Statistical Analysis Plan
Final version 1.2

Study Title	**A study to assess the clinical and cost-effectiveness of aphasia computer treatment versus usual stimulation or attention control long term post stroke**
Short title	**Big CACTUS**
Funding body / Reference	National Institute for Health Research (NIHR) Health Technology Assessment (HTA) / HTA 12/21/01
Sponsor	**University of Sheffield**

Amended by

_____ ___/___/_____

Munya Dimairo *Date*
Research Fellow in Medical Statistics
CTRU, University of Sheffield

Approved by

_____ ___/___/_____

Steven Julious *Date*
Professor of Medical Statistics
University of Sheffield

_____ ___/___/_____

Derrick Bennett *Date*
TSC Senior Statistician
CTSU, University of Oxford

_____ ___/___/_____

Rebecca Palmer *Date*
Chief Investigator
University of Sheffield

Table of Contents

List of Figures

List of Tables

List of abbreviations used

AC	Attention control
ACT NoW	Assessing Communication Therapy in the North West
AE	Adverse event
ANCOVA	Analysis of Covariance
Carer QoL	Carer Quality of Life measure
CaCOAST	Carer Communication Outcomes After Stroke
CAT	Comprehensive Aphasia Test
CC	Complete case
CI	Confidence interval
COAST	Communication Outcomes After Stroke
CRF	Case report form
CSLT	Computerised speech and language therapy
CTRU	Clinical Trial Research Unit
DMEC	Data Monitoring and Ethics Committee
EQ-5D	EuroQol health utility questionnaire
GCP	Good Clinical Practice
HTA	Health Technology Assessment
ICH	International Conference on Harmonisation
ICW	Information carrying words
IQR	Interquartile range
ITT	Intention-to-treat
Lowess	Locally weighted scatterplot smoothing
MCA	Middle cerebral artery
MDC	Mean difference in change
MI	Multiple imputation
MICE	Multiple imputation using chained equations
mITT	Modified Intention-to-Treat
mths	Months
MVI	Mean value imputation
NIHR	National Institute for Health Research
OR	Odds ratio
PI	Principal Investigator
PP	Per-protocol
QALY	Quality adjusted life year
SAE	Serious adverse event
SD	Standard deviation
SE	Standard error
SLT	Speech and language therapy/therapist
TMG	Trial Management Group
TOMS	Therapy Outcome Measures
TSC	Trial Steering Committee
UC	Usual Care
VAS	Visual analogue scale

1 Introduction, study design, and key trial objectives

1.1 Study outline

The Big CACTUS study is a pragmatic, three-arm parallel group, single-blind randomised controlled trial which will compare outcomes for people with persistent aphasia using computerised speech and language therapy (CSLT) at home with those having usual care, or attention control. Detailed background of the study is available in the protocol.

This statistical analysis plan is written to conform with the International Conference on Harmonisation (ICH) topic E9 (ICH, 1998), applicable standard operating procedures from the University of Sheffield Clinical Trials Research Unit (CTRU) and trial documents (Palmer, 2015). The trial will be conducted in accordance with Good Clinical Practice (GCP) in clinical trials (ICH, 2005). The trial is funded by the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) programme (Ref: HTA 12/21/01). The trial registration number is ISRCTN 68798818.

1.2 Primary and secondary objectives

The primary trial objectives are to:

1) Establish whether self-managed CSLT for word finding increases the ability of people with aphasia to use vocabulary of personal importance,
2) Establish whether self-managed CSLT for word finding improves functional communication ability in conversation,
3) Investigate whether patients receiving self-managed CSLT perceive greater changes in social participation in daily activities and quality of life,
4) Establish whether self-managed CSLT is cost effective for persistent aphasia post stroke,
5) Identify whether any effects of the interventions are evident 12 months after therapy has begun.

Secondary trial objectives include:

1. Investigating the generalisation of treatment to finding of untreated words;
2. Investigating the carer perception of communication effectiveness and the impact on the carer quality of life;
3. Identification of any possible adverse events.

2 Outcome measures

The objectives of the trial will be evaluated by the following endpoints and for CSLT compared to Usual Care (UC), or Attention Control (AC). In addition to baseline measures, all outcome variables will be measured at

6 (end of treatment), 9 and 12 months by speech language therapists (SLTs) at each centre that are blinded to treatment allocation.

2.1.1 Primary endpoints

The co-primary endpoints assessed at 6 months from baseline are:

1) The change in word finding ability of words personally relevant to the participant at 6 months from baseline. Word finding ability is measured by a picture naming task using percentage scores based on a set of 100 personally selected words as detailed in Section 7.2.2;
2) Change in functional communication at 6 months measured by blinded ratings of video recorded conversations between a speech language therapist (SLT) and participants, using the activity scale of the Therapy Outcome Measures (TOMS).

2.1.2 Secondary endpoints

The key secondary endpoint is:

1) Change in patient perception of communication and quality life at 6 months. This is assessed using the Communication Outcomes After Stroke (COAST) – a patient-reported measure of communication-related activity, participation, and quality of life validated for evaluating SLT interventions in the HTA Assessing Communication Therapy in the North West (ACT NoW) project (Bowen et al., 2012).

Other secondary endpoints measured at 6 months are:

2) Use of learnt vocabulary in the context of conversation at 6 months measured using a checklist of target words during rating of the videoed conversations.
3) Quality-adjusted life years (QALYs) for cost-effectiveness – EuroQoL health utility questionnaire (EQ-5D) for patient and carer;
4) Generalisation to untreated words measured using the naming test from the Comprehensive Aphasia Test (CAT);
5) Carer perception of communication effectiveness measured using the first 15 questions from the Carer COAST;
6) Impact on carers' quality of life measured using the last 5 questions of the Carer COAST and the Carer Quality of Life measure (Carer QoL). The latter is for health economics evaluation.

All primary and secondary outcomes will also be reported at 9 and 12 months post randomisation to identify any longer term effect of the intervention.

Safety endpoints:

1) Negative effects of treatment – patient diary to record any difficulties/negative impacts of the

intervention. These are recorded every month for the 6 month treatment period;

2) Serious adverse events/adverse events (SAEs/AEs) may be recorded at any time during the study period. Formal checks will be carried out by SLTs every 3 months.

3 Sample size

The study aims to recruit 285 participants (95 per arm) across 20-24 SLT departments (study sites/centres). The target for each site is 15 participants in total with 5 randomised to each of the three study arms. The sample size of 285 participants in total (95 per arm) is the maximum sample size estimate across the two co-primary endpoints (word finding of personally selected words and functional communication – conversation) for 90% power and a 5% two-sided significance level.

3.1.1 Assumptions for the sample size calculation

For assessment of conversation (functional communication – activity scale of the TOMS) the estimated effect size is 0.45 of a standard deviation (SD) (with a correlation between baseline and outcome of 0.5 previously observed in the ACT NoW study, personally communicated by Prof Andy Vail, University of Manchester). The variance is adjusted to account for the baseline using $(1-\rho^2)$ such that $(1 \times (1-\rho^2)$ is used in the sample size calculations, where $\rho = 0.5$ is the correlation between baseline and outcome (Frison and Pocock, 1992). For change in word finding of personally selected words for treatment, the estimated effect size is 10% with an SD of 17.38% obtained from an analysis of covariance model based on results of the pilot study (Palmer et al., 2012). For patient-rated improvement using the COAST questionnaire the estimated effect size is 7.2%, with a SD of 18% based on external supplied data (with an assumed correlation between baseline and outcome of 0.5). The sample size of 285 participants has 83% power for the COAST (a key secondary endpoint).

The observed dropout rate was 5/33 (15%; 95% CI: 5% to 32%) in the pilot study which translated to a completion rate of 28/33 (85%; 95% CI: 68% to 95%) (Palmer et al., 2012), therefore the sample size has been inflated to account for 15% attrition. A summary of the calculations is shown in Table 1.

Table 1: Sample size assumed parameters and estimates

Outcome	Effect Size	Standard Deviation	Correlation	Power	Significance level	Evaluable Sample Size	Total Sample Size
Conversation	0.45	1*	0.5	90%	5%	78	92
Word finding	10%	17.38%	N/A	90%	5%	73+	86
COAST	7.2%	18%	0.5	83%	5%	81	95

* a standardised difference is used for the sample size calculation;

+ an inflation factor of 1.14 used to account for the fact variance comes from a pilot (Julious, 2004; Julious and Owen, 2006).

4 Randomisation, blinding, and outcome assessments

Participants will be randomised to one of the three trial arms using 1:1:1 allocation ratio. Randomisation will be performed by a web-based online randomisation system hosted by the Sheffield CTRU. The randomisation list will be stratified by centre and by the severity of word finding at baseline based on scores on the naming test of the CAT; mild (31 to 43), moderate (18 to 30), severe (5 to 17).

The SLTs will randomise participants in their homes using the Sheffield CTRU web-based online randomisation system and disclose their allocation. If no internet connection is available the SLT will phone the research team at the Sheffield CTRU, who will randomise online and give the allocation immediately over the phone to the SLT to disclose to the participant.

This is a single blind study recognising that participants cannot be blinded to their treatment allocation. The outcome assessors are SLTs with no previous involvement in the conduct of the trial. These are trained, via a webinar session run by the central team, to remain unaware of the allocation of the participants they will be assessing. Principal Investigators (PIs) will be asked not to disclose baseline case report forms, not openly discuss participants with colleagues in open plan offices, and to remind their participants not to discuss their activities on the trial with any other SLTs they may come into contact with, as it is 'a secret'.

When outcome assessors contact participants and conduct their assessments they will be advised to remind participants that their activity on the trial is 'a secret'. It is possible that during a conversation with the participant or carer, outcome assessors could become unblinded by the participant or their carer. If this occurs on the telephone, before the assessment takes place, then the assessment will be carried out by a different blinded assessor. If this occurs at the end of the visit, when the assessment is complete, then this will not be classed as an unblinded assessment, as the actual assessment was carried out when the assessor was still blinded. In the event of unblinding of the SLT occurring in this manner, the next assessment will be carried out by a different blinded assessor. All sites will have a minimum of two trained SLTs assessors that are blinded to the outcome to allow for unblinding issues. The intention is that the same assessor will carry out all outcome assessments for consistency, but if unblinding has occurred then an alternative assessor will be used as blinded assessments will take priority over assessments by the same assessor. If treatment arm allocation is disclosed during an assessment, then the outcome assessor will continue with the assessment but subsequently will alert the PI and complete an unblinding form. The unblinding form asks the assessor to record what they believe the participant's treatment allocation to be ("the suspected allocation"). In some instances, the assessor will guess the treatment allocation incorrectly, so the central team will report as "suspected unblinding" only. For example, the SLT may believe the treatment allocation to be UC but in fact, the participant is allocated to AC. Descriptive summaries of circumstances surrounding the unblinding of SLTs cases will be reported as described in Section 7.6.

5 Trial management and monitoring committees

The Trial Steering Committee (TSC) and Trial Management Group (TMG) will be established to govern the conduct of this study guided by GV001 and GV002 internal SOPs (Sheffield CTRU, 2016a, 2016b). The TMG

will be blinded to any data which could lead to them becoming unblinded such as resource use for individual participants. Three members of the TSC (Independent Chair, Independent Stroke Physician, Independent Statistician) constitute the Data Monitoring Committee and will review the accumulating unblinded data throughout the duration of the trial.

6 Data sources, protocol non-compliances, and analysis populations

6.1 Data sources

The data used in this study will come from data entered onto the Case Report Forms (CRFs) version 1 developed as guided by DM003 internal SOP (Sheffield CTRU, 2016c). Data will be stored on the CTRU database platform with the exception of the randomisation list which is held on the CTRU's secure randomisation system, developed and maintained by EpiGenesys. Electronic data will be extracted from the system at regular intervals in order to facilitate validation of the data and monitoring of the trial progress. Any spurious data will be queried and checked for consistency with data management before data lock. Personal records will not be accessible to CTRU staff. Where the randomised intervention as recorded on the randomisation list differs from the intervention group as recorded on the CRF, the randomisation list will be assumed to be the correct data source.

In the case of discrepancies between data relating to stratification factors recorded on the randomisation list and CRF, the data management team will investigate the source of discrepancies between the two data sources and the correct data source (based on their conclusions) will be used.

6.2 Protocol non-compliances

For the purposes of the analyses, participants who are deemed not to have adhered to the randomised intervention will be considered as protocol non-compliances. The number and proportion of participants who were randomised to CSLT or AC but technically received UC only will be summarised. This classification will be based on participants who did not practice any computer therapy sessions or puzzle book. Participants allocated to UC may purchase puzzle books or engage in word finding exercises on a computer. However, CSLT and AC are complex interventions including words and puzzles tailored to individual requirements in addition to access to trained volunteer support specifically for this study. Therefore, even if participants were to acquire puzzle books or engage in word finding exercises on a computer they would not have access to bespoke trained volunteer support – so "drop-ins" to CSLT or AC is unlikely.

6.3 Analysis populations

6.3.1 Modified Intention–to-Treat (mITT) analysis set

The mITT will be used as the primary set for analysis and any other sets for sensitivity analysis. This includes:

1) all participants for whom consent is obtained and;

2) treatment assignment as per randomised list regardless of 'circumstances' after randomisation, such as protocol non-compliance and treatment switching, and**;**
3) all participants with primary outcome data at 6 months from baseline.

Note that the mITT denominator for baseline tables of demographics and characteristics of participants illustrated in Table 2 and Table 3 will be primarily based on participants with either word finding or functional communication outcomes. For participants with primary outcome data at 6 months but missing data relating to the primary outcome assessed at baseline, the mean value of those with available baseline data will be used to impute missing baseline data.

We recognise the issue with missing data for this study where the first assessment of the primary outcome is at 6 months. From our experience during the pilot trial, participants that were dissatisfied with the treatment just stopped using the computer but stayed in the study. These participants will be included in the mITT but not in the Per-Protocol (PP) analysis set. Although there were no withdrawals due to treatment in the pilot trial, the common reasons for withdrawal were due to poor health, moving away, or not wanting to have outcome assessment performed. However, the study protocol makes provision for potential withdrawal from intervention by defining two types of withdrawal:

1) withdrawal from the intervention and,
2) withdrawal from the whole study.

