CognIFiEd Statistical Analysis Plan

A pilot randomised controlled trial of culturally adapted compensatory cognitive training for young adults with first episode psychosis in Nigeria

Administrative item	Details
Principal Investigator	Professor Abiodun O. Adewuya, Lagos State University College of Medicine / Lagos State University Teaching Hospital
Trial registration	ISRCTN44794154
Funding	Wellcome Trust, grant 226720/Z/22/Z
Protocol source	Adewuya et al. BMJ Open 2026;16:e115815. doi:10.1136/bmjopen-2025-115815. Published 12 March 2026.

Trial design and objectives

CognIFiEd is a multicentre, assessor-blinded, individually randomised, parallel-group pilot trial. The target is 180 participants, allocated 1:1 to culturally adapted compensatory cognitive training (CCT) or Enhanced Recreational Therapy (ERT). Both arms receive 12 weekly group sessions lasting 60–90 minutes, generally in groups of six to eight, alongside usual clinical care. Therapy-group clustering arises from delivery; hospitals and therapy groups are not the units of randomisation.

Sites are LASUTH, Federal Neuro-Psychiatric Hospital Yaba and Neuropsychiatric Hospital Aro, Abeokuta. Randomisation uses variable blocks of four, six and eight, stratified by site and age band (18–21, 22–26 and 27–30 years). Eligibility includes age 18–30, DSM-5 psychotic disorder with first episode within five years, clinical stability and objectively assessed cognitive impairment, with the remaining criteria as stated in the protocol.

Primary objectives concern recruitment, retention, CCT adherence, participant satisfaction and delivery fidelity. Secondary objectives estimate cognitive, functional, symptom, disability and quality-of-life differences. Social cognition and economic outcomes are exploratory. The sample size is fixed by the pilot design; no retrospective power calculation or significance-based sample-size revision will be used. Estimates and uncertainty will inform a future definitive trial, alongside external evidence.

Analysis timing and general principles

The final analysis will occur after the last participant completes the 12-month follow-up or reaches their final documented study disposition, outstanding queries are resolved and the database is locked. Interim operational summaries will identify their cutoff date and include only participants whose relevant assessment is due when estimating interim retention. No efficacy stopping boundary is introduced by this SAP. Safety oversight continues independently under trial procedures.

All estimates will have two-sided 95% confidence intervals where estimable. Exploratory clinical P values will be two-sided and reported exactly to three decimals, with values below 0.001 shown as $P<0.001$. No multiplicity-adjusted confirmatory claims will be made. Primary feasibility outcomes will not be subjected to efficacy-style null-hypothesis tests. Baseline significance tests and stepwise covariate selection will not be used.

Continuous measures will be summarised by n, mean, SD, median, quartiles and range as appropriate. Categorical measures will show counts, explicit denominators and percentages; missing values will be shown separately. Proportions will use Wilson 95% intervals unless a clustering-aware method is specified below. Analyses will use R version 4.3.0 or later, or Stata version 18 or later; exact software and package versions, seeds and scripts will be archived.

Analysis populations and estimands

The screening population comprises unique individuals entered in the screening log, with eligibility, refusal and randomisation disposition reconciled. The intention-to-treat population includes every randomised participant in their assigned arm, regardless of initiation, adherence, crossover or later eligibility concerns. Consent restrictions on use of data will be respected and disclosed. Participants

without measured outcomes remain in the flow diagram and feasibility denominators; the model cannot create information where all outcomes are absent.
The per-protocol sensitivity population comprises participants attending at least nine of the twelve assigned sessions in either arm. The protocol does not prescribe further exclusions, so this plan does not silently add them. Major deviations will be tabulated separately. A descriptive exposure population includes participants attending at least one session; safety will primarily be reported for all randomised participants by allocation, with exposure summaries supplementary.
For feasibility, estimands are the proportions or means defined in Section 6 over the specified screening, randomised, participant-response or delivered-session populations. The preliminary clinical estimand is the adjusted CCT-minus-ERT mean difference at 12 months in the randomised population. For nonfatal intercurrent events, a treatment-policy strategy applies: observations after nonadherence, discontinuation, medication changes, admission or crossover are retained where permitted. These events are described and are not covariate-adjusted away in the main model.
Death makes subsequent clinical scores undefined. Deaths will be reported separately and will not be assigned or imputed ordinary clinical scores. If deaths occur, clinical contrasts will be explicitly interpreted as contrasts among participants alive at assessment under each allocation, rather than an unconditional all-randomised treatment effect. A supplementary worst-rank analysis at 12 months will rank deaths below surviving outcomes for BACS and UPSA-N, stratified by site and age where feasible. This changes the estimand and will be labelled supplementary; it will not be substituted for the planned mean comparison.

