. University of Sheffield, UK

2. Rotherham Doncaster and South Humber NHS Foundation Trust, UK

3. Sheffield Health and Social Care NHS Foundation Trust, UK

*Corresponding author: Dr Chris Gaskell, c.gaskell@sheffield.ac.uk

Abstract

Objectives. The literature regarding the effectiveness of long-term psychological interventions delivered in tertiary care is scarce. This study sought to quantify and evaluate outcomes delivered in a UK tertiary care psychotherapy service against equivalent service benchmarks. *Design.* A retrospective analysis of outcomes on the Outcome Questionnaire-45 (OQ-45) over a 10-year period in a tertiary care psychotherapy service. The modalities evaluated were cognitive-behavioural, cognitive-analytic and psychoanalytic psychotherapies. *Methods.* Effectiveness was calculated at the service level and for each modality using pre-post effect sizes and recovery rates. Benchmarking included a random effects meta-analysis. Trajectories of change for each modality were examined using growth curve modeling. *Results.* Baseline distress on the OQ-45 was higher than comparative norms (M = 102.57, SD = 22.79, N = 364). The average number of sessions was 48.68 (SD = 42.14, range = 5-335). There was a moderate pre-post treatment effect (*d* = 0.46, 95% CI = 0.37-0.55) which was lower than available benchmarks. The modalities differed in duration but were largely equivalent in terms of outcome. The reliable improvement rate was 29.95% and the recovery rate was 10.16% and change over time was best explained using a nonlinear (cubic) time trend. *Conclusions.* The elevated distress at baseline appears to create the conditions for relatively lengthy interventions and attenuated clinical outcomes. Suggestions are made regarding the clinical role, function, and evaluation of tertiary care psychotherapy services.

Keywords: 'psychotherapy,' 'effectiveness,' 'tertiary-care,' 'growth curves.'

Practitioner Points:

(i) Evidence indicates that a sub-sample of tertiary care patients do not respond to psychological treatment. For these patients, there is limited evidence for extending therapy beyond 100 sessions.

(ii) Outcome monitoring within supervision may provide a suitable means of reviewing treatment progress to rectify difficulties or to avoid protracted and ineffective interventions.

(iii) There was limited evidence favouring the outcomes of one psychological modality over another, since the confidence intervals for treatment-specific effect sizes overlapped.

Long-Term Psychotherapy in Tertiary Care: A Practice-Based Benchmarking Study

Within the National Health Service (NHS) in the United Kingdom (UK) psychological healthcare delivery is organised into a hierarchy of sectors according to patient need and complexity across primary, secondary, and tertiary tiers. In primary care, the Improving Access to Psychological Therapies (IAPT) programme provides brief and time limited interventions for mild-to-moderate (low intensity psychological interventions) and moderate-to-severe (high intensity psychological interventions) anxiety, depression, and often other comorbid mental health difficulties (Clark, 2018). Patients with greater complexity and risk, or who have not responded to treatment in IAPT are signposted to secondary care psychological provision, often delivered in multi-disciplinary community mental health teams (Burns, 2004). Patients that do not respond to primary or secondary care are referred for specialist and long-term psychological interventions offered in tertiary care psychotherapy services (*Taylor et al., 2012*). Tertiary care therapy services in the NHS are rare, they cover wide geographical regions and offer resource intensive and typically lengthy psychological interventions. Examples of the psychotherapies provided in tertiary care services include: dynamic interpersonal therapy (DIT, *Douglas et al., 2016*), intensive short-term dynamic interpersonal therapy (ISTDP, *Johansson et al., 2014*), cognitive analytic therapy (CAT, *Ryle & Kerr, 2020*), psychodynamic-interpersonal therapy (PIT, *Paley et al., 2008*) and psychoanalytic psychotherapy (*Warden et al., 2008*).

Tertiary care psychotherapy services have been characterised as being in high demand from referrers and commissioners, but often lacking in resources and relevant practice-based evidence for their effectiveness (*Warden et al., 2008*). Reasons for the lack of evidence include services being few, outcome studies failing to name the clinical context, and very high levels of missing data in the studies that have been conducted (e.g., up to 95%; *Firth et al., 2020*). Such

levels of missing data unfortunately lead researchers to omit tertiary care samples from multi-sector analyses (see *Stiles et al., 2006* for an example). The few studies which evaluated tertiary outcomes in the NHS have tended to have very small sample sizes, therefore undermining the reliability of the evidence base. For example, *Paley et al.* (*2008*) reported on the effectiveness of psychodynamic interpersonal therapy with 47 tertiary care patients and *Douglas et al.* (*2016*) explored the effectiveness of dynamic interpersonal therapy with 28 patients. It is likely that the heterogeneity of patients being referred to these services complicates the process of grouping clients in a meaningful way while maintaining a sufficient sample size. Other studies conducted in tertiary care services have evaluated the effectiveness of intensive short-term dynamic psychotherapy (*Abbass et al., 2008*; *Johansson et al., 2014*; *Lilliengren et al., 2020*; *Nowoweiski et al., 2020*), interventions for chronic fatigue (*Heins et al., 2011*; *Worm-Smeitink et al., 2016*) and for psychological distress in the context of autism spectrum disorders (*Blainey et al., 2017*).

Because of the relatively lengthy psychological treatments delivered in tertiary care, understanding the dose-response relationship in this context is of clinical and empirical interest (i.e., where the "dose" refers to the number of sessions and "response" refers to the clinical outcome; *Howard et al., 1986*). The dose-response relationship during routine psychological interventions is often found to be curvilinear, with most of the changes observed during earlier stages of treatment (see review by *Robinson et al., 2020*). This systematic review suggests that curvilinear relationships between treatment duration and outcomes are commonly observed in psychotherapy studies, and estimated that an optimal number of sessions to attain symptomatic improvement was 4-26 sessions, but that the range varied according to setting/clinical population/outcome measure. In particular, this review highlighted the absence of tertiary level care evidence. The impact of patient complexity on the dose-response effect is supported by

studies demonstrating that patients with chronic and characterological symptoms require longer treatments to reach comparable response rates (*Howard et al., 1986*) and that interpersonal problem resolution lags behind symptom improvement (*Kopta et al., 1994*). In a rare examination of the dose-response effect in long-term and open-ended psychotherapy for patients with severe psychopathology, Nordmo et al. (2021) reported that the degree of improvement was linearly associated with treatment duration and moderated by intake severity (less severe cases improved sooner). Since tertiary-level psychotherapy services typically offer a range of modalities to patients with severe and complex presentations, it could be that these patients show a different dose-response pattern (e.g., linear vs. nonlinear) to that typically observed in other clinical populations.

To summarise, there is a lack of practice-based evidence regarding the effectiveness of tertiary care psychotherapy services and few dose-response studies of long-term and open-ended psychotherapies. The aim of this study was to evaluate the effectiveness of tertiary care psychological interventions delivered in a routine clinical service. Specific objectives were [1] to index effectiveness at both the service (i.e., via effect sizes) and individual level (i.e., via calculating rates of various indices of change) and then [2] to benchmark service-level outcomes against the tertiary level care evidence base. Further objectives were [3] to explore how change occurs over time (i.e., using growth trajectories) and [4] to compare outcomes between three routinely delivered tertiary care psychotherapies, with the hypothesis that there would be no between-modality differences; in line with the frequently reported finding (e.g., Wampold et al., 1997) that different psychotherapy treatments do not systematically produce different rates of effectiveness (i.e., the 'dodo bird paradox', Rosenzweig, 1936).

Method

Design, ethical approval and service description

This was a retrospective analysis of naturalistic therapy outcomes data collected from a tertiary care specialist psychotherapy service. Approvals were obtained from an NHS research ethics committee (ref; 19/NW/0753), by the NHS Confidentiality Advisory Group (CAG) and the Health Research Authority (HRA) to access and analyse de-identified demographic (i.e., gender, age, employment, ethnicity), treatment (i.e., dosage, treatment received, clinician, completion status) and service utilisation information.

The entry criteria for the tertiary care service (TCS) were that patients were (1) aged 16 years or older, (2) not currently experiencing a mental health crisis (i.e., acute period of suicidality or risk to others), but in psychological distress, (3) had failed to benefit from psychological interventions in primary and/or secondary care services. The staff team was multi-disciplinary (i.e., clinical psychologists, psychotherapists, and medical psychotherapists), with each therapist receiving regular clinical supervision matched to their therapeutic modality. The modalities offered were cognitive-behavioural therapy (CBT), cognitive-analytic therapy (CAT) and psychoanalytic therapy (PAT). Patients are allocated to treatments based on team/clinician judgment of model suitability as decided during a multi-disciplinary team meeting. If a patient is known to have not respond to a particular modality in the past, then this may inform treatment selection. The CBT intervention involved 6-months of weekly sessions, the CAT intervention involved 8, 16 or 24 weekly sessions (according to patient complexity) and the PAT intervention involved weekly therapy sessions of 1-2 years, either one-to-one or in a group.