If participants are in the first category, therapists at the site have been asked to check with the participants if they agree to the SLT visiting them to conduct outcome measurements at 6, 9 and 12 months. These participants will be included in the mITT analysis set.

For those participants in the second category, available data from these participants prior to their withdrawal will be used in the analysis. Reasons for withdrawal will be summarised as described in Section 7.3.1. Sensitivity analysis to assess the impact of missing data on the treatment effect estimates will be explored using imputation methods highlighted in Section 6.3.3 and detailed in Section 7.16.

6.3.2 Per Protocol (PP) analysis sets

Participants that are included in the PP analysis set are those for whom key components of the intervention were adhered to, including achieving the minimum amount of practice recommended and having access to support up to and including their 6 months visit. The objective is to explore the effectiveness of the intervention among participants who adhered to key components of this intervention as intended. PP classification will only be done for the CSLTand AC groups and it should be noted that some participants may meet PP classification described below without complete follow-up data (such as at 6 months).

In this study, four PP analysis sets will be considered that vary the minimum threshold of the total computer practice time required and the number of contacts with the volunteer or assistant if they wished. PP analysis sets for the CSLT group will include those participants who:

1) practiced computer therapy exercises for a minimum total of *26 hours* over a period of 6 months from randomisation (PP1 CSLT);
2) practiced computer therapy exercises for a minimum total of *10 hours* over a period of 6 months from randomisation, which is consistent with other studies (Breitenstein et al., 2017; Katz et al., 2000) (PP2 CSLT);
3) practiced computer therapy exercises for a minimum total of *26 hours* over a period of 6 months from randomisation and had a minimum of *4 contacts* from a volunteer or assistant unless it had been documented that they did not wish to have volunteer/assistant support (PP3 CSLT);
4) practiced computer therapy exercises for a minimum total of *10 hours* over a period of 6 months from randomisation and had a minimum of 4 *contacts* from a volunteer or assistant unless it had been documented that they did not wish to have volunteer/assistant support (PP4 CSLT).

Section 7.2.1 details how participants' computer practice time and the number of contacts with the volunteer or assistant (if they wished) are computed.

Participants will be excluded from the PP analysis sets for the CSLT group for one or more of the following reasons:

- practised computer therapy for less than 26 hours (PP1 CSLT and PP3 CSLT referred to above)
- practiced computer therapy for less than 10 hours (PP2 CSLT and PP4 CSLT referred to above)
- the participant was not offered volunteer/assistant support for a minimum of 4 contacts if they wished (PP3 CSLT and PP4 CSLT referred to above);
- the participant was randomised in error, for example, the participant did not meet inclusion criteria but was randomised or received treatment which they were not allocated to.

For the AC group, the two PP sets will include those participants who:

1) were sent at least 6 puzzle books irrespective of the number of contacts made. A puzzle book is only sent if a participant reported use of the previous one indicating regular engagement in puzzle book activity, or in a small number of instances where a person was not able to be contacted by phone. Participants were excluded from the PP set if books had been sent with no evidence from future phone calls of the participant's engagement with the book (PP1 AC);
2) were sent at least 6 puzzle books (as per PP1 AC) and were contacted at least 4 times between randomisation and at 6 months follow-up (PP2 AC).

For comparability, CSLT and AC PP analysis sets defined above will be matched with respect to: a) minimum total practice time and minimum puzzle books sent; b) minimum total practice time, minimum puzzle books sent, and minimum contact. This will result in four pairwise comparisons of the treatment effect: PP1 CSLT versus PP1 AC; PP2 CSLT versus PP1 AC; PP3 CSLT versus PP2 AC; and PP4 CSLT versus PP2 AC.

There will be no PP classification required for the UC group relating to the 'intervention' aspects. However, across treatment groups participants whose outcome measures were assessed 14 days before or 31 days after the expected 6 months assessment will be excluded in the PP analysis.

6.3.3 Missing data

In Section 6.3.1, the mITT excludes randomised participants without primary outcome data at 6 months. Additional sensitivity analyses on the primary endpoint will be undertaken by including these participants with missing data via the use of imputation of the missing data as detailed in Section 7.16 using:

- mean value imputation (MVI) or linear interpolation (LI),
- last observation carried forward (LOCF) where appropriate and,
- Multiple Imputation (MI).

6.3.4 Complete Case set (CC)

These are participants who completed their outcome measurements at different time points (baseline, 6, 9 or 12 months). This will be utilised for the subsidiary analysis on the co-primary outcome measures at 9 and 12 months and for mean profile response stratified by treatment group.

6.3.5 Safety set

The analysis of safety outcomes detailed in Section 7.15 will include all randomised participants with informed consent and treatment allocation for analysis and will be based on the actual treatment received. For instance, if a participant was randomised to receive UC but received CSLT for some reason, then that participant will be assigned to CSLT for safety-related analysis. Similarly, participants who were randomised to either CSLT or AC but technically received UC only will be reassigned to UC group for the investigation of safety outcomes.

7 Outline of statistical analyses

7.1 General considerations

The naming test from the CAT, activity scale of the TOMS, COAST, EQ-5D, Carer COAST (CaCOAST) and Carer QoL measures will be undertaken pre-treatment (baseline), and at 6, 9 and 12 months post-baseline. Resource usage data will be collected every 3 months. Any adverse events (AEs) may be reported at any time during the study period, with formal checks carried out by the SLTs at 3-month intervals. The use of outcome measures for health economics evaluation only such as Carer QoL and resource usage is beyond the scope of this SAP and will be described elsewhere.

7.2 Definitions, data manipulation, and dealing with missing data

This section deals with data manipulation of key variables and computation of summaries from outcome measures for analysis including how missing data will be dealt with. Outcomes are presented in order of their importance in relation to trial aims and objectives.

7.2.1 General derived variables

The baseline date will be considered as the date of randomisation. The centre will be defined as the place from which the participant was identified. Age in years is computed as (date of randomisation - date of birth)/365.25.

The severity of word finding difficulty will be categorised based on the CAT Naming Test using a severity rating of:

- mild (31-43),
- moderate (18-30),
- severe (5-17).

Comprehension ability, a categorical variable, will be generated using total scores from the CAT Comprehension of Spoken Words using the following classification system:

- severe (0 to 8); inconsistently understanding at 2 information carrying words (ICW) level,
- moderate (9 to 17); consistently understanding at 2-3 ICW level/simple sentence structures but not complex sentence structures,
- mild (18 to 26); some understanding of complex sentence structures but not consistent;
- within normal limits (27 to 32) based on CAT cut-off score for normal /aphasic.

Time variables will be computed based 365.25 days in a year and number of days in a month will be rounded upwards. Computer usage as assessed by the total time spent practising will be based on practising time over the first 6 months from randomisation (approximately the first 183 days).

Data from therapy assistant/volunteer time to calculate how many contacts a participant had between randomisation and 6 months follow-up. For each new date recorded, a participant can have up to 2 contacts. If at least one of the activity type and duration subcategories are selected, this counts as a single contact. If the participant declined to receive support, then that participant will still be included in the PP analyses.

7.2.2 Word finding of personally selected words for treatment

Personal Vocabulary Naming Test is used to assess word finding ability based on 100 personally selected words for treatment. For each personally selected word, word finding ability is then assessed using the following scoring system:

- 0 for an incorrect or no response,
- 1 for a correct word named correctly after a delay of 5 seconds and/or for a self-correction,
- 2 for a correct prompt answer within 5 seconds.

This scoring system yields a potential maximum score of 200. It should be noted that some participants may be assessed based on less than 100 words due to some reasons such as tiredness although all participants are expected to be assessed based on 100 personally selected words. If less than 100 words but more than 70 words are assessed, the word finding ability for participant *k* (Y_k), expressed as a percentage will be calculated based on the total score relative to the potential maximum score given by:

$$Y_k = \frac{\sum_{i=1}^{\mathrm{V}} item_i}{2\mathrm{V}} * 100$$

Where $i = \{1, 2, \dots, V\}$ is the picture item considered for pesonal vocabulary naming and V is the total number of personally selected words assessed.

7.2.3 Functional communication – TOMS

Only the activity dimension of the TOMS instrument will be used to assess functional communication rating (conversation) which is one of the coprimary outcomes. This activity dimension is measured on a six-point ordinal scale ranging from 0 (unable to communicate in any way) to 5 (communicates effectively in all situations). The rating system also allows scoring between ordinal descriptors such as 0.5, 1.5, 2.5, 3.5, and 4.5. Thus the rating scores yield 11 ordinal possibilities which will be treated as a continuous outcome.

There is a ceiling effect for participants who are able to communicate effectively in all situations at baseline, with a TOMS rating of 5. The number (and percentage if appropriate) of these participants will be reported by treatment group. Their baseline level of word finding ability for their personally selected words for treatment will also be noted.

7.2.4 COAST scoring and computation of summary measure

The COAST is a patient-centred measure used to assess self-perceived communication effectiveness and impact on quality of life for people with aphasia and/or dysarthria (Long et al., 2008). The measure has 20 items and each item is measured on a scale of 0 to 4 and a summary measure is calculated to translate into a percentage score. A procedure is then applied to compute a percentage score under a number of scenarios; all applicable and answered items, the existence of 'not applicable' items, the existence of 'unclear' or 'no response' items. The overall percentage is computed using a validated algorithm summarised as follows (Bowen et al., 2009):

1) When all 20 items are applicable and the responses recorded, a percentage summary measure is calculated using the following equation:

$$COAST\ (\%) = \frac{\sum_{i=1}^{20} item_i}{80} * 100$$

Where *i* is the item indicator,

2) When there are 'not applicable' items and all applicable items have been answered, a percentage summary measure is calculated by accounting for in the numerator and denominator using the following

equation:

$$COAST\ (\%) = \frac{\sum_{i=1}^{\mathrm{J}} item_i}{4\mathrm{J}} * 100$$

Where $J \leq 20$ is the total number of applicable items and the corresponding maximum possible score is $4J$. For example, 18 applicable items will have a corresponding denominator of 72 scores.

3) When there are 'not applicable' and 'unclear' or 'no response' items, the following approach will be undertaken:
    a. The summary measure will be deemed invalid and no computation will be done if more than 10% of the applicable items are missing ('unclear' or 'no response')
    b. If there are less than or equal to 10% of applicable items with missing values ('unclear' or 'no response'), implement the following;
        i. Calculate the mean response score of applicable and completed items,
        ii. Replace the missing ('unclear' or 'no response') item score by the mean response,
        iii. Compute the following percentage summary measure as follows,

$$COAST\ (\%) = \frac{(\sum_{i=1}^{K} item_i) + (J - K)M}{4J} * 100$$

Where K is the number of applicable items with complete responses and M is their corresponding mean response. Thus, J-K is the number of applicable items with missing responses ('unclear' or 'no response'). This equation is generalisable, reduces to cases 2) and 3) so it can be used to compute percentage summary measures under all the described scenarios.

7.2.5 CaCOAST scoring and computation of summary measure

The CaCOAST assesses carer perception of patient's communication effectiveness and impact on their quality of life (Long et al., 2009). The measure has 20 items and each item is measured on a scale of 0 to 4 and a percentage summary measure is calculated. The CaCOAST was administered by the research therapists as one questionnaire, however, the first 15 items and the last 5 items will be analysed separately as they address two different research questions (items 5 and 6 in Section 2.1.2) The first 15 items assess carer perception of communication while the last 5 items measure the impact of the patient's communication difficulties on the carer's quality of life. Although the original scoring algorithm is based on all 20 items, the research team considered the first 15 items and the last 5 items separately to assess different aspects. We, therefore, adapted the scoring algorithm which is consistent with the original scoring system using 20 items (Bowen et al., 2009) to compute the $CaCOAST_{15}$ (%) and $CaCOAST_5$ (%) but based on the first 15 and last 5 items, respectively. This uses the same scoring algorithm as described in Section 7.2.4 but account for missing data as other aspects are uninformative ('not applicable' and 'unclear').

7.2.6 CAT: Naming Test and Comprehension of Spoken Words

The Naming Test of the CAT consists of 24 (picture naming tasks) words and assesses generalisation of treated to untreated words measured at 6, 9, and 12 months, including baseline. For each picture naming task, the following scoring system is used depending on participant response:

- 0 for an incorrect response,
- 1 for an accurate response after a delay of more than 5 seconds,
- 2 for an accurate and prompt answer.

The total score, which ranges from 0 to 48 is then generated to assess word finding ability of untreated words.

The CAT Comprehension of Spoken Words is only assessed at baseline with total scores ranging from 0 to 32. Baseline comprehension ability, a categorical variable, will be generated as detailed in Section 7.2.1.

Missing information (item level or all items) is possible due to tiredness or being unable to complete the tests. For missing items, summary measures from the CAT Naming Test and Comprehension of Spoken Words will be calculated assuming conservative worst case scenario. That is, a zero score will be assumed for missing item scores when computing the summary measures (total scores). No summary measure will be calculated if all items are missing.

7.2.7 Word finding of treated words used in conversation

The use of vocabulary in the context of the conversation is assessed using a checklist of target words during ratings of videoed conversations at 6 months. Out of the 100 treated words (personally selected for treatment), the number of words retrieved during videoed conversations will be counted. Note that a word correctly retrieved will be counted only once regardless of the number of times it has been retrieved during the conversation. The total score will range from 0 to 100.

7.2.8 EQ5D computation of summary measure

The study uses the EQ-5D-5L version to assess health status and produces a single index value for health status for use in the calculation of quality-adjusted life years to inform health economics evaluation of investigative interventions (van Reenen and Janssen, 2015). The instrument consists of an EQ-5D-5L descriptive system and an EQ-5D-5L VAS. The descriptive system has 5 dimensions assessing mobility, self-care, usual activity, pain/discomfort, and anxiety. Each of these dimensions has 5 levels of severity which participants are asked to select one of them to best describe their health status ‘today’: no problems, slight problems, moderate problems, severe problems, and extreme problems. Based on participants’ responses from these 5 dimensions, a single index value will be calculated as detailed by Devlin et al (2016). The single index values are on a scale of 0 (full health) to 1 (state equivalent to dead) and health states considered to be worse than dead attain negative values (<0).

