ECLIPSE (CIRCUITS study)

A randomised controlled trial of cognitive remediation therapy in patients with non-affective psychosis

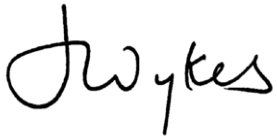
Statistical Analysis Plan for End of Trial
Version 1.0
Version 1.0 started: 27/07/2020

ISRCTN: ISRCTN14678860

Trial Statistician: Dominic Stringer

Signature  Date: 23/09/2020

Chief Investigator: Professor Dame Til Wykes

Signature:  Date: 23rd September 2020

Trial Steering Committee Chair: Professor David Kingdon

Signature.....  ..... Date....26/08/20.....

This document details the presentation and analysis strategy for the primary paper reporting results from the ECLIPSE (CIRCuiTS study) trial. It is intended that the results reported in these papers will follow this strategy; subsequent papers of a more exploratory nature will not be bound by this strategy but will be expected to follow the broad principles laid down for the principal paper(s). These principles are not intended to curtail exploratory analysis or to prohibit sensible statistical and reporting practices,but are intended to establish the strategy that will be followed as closely as possible, when analysing and reporting the trial. Reference was made to the trial protocol (ECLIPSE Study9 Research Protocol V1.5 18 12 2018), ICH [1] guidelines on Statistical Principles (E9) and CONSORT [2] guidelines.

QUANTITATIVE ANALYSIS PLAN

Investigators
Dr Clare Reeder
Dr Diana Rose
Dr Matteo Cella
Dr Gregory Aarons
Prof Sue Dopson
Prof Sonia Johnson
Prof Max Birchwood
Prof David Fowler
Dr Jesus Perez
Prof Andrew Pickles
Prof Kathy Greenwood
Prof Rachel Upthegrove
Dr Andrew Thompson
Dr Rosa Ritunnano

Joint Chief investigators
Prof Til Wykes
Prof Eileen Joyce

Trial manager
Dr Emese Csipke

Trial statistician
Prof Andrew Pickles
Dominic Stringer

Trial health economist

Prof Paul McCrone

Description of the trial

ECLIPSE is a three-year trial of computerised cognitive remediation (CIRCUITS). This will be conducted in the Early Intervention Services (EIS) of NHS Mental Health Trusts with catchment areas ranging from high density urban to rural. Three different methods of providing CR will be evaluated: intensive, group, independent and a comparator of treatment as usual.

Principal research objectives to be addressed

Primary objectives

To determine the best way of introducing cognitive remediation therapy (CR) for psychosis into NHS early intervention services in order to optimise individual functional outcomes.

To do this we will conduct a randomised controlled trial of 3 CR implementation methods and treatment as usual to evaluate:

The effectiveness of each CR method in the achievement of personal goals as measured by the total score on the Goal Attainment Scale.

Secondary objectives

The effectiveness of each CR method in improving individual components that might contribute to goal attainment (cognition, social function, self-esteem and symptoms)

The cost-effectiveness of each implementation method by combining service costs (derived from administrative data and the Client Service Receipt Inventory) and quality-adjusted life years (measured by EQ-5D)

To compare service use and costs at 15 weeks post-randomisation (post-therapy assessment)

Secondary objectives 3 and 4 are objectives of the economic analysis and will not be covered in this document.

This SAP relates only to the main paper resulting from the main outcome (GAS) results of study 9. This trial is part of a larger programme grant on the advice that can be given to the NHS to implement cognitive remediation. Information gleaned from the costs of treatment, cost-effectiveness as well as participant and staff satisfaction will contribute to the overall final recommendations and conclusions. How this recommendation will be formed is not part of this SAP.

Trial design including blinding

ECLIPSE is a randomised four-arm trial of three different implementation modes of Cognitive Remediation Therapy (CR) compared to Treatment as Usual in people presenting with non-affective psychosis. However, from 19/02/2019, the Independent and Treatment arms were dropped, and the trial is continuing to randomise to the Group CR and Intensive CR arms only.

Clinicians and research workers completing baseline and follow up assessments (including outcome assessment) will be blind to group allocation. Blinding will be maintained by ensuring patient data are stored in separate offices, locating research and therapy staff in separate offices and, where possible, third party management of appointments to avoid appointment clashes. In the case of a research worker becoming unblinded, another researcher will be sent to the site to complete the rest of the assessments and this will be recorded.

Therapists and patients will necessarily not be blind to group allocation (following randomisation).

The trial has four groups, although the Independent and TAU arms were dropped as of 19/02/2019 and the trial will randomise only to the Intensive CR and Group CR arms from this date.

1. Intensive CR

(a) All participants are offered 10.5 weeks of twice weekly individual therapy.

(b) Sessions (60-180 minutes duration) are in 3 parts: (1) 20-60 minutes of CR with a therapist; (2) 20-60 minutes of in vivo transfer work (i.e. putting CR strategies into real life) with a therapist; (3) 20-60 minutes of independent CR, set up by the therapist on site, or done off-site in the patient's own time.