Outcome measure

The Outcome Questionaire-45 (OQ-45) is a self-report measure of global psychological distress (*Lambert et al., 2004*). Each item provides a 5-point Likert scale ranging from 0 (never) to 4 (almost always) with the cumulative score across the 45 items providing a global distress score with a maximum of 180. Embedded within the OQ-45 are three additional sub-scales. *Symptom distress* (range = 100, cut-off = 36, reliable change = 10) is designed to map onto symptoms of common mental health disorders (i.e., anxiety and affective disorder symptoms, Lambert, 2004). *Interpersonal relationship* (range = 1-36, cut-off = 15, reliable change = 8) explores complaints of conflict, loneliness, and family difficulty. Finally, *social role* (range = 44, cut-off = 12, reliable change = 7) is the extent to which the individual patient experiences difficulties relating to occupational and functional independence.

Normative data for community and clinical populations in the United States (*Lambert et al., 1996*) has provided a clinical cut-off point of 64, and a reliable change score of 14 (*Lambert, 2004*). The psychometric properties of the OQ-45 have been established (*Lambert et al., 1996*; *Vermeersch et al., 2000*). The internal consistency of the OQ-45 total score has been estimated at $r = .94$ (symptom distress = .93, interpersonal relationships = .78, social tole = .70, Boswell et al., 2013). When scoring the OQ-45, missing values at the item level were pro-rated when there was ≤5 missing values. In situations when there was >5 missing values for a single OQ-45 then the measure was discarded.

Outcome monitoring, data extraction, sample size and analyses

Patients completed the OQ at various stages of therapy (i.e., assessment, during therapy, final appointment); however, as outcome measurement was not mandatory, this was not implemented in a consistent way. Historical data spanning a 10-year period (2011-2021) was analysed consisting of 4,203 OQ-45 administrations from N=1,027 patients, treated by 53 therapists.

When limiting data to the first care episode (referral to discharge), 3,198 OQ-45 measures remained. There were 2,639 OQ-45 measures across 364 patients for the effectiveness and recovery analyses and the growth curve analyses were conducted using data from 298 eligible patients. For patients who had multiple recorded care episodes, the first care episode recorded within the OQ-45 database was used. For some patients, a single care episode included multiple treatments within a single modality. As concurrent treatments (based on the information available) could not be accurately disentangled, the current study analysed change across care episodes. Change was considered by comparing the first and final OQ-45 treatment session (exclusive of follow-up appointments) within a single care episode. Last observation carried forward was used for patients who dropped out or when data had not been collected after initial assessment. As patients complete the OQ-45 prior to treatment sessions, the study only reported on change in those patients who had received at least two treatment sessions, to ensure that treatment had begun. This sample was used for analysis of effectiveness and recovery rates. Longitudinal multi-level (mixed effects) modeling examined trajectories of change. Only those patients with a minimum of three completed OQ-45 measures (including at least one treatment session) were included in this analysis to allow for the assessment of higher-order polynomial change trajectories (e.g. quadratic, cubic trends).

Service outcomes

Effect sizes were calculated for the OQ-45 total score and each sub-scale using standardised mean change (*Cohen, 1988*) and using the formula advocated for benchmarking studies (*Minami et al., 2008*). This approach subtracted the mean end-of-therapy score from the mean pre-treatment score, before dividing it by the pre-treatment standard deviation. Regression to the mean was accounted for by adjusting confidence intervals by the correlation (Pearson's *r*)

between pre and post-treatment measures. Effect-sizes were interpreted as 'small' (0.2–0.5), 'moderate' (0.5–0.8) or 'large' (> .08).

Benchmarking

The benchmarking approach was used to contextualise the outcomes (*Minami et al., 2008*), as this compares clinical outcome data to established reference points from efficacy trials or practice-based outcome studies (*Delgadillo et al., 2014*; *Minami et al., 2008*). This allows services to compare their performance against services which are similar in design, or against aggregated study benchmarks (*Department of Health, 2004*). To benchmark service intake scores, (i.e., baseline distress) a range of US OQ-45 intake severity comparators were used to create a rounded comparison. These included an employee assistance program, university outpatient clinic, community health centre (all from Lambert et al., 1996), and acute short-stay inpatient setting (*Doerfler et al., 2002*). For pre-post change, a variety of effectiveness benchmarks were used. Data from a large Canadian tertiary care outpatient outcome study (*Johansson et al., 2014*) was used as the tertiary-care effectiveness benchmark.

Meta-analytic Benchmark.

To further the benchmarking effort, a meta-analytic benchmark was developed based on other studies reporting outcomes using the OQ-45 from routine practice. These studies were identified from a recent systematic search of practice-based studies (Gaskell et al., in press) using the OQ-45. In total, 13 studies were identified, for which Cohen's *d* effect-sizes were entered into a random effects meta-analysis to provide an aggregated benchmark of OQ-45 data, forming a suitable comparison to the current study. Most of the studies included in the meta-analytic benchmark (total sample = 12,263) came from the USA (n = 8), with the remaining studies

coming from Switzerland (n = 2), Norway (n = 1) and Israel (n = 1). The aggregated OQ-45 pre-post therapy effect-size was medium ($d = 0.58$, $k = 13$, CI = 0.42-0.75 $p = < .001$)[1].

Individual outcomes

This study calculated rates of improvement and deterioration using reliable and clinically significant change indices (*Jacobson & Truax, 1991*) for the whole service and for each of the three treatment modalities. A *reliable change* (improvement or deterioration) occurred when a patient met the minimum pre-determined OQ-45 change score based on a magnitude of change exceeding the reliable change index (i.e., change of 14 or more). A *clinically significant change* occurred when a patient moved from above the threshold for clinical distress on the OQ-45 before treatment to below this threshold after treatment. Those patients meeting both reliable improvement and clinically significant change criteria were labelled as 'recovered.' Each patient's treatment outcome was therefore classed as either recovered, reliably improved, no reliable change, or reliably deteriorated. Patients scoring below the OQ-45 clinical threshold at baseline assessment were unable to achieve full recovery. The modality-specific recovery rate was then compared to established OQ-45 recovery benchmarks (*Hansen & Lambert, 2002*) in US care sectors (i.e., employee assistance programs, community mental health centres, health maintenance organisations).

---

[1] Two studies in the meta-analysis were statistical outliers (*Goldberg et al., 2016*; *Lunnen et al., 2008*), but were retained as preliminary sensitivity removal did not substantially alter the effect-size ($d = 0.60$, $k = 11$, 95% CI = 0.48-0.73, $p = < .001$).

Dose-response analysis

As OQ-45 measures were not always administered at the start of treatment, this means that the total number of sessions attended often differed from the number of sessions with an available OQ-45 measure. To estimate the approximate number of sessions, the last OQ-45 session number minus the first OQ-45 session number was used. Patients were ordered by number of sessions and split into 10 groups (i.e., deciles) before being compared on statistical change (Cohen's *d*) and clinical change (recovery rates).

Growth curve modelling

Growth curves were developed for the OQ-45 subscale and total scores following the *Singer and Willett (2003*) guidelines to explore change trajectories, variance in change and the factors influencing change. The approach is robust to missing data inherent in practice-based datasets. The hierarchical data structure was repeated OQ-45 measures (level-1) nested within individual patients (level-2).

Time

As the OQ-45 administration number did not necessarily correspond to session number, this was not a viable temporal predictor, instead number of estimated *sessions* (i.e., contacts) since the first recorded session (in that care episode) was used (centred on the first session). Additional sessions variables were created to assess polynomial and log-linear trends.

Model building

To select a best-fitting covariance structure, random slopes unconditional models were estimated fitting a series of alternative covariance structures (standard, unstructured, compound symmetry, auto-regressive1, topeplitz). Topeplitz provided the best fit for total OQ-45 score, symptom

distress and interpersonal relationships (see Table in online supplementary materials). A standard covariance structure provided the best fit for social role. In terms of polynomial change, a cubic form was the best fit for the total score, interpersonal and symptom distress outcomes. Log-linear form was the best fit for social role domain. Final unconditional and conditional models utilised random intercepts and slopes, optimal covariance structures and optimal time trends. As topeplitz covariance structures are fit using generalised least squares there were no random effects to report. Growth curves for unconditional models were visualised using scatter plots. Following the unconditional models (described above), a conditional model was developed using a single predictor of therapeutic modality. As the CAT modality was much smaller than the CBT and PAT modalities, CAT and PAT were merged to form a single '*relational*' treatment group. There were no significant differences between these CAT and PAT on average baseline distress ($p = 0.177$) or number of sessions ($p = 0.115$); and there was no significant difference in baseline distress between the three modalities (see Table 1) so no adjustments were necessary.

Statistical software

All analyses were performed using R (*R Core Team, 2020*, v 4.0.2). Multi-level modeling was conducted using the nlme (*Pinheiro et al., 2020*) package while growth-curve plots were developed using ggplot2 (*Wickham, 2016*). Effect-sizes (Cohen's *d*) and random-effects meta-analysis were computed using the metafor package (*Viechtbauer, 2010*), while forest plots were made using the meta package (*Gordon & Lumley, 2020*).

