RESEARCH PROTOCOL
UpLift-X Trial 1: A cluster factorial randomised controlled trial of digitally-enabled group psychotherapy for common mental disorders

Short title of study	Protocol version and date
UpLift-X Trial 1	(v5.1) 06.06.25 (RDaSH0348)
Sponsor organisation	**Funder**
Rotherham Doncaster and South Humber NHS Foundation Trust	Innovate UK
Research Ethics Committee (REC) reference	**Controlled trials registration number**
24/YH/0182	NCT06722781 OSF pre-registration: https://osf.io/vup9c/

Research team

Trial Management Group
Chief investigator: Jaime Delgadillo[1] (email: jaime.delgadillo@kcl.ac.uk)
Trial Manager: Ben Lorimer[2] (email:rdash.upliftx@nhs.net)
Clinical Lead: Stephen Kellett[2] (email: stephen.kellett@nhs.net)
Cognitive Behavioural Therapist: Dennis Convery[1,2]

Research Team Clinical Studies Officer: Katie Marlow[2]
Local Principal Investigator: Kyla Bowman[2]

Scientific Advisory Board (SAB)
Wolfgang Lutz[3], Christian Webb[4], Juan Martin Gomez Penedo[5,6], Shehzad Ali[7], Taposhri Ganguly[8]

Clinical Advisory Board (CAB)
Glenn Waller[1], Jennie Hague[1], Shannon Sauer-Zavala[9], Carmen Schäuffele[10], Jorge Osma[11], Arnon Rolnick[12], Dennis Convery[1,2]

Affiliations
1. University of Sheffield, UK
2. Rotherham Doncaster and South Humber NHS Foundation Trust, UK
3. University of Trier, Germany
4. Harvard Medical School, USA
5. University of Kassel, Germany
6. University of Osnabrück, Germany
7. Western University, Canada
8. MindLife, UK
9. University of Kentucky, USA
10. Freie Universität Berlin, Germany
11. University of Zaragoza, Spain
12. Rolnick's Institute for Advanced Psychotherapy, Israel

Industry partners

MindLife UK Ltd.

Version control

Protocol version and date	Summary of modifications
(v1) 13.06.24	Developed with reference to transparent reporting guidelines for Factorial Randomized Trials. https://doi.org/10.1001/jama.2023.19793
(v2) 05.07.24	Integrated comments from SAB and CAB members.
(v3) 12.07.24	Revised sections 1-4 based on feedback from SAB & CAB.

(v4) 04.10.24	Add research team members and REC and OSF registration numbers. Updated data controller to RDASH.
(v5) 28.02.25	Amendment to allow recording of therapy sessions for treatment integrity checks.
(v5.1) 06.06.25	Recording of therapy sessions will be via audio recording not videoed.

Synopsis of the study	
Short study title	UpLift-X Trial 1
Design	Pragmatic, multi-site, factorial cluster randomised controlled trial
Setting	NHS Talking Therapies services
Study Participants	Patients with common mental disorders
Aim	To develop methods to personalise digitally-enabled group psychotherapy
Hypotheses	1. Altering the sequence of therapy modules will not be significantly associated with overall group-level treatment outcomes or dropout rates 2. Patients with specific profiles (based on their psychometric characteristics) will benefit more from specific treatment modules and will experience significantly greater symptom reductions in the early phase of treatment if they received their optimal treatment module before other modules 3. There will be no statistically significant differences in overall treatment outcome or dropout rates when comparing digitally-enabled group therapy vs. individual cognitive behavioural therapy (CBT)
Objectives	1. To assess if altering the sequence of therapy modules is generally associated with outcomes or dropout 2. To develop a personalized module selection method 3. To evaluate the clinical and cost-effectiveness of digitally-enabled group therapy vs. individual CBT
Measures	*Primary outcome:* Post-treatment Negative Affectivity (NA) scale *Other measures:* Battery of demographic, diagnostic and psychometric measures
Intervention	12-session, transdiagnostic, digitally-enabled group psychotherapy based on Barlow's unified protocol for the treatment of emotional disorders, including a self-help app with videocall interface
Randomization	Using cluster randomisation, participating services (cluster) will be randomly allocated to deliver different versions of the treatment, where the sequence of modules is altered
Data collection	Consenting patients will initially complete a battery of baseline (pre-treatment) questionnaires, after which they will complete the NA scale on a weekly basis (prior to starting each therapy session). All data will be collected via electronic surveys.
Planned Sample Size	660 participants
Data analysis method	Objective 1: Multilevel model predicting post-treatment NA, controlling for baseline NA, with group (5 alternative module sequences) as the independent variable. Objective 2: Machine learning analysis based on a 60:40 split for training/validation. Objective 3: Comparing post-treatment NA between group therapy vs. individual CBT cases, using doubly robust estimation.
Study Period	24 months

1. Introduction

Background and rationale

Common mental disorders (CMD) related to depression and anxiety affect around 17% of adults each year, leading to adverse personal and societal consequences (Steel et al., 2014). Most countries around the world lack sufficient mental health provision to meet the needs of people with CMD. There is a strong economic case to scale-up the global availability of affordable and evidence-based interventions, given the impact of CMDs on unemployment and lost productivity (Chisholm et al., 2016). Furthermore, in healthcare settings where such treatments are available, there is convincing evidence that only a fraction of patients experience a full and lasting improvement. In England, evidence-based psychological interventions are available at a country-wide level through a publicly funded health care system called NHS Talking Therapy for anxiety and depression (NHS-TT). According to national evaluation reports, over half a million (~664,087) patients complete treatment in NHS-TT services each year, of whom only ~50% of patients recover from their symptoms, despite having access to one or more interventions including other psychological treatments and concurrent pharmacotherapy (NHS England, 2022). In summary, there is a global dearth of access to evidence-based treatments for CMD and currently available treatments are only moderately effective. In other countries the situation is much worse.

Four developments in the field of psychotherapy have great potential to simultaneously address the deficits in access and effectiveness of treatment. [1] First, group-based psychological interventions have been shown to be as effective as traditional individual-based therapies and pharmacotherapies (Barkowski et al., 2016). However, group interventions can be considerably more efficient because they are delivered to many patients simultaneously, such as in the example of stress control groups delivered in England (Dolan et al., 2021) to large groups (average of N=39) of patients. [2] Second, effective *transdiagnostic* interventions have been developed to ensure they can be relevant and helpful for patients with a wide variety of CMDs such as depression, panic disorder, social anxiety disorder, post-traumatic stress disorder, obsessive-compulsive disorder, etc. The "Unified Protocol for transdiagnostic treatment of emotional disorders" (UP) is an example of an effective transdiagnostic intervention proven to be as effective as other well-established treatments for several CMDs (Carlucci et al., 2021; Longley & Gleiser, 2023), which can be easily adapted to a group setting (Ayuso-Bartol et al., 2024). [3] Third, in recent decades computerised and internet-delivered psychological interventions have been proven to be effective for the treatment of CMDs (Karyotaki et al., 2017), reducing the need for patients to travel to a clinic in order to access mental health care. [4] Finally, recent studies have shown that personalized psychological interventions, which tailor the specific treatment components or levels of treatment intensity (e.g., shorter or longer treatments) for each patient, are more effective than standard psychological interventions (Nye et al., 2023). Taken together, these developments indicate that psychological treatments could be made more accessible and cost-effective if they are delivered online, in groups, and integrating transdiagnostic treatment strategies to ensure the content of therapy is tailored to the needs and characteristics of patients with a wide range of clinical problems. UpLift-X is a research programme that aims to integrate all of these four advances into a single and comprehensive digital health platform, which we expect will be a cost-effective solution to maximize access to highly effective treatment for a large population of patients with CMDs. This protocol describes the first work package within the wider UpLift-X research programme.

The overall aim of this study is to test the effectiveness of a modularised, transdiagnostic, and personalized digitally-enabled group intervention for the treatment of CMD. The intervention will be based on the transdiagnostic UP (Barlow et al., 2010), based on principles of cognitive behavioural therapy (CBT). This intervention includes eight distinctive modules with specific content and treatment strategies that have been found to be effective for the treatment of several mental disorders (Longley & Gleiser, 2023). There is evidence that each of these modules target and help to change specific processes that are theoretically posited to be maintaining factors for emotional disorders (Sauer-Zavala et al., 2017). The modularised nature of this intervention could potentially make it possible to offer it in a personalised way, for example by selecting specific modules based on patients' target symptoms (Fisher et al., 2019), their strengths and weaknesses related to emotion regulation (Sauer-Zavala et al., 2019), or their personality traits (Osma et al., 2021). Such an approach would mean that different patients might benefit from selected modules, delivered in different sequences depending on their personal characteristics and needs. Following developments in the area of psychotherapy personalisation, this study will examine the effectiveness of this intervention both in a global (overall effect after

treatment) and granular way (effect after exposure to each module), in order to develop a data-driven methodology to recommend personalised treatment sequences.