(c) Each patient receives up to 42 hours of CR (21 with therapist, 21 independently) and 21 hours of in vivo transfer work (with a therapist).

2.Group CR

(a) All participants attend 14 weeks of three times weekly group therapy (up to 42 hours of CR in total).

(b) Group sessions last 1 hour, with attendance for at least 20 minutes considered have completed a session.

(c) Participants will join the group as soon as possible following randomisation. Each group will have one therapist.

(d) Group sessions begin and end with group activities, relating to goal-setting and metacognition. During the rest of the session, patients work independently on CIRCuiTS tasks (at the same time) with the therapist offering help and support to individuals on an as-needed basis.

3. Independent CR

(a) All participants are offered one individual session to get started.

(b) Following this, participants are offered up to 41 independent sessions (up to 42 hours of CR in total).

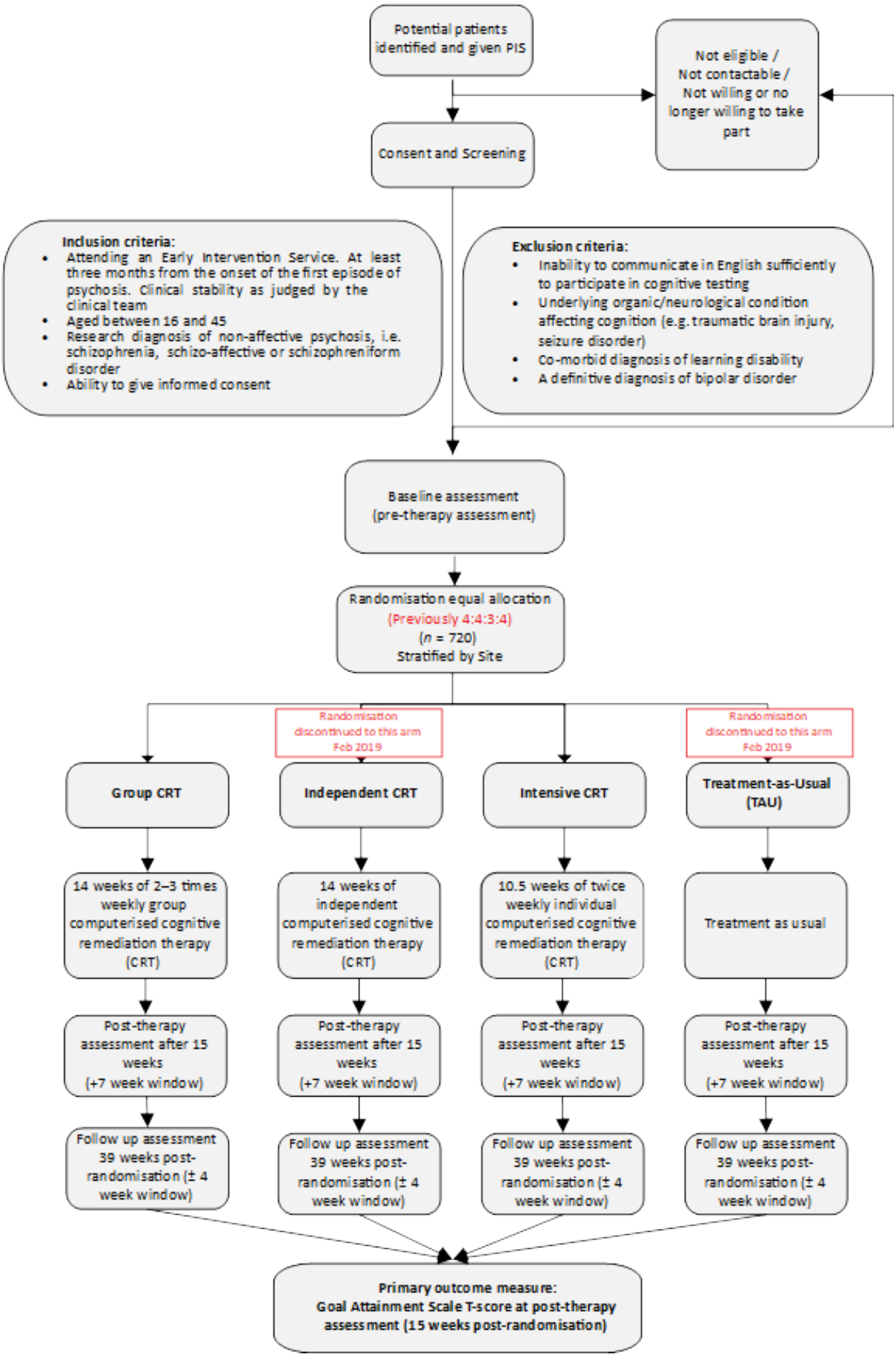
(c) To support the independent sessions, the therapist offers telephone contact and/or attendance at daily drop-in sessions on an as needed basis (estimated average therapist time: 1 hour per fortnight).