<div align="center">

Results

</div>

Sample characteristics

364 patients were included in Tertiary Care Service (TCS) evaluation. Almost a third of referrals came from primary care (31.32%). The average age across participants was 42.16 (SD = 11.78) and there was a greater representation of female patients (60.71%). A substantial proportion of patients did not have a recorded employment status (36.26%). Most patients identified as white British (82.69%). Additional sample characteristics are displayed in Table 1. Details regarding primary diagnosis were not recorded during the evaluation however the sample is represented by the four sub teams of TCS, and therefore the sample is likely to represent a blend of the presenting problems that these teams service. These include: (i) the Personality Disorder Team, (ii) the Anxiety and/or and Post-Traumatic Stress Disorder (PTSD) Team, (iii) the Focused Depression Team, and finally (iv) the Obsessive-Compulsive Disorder (OCD) and/or Body Dysmorphic-Disorder (BDD) Team.

[[[-- **Insert Table 1** --]]]

Service benchmarking; baseline severity and effectiveness

The average OQ-45 baseline score was 102.57 (SD = 22.79), which was markedly higher than any of the OQ-45 baseline distress benchmarks (range 73.02-89.17). Effect-sizes (total and subscales) and the selected tertiary and OQ-45 benchmarks are shown in Figure 1. The overall effect-size was small ($d = 0.46$, 95% CI = 0.37- 0.55). CAT produced the largest effect size (medium, $d = 0.64$, 95% CI = 0.32- 0.96) relative to CBT ($d = 0.45$, 95% CI = 0.34- 0.56) and PAT ($d = 0.45$, 95% CI = 0.28- 0.61), but all confidence intervals overlap – indicating no significant differences. Furthermore, none of the CIs for any modality exceeded the service pooled effect-size. When compared to the tertiary care specific benchmarks, the current study effect-size fell below the benchmark (no overlap in CI range). This was also the case when

comparing to the meta-analytic benchmark of practice-based OQ-45 studies, with the benchmark outperforming the current study (no CI overlap).

Treatments; delivery and duration

For the 364 patients included in the effectiveness and recovery analysis, there were an average of 5.64 OQ-45 administrations (SD = 4.41, median = 4.00, range = 2-29) during care episodes lasting 138.7 weeks (SD = 64.69, median = 126.5, range = 15.57- 424.86). The number of completed OQ-45s represented 14.15% of the total number of attended sessions for this cohort of patients. Only 14 (3.85%) patients had care periods shorter than 12-months in duration. The estimated mean number of sessions was 48.68 (SD = 42.14, median = 38.5, range = 5-335) and across all care episodes a mean of 62.47 sessions (SD = 61.54, median = 45.50, range = 5-503). CBT was the most frequently delivered (n = 248), followed by PAT (n = 86) and then CAT (n = 30). In terms of the durations of treatments, almost all included patients received at least 10 sessions (*n* = 354, 97.25%). There was an expected trend for fewer numbers of patients receiving longer treatments, with 308 patients (84.62%) receiving ≥20 sessions, 175 (48.08%) receiving ≥40 sessions, 22 (6.04%) receiving ≥100 sessions, and 10 (2.75%) receiving ≥150 sessions, and 8 (2.20%) receiving ≥200 sessions. There was a statistically significant difference in the average number of sessions delivered between the modalities (F [2, 361], = 10.64, p = <0.001) with PAT being the lengthiest (64.56, SD = 62.38, median = 47.00, range = 5-335), followed by CBT (45.58, SD = 33.33, median = 38.50, range = 5-292) and CAT (28.83, SD = 10.87, median = 27.50, range = 9-63). PAT delivered significantly more sessions than CBT (*p* = 0.001) and CAT (*p* = <0.001); CAT and CBT did not differ in number of sessions delivered (*p* = 0.089).

[[[--- **Insert Figure 1** ---]]]

Individual recovery

Recovery rates are shown in Table 2. The associated Jacobson plot (Figure 2) shows two dense subsamples of patients who were either classed as 'no change' or being 'improved'. There were 18 (4.95%) patients who had sub-clinical scores at baseline, and so these could not reach all criteria required for clinical recovery. In terms of post-treatment status, 37 (10.16%) patients recovered, 109 (29.95%) patients reliably improved, and 29 (7.97%) patients reliably deteriorated. When deterioration and no change rates were combined (i.e., poor outcomes), this accounted for 59.89% of the sample. There were little differences in recovery rates between modalities. In comparison to the pooled benchmark, fewer patients recovered and more had no reliable change. Rates of recovery and no-change were very similar between the service (recovery = 10.16%, no reliable change = 51.92%) and the community mental health centre (CMHC) benchmark (recovery = 8.60%, no reliable change = 60.60%). By comparison to benchmark data, the tertiary care service showed less deterioration (TCS = 7.97%, CMHC = 10.2%) and greater rates of reliable improvement (TCS = 29.95%, CMHC = 20.5%). There was a statistically significant difference in response rates between the service and the overall recovery benchmarks ($p = .001$), but no difference relative to the more closely related community mental health centre benchmarks ($p = 0.098$).

[[[--- **Insert Figure 2** ---]]]

[[[--- **Insert Table 2** ---]]]

Dose-response analysis

Figure 3 shows a bar chart for treatment response in relation to the estimated number of attended treatment sessions. Patients receiving brief treatments (i.e., less than 10 sessions) were highly unlikely to respond (6.04% recovered). Response increased in-line with sessions, until approximately 40 sessions. This trend was mirrored also in term of effectiveness (Table 3). Cohen's *d* pre-post treatment effect-sizes were particularly small for patients receiving less than 10 sessions, and this increased with the number of sessions, until approximately 40 sessions, after which the relationship between treatment duration and response became attenuated.

[[[--- **Insert Figure 3** ---]]]

[[[--- **Insert Table 3** ---]]]

Growth curves

Fixed effects and goodness-of-fit statistics for unconditional and conditional models are shown in Table 4. For the OQ-45 total score, the final unconditional model demonstrated a significant main effect for the intercept (initial score, $\gamma = 104.18$, F = 2100, $p = <0.001$). For sessions, the significant time trends included linear ($\gamma = -0.5$, F = 38.04, $p = <0.001$), quadratic ($\gamma = 0.005$, F = 25.15, $p = <0.001$) and cubic ($\gamma = -0.00001$, F = 21.39, $p = <0.001$). In other words, OQ-45 total score dropped by 0.5 per session, however this was gradually reversed by the curvilinear terms. The growth curve for the OQ-45 total score (Figure 4) shows that improvements begin to dissipate at approximately the 150th session. In the conditional model (including treatment modality), there was no significant main effect for treatment modality ($\gamma = -3.1$, F = 6.42, $p = 0.149$), or treatment-by-sessions interaction ($\gamma = -0.07$, F = 1.88, $p = 0.171$). However, the conditional model did provide a significantly improved model fit ($\chi^2 = 8.24$, $p = 0.016$). As shown in Table 4, the fixed effect for therapy modality (represented by 'Analytic') was not

17

statistically significant for any of the OQ-45 sub-scales. Conditional models for symptom

distress and social role demonstrated significant improved model fit when compared to

unconditional models however this was not the case for interpersonal. Log-linear (social role)

and cubic (symptom distress, interpersonal) growth curves are illustrated in growth curve plots in

Figure 4.

[[[--- **Insert Table 1** ---]]]

[[[--- **Insert Table 4** ---]]]

[[[--- **Insert Figure 4** ---]]]

Discussion

The aim of this study was to explore the effectiveness of treatment in a tertiary care

psychotherapy service designed to meet the needs of patients with difficulties that have been

unresponsive to primary and secondary care psychological interventions. The sample was large

compared to previous UK-based tertiary care studies (n = <50, *Douglas et al., 2016*; *Paley et al.,*

2008) and only two studies were larger in the meta-analytic benchmarking exercise (*Baldwin et*

al., 2009; *Goldberg et al., 2016*). The pooled pre-post therapy effect-size for the study sample

was lower than the service and research benchmarks however we note that baseline severity of

initial distress was higher than any of the available benchmarks. The rates of reliable

improvement and recovery were comparable to a US OQ-45 community mental health centre

benchmark (*Hansen & Lambert, 2002*). The three treatment modalities were equally effective.

Growth curve trajectories demonstrated that OQ-45 scores reduced by 0.5 points per session, but

that this rate declined in line with significant curvilinear trends (quadratic and cubic).

Contribution to the evidence base

Because tertiary care services differ markedly from primary/secondary care services in terms of the type and duration of treatments, it is all the more important to define the clinical population served and to assess how the effectiveness of associated interventions is affected. The intake severity analysis indicates that this patient population was more severely distressed in comparison to US acute/short stay hospital inpatients (e.g., *Doerfler et al., 2002*). This also fits with an evidence base suggesting greater UK intake levels of distress regardless of the measure used (*Francis et al., 1990*; *Ryan, 2007*). The overall effect-size was less than both *Paley et al.* (*2008*) and *Johansson et al.* (*2014*) tertiary care benchmarks and was likely suppressed by those patients having long-treatments but not responding (hence the cubic trend).