Objectives

1. To assess if altering the sequence of UP therapy modules is associated with overall treatment outcomes or drop out.
2. To develop a personalized module selection method.
3. To evaluate the effectiveness of digitally-enabled UP group therapy vs. individual CBT.

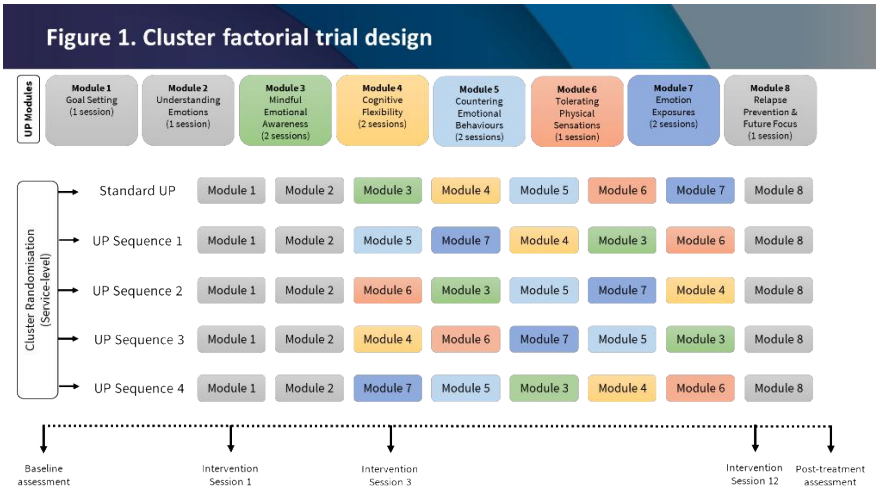
Hypotheses

1. Altering the sequence of therapy modules will not be significantly associated with overall group-level treatment outcomes (e.g., mean-level of symptom severity in each arm of the trial).
2. Altering the sequence of therapy modules will not be significantly associated with overall group-level dropout rates (e.g., % of dropout cases per arm of the trial).
3. Some patients will benefit more from specific treatment modules and will experience significantly greater symptom reductions in the early phase of treatment if they received their optimal treatment module before other modules. In other words, we expect that specific types of patients will benefit from a specific order of modules.
4. There will be no statistically significant differences in overall (group-level, as defined above) treatment outcome or dropout rates when comparing digitally-enabled group therapy vs. individual cognitive behavioural therapy (CBT).
5. Due to the number of patients that can be treated in a group, we expect that digitally-enabled group therapy will be more cost-effective than individual CBT.

2. Methods

Trial design

This will be a pragmatic, multi-site, cluster factorial randomised controlled trial, designed following transparent reporting guidelines for this methodology (Kahan et al., 2023). The unit of randomisation (cluster) will be each participating NHS-TT service. NHS Trusts are wider organisations that manage one or more NHS-TT services, so the unit of randomisation is at this lower level of organisation (e.g., local area services). Therefore, all patients treated in the same local area service will be members of the same cluster. Participating services will be randomly allocated to one of 5 arms (see Figure 1). Each arm will include the same treatment but organised in a different sequence. As shown in the figure, the first two modules will remain fixed at the start of the intervention, as these introduction sessions are a necessary prerequisite to orient all patients to the treatment and to establish a cohesive group process. Similarly, the final module is fixed at the end of the intervention, as it includes an overall review of all coping skills and it focuses on relapse prevention. Modules 3 to 7 cover specific emotion regulation skills and are most amenable to personalisation (e.g., some may be more or less effective for specific patients). Hence, these *prescriptive* modules will be delivered in different sequences across each of the trial arms.



The rationale for altering the sequence of these prescriptive modules is as follows:

- Currently available evidence does not strongly support the personalisation of modules based on clinical theory (Sauer-Zabala et al., 2022). Literature in the field of psychotherapy personalisation indicates that there is little evidence that tailoring treatment based on clinical theory and judgment improves treatment outcomes (Cohen et al., 2021). More recently, data-driven personalised interventions have been shown to outperform standard psychological interventions in clinical trials (see meta-analysis by Nye et al., 2023). These studies use clinical prediction models to prescribe specific treatments to each patient based on their unique characteristics to maximise the probability of improvement. Hence, we will take a data-driven approach to personalisation, which would require a random allocation of patients to alternative sequences, without making theoretical assumptions about clinically "appropriate" or "optimal" sequences.
- Current evidence indicates that reordering the sequence of these modules results in similar overall clinical outcomes and satisfaction ratings from patients (Sauer-Zavala et al., 2019), and hence there is no empirical evidence that altered sequences may have any adverse or disadvantageous effects overall. However, given that these modules target specific mechanisms of change (Sauer-Zavala et al., 2017), some of which may be more relevant to some patients and not others, it is plausible that a personalised sequence could lead to more rapid improvements, and hence more parsimonious and cost-effective treatments.
- One of the challenges for psychotherapy personalisation research is that a substantial number of patients (up to 48%) experience considerable symptomatic improvements during the earliest sessions of treatment (see meta-analysis by Beard & Delgadillo, 2019). According to meta-analytic evidence, patients with *early response* have better overall treatment outcomes across a variety of psychotherapies with different underpinning theories and techniques. Given that the early response phenomenon is not unique to specific forms of therapy, researchers have proposed that it may be related to a general increase of hope and expectations of improvement, rather than specific treatment techniques (Howard et al., 1993). Therefore, making claims about the effects of specific treatment techniques requires controlling for the early response effect, especially when the techniques under study occur in the early phase of treatment (e.g., first 4 sessions). Therefore, altering the sequence of each of the prescriptive modules could enable us to isolate module-specific effects and early response effects.
- In this trial design, it will be possible to investigate if exposure to specific modules leads to contiguous symptomatic improvements (while isolating early response effects) and whether or not the personalised sequencing of modules has additional benefits (e.g., offering more "personally relevant" modules sooner).

Setting and location

This will be a multi-centre study, involving National Health Service (NHS) Talking Therapies services for anxiety and depression (formerly known as IAPT services) in England. NHS-TT services offer evidence-based

psychological interventions for common mental health problems following clinical guidelines (National Institute for Health and Care Excellence, 2011). Following a stepped care model, these services offer low intensity interventions to most patients as an initial step, which involves up to 8 sessions of guided self-help with a psychological wellbeing practitioner. Patients who remain symptomatic after this initial step have the option to access high intensity psychotherapies delivered by qualified therapists as a next step in their treatment pathway. Furthermore, some patients with specific conditions (e.g., post-traumatic stress disorder, social anxiety disorder, body dysmorphic disorder) or those with severe impairment can be referred directly to high intensity therapies, in accordance with national guidelines (National Collaborating Centre for Mental Health, 2024). After obtaining approval from the Health Research Authority (HRA), we will recruit NHS-TT services through national and regional NHS-TT professional networks and mailing lists that are accessible to our team (as members of these networks).

Participants

The digitally-enabled group intervention will be available to patients with common mental disorders (and comorbidities) that are eligible to receive high intensity CBT in NHS-TT services, following national guidelines (National Collaborating Centre for Mental Health, 2024). Eligibility for high intensity CBT is established through semi-structured interviews conducted in routine care by qualified mental health practitioners, and following guidelines cited above. These are not structured diagnostic interviews, but they do include the use of validated psychometric measures to establish the patient's primary presenting problem (these measures are described below). Therefore, eligibility will be determined by qualified mental health practitioners as part of routine practice assessment interviews. Detailed inclusion and exclusion criteria are listed below.