4. Treatment as usual

This is defined as multi-modal treatment and will consist of different therapies as defined as necessary by the treating team.

Figure 1: Description of the participant flow and assessment protocol

Trial design flow diagram



Method of allocation of groups

Originally, consented patients were randomised in blocks of 15 stratified by research site with randomisation in proportions 4:4:3:4 (group CR/independent CR/ intensive CR/ treatment-as-usual). Alternative proportions to 4:4:3:4 were used for blocks of less than 15 participants.

Following difficulties in timely recruitment of complete blocks of patients, the treatment allocation process was then changed to randomisation of individual patients with equal allocation to the 4-arms, stratified by site, using a random sequence of blocks of variable size. This was further changed to be randomisation to 2-arms (otherwise with the same specification) when the independent and treatment as usual arms were dropped.

This was implemented using an independent web-based randomisation service at the UKCRC registered King's Clinical Trials Unit (CTU). Participant allocation will be communicated only to therapists by email.

Duration of the treatment period

The treatment period will be for 14 weeks from 0 weeks post-randomisation to 14 weeks post-randomisation. Participants in the Intensive CR arm are offered 10.5 weeks of twice weekly individual therapy. Participants in the Group CR arm attend 14 weeks of three times weekly group therapy.

Participants who were previously randomised the Independent CR arm were offered up to 41 independent sessions over the 14 weeks. Participants who were previously randomised the Treatment as usual arm continued to receive treatment as usual over the 14 weeks.

Frequency and duration of follow-up

Participants will complete follow up measures following therapy (15 weeks post-randomisation) and at 6 months post-therapy (39 weeks post-randomisation).

Visit windows

The following visit windows apply to the assessments carried out in the trial:

Pre-therapy (baseline) assessment; 12 to 0 weeks pre-randomisation.

Post-therapy assessment; 15 weeks post randomisation +4-week window, i.e. 15- and 22-weeks post-randomisation.

6-month post-therapy assessment; 39 weeks post-randomisation with +/- 4-week window i.e. 35 to 44 weeks post-randomisation.

We will however utilise all available data in the analyses, regardless of when visit actually occurred. A sensitivity analysis will be carried out on the effect of using only data from assessments that complied with the visit windows (see Section 3.14)

Data collection

Eligibility screening

Inclusion criteria

Attending an Early Intervention Service or any individual within 5 years of their first episode of psychosis. At least three months from the onset of the first episode of psychosis. Clinical stability as judged by the clinical team;

Aged between 16 and 45;

A research diagnosis of non-affective psychosis, i.e. schizophrenia, schizo-affective or schizophreniform disorder;

Ability to give informed consent.

Exclusion criteria

Inability to communicate in English sufficiently to participate in cognitive testing;

Underlying organic/neurological condition affecting cognition (e.g. traumatic brain injury, seizure disorder);

Co-morbid diagnosis of learning disability.

A definitive diagnosis of bipolar disorder

Measures

A timeline of data collected is given in the Schedule of Assessments and Measures (section B of this document). What follows is a brief overview to aid understanding of the analysis plan.

Measures collected at Baseline only

Eligibility and Consent

Mini International Neuropsychiatric Interview (MINI)

Sociodemographics (ethnicity, employment, living situation, and relationship status)

Background Trauma and Risk Information

Duration of Untreated Psychosis (DUP) assessed using the relevant sections of the Nottingham Onset Schedule (NOS)

The Wechsler Test of Adult Reading

Wechsler Abbreviated Scale of Intelligence II

Primary outcome measure at 15 weeks post-randomisation (also collected at Baseline)

The Goal Attainment Scale (GAS)

Secondary outcome measures at 15 weeks and 39 weeks post-randomisation (also collected at Baseline)

The following measures will form part of the Primary outcome paper:

The Goal Attainment Scale (GAS) (at 39 weeks)

Social and Occupational Functioning Assessment Scale (SOFAS)

The Time Use Survey – total hours in structured activity

CAINS total score

Composite Cognitive score as measured using the Cambridge Neuropsychological Test Automated Battery (CANTAB). The included CANTAB measures will be:

1. Motor Screening Task: MOT
2. Reaction Time: RTI
3. Paired Associated Learning: PAL
4. Spatial Working Memory
5. One-touch Stockings of Cambridge: OTS
6. Emotion Recognition Task: ERT
7. Rapid Visual Information Processing: RVP
8. Attention Switching Task: AST

Self-esteem as measured using the Rosenberg Self Esteem Scale

The following measures are process outcomes (not collected in TAU arm) rather than necessarily secondary outcomes and will be compared between arms in the primary paper:

Bespoke satisfaction measure (service users)

Bespoke satisfaction measure (staff)

Engagement with therapy

The following measures relate to the Economic analyses to be carried out by the Health economists (will not be covered by this SAP but will be in primary paper):

The Client Service Receipt Inventory

The EQ-5D

The following measures will be exploratory or potential moderators/mediators and will not be analysed in the primary outcome paper (will not be covered by this SAP)

Adapted History of Substance Use (alcohol and Drugs)

Computerised Wisconsin Card Sorting Task (WCST)

Digit Span task

Rey Auditory Verbal Learning Test

Rey Osterrieth Complex Figure

IoPPN Narrative Metacognition Task

Metacognition Assessment Scale for the Rey Osterrieth Complex Figure

Self-appraisal Scale (devised for the purposes of this study).