As for the EQ-5D-5L VAS, participants are asked to rate how good or bad their health is 'today' on a scale of 0 (the worst health imaginable) to 100 (best health imaginable). The scores from this continuous scale will be used to assess change in overall self-rated health status.

7.3 Demographics and baseline characteristics

Demographics and baseline characteristics of participants will be summarised within each treatment arm and overall and assessed for comparability based on the mITT analysis set in accordance with the CONSORT guideline for parallel group, individually randomised trials (Schulz et al., 2010). The mITT denominator is the number of participants with 6 months primary outcome(s) data either word finding or functional communication. For continuous variables, summaries will comprise the number of participants and either of the following depending on the distribution of the data:

1) Mean, SD, minimum (min) and maximum (max) or;
2) Median, interquartile range (IQR), min and max.

Summaries for categorical outcomes will comprise the number of participants and their respective proportion as a percentage in that category. Variables such as aphasia type, gender, recruitment centre, the severity of aphasia, type and location of the stroke, and evidence of apraxia of speech will be treated as categorical variables. The number of missing responses will also be provided for both continuous and categorical variables. The results will be presented as shown in Table 2 and Table 3. No statistical significance testing will be done to test for a difference in the baseline characteristics but any noted differences will be descriptively reported (de Boer et al., 2015; Pocock et al., 2002; Senn, 1994).

Summaries Table 2 and Table 3 will also be produced using all randomised participants for whom consent was obtained with or without 6 months co-primary outcome data (strict ITT) to facilitate the interpretation of results from sensitivity analyses using all participants based on imputation methods.

Table 2. Demographic and baseline characteristics of participants at baseline

Variable	**Scoring**	**UC (n=xx)**	**AC (n=xx)**	**CSLT (n=xx)**	**All (N=XX)**
Recruitment centre	Sheffield	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Hull	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Newcastle	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Northern	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Belfast	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	South Beds	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	…	...	…	…	…
Age (years)	N	xx	xx	xx	xx
	Mean (SD)	xx (xx)	xx (xx)	xx (xx)	xx (xx)
	Median (IQR)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx	xx to xx
Gender	Male	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Female	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
Aphasia type	Anomic	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Non-fluent	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Mixed non-fluent	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Fluent	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)

Severity of word finding (CAT Naming Test)	Mild (31-43)	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Moderate (18-30)	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Severe (5-17)	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
Evidence of apraxia of speech	Yes	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	No	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
Type of stroke	Infarct	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Haemorrhage	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Not known	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
Location of stroke	Brain stem	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	MCA	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Frontal lobe	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Temporal lobe	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Parietal lobe	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Occipital lobe	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Cerebellum	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Not known	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
Lateralisation (if not brain stem)	Left	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Right	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Not known	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
Comprehension of Spoken Sentences from CAT	N	xx	xx	xx	xx
	Mean (SD)	xx (xx)	xx (xx)	xx (xx)	xx (xx)
	Median (IQR)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx	xx to xx
CAT comprehension ability	Within normal limits (27 to 32)	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Mild (18 to 26)	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Moderate (9 to 17)	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)
	Severe (0 to 8)	xx (xx %)	xx (xx %)	xx (xx %)	xx (xx %)

MCA: Middle Cerebral Artery; SD: standard deviation, IQR: interquartile range; Min: minimum; Max: maximum; CAT: Comprehensive Aphasia Test; Recruitment centre is the site recorded at randomisation.

Table 3. Distribution of continuous outcomes measured at baseline

Variable	Scoring	UC (n=xx)	AC (n=xx)	CSLT (n=xx)	All (N=xx)
Word finding ability (%)	N	xx	xx	xx	xx
	Mean (SD)	xx (xx)	xx (xx)	xx (xx)	xx (xx)
	Median (IQR)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx	xx to xx
Functional communication (TOMS) [a]	N	xx	xx	xx	xx
	Mean (SD)	xx (xx)	xx (xx)	xx (xx)	xx (xx)
	Median (IQR)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx	xx to xx
COAST (%) [b]	N	xx	xx	xx	xx
	Mean (SD)	xx (xx)	xx (xx)	xx (xx)	xx (xx)
	Median (IQR)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx	xx to xx
$CaCOAST_{15}$ (%): first 15 items [c]	N	xx	xx	xx	xx
	Mean (SD)	xx (xx)	xx (xx)	xx (xx)	xx (xx)
	Median (IQR)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx	xx to xx
$CaCOAST_5$ (%): last 5 items [d]	N	xx	xx	xx	xx
	Mean (SD)	xx (xx)	xx (xx)	xx (xx)	xx (xx)
	Median (IQR)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx	xx to xx
Word finding of untreated words (Score from CAT Naming Objects test)	N	xx	xx	xx	xx
	Mean (SD)	xx (xx)	xx (xx)	xx (xx)	xx (xx)
	Median (IQR)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx	xx to xx
EQ-5D-5L index value (patients-aphasia friendly) [e]	N	xx	xx	xx	xx
	Mean (SD)	xx (xx)	xx (xx)	xx (xx)	xx (xx)
	Median (IQR)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx	xx to xx
EQ VAS score (patients-aphasia friendly) [f]	N	xx	xx	xx	xx
	Mean (SD)	xx (xx)	xx (xx)	xx (xx)	xx (xx)
	Median (IQR)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx	xx to xx
EQ-5D-5L index value (patients – proxy) [e]	N	xx	xx	xx	xx
	Mean (SD)	xx (xx)	xx (xx)	xx (xx)	xx (xx)
	Median (IQR)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx	xx to xx
EQ VAS score (patients -proxy) [f]	N	xx	xx	xx	xx
	Mean (SD)	xx (xx)	xx (xx)	xx (xx)	xx (xx)
	Median (IQR)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx	xx to xx
EQ-5D-5L index value (carers) [e]	N	xx	xx	xx	xx
	Mean (SD)	xx (xx)	xx (xx)	xx (xx)	xx (xx)
	Median (IQR)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx	xx to xx
EQ VAS score (carers) [f]	N	xx	xx	xx	xx
	Mean (SD)	xx (xx)	xx (xx)	xx (xx)	xx (xx)
	Median (IQR)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)	xx (xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx	xx to xx

[a] higher TOMS scores mean improved functional communication; [b] higher score indicates positive self-perceived communication and impact on patient's quality of life; [c] higher score indicates positive carer's perception of patient's

communication ability; [d] higher score indicates positive carer perception of the impact of the patient's communication ability on the carer's quality of life; [e] higher score indicates higher health-related quality of life; [f] higher score indicates positive perception of health status, with 0 and 100 meaning worst and best health status imaginable. Computation of the COAST (%), $CaCOAST_{15}$ (%), $CaCOAST_{5}$ (%) and EQ-5D-5L is detailed in Sections 7.2.4, 7.2.5, and 7.2.8, respectively.

7.3.1 Recruitment and data completeness

The following summaries will be presented for all participants screened for entry to the study, by identification or recruitment source and overall. For the purpose of recruitment, the following summaries will be collected:

1) The number of participants screened and screening rate (per month),
2) The number of participants recruited and recruitment rate (per month),
3) Number and percentage of participants not recruited and the reasons for non-recruitment.

Relevant summaries on recruitment, consent and data completeness during follow-up will be presented in a CONSORT flowchart as shown in Figure 2 (Schulz et al., 2010). These summaries on recruitment and data completeness will be made available to members of the TSC and TMG during the course of the trial. Reasons for withdrawal at different follow-up times will also be summarised by treatment group.

7.4 Dealing with deaths

In this trial population, few deaths during the trial are expected. The TMG discussed on how to handle deaths during analysis. There was agreement that the influence of the intervention on mortality such as increasing the risk of mortality is very unlikely. In addition, the interpretation of imputed data for this study, such as word finding and functional communication for participants who have died is difficult. In this regard, missing data due to deaths will not be imputed. Thus, deaths prior to a follow-up assessment will be excluded in any analysis after this point. However, for multiple imputation, participants with baseline and 6 months data will be included in the imputation model and imputed values post-death will be excluded from statistical analyses. However, the number of individuals who have died during the trials will be reported by treatment group and presented in the CONSORT flowchart.

7.5 Differential characteristics of completers versus non-completers at baseline

This section aims to explore whether completers differ systematically at baseline from non-completers with respect to their key characteristics. Completers are participants with primary outcome data at 6 months (either word finding or functional communication) and non-completers are those with missing primary outcome data for any reason, excluding death. Exploring the patterns of missing data is important to aid interpretation of results as well as to inform the imputation model (e.g covariates to include) used in subsequent multiple imputations.

First, the reasons for missing data will be summarised appropriately. Second, descriptive statistics will be used to further this objective and presented as illustrated in Table 4 and Table 5 for continuous and categorical variables, respectively, as appropriate. Participants who have died prior to 6 months follow-up assessment will be excluded in this exploratory analysis since they will not be included in any subsequent analysis. Third, baseline variables associated with the co-primary endpoints will be descriptively explored among those with available data (CC analysis set). For example, using scatterplots of co-primary endpoints against baseline variable stratified by the treatment allocation.

Table 4: Continuous baseline characteristics of completers versus non-completers

Baseline variable	Scoring	Completers			Non-completers		
		UC	AC	CSLT	UC	AC	CSLT
Word finding ability (%)		(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)	xx(xx)	xx(xx)	xx(xx)
	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
Functional communication (TOMS)		(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)	xx(xx)	xx(xx)	xx(xx)
	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
COAST (%)		(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)	xx(xx)	xx(xx)	xx(xx)
	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
$CaCOAST_{15}$ (%)		(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)	xx(xx)	xx(xx)	xx(xx)
	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
$CaCOAST_{5}$ (%)		(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)	xx(xx)	xx(xx)	xx(xx)
	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
Word finding of untreated words		(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)
(score CAT naming objects test)	Mean(SD)	xx(xx)	xx(xx)	xx(xx)	xx(xx)	xx(xx)	xx(xx)
	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
Severity of word finding		(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)
(Comprehension of Spoken	Mean(SD)	xx(xx)	xx(xx)	xx(xx)	xx(xx)	xx(xx)	xx(xx)
Sentences)	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
Age (years)		(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)
	Mean(SD)	xx(xx)	xx(xx)	xx(xx)	xx(xx)	xx(xx)	xx(xx)
	Median(IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)

Note: Participants who have died prior 6 months follow-up assessment will be excluded. "Non-completers" may be replaced by "completers and non-completers" (excluding deaths prior to 6 months follow-up) at the discretion of the Trial Statistician.

Table 5: Categorical baseline characteristics of completers versus non-completers.

Baseline variable	Scoring	Completers			Non-completers		
		UC (n=xx)	AC (n=xx)	CSLT (n=xx)	UC (n=xx)	AC (n=xx)	CSLT (n=xx)
Gender	Female	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
	Male	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
Severity of word finding	Mild (31-43)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
	Moderate (18-30)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
	Severe (5-17)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
Evidence of apraxia speech	Yes	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
	No	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
Type of stroke	Infarct	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
	Haemorrhage	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
	Unknown	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
Location of stroke	Brain stem	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
	MCA	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
	Frontal lobe	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
	Temporal lobe	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
	Parietal lobe	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
	Occipital lobe	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
	Cerebellum	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)
	Not known	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)	xx(xx%)

Note: Participants who have died prior 6 months follow-up assessment will be excluded. “Non-completers” may be replaced by “completers and non-completers” (excluding deaths prior to 6 months follow-up) at the discretion of the Trial Statistician.

7.6 Unblinding of outcome assessors at 6 months

Although outcome assessors will be blinded to the treatment allocation, it could be possible that they could become unblinded for some reason, as highlighted in Section 4. In this section, the following descriptive summaries (total and by treatment group) will be reported focusing on 6 months assessments:

- the number and percentage of participants with reported "suspected unblinding" of assessment,
- the number and percentage of participants with reported unblinding assessment.

7.7 Description of Usual Care

This section aims to explore variation in UC treatment received 3 months prior to randomisation and 3 months prior to the intended follow-up times at 3, 6, 9 and 12 months. Proxy data to measure the UC treatment received includes three types of treatment received and duration of each treatment (in minutes); impairment (rehabilitation), compensatory (enabling), and support (supportive). Profiles of the mean duration of UC treatment received by treatment group (UC, AC, or CSLT) will be graphically presented by:

- type of treatment received and,
- overall

This will be based on patients with baseline and follow-up measures. The research team highlighted that the time post-stroke influences the duration of UC treatment patients receive. These relationships may be explored using scatter plots.

7.8 Analysis of the primary and key endpoints at 6 months

7.8.1 Co-primary endpoints: word finding and conversation

The co-primary endpoints assessed following 6 months of treatment (from baseline) will be the change in word finding of personally selected words and change in functional communication (conversation), measured using the activity scale of the TOMS. The primary analysis will be based on mITT set defined in Section 6.3.1. The denominators may differ across endpoints depending on primary outcome data completeness at 6 months.

For the change in word finding of personally selected words at 6 months from baseline, the measure of intervention effect will be the mean difference in change in word finding of personally selected words between the CSLT and UC groups, and the CSLT and AC groups. The primary analyses will utilise a multiple linear regression model adjusted for baseline word finding of personally selected words and stratification factors (centre and severity of word finding) as fixed effects (Kahan and Morris, 2012). The outcome will be modelled as a function of:

- word finding of personally selected words at baseline,
- treatment group (UC, AC, and CSLT),
- centre as a fixed effect (random effect may be used depending on model fit) and,

- the severity of word finding as a fixed effect (mild, moderate, and severe).

