Behavioural therapy for inter-episode bipolar symptoms: A multiple baseline case series evaluation (STABILISE)

Statistical Analysis Plan

Trial ID: ISRCTN19416314

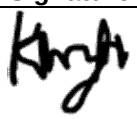
This statistical analysis plan (SAP) was finalised on 15.6.23; version number 1.2

This SAP builds on the information provided in the study protocol, version 1.2 27.4.23

Revisions to this SAP and details of the revisions are provided below.

SAP revision number	**Date of revision *(DD-MMM-YYYY)***	**Timing of SAP revision in relation to analyses, eg: before interim analysis**	**Details of the revision**	**Justification for the revision**
1				
2				
3				
…				

This SAP was produced by members of the trial team as described in the table:

Name	**Affiliation(s)**	**Role**	**Signature**
Kim Wright	University of Exeter	CI	

1. Introduction

1.1. Background and rational

Bipolar Spectrum Disorders typically involve periods of depression, and periods of very high energy and mood (hypomania or mania). Bipolar Disorders (BDs) are common, affecting around 1 in 20 people at some point in their lifetime. They can be very distressing and disruptive to the people who experience them, and for friends and relatives. Many people with these conditions have mood issues outside depressive or manic episodes. Often these issues include ongoing low mood, and/or frequent swings in mood or emotions. These are sometimes called “inter-episode bipolar symptoms” (IEBS).

There are psychological (talking) therapies designed to help people with Bipolar Disorders, but these tend to be aimed at preventing a relapse of depression or mania, or they are aimed at helping people recover from a period of depression. There is no universally accepted talking therapy for helping people who have ongoing low mood or mood swings in between full episodes.
The aim of this study is to provide an initial evaluation of a psychological therapy designed to address these ongoing symptoms. A total of 12 people, in two consecutive cohorts of six to allow refinement of the therapy in between, will be offered the therapy and will be randomly allocated to a waiting period of 3-5 weeks after their baseline assessment. They will then repeat some of the assessment measures and commence the therapy which will be delivered over a period of approximately seven months. At the end of therapy, they will complete the set of assessment measures again and complete a three week post therapy monitoring period. Throughout the baseline, therapy and post-therapy periods they will be invited to complete some brief measures of symptoms on a weekly basis. At two points (during the first two weeks of the baseline period and of the post therapy period) they will be invited to complete a block of mood and activity monitoring when they will be asked briefly about mood and key activities five times per day for 14 days. Views on the acceptability of the therapy will be collected from participants in terms of numerical ratings of acceptability, written feedback and through interviews.

1.2. Objectives

i) To allow initial evaluation of intervention safety, feasibility and acceptability.
ii) To investigate whether the pattern of change in symptoms pre to post intervention is consistent with the potential of the intervention to deliver benefit.
iii) To refine the therapy protocol and procedures for training and supervising therapists.
iv) To develop an initial therapy competence and adherence measure.

2. Study methods

2.1. Design

Two cohort randomised multiple baseline ABA case-series design. Randomisation is to a wait period of 3,4 or 5 weeks before starting therapy. The intervention is an adapted form of behavioural therapy delivered on a one-to-one basis over up to 22 sessions, over up to 7 months.

2.2. Randomisation

Participants will be randomised to different baseline assessment lengths (between 3 and 5 weeks). The randomisation sequence will be generated and administered by an individual independent of the study on a 1:1:1 allocation sequence (block randomisation) with one block per cohort of 6. The randomisation sequence will be generated by Fiona Warren (Exeter CTU statistician) and then passed to a researcher independent of the project and research team. The research team will contact this person after eligibility is confirmed for each participant to obtain the allocation next in the sequence.

2.3. Sample Size

It is good practice to have at least 3 replications of the pattern of change across different cases (inter-case replication) as well as more than one therapist treating patients. Patients are treated in two cohorts within the current study: a total of 12 participants, 6 in each cohort, would allow these criteria to be met within each cohort (with at least two therapists seeing patients in each cohort).
Participants who exit the study after randomisation, but before completing the pre-therapy measures will be replaced (with new random allocation) to give N=12 commencing the therapy period.

2.4. Interim assessment of feasibility criteria and progression rules

Because of the intention to rapidly proceed to a feasibility trial if the therapy protocol appears promising, we will assess the data against the minimum thresholds for progression after those in cohort one have each attended at least eight therapy sessions (or have discontinued therapy, if before session 8). If the progression thresholds are met, we will apply for approval for a feasibility trial at this point. The trial preparation and approval process will take around six months; over this period learning from the remainder of the case series will continue to inform the therapy protocol. The progression rules are in place simply to ensure we do not progress with a therapy that is likely to be unsafe or unacceptable and this cannot be rectified in time prior to the trial commencing. If the progression thresholds are not met at the end of cohort one we will alter the research and therapy protocols as required (with support from the PPI panel and TSC) and progress with cohort 2. The rules will be re-examined in relation to cohort 2 (in relation to whether to proceed to a feasibility trial) once all six participants have attended at least 8 sessions or discontinued therapy.