The IoPPN Narrative Metacognition Task and the Self-Appraisal scale will also be done after the Spatial Working Memory (SWM) subtest of the CANTAB

Measure of Insight into Cognition - Self Report (MIC-SR)

Cognitive reserve (measured from WTAR and WASI 11)

Duration of untreated psychosis (NOS)

Premorbid social adjustment (PAS) (removed in most recent visit schedule so not collected on all participants)

Working alliance inventory

Brief Core Schema Scale (removed in most recent visit schedule so not collected on all participants)

Calgary Depression Scale total score (removed in most recent visit schedule so not collected on all participants)

Frontal Systems Behavior Scale

Adverse events

Reported adverse events

Additional post-randomisation (follow-up) measures (to be described only)

Withdrawal from follow up

Psychiatric medication being taken at time of assessment

PANSS total score

Sample size estimation (including clinical significance)

Sample size

A total of 720 first episode psychosis patients will be recruited.

Power

We have the capacity to recruit 900 patients (from 1500 attending 10 services for 3 years) and have allowed for a 20% drop-out pre-randomisation.

Using a design with parallel arms of equal size with 180 patients per arm provides approximately 80% power for a simple group ES difference of 0.3 with alpha=0.05. This increases to 91% for outcomes that correlate 0.5 with baseline (both calculated using sampsi in Stata).

This power calculation was based on comparisons between each of the active arms and TAU. While testing against TAU would be expected to require only a modest sample size, the effect size differences among the active arms are likely to be less substantial.

This power calculation was not controlled for multiple comparisons. Friedlin et al [3] suggest no great advantage in accounting for multiple testing in a multi-arm trial, and also that the advantages of a larger TAU arm are more slight than commonly assumed. Interaction among patients in group delivery is very slight so no allowance for clustering was thought necessary.

The power calculation is based on arms of equal size; the difference in power as a result of the initial unequal allocation is likely to be small as the use of modestly unequal randomisation ratios only very slightly reduces the power of a study ([4]).

Interim analysis

Planned interim analysis at start of trial

A planned interim analysis was to be undertaken by the health economist after recruitment of the first 195 patients (using the post-therapy data at 15 weeks post-randomisation). The economist was not to be blind to trial arm because the costs were to be specific to the intervention provided. That analysis might have resulted in one of the trial arms being closed, with an immediate impact on the randomisation of the next patients.

The decision was to depend on the cost of therapy and the outcomes achieved from it. The relevant outcomes were to be time use, goal attainment, and service user satisfaction. The costs were to be confined to those of the direct therapy inputs and not the cost of other services derived from the Client Service Receipt Inventory. The direct therapy costs were to be calculated from data on the number and length of sessions, number of attendees (for group therapy), and unit costs based on staff grade and overheads. Missing outcome data were to be handled as described for the main analysis. Cost-effectiveness planes were to be generated by plotting 1000 incremental cost-outcome combinations derived using bootstrapping from the sample and comparing each pair of therapies. This would tell us the probability that one therapy had (i) lower costs and better outcomes, (ii) lower costs and worse outcomes, (iii) higher costs and better outcomes, and (iv) higher costs and better outcomes than a comparator. Repeated for each outcome measure, a decision to drop an arm would have been taken if the probability of a good result fell below 25% for each.

Actual interim analysis and contingent recruitment decisions

In the light of recruitment falling short of the target for the interim analysis, and for the need to concentrate remaining recruitment time on fewer arms, an ad-hoc interim analysis was undertaken. Undertaken in November 2018 on the 100 participants with endpoint data at that time, evidence for each trial arm was reviewed by the DMC against the following criteria

Treatment engagement – an arm that has more than 50% of individuals receiving therapy for less than 5 hours.

Cost-effectiveness –more than **£500 increase in costs in an arm** per one point increase in cognition (visual and verbal memory) or for one hour of structured activity

Participant satisfaction –25% of participants disagree with the statement, "Overall I was satisfied with the CIRCuiTS therapy"

Treatment engagement and satisfaction had a high clinical value when considering dropping an arm. Cost-effectiveness was be considered but a key question for the NHS is how to differentiate between the two more costly arms (group treatment and intensive treatment) so the costs for improvement would have to be much greater between these two comparison arms to be confident in dropping one of them. Cost-effectiveness would therefore have less clinical value in making a decision than the other two criteria. If the two high value criteria differ in their conclusions (i.e. suggest different arms to drop) then a clinical judgement would favour making a decision on the basis of user led direct information – satisfaction.