Inclusion criteria	Exclusion criteria
Age ≥17	Age ≤ 16
Literate and fluent in the English language, which would enable group interactions and learning via online materials	Not able to speak or read English fluently, which would require more individualised support with interpreters
With access to an internet-connected device (e.g., computer, phone) and confidential space	No access to internet-connected device or confidential space to engage with digital intervention
Referred to NHS-TT services	Not involved with NHS-TT services
Assessed by a qualified psychological professional and deemed to be eligible for high intensity CBT (this includes low risk of harm to self or others)	Assessed by a qualified psychological professional and not deemed to be eligible for high intensity CBT on the basis of mental health condition and/or risk assessment (e.g., acute risk of suicide at the time of assessment)
Currently on waiting list for high intensity CBT and not accessing any other psychological interventions. This includes patients accessing pharmacotherapy but who are additionally seeking psychological treatment.	Already accessing a high intensity psychotherapy or other forms of psychological treatment (e.g., private therapy, counselling)
Presenting symptoms of one or more internalizing disorders (major depressive disorder, generalised anxiety disorder, panic disorder, agoraphobia, specific phobias, somatoform disorder, obsessive-compulsive and related disorders, body dysmorphic disorder, post-traumatic stress and related disorders)	Presenting with symptoms of eating disorders, substance use disorders, psychotic disorders, personality disorders, bipolar disorders identified at the time of initial assessment.

Interventions

Primary intervention

The UP is a transdiagnostic treatment that was designed to be relevant to a wide range of emotional disorders, given the common co-occurrence of depression, anxiety, somatoform and dissociative symptoms (Barlow et al 2010). It is based on CBT and emotion regulation theories, and the aim is to support patients to apply a series of emotion regulation skills that target the core common vulnerabilities that underlie a range of mental health problems: neuroticism, low perceived control, and overestimation of threat. The protocol follows the principle of parsimony, by including the minimum number of empirically validated techniques that would make it accessible, uncomplicated, and effective for the maximum number of patients with various combinations of symptoms and problems. Contents are organised across 8 modules, within 10-18 sessions (1hr) which are summarised in Table 1 below. It can be delivered individually, in groups and via internet. Although the original protocol proposes a specific sequence, case studies have found that this intervention is equally acceptable and effective when the modules are delivered in different sequences (Sauer-Zavala et al 2019). Given that it has been shown to be helpful across a wide range of clinical diagnoses and populations, the UP is now widely recognised as an evidence-based psychotherapy and has been implemented in at least 11 countries (Cassiello-Robbins et al., 2020).

Over 70 studies have tested the effectiveness of the UP using experimental or observational designs (Cassiello-Robbins et al., 2020). Large effect sizes for change in anxiety (g = .80) and depression (g = 0.74) are found in observational studies after treatment, and at 3-6 months follow-up (g = 1.13) indicating that people continue to improve over time (Carlucci et al., 2021). According to meta-analytic evidence from randomised controlled trials, the UP was found to be more effective than passive controls such as waitlist (g = .59) and active controls (g = .38) including other psychological interventions (Carlucci et al., 2021). Current evidence indicates that the UP is equally as effective as disorder-specific CBT treatment protocols for depression and slightly more effective for anxiety (Longley & Gleiser, 2023). According to moderator analyses, UP interventions delivered by more experienced therapists (Carlucci et al., 2021) and with a larger number of sessions (Longley & Gleiser, 2023) are more effective than shorter versions. The UP is equally effective delivered individually, in groups or via internet (Carlucci et al., 2021).

The "UPLift" digitally-enabled group intervention is based on the UP, and it will include a combination of internet-based resources (self-monitoring questionnaires, psychoeducational videos, skills practices linked to each module) and weekly group therapy sessions delivered by a qualified cognitive behavioural therapist. Each group will have an average of 10 patients who will participate via a secure and confidential video conference platform that is integrated within the UPLift-X digital platform. The full treatment will last 12 sessions, delivered once per week, each lasting up to 90 minutes (overall equivalent of 18 hours of therapist-facilitated therapy). Patients will complete a series of questionnaires (described below) using the UPLift-X website prior to each group therapy session, which will help them and their therapists to monitor changes in their symptoms over time. Once they log in to their scheduled therapy session, they will be able to interact with the facilitator and other patients, as part of a structured process that has a specific agenda and content for each session (based on each UP module). In general terms, each session covers specific coping skills with examples of how to apply them in daily life and it ends with an agreed therapy "skills practice" that patients in between sessions with the aid of the content available in the UPLift-X website. Group therapy sessions will be audio recorded to check treatment integrity and support the ongoing monthly clinical supervision of the therapists. The sessions will be rated on a measure that can rate each session for treatment differentiation, adherence to the Unified Protocol and competency in facilitating the group. At least one session per participating therapist will be rated for feedback during delivery of groups (more may be required if, for example, an initial review by the clinical supervisor leads to a recommendation for further observation and feedback for specific therapists) Ratings of sessions will be included in the method of the main trial to index that the UP is being delivered as intended.

Table 1. Summary of UP modules and content

Module	***Content of the Unified Protocol for Transdiagnostic Treatment of Emotional Disorders***
1	***Motivational Reinforcement***
	The module aims at increase the patient's motivation for change by promoting a sense of self-efficacy and a belief in his own ability to succeed. This module could be used throughout the course of treatment to maintain high motivation for behavioural changes.
2	***Psychoeducation and Treatment Rationale***
	The module aims at explaining to the patients the functional nature of emotions and its multi-components (behavioural, physiological and cognitive).
3	***Emotional Awareness Training***
	The module aims at foster a full awareness of one's emotional experience in the present moment by monitoring the interaction between feelings, thoughts, body sensations and behaviours. This is missing in alexithymic people who experienced difficulty in identifying and describing emotions.
4	***Cognitive Reappraisal***
	The module focuses on patients' cognitions (thoughts, beliefs, automatic appraisals) about emotional experiences, and how cognitions influence their physiological sensations and behaviours, teaching patients to create alternative ways of thinking when they experience intense emotions (cognitive flexibility).
5	***Emotion Driven Behaviours and Emotional Avoidance***
	The module focuses on the emotional avoidance patterns and on strategies adopted in emotional situations, that patients experienced as uncontrollable and threatening. Patients learn to identify the maladaptive Emotion-Driven Behaviours and work to change concurrent patterns of emotional responding.
6	***Awareness and Tolerance of Physical Sensation***

	In the following module the patient is engaged in interoceptive exercises designed to evoke physical sensations similar to those typically associated with anxiety and distress. This module aims at fostering an awareness of the somatic component of the emotion such as to increase tolerance.
7	***Interoceptive and Situational Exposure***
	The module focuses on the emotional experiences that arise in situations and take the form of in-vivo, imaginal and in-session exposure. In this module patients are graded exposed to internal and external emotional triggers to increase their tolerance of emotions and to promote new contextual learning.
8	***Relapse Prevention***
	During the module patients are reminded that a return of symptoms is not a sign of relapse and that, with the skills learned during the therapy, they can cope with emotional oscillations and respond adaptively.

Secondary interventions of interest

Patients who meet the eligibility criteria listed above usually access high intensity CBT. We will collect anonymised clinical health records for all patients accessing individually-delivered CBT in the participating services during the time period of the clinical trial. These data will include demographics (e.g., age, gender, ethnicity, etc.), clinical care (e.g., treatments received, number of sessions attended) and clinical outcomes (routinely collected measures described below). This will enable us to undertake comparisons of clinical and cost-effectiveness using a case-control matching strategy that is described in the data analysis section.

Measures

Routine outcome monitoring in NHS-TT services

NHS-TT services implement a routine outcome monitoring system in which patients complete a series of validated questionnaires on a regular basis, prior to the start of each therapy session. Currently, many services use electronic surveys that are sent to patients via text or email within 24 hours of their next scheduled appointment, making data collection efficient and reliable. All patients routinely complete three questionnaires described below.

The PHQ-9 is a measure of depression symptoms, where each of 9 questions is rated using a Likert scale from 0 to 3, yielding an overall severity score between 0 and 27 (Kroenke, Spitzer, & Williams, 2001). A cut-off of ≥10 has been recommended as providing the best trade-off between sensitivity (88%) and specificity (88%) for a diagnosis of major depressive disorder (Kroenke et al., 2001). A difference of ≥6 points between measurements has been recommended to assess statistically reliable change (National Collaborating Centre for Mental Health, 2024). The GAD-7 is a 7-item questionnaire used to identify anxiety disorders; each item is also rated between 0 and 3, with a total severity score between 0 and 21 (Spitzer et al., 2006). A cut-off score ≥8 is recommended to identify clinically important anxiety symptoms, with adequate sensitivity (77%) and specificity (82%) (Kroenke et al., 2007). A change of ≥4 points has been recommended to assess reliable change (National Collaborating Centre for Mental Health, 2024). The Work and Social Adjustment Scale (WSAS) is a measure of functional impairment due to poor mental health, which rates overall functioning across 5 domains including: work, home management, social life, private leisure activities, and family relationships (Mundt et al., 2002). Responses are captured on a nine-point scale ranging from “not at all” (0) to “very severely impaired” (8), yielding a total severity score between 0 and 40. Copies of all three questionnaires can be found in Appendix 1.