Assessment timing and scoring
Baseline is the last valid pre-randomisation assessment within 14 days before randomisation. Follow-up targets are calendar dates three, six and twelve months after randomisation, with protocol windows of ±14, ±14 and ±28 days respectively. Window endpoints are inclusive. If several assessments exist in a window, select the one nearest the target, using the earlier assessment for ties. Corrections to the same encounter use the adjudicated final record. Actual assessment dates and deviations are retained. Main longitudinal analyses will include valid assessments recorded for the intended visit even when outside its window, provided they are not duplicate visits or incorrectly labelled encounters. A sensitivity analysis will restrict to protocol windows. A visit will not contribute twice. Ambiguous visit assignments must be resolved from source records before lock, without considering outcome magnitude or treatment differences.

Measure	Schedule	Scoring and direction
BACS composite	Baseline, 3, 6, 12 months	Composite z score; higher is better. Freeze the existing normative reference and algorithm; do not restandardise separately by arm or follow-up.
UPSA-N	Baseline, 3, 6, 12 months	Use the adapted instrument's approved total-score algorithm; higher is better.
BPRS-18	Baseline, 3, 6, 12 months	Sum valid items using the administered rating convention; higher is worse. Record whether items are coded 0–6 or 1–7.
WHODAS 2.0 twelve items	Baseline, 3, 6, 12 months	Use simple scoring, 0–4 per item, sum 0–48; recode 1–5 entries if needed. Higher is worse. If the operative manual specifies complex scoring, reconcile before approval.
MANSA	Baseline, 3, 6, 12 months	Mean of the twelve subjective satisfaction items, 1–7; higher is better. Objective items are not included in this mean.
WHOQOL-BREF global items	Baseline, 3, 6, 12 months	Report overall quality of life and general health satisfaction separately, each 1–5; no invented global sum.

CSQ-8	3 months	Eight items scored 1–4 in the satisfaction direction; sum 8–32. Higher is better.
SCSQ-N	Baseline, 12 months	Use the adapted scoring manual and prespecified total/subscales; confirm direction for each before approval.
EQ-5D-5L and CSRI	Baseline, 12 months	Utility, descriptive health states, VAS if collected, and resource-use categories; economic rules in Section 12.

The trial statistician will sign a scoring dictionary naming instrument versions, variable mappings, permissible ranges, reverse-coded items, normative references and item-missing rules. Published or approved instrument rules take precedence. In their absence, no ad hoc proration will be introduced: incomplete totals/composites are missing. UPSA-N, SCSQ-N, the fidelity checklist and the CSRI recall period require the operative manuals or forms before final approval.

Primary feasibility and acceptability analyses

Recruitment proportion is the number of eligible individuals who consent and are randomised divided by all unique individuals confirmed eligible during recruitment, including eligible refusals. Screening yield (randomised/all screened) and consent rate will be reported separately. Unknown eligibility will not be counted as eligible or ineligible; its frequency and reasons will be shown, with a conservative recruitment sensitivity denominator that includes unknown eligibility. Screening counts must not substitute for the eligible denominator.

Recruitment speed will be described by site and month using actual opening and closing dates, participants per active site-month, and elapsed time to the target of 180. The protocol refers to a nine-month recruitment period but gives 1 February to 30 November 2026, which spans ten calendar months. Both the stated target and actual dates will be disclosed; the inconsistency will not be resolved retrospectively by selecting the more favourable benchmark.