The DMEC recommendations were as follows:

Drop Independent CR arm

Provide power calculations for dropping the Independent CR arm and retaining TAU as well as dropping both and seek opinion from NIHR regarding dropping or retaining TAU on this basis. If NIHR require a clear opinion from the DMEC then the balance of opinion was in favour of retaining TAU.

Programme Steering Committee and Patient Advisory Board involvement

The draft report was sent to the Patient Advisory Board and the Eclipse Programme Steering Committee who supported the proposal to **drop both Independent arm and TAU** which was the preferred option of the study team and was subsequently approved by NIHR.

Revised Power Calculation

Potential sample sizes were calculated following alternative decisions using an overall expected sample size of 438 participants. For the scenario with retaining the Group and Intensive arms, this gave an expected total of 158 and 141 participants in these arms respectively at the end of the trial. For the contrast of Group vs Intensive, assuming 80% with endpoint and follow-up data, with plausible correlation structure (correlation between follow up measures=0.5 and correlation between baseline and follow up =0.2), but making no allowance for clustering and retaining the effect size of 0.3 for a comparison of active-arms gave nominal 79% power (two-tailed alpha=.05).

Impact of the ad-hoc interim analysis on the proposed main analysis

The primary outcome for the main analysis was to be the Goal-Attainment Scale (GAS). While it would be usual to account for the impact of an interim analysis on trial analysis undertaken at the end of recruitment, the relationship between the ad-hoc interim analysis and final analysis is not one that could be easily formalised. We therefore propose undertaking a naïve analysis of the trial, that assumes that decisions made on arm-specific recruitment stoppage were made independent of GAS scores on the available sample. Additionally, two sensitivity simulations would then be undertaken, based on the baseline data, and the observed missing data pattern.

Data analysis plan – Data description

Recruitment and representativeness of recruited patients

A CONSORT flow chart will be constructed. This will include the number of eligible patients, number of patients agreeing to enter the trial, number of patients refusing, numbers randomised to each treatment arm: the number of patients who received at least 1 therapy session, the number continuing through the trial, the number withdrawing, the number lost to follow-up and the numbers excluded/analysed.

Baseline comparability of randomised groups

Baseline descriptions of participants by treatment arm and overall: means and standard deviation or numbers and proportions as appropriate. No significance testing will be used to test baseline differences between the randomised treatment groups.[5]

All baseline variables listed under measures in section 1.7 will be reported overall and by trial arm. Baseline values of primary and secondary outcomes will also be summarised overall and by trial arm.

Adherence to allocated treatment and treatment fidelity

Compliance will be a continuous measure of number of valid sessions the participant attended and will be described by treatment arm and in terms of baseline variables. Compliance will be ascertained from the therapy audit forms.

Loss to follow-up and other missing data

It is the aim of the trial to minimise withdrawal of participants from treatment and follow-up. Completion of a withdrawal from trial is regarded as withdrawal from data collection/follow-up in this analysis.

Withdrawal from trial will be reported by intervention group. The proportions of participants missing each variable will be summarised in each arm and at each time point. The numbers, proportions and reasons for withdrawal from trial will be summarised by treatment arm. The distribution of times between randomisation and withdrawal from follow-up will be summarised using a histogram.

The baseline characteristics of those missing follow-up will be compared to those with complete follow-up. The relationship between baseline characteristics and missing data will be investigated graphically.

Adverse event reporting

Adverse events (AE) and serious adverse events (SAE) will be summarised by treatment arm.

Assessment of outcome measures (unblinding)

Outcome assessors (research workers) and the senior trial statistician are being kept blind to treatment allocation.

Descriptive statistics for outcome measures

Each of the outcome measures will be described by treatment group. Means and standard deviations or medians and interquartile ranges will be used for continuous variables; Q-Q plots will be used to assess whether the distribution of a variable is normal. Frequencies and proportions will be used to describe categorical variables.

Data analysis plan – Inferential analysis

Main analysis of treatment differences

All analyses will use the intention-to-treat population unless otherwise specified (i.e. all randomised participants included according to allocated randomised trial arm irrespective of treatment received). The formal statistical analyses will use the following sequence of contrasts for the primary and all secondary outcomes:

Group versus Intensive

Independent versus TAU

Group+Intensive vs TAU

Estimates of differences between groups and associated 95% confidence intervals will be reported.

The significance level will be 5% (two-sided) for all outcomes. No adjustment will be made for the multiple contrasts. Sensitivity analyses will be used to assess the robustness of conclusions to non-ignorable missing outcome data.