In addition to these three routinely collected measures, patients with specific mental health problems are also asked to complete a relevant “anxiety disorder specific measure” (ADSM). These are also psychometrically validated questionnaires that are recommended in clinical guidelines to monitor changes pertaining to particular conditions such as post-traumatic stress disorder, obsessive compulsive disorder, social anxiety disorder, panic disorder, health anxiety, and others. A list of ADSM questionnaires can be found in Appendix 2.

Primary outcome

The present study will include patients with a wide range of diagnoses and symptoms. As such, it is necessary to select a primary outcome that is relevant to a heterogeneous clinical sample with symptoms of common mental disorders, social and interpersonal problems. Based on this rationale, the primary outcome of interest will be *Negative Affectivity* (NA), a transdiagnostic construct that includes commonly occurring symptoms of low mood, negative thoughts, fear/anxiety, social and interpersonal difficulties. NA can be measured by pooling all items from the PHQ-9, GAD-7 and WSAS into a single scale which attributes specific weights to each item according to their strength of association with an underlying dimension of general psychological distress. To derive this NA

scale, we will apply the item-weighting methodology proposed by (Böhnke et al., 2014). This involves multiplying the raw item scores across all three questionnaires with the non-standardized factor loadings presented in the last column of Table 2 of the publication by Böhnke et al. (2014). This yields a continuous NA severity score ranging between 0 and 69.58, with a reliable change index of >12.44. This NA measure will be completed by participants at a baseline (pre-treatment) assessment, prior to every group therapy session, and at 6-months follow-up.

Secondary outcomes

Two secondary outcomes of interest will be completed by participants at a baseline (pre-treatment) assessment, at the final scheduled therapy session (week 12 of treatment), and at 6-months follow-up. Further information about these measures is available in Appendix 3.

The EuroQol Group 5-Dimension self-report questionnaire (EQ-5D) measures health-related quality of life across five domains: mobility, self-care, daily activities, pain/discomfort, and depression/anxiety (TEQ Group, 1990). This measures yields a continuous score with a range between −0.446 to 1.00, where higher scores indicate better quality of life.

The Short-form Warwick and Edinburgh Mental Wellbeing Scale (SWEMWBS) is a 7-item self-report scale that measures psychological well-being, with a continuous score ranging between 7 and 35, where a higher score indicates better well-being (Stewart-Brown et al., 2009).

Profiling variables

In order to meet objective 2 of the study, participants will be asked to complete a baseline electronic survey which will include a battery of questionnaires. The purpose of these baseline measures is to enable us to identify the *profiles* (e.g., combination of features) that are associated with symptomatic improvements that occur after exposure to each of the UP modules. A full list and corresponding references for these profiling variables is available in Appendix 3. The baseline questionnaires will cover information organised in six domains:

1. *Sociodemographics* (e.g., age, gender, ethnicity, employment status, socioeconomic status, etc.)
2. *Mental health history* (e.g., childhood adversity, chronicity, previous episodes, etc.)
3. *Current symptoms & functioning* (e.g., problem descriptor, routine outcome measures described above)
4. *Relational & characterological* (social support, personality)
5. *Expectancy* (expectations of improvement)
6. *UP module-matched constructs* (e.g., experiential avoidance, mindfulness, beliefs about emotions, dysfunctional attitudes, etc.)

Additional measures

Two additional measures will be completed by participants to support the analysis. To control for group alliance levels, participants will complete the 12-item Group Climate Questionnaire-Short (GCQ-S; MacKenzie, 1983) as a measure of group cohesion after session 3 (week 3 of treatment). To capture service utilisation for the health economic analysis, participants will complete the Modified Adult Service Use Schedule (AD-SUS) at baseline and 6-month follow-up Further information about these measures is available in Appendix 3.

Sample size

The sample size for this study has been calculated following guidelines for multilevel cluster factorial trials (Nahum-Shani et al., 2018) and with reference to the primary data analysis for objective 1 and hypothesis 1; the corresponding statistical analysis is explained in further detail in section 4 below. The calculation has been made based on the following assumptions and parameters:

- Powered to detect a small effect size difference of d = .20 for a non-inferiority comparison
- Power = 80%, alpha = 0.05
- Intracluster correlation coefficient (ICC) = .036 for group-level clustering, based on the ICC for group interventions in NHS-TT services reported by Delgadillo et al. (2016)
- Design effect = 1.324
- Average cluster size = 10 group participants
- Attrition = 30%
- Multiple comparisons = 4

The sample size calculation based on these parameters equals 659, so we will aim to recruit a minimum of 660 participants who complete a baseline assessment.

Interim analyses and stopping guidelines

No interim analyses or trial progression/stopping guidelines will be applied, given that the UP intervention has a well-established evidence base demonstrating its efficacy, acceptability and safety (Carlucci et al., 2021; Cassiello-Robbins et al., 2020; Longley & Gleiser, 2023). In addition, the feasibility and effectiveness of group CBT interventions in NHS-TT services is also supported by practice-based evidence (Dolan et al., 2021).

Randomisation

Participating therapists (clustered by site) will be randomly allocated to one of 5 trial arms, as shown in Figure 1. The random allocation sequence will be generated by a computerised random integer generator. Concealment is not possible in this trial design, since each therapist must be aware of their allocation in order to deliver the treatment in the specified sequence. A research assistant who is not involved in recruitment, treatment delivery or data analysis, will generate the randomisation sequence and will communicate the allocation to the researcher in charge of communicating this (via email) to participating therapists.

3. Procedures

Identification and recruitment

Participating NHS-TT services will identify potentially eligible participants through weekly reviews of the waiting list for high intensity CBT. The waiting list will be reviewed by administrative and/or research staff employed by each NHS Trust. This will necessarily require these NHS staff members and members of the research team to access personally identifiable information (name, contact details) in order to provide information about the study to potential participants and to communicate with them during the study. The process will follow this sequence:

1. Access the list of patients who have been assessed as eligible to access high intensity CBT and who are currently waiting to start their therapy.
2. Sort the list by referral date.
3. Check the patient's record to identify their recorded primary presenting problem and language. Make a note of the cases that match up with the study exclusion criteria (e.g., primary problem is in the list of excluded diagnoses, the patient requires an interpreter).
4. Promote the study with all eligible patients on the high intensity CBT waiting list, in consecutive order based on referral dates, following a three-step strategy outlined below.
5. The first step is to send an invitation text message which will contain a weblink to a promotional video that briefly explains the study and which prompts patients to access an online survey that will include the participant information sheet and consent form.
6. If the person has not provided consent within 1 week of the initial text, send a second text message offering an opportunity to have a telephone call to discuss the study.
7. If the person consents to the phone call, the third step is to call them in order to explain the study and how the person can consent if they would like to.
8. Keep a record in a recruitment spreadsheet of the dates when each person on the waiting list has been contacted, how (text, email, phone call) and how many times, following the steps listed above.

In order to manage the above process, administrative/research staff in each trial site will maintain a site-specific recruitment spreadsheet to record all steps in the identification and recruitment process. They will communicate regularly with the Trial Manager to confirm which patients have been contacted each week, and the Trial Manager will access the secure electronic survey used to obtain informed consent in order to verify whether or not each contacted patient has consented. This coordination will ensure that the research team will be able to monitor and manage the recruitment process on a weekly basis. The Trial Manager may delegate some of these tasks to research assistants in the research team. Template text messages for contacts listed in points 2-3 above are available in Appendix 4.

Furthermore, clinicians who undertake routine initial assessments (e.g., with newly referred patients seeking treatment) or post-treatment assessments (e.g., with patients who complete a low intensity intervention and who require additional treatment) will make potentially eligible patients aware of the study by providing advanced notification of the available treatment they will be contacted about while on the waiting list. A script that therapists will use for this purpose is available in Appendix 5.

Contact strategy

After completing the electronic consent form, participants will be contacted via email by the research team at several timepoints:

1. Within 1 week of completing the consent form, to guide them on how to access the UPLift website, where they will view introductory videos that will prepare them to start their group therapy.
2. One week prior to their scheduled initial therapy session, to remind them and motivate them to attend.
3. After session 3 (week 3), to complete a self-report measure of group cohesion.
4. One week prior to the last scheduled therapy session (week 12), to remind them that they will be asked to complete a follow-up survey that is very important for the study. This reminder is particularly important to ensure that patients who have stopped attending therapy sessions (e.g., dropped out) still complete the follow-up outcome measure, as this also contains secondary measures of quality of life and well-being.
5. At the 12th week of their scheduled therapy session, to collect the primary outcome measurement.
6. One week before the 6-month follow-up measurement point, with a reminder to complete the questionnaire they will receive the following week.
7. At the 6-month follow-up time point.