The non-linear trends found in the growth curve analysis are consistent with the dose-response evidence base (see review by *Robinson et al., 2020*). The time trends of best fit for the OQ-45 total, symptom distress and interpersonal outcomes were all cubic, suggesting that some patients who remain in therapy for very long periods are non-responders or are at risk of deterioration. More specifically, patients receiving over 100 sessions had a smaller pooled effect-size. This suggests that this patient group tends to make smaller overall improvements, and those improvements mostly occur in the early stages of treatment; although, it should be noted that this is based on a limited sample of patients receiving over 100 appointments ($n = 22$, 6.04%). For social role outcomes (log-linear), improvements were rapid in the early stage of treatment, while further improvements were made with a negatively decelerating rate.

It is unclear why social role was the only sub-scale that continued to make improvements (albeit at a reducing rate), or why interpersonal and symptom distress did not. This pattern is

different to patterns previously identified (*Schilling et al., 2020*; *White et al., 2015*) and may be more characteristic of patients who access UK tertiary care therapy. Interpersonal problems being a predictor of poor treatment outcomes was consistent with prior evidence (*Probst et al., 2020*). Whilst two of the therapies delivered (i.e., PAT and CAT) focus on interpersonal issues, this does not seem to have led to a differential treatment response in this domain. Based on the growth-curve plots of OQ-45 total scores, there was evidence that growth became negative (e.g., indicating deterioration) from approximately 150 sessions. Negative growth was influenced by two sub-scales with cubic trends (symptom distress and interpersonal). In other words, patients in the latter stages of exceptionally long treatments show particularly elevated rates of interpersonal distress and symptom distress.

The evidence of positive response rates for some patients challenges the notion that patients seen in tertiary care are 'treatment resistant' (*Taylor et al., 2012*). However, the study did find that nearly 60% had a deterioration or no change outcome and this may be due to a range of factors prevalent within complex client groups (e.g. *van der Kolk, 1989*), including the influence of social disadvantage, life circumstances and multi-trauma backgrounds (*Finegan et al., 2018*). There was a significant difference between the service recovery rate and the OQ-45 recovery benchmarks, as less patients made full recovery, but more reliably improved. This difference is likely a reflection of the higher initial baseline distress in the sample (i.e., a greater amount of change is required to fall under the clinical threshold to meet full recovery). When compared to the community mental health centre sub-group of the OQ-45 recovery benchmarks, which was suspected to be most comparable to TCS, there was no significant difference in recovery rates.

There was limited evidence for significant differential rates of effectiveness for any single modality. CAT had the highest indices of average (within-group) change (Cohen's *d*), but CIs overlapped with PAT and CBT. Both CAT and CBT showed greater levels of 'service efficiency' based on comparable clinical effectiveness, but in the context of significantly shorter treatment contracts. Some PAT patients received very long treatments, and this skewed the mean upwards. The comparable rates of effectiveness provides further support to the equivalence paradox; that is, bone fide psychotherapies delivered in routine care tend not to significantly differ in terms of effectiveness (Wampold, et al., 1997). However, it is evident that considerably lengthy interventions (>40 sessions) had low improvement rates and may be less cost-effective than briefer interventions, from a health economic perspective.

Limitations

There are a number of study limitations that should be considered. First, given the observational cohort design, it is impossible to say to what extent the observed outcomes were due to the allocated treatment, or due to other potential confounds (e.g., regression to the mean, spontaneous remission). As the OQ-45 was not consistently administered at every session, the level of detail regarding change over time was limited in granularity. The pre-post treatment effect-size calculation did not include patients offered treatment at screening, but who then did not attend any treatment sessions. Concurrent treatments (e.g., pharmacological treatment) were not recorded. There was an absence of treatment fidelity measures; so the extent to which these interventions were delivered with fidelity to bona fide empirically supported procedures is unknown, and the frequency of clinical supervision was not reported. The lack of UK normative data for the OQ-45 means that recovery rates may be inaccurate. The findings were exclusively based upon a single self-report measure, and other relevant outcomes such as wider service

utilisation and adverse incidents were not recorded. Details regarding primary diagnosis were not recorded and subsequently we are unable to determine if there is a differential treatment response based on condition. Finally, all data included in the current study was collected during treatment (i.e., no follow-up) and therefore we are unable to comment on the degree to which the improvements were maintained after treatment ended.

Implications for policy & practice

The finding that initial distress was higher than benchmarks for other sectors fits with the intention that tertiary care services cater for patients with particularly high levels of distress, complexity and impairment. Effect-sizes were likely suppressed by a sizeable sub-group of patients who did not respond well to treatment (around 60%), despite considerable numbers of sessions. On average, improvement rates plateau after around 40 sessions, and growth appears to become negative from approximately 150 sessions in tertiary care patients. Non-responders need to be identified as early as possible, particularly as tertiary care psychotherapy services are offered to those that have not benefited from interventions in primary and secondary care. Alternative management strategies need to be developed and tested for this population. Utilising routine outcome monitoring and feedback systems during clinical supervision could potentially help to rectify trajectories of poor treatment response (*de Jong et al., 2021*), or allow for consideration of earlier necessary treatment termination and referral to alternative support options or services, which may subsequently improve the efficiency of treatments and reduce waiting times for patients who are more likely to respond to briefer psychotherapeutic interventions (e.g., up to 40 sessions).

Conclusions

Despite the delivery of lengthy interventions, treatment outcomes were modest. Service-level indices of effectiveness were seemingly reduced by the inclusion of patients receiving very long treatments but not improving during these interventions, irrespective of the type of treatment modality offered. On this basis, tertiary care services could benefit from adopting more frequent routine outcome monitoring and feedback systems, which are known to help to prevent deterioration and to improve the efficiency of psychological care.

Table 1.
Sample characteristics of patients included in the current study broken down by treatment modality.

		CBT	CAT	PAT	Total	*p*
Patients	N	248	30	86	364	
Age	Mean	42.61	39.67	41.74	42.16	0.404
	SD	11.63	11.24	12.37	11.78	
	Range	18-74	21-61	17-73	17-74	
Baseline OQ-45 Severity	Total Mean	104.04	102.27	98.44	102.57	0.145
	Total SD	22.86	19.32	23.42	22.79	
	SD Mean	65.78	63.53	60.17	64.27	0.007
	SR Mean	16.12	17.2	16.72	16.35	0.494
	IR Mean	22.58	22.73	21.97	22.45	0.757
Sessions in Care Period	Mean	45.58	28.83	64.56	48.68	<0.001
	SD	33.33	10.87	62.38	42.14	

23

Table 1.
Sample characteristics of patients included in the current study broken down by treatment modality.

		CBT	CAT	PAT	Total	*p*
Weeks in Care Period	Weeks	143.82	93.17	141.64	138.7	<0.001
	Range	16-382	41-162	28-425	16-425	
Gender	Female	145 (58.47%)	19 (63.33%)	57 (66.28%)	221 (60.71%)	0.422
	Male	103 (41.53%)	11 (36.67%)	29 (33.72%)	143 (39.29%)	
Ethnicity	White British	207 (83.47%)	26 (86.67%)	68 (79.07%)	301 (82.69%)	0.087
	Any other	11 (4.44%)	1 (3.33%)	3 (3.49%)	15 (4.12%)	
	Not Stated	12 (4.84%)	0 (0.00%)	3 (3.49%)	15 (4.12%)	
	Black	9 (3.63%)	2 (6.67%)	3 (3.49%)	14 (3.85%)	
	Asian	8 (3.23%)	0 (0.00%)	3 (3.49%)	11 (3.02%)	
	White Other	1 (0.40%)	1 (3.33%)	6 (6.98%)	8 (2.20%)	
Work Status	Not Known/Other	92 (37.10%)	8 (26.67%)	32 (37.21%)	132 (36.26%)	0.006
	Employed	55 (22.18%)	9 (30.00%)	33 (38.37%)	97 (26.65%)	
	Unemployed	53 (21.37%)	8 (26.67%)	7 (8.14%)	68 (18.68%)	
	Sick/Disabled	31 (12.50%)	3 (10.00%)	3 (3.49%)	37 (10.16%)	
	Student	14 (5.65%)	2 (6.67%)	7 (8.14%)	23 (6.32%)	

Table 1.

Sample characteristics of patients included in the current study broken down by treatment modality.

		CBT	CAT	PAT	Total	*p*
	Retired	3 (1.21%)	0 (0.00%)	4 (4.65%)	7 (1.92%)	
Marital Status	Married or settled	99 (39.92%)	14 (46.67%)	32 (37.21%)	145 (39.84%)	0.661
	Single	96 (38.71%)	12 (40.00%)	33 (38.37%)	141 (38.74%)	
	Other	37 (14.92%)	1 (3.33%)	15 (17.44%)	53 (14.56%)	
	Divorced or separated	16 (6.45%)	3 (10.00%)	6 (6.98%)	25 (6.87%)	
Referrer	Other	120 (48.39%)	13 (43.33%)	40 (46.51%)	173 (47.53%)	0.819
	GP	72 (29.03%)	12 (40.00%)	30 (34.88%)	114 (31.32%)	
	Psychiatry	32 (12.90%)	3 (10.00%)	11 (12.79%)	46 (12.64%)	
	IAPT	17 (6.85%)	2 (6.67%)	3 (3.49%)	22 (6.04%)	
	External Hospital	3 (1.21%)	0 (0.00%)	2 (2.33%)	5 (1.37%)	
	Dental	4 (1.61%)	0 (0.00%)	0 (0.00%)	4 (1.10%)	

Note. SD = Symptom Distress, SR = Social Role, IR = Interpersonal

Table 2.
Rates of reliable change, recovery, and deterioration for the current study sample and for selected benchmarks (Hansen and Lambert 2002).