The senior trial statistician will remain blind until the main analyses are completed. Any analyses that cannot be performed blind will be done at the end of the final analysis in order to preserve blinding for as long as possible.

Analysis of primary outcomes

The analysis population will include all patients and intended to recover ITT estimates of effects. Patients with missing baseline measurements will be included using the mean imputation and dummy variable approach (White and Horton [7]).

The primary outcome is Goal Attainment Scale (GAS) T-score at 15 weeks post-randomisation (post-therapy assessment). The Goal Attainment Scale T-score is a standardized measure which will be calculated from the goals set in the GAS weighted by importance and difficulty using a formula shown below as specified in the GAS Practical Guide [6] . If goals are set in an unbiased fashion, we would expect this measure to be normally distributed [6].

GAS Scoring for primary outcome

We will calculate the GAS weighted T-score by applying the following formula (as specified in the GAS Practical Guide [6]):

$$\frac{10\sum(w_i x_i)}{[(1-\rho)\sum {w_i}^2 + \rho\left(\sum(w_i)^2\right)]^{1/2}}$$

Where:

w_i = the weight assigned to the *i*th goal

x_i = the numerical value achieved (between -2 and +2)

ρ = the expected correlation of the goal scales, we will use 0.3 as recommended by Kirusek and Sherman [10] as this is most common approximation.

A linear mixed model will be used to estimate difference in mean GAS T-score at 15 weeks between arms. Linear mixed modelling utilises all available information (including the 40-week follow-up assessment), leading to more precise estimates of the treatment effect. This technique will allow the simultaneous modelling of the repeated outcome time points.

In such models the outcome variable measured at the post treatment time points (here post-therapy and follow up) features as the dependent variable, with treatment arm, time (post-therapy or follow

up), a time by treatment arm interaction, baseline GAS T-score and the randomization stratifier (Site) included as independent variables and a random patient-specific intercept. We will additionally include period as a covariate as described below in section 3.5.

All arms will be modelled simultaneously, and post-estimation commands used to obtain separate estimates for each of the specified contrasts at the post-therapy and follow up timepoints.

Analysis of secondary outcomes

Treatment effects on all secondary outcomes that were measured repeatedly over the follow-up period will lend themselves to the same analysis as described above for the primary outcome. These secondary outcomes will be assessed using similar modelling techniques, employing generalisations to non-normal data where necessary or transformation of the outcome variable. Where secondary outcomes are only measured once over the follow up period, generalised linear models will be used.

Time points

The primary and secondary analyses use all available data from post-therapy (15 weeks post-randomisation), post therapy (39 weeks post-randomisation) and baseline assessments. Deviations of measurements from planned time points will be summarised by treatment group.

Stratification and clustering

Randomisation is stratified by site; therefore, this variable will be included as a factor in the modelling process as detailed in 3.2. There is limited interaction expected between participants in the group therapy arm and so we will not account for any clustering effects within the group arm; however, a sensitivity analysis may be conducted to test this assumption if data allows.

We will additionally adjust for period as a factor in all of the models. Period will be a binary variable of whether participant was randomised before or when the Independent and TAU arms were dropped (following the interim analysis). This is primarily as randomised participants in the two periods may not be exchangeable under their respective null hypotheses, as well as for reasons of potential bias as outlined further in Section 3.12.

Missing items in scales and subscales

The number (%) with complete data will be reported. Where present, missing value guidance provided for scales will be used. Where this is not possible, scales will be pro-rated for an individual if 20% or fewer items are missing. For example, in a scale with 10 items, prorating will be applied to individuals with 1 or 2 items missing. The average value for the 8 or 9 complete items will be calculated for that individual and used to replace the missing values. The scale score will be calculated based on the complete values and these replacements.

Missing baseline data

Missing baseline data should not be an issue for the primary analysis. Some extensions to this analysis may use other baseline variables; if these contain missing data, the number with complete data will be reported and they will be imputed using a method suitable to the variable as per the recommendations of White and Thompson [7] .

Missing outcome data

Analyses will be undertaken assuming outcomes are missing-at-random and using all available data. This allows drop-out to be related to treatment group, stratification factors, period and baseline severity. Sensitivity analyses will be carried out to assess (1) the association of drop-out with baseline demographic variables and their inclusion as additional covariates (2) last observation carried forward as an alternative assumption.

Method for handling multiple comparisons

No formal adjustment of p-values for multiple testing or as a consequence of multiple comparisons will be made, However, care will be given to the interpretation of inference for the numerous secondary outcomes and with respect to multiple contrasts. The absence of any correction will be reported.

Method for handling non-compliance (per protocol/CACE analyses)

In addition to the primary intention-to-treat analysis the effect of actually receiving treatment will be estimated. Compliance will be a continuous measure of the number of valid therapy sessions attended. Local average or complier average treatment effects per hour of active CR compared to TAU will be estimated. On an assumption of a common effect per hour of active CR, evidence for differences in treatment mode not explained by treatment hours of active CR will be presented.