The research team will regularly monitor responses to the electronic surveys at the relevant time-points, and will contact participants who have not completed the survey with up to 3 text/email reminders within a maximum period of 2 weeks after each measurement point.

Data collection

All data will be collected via industry-standard, secure electronic surveys. The measurement schedule will be as follows:

1. The baseline assessment will include all measures (primary, secondary, profiling variables, service utilisation) described above. This assessment will be included in the same electronic survey that includes the participant information sheet and consent form, so that baseline measures can be collected at the time of recruitment.
2. Patients will be prompted to complete the primary outcome measure at least 24 hours before each of the 12 scheduled therapy sessions.
3. After session 3 (treatment week 3) patients will be prompted to complete the group cohesion measure.
4. The 12th sequential pre-session survey will include the primary and secondary outcome measures.
5. The six-month follow-up survey will include the primary and secondary outcome and service utilisation measures.

All patients will be registered into the electronic data collection survey using a pseudonym (e.g., a series of letters and numerical characters that do not reveal the person's name or other identifiable details) which we refer to as a patient ID. This patient ID will enable us to link the trial participant's responses to the electronic surveys with wider (anonymized) healthcare records that will be collected from participating NHS-TT services. This will enable us to identify and differentiate group therapy cases vs. individual CBT cases, in order to compare their treatment outcomes and costs.

Trial endpoint criteria

This clinical trial will end when the last consenting patient has provided a 6-month follow-up measure (or has met the 3 follow-up contact criteria for those who fail to complete the electronic survey).

4. Statistical analysis plan

The statistical analysis plan will be carried out in three stages linked to each of the objectives of the study.

Objective 1 – compare outcomes of different UP module sequences

Variables and data pre-processing

- *Dependent variable*: The dependent variable will be a continuous Negative Affectivity (NA) score measured at the 12th (last) scheduled session of the group intervention. Coding and calculation rules for the NA score are described above in the measures section.
- *Independent variables*: The primary (hypothesis-testing) variable of interest will be a categorical "arm" variable where the reference category (Standard UP sequence) is compared to each of the 4 alternative sequences. The baseline (pre-treatment) NA score will be controlled as a potential confounder.
- *Imputation*: Cases with missing post-treatment (12th session) data-points in the dependent variable will be imputed using the MissForest R package (Stekhoven & Bühlmann, 2012) using all baseline variables as predictors. Imputations will be carried out separately for each trial arm.

Analysis

This will be the primary analysis, based on which the sample size calculation is based. The analysis will apply a linear multilevel (mixed effects) model, where patients will be nested within groups, entering a random effect for the group-level. The dependent variable will be the post-treatment NA score (a continuous variable), entering *arm* as an independent variable (a categorical variable), controlling for baseline NA (continuous variable).

Following conventional model-building guidelines (Raudenbush, 1993), continuous predictors will be grand mean-centred and modelling will be performed in sequential steps, starting with single-level models and eventually developing multi-level and covariate-adjusted models that optimise goodness-of-fit. Model fit indices (AIC, BIC, -2LL) will be examined after each modelling step. We will retain and interpret the best-fitting and most parsimonious model achieved through this stepwise process, using the -2 loglikelihood ratio test to compare models. In the first modelling step, we will compare whether entering the raw NA outcome measure or entering a log-transformed measure (to deal with skewness) offers better goodness-of-fit, prior to developing covariate adjusted models. The intracluster correlation coefficient (ICC) will be reported as an index of variability in treatment outcomes attributable to the group-level. Between-group effect sizes (Cohen's *d*) will be calculated and reported for each contrast (e.g., standard UP sequence vs. other sequences). As a sensitivity analysis, the fully adjusted 2-level model will be fitted again after winsorizing data-points for extreme outliers (outside of 2 standard deviations from the mean) in the dependent variable.

Secondary analysis

The above analysis plan will be repeated using a logistic multilevel model with dropout as the dependent variable. The dropout variable will be derived from group therapy attendance records, where cases that attended more than half (≥6) sessions will be grouped into the reference category (code = 0) and cases that attended less than half (≤5) will be classed as cases that dropped out (code = 1).

Objective 2 – develop a personalised module selection method

Variables and data pre-processing

- *Dependent variable*: Reliable improvement (RI) refers to a reduction in symptom severity in between two time-points, which is greater in magnitude than the reliable change index for the psychometric measure (Jacobson & Truax, 1991). Following this definition, we will class cases using a binary variable denoting whether they did (code = 1) or did not (code = 0) have reliable improvement after exposure to a *prescriptive module* (colour coded in Figure 1). Since there are 5 prescriptive modules, there will be 5 separate dependent variables of interest. To be coded as "1" in this variable, a patient's NA score measured after completion of the target module should be reliably smaller than the score measured immediately prior to the first session pertaining to that module (e.g., a pre-post difference greater than the reliable change index for the NA scale, which is 12.44). This coding will enable us to identify cases that experience reliable symptomatic improvements after exposure to each of the target modules, regardless of where the module is located (e.g., towards the start, middle or end of the intervention) in the group intervention they have been allocated to. This methodology will enable us to control for

sequencing and early response effects, to identify the profile of patients who have a favourable response to each specific module.

- *Independent variables*: All profiling variables collected at baseline (pre-treatment) assessment.
- *Sample partitioning*: The full trial sample will be randomly partitioned into training and validation subsets using a 60:40 split, with a balanced number from each of the 5 trial arms in each training-partition split. This will ensure that 396 cases will be randomly selected into the training partition, exceeding the minimum sample size of 300 necessary to train a reliable clinical prediction model using supervised machine learning (Luedtke et al., 2019).
- *Imputation*: Cases with missing post-treatment (last attended session) data-points in the dependent variable will be imputed using longitudinal multilevel modelling (e.g., growth curve analysis using available session-by-session measures). Imputations will be carried out separately for each trial arm, with separate imputations in the training/validation partitions. Baseline measures will not be missing, since participants will complete these measures immediately after completing the online consent form, and completing the measures is a prerequisite to participation.

Model development and optimisation (in the training partition)

Three types of supervised machine learning models will be developed in the training sample: a regularized regression (LASSO), a decision tree ensemble (random forest), and a Bayesian classifier (tree augmented naïve Bayes). The two latter models do not perform variable selection, so additional versions of these models will be produced using the variables selected by LASSO regularization. Hyperparameter tuning will be automated using grid search procedures. The predictive accuracy of these three alternative models will be compared in the training sample using 5-fold cross-validation. The pooled area-under-the-curve (AUC) index across all iterations of the k-fold cross-validation process will be used to determine which type of machine learning model achieves the best trade-off between sensitivity and specificity. Additional performance indices to compare models will be the balanced accuracy, positive and negative predictive values. The best-performing model will be chosen among these alternative methods and will be applied in the full training sample to train a final machine learning model to predict the outcome of interest (reliable improvement after exposure to a target UP module), making best use of the full sample size in the training sample, and minimising complexity by selecting one preferred machine learning approach (rather than multiple modelling approaches).