Retention at each follow-up is the number of randomised participants completing a follow-up outcome visit divided by all randomised participants. For reproducibility, completion means at least one valid protocol clinical outcome score at that visit; contact alone does not count. Report additionally completion of BACS, UPSA-N, both measures, the full scheduled battery and completion within the protocol window. The 12-month retention proportion is the principal retention measure. Death, withdrawal and loss to follow-up remain denominator events and are shown separately.

CCT adherence is the proportion of all participants allocated to CCT attending at least nine distinct scheduled sessions. Equivalent ERT attendance will be summarised descriptively. Report initiation, zero attendance, number of sessions, proportions attending ≥9 and all 12, and discontinuation reasons. One curriculum session counts at most once; a documented make-up covering the same scheduled content may count once. Cancelled sessions are delivery failures and do not reduce the denominator of twelve. Unresolved attendance records will count as not documented attended, with a sensitivity analysis treating them as attended.

CSQ-8 acceptability is the mean valid total at three months among CCT respondents, with SD, range, median and a 95% CI. Report the number responding out of all allocated participants and out of those initiating treatment, including reasons for missingness. ERT results are supplementary. The respondent mean is not an all-randomised satisfaction estimand. Complete-case CSQ-8 is primary; bounds assigning missing total scores 8 and 32 will show how nonresponse could affect interpretation.

For recruitment, report overall and site-specific Wilson intervals. For retention and adherence, show Wilson intervals for transparency and a therapy-group cluster-bootstrap interval where group assignment can be reconstructed for all participants. Resample groups within site and arm, retaining their participants, for 2,000 replicates. For overall retention, combine arm/site estimates using fixed original sample weights. For CSQ-8 use the same group-resampling principle within CCT, alongside a conventional t interval. Nonstarters retain planned group membership; participants without planned membership remain individually identifiable units and their number will be disclosed. If too few independent groups or sparse strata prevent a defensible bootstrap, report that limitation and do not manufacture a clustering-adjusted interval.

Fidelity and progression criteria

The fidelity sample is the protocol's randomly selected 20% of audio-recorded CCT sessions. Preserve the sampling frame, selection seed, eligible recordings, selected sessions and reasons any selected session could not be coded. Report the percentage of all delivered sessions recorded, because a random sample of available recordings does not guarantee representation of unrecorded sessions.

Using the approved checklist, calculate adherence as achieved applicable content points divided by possible applicable content points, expressed as a percentage. Competence will be scored separately according to its manual; any conversion to a percentage and competency pass threshold must be documented before approval. Do not treat a raw Likert mean as a percentage without an agreed conversion. The principal fidelity estimate is the mean session-level percentage, with CIs resampling therapy groups. Report adherence and competence separately by site, group and facilitator, without identifying individual staff publicly. Where duplicate ratings exist, report the absolute-agreement ICC for continuous scores and agreement for categorical items, with uncertainty.

Domain	Green	Amber	Red
Recruitment	≥60%	≥40% to <60%	<40%
Retention at 12 months	≥70%	≥50% to <70%	<50%
CCT attendance ≥9 sessions	≥75%	≥60% to <75%	<60%
Mean CCT CSQ-8	≥24	≥20 to <24	<20
CCT fidelity	≥80%	≥65% to <80%	<65%

These continuous interval boundaries operationalise the published rounded categories without gaps. The protocol's analysis section says that lower confidence limits will be evaluated against the thresholds, whereas Table 3 presents outcome-value thresholds. Accordingly, report two classifications: the observed point-estimate band and the band containing its lower 95% confidence limit. The latter is a conservative evidence assessment, not an undisclosed replacement of the published table. Where a clustering-adjusted interval is available, it governs the lower-bound assessment; otherwise identify the conventional interval and its limitations.

For fidelity, green requires both adherence and competence to meet the 80% benchmark on valid comparable percentage scales. Lower-bound support requires both lower limits to meet that benchmark. If the competence scale cannot support this interpretation, competence is reported separately and no overall fidelity-green designation is assigned until the criterion is reconciled in an approved amendment. Unavailable outcomes receive an unevaluable status, not green.