Model assumption checks

The models assume normally distributed outcomes; residual plots will be checked for normality and outliers and if substantial departures occur, transformations will be applied.

Sensitivity to interim analysis

Sensitivity analyses will be carried out to assess the impact of dropping the Independent and TAU arms consequent to the interim analysis. There are at least 3 mechanisms by which this may affect these results:

The interim analysis was partially based on the primary outcome. This could theoretically lead to bias in the treatment estimate as the primary outcome at the interim for the dropped arms may have been lower than the continuing arms by chance.

Participants interested in participating and consequently recruited to the 2 arm version of the trial may differ in characteristics from than those that were interested in participating and were recruited to the 4 arm version (e.g. the latter included a Treatment as usual arm which may have dissuaded participants from participating).

There may be a "time of recruitment" effect as comparisons between continuing arms and dropped arms will include non-contemporaneous participants.

We will report simulation results under two scenarios:

Simulate all 4 arms under the global null hypothesis to the interim point

Select those scenarios where the 2 arms selected are the best two according to the primary outcome. Whilst this does not reflect the decision criteria that was used at the interim analysis, this should provide an upper bound on the degree of bias. Progress simulation to end of recruitment. Estimate effects as per SAP

Continue simulation to end of recruitment for all scenarios. Estimate effects as per SAP

Compare distributions of the average treatment effect estimates under a and b

As above but simulated under parameters of the naïve analysis. This is equivalent to simulating under a "global alternative hypothesis".

Sensitivity to circumstances surrounding COVID-19 pandemic

We believe it is unlikely that the effects of the pandemic/lockdown (which has led to changes in the way outcomes were recorded) will introduce bias for the comparison of contemporaneous arms, as any effects are unlikely to differ by arm.

We expect the pandemic/lockdown will affect comparisons between the arms that were dropped at interim (TAU and Independent CR) and the ongoing arms (Intensive and Group CR). This is because the ongoing arms will have participants whose follow up measures were affected by the pandemic and the dropped arms won't. However, as described in Section 3.12, there are already other potential sources of bias in these comparisons. We are already attempting to account for these by including period (before or after interim) as a covariate in the analysis and assessing further impact by carrying out simulations. Any potential for bias in these comparisons due to the pandemic will be assessed in the same way.

Sensitivity to visit windows

As per Section 1.6, visit windows are defined in the protocol for the post-therapy and 6-month post-therapy follow up assessments anchored to date of randomisation. However, where visits have occurred outside these windows, data has still been recorded and will be used for the primary analysis. A sensitivity analysis will be carried out (on the primary outcome only) using only data from visits that occurred within the visit windows.

Planned subgroup analyses

No subgroup analyses are planned for the primary paper. The study is not powered to investigate interaction effects.

Exploratory analyses

This analysis plan does not cover exploratory analyses. Exploratory mediator and moderator analyses may be performed after the primary trial data analysis.

Software

Data management: An online data collection system for clinical trials (MACRO; InferMed Ltd) will be used. This is hosted on a dedicated server at KCL and managed by the KCTU. The KCTU will extract data periodically as needed and provide these in comma separated (.csv) format.

Statistical analysis: Stata 15 [8] will be used for data description and inferential analyses. R [9] may additionally be used for data description and production of graphs, tables and reports.

SCHEDULE OF ASSESSMENTS AND MEASURES

Table 1: Schedule of assessments and measures

#	Form	Screening assessment	Pre- therapy assessment	Post-therapy assessment	6-month Follow-up	Type	Administered to	Administration time
MAIN DATABASE								
1	**Eligibility form**	x				Examination of case notes	Completed by RW	n/a (completed by RW)
2	**Mini International Neuropsychiatric Interview (MINI)**	x			x	Diagnostic checklist	Participant/Notes	**5 min** (3 sections only)
3	**Registration and Demographics Form**	x				Questionnaire	Participant	**2min**
4	**Nottingham Onset Schedule (NOS)**	x				Semi-structured interview	Participant *(&case notes)*	**5min**
5	**Social and Occupational Functioning Assessment Scale (SOFAS)**		x	x	x	Rating scale	Completed by RW	n/a (completed by RW)
6	**Adapted Substance Misuse Questionnaire**		x	x	x	Questionnaire	Participant	**5min**
7	**Positive and Negative Symptom Scale (PANSS)**		x	x	x	Semi-structured interview	Participant	**20-30min**
8	**Clinical Assessment Interview for Negative Symptoms (CAINS)**		x	x	x	Semi-structured interview	Participant	**20min**
9	**The Time Use Survey**		x	x	x	Questionnaire	Participant	**5-10min**
10	**The Client Service Receipt Inventory**		x	x	x	Questionnaire	Participant	**10-15min**