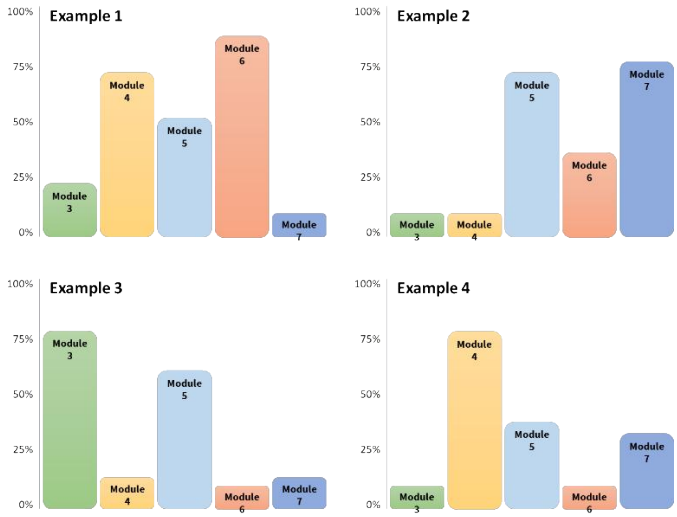
Therefore, we will produce a total of 5 machine learning prediction models, one for each prescriptive UP module. Each trained model will therefore process inputs (patients' baseline features) and will output [a] a predicted classification and [b] a corresponding predicted probability pertaining to the classification. The logic of these algorithms is to answer the following question: *Given these patient characteristics, how likely is it that this specific patient will experience reliable improvement after accessing module X?*

Cross-validation strategy (in the validation sample)

The statistical accuracy and clinical utility of the machine learning models will be evaluated. First, the trained algorithms (one for each of the 5 prescriptive modules) will be applied to make predictions for patients whose data has been randomly included in a statistically independent validation sample (including 40% of cases). The statistical accuracy of the model predictions will be assessed against the observed outcome (e.g., actual events of reliable improvement occurring after exposure to the target module). This will be formally evaluated using the AUC, balanced accuracy, and calibration plots. The prediction shrinkage will also be reported, by subtracting the magnitude of the AUC indices observed in the training and validation samples.

Next, the potential clinical utility of the machine learning models for the purpose of personalised module selection will also be evaluated. Each patient in the validation sample will have a total of 5 predictions – one for each prescriptive module – expressed on a probability scale (0% to 100% probability of reliable improvement). These predictions will be taken as a proxy indicator of the clinical relevance of each module to that individual patient. Hence, each patient would have a data-driven ranking of clinical relevance for all 5 modules. The module with the highest predicted probability would be taken as the "optimal" module for that patient, and other modules would be ranked thereafter in an order of decreasing relevance. Figure 2 shows examples of personalised sequence prescriptions, illustrating how the trained machine learning algorithm would output personalised recommendations for four different patients.

Figure 2. Examples of personalised module prescriptions for four different patients



To test whether or not this data-driven ranking/sequencing of modules could potentially guide treatment selection, we will test it in two ways described below. We will apply the same multilevel (mixed effects) modelling strategy described for Objective 1 above to compare outcomes between patients who were randomly assigned to their "optimal module" as the first one in the sequence of prescriptive modules, versus all other cases. The dependent variable of interest in this analysis will be the NA score measured after exposure to the first prescriptive module (e.g., NA score measured at session 5), and the analysis will control for baseline NA. For example, for patients whose "optimal" module is module 4, we would expect them to have better outcomes early in treatment if they were randomly assigned to sequence 3 (see Figure 1), if indeed the content of module 4 is more advantageous than any other randomly assigned module, and if such content adds value over and above early response effects.

In a secondary analysis following a similar logic to the above step, we will use the same multilevel modelling strategy to predict NA scores observed by session 8, controlling for baseline NA. In this analysis, the hypothesis testing variable will be an ordinal variable that denotes the similarity between the data-driven sequence for the top three most advantageous modules for each patient and the actual (randomly assigned) sequence that they actually received. In this "prescription similarity" variable, 0 = not at all similar, 1 = only one relevant module was accessed during this period, 2 = two relevant modules were accessed during this period, 3 = the three most relevant modules were accessed during this period but in the reverse order, 4 = the three most relevant modules were accessed during this period almost in the prescribed order, 5 = the three most relevant modules were accessed during this period in the prescribed order. If a personalised sequence of modules is indeed more effective than a randomised sequence, we would expect that this ordinal variable would be statistically significant and inversely associated with NA scores measured at the 8th session (e.g., the greater the similarity between the prescribed and the received sequence of modules, the lower the NA score will be after exposure to the first 3 prescriptive modules).

Secondary analysis

The above machine learning analysis and cross-validation strategy will be repeated to train a binary classifier with dropout as the dependent variable. The dropout variable will be derived from group therapy attendance records, where cases that attended more than half (≥6) sessions will be grouped into the reference category (code = 0) and cases that attended less than half (≤5) will be classed as cases that dropped out (code = 1).

Objective 3 – evaluate the effectiveness of digitally-enabled UP group therapy vs. individual CBT

Variables and data pre-processing

- *Sample selection*: This analysis will include [a] a subgroup of clinical trial participants who were randomly assigned to the arm where the UP protocol was delivered in the standard sequence, and [b] a matched sample of patients who accessed individual high intensity CBT in the participating NHS-TT services. The data for the individual CBT cases will be sampled from electronic health records. CBT cases that meet the inclusion criteria listed in page 7 will be selected for this analysis, only including cases where the primary presenting problem is identifiable in clinical records and where baseline severity scores (PHQ-9, GAD-7, WSAS) are available at their first attended CBT session.

- *Dependent variable*: The dependent variable will be a continuous Negative Affectivity (NA) score measured at the 12th (last) scheduled session of the group intervention.
- *Independent variables*: The primary (hypothesis-testing) variable of interest will be a categorical "modality" variable where 0 = individual CBT cases vs. 1 = cases accessing group therapy following the standard UP protocol sequence. The analysis will control for the following potential confounders, which are collected as part of routine clinical health records: baseline severity of routinely collected outcome measures (PHQ-9, GAD-7, WSAS), age, ethnicity, employment status, index of multiple deprivation, disability.
- *Imputation*: Cases with missing post-treatment (last attended session) data-points in the dependent variable will be imputed using longitudinal multilevel modelling (e.g., growth curve analysis using available session-by-session measures). Imputations will be carried out separately for each group (individual CBT; standard UP group therapy).

Analysis

Post-treatment clinical outcomes (NA scores) will be compared between cases accessing the standard group UP protocol versus individual CBT cases using *doubly robust estimation*, which combines an adjustment for non-random allocation to treatments, and a covariate-adjusted regression model to test the causal effect of exposure to an intervention. When used individually to estimate a causal effect, both outcome regression and propensity score methods are unbiased only if the statistical model is correctly specified. The doubly robust estimator combines both approaches, such that only 1 of the 2 models need be correctly specified to obtain an unbiased effect estimator (Funk et al., 2011). In this analysis, we will apply *inverse probability treatment weighting* (IPTW) based on the propensity score that denotes the probability of assignment to group therapy, in order to control for confounding by indication and/or patients' preferences. IPTW will be calculated using the methodology proposed by Emsley et al. (2008).

Secondary analyses

The above analysis will be repeated to compare outcomes using disorder-specific measures (PHQ-9, GAD-7, ADSM) for cases with the relevant diagnoses. Between-group effect sizes (Cohen's *d*) will be calculated and reported for each comparison. As a sensitivity analysis, the model will be fitted again after winsorizing data-points for extreme outliers in the dependent variables. A logistic regression model will also be used to compare dropout rates as a secondary outcome of interest.

If the analysis for Objective 2 does indeed indicate that a personalised sequence of UP modules is more advantageous than a randomly allocated sequence, we will compare treatment outcomes (NA scores) for individual CBT cases vs. group therapy cases that have a high "prescription similarity score" ≥ 3 (as described above). This would enable us to examine whether a personalised group intervention may potentially lead to faster improvements compared to individual CBT cases. This analysis will compare the slope of change in NA scores over time, using longitudinal multilevel modelling, where session NA scores (level 1) are nested within cases (level 2). This model will enter session-by-session NA scores as the dependent (time-series) variable, and it will model *Time* (sessions) using the best-fitting growth trend (linear, quadratic, cubic, or logarithmic). The hypothesis testing variable in this model with be the interaction between *Time* and *Modality*, controlling for baseline NA (grand mean-centred) and the main effect of Modality (group vs. individual CBT). This model will apply IPTW, as described above. A sensitivity analysis will also examine the results using a 3-level model (NA scores nested within cases, nested within therapists).

Health economic analysis

An economic analysis will be conducted from the NHS and Personal Social Services perspective. A cost-utility analysis will use quality-adjusted life years (QALYs derived from the EQ-5D questionnaire, and tariff based on the UK public value set) as the measure of quality of life, using post-treatment (last-attended session) data on the NA scale in the primary analysis, and 3-month follow-up data in a secondary analysis. Health and social services resource use will be valued using NHS reference costs and the personal and social services resource use database (PSSRU). We will estimate the incremental cost effectiveness ratio for group therapy versus individual CBT, using bootstrapping to estimate confidence intervals (Briggs et al., 1997). Decision uncertainty will be presented on a cost-effectiveness acceptability curve (Fenwick et al., 2005). Additional sensitivity analyses will be conducted for resource use and unit costs. A sensitivity analysis will control for baseline costs. Scenario analyses will explore alternative costing perspectives; that is, NHS and NHS/PSS perspectives. Results of the cost-effectiveness analysis will be reported in line with the Consolidated Health Economic Evaluation Reporting Standards 2022 Statement (Husereau et al., 2022).