Study		Recovered	No Change	Deterioration	Improved	Total
Hansen (2002)	Total	681 (14.3%)	2709 (56.9%)	377 (7.9%)	994 (20.9%)	4761
	CMHC	31 (8.6%)	219 (60.6%)	37 (10.2%)	74 (20.5%)	361
TCS	Total	37 (10.16%)	189 (51.92%)	29 (7.97%)	109 (29.95%)	364
	CBT	22 (8.87%)	128 (51.61%)	23 (9.27%)	75 (30.24%)	248
	PDT	11 (12.79%)	46 (53.49%)	4 (4.65%)	25 (29.07%)	86
	CAT	4 (13.33%)	15 (50.00%)	2 (6.67%)	9 (30.00%)	30

Note. 18 patients fell within in the non-clinical range at baseline. No change = no reliable changes Improved = Reliable Improvement Deterioration = reliable Deterioration

Table 3.
Non-cumulative, differential rates of statistical and clinical change based on different dosage groups.

Sessions	n	*d*	Ci	Recovered	No Change	Deteriorated	Improved
1-10	81	0.15	-0.04-0.35	4 (4.94%)	50 (61.73%)	9 (11.11%)	18 (22.22%)
10-19	88	0.51	0.31-0.71	8 (9.09%)	48 (54.55%)	7 (7.95%)	25 (28.41%)
20-29	77	0.63	0.41-0.85	14 (18.18%)	32 (41.56%)	9 (11.69%)	22 (28.57%)
30-39	37	0.94	0.58-1.3	5 (13.51%)	16 (43.24%)	0 (0.00%)	16 (43.24%)
40-49	26	0.54	0.17-0.91	0 (0.00%)	14 (53.85%)	1 (3.85%)	11 (42.31%)
50-69	31	0.25	-0.07-0.57	3 (9.68%)	18 (58.06%)	2 (6.45%)	8 (25.81%)
70-99	16	0.67	0.17-1.16	1 (6.25%)	9 (56.25%)	0 (0.00%)	6 (37.50%)
Over 100	8	0.46	-0.2-1.12	2 (25.00%)	2 (25.00%)	1 (12.50%)	3 (37.50%)

Table 4
Fixed effects and goodness-of-fit statistics for optimal unconditional and final conditional models for the OQ-45 and each of the three sub-scales.

	Total Score		Symptom Distress		Social Role		Interpersonal	
	OQ Cubic	**OQ Cond.**	**SD Log**	**SD Cond.**	**SR Log**	**SR Cond.**	**IR Quad**	**IR Cond.**
Fixed Effects								
Intercept	104.180*	105.453*	64.725*	65.403*	17.668*	17.847*	22.973*	23.151*
	(1.497)	(1.693)	(0.962)	(1.246)	(0.407)	(0.515)	(0.434)	(0.496)
Linear/Log	−0.496*	−0.494*	−0.293*	−0.275*	−0.949*	−1.021*	−0.103*	−0.102*
	(0.057)	(0.058)	(0.036)	(0.070)	(0.129)	(0.165)	(0.017)	(0.017)
Quadratic	0.005*	0.005*	0.003*	0.003*			0.001*	0.001*
	(0.001)	(0.001)	(0.000)	(0.001)			(0.000)	(0.000)
Cubic	−0.000*	−0.000*	−0.000*	−0.000*			−0.000*	−0.000*
	(0.000)	(0.000)	(0.000)	(0.000)			(0.000)	(0.000)
Analytic		−3.102		−1.107		−0.425		−0.424
		(2.134)		(1.641)		(0.768)		(0.634)
Goodness of fit								
R^2	0.014	0.042	0.008	0.049			0.009	0.016
AIC	17216.0	17211.7	15199.3	15193.4	12113.7	12117.7	12166.7	12168.6
BIC	17261.2	17268.2	15244.5	15255.6	12147.6	12162.9	12211.9	12225.1
Log-Likelihood	-8599.99	-8595.87	-7591.66	-7585.72	-6049.34	-6049.08	-6075.34	-6074.32
p value		0.016		0.008		0.774		0.362

$* \, p < 0.05$
AIC = Akaike information criterion; BIC = Bayesian information criterion;
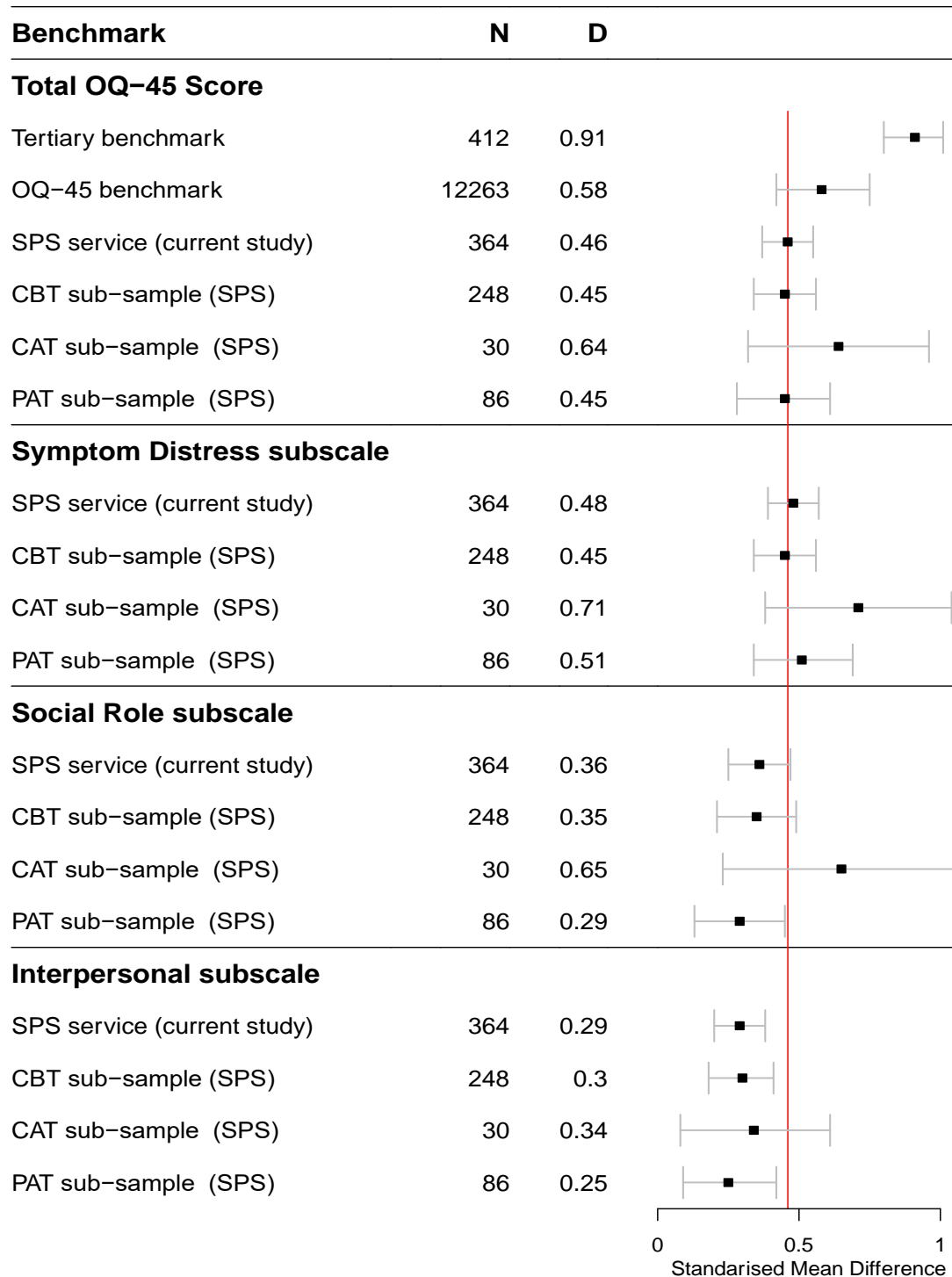OQ = Outcome questionnaire; SD = Symptom distress;
SR = Symptom Role; IR = Interpersonal relationships;
Cond = Conditional model (including treatment); Log = Logarithmic model.
Quad = Quadratic model

Figure 1.

Forest plot of pre-post therapy effect sizes for the current study sample and selected effectiveness benchmarks. The square boxes depict individual Cohen's d effect sizes and error bars display 95 percent confidence intervals. The red horizontal line represents the Cohen's d effect-size for the pooled TCS sample.

