Statistical Analysis Plan
TherapyMatch-D: A pilot cluster randomised controlled trial of psychological treatment selection for depression

Short title of study	
TherapyMatch-D	
Protocol version and date	**Research Ethics Committee (REC) reference**
(v2) 11th August 2023	23/LO/0487
Sponsor organisation	**Controlled trials registration number**
Rotherham Doncaster and South Humber NHS Foundation Trust	ISRCTN21721966

Research team

Jaime Delgadillo[1,2], Paulina Gonzalez Salas Duhne[1,2], Michael Barkham[1], David Saxon[1]

1 – Clinical and Applied Psychology Unit, Department of Psychology, The University of Sheffield, UK

2 – Grounded Research Team, Doncaster and South Humber NHS Foundation Trust, UK

Chief Investigator contacts details

J. Delgadillo

Clinical and Applied Psychology Unit, Department of Psychology, The University of Sheffield

Cathedral Court, Floor F, 1 Vicar Lane, Sheffield S1 1HD

Tel: 0114 222 6614; Email: j.delgadillo@sheffield.ac.uk

Sample size calculation

Pilot trials do not require a formal sample size calculation. However, a sample of at least 64 participants per arm is desirable (Teare et al., 2014) for a binary outcome in a pilot trial. We will aim to recruit 6-10 clusters (3-5 sites randomised in each arm) to allow subsequent sample size calculation for a definitive RCT. Within each Talking Therapies site, we will aim to recruit a minimum of 1 therapist that carry out routine assessment (maximum of 2 therapists per site). Each Talking Therapies site will aim to recruit 64 patients, For this, a total of 432 patients would be screened (216 patients per arm). As we expect approximately 30% of patients across sites to have a differential response, we estimate there would need to be 216 patients screened (64 patients per arm; 216*30% = 64). This calculation is based on the assumption of equal availability for treatment. As a pilot trial, any differences in treatment offer and waiting time for the different treatments will be reported and considered for potential subsequent research.

Overall, we will aim to recruit a minimum of 6-10 therapists that carry out routine assessments across the 6-10 Talking Therapies services (1 to 2 per Talking Therapies service). Between the 6 to 20 therapists, we expect that they will assess at least 432 during a 1-year study period, which would require each therapist to assess 1 to 2 cases per week on average.

Primary analysis

Patient-level clinical outcome data will be analysed using logistic multilevel modelling conducted according to the intention-to-treat principle according to CONSORT guidelines for Cluster RCTs. Researchers will be blinded to the group allocation while analysing the trial data.

The primary outcome will be defined as reliable and clinically significant improvement (RCSI) in depression symptoms (PHQ-9) after treatment, based on the method described by Jacobson and Truax (1991). A 3-level model will be applied, with patients nested within therapists, nested within Talking Therapies sites, and post-treatment RCSI in depression (PHQ-9) symptoms as the dependent variable. Group (TherapyMatch-D vs. usual care) will be entered as a level-2 predictor, along with baseline PHQ-9 as a level-one covariate. This method will enable us to determine whether TherapyMatch-D is associated with a greater treatment effect (depression symptom reduction) by comparison to usual care and will be specifically run in the target sample of patients identified as having a differential response to treatment. The three-level model will account for the nested structure of the data, as appropriate within a cluster trial design. Given the small sample sizes of this pilot trial, it is unlikely the estimates of random effects for site and therapist will be reliable, but this data may be used to guide subsequent research. If the site and therapist random intercepts are not statistically significant, a one-level parsimonious multivariate regression model will be estimated as a sensitivity analysis. Adjusted odds ratios and confidence intervals will be reported as a primary effect sizes. In addition, we will also report the results in the best fitting model following analyses of goodness of fit to find the most parsimonious model.

We acknowledge that some patients might decide to switch from one therapy to another following allocation in either arm of the trial (i.e., patient allocated to CBT after consenting to treatment asks to switch to PCE-CfD), as occurs in routine practice occasionally. In this pilot trial, we will report the frequency of this occurrence and circumstances. Cases will be included in the primary analysis (as part of the allocation group to which they were originally allocated to), and secondary analyses will exclude these cases to examine the extent to which this data might contaminate the results. The same

reporting and inclusion of cases for primary analyses will be followed in cases where patients and clinicians select another High Intensity Treatment beyond the two examined in this trial.

Secondary analyses

The above modelling strategy will be repeated to compare the rate of improvement (RCSI) in GAD-7 anxiety symptoms, and RCSI in depression symptoms only in the cases who attended at least one treatment session. We will also compare outcome expectancy, dropout rates and adverse effects between groups, and compare characteristics between those who attended at least one treatment session versus those who dropped out following assessment. For both TherapyMatch-D and control groups, we will estimate the proportions of participants who had RCSI post-treatment, and who completed/dropped out of treatment. We will also calculate and report effect sizes for each of the symptom measures (PHQ-9, GAD-7, WSAS) using Cohen's *d* metric (Cohen, 1988), based on mean differences between groups. Treatment attendance, completion, and dropout rates will be presented diagrammatically and based on CONSORT guidelines. To quantify any potential adverse effects related to the trial, we will also calculate a rate of reliable deterioration in patients across both arms of the trial. Further, we will use the data collected to identify complex cases (see StratCare Trial from Delgadillo et al., 2021) and control for variability in complex cases across Talking Therapies sites.

We will examine adherence to the TherapyMatch-D model using the full study sample of participants with a differential treatment response across both arms. Each therapist's Treatment-Matching Precision Score (TMaP) will be calculated, which is a statistical measure of agreement between their observed treatment selection and the recommendations made using the prognostic algorithm. In the experimental group, this would represent the agreement of patients accepting the recommended treatment. In the control group, this is the post-hoc analysis of the agreement the optimal treatment that the prescriptive algorithm would recommend (which patients did not receive, since they were in the control group) and the actual treatment selected via shared decision making. A comparison of mean TMaP scores between groups will be done using parametric or non-parametric tests, depending on the distribution of TMaP scores.

The wider cohort dataset (i.e., non-differential responders who consented to take part in this trial) will be used to examine changes in treatment recommendation patterns over time and treatment outcomes, modelling longitudinal changes in the prevalence of treatment recommendation. These additional analyses with non-differential responders will serve only to maximize the use of available data to inform a future trial, but the results will not be included in the main analyses.

Basic demographic data collected from therapists providing the two psychological treatments to describe the overall workforce (e.g., age, gender, caseload, and qualifications) will be reported in the study, to identify any potential differences in staff providing the intervention. No further analysis of this data is planned a priori. If large differences were found between therapists providing interventions, post-hoc exploratory analysis may be conducted as part of a feasibility trial that may inform future research.