Results will be reported and presented as adjusted mean difference in change in word finding of personally selected words between the CSLT and UC, CSLT and AC (the pre-specified comparisons), and AC and UC (an exploratory comparison), with its associated 95% CI and associated P-value.

Improvement in functional communication at 6 months assessed using activity domain of the TOMS, which is a co-primary endpoint will be analysed as the change in functional communication adjusted for baseline functional communication in addition to stratification factors. The results of these primary endpoints will be jointly reported as shown in Table 6.

Table 6: Co-primary endpoints results at 6 mths from randomisation (mITT)

Co-primary outcomes at 6 months	**UC**		**AC**		**CSLT**		**CSLT versus UC** †		**AC versus UC** †		**CSLT versus AC** ‡	
	n	Mean(SD)	n	Mean(SD)	n	Mean(SD)	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value
Change in word finding of personally selected words (%) [1]	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx [a]	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx [c]
Change in functional communication [2]	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx [b]	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx [d]

Main results from multiple linear regression models adjusted for baseline measures and fixed stratification factors (centre and severity of word finding)

[a, b, c,d] are referenced in Figure 1 to aid interpretation of Hochberg sequential and hierarchical hypotheses testing procedure for decision-making

† UC as the reference group; ‡ AC as the reference group; UC: Usual care; AC: Attention control; CSLT: Computerised speech and language therapy; MDC: mean difference in change; SD: standard deviation

Interpretation: [1] higher scores indicate improved personal vocabulary of personal importance; [2] higher scores indicate improved functional communication ability in conversation

7.8.2 Sensitivity analyses on the primary endpoints: word finding and conversation

A sensitivity analysis will be conducted using a multiple linear regression model adjusted for baseline word finding of personally selected words, stratification factors (centre and severity of word finding), and potential confounders pre-specified by the research team (length of time post stroke and location of stroke). In the case of marked residual baseline imbalances, additional covariates may be included in the model at the discretion of the research team. The outcome will be modelled as a function of:

- word finding of personally selected words at baseline;
- treatment group (UC, AC, and CSLT);
- centre as a fixed effect (random effect may be used depending on model fit);
- the severity of word finding as a fixed effect (mild, moderate, and severe);
- the length of time post stroke and;
- the location of stroke (yes or no); middle cerebral artery (MCA), frontal lobe, parietal lobe, and temporal lobe.

A similar analysis strategy will be undertaken for the change in functional communication (TOMS) at 6 months adjusted for baseline functional communication. The results of these sensitivity analyses will be jointly reported as shown in Table 7.

Table 7: Sensitivity analyses – co-primary endpoints at 6 mths adjusted for additional potential confounders

Co-primary outcomes at 6 months	**UC**		**AC**		**CSLT**		**CSLT versus UC †**		**AC versus UC †**		**CSLT versus AC ‡**	
	n	Mean(SD)	n	Mean(SD)	n	Mean(SD)	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value
Change in word finding of personally selected words [1]	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Change in functional communication [2]	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx

Sensitivity analysis results from multiple linear regression models adjusted for baseline measures, fixed stratification factors (centre and severity of word finding), and potential confounders (length of time post stroke and location of stroke)

† UC as the reference group; ‡ AC as the reference group; UC: Usual care; AC: Attention control; CSLT: Computerised speech and language therapy; MDC: mean difference in change; SD: standard deviation

Interpretation: [1] higher scores indicate improved personal vocabulary of personal importance; [2] higher scores indicate improved functional communication ability in conversation.

7.8.3 Key secondary endpoint: patient perceptions of communication and its impact on their life

The COAST questionnaire is used to assess patient rated communication effectiveness and impact on their quality of life. The key secondary endpoint is the change in patient perception of communication and impact on their quality of life at 6 months from baseline.

The computation of COAST (%) summary measures based on all 20 items is detailed in Section 7.2.4. Based on the mITT set, the change in COAST (%) at 6 months from baseline will be modelled using a multiple linear regression model adjusted for:

- baseline COAST (%);
- treatment group (UC, AC, and CSLT);
- centre as a fixed effect (a random effect may be used depending on model fit);
- the severity of word finding as a fixed effect (mild, moderate, and severe).

Results will be reported as shown in Table 8 and will be the basis for the Hochberg testing procedures as described in Section 7.8.4.

A sensitivity analysis will be undertaken using a multiple linear regression model adjusted for baseline COAST (%), fixed stratification factors (centre and severity of word finding), and potential confounders pre-defined by the research team (length of time post stroke and location of stroke). In the case of marked residual baseline imbalances, additional covariates may be included in the model at the discretion of the research team.The change in COAST (%) will be modelled as a function of, and reported as shown in Table 8 together with main results for comparability:

- baseline COAST (%);
- treatment group (UC, AC, and CSLT);
- centre as a fixed effect (random effect may be used depending on model fit);
- the severity of word finding as a fixed effect (mild, moderate, and severe);
- the length of time post stroke and;
- the location of stroke (yes or no); MCA, frontal lobe, parietal lobe, and temporal lobe.

Table 8: Key secondary outcome – COAST results at 6 mths (mITT)

Key secondary outcome	UC		AC		CSLT		CSLT versus UC †		AC versus UC †		CSLT versus AC ‡	
	n	Mean(SD)	n	Mean(SD)	n	Mean(SD)	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value
Change in COAST (%) [1]	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx [e]	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx [f]
Change in COAST (%) [2]	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx

[1] Main results from multiple linear regression models adjusted for baseline measures and fixed stratification factors (centre and severity of word finding)

[2] Sensitivity analysis results from multiple linear regression models adjusted for baseline measures, fixed stratification factors (centre and severity of word finding), and additional potential confounders (length of time post stroke and location of stroke).

[e, f] are referenced in Figure 1 to aid interpretation of Hochberg sequential and hierarchical hypothesis testing procedure

† UC as the reference group; ‡ AC as the reference group; UC: Usual care; AC: Attention control; CSLT: Computerised speech and language therapy; MDC: mean difference in change; SD: standard deviation

Interpretation: [1,2] higher percentage score indicates improved patient perception of communication effectiveness and its impact on their quality of life.

7.8.4 Application of Hochberg hypotheses testing procedure: co-primary and key secondary endpoints

Co-primary and key secondary outcomes results reported as shown in Table 6 and Table 8 will be interpreted using a Hochberg testing procedure to control the chances of falsely declaring statistically significant results (at 5%) due to multiple hypothesis testing; multiple endpoints (co-primary and key secondary) and multiple treatment comparisons (Hochberg and Tamhane, 1987).

Figure 1 illustrates the interpretation strategy of the results from Table 6 and Table 8 (Sections 7.8.1 and 7.8.3) in order to claim statistical significance and superiority. Significance will be declared if changes in both word finding of personally selected words and functional conversation at 6 months are significant at the 5% level or if either comparison is significant at 2.5%. COAST ratings will be assessed at 5% if both changes in functional communication (conversation) and word finding of personally selected words are significant. Comparisons will be made firstly between the CSLT and the UC groups, and then between the CSLT and AC groups.

If and only if significance is declared for both co-primary outcomes, a similar comparison of AC to intervention will be made. Significance will be declared for the comparison of AC to intervention if and only if changes in both word finding of personally selected words and functional communication (conversation) are significant at the 5% level or if either comparison is significant at 2.5%.

If and only if significance is declared for the comparison of AC to intervention for both comparisons will the key secondary outcome measure (patient perception of communication and impact on their quality of life, measured using the COAST rating questionnaire) be used in a further comparison of UC to intervention. If and only if this comparison is significant at the 5% level will the intervention be compared to AC based on COAST rating questionnaire. Results will be presented as shown in Table 6 and Table 8.

Table 29 to Table 31 (Appendix) are examples to help with the interpretation of results using the Hochberg multiple testing procedure.

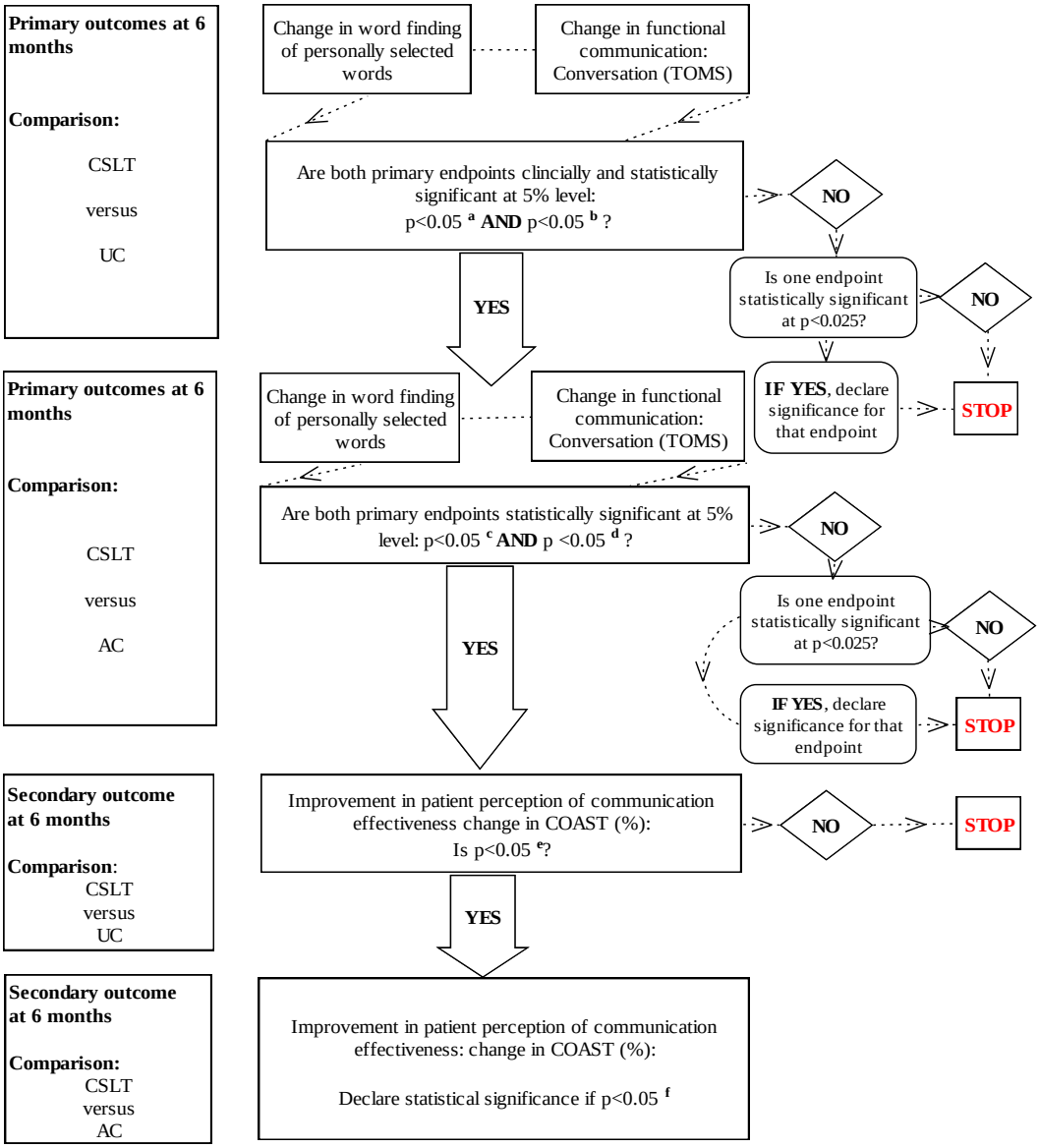

Figure 1: Interpretation of the Hochberg hierarchical sequential hypotheses testing strategy
Superscripts [a, b, c, d, e, f] are referenced in Table 6 and Table 8.

Exploratory analyses of the co-primary and key secondary outcomes between the AC group and the UC will be undertaken and results presented together with other comparisons, for instance, as shown in Table 6 to Table 8.

7.8.5 Attrition and protocol compliance sensitivity analyses: co-primary and key secondary endpoints

This section aims to explore the impact of attrition and adherence to the protocol on the co-primary outcomes (change in word finding of personally selected words and change in functional communication) and a key secondary outcome (patient rated communication and its impact on quality of life) results at 6 months. Sensitivity analyses will be undertaken under various assumptions to predict missing data (MI in Section 6.3.3) and PP compliance (Section 6.3.2). These outcomes will only be modelled using a multiple linear regression

model as a function of their baseline measures and stratification factors (centre and severity of word finding). Sensitivity analyses results will be reported as shown in Table 9 together with main results based on mITT set from Table 6 for comparability. LOCF will not be considered for the analysis of 6 months outcomes because of the model structure adjusted for baseline.

Table 9: Additional sensitivity analyses on the co-primary and key secondary outcomes at 6 mths

Outcomes at 6 months	Analysis set	UC		AC		CSLT		CSLT versus UC †		AC versus UC †		CSLT versus AC ‡	
		n	Mean(SD)	n	Mean(SD)	n	Mean(SD)	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value
Co-primary:													
Change in word finding of personally selected words [1]	mITT ‡‡	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	MI ‡‡‡	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	PP CC	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	PP MI	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Change in functional communication (TOMS) [2]	mITT ‡‡	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	MI ‡‡‡	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	PP CC	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	PP MI	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Key secondary:													
Change in COAST (%) [3]	mITT‡‡	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	MI ‡‡‡	xx	xx(xx)	xx	xx(xx)	Xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	PP CC	xx	xx(xx)	xx	xx(xx)	Xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	PP MI	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx

Sensitivity analysis results from multiple linear regression models adjusted for baseline measures and fixed stratification factors (centre and severity of word finding); CC: Complete case; MI: Multiple Imputation; PP: Per-protocol (four analysis sets); † UC as the reference group; ‡ AC as the reference group; UC:Usual care; AC:Attention control; CSLT:Computerised language speech therapy; MDC: mean difference in change; SD: standard deviation

‡‡ Results from Table 6 for comparability ; ‡‡‡ Covariates used in the multiple imputation models include age (years), gender, treatment group (as received not allocated), presence of a carer (yes/no), severity of word finding (total score), severity of comprehension ability (total score), and baseline outcome measures under consideration;

Interpretation: [1] higher scores indicate improved personal vocabulary of personal importance; [2] higher scores indicate improved functional communication ability in conversation; [3] higher percentage score indicates improved patient perception of their communication and its impact on their quality of life.

7.8.6 Subgroup evaluation for the primary and key secondary outcomes at 6 months

The objective of this section is to explore the effect of the intervention among certain pre-specified subgroups on the change in word finding ability of personally selected words, change in functional communication, and change in COAST (%). These subgroups are:

1) the severity of word finding difficulty: mild (31 to 43), moderate (18 to 30), and severe (5 to 17);
2) Baseline comprehension ability based on the CAT sentence comprehension scores: within normal limits (27 to 32) mild (18 to 26), moderate (9 to 17), and severe (0 to 8);
3) the length of time post-stroke; there is no existing literature to guide its classification so categorisation will be distribution based – taking the underlying risk profile into account. A Trial Statistician blinded to treatment allocation will provide plotted data on the overall distribution of the primary outcomes and key secondary outcome against the length of time post-stroke to the TMG members blinded to the treatment allocation and trials results. Based on the observed relational distribution, the TMG will make a decision about cut-off points to categorise the length of time post-stroke for subgroup analyses. An alternative distribution based approach such as using quintiles or tertiles will be considered if the first approach is uninformative.

Analyses will be performed for these three pre-specified subgroups based on the mITT set. The number of participants and mean change in the word finding ability of personally selected words will be reported stratified by treatment received and subgroup category. Effect modification between the treatment group and subgroup will be assessed using a multiple linear regression model that includes an interaction term between the treatment and the subgroup of interest. Overall p-values from interaction tests will be used to explore the strength of evidence for heterogeneity. Results will be reported as shown in Table 10 to Table 12.

To aid visual interpretation, results will also be presented on a forest plot (Figure 4); for example, showing mean difference in change in word finding ability of personally selected words (adjusted for baseline and stratification factors) between the CSLT and UC, and CSLT and AC with its 95% CI grouped by subgroup category as shown in Table 10 to Table 12.

Similar analyses will be undertaken for the endpoints of the change in functional communication (measured using the TOMS) and the change in patient perception of communication and impact on quality of life (assessed using the COAST).

Table 10: Word finding severity subgroup analysis – co-primary & key secondary outcomes at 6 mths (mITT)

Outcomes at 6 months	Subgroup	UC		AC		CSLT		CSLT versus UC †		CSLT versus AC ‡	
		n	Mean(SD)	n	Mean(SD)	n	Mean(SD)	Adjusted MDC (95% CI)	Overall interaction p-value	Adjusted MDC (95% CI)	Overall interaction p-value
Co-primary:	**Word finding severity**										
change in word finding of personally selected words [1]	Mild	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Moderate	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Severe	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Change in functional communication [2]	Mild	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Moderate	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Severe	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Key secondary:											
Change in COAST (%) [3]	Mild	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Moderate	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Severe	xx	xx(xx)	xx	xx(xx)	Xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx

† UC as the reference group; ‡ AC as the reference group; UC: Usual care; AC: Attention control; CSLT: Computerised speech and language therapy; MDC: mean difference in change; SD: standard deviation

Interpretation: [1] higher scores indicate improved personal vocabulary of personal importance; [2] higher scores indicate improved functional communication ability in conversation; [3] higher percentage score indicates improved patient perception of their communication and its impact on their quality of life

Table 11: Comprehension ability subgroup analysis – co-primary & key secondary outcomes at 6 mths (mITT)

Outcomes at 6 months	Subgroup	UC		AC		CSLT		CSLT versus UC †		CSLT versus AC ‡	
		n	Mean(SD)	n	Mean(SD)	n	Mean(SD)	Adjusted MDC (95% CI)	Overall interaction p-value	Adjusted MDC (95% CI)	Overall interaction p-value
Co-primary:	**CAT comprehension ability**										
Change in word finding of personally selected words [1]	Within normal limits	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Mild	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Moderate	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Severe	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Change in functional communication [2]	Within normal limits	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Mild	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Moderate	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Severe	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Key secondary:											
Change in COAST (%) [3]	Within normal limits	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Mild	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Moderate	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Severe	xx	xx(xx)	xx	xx(xx)	Xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx

† UC as the reference group; ‡ AC as the reference group; UC: Usual care; AC: Attention control; CSLT: Computerised speech and language therapy; MDC: mean difference in change; SD: standard deviation

Interpretation: [1] higher scores indicate improved personal vocabulary of personal importance; [2] higher scores indicate improved functional communication ability in conversation; [3] higher percentage score indicates improved patient perception of their communication and its impact on their quality of life; CAT comprehension: severe (0 to 8) – inconsistently understanding at 2 ICW level; moderate (9 to 17) – consistently understanding at 2-3 ICW level/simple sentence structures but not complex sentence structures; mild (18 to 26) – some understanding of complex sentence structures but not consistent; within normal limits (27 to 32) based on CAT cut-off score for normal /aphasic.

Table 12: Time post-stroke subgroup analysis – co-primary & key secondary outcomes at 6 mths (mITT)

Outcomes at 6 months	Subgroup	UC		AC		CSLT		CSLT versus UC †		CSLT versus AC ‡	
		n	Mean(SD)	n	Mean(SD)	n	Mean(SD)	Adjusted MDC (95% CI)	Overall interaction p-value	Adjusted MDC (95% CI)	Overall interaction p-value
Co-primary:	**Time post-stroke**										
Change in word finding of personally selected words [1]	T1 (xx to xx)	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	….	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Tn (xx to xx)	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Change in functional communication (TOMS) [2]	T1 (xx to xx)	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	….	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Tn (xx to xx)	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Key secondary:											
Change in COAST (%) [3]	T1 (xx to xx)	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	….	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)		xx (xx to xx)	
	Tn (xx to xx)	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx

† UC as the reference group; ‡ AC as the reference group; UC=Usual care; AC: Attention control; CSLT: Computerised speech and language therapy; MDC: mean difference in change; SD: standard deviation

Interpretation: [1] higher scores indicate improved personal vocabulary of personal importance; [2] higher scores indicate improved functional communication ability in conversation; [3] higher percentage score indicates improved patient perception of their communication and its impact on their quality of life.

7.8.7 Assessment of the association between co-primary/key secondary outcomes and other variables

This section aims to explore whether a change in word finding of personally selected words, a change in functional communication (conversation), and a change in patient-rated communication effectiveness and impact on quality of life (COAST) outcomes at 6 months stratified by severity of word finding (mild, moderate, and severe) are associated with:

- time spent practising (total),
- the length of time post stroke (years),
- age (years)

These variables will be calculated as described in Section 7.2.1.

Locally weighted scatterplot smoothing (LOWESS) will be used for exploratory analyses. For example, change in word finding of personally selected words will be plotted against time spent practising stratified by treatment group. These scatter plots will be produced by the severity of word finding and presented side by side. The same approach will be used for change in functional communication and change in COAST (%). Total time spent practising will be based on the first 6 months from randomisation (approximately 183 days).

7.9 Analysis of additional secondary outcomes at 6 months

This section details the analysis of other secondary outcomes assessed at 6 months from baseline in order to explore the effect of the intervention further.

7.9.1 CaCOAST: carer rated patient communication and impact on carer's life

The impact of CSLT on carer perception of patient's communication effectiveness and impact on their quality of life at 6 months is another secondary outcome assessed using the first 15 items and last 5 items of the CaCOAST questionnaire, respectively. The computation of $CaCOAST_{15}$ (%) and $CaCOAST_5$ (%) summary measures used for the analysis for the first 15 items and the last 5 items, respectively is detailed in Section 7.2.5. Based on mITT population, the change in $CaCOAST_{15}$ (%) at 6 months from baseline based on the first 15 items (%) will be modelled using a multiple linear regression model adjusted for:

- baseline $CaCOAST_{15}$ (%),
- treatment group (UC, AC, and CSLT),
- centre as a fixed effect (random effect may be used depending on model fit),
- the severity of word finding as a fixed effect (mild, moderate, and severe).

The effect of the CSLT on the effect of patient's communication difficulties on the carer's life will be examined using the last 5 items of the COAST questionnaire. The analysis of change in the $CaCOAST_5$ (%) will be performed in a similar manner but adjusted for baseline $CaCOAST_5$ (%) in addition to stratification factors. Results will be reported as shown in Table 13.

A sensitivity analysis on the change in $CaCOAST_{15}$ (%) will be undertaken using a multiple linear regression model as a function of, and reported as shown in Table 14:

- baseline $CaCOAST_{15}$ (%),
- treatment group (UC, AC, and CSLT),
- centre as a fixed effect (random effect may be used depending on model fit),
- the severity of word finding as a fixed effect (mild, moderate, and severe),
- the length of time post stroke and,
- location of stroke (yes or no); MCA, frontal lobe, parietal lobe, and temporal lobe.

Likewise, a sensitivity analysis on the change in $CaCOAST_5$ (%) will be undertaken in a similar manner but adjusted for baseline $CaCOAST_5$ (%) rather than $CaCOAST_{15}$ (%), same stratification factors and additional a priori defined covariates. In case of marked residual baseline imbalances, additional covariates may be included in the model at the discretion of the research team

Table 13: Secondary results – change in CaCOAST at 6 mths (mITT)

Secondary outcome at 6 months	UC		AC		CSLT		CSLT versus UC †		AC versus UC †		CSLT versus AC ‡	
	n	Mean(SD)	n	Mean(SD)	n	Mean(SD)	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value
Change in $CaCOAST_{15}$ (%) [1]	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Change in $CaCOAST_{5}$ (%) [2]	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx

Results from multiple linear regression models adjusted for baseline measures and fixed stratification factors (centre and severity of word finding);

† UC as the reference group; ‡ AC as the reference group; UC: Usual care; AC: Attention control; CSLT: Computerised speech and language therapy; MDC: mean difference in change; SD: standard deviation

Interpretation: [1] higher percentage score indicates improved carer perception of patient's communication; [2] higher percentage score indicates a positive effect of patient's communication difficulties on the carer's life.

Table 14: Secondary results – change in CaCOAST at 6 mths adjusted for additional confounders (mITT)

Secondary outcome at 6 months	UC		AC		CSLT		CSLT versus UC †		AC versus UC †		CSLT versus AC ‡	
	n	Mean(SD)	n	Mean(SD)	n	Mean(SD)	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value
Change in $CaCOAST_{15}$ (%) [1]	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Change in $CaCOAST_{5}$ (%) [2]	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx

Results from multiple linear regression models adjusted for baseline measures, fixed stratification factors (centre and severity of word finding), and additional potential confounders (length of time post stroke and location of stroke);

† Usual care as the reference group; ‡ Attention control as the reference group; UC: Usual care; AC: Attention control; CSLT: Computerised speech and language therapy; MDC: mean difference in change; SD: standard deviation

Interpretation: [1] higher percentage score indicates improved carer perception of patient's communication; [2] higher percentage score indicates a positive effect on patient's communication difficulties on carer's life.

7.10 Analysis of the co-primary and key outcomes data at 9 and 12 months

Further exploratory analyses of the medium to long term effect of the intervention on the co-primary outcomes (change in word finding of personally selected words and change in functional communication) and a key secondary outcome (change in COAST (%)) assessed at 9 and 12 months from randomisation will be undertaken. The same statistical analysis approach using a multiple linear regression model adjusted for baseline measures and stratification factors as described in Sections 7.8.1 will be adopted. Analysis based on CC set will be the focus here, however, the impact of attrition will be explored using MI and LI approaches detailed in Section 7.16. The MI model will also take into account the longitudinal nature of the data using chained equations (van Buuren, 2007; Buuren and Oudshoorn, 2000) via Stata ***mi*** command. This multiple imputation using chained equations (MICE) analysis will be conducted and reported in accordance to the guidance provided by White et al (2011). Results will be reported as shown in Table 15 and Table 16. LOCF may be considered if profile plots over time suggest that the underlying assumptions are reasonable.

Table 15: Co-primary and key secondary outcomes result at 9 mths

Primary outcomes at 9 months	Analysis set	UC		AC		CSLT		CSLT versus UC †		AC versus UC †		CSLT versus AC ‡	
		n	Mean(SD)	n	Mean(SD)	n	Mean(SD)	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value
Co-primary*:*													
Change in word finding of personally selected words [1]	CC	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	MI ǂǂ	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	LI/MVI	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Change in functional communication (TOMS) [2]	CC	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	MI ǂǂ	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	LI/MVI	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Key secondary*:*													
Change in COAST (%) [3]	CC	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	MI ǂǂ	xx	xx(xx)	xx	xx(xx)	Xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	LI/MVI	xx	xx(xx)	xx	xx(xx)	Xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx

Results from multiple linear regression models adjusted for baseline measures and fixed stratification factors (centre and severity of word finding)

MI: Multiple imputation; LI: Linear Interpolation; MVI: Mean Value Imputation ǂǂǂ Covariates for multiple imputation models include age (years), gender, treatment group (as received not allocated), presence of a carer (yes/no), severity of word finding (total score), severity of comprehension ability (total score), and longitudinal outcome measures under consideration (baseline and/or 6 months)

† UC as the reference group; ‡ AC as the reference group; UC: Usual care; AC: Attention control; CSLT: Computerised speech and language therapy; MDC: mean difference in change; SD: standard deviation

Interpretation: [1] higher scores indicate improved personal vocabulary of personal importance; [2] higher scores indicate improved functional communication ability in conversation; [3] higher percentage score indicates improved patient perception of their communication and its impact on their quality of life.

Table 16: Co-primary and key secondary outcomes result at 12 mths

Primary outcomes at 12 months	Analysis set	UC		AC		CSLT		CSLT versus UC †		AC versus UC †		CSLT versus AC ‡	
		n	Mean(SD)	N	Mean(SD)	n	Mean(SD)	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value
Co-primary:													
Change in word finding of personally selected words [1]	CC	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	MI ‡‡	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	LI/MVI	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Change in functional communication (TOMS) [2]	CC	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	MI ‡‡	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	LI/MVI	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Key secondary:													
Change in COAST(%) [3]	CC	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	MI ‡‡	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	LI/MVI	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx

Results from multiple linear regression model adjusted for baseline measures and fixed stratification factors (centre and severity of word finding)

MI: Multiple imputation; LI: Linear Interpolation; MVI: Mean Value Imputation; ‡‡ Covariates for multiple imputation models include age (years), gender, treatment group (as received not allocated), presence of a carer (yes/no), severity of word finding (total score), severity of comprehension ability (total score), and longitudinal outcome measures under consideration (baseline, and/or 6, and/or 9 months);

† Usual care as the reference group; ‡ Attention control as the reference group; UC: Usual care; AC: Attention control; CSLT: Computerised speech and language therapy; MDC: mean difference in change; SD: standard deviation

Interpretation: [1] higher scores indicate improved personal vocabulary of personal importance; [2] higher scores indicate improved functional communication ability in conversation; [3] higher percentage score indicates improved patient perception of their communication and its impact on their quality of life.

7.11 Exploration of profile response of word finding ability with time

This section aims to graphically explore individual and mean profile of word finding ability of personally selected words during the course of the study. Exploration of the individual profile of word finding ability at baseline, 6, 9, and 12 months using spaghetti plots will be undertaken. This will be based on all randomised participants as follows:

- stratified by treatment group,
- by treatment group but the severity of word finding (mild, moderate, and severe) at baseline,
- CSLT treatment group only stratified by the continuation of the programme after 6 months (yes or no).

Furthermore, the average profiles of word finding ability of personally selected words will be graphically displayed showing the mean word finding ability at baseline, 6 months, 9 months, and 12 months. The following will be considered and repeated using the CC set and only participants with data at all timepoints:

- stratified by treatment group,
- In the CSLT group only, but stratified by the continuation of the programme after 6 months (yes or no).

7.12 Analysis of the CaCOAST: secondary outcomes at 9 and 12 months

For exploratory purposes, the effect of the intervention on the carer rated communication effectiveness ($CaCOAST_{15}$ (%)) and impact on carer's quality of life ($CaCOAST_5$ (%)) at 9 and 12 months after baseline will be evaluated using a multiple linear regression model adjusted for baseline measures and stratification factors as described in Section 7.9.1. No further adjustment for additional potential confounders will be undertaken. In addition, only CC population will be utilised for this exploratory analysis. Results will be reported as shown in Table 17.

Table 17: CaCOAST secondary outcomes at 9 and 12 mths (CC)

Secondary outcomes during follow-up	UC		AC		CSLT		CSLT versus UC †		AC versus UC †		CSLT versus AC ‡	
	n	Mean(SD)	n	Mean(SD)	n	Mean(SD)	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value
9 months												
Change in $CaCOAST_{15}$ (%) [1]	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Change in $CaCOAST_{5}$ (%) [2]	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
12 months												
Change in $CaCOAST_{15}$ (%) [1]	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Change in $CaCOAST_{5}$ (%) [2]	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx

Results from multiple linear regression models adjusted for baseline measures and fixed stratification factors (centre and severity of word finding);

† Usual Care as the reference group; ‡ Attention Control as the reference group; UC: Usual care; AC: Attention control; CSLT: Computerised speech and language therapy; MDC: mean difference in change; SD: standard deviation

Interpretation: [1] Higher percentage score indicates improved carer perceived communication effectiveness; [2] Higher percentage score indicates positive impact on carer's quality of life;

7.13 Analysis of generalisation of treatment at 6, 9, and 12 months

7.13.1 Generalisation to untreated words

This section aims to contribute to the debate on whether there is a generalisation of word finding ability from treated words (of personal importance) to untreated words. Generalisation to untreated words (word finding of untreated words) will be measured at baseline, 6, 9 and 12 months using the CAT Naming Test as detailed in Section 7.2.6.

The following approaches will be used to graphically display the mean profile in word finding of untreated words at baseline, 6, 9, and 12 stratified by treatment group:

a) Using all participants with available data on word finding of untreated words at different timepoints,

b) Using only participants with data on word finding of untreated words measured at all timepoints.

The effect of the intervention on word finding of untreated words will be assessed using a multiple linear regression model adjusted for baseline word finding of untreated words (scores) and stratification factors (centre and severity of word finding). The outcome will be the change in word finding of untreated words from baseline and analysed at 6, 9 and 12 months. In addition to the CC set, MI and LI/MVI will be used for sensitivity analysis to assess the influence of missing data. The results will be reported as shown in Table 18.

Table 18: Generalisation to untreated words at 6, 9 &12 mths

Change in word finding of untreated words	Analysis Set	UC		AC		CSLT		CSLT versus UC †		AC versus UC †		CSLT versus AC ‡	
		n	Mean(SD)	n	Mean(SD)	N	Mean(SD)	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value
6 months	CC	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	MI ⌗	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	LI/MVI	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
9 months	CC	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	MI ⌗	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	LI/MVI	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
12 months	CC	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	MI ⌗	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
	LI/MVI	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx

Results from multiple linear regression models adjusted for baseline measures and fixed stratification factors (centre and severity of word finding);

MI: Multiple imputation; LI: Linear Interpolation; ⌗ Covariates for multiple imputation models include age (years), gender, treatment group, presence of a carer (yes/no), severity of word finding (total score), severity of comprehension ability (total score), and longitudinal outcome measures under consideration (baseline, and/or 6, and/or 9 months);

† Usual care as the reference group; ‡ Attention Control as the reference group; UC: Usual care; AC: Attention control; CSLT: Computerised Speech and language therapy; MDC: mean difference in change; SD: standard deviation

Interpretation: higher change scores indicate improved generalisation of untreated words (general vocabulary).

7.13.2 Clinical word finding improvement in treated and untreated words

The objective of this section is to explore the effect of the intervention on the proportion of participants achieving 'clinical improvement' in both word finding of personally selected (treated) and untreated words. The following criteria have been a priori adopted by the research team to define clinical improvement:

a) change from baseline of ≥ 10% in both word finding of personally selected words (co-primary outcome) **AND** word finding ability of untreated words (scores from CAT Naming Test);
b) change from baseline of ≥ 5% in both word finding of personally selected words (co-primary outcome) **AND** word finding of untreated words (scores from CAT Naming Test). The rationale is to pick up small improvements, which may indicate some clinical benefit;
c) change from baseline of ≥ 10% and ≥ 5% in word finding of personally selected words (co-primary outcome) and word finding of untreated words (scores from CAT Naming Test), respectively.

Descriptive summaries of clinical improvement patterns in treated and untreated words under pre-defined thresholds will be reported as illustrated in Table 19 to Table 21

Table 19: Clinical improvement patterns in word finding of treated & untreated words at 6 mths

Change in word finding of personally selected words	Change in word finding of untreated words	UC (n=xx)	AC (n=xx)	CSLT (n=xx)
≥ 10%	**≥ 10%**			
Yes	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Yes	No	xx(xx%)	xx(xx%)	xx(xx%)
No	Yes	xx(xx%)	xx(xx%)	xx(xx%)
No	No	xx(xx%)	xx(xx%)	xx(xx%)
≥ 10%	**≥ 5%**			
Yes	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Yes	No	xx(xx%)	xx(xx%)	xx(xx%)
No	Yes	xx(xx%)	xx(xx%)	xx(xx%)
No	No	xx(xx%)	xx(xx%)	xx(xx%)
≥ 5%	**≥ 5%**			
Yes	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Yes	No	xx(xx%)	xx(xx%)	xx(xx%)
No	Yes	xx(xx%)	xx(xx%)	xx(xx%)
No	No	xx(xx%)	xx(xx%)	xx(xx%)

Table 20: Clinical improvement patterns in word finding of treated & untreated words at 9 mths

Change in word finding of personally selected words	Change in word finding of untreated words	UC (n=xx)	AC (n=xx)	CSLT (n=xx)
≥ 10%	**≥ 10%**			
Yes	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Yes	No	xx(xx%)	xx(xx%)	xx(xx%)
No	Yes	xx(xx%)	xx(xx%)	xx(xx%)
No	No	xx(xx%)	xx(xx%)	xx(xx%)
≥ 10%	**≥ 5%**			
Yes	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Yes	No	xx(xx%)	xx(xx%)	xx(xx%)
No	Yes	xx(xx%)	xx(xx%)	xx(xx%)

No	No	xx(xx%)	xx(xx%)	xx(xx%)
≥ 5%	**≥ 5%**			
Yes	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Yes	No	xx(xx%)	xx(xx%)	xx(xx%)
No	Yes	xx(xx%)	xx(xx%)	xx(xx%)
No	No	xx(xx%)	xx(xx%)	xx(xx%)

Table 21: Clinical improvement patterns in word finding in treated & untreated words at 12 mths

Change in word finding of personally selected words	Change in word finding of untreated words	UC (n=xx)	AC (n=xx)	CSLT (n=xx)
≥ 10%	**≥ 10%**			
Yes	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Yes	No	xx(xx%)	xx(xx%)	xx(xx%)
No	Yes	xx(xx%)	xx(xx%)	xx(xx%)
No	No	xx(xx%)	xx(xx%)	xx(xx%)
≥ 10%	**≥ 5%**			
Yes	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Yes	No	xx(xx%)	xx(xx%)	xx(xx%)
No	Yes	xx(xx%)	xx(xx%)	xx(xx%)
No	No	xx(xx%)	xx(xx%)	xx(xx%)
≥ 5%	**≥ 5%**			
Yes	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Yes	No	xx(xx%)	xx(xx%)	xx(xx%)
No	Yes	xx(xx%)	xx(xx%)	xx(xx%)
No	No	xx(xx%)	xx(xx%)	xx(xx%)

These exploratory outcomes are binary (yes or no) and will be modelled using an exact multiple logistic regression model only adjusted for stratification factors (centre and severity of word finding). The numbers and proportion of participants meeting each clinical improvement criterion will be reported by intervention group together with exact Odds Ratio (OR) and associated 95%CI and p-values, as shown in Table 22. This will be undertaken and presented based on two scenarios considering:

a) only participants with complete data on both outcomes and,
b) all randomised participants, but assuming that those with missing data failed to achieve clinical improvement (worst case scenario).

Table 22: Clinical improvement in word finding of treated & untreated words at 6, 9 &12 mths (CC)

Change in word finding of treated and untreated words	**UC**	**AC**	**CSLT**	**CSLT versus UC †**		**AC versus UC †**		**CSLT versus AC ‡**	
	(n=xx)	(n=xx)	(n=xx)	Exact OR (95% CI)	P-value	Exact OR (95% CI)	P-value	Exact OR (95% CI)	P-value
6 months									
≥ 10%	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
≥ 5%	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
≥ 10% & ≥ 5% respectively	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
9 months									
≥ 10%	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
≥ 5%	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
≥ 10% & ≥ 5% respectively	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
12 months									
≥ 10%	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
≥ 5%	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
≥ 10% & ≥ 5% respectively	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx

Results from exact multiple logistic regression models adjusted for stratification factors (centre and severity of word finding);

† UC as the reference group; ‡ AC as the reference group; OR: Odds Ratio; UC: Usual care; AC: Attention control; CSLT: Computerised speech and language therapy.

Table 23: Clinical improvement in word finding of treated & untreated words at 6, 9 &12 mths (worst case)

Change in word finding of untreated and untreated words	**UC** (n=xx)	**AC** (n=xx)	**CSLT** (n=xx)	**CSLT versus UC †** Exact OR (95% CI)	P-value	**AC versus UC †** Exact OR (95% CI)	P-value	**CSLT versus AC ‡** Exact OR (95% CI)	P-value
6 months									
≥ 10%	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
≥ 5%	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
≥ 10% & ≥ 5% respectively	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
9 months									
≥ 10%	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
≥ 5%	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
≥ 10% & ≥ 5% respectively	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
12 months									
≥ 10%	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
≥ 5%	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
≥ 10% & ≥ 5% respectively	xx(xx%)	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx

Results from exact multiple logistic regression models adjusted for stratification factors (centre and severity of word finding);

† UC as the reference group; ‡ AC as the reference group; OR: Odds Ratio; UC: Usual care; AC: Attention control; CSLT: Computerised speech and language therapy

7.13.3 Generalisation of treated words used in conversation

This section explores the effect of the intervention on the use of learnt vocabulary (treated words finding retrieved) in conversation, which is calculated from video conversations (out of the 100 treated words) and captured in the database, as described in Section 7.2.7. The secondary outcomes of interest are the change in the number of treated words retrieved in conversation from baseline assessed at 6, 9, and 12 months. The outcome will be modelled using a multiple linear regression model as a function of:

- baseline number of words retrieved,
- treatment group (UC, AC, and CSLT),
- centre as a fixed effect (a random effect may be used depending on model fit),
- the severity of word finding as a fixed effect (mild, moderate, and severe),

Only CC set will be used for this exploratory analysis and results will be reported as shown in Table 24.

The research team also wants to explore the effect of having a volunteer or assistant on the generalisation of treated words used in conversation only for participants in the computer arm. Access to a volunteer or assistant will be classified according to the following approaches:

- binary outcome (yes/no) – yes if a participant had access to a volunteer or assistant for a minimum of 4 visits (including scheduled and unscheduled visits);
- count outcome – the number of times/visits a participant had a volunteer or assistant.

For both cases, the distribution of change in the number of words retrieved during a conversation will be displayed graphically displayed stratified by access to volunteer or assistant, and summarised appropriately depending on the observed distribution.

Table 24: Generalisation of untreated words used in conversation (CC)

Change in the number of untreated words retrieved	**UC**		**AC**		**CSLT**		**CSLT versus UC †**		**AC versus UC †**		**CSLT versus AC ‡**	
	n	Mean(SD)	n	Mean(SD)	n	Mean(SD)	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value	Adjusted MDC (95% CI)	P-value
6 months	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
9 months	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
12 months	xx	xx(xx)	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx

† Usual care as the reference group; ‡ Attention Control as the reference group; UC: Usual care; AC: Attention control; CSLT: Computerised speech and language therapy; MDC: mean difference in change; SD: standard deviation

Interpretation: higher scores indicate improved use of learnt vocabulary (untreated words) in conversation

7.13.3.1 Patient and Carer COAST

A graphical display using spaghetti plots showing the patient COAST scores at baseline, 6 months, 9 months and 12 months will be produced by treatment group. The average Carer COAST scores (last 5 questions) at baseline, 6 months, 9 months and 12 months will also be displayed graphically by treatment group.

In the pilot study, the participants and carers were interviewed about the benefits of the computer therapy. Seven themes emerged, that can be mapped on to eight items of the COAST are shown in Table 25.

Table 25: Themes and relation to the COAST items for subgroup exploration

Interview theme	Related COAST item for sub-analysis
Improved Word finding	Item 3 – chat with someone you know well Item 4 – Short conversation with unfamiliar person
Improved Conversation	Item 3 – chat with someone you know well Item 4 – Short conversation with unfamiliar person
Improved sentence production	Item 6 – make self-understood in longer sentences
Improved spelling	Item 11 – how well can you write
Increased confidence	Item 15 – confidence
Increased participation	Item 16 – family Item 17 – social life
Increased wellbeing	Item 20 – quality of life

Each of the above items (items 3, 4, 6, 11, 15, 16, 17, 20) will be analysed using a bootstrapping procedure (Efron, 1979). The median (IQR) by the treatment group will be calculated and the median difference in rating with its 95% CI will be estimated. The aim is to provide quantitative data on similar outcomes as identified in the qualitative data previously published (Palmer et al., 2013). The analysis will be performed using both patient and Carer COAST and for the 6, 9 and 12-month timepoints. Participants for whom Items are coded as not applicable, unclear or no response will not be included in the analysis of those items. Results will be presented as shown in Table 26 and Table 27. P-values will not be reported.

Table 26: COAST items for subgroup exploration at 6, 9 &12 mths.

COAST	UC		AC		CSLT		CSLT versus UC	CSLT versus AC
	n	Median (IQR)	n	Median (IQR)	n	Mean (IQR)	Median Difference (95% CI)	Median Difference (95% CI)
Item 3	xx	xx(xx to xx)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx (xx to xx)	xx (xx to xx)
Item 4	xx	xx(xx to xx)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx (xx to xx)	xx (xx to xx)
Item 6	xx	xx(xx to xx)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx (xx to xx)	xx (xx to xx)
Item 11	xx	xx(xx to xx)	Xx	xx(xx to xx)	xx	xx(xx to xx)	xx (xx to xx)	xx (xx to xx)
Item 15	xx	xx(xx to xx)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx (xx to xx)	xx (xx to xx)
Item 16	xx	xx(xx to xx)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx (xx to xx)	xx (xx to xx)
Item 17	xx	xx(xx to xx)	Xx	xx(xx to xx)	xx	xx(xx to xx)	xx (xx to xx)	xx (xx to xx)
Item 20	xx	xx(xx to xx)	Xx	xx(xx to xx)	xx	xx(xx to xx)	xx (xx to xx)	xx (xx to xx)

Table 27: CaCOAST items for subgroup exploration at 6, 9 & 12 mths

CaCOAST (%)	UC		AC		CSLT		CSLT versus UC	CSLT versus AC
	n	Median (IQR)	n	Median (IQR)	n	Mean (IQR)	Median Difference (95% CI)	Median Difference (95% CI)
Item 3	xx	xx(xx to xx)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx (xx to xx)	xx (xx to xx)
Item 4	xx	xx(xx to xx)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx (xx to xx)	xx (xx to xx)
Item 6	xx	xx(xx to xx)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx (xx to xx)	xx (xx to xx)
Item 11	xx	xx(xx to xx)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx (xx to xx)	xx (xx to xx)
Item 15	xx	xx(xx to xx)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx (xx to xx)	xx (xx to xx)
Item 16	xx	xx(xx to xx)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx (xx to xx)	xx (xx to xx)
Item 17	xx	xx(xx to xx)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx (xx to xx)	xx (xx to xx)
Item 20	xx	xx(xx to xx)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx (xx to xx)	xx (xx to xx)

7.14 Computer usage and association with change in communication ability

This section aims to contribute to the debate about the level of therapy practice required to achieve a good outcome. The additional exploratory analysis will be undertaken to look at the association between change in communication ability and pattern of practice in the CSLT intervention group. It is recommended that the participants use the computer for word finding practice for 20-30 minutes per day. Data on the total time used (based on the first 6 months from randomisation), the number of sessions used for/how many times a week and length of time per session will be collected. Descriptive statistics will be performed on the data so that the mean, median and range for each parameter will be given for each individual and for the participants as a group.

Graphs of computer use (focusing on the total time used) against percentage change in word finding ability will be produced stratified by severity.

7.15 Safety outcomes

Safety analysis will be based on the population defined in Section 6.3.5.

7.15.1 AEs and SAEs

AEs will be reported as number and percentage of patients overall and compared between treatment groups but no formal statistical analysis is planned. The following summaries will be presented:

1. The number and percentage of participants reporting at least one AE,
2. The number and percentage of participants reporting at least one SAE,
3. The number and percentage of participants reporting a treatment-related AE.

It should be noted that the number of events may depend on the patient's follow-up time. Therefore, for a fairer treatment comparison, it is important to account for follow-up time when reporting AEs and SAEs summaries by treatment group. For each patient, the exposure in months will be calculated as ((last known study follow-up date-date of randomisation)/365.25)*12. The incidence of AEs and SAEs will be reported per treatment group as an average number of events experienced per patient per month. A Poisson or Negative Binomial

regression model will be used with follow-up time as an offset to account for repeated events depending on the observed distribution of these events.

A listing of all AEs will also be presented and will include the following information:

- Description/Site/Signs and Symptoms
- Severity
- Relationship
- Action taken
- Outcome
- Seriousness

7.15.2 Negative effects of computer use

Any perceived negative effects of the intervention, as recorded in the patient diaries on monthly basis, will also be reported. This will only happen in the computer therapy arm. The following negative effects data will be reported regarding effects on;

1. tiredness,
2. vision,
3. headaches,
4. anxiousness or worrisome.

For each participant, the total number of repeated events experienced per negative effect category and number of months contributing to the data will be used to calculate the average number of repeated events per negative event category per month. The distribution of the average number of repeated events per negative events per month per patient will be graphically displayed and summarised appropriately depending on the observed distribution.

7.16 Handling missing data

Missing data patterns will be examined visually. For the repeated measures of the 100 personal words, the scores from 24 non-treatment words and the score from the activity scale of the TOMs, the following imputation methods will be considered and the analyses repeated as a sensitivity analysis:

a) By deterministic imputation (linear interpolation), where the method depends on the type of missing data:

Where data are missing at a time point t_i but valid data are available at previous (t_{i-1}) and future (t_{i+1}) time points, the missing value will be linearly interpolated by the formula

$$y_{missing} = y_{i-1} + (y_{i+1} - y_{i-1})\frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}}$$

Where all data are missing from a time point t onwards, imputation will be by last observation carried forward, i.e. using data from the previous time point $t - 1$.

b) By multiple imputation, in which baseline covariates are among predictors of the missing data. Multiple imputation will only be undertaken on the co-primary and key secondary outcomes. Age, gender, treatment group (as received and not allocated), the presence of a carer, severity of word finding (total score), and severity of comprehension ability (total score) and baseline primary outcome measure under consideration will be mandatory covariates for the imputation models. Other covariates may be included in the imputation models at the discretion of the Trial Statistician depending on the results from Section 7.5.

The multiple imputation will be undertaken as a sensitivity analysis to investigate the robustness of the primary analysis. Twenty datasets will be imputed and combined as follows (Yuan, 2010):

- ✓ To calculate the point estimate of parameter Q (e.g. mean difference in % change in word finding), let $\hat{Q}_i$ be the point estimate of the ith imputed dataset, $i = 1, 2, \dots, 20$. The point estimate for Q is the average of the 20 imputed dataset estimates:

$$\bar{Q} = \frac{1}{20}\sum_{i=1}^{20} \hat{Q}_i.$$

- ✓ To calculate the variance estimate associated with $\bar{Q}$ we first calculate the within and between imputation variances. The within-imputation variance (W) is the average of the 20 imputed dataset variances:

$$W = \frac{1}{20}\sum_{i=1}^{20} \widehat{W}_i,$$

where $\widehat{W}_i$ is the variance estimate of the ith imputed dataset, $i = 1, 2, \dots, 20$. The between-imputation variance (B) is given by

$$B = \frac{1}{19}\sum_{i=1}^{20} (\hat{Q}_i - \bar{Q})^2.$$

The variance estimate associated with $\bar{Q}$ is the total variance given by

$$T = W + \left(1 + \frac{1}{20}\right) B.$$

The number of imputations required may be changed depending on the observed percentage of missing data (White et al., 2011).

7.16.1 Model Diagnostics

7.16.1.1 Co-primary and key secondary outcomes

For the co-primary and secondary outcomes analyses that are based on multiple linear regression models, the assumptions will be assessed by inspection of residual plots. Homogeneity of variance will be assessed by plotting the studentised residuals against the predicted values from the model, whilst Normality will be assessed using Normal probability plots. If the assumptions for the analysis of covariance are violated either an appropriate transformation will be applied or a non-parametric procedure with less stringent assumptions will be utilised.

7.16.1.2 Safety outcomes

The ratio of the mean and variance of the outcome data under consideration with the aid of graphical plots (such as histograms) will be used to investigate overdispersion. If the results suggest the existence of overdispersion which violates the assumption of the Poisson regression model, then a Negative Binomial regression model will be used to model repeated count outcome data to account for overdispersion (Agresti, 2007)

7.17 Implementation of the SAP

This SAP will be used as a work description of the Trial Statistician in consultation with the Senior Statistician. No analysis will be undertaken until after the sign-off of this SAP by relevant personnel. Any economic related analysis will be the responsibility of the Health Economist. Data will be released by the data management group after sign-off by the relevant personnel to the Trial Statistician. There will be a period of data cleaning in order to query any spurious data and initiate final data lock before the conduct of the analysis. At this point, no changes will be allowed on the database. The Trial Statistician will produce an unblinded report for the DMEC members and the Senior Statistician. A concurrent blinded report (that uses a dummy version of the real treatment codes) will be supplied to the Chief Investigator in order to identify any anomalies or issues that need to be rectified and investigated further. Any additional further exploratory analyses can also be suggested by the Chief Investigator based on the blinded results. Once all relevant parties (Trial Statistician and Senior Statistician) are agreed on the final analyses and the contents of the unblinded report, the unblinded report can then be shared with the Chief Investigator and other members of the TMG following appropriate quality control measures.

8 Appendix: Tables and figures to aid presentation and interpretation of results

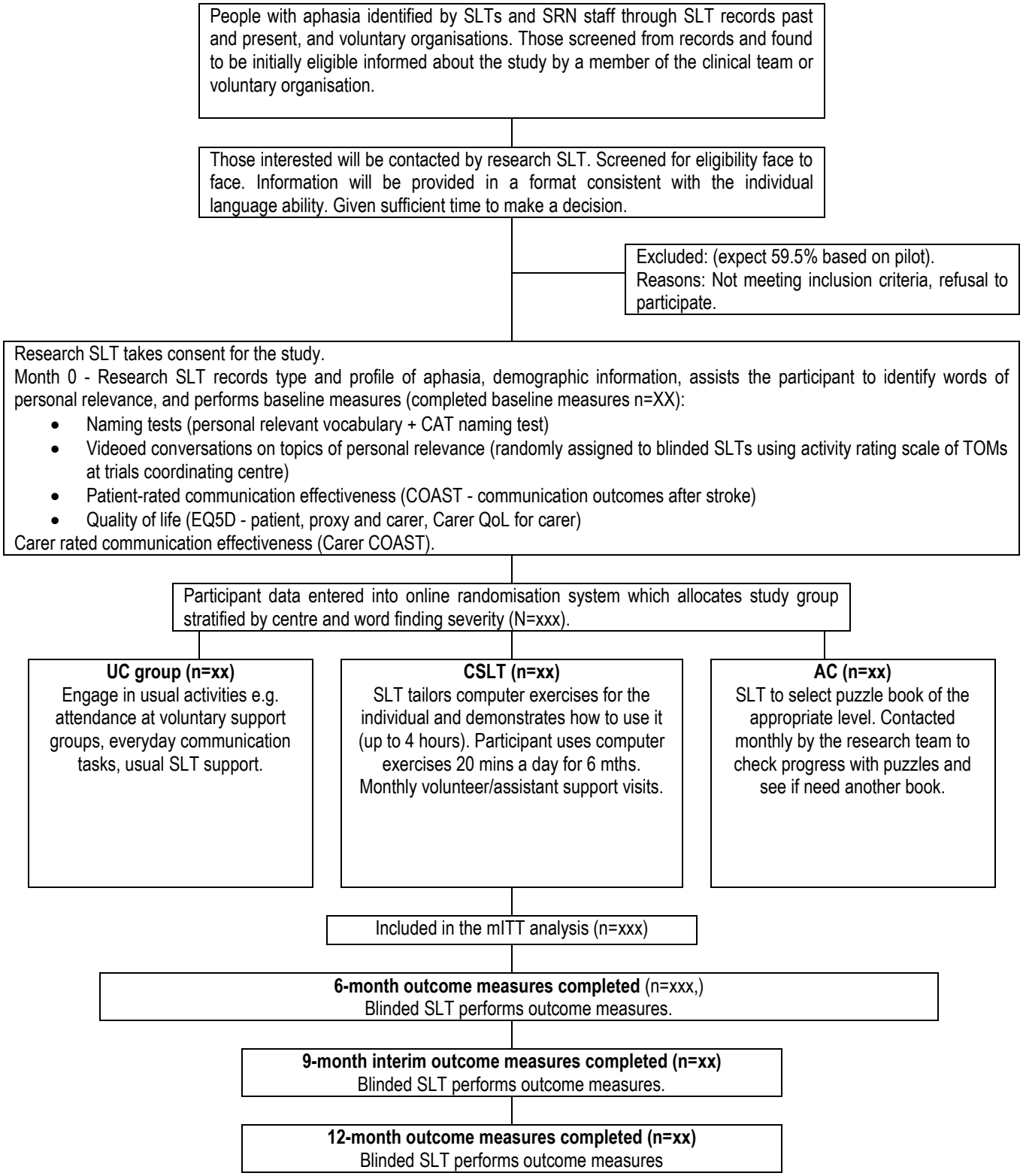

Figure 2: CONSORT flowchart of trial participants through the trial

Note: this will also include reasons for withdrawal and deaths at each follow-up.

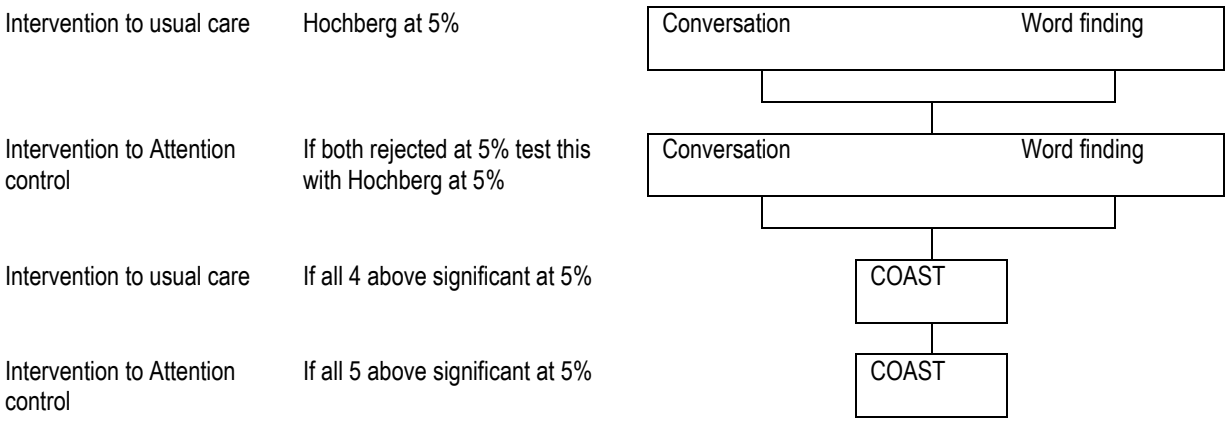

Figure 3: Hochberg testing procedure diagram

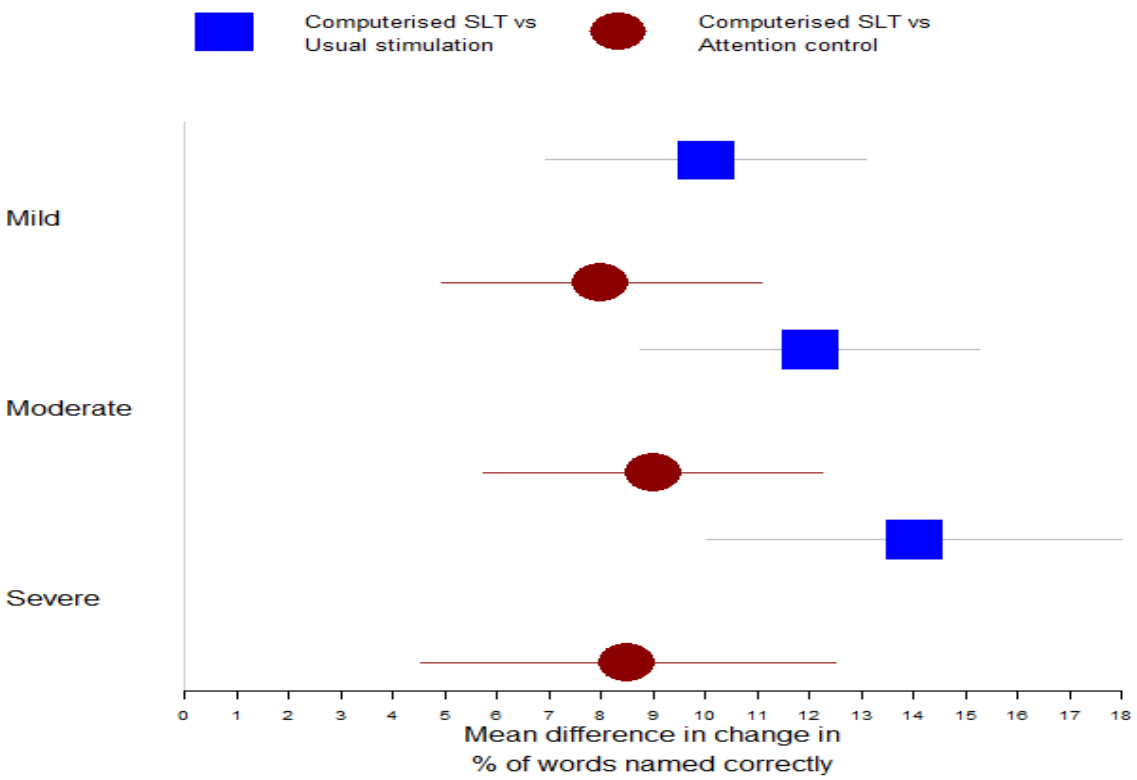

Figure 4: Example of a forest plot

Table 28: Hochberg testing procedure

	Conversation	**Word finding**
Intervention to usual care	P = x.xxx	P = x.xxx
*if and only if **both** significant at 5% then…*		
	Conversation	**Word finding**
Intervention to attention control	P = x.xxx	P = x.xxx
*if and only if **both** significant at 5% then…*		
	COAST	
Intervention to usual care	P = x.xxx	
if and only if significant at 5% then…		
	COAST	
Intervention to attention control	P = x.xxx	

Table 29: Example 1 of the Hochberg multiple testing procedure

	Conversation	**Word finding**
Intervention to usual care	P = 0.04	P = 0.03
Both conversation and word finding significant at 5%		
	Conversation	**Word finding**
Intervention to attention control	P = 0.1	P = 0.01
Conversation not significant at 5%. Word finding significant at 2.5%.		
	COAST	
Intervention to usual care	Test not performed since the conversation was not significant at the 5% level. Conclude that conversation and word finding significant when comparing the intervention to usual care, and word finding significant when comparing the intervention to attention control.	
	COAST	
Intervention to attention control	Not performed.	

Table 30: Example 2 of the Hochberg multiple testing procedure

	Conversation	**Word finding**
Intervention to usual care	P = 0.02	P = 0.07
Conversation significant at 2.5%. Word finding not significant at 5%.		
	Conversation	**Word finding**
Intervention to attention control	Test not performed since word finding was not significant at the 5% level. Conclude that conversation is significant when comparing the intervention to usual care.	
	COAST	
Intervention to usual care	Not performed.	
	COAST	
Intervention to attention control	Not performed.	

Table 31: Example 3 of the Hochberg multiple testing procedure

	Conversation	**Word finding**
Intervention to usual care	P = 0.04	P = 0.03
Both conversation and word finding significant at 5%.		
	Conversation	**Word finding**
Intervention to attention control	P = 0.02	P = 0.01
Both conversation and word finding significant at 5%.		
	COAST	
Intervention to usual care	P = 0.2	
COAST comparing the intervention to usual care not significant at 5%.		
	COAST	
Intervention to attention control	Test not performed since previous test not significant at 5%.	

9 References

Agresti, A. (2007), "Generalised Linear Models", *An Introduction to Categorical Data Analysis*, 2nd Ed., John Wiley & Sons, Inc, pp. 65–98.

de Boer, M.R., Waterlander, W.E., Kuijper, L.D.J., Steenhuis, I.H.M. and Twisk, J.W.R. (2015), "Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate.", *The International Journal of Behavioral Nutrition and Physical Activity*, Vol. 12 No. 1, p. 4.

Bowen, A., Hesketh, A., Long, A. and Patchick, E. (2009), "Scoring instructions for COAST and CaCOAST", available at: http://www.click2go.umip.com/i/coa/coast.html (accessed 8 August 2016).

Bowen, A., Hesketh, A., Patchick, E., Young, A., Davies, L., Vail, A., Long, A., et al. (2012), "Clinical effectiveness, cost-effectiveness and service users' perceptions of early, well-resourced communication therapy following a stroke: a randomised controlled trial (the ACT NoW Study).", *Health Technology Assessment*, Vol. 16 No. 26, pp. 1–160.

Breitenstein, C., Grewe, T., Flöel, A., Ziegler, W., Springer, L., Martus, P., Huber, W., et al. (2017), "Intensive speech and language therapy in patients with chronic aphasia after stroke: a randomised, open-label, blinded-endpoint, controlled trial in a health-care setting", *The Lancet*, Vol. 389 No. 10078, pp. 1528–1538.

van Buuren, S. (2007), "Multiple imputation of discrete and continuous data by fully conditional specification", *Statistical Methods in Medical Research*, Sage PublicationsSage UK: London, England, Vol. 16 No. 3, pp. 219–242.

Buuren, S. van and Oudshoorn, C.G.M. (2000), "Multivariate Imputation by Chained Equations : Mice V1.0 User's manual", TNO.

Devlin, N., Shah, K., Feng, Y., Mulhern, B. and Hout, B. van. (2016), "Valuing health-related quality of life : An EQ-5D-5L value set for England", *Office of Health Economics*, Vol. 16 No. 1, pp. 1–22.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife", *The Annals of Statistics*, Institute of Mathematical Statistics, Vol. 7 No. 1, pp. 1–26.

Frison, L. and Pocock, S.J. (1992), "Repeated measures in clinical trials: Analysis using mean summary statistics

and its implications for design", *Statistics in Medicine*, Wiley Subscription Services, Inc., A Wiley Company, Vol. 11 No. 13, pp. 1685–1704.

Hochberg, Y. and Tamhane, A.C. (1987), *Multiple Comparison Procedures*, edited by Hochberg, Y. and Tamhane, A.C., John Wiley & Sons, Inc., Hoboken, NJ, USA, available at:http://doi.org/10.1002/9780470316672.

ICH. (1998), *ICH E9: Guidance on Statistical Principles for Clinical Trials*.

ICH. (2005), *HARMONISED TRIPARTITE GUIDELINE: GUIDELINE FOR GOOD CLINICAL PRACTICE E6(R1)*.

Julious, S. (2004), "Letter to the Editors", *Biometrics*, Blackwell Publishing, Vol. 60 No. 1, pp. 284–284.

Julious, S.A. and Owen, R.J. (2006), "Sample size calculations for clinical studies allowing for uncertainty about the variance", *Pharmaceutical Statistics*, John Wiley & Sons, Ltd., Vol. 5 No. 1, pp. 29–37.

Kahan, B.C. and Morris, T.P. (2012), "Improper analysis of trials randomised using stratified blocks or minimisation.", *Statistics in Medicine*, Vol. 31 No. 4, pp. 328–40.

Katz, R.C., Hallowell, B., Code, C., Armstrong, E., Roberts, P., Pound, C. and Katz, L. (2000), "A Multinational Comparison Of Aphasia Management Practices", *International Journal of Language & Communication Disorders*, Vol. 35 No. 2, pp. 303–314.

Long, A., Hesketh, A. and Bowen, A. (2009), "Communication outcome after stroke: a new measure of the carer's perspective", *Clinical Rehabilitation*, SAGE PublicationsSage UK: London, England, Vol. 23 No. 9, pp. 846–856.

Long, A., Hesketh, A., Paszek, G., Booth, M. and Bowen, A. (2008), "Development of a reliable self-report outcome measure for pragmatic trials of communication therapy following stroke: the Communication Outcome after Stroke (COAST) scale.", *Clinical Rehabilitation*, Vol. 22 No. 12, pp. 1083–94.

Palmer, R. (2015), *Big CACTUS Protocol Version 4.0*, Sheffield.

Palmer, R., Enderby, P., Cooper, C., Latimer, N., Julious, S., Paterson, G., Dimairo, M., et al. (2012), "Computer Therapy Compared With Usual Care for People With Long-Standing Aphasia Poststroke: A Pilot Randomized Controlled Trial", *Stroke*.

Palmer, R., Enderby, P. and Paterson, G. (2013), "Using computers to enable self-management of aphasia therapy exercises for word finding: the patient and carer perspective", *International Journal of Language & Communication Disorders*, Vol. 48 No. 5, pp. 508–521.

Pocock, S.J., Assmann, S.E., Enos, L.E. and Kasten, L.E. (2002), "Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems.", *Statistics in Medicine*, Vol. 21 No. 19, pp. 2917–30.

van Reenen, M. and Janssen, B. (2015), "EQ-5D-5L User Guide: Basic information on how to use the EQ-5D-5L instrument", available at: http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-5L_UserGuide_2015.pdf (accessed 15 February 2017).

Schulz, K.F., Altman, D.G. and Moher, D. (2010), "CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials.", *Annals of Internal Medicine*, Vol. 152 No. 11, pp. 726–32.

Senn, S. (1994), “Testing for baseline balance in clinical trials.”, *Statistics in Medicine*, Vol. 13 No. 17, pp. 1715–26.

Sheffield CTRU. (2016a), *GV001: Trial Management Group*.

Sheffield CTRU. (2016b), *GV002: Trial Steering Committee*.

Sheffield CTRU. (2016c), *DM003: Case Report Form Development*.

White, I.R., Royston, P. and Wood, A.M. (2011), “Multiple imputation using chained equations: Issues and guidance for practice”, *Statistics in Medicine*, John Wiley & Sons, Ltd., Vol. 30 No. 4, pp. 377–399.

Yuan, Y.. (2010), “Multiple Imputation for Missing Data: Concepts and New Development (Version 9.0)”, *SAS Institute Inc*.

1)