11	**EQ-5D-5L**		x	x	x	Questionnaire	Self-report	**2min**
12	**Rosenberg Self Esteem Scale**		x	x	x	Questionnaire	Self-report	**2 min**
13	**CANTAB - Motor Screening Task: MOT**		x	x	x	Computerised test	Participant	**7min**
14	**CANTAB - Reaction Time: RTI**		x	x	x	Computerised test	Participant	**8min**
15	**CANTAB - Paired Associates Learning (PAL)**		x	x	x	Computerised test	Participant	**10min**
16	**CANTAB – Spatial Working Memory**		x	x	x	Computerised test	Participant	**5min**
17	**CANTAB - One-touch Stockings of Cambridge: OTS**		x	x	x	Computerised test	Participant	**6min**
18	**CANTAB - Emotion Recognition Test (ERT)**		x	x	x	Computerised test	Participant	**10min**
19	**CANTAB - Rapid Visual Information Processing: RVP**		x	x	x	Computerised test	Participant	**10min**
20	**CANTAB - Attention Switching Task: AST**		x	x	x	Computerised test	Participant	**10min**
21	**Computerised Wisconsin Card Sorting Task (WCST)**		x	x	x	Computerised test	Participant	**10min**
22	**Digit Span task**		x	x	x	Cognitive test	Participant	**5min**
23	**Rey Auditory Verbal Learning Test**		x	x	x	Cognitive test	Participant	**10-15min** (30min delay)
24	**Rey Osterrieth Complex Figure**		x	x	x	Cognitive test	Participant	**10-15min** (copy/trial, 3min later, 30 min later)
25	**The Wechsler Test of Adult Reading**		x			Cognitive test	Participant	**10min**
26	**Wechsler Abbreviated Scale of Intelligence II**		x			Cognitive test	Participant	**20-30min**
27	**Frontal Systems Behavior Scale**		x	x	x	Questionnaire	Completed by a relative/carer	n/a

								(completed by relative/carer)
28	**IoPPN Narrative Metacognition Task**		2x Completed twice (after Rey & SWM)	2x Completed twice (after Rey & SWM)	x	Semi-structured interview	Participant	**2min**
29	**Metacognition Assessment Scale for the Rey Osterrieth Complex Figure**		x	x	x	Questionnaire	Self-report	**1min**
30	**Self-appraisal Scale**		2x Completed twice (after Rey & SWM)	2x Completed twice (after Rey & SWM)	x	Questionnaire	Self-report	**1min**
31	**Measure of Insight into Cognition - Self Report (MIC-SR)**		x	x	x	Questionnaire	Self-report	**1min**
32	**Goal Attainment Scale**		x (goals set)	x (rated)	x (rated)	Rating scale	Participant	**10-15min (?)**
33	**Adverse events**	When required				n/a	Completed by RW	n/a
34	**Withdrawal form**	When required				n/a	Completed by RW	n/a
	THERAPY DATABASE and Other measures							
35	**Therapy audit**	Ongoing				n/a	Completed by the therapist	n/a
Other 36	**Satisfaction measure**			x		Questionnaire	Self-report/Online survey	**10min**
Other 37	**Working alliance inventory**			x		Questionnaire	Self-report/Online survey	**10min**

REFERENCE LIST

[1] EMEA, “ICH Topic E 9 Statistical Principles for Clinical Trials.” 1998.

[2] K. F. Schulz, D. G. Altman, and D. Moher, “CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials,” *BMJ*, vol. 340, p. c332, Mar. 2010.

[3] B. Freidlin, E. L. Korn, R. Gray, and A. Martin, “Multi-Arm Clinical Trials of New Agents: Some Design Considerations,” *Clin. Cancer Res.*, vol. 14, no. 14, pp. 4368–4371, Jul. 2008.

[4] S. J. Pocock, *Clinical trials: a practical approach*. Chichester: John Wiley and Sons, 1995.

[5] S. F. Assmann, S. J. Pocock, L. E. Enos, and L. E. Kasten, “Subgroup analysis and other (mis)uses of baseline data in clinical trials,” *Lancet Lond. Engl.*, vol. 355, no. 9209, pp. 1064–1069, Mar. 2000.

[6] L. Turner-Stokes, “Goal attainment scaling (GAS) in rehabilitation: a practical guide,” *Clin. Rehabil.*, vol. 23, no. 4, pp. 362–370, Apr. 2009.

[7] I. R. White and S. G. Thompson, “Adjusting for partially missing baseline measurements in randomized trials,” *Stat. Med.*, vol. 24, no. 7, pp. 993–1007, Apr. 2005.

[8] StataCorp, *Stata Statistical Software: Release 15.* College Station, TX: StataCorp LLC, 2017.

[9] R Development Core Team, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008.

[10] T. J. Kiresuk and R. E. Sherman, “Goal attainment scaling: A general method for evaluating comprehensive community mental health programs,” *Community Ment. Health J.*, vol. 4, no. 6, pp. 443–453, Dec. 1968.