2.4.1. Discontinuation criteria for individual participants:

1. Participant does not wish to continue with the intervention and study (participants will be considered to be remaining in the study if they discontinue therapy, unless they opt to also leave the study).
2. The participant experiences a serious adverse reaction (mental or physical health event that results in significant impairment, hospitalisation or death) that is judged to be the direct result of the intervention or trial participation.
3. Participant and / or therapy / research team believe that the intervention or trial participation will result in, or is likely to result in, a serious adverse reaction if continued.

4. The participant does not attend more than three consecutive therapy sessions without explanation despite attempts to contact them (sometimes referred to as “DNA”): this will be judged to indicate discontinuation of therapy.

2.4.2. Discontinuation of the trial:

1. Should an unexpected serious adverse reaction occur to either the therapy or the trial procedures, and this is judged to be directly related to trial participation or to the therapy, the trial will be temporarily halted pending investigation and analysis of the extent to which future risk can be mitigated against. If it is judged that this is not possible, the trial will be discontinued. This process will be led by the sponsor in collaboration with the TSC chair and CI.
2. Should information come to light that indicates that the therapy intervention or trial procedures are unsafe, the process outlined in (1) above will be followed.

2.5. Timing of final analysis

Final analysis will take place once the final data collection point has passed for the final participant and data have been entered and cleaned.

2.6. Timing of outcome assessments

Table 2.6.1: Timings of outcome assessments

Measure	**Intake**	**Pre-treatment**	**Weekly***	**Immediately post-treatment**	**7 months (30 weeks) post treatment onset**	**6 months (26 weeks) post end of treatment**
PHQ-9	X	X	X		X	
ASRM	X	X	X		X	
Depression-elation subscale of ALS short form			X			
ALS short form	X	X			X	
Brief QoL.BD	X	X			X	
BRQ	X	X			X	
GAD-7	X	X			X	
Momentary assessment block	X			X		
Post therapy feedback questionnaire				X		

Post therapy interview				X		
Booster sessions feedback questionnaire						X

Data that are collected outside of the time-windows shown in Table 2.6.2 will be excluded from the primary analysis but included in a secondary sensitivity analysis (quantitative measures) or will be included but with discrepancy in collection time noted (qualitative).

Table 2.6.2: Time windows for data collection per assessment point

Timepoint	**Data collection window**	**Notes**
Intake (prior to randomisation)	Not applicable	Required for eligibility
Weekly from intake to 3 weeks post end of therapy (randomisation occurs between weeks 0 and 1)	+7 days	late returns within 5 days will be accepted and subsequent time point will continue as planned
5x per day for 14 days starting day after eligibility confirmed; second 14 day block starting day after end of treatment confirmed (randomisation occurs during first week of block 1)	+3 hours	late returns within this window will be accepted and subsequent timepoint will continue as planned
Pre-treatment (post randomisation)	+28 days	
7 months post pre-treatment	+28 days	
Post end of treatment	+28 days	
6 months post end of treatment	+28 days	

3. Statistical principles

3.1. Confidence intervals and p-values

In accordance with the feasibility aims of the study, all outcomes will be reported descriptively. All quantitative outcomes are continuous measures and will be reported at each timepoint measured using the mean and standard deviation. Descriptive analyses will include all reported data collected within the time window specified (Section 2.6). In addition, as a sensitivity analysis, we will also report descriptive analyses including all reported data (irrespective of whether within the time window).

In accordance with the aims of the study, which primarily concern feasibility, inferential statistical tests will be limited. Where these are carried out (calculation of pre-post change in terms of reliable and reliable and clinically significant change, and group-level, pre-post effect size), effect sizes and two-sided 95% and 90% confidence intervals will be reported.

3.2. Intervention adherence and protocol deviations

As this is a feasibility study we will not set a minimum threshold for therapy attendance in order for the participant to be considered to have had an adequate minimum dose. Instead we will explore attendance rates and present these in terms of frequencies, minimum, maximum and median. We will also consider these narratively in relation to quantitative and qualitative outcomes on a per-participant basis, in order to inform decisions about minimum adequate dose in subsequent studies of the intervention. No sensitivity analyses, i.e. by exclusion of participants who do not complete a pre-specified minimum number of sessions, will be performed.

Protocol deviations will be recorded by the study team; major deviations will be reported in terms of count and narratively in the final report.

3.3. Study populations

Analyses will use observed data only (i.e. without imputation). Data on session attendance rates will separate out any participants who left the study before the pre-treatment point and thus did not reach the point in the study where therapy commenced.