The TSC will consider the entire profile, precision, safety, remediable delivery barriers and qualitative findings. No arithmetic composite score or automatic majority vote is introduced. Qualitative findings will distinguish remediable barriers from fundamental barriers. Any recommendation inconsistent with a numerical band will state the rationale and proposed modifications. Clinical P values will not determine progression.

Preliminary clinical outcome models

Each repeated continuous outcome will be modelled separately using post-baseline observations at three, six and twelve months. Fixed effects are allocated treatment, categorical visit, treatment-by-visit interaction, the corresponding baseline outcome, site and randomisation age stratum. A baseline-by-visit interaction will allow the baseline association to vary over follow-up. Random intercepts for participant and therapy group will account for repeated measurement and delivery clustering in both arms. Site is fixed because there are only three sites.

In notation, $Y_{igt} = \beta_0 + \beta_1 T_i + \beta_2 Visit6 + \beta_3 Visit12 + \beta_4 T_i \times Visit6 + \beta_5 T_i \times Visit12 + \gamma_t Y_{i0} + \delta Site_i + \eta AgeStratum_i + u_g + b_i + \varepsilon_{igt}$.

The three-month visit is the reference.

The principal clinical contrast is $\beta_1 + \beta_5$, the adjusted CCT-minus-ERT difference at twelve months. Three- and six-month differences are supplementary.

Report estimates, standard errors, 95% CIs and exploratory P values. Higher scores favour CCT for BACS, UPSA-N and quality of life; lower scores favour CCT for BPRS and WHODAS.

Use restricted maximum likelihood and Kenward–Roger degrees of freedom where supported, otherwise Satterthwaite inference. Report therapy-group and participant variance components, convergence diagnostics, the number and sizes of groups, and the estimated group ICC with its uncertainty where estimable. Retain planned first therapy group for participants who never attend; document transfers and cross-group attendance. Group identifiers must be unique across sites and arms.

If the main model is singular or fails to converge, first check data, coding and optimiser settings. A variance component estimated at zero may be removed with disclosure. If a stable mixed model remains unavailable, use visit-specific baseline-adjusted linear models with site and age stratum, and CR2 small-sample corrected group-robust standard errors with Satterthwaite degrees of freedom. Do not remove nonzero group clustering solely to obtain smaller P values. Any further simplification will be documented and inference presented cautiously.

Inspect residual distributions, fitted-versus-residual plots and influential observations. Valid outliers are retained; errors are corrected through queries. Strong skewness or ceiling/floor effects will prompt supplementary robust or bootstrap analyses on the original scale, not a data-driven replacement of the specified outcome. No covariate will be selected because of baseline significance or its effect on the treatment P value.

For BACS and UPSA-N, report adjusted mean differences divided by the pooled within-arm baseline SD as baseline-standardised Cohen-type effect sizes. Report the denominator and its value. CIs will be obtained by group bootstrap, re-estimating both numerator and denominator where feasible; an interval treating the SD as fixed will be clearly identified if used. A zero baseline SD makes standardisation undefined. Other outcomes will emphasise their original scales.

SCSQ-N is measured only at baseline and twelve months. Fit a baseline-adjusted linear model with treatment, site and age stratum and a therapy-group random intercept, using the same small-sample inference and fallback principles. Analyse only the total and subscales supported by the approved adapted scoring scheme.

Missing data and sensitivity analyses

Report missing visits and missing items by arm, site and time, distinguishing death, withdrawn data permission, inability to assess, loss to follow-up and administrative omission. Describe associations with observed baseline and previous outcomes without claiming that missing-at-random can be established empirically. The main repeated-measures model uses available outcomes under a missing-at-random assumption conditional on included variables. Last observation carried forward and mean substitution will not be used.

The protocol requires chained-equations multiple imputation when missing data exceed 15% for key outcomes. Trigger this sensitivity analysis when BACS or UPSA-N is missing for more than 15% of participants alive at a scheduled follow-up, and apply the same approach to other repeatedly measured outcomes exceeding this threshold. Report missingness among all randomised participants separately. If a baseline outcome or required covariate is missing, imputation will also be used as a sensitivity analysis even below the threshold, because complete-covariate modelling could exclude participants.