Benchmark	N	D
Total OQ−45 Score		
Tertiary benchmark	412	0.91
OQ−45 benchmark	12263	0.58
SPS service (current study)	364	0.46
CBT sub−sample (SPS)	248	0.45
CAT sub−sample (SPS)	30	0.64
PAT sub−sample (SPS)	86	0.45
Symptom Distress subscale		
SPS service (current study)	364	0.48
CBT sub−sample (SPS)	248	0.45
CAT sub−sample (SPS)	30	0.71
PAT sub−sample (SPS)	86	0.51
Social Role subscale		
SPS service (current study)	364	0.36
CBT sub−sample (SPS)	248	0.35
CAT sub−sample (SPS)	30	0.65
PAT sub−sample (SPS)	86	0.29
Interpersonal subscale		
SPS service (current study)	364	0.29
CBT sub−sample (SPS)	248	0.3
CAT sub−sample (SPS)	30	0.34
PAT sub−sample (SPS)	86	0.25

Standarised Mean Difference

SPS = Specialist Psychotherapy Service (i.e., current study service).

28

Figure 2.

Jacobson plot to show the rates of patient response. Points to the right of the vertical dashed line represent patients who started treatment as clinically distressed. Points beneath the horizontal dashed line represent patients who finished treatment in the non-clinically distressed range.
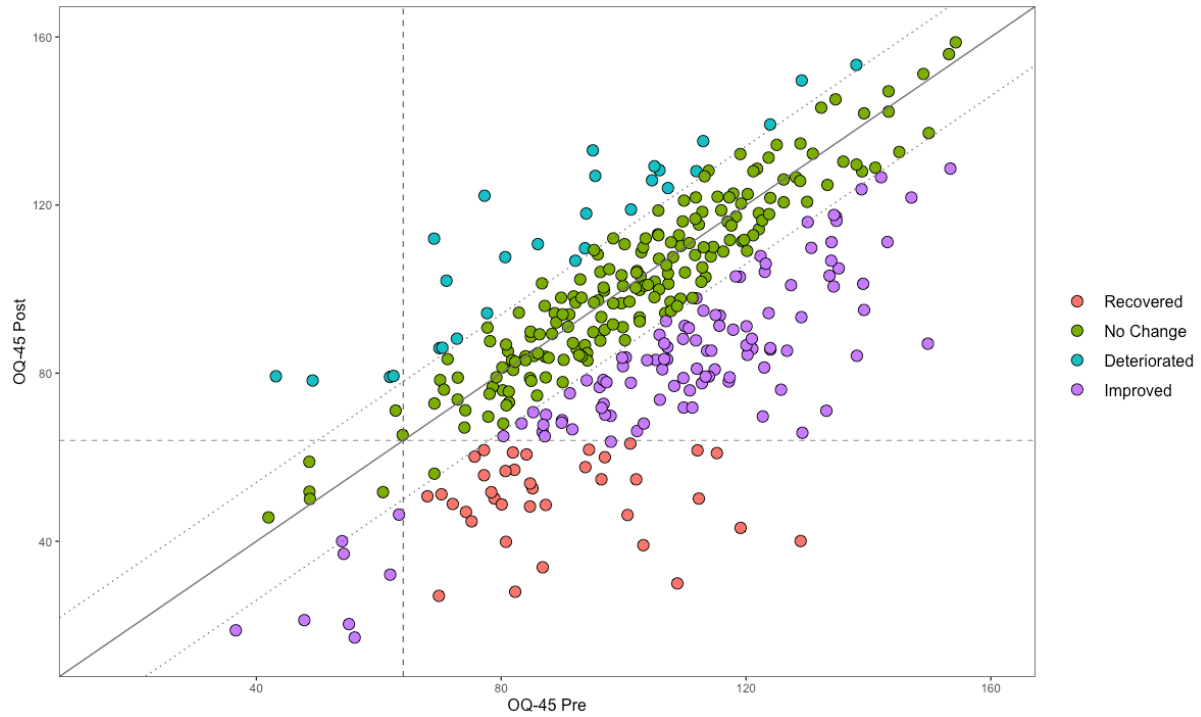
Figure 3.

Rates of response to tertiary care therapy, based on number of sessions received. Response here is the sum of patients who showed reliable improvement. Bars show rate of improvement by number of sessions received. The red line/text denotes the cumulative number of patients who had improved by the number of sessions received.
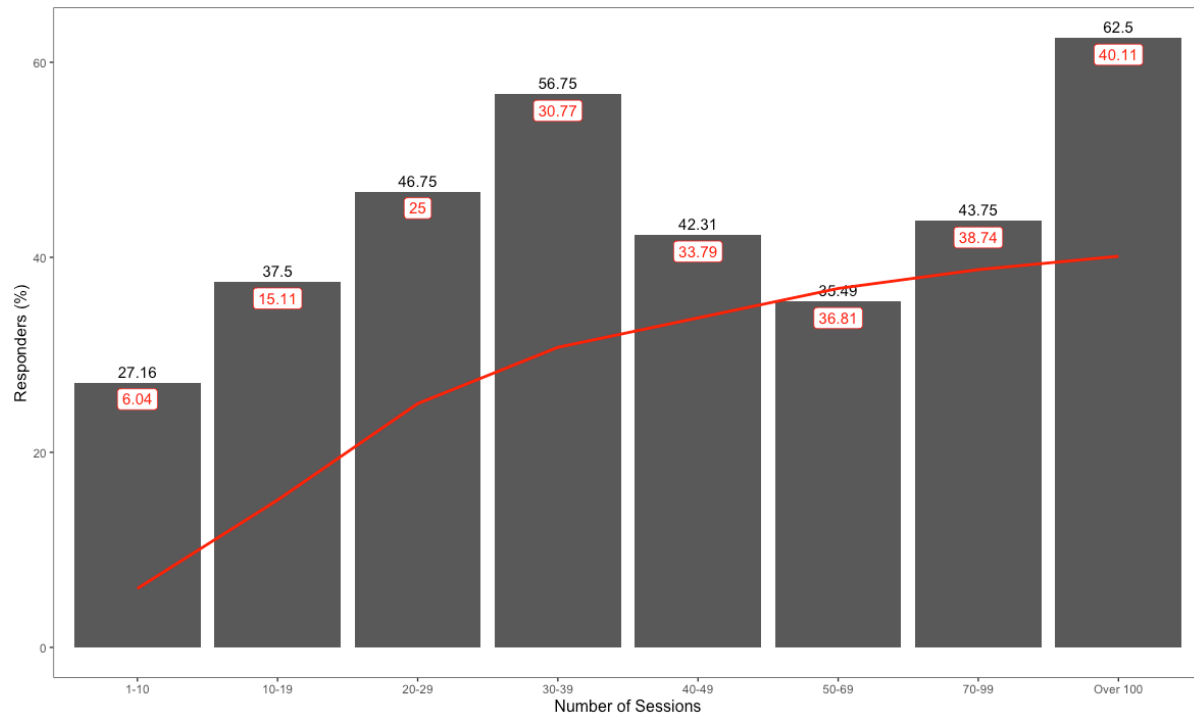
Figure 4.

Growth curves for optimal unconditional models. Blue lines represent growth curve trajectories. Grey shaded regions represent 95% confidence interval regions. Trends line types represent cubic for OQ-45 total score, interpersonal and symptom distress. Social role is represented by a log-linear trend.
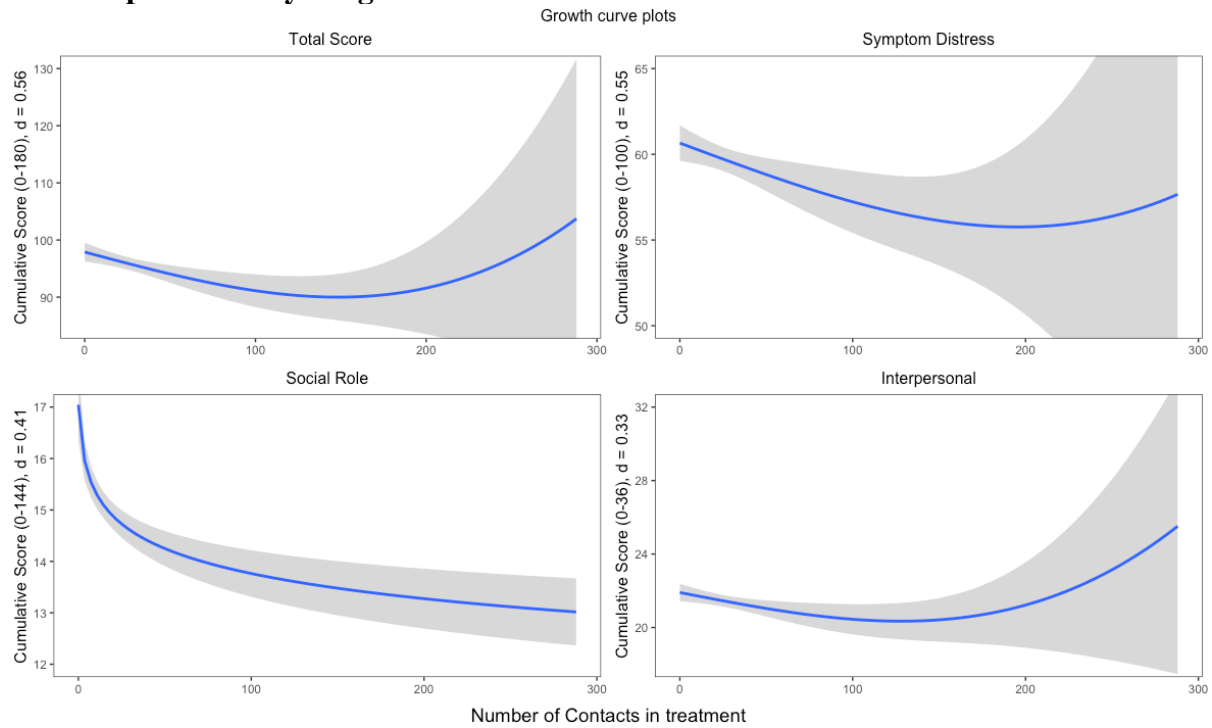
References

Abbass, A. A., Joffres, M. R., & Ogrodniczuk, J. S. (2008). A naturalistic study of intensive short-term dynamic psychotherapy trial therapy. *Brief Treatment and Crisis Intervention*, *8*(2), 164–170. *https://doi.org/10.1093/brief-treatment/mhn001*

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. *Journal of Consulting and Clinical Psychology*, *77*(2), 203–211. *https://doi.org/10.1037/a0015235*

Blainey, S. H., Rumball, F., Mercer, L., Evans, L. J., & Beck, A. (2017). An evaluation of the effectiveness of psychological therapy in reducing general psychological distress for adults with autism spectrum conditions and comorbid mental health problems. *Clinical Psychology & Psychotherapy*, *24*(6), 474–484. *https://doi.org/10.1002/cpp.2108*

Bone, C., Delgadillo, J., & Barkham, M. (2021). A systematic review and meta-analysis of the good-enough level (GEL) literature. *Journal of Counseling Psychology*, *68*(2), 219–231. *https://doi.org/10.1037/cou0000521*

Burns, T. (2004). Community mental health teams. *Psychiatry, 3*, 11-14. *https://doi.org/10.1383/psyt.3.9.11.50258*

Cahill, J., Barkham, M., & Stiles, W. (2010). Systematic review of practice-based research on psychological therapies in routine clinic settings. *The British Journal of Clinical Psychology*, *49*(4), 421–453. *https://doi.org/10.1348/014466509X470789*

Clark, D. M. (2018). Realising the mass public benefit of evidence-based psychological

    therapies: The IAPT program. *Annual Review of Clinical Psychology, 14*, 159-183.

    https://doi.org/10.1146%2Fannurev-clinpsy-050817-084833

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum

    Associates.

de Jong, K., Conijn, J. M., Gallagher, R. A. V., Reshetnikova, A. S., Heij, M., & Lutz, M. C.

    (2021). Using progress feedback to improve outcomes and reduce drop-out, treatment

    duration, and deterioration: A multilevel meta-analysis. *Clinical Psychology Review*, *85*,

    102002. *https://doi.org/10.1016/j.cpr.2021.102002*

Delgadillo, J., Asaria, M., Ali, S., & Gilbody, S. (2016). On poverty, politics and psychology:

    The socioeconomic gradient of mental healthcare utilisation and outcomes. *The British*

    Journal of Psychiatry, *209*(5), 429–430. *https://doi.org/10.1192/bjp.bp.115.171017*

Delgadillo, J., & Gonzalez Salas Duhne, P. (2020). Targeted prescription of cognitivebehavioral

    therapy versus person-centered counseling for depression using a machine learning

    approach. *Journal of Consulting and Clinical Psychology*, *88*(1), 14–24.

    https://doi.org/10.1037/ccp0000476

Delgadillo, J., McMillan, D., Leach, C., Lucock, M., Gilbody, S., & Wood, N. (2014).

    Benchmarking routine psychological services: A discussion of challenges and methods.

    Behavioural and Cognitive Psychotherapy, *42*(1), 16–30.

    https://doi.org/10.1017/S135246581200080X

Department of Health. (2004). *Organising and delivering psychological therapies*. Department

of Health.

Doerfler, L. A., Addis, M. E., & Moran, P. W. (2002). Evaluating mental health outcomes in an

inpatient setting: Convergent and divergent validity of the OQ-45 and BASIS-32. *The

Journal of Behavioral Health Services & Research*, *29*(4), 10.

Douglas, A., Ablett-Tate, N., & Chadd, N. (2016). Dynamic interpersonal therapy in an NHS

tertiary level specialist psychotherapy service. *Psychoanalytic Psychotherapy*, *30*(3),

223–239. *https://doi.org/10.1080/02668734.2016.1198415*

Finegan, M., Firth, N., Wojnarowski, C., & Delgadillo, J. (2018). Associations between

socioeconomic status and psychological therapy outcomes: A systematic review and

meta-analysis. *Depression and Anxiety*, *35*(6), 560–573.

https://doi.org/10.1002/da.22765

Firth, N., Saxon, D., Stiles, W. B., & Barkham, M. (2020). Therapist effects vary significantly

across psychological treatment care sectors. *Clinical Psychology & Psychotherapy*, *27*(5),

770–778. *https://doi.org/10.1002/cpp.2461*

Francis, V. M., Rajan, P., & Turner, N. (1990). British community norms for the Brief Symptom

Inventory. *British Journal of Clinical Psychology*, *29*(1), 115–116.

https://doi.org/10.1111/j.2044-8260.1990.tb00857.x

Goldberg, S. B., Miller, S. D., Nielsen, S. L., Rousmaniere, T., Whipple, J., Hoyt, W. T., &

Wampold, B. E. (2016). Do psychotherapists improve with time and experience? A

longitudinal analysis of outcomes in a clinical setting. *Journal of Counseling Psychology*, *63*(1), 1–11. *https://doi.org/10.1037/cou0000131*

Gordon, M., & Lumley, T. (2020). *Forestplot: Advanced forest plot using 'grid' graphics*. *https://CRAN.R-project.org/package=forestplot*

Hallam, C., Simmonds-Buckley, M., Kellett, S., Greenhill, B., & Jones, A. (2021). The acceptability, effectiveness, and durability of cognitive analytic therapy: Systematic review and meta-analysis. *Psychology and Psychotherapy*, *94*(1), 8–35. *https://doi.org/10.1111/papt.12286*

Hansen, N. B., & Lambert, M. J. (2002). An evaluation of the doseresponse relationship in naturalistic treatment settings using survival analysis. *Mental Health Services Research*, *5*(1), 1–12.

Heins, M. J., Knoop, H., Lobbestael, J., & Bleijenberg, G. (2011). Childhood maltreatment and the response to cognitive behavior therapy for chronic fatigue syndrome. *Journal of Psychosomatic Research*, *71*(6), 404–410. *https://doi.org/10.1016/j.jpsychores.2011.05.005*

Howard, K., Kopta, S., Krause, M., & Orlinsky, D. (1986). The doseeffect relationship in psychotherapy. *American Psychologist*, *41*(2), 159–164.

Jacobson, N. S., & Truax, P. (1992). Clinical significance : A statistical approach to defining meaningful change in psychotherapy research. In *Methodological issues & strategies in clinical research* (pp. 631–648). American Psychological Association. *https://doi.org/10.1037/10109-042*

Johansson, R., Town, J. M., & Abbass, A. (2014). Davanloo's intensive short-term dynamic

    psychotherapy in a tertiary psychotherapy service: Overall effectiveness and association

    between unlocking the unconscious and outcome. *PeerJ*, *2*, Article e548.

    https://doi.org/10.7717/peerj.548

Johns, R. G., Barkham, M., Kellett, S., & Saxon, D. (2019). A systematic review of therapist

    effects: A critical narrative update and refinement to Baldwin and Imel's (2013) review.

    Clinical Psychology Review, *67*, 78–93. *https://doi.org/10.1016/j.cpr.2018.08.004*

Jong, K. de, Nugter, M. A., Polak, M. G., Wagenborg, J. E. A., Spinhoven, P., & Heiser, W. J.

    (2007). The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural

    validation. *Clinical Psychology & Psychotherapy*, *14*(4), 288–301.

    https://doi.org/10.1002/cpp.529

Kolly, S., Kramer, U., Maillard, P., Charbon, P., Droz, J., Fresard, E., Berney, S., & Despland,

    J.-N. (2015). Psychotherapy for personality disorders in a natural setting: A pilot study

    over two years of treatment. *The Journal of Nervous and Mental Disease*, *203*(9), 735–

    738. *https://doi.org/10.1097/NMD.0000000000000356*

Kopta, S. M., Howard, K. I., Lowry, J. L., & Beutler, L. E. (1994). Patterns of symptomatic

    recovery in psychotherapy. *Journal of Consulting and Clinical Psychology*, *62*(5), 1009–

    1016. *https://doi.org/10.1037/0022-006X.62.5.1009*

Kwok, O.-M., Underhill, A. T., Berry, J. W., Luo, W., Elliott, T. R., & Yoon, M. (2008).

    Analyzing longitudinal data with multilevel models: An example with individuals living

    with lower extremity intra-articular fractures. *Rehabilitation Psychology*, *53*(3), 370–386.

    https://doi.org/10.1037/a0012765

Lambert, M. J. (2004). *Administration and scoring manual for the OQ-45.2 (Outcome questionnaire)*. OQ Measures, LLC.

Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C., & Yanchar, S. C. (1996). The reliability and validity of the outcome questionnaire. *Clinical Psychology & Psychotherapy*, *3*(4), 249–258. *https://doi.org/10.1002/(SICI)1099-0879(199612)3:4<249::AID-CPP106>3.0.CO;2-S*

Lambert, M. J., Gregersen, A. T., & Burlingame, G. M. (2004). The Outcome Questionnaire-45. In M. E. Maurish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment: Instruments for adults* (Vol. 3, pp. 191–234). Lawrence Erlbaum Associates Publishers.

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology*, *69*(2), 159–172. *https://doi.org/10.1037/0022-006X.69.2.159*

Lilliengren, P., Cooper, A., Town, J. M., Kisely, S., & Abbass, A. (2020). Clinical- and cost-effectiveness of intensive short-term dynamic psychotherapy for chronic pain in a tertiary psychotherapy service. *Australasian Psychiatry*, *28*(4), 414–417. *https://doi.org/10.1177/1039856220901478*

Lunnen, K. M., Ogles, B. M., & Pappas, L. N. (2008). A multiperspective comparison of satisfaction, symptomatic change, perceived change, and end-point functioning. *Professional Psychology: Research and Practice*, *39*(2), 145–152. *https://doi.org/10.1037/0735-7028.39.2.145*

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92. *https://doi.org/10.1027/1614-1881.1.3.86*

Mallinckrodt, C. H., Clark, W. S., & David, S. R. (2001). Accounting for dropout bias using mixed-effects models. *Journal of Biopharmaceutical Statistics*, *11*(1-2), 9–21. *https://doi.org/10.1081/BIP-100104194*

Minami, T., Wampold, B. E., Serlin, R. C., Hamilton, E. G., Brown, G. S. J., & Kircher, J. C. (2008). Benchmarking the effectiveness of psychotherapy treatment for adult depression in a managed care environment: A preliminary study. *Journal of Consulting and Clinical Psychology*, *76*(1), 116–124. *https://doi.org/10.1037/0022-006X.76.1.116*

Morley, S., & Dowzer, C. (2014). *Manual for the Leeds Reliable Change Indicator: Simple Excel(tm) applications for the analysis of individual patient and group data.*

Nordmo, M., Monsen, J. T., Høglend, P. A., & Solbakken, O. A. (2021). Investigating the dose–response effect in open-ended psychotherapy. *Psychotherapy Research, 31*(7), 859-869. https://doi.org/10.1080/10503307.2020.1861359

Nowoweiski, D., Abbass, A., Town, J., Keshen, A., & Kisely, S. (2020). An observational study of the treatment and cost effectiveness of intensive short-term dynamic psychotherapy on a cohort of eating disorder patients. *Journal of Psychiatry and Behavioral Sciences*, *3*(1), Article 1030.

Paley, G., Cahill, J., Barkham, M., Shapiro, D., Jones, J., Patrick, S., & Reid, E. (2008). The effectiveness of psychodynamic-interpersonal therapy (PIT) in routine clinical practice:

A benchmarking comparison. *Psychology and Psychotherapy: Theory, Research and Practice*, *81*(2), 157–175. *https://doi.org/10.1348/147608307X270889*

Pinheiro, J., Bates, D., & R-core. (2020). *Nlme: Linear and nonlinear mixed effects models*. *https://svn.r-project.org/R-packages/trunk/nlme/*

Probst, T., Kleinstäuber, M., Lambert, M. J., Tritt, K., Pieh, C., Loew, T. H., Dahlbender, R. W., & Delgadillo, J. (2020). Why are some cases not on track? An item analysis of the Assessment for Signal Cases during inpatient psychotherapy. *Clinical Psychology & Psychotherapy*, *27*(4), 559–566. *https://doi.org/10.1002/cpp.2441*

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. *https://www.R-project.org/*

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. SAGE.

Robinson, L., Delgadillo, J., & Kellett, S. (2020). The dose-response effect in routinely delivered psychological therapies: A systematic review. *Psychotherapy Research*, *30*(1), 79–96. *https://doi.org/10.1080/10503307.2019.1566676*

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

Ryan, C. (2007). British outpatient norms for the Brief Symptom Inventory. *Psychology and Psychotherapy: Theory, Research and Practice*, *80*(2), 183–191. *https://doi.org/10.1348/147608306X111165*

Ryle, A., & Kerr, Ian. B. (2020). The main features of CAT. In *Introducing cognitive analytic therapy* (2nd ed, pp. 9–29). John Wiley & Sons, Ltd. *https://doi.org/10.1002/9781119375210.ch2*

Schilling, V. N. L. S., Zimmermann, D., Rubel, J. A., Boyle, K. S., & Lutz, W. (2020). Why do patients go off track? Examining potential influencing factors for being at risk of psychotherapy treatment failure. *Quality of Life Research.* *https://doi.org/10.1007/s11136-020-02664-6*

Simmonds-Buckley, M., Osivwemu, E.O., Kellett, S., & Taylor, C. (2022). The acceptability of cognitive analytic therapy (CAT): Meta-analysis and benchmarking of treatment refusal and treatment dropout rates. *Clinical Psychology Review, 96:*102187. doi: 10.1016/j.cpr.2022.102187.

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

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, *32*(9), 752–760. *https://doi.org/10.1037/0003-066X.32.9.752*

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*, *77*(4), 595–606. *https://doi.org/10.1037/a0016032*

Stiles, W. B., Barkham, M., Twigg, E., Mellor-Clark, J., & Cooper, M. (2006). Effectiveness of cognitive-behavioural, person-centred and psychodynamic therapies as practised in UK

National Health Service settings. *Psychological Medicine*, *36*(4), 555–566.

https://doi.org/10.1017/S0033291706007136

Taylor, D., Carlyle, J., McPherson, S., Rost, F., Thomas, R., & Fonagy, P. (2012). Tavistock

Adult Depression Study (TADS): A randomised controlled trial of psychoanalytic

psychotherapy for treatment-resistant/treatment-refractory forms of depression. *BMC*

Psychiatry, *12*(1), 60. *https://doi.org/10.1186/1471-244X-12-60*

van der Kolk, B. A. (1989). The compulsion to repeat the trauma: Re-enactment, revictimization,

and masochism. *Psychiatric Clinics of North America*, *12*(2), 389–411.

https://doi.org/10.1016/S0193-953X(18)30439-8

Vermeersch, D. A., Lambert, M. J., & Burlingame, G. M. (2000). Outcome Questionnaire: Item

sensitivity to change. *Journal of Personality Assessment*, *74*(2), 242–261.

https://doi.org/10.1207/S15327752JPA7402_6

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

Statistical Software, *36*(3), 1–48. *https://www.jstatsoft.org/v36/i03/*

Wakefield, S., Kellett, S., Simmonds-Buckley, M., Stockton, D., Bradbury, A., & Delgadillo, J.

(2021). Improving Access to Psychological Therapies (IAPT) in the United Kingdom: A

systematic review and meta-analysis of 10-years of practice-based evidence. *British*

Journal of Clinical Psychology, *60*(1), 1–37. *https://doi.org/10.1111/bjc.12259*

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-

analysis of outcome studies comparing bona fide psychotherapies: Empiricially, "all must

have prizes". *Psychological Bulletin*, *122*(3), 203–215. *https://doi.org/10.1037/0033-2909.122.3.203*

Warden, S., Ricketts, T., Saxon, D., Houghton, S., St. Ledger, S., Curran, J., & Fitzgerald, G. (2008). When to offer cognitive behavioural or psychoanalytic psychotherapy in an integrated psychotherapy service: Are everyday allocation decisions theoretically congruent? *Counselling and Psychotherapy Research*, *8*(2), 102–109. *https://doi.org/10.1080/14733140801972604*

Wennberg, P., Philips, B., & Jong, K. de. (2010). The Swedish version of the Outcome Questionnaire (OQ-45): Reliability and factor structure in a substance abuse sample. *Psychology and Psychotherapy: Theory, Research and Practice*, *83*(3), 325–329. *https://doi.org/10.1348/147608309X478715*

White, M. M., Lambert, M. J., Ogles, B. M., Mclaughlin, S. B., Bailey, R. J., & Tingey, K. M. (2015). Using the assessment for signal clients as a feedback tool for reducing treatment failure. *Psychotherapy Research*, *25*(6), 724–734. *https://doi.org/10.1080/10503307.2015.1009862*

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. *https://ggplot2.tidyverse.org*

Worm-Smeitink, M., Nikolaus, S., Goldsmith, K., Wiborg, J., Ali, S., Knoop, H., & Chalder, T. (2016). Cognitive behaviour therapy for chronic fatigue syndrome: Differences in treatment outcome between a tertiary treatment centre in the United Kingdom and the Netherlands. *Journal of Psychosomatic Research*, *87*, 43–49. *https://doi.org/10.1016/j.jpsychores.2016.06.006*

Youn, S. J., Xiao, H., McAleavey, A. A., Scofield, B. E., Pedersen, T. R., Castonguay, L. G.,

    Hayes, J. A., & Locke, B. D. (2019). Assessing and investigating clinicians' research

    interests: Lessons on expanding practices and data collection in a large practice research

    network. *Psychotherapy, 56*(1), 67–82. *https://doi.org/10.1037/pst0000192*