4. Trial population

4.1. Screening Data

Numbers of participants expressing an interest in the study, undertaking the initial screening process, giving consent, found to be eligible and entering the study will be reported.

4.2. Eligibility Criteria

Participants will be adults who i) meet research diagnostic criteria for Bipolar I or II Disorder, Other Specified Bipolar Disorder or Cyclothymic Disorder; ii) do not meet criteria for a manic or severe depressive episode; iii) have IEBS, defined as at least mild depressive symptoms (Patient Health Questionnaire [PHQ9] >=5) or above-average bipolar mood instability defined as >=1.3 on the brief Affective Lability Scale (ALS) depression-elation scale; iv) are willing to engage in psychological work addressing IEBS or its impact on functioning; v) sufficient English to complete questionnaires without translation; vi) have completed the intake measures.
Exclusion criteria include: i) current substance dependence according to ICD-11 criteria; ii) frequent and serious self-harm that cannot be safely managed in a community outpatient setting; iii) currently engaged in another psychological therapy for bipolar disorder.

4.3. Recruitment

The CONSORT diagram will include: number expressing an interest, number completing screening and reasons for any exclusion, number giving consent and assessed for eligibility, number eligible and randomised and reasons for any exclusions, number completing therapy, number giving post

therapy feedback, number completing 7 month follow up assessment, number giving feedback post booster sessions.

4.4. Withdrawal / follow-up

Withdrawal from the research aspect of the study will be reported separately to withdrawal from therapy as participants may withdraw from one but not the other. Participants who do not complete the intervention will be given the opportunity to complete the follow-ups. With respect to both, reasons for withdrawal, where available, will be reported. Numbers completing each measure at each time point will be reported.

4.5. Baseline patient characteristics

Demographic and clinical characteristics of the sample will be reported: gender, age ethnicity self-perceived financial status, psychiatric medication prescribed currently (Yes / No and numbers prescribed each type), research diagnosis (bipolar I, bipolar II, cyclothymic disorder, other specified bipolar), illness history (number of episodes, age of onset), current presenting difficulty (mood instability, low mood or both), currently held within primary care only / secondary care scores on study measures at intake. Binary and categorical characteristics will be reported as numbers and percentages in each categories; continuous characteristics will be reported using the mean and standard deviation (or median and IQR if not normally distributed), and range for age characteristics and number of previous episodes.

5. Statistical analysis

Analyses will be conducted using SPSS and STATA.

5.1. Outcome definitions

Means of assessing primary (feasibility) outcomes

1. Safety will be measured by rates of therapy-related adverse events across the study period, and rates of reliable deterioration from the pre-treatment point on the Patient Health Questionnaire (PHQ), Altman Scale for Rating Mania (ASRM), Affective Lability Scale (ALS), Generalised Anxiety Disorder questionnaire (GAD), Quality of Life in Bipolar Disorder scale (QoL.BD) and Bipolar Recovery Questionnaire (BRQ) at 7 months (30 weeks) follow-up.
2. Acceptability will be measured by therapy uptake and completion rates across the study period, and quantitative and qualitative feedback from participants and therapists at the post-therapy point.
3. Potential of the intervention to deliver benefit will be measured by rates of reliable, and reliable and clinically significant, change in the Patient Health Questionnaire and the bipolar-depression scale of the Affective Lability Scale, from the baseline period (mean score) to the 3 week post-therapy period (mean score).

Secondary outcome measures

1. Rates of reliable, and reliable and clinically significant, change on the following measures: Bipolar Disorder Recovery Questionnaire, Quality of Life in Bipolar Disorder Scale and Generalised Anxiety Disorder assessment scale from pre-therapy to 7-month follow-up.
2. Rates of reliable, and reliable and clinically significant, change on the Altman Scale for Rating Mania from the baseline period (mean score) to the 3 week post-therapy period (mean score).
3. Group-level change in PHQ, ALS, ASRM, BRQ, QoL.BD, GAD from baseline to 6-month follow-up and pre-therapy to 7-month follow-up.
4. Mood instability will be measured (in addition to the ALS) by a visual analogue scale of mood at five timepoints per day for 14 days on two occasions (pre and post therapy) for each participant. This allows for the calculation of mood instability before and after therapy.

5.1.1. Scoring rules and handling of individual item missing data

Measure	Scoring	Treatment of missing data
PHQ9	Sum of items	Imputation based on mean of remaining items if number of missing items <=2
ALS	sum of three subscales (elation, anxiety, anger); sum of these to give total score	Imputation of subscale score based on mean of remaining items if less than 25% of items are missing
ASRM	Sum of items	Imputation based on mean of remaining items if number of missing items >2
BRQ	Sum of items (with reversal of selected items)	Imputation based on mean of remaining items if number of missing items <9.
QoL.BD	Sum of items	Imputation based on mean of remaining items if number of missing items <3.
GAD7	Sum of items	Imputation based on mean of remaining items if number of missing items >2.