Generate at least the protocol-specified 20 imputed datasets; increase to 50 or more if Monte Carlo error is material. The imputation model will include treatment, site, age stratum, all scheduled measurements of the analysed outcome, baseline demographic and clinical predictors, observed attendance and documented missingness predictors. Use appropriate continuous, binary or ordinal conditional models and preserve score bounds. Imputation must accommodate therapy-group clustering and treatment-by-time relationships, using multilevel chained equations or a documented compatible alternative. Pool estimates with Rubin’s rules. Inspect convergence, distributions and Monte Carlo stability. Do not impute values after death or data that cannot lawfully be retained.

For BACS and UPSA-N, evaluate departures from missing-at-random through delta-adjusted imputation of missing follow-up outcomes. Use shifts of 0, −0.25 and −0.50 pooled baseline SD in each arm, considering combinations that disadvantage either arm. Negative shifts represent poorer unobserved outcomes. Apply shifts only to imputed outcomes and respect possible score limits. Report changes in

estimated differences and uncertainty, rather than seeking a favourable significance threshold. If sparse data prevent the analysis, explain why.
Additional sensitivity analyses comprise the ≥9-session per-protocol population, assessments within protocol windows, and twelve-month complete-case ANCOVA with group-aware inference. Per-protocol differences are potentially confounded by adherence and are not causal effects of treatment receipt. Where transfers are material and records support it, examine actual principal group membership as a supplementary clustering specification. No unplanned mediation or complier-average causal analysis is designated confirmatory.

Exploratory effect modification
Assess baseline cognitive severity using continuous baseline BACS and age using the three randomisation strata. Add the appropriate modifier-by-treatment-by-visit terms and all required lower-order terms. For the BACS model, baseline BACS already appears as an adjustment covariate and will not be duplicated. Present twelve-month interaction estimates or an omnibus interaction test, CIs and descriptive contrasts. Do not derive severity cutoffs from observed outcomes. Small cells or unstable models will be reported descriptively. These analyses are hypothesis-generating and carry no multiplicity-controlled claim.

Safety deviations and assessor blinding
Using the existing safety definitions and adjudicated event log, report deaths, serious adverse events, psychiatric admissions, symptom deterioration and other recorded adverse events from randomisation to twelve months or last observed contact. Present participants with at least one event and total events by allocated arm, seriousness, severity and assessed relatedness. Recurrent events and participants are separate counts. Describe exposure and follow-up duration. No event is inferred from absence of a completed form; no-event status requires a documented record.
Summarise protocol deviations including eligibility errors, allocation errors, intervention crossover, contamination, session noncompletion, visit-window deviations and assessor unblinding. All classifications will be finalised before comparative analysis. Report the twelve-month assessor allocation guesses as correct, incorrect or unknown, with denominators and 95% CIs; show results by actual arm. Correct guesses alone cannot establish failed masking because they may reflect perceived response. Unblinded assessments remain in the main analysis and may be examined separately descriptively.

Exploratory economic evaluation
The primary perspective is the public healthcare provider; the secondary perspective is societal. The horizon is twelve months from randomisation. Costs will be expressed in 2026 Nigerian Naira. International-dollar conversion will use a documented purchasing-power-parity source and reference year, not a nominal exchange rate labelled as PPP. No discounting is needed within a one-year horizon. Micro-costing will include intervention training, staff delivery and supervision time, materials, facilities and relevant infrastructure. Separate research-only activities from replicable service delivery. Allocate group costs across intended group capacity for a standardised delivery-cost estimate and across actual recipients for observed cost per recipient, explaining each denominator. Training and reusable capital will be allocated using documented useful lives and expected service volumes; alternative allocations will be sensitivity analyses. Usual-care and other service use will be valued with documented Nigerian unit-cost sources. Societal analyses will additionally include participant transport, informal care and productivity losses without double counting.
The CSRI recall period must be established from the actual forms. Baseline recalled use is a pre-randomisation covariate, not a post-randomisation cost. If the twelve-month CSRI does not cover the full follow-up year, full-year costs will not be asserted without an explicit extrapolation assumption. Report observed-period cost-consequence results and test any proposed annualisation separately.
EQ-5D-5L health states will be converted to utilities with one named, published value set fixed before approval. Whether an appropriate Nigerian value set is available must be verified; if not, justify the chosen external tariff and, where feasible, an alternative tariff sensitivity analysis. No tariff is silently