5. Transparent reporting of results

Intention-to-treat principles
This clinical trial will collect and analyse data using intention-to-treat principles. As such, data from cases that dropped out of the study and/or treatment procedures will be retained and used in the analyses described above, to overcome the biases known to be present in analyses that only include treatment completers. We will only remove data for participants who decide to withdraw from the study and who request for their data to be deleted.

Participant flow
A CONSORT diagram with adjustments for factorial trials will be used to report the flow of participants including the numbers approached, consented, randomised to each arm, treated, imputed and analysed.

Losses and exclusions
We will report attrition at all stages of the study, as well as the number of patients who withdraw from the study for any reason.

Baseline data
Sample characteristics at the time of consent and baseline assessments will be reported using descriptive statistics for each trial arm, as well as for the full and selected subsample of individual CBT cases (from electronic health records).

Adverse events
Any adverse events recorded in electronic health records will be reported.

6. Ethical considerations

Identification of potential participants

It will be necessary to access personally identifiable information to identify and to contact potentially eligible patients on waiting list for high intensity CBT in the participating NHS-TT services. This will be carried out by members of the direct care team (NHS-TT) and/or research staff employed by the relevant NHS Trust who have obtained a research passport to enable this. This means that research members outside of the direct care team will not have access to personally identifiable information for any patients who have not yet provided informed consent.

Informed consent

In order to obtain informed consent in line with good practice guidelines, we will take the following steps:

- Potential participants will have 1 week to consider their participation, so that they do not feel unduly pressured to consent at the time when they initially receive the study promotional information and electronic information consent sheet.
- Potential participants will be advised of their right to withdraw from the study at any stage and the right to request their data to be deleted from the study dataset.

We will also be collecting fully pseudonymised patient-level data described above. We consider that our proposed method for collecting and storing fully pseudonymised patient data is congruent with the NHS information governance policy and good practice guidelines. We will also obtain informed consent from patients and they will have immediate access to information on how to withdraw their data from the study if they wish to do so.

Patient preference

The recruitment process is designed in such a way that patients' preferences will be respected and followed. Some patients may not find group therapy acceptable for various reasons (e.g., severe social anxiety, preference to maintain their access to a mental health service anonymous, etc.). For these reasons, we will only recruit patients who find group therapy acceptable and who provide informed consent. Potential self-selection bias introduced by this process will be adjusted for statistically, as described above.

Clinical risk management

Participants in this clinical trial will receive an intervention that has a well-established evidence base for its safety, efficacy and acceptability. The treatment principles that guide the UP protocol are similar to those that underpin usual high intensity CBT interventions delivered in routine care. Therefore, we have no reason to believe that the intervention may be any less acceptable, safe and effective as usual CBT in the participating services. As explained above, we will only recruit participants who find group therapy acceptable, and who have been assessed as having low risk of harm (eligibility criteria for treatment in NHS-TT services), so we do not expect that this would carry any higher risks than usual care.

The intervention will be delivered by qualified and experienced cognitive behavioural therapists employed by NHS-TT services. Participating therapists will receive additional training (2 days) by experts in the field and will receive specialist clinical supervision to ensure that the group therapy is delivered in a safe and effective way, with due attention to any risk issues that may become obvious to the therapists as they will have access to routine self-monitoring data on the NA scale – which will be used as a mechanism to identify cases that [a] are showing increasing risk related to suicidal ideation, [b] increased symptom severity, or [c] increased functional impairment. The UPLift website will automatically identify and flag these three types of clinical risk indicators to participating therapists, who will then discuss these cases with clinical supervisors to decide on an appropriate risk management plan – as per routine clinical care.

Data protection and data sharing conditions

Participating therapists will have access to personally identifiable information from participating patients, in order to maintain contact with them throughout treatment and to keep their NHS clinical health records up-to-date. Therapists already have access to identifiable information for these reasons for all patients accessing NHS-TT services. Therapists in this clinical trial will have access to the UPLift website and digital platform, which will provide them with information about their patients' treatment pathway, such as the number of group sessions attended and responses to the questionnaires and interactive skills practices that they complete during the course of treatment. The UpLift platform will link these records to a unique pseudonym for each participant (a "patient ID"). It will be necessary for this website to temporarily store each participant's mobile telephone number, so that the system can automatically send text message reminders to each participant 24 hours prior to their next scheduled group therapy session. These phone numbers will be permanently and irretrievably deleted after the endpoint criteria for the study have been met, as defined above. Group therapy sessions will be audio recorded to check how the therapy is being conducted and provide feedback to therapists. These treatment integrity checks will be done by a member of the research team who is a qualified NHS health professional and will be providing clinical supervision to the therapists in the trial. These members of the research team are well experienced in rating the competency of CBT therapists and will use a session rating tool that has been developed to be sensitive to the group and UP context. This treatment integrity check is a highly common and expected quality control procedure in clinical trials of psychological interventions. The sessions will be recorded by the secure therapy website and will be transferred between the participating NHS trust and the NHS research team using encrypted email (e.g., Egress). The method of the recording will be unobtrusive to the group process. The recordings will be stored on a secure RDASH NHS Trust computer drive only accessible by the research team. Recordings will be destroyed on all other devices.

Data collected as part of all stages of the study (apart from therapy session recordings) will be stored in a secure network drive at the University of Sheffield, with restricted access to members of the research team. Copies of essential records and documentation will be stored in an NHS network drive, to facilitate communications with research assistants that are supporting the participants identification and recruitment process. RDASH NHS Trust will be the data controller, in accordance with the General Data Protection Regulation (GDPR),and will retain copies of the dataset for 10 years from the pre-registered endpoint date for the study. The University of Sheffield will have access to the personal data provided by consenting participants and the anonymised health economic data will be sent to a member of the research team in Canada for analysis. The study dataset will be only accessible to collaborators named in this study protocol, in a fully pseudonymised

format. Study data will not be made available in a publicly accessible repository. Requests for data access are to be made in writing to the Chief Investigator, and will only be granted to qualified academic researchers who provide a study protocol and after this protocol has been pre-registered in a public repository such as the open science framework.

7. Registration

The study protocol will be pre-registered in the open access repository Open Science Framework (OSF; https://osf.io/vup9c/) and registered on the clinical trials registry https://www.clinicaltrials.gov/.

8. Funding

This study has been funded by a research and development grant by Innovate UK – Biomedical Catalyst 2023, Round 1: Industry-led R&D.

- Award reference: 10070872
- Award date: 15 September 2023

9. Dissemination plan

After the conclusion of data analysis, we plan to disseminate findings about this study using a variety of forms of communication, including:

- Scientific journal publications
- Newsletter in lay terminology
- Mental health conferences in the UK and abroad
- NHS Trust communications newsletter and email

10. References

Ayuso-Bartol, A., Gómez-Martínez, M. Á., Riesco-Matías, P., Yela-Bernabé, J. R., Crego, A., & Buz, J. (2024). Systematic review and meta-analysis of the efficacy and effectiveness of the unified protocol for emotional disorders in group format for adults. *International Journal of Mental Health and Addiction.* Advance online publication. https://doi.org/10.1007/s11469-024-01330-z

Barkowski, S., Schwartze, D., Strauss, B., Burlingame, G. M., Barth, J., & Rosendahl, J. (2016). Efficacy of group psychotherapy for social anxiety disorder: A meta-analysis of randomized-controlled trials. *Journal of Anxiety Disorders, 39*, 44-64. https://doi.org/10.1016/j.janxdis.2016.02.005

Barlow, D. H., Ellard, K. K., & Fairholme, C. P. (2010). Unified protocol for transdiagnostic treatment of emotional disorders: Workbook. Oxford University Press.

Beard, J. I., & Delgadillo, J. (2019). Early response to psychological therapy as a predictor of depression and anxiety treatment outcomes: A systematic review and meta-analysis. *Depression and Anxiety, 36*(9), 866-878. https://doi.org/10.1002/da.22931

Böhnke, J. R., Lutz, W., & Delgadillo, J. (2014). Negative affectivity as a transdiagnostic factor in patients with common mental disorders. *Journal of Affective Disorders, 166*, 270-278. https://doi.org/10.1016/j.jad.2014.05.023

Briggs, A. H., Wonderling, D. E., & Mooney, C. Z. (1997). Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Economics, 6*(4), 327-340. https://doi.org/10.1002/(SICI)1099-1050(199707)6:4%3C327::AID-HEC282%3E3.0.CO;2-W

Carlucci, L., Saggino, A., & Balsamo, M. (2021). On the efficacy of the unified protocol for transdiagnostic treatment of emotional disorders: A systematic review and meta-analysis. *Clinical Psychology Review, 87*, 101999. https://doi.org/10.1016/j.cpr.2021.101999

Cassiello-Robbins, C., Southward, M. W., Tirpak, J. W., & Sauer-Zavala, S. (2020). A systematic review of Unified Protocol applications with adult populations: Facilitating widespread dissemination via adaptability. *Clinical Psychology Review, 78*, 101852. https://doi.org/10.1016/j.cpr.2020.101852

Chisholm, D., Sweeny, K., Sheehan, P., Rasmussen, B., Smit, F., Cuijpers, P., & Saxena, S. (2016). Scaling-up treatment of depression and anxiety: a global return on investment analysis. *The Lancet Psychiatry, 3*(5), 415-424. https://doi.org/10.1016/S2215-0366(16)30024-4

Cohen, Z.D., Delgadillo, J., DeRubeis, R.J. (2021). Personalized treatment approaches. In: M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Handbook of Psychotherapy and Behavior Change* (7th ed.). Wiley.

Delgadillo, J., Kellett, S., Ali, S., McMillan, D., Barkham, M., Saxon, D., Donohoe, G., Stonebank, H., Mullaney, S., Eschoe, P., Thwaites, R.,& Lucock, M. (2016). A multi-service practice research network study of large group psychoeducational cognitive behavioural therapy. *Behaviour Research and Therapy, 87*, 155-161. https://doi.org/10.1016/j.brat.2016.09.010

Dolan, N., Simmonds-Buckley, M., Kellett, S., Siddell, E., & Delgadillo, J. (2021). Effectiveness of stress control large group psychoeducation for anxiety and depression: Systematic review and meta-analysis. *British Journal of Clinical Psychology, 60*(3), 375-399. https://doi.org/10.1111/bjc.12288

Emsley, R., Lunt, M., Pickles, A., & Dunn, G. (2008). Implementing double-robust estimators of causal effects. *The Stata Journal, 8*(3), 334-353. https://doi.org/10.1177/1536867X0800800302

Fenwick, E., & Byford, S. (2005). A guide to cost-effectiveness acceptability curves. *The British Journal of Psychiatry, 187*(2), 106-108. https://doi.org/10.1192/bjp.187.2.106

Fisher, A. J., Bosley, H. G., Fernandez, K. C., Reeves, J. W., Soyster, P. D., Diamond, A. E., & Barkin, J. (2019). Open trial of a personalized modular treatment for mood and anxiety. *Behaviour Research and Therapy, 116*, 69-79. https://doi.org/10.1016/j.brat.2019.01.010

Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., & Davidian, M. (2011). Doubly robust estimation of causal effects. *American Journal of Epidemiology, 173*(7), 761-767. https://doi.org/10.1093/aje/kwq439

Howard, K. I., Lueger, R. J., Maling, M. S., & Martinovich, Z. (1993). A phase model of psychotherapy outcome: Causal mediation of change. *Journal of Consulting and Clinical Psychology, 61*(4), 678–685. https://doi.org/10.1037/0022-006X.61.4.678

Husereau, D., Drummond, M., Augustovski, F., de Bekker-Grob, E., Briggs, A. H., Carswell, C., ... & Staniszewska, S. (2022). Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. *MDM Policy & Practice, 7*(1), 23814683211061097. https://doi.org/10.1177/23814683211061097

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12–19. https://doi.org/10.1037/0022-006X.59.1.12

Kahan, B.C., Hall, S.S., Beller, E.M., Birchenall, M., Chan, A.W., Elbourne, D., Little, P., Fletcher, J., Golub, R.M., Goulao, B., Hopewell, S., Islam, N., Zwarenstein, M., Juszczak, E., Montgomery, A.A. (2023). Reporting of Factorial Randomized Trials: Extension of the CONSORT 2010 Statement. *JAMA, 330*(21), 2106-2114. https://doi.org/10.1001/jama.2023.19793

Karyotaki, E., Riper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., Mira, A., ... & Cuijpers, P. (2017). Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. *JAMA Psychiatry, 74*(4), 351-359. https://doi.org/10.1001/jamapsychiatry.2017.0044

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(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. *Annals of Internal Medicine, 146*(5), 317–325. https://doi.org/10.7326/0003-4819-146-5200703060-00004

Longley, S. L., & Gleiser, T. S. (2023). Efficacy of the Unified Protocol: A systematic review and meta-analysis of randomized controlled trials. *Clinical Psychology: Science and Practice, 30*(2), 208–221. https://doi.org/10.1037/cps0000141

Luedtke, A., Sadikova, E., & Kessler, R. C. (2019). Sample size requirements for multivariate models to predict between-patient differences in best treatments of major depressive disorder. *Clinical Psychological Science, 7*(3), 445-461. https://doi.org/10.1177/2167702618815466

MacKenzie, K. R. (1983). The clinical application of a Group Climate measure. In R. R. Dies & K. R. MacKenzie (Eds.), Advances in group psychotherapy: Integrating research and practice (pp. 159-170). New York: International Universities Press.

National Collaborating Centre for Mental Health (2024). *The Improving Access to Psychological Therapies Manual, Version 7*. https://www.england.nhs.uk/publication/the-improving-access-to-psychological-therapies-manual/

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

Nahum-Shani, I., Dziak, J. J., & Collins, L. M. (2018). Multilevel factorial designs with experiment-induced clustering. *Psychological Methods, 23*(3), 458–479. https://doi.org/10.1037/met0000128

NHS England. (2022). *Psychological Therapies, Annual report on the use of IAPT services, 2021-22*. https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-annual-reports-on-the-use-of-iapt-services/annual-report-2021-22

Nye, A., Delgadillo, J., & Barkham, M. (2023). Effectiveness of personalized psychological interventions: A systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology, 91*(7):389-397. https://doi.org/10.1037/ccp0000820

Osma, J., Peris-Baquero, O., Suso-Ribera, C., Sauer-Zavala, S., & Barlow, D. H. (2021). Predicting and moderating the response to the unified protocol: Do baseline personality and affective profiles matter? *Cognitive Therapy and Research, 45*, 817-830. https://doi.org/10.1007/s10608-021-10208-6

Raudenbush, S. W. (1993). Hierarchical linear models and experimental design. In L. Edwards (Ed.), *Applied analysis of variance in behavioral science* (pp. 459-496). New York: Marcel Dekker.

Sauer-Zavala, S., Cassiello-Robbins, C., Conklin, L. R., Bullis, J. R., Thompson-Hollands, J., & Kennedy, K. A. (2017). Isolating the unique effects of the unified protocol treatment modules using single case experimental design. *Behavior Modification, 41*(2), 286-307. https://doi.org/10.1177/0145445516673827

Sauer-Zavala, S., Cassiello-Robbins, C., Ametaj, A. A., Wilner, J. G., & Pagan, D. (2019). Transdiagnostic treatment personalization: The feasibility of ordering unified protocol modules according to patient strengths and weaknesses. *Behavior Modification, 43*(4), 518-543. https://doi.org/10.1177/0145445518774914

Sauer-Zavala, S., Southward, M. W., Stumpp, N. E., Semcho, S. A., Hood, C. O., Garlock, A., & Urs, A. (2022). A SMART approach to personalized care: preliminary data on how to select and sequence skills in transdiagnostic CBT. *Cognitive Behaviour Therapy, 51*(6), 435-455. https://doi.org/10.1080/16506073.2022.2053571

Spitzer, R.L., Kroenke, K., Williams, J.B., and Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine, 166*(10), 1092-1097. doi:10.1001/archinte.166.10.1092

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics, 28*(1), 112–118. https://doi.org/10.1093/bioinformatics/btr597

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: a systematic review and meta-analysis 1980–2013. *International Journal of Epidemiology, 43*(2), 476-493. https://doi.org/10.1093/ije/dyu038

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., & Stewart-Brown, S. (2007). The Warwick-Edinburgh mental well-being scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes, 5*(1), 63. https://doi.org/10.1186/1477-7525-5-63

TEQ Group. (1990). EuroQol-a new facility for the measurement of health-related quality of life. *Health Policy, 16*(3), 199-208. https://doi.org/10.1016/0168-8510(90)90421-9