5.2. Analysis methods

In relation to aim (i) which concerns safety, feasibility and acceptability, we will report descriptively the number of adverse events, serious adverse events and the number of each judged to be related to the therapy, and rates of reliable deterioration. We will also report the number of participants experiencing at least one adverse event. We will report the number of participants expressing an interest in the study, giving consent, found to be eligible, commencing treatment and completing treatment, as well as the median, mode and range of sessions attended during treatment. We will report descriptively the average and distribution of scores on the therapy feedback items.

Aim (ii) investigates the potential for clinical efficacy. Analytical approaches to examining clinical outcome data in a case series can include visual and statistical methods, and focus on change at the

level of the individual participants, or aggregated across the sample. In this population the extent to which a stable baseline can be expected is not clear, therefore we will use visual methods combined with descriptive statistics (Lane & Gast, 2013) to allow examination of overall patterns of change. Assessment of symptom scores over time for each participant will be made by two independent raters who will then reach consensus on each pattern (improvement, deterioration, no change) and will include calculation of overlap between weekly scores in the baseline, therapy and post-therapy periods through calculation of baseline-corrected Tau following the steps described by Tarlow, (2017) whereby the baseline trend is removed before performing Tau analysis if it is statistically significant.
In terms of statistical analyses, reliable change index scores (RC) will be calculated to assess for the statistical reliability of the changes on the clinical outcome variables between each phase of the case-series for each participant. A participant is said to show reliable change when their change score from pre to post therapy is more than 1.96x the standard error of the difference (Ferguson, Robinson, & Splaine, 2002), the latter being calculated using values from the literature in most comparable samples. RC scores greater than the z-score level of 1.96 are statistically significant at p>.05. We will report the proportion of participants showing: i) reliable improvement /no change / reliable deterioration in the clinical outcome measures; ii) improvement that is both reliable and clinically significant, where clinically significant change is defined as the scoring falling below / above the established threshold for clinical significance (dependent upon whether decrease / increase in score is considered to represent clinical improvement). Where established thresholds are not available we will use Jacobson and colleagues (1984) criteria a, b or c to define clinically significant change depending upon the information available in the literature for each measure. Criterion a requires the participant's score to have moved more than 2 SDs from the mean of the patient group, criterion b requires the participant's score to have moved to within 2 SDs of the unaffected reference group, and criterion c requires the score to have moved closer to the mean of the unaffected group than to the mean of the clinical group.
Using continuous scores on the outcome measures, effect sizes and 95% confidence intervals will be computed to obtain a preliminary estimate of the potential magnitude of change on each measure from baseline to post therapy and pre to post therapy: these statistics can be used by other researchers in future meta-analyses seeking to synthesise the outcomes of a number of studies (Manolov et al., 2022).
Relevant to aim (iv) which is to develop an initial therapy competence and adherence measure, using the version of the measure developed during the case series two therapists will independently assess the same 12 recordings of therapy sessions to estimate inter-rater agreement on the measure by calculating the intraclass correlation coefficient using a two-way random effects model, for both consistency and absolute agreement.

Planned secondary analyses

Relevant to aim (i), qualitative interviews will be audio-recorded and then transcribed. Participants' responses will be read closely and coded using a framework approach. Data relevant to each category will be compared and summarised and non-conforming cases examined closely in order to

understand similarities and differences in perspective. Interview recordings will be transcribed by a member of the study team, or a suitable individual working with the team who has signed a confidentiality agreement.

Relevant to aim (ii), data from the momentary assessment blocks will be used to examine change in mood variability from the baseline to post therapy phases, by calculating mean variability (SD) across each of the two blocks.

5.3. Harms

We will report descriptively the number of adverse events, serious adverse events and the number of each judged to be related to the therapy, and rates of reliable deterioration. Instances of adverse events will be reviewed throughout the trial by the chief investigator (and the TSC and sponsor if required according to the relevant SOPs), following the relevant SOPs with regard to classification, response and reporting.

6. Abbreviations and definitions

Abbreviation	Meaning	Detail
SAP	Statistical Analysis Plan	
SOP	Standard Operating Procedures	SOPs document a series of steps to be followed to accomplish a process. They include the purpose and scope of a process, describes procedure (i.e. what, where, and when), assigns responsibility for action and decisions (i.e. who), identify records to be kept, and identify associated and reference documents (as applicable)
TSC	Trial Steering Committee	The role of the TSC is to provide overall supervision for a study on behalf of the Sponsor and Funder and to ensure that the project is conducted in line with regulatory requirements and to the rigorous standards set out in the Department of Health's Research Governance Framework for Health and Social Care and Guidelines for Good Clinical Practice.