selected in this draft. With baseline and twelve-month utilities only, QALYs will use trapezoidal area under the curve with linear interpolation and actual elapsed time in years, capped at the twelve-month horizon. For deaths, assign utility zero from the date of death and interpolate from the preceding observed utility to zero at death. Disclose the limitation of sparse utility measurement.
Report mean cost and QALY by arm, incremental cost and incremental QALYs, with baseline utility adjustment for QALYs and baseline resource-use adjustment for costs where compatible. Models will account for site, age stratum and therapy group. Present a cost-consequence table first. Report incremental cost per QALY only when interpretable, clearly distinguishing dominance and dominated results from ratios with negative denominators.
Use 2,000 group-bootstrap replicates stratified by site and arm to quantify uncertainty jointly in costs and QALYs. Present a cost-effectiveness plane and acceptability curve over a justified range of willingness-to-pay values in Naira per QALY. Do not imply an established Nigerian threshold without a source. For extensive missing economic data, use compatible joint imputation of utilities and cost components or report a clearly limited complete-case evaluation if a defensible model cannot be fitted. Economic results remain exploratory and will not establish affordability or cost effectiveness conclusively.

Reproducibility and quality assurance
Before lock, reconcile screening, consent, allocation, session, group, outcome and safety logs; verify duplicate IDs, dates, permitted ranges and scoring. Maintain a query trail and record frozen extracts, dataset checksums, code versions and analysis seeds. The allocation key will be controlled separately where blinding remains feasible. Independent code or output review will verify population counts, principal feasibility denominators, scoring and twelve-month treatment contrasts. Only de-identified data will enter public outputs.
The report will follow pilot-trial CONSORT principles, separating feasibility conclusions from uncertain clinical signals. The SAP, approved amendments, analysis code where sharing is permitted, and a data dictionary will accompany appropriate dissemination. Reporting will acknowledge Wellcome Trust grant 226720/Z/22/Z.

Planned tables figures and listings

Output	Required content
Figure 1	CONSORT flow by allocation, follow-up and reasons for exclusion or loss
Table 1	Baseline characteristics by allocation, with denominators and missingness
Table 2	Recruitment by site/month and eligible denominator reconciliation
Table 3	Retention, attendance, satisfaction and fidelity with 95% CIs and progression bands
Table 4	Observed clinical outcomes by arm and visit
Table 5	Adjusted differences at 3, 6 and 12 months, with primary emphasis on 12 months
Table 6	Missingness, sensitivity analyses and exploratory interactions
Table 7	Adverse events, deviations, contamination and assessor guesses
Table 8	Costs, resource use, utilities, QALYs and incremental estimates
Figures 2 and 3	Clinical trajectories with CIs; economic uncertainty plots where estimable
Supplementary listings	Scoring dictionary, group membership rules, deviations, amendments and model diagnostics

Items to resolve before final approval
The following are limited technical and governance completions, not requests to redesign the trial: identify the trial statistician and operative protocol version; complete the outcome-data-access declaration; supply the adapted scoring manuals, fidelity checklist and CSRI form; agree the documented progression interpretation; select and justify the EQ-5D tariff and PPP source; verify planned therapy-group identifiers for nonstarters; and enter the actual approval and upload dates. These fields must be resolved before this draft is described as the final approved registry SAP.

REFERENCES

1. Adewuya AO, Okewole A, Ola B, et al. CognIFied: protocol for a pilot randomised controlled trial of a culturally adapted, task-shifted compensatory cognitive training intervention for young adults with first-episode psychosis in Nigeria. BMJ Open. 2026;16:e115815. doi:10.1136/bmjopen-2025-115815.
2. Gamble C, Krishan A, Stocken D, et al. Guidelines for the content of statistical analysis plans in clinical trials. JAMA. 2017;318(23):2337–2343. doi:10.1001/jama.2017.18556.
3. International Council for Harmonisation. ICH E9(R1) Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. 2019. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf