Southampton NHS
University Hospitals NHS Trust

**PREVENTION OF MORBIDITY IN SICKLE CELL DISEASE PHASE II:**
**POMS 2b – Statistical Analysis Plan**

# Prevention of morbidity in sickle cell disease Phase II (Improvement of Pain and Quality of Life in Patient with Sickle Cell Disease with Auto-adjusting Continuous Positive Airways Pressure: Phase II) (POMS 2b): Child Cohort

ISRCTN46012373

TOC

# Contents

# 1. INTRODUCTION

In addition to pain, sickle cell anaemia (HbSS) complications include neurocognitive difficulties in attention and processing speed associated with low daytime and night-time oxygen saturation. These effects can be compounded by obstructive sleep apnoea (OSA). Continuous Positive Airways Pressure (CPAP) is an accepted treatment for OSA in the general population. The aim of this single-blind, randomised, controlled phase II trial is to compare Auto-adjusting CPAP (APAP) with standard care to standard care alone in subjects with HbSS to determine whether the intervention improves attention and processing speed, brain structure, pain and quality of life.

Eligibility criteria include ability to provide informed consent, age ≥ 8 years, diagnosis of HbSS and mean overnight saturation of <90% for <30% of the night (i.e. not meeting current criteria for overnight oxygen therapy). Key exclusion criteria are overnight respiratory support, respiratory or decompensated cardiac failure, chronic transfusion or contra-indications to APAP therapy or MRI.

## 1.1 Trial Objective

To determine whether Auto-adjusting CPAP (APAP) improves outcomes compared to standard of care.

### 1.1.1 Primary Objectives

The aim of this Phase II trial is to compare Auto-adjusting Continuous Positive Airways Pressure (APAP) with standard care in patients with Sickle Cell Anaemia (SCA) to determine whether the intervention improves Cancellation, a measure of attention and processing speed.

### 1.1.2 Major Secondary Objectives

- To assess whether APAP therapy has any effect on cognition, pain intensity, and/or quality of life in children and adult SCD patients
- To assess whether there are any physiological effects of APAP therapy e.g. on daytime oximetry, brain MRI, and laboratory investigations
- To assess side effects of APAP and safety, specifically with respect to haematological investigations and hospitalisations

- To assess the feasibility of using smart phone technology to collect information on site and severity of pain daily for 2 week periods.
- To identify main cost drivers and potential cost implications of providing the intervention.

### 1.1.3 Exploratory

To determine whether Auto-adjusting CPAP (APAP) has any impact on biomarkers, sleep and cardiac investigations. Also to determine whether biomarkers are related to SCD outcomes.

## 1.2 Trial Design

Sixty subjects with HbSS (30 children and 30 adults) will be randomised to standard care + APAP or standard care alone for 6 months. Minimisation factors are age group (8-11, 12-15, 16-22 and >23 years), silent infarction on MRI, minimum overnight oxygen saturation ≥ or < 90%, and hydroxyurea use. For APAP subjects, the intervention will be administered at home. Adherence and effectiveness will be recorded using software documenting hours of use each night and overnight oximetry. Participant support in terms of appropriate facemask and facilitating adherence will be provided by an unblinded sleep physiologist. According to the trial visit schedule, endpoints are evaluated through 6 months after randomisation.

### 1.2.1 Statistical Hypotheses for Trial Objectives

#### *1.2.1.1 Primary Hypothesis*

**$H_0$:** Mean change in cancellation from baseline to 6 months is equal for the APAP and control groups

*Versus*

**$H_A$:** Mean change in cancellation from baseline to 6 months is not equal for the APAP and control groups

#### *1.2.1.2 Secondary Outcomes*

Secondary and exploratory outcomes are described in their respective sections below.

### 1.2.2 Sample Size Justification

This study consists of two separate and independent cohorts. The first cohort will consist of 30 children (aged 8-16) randomised 1:1 to either APAP or standard of care. The second cohort will consist of adults (≥16 years) randomised 1:1 to either APAP or standard of care. The child cohort represents an independent trial with 90% power to detect a difference in cancellation rate at a significance level of 0.05. Since the cohorts are independent and each has a type I error allocation of 5%, no adjustments for multiple comparisons for two cohorts are needed. Any analyses pooling the child and adult cohorts will be considered exploratory. The calculation of the sample size for the cohort of children is described below. In the absence of pilot data for adults, the effect size was assumed to be similar in the adult cohort. As a result, assessment of the impact of APAP in the child cohort is considered the primary goal of this study, and the adult cohort is of secondary interest.

In the previous pilot study in children, the results of the cancellation task within each subject group was normally distributed with standard deviation 2, resulting in a mean difference for the scaled score between APAP treatment and standard care at six weeks of 2.6, 95% CI (0.83, 4.3). Based on this finding, 24 evaluable children will be sufficient to determine a difference in the primary outcome of cancellation of 2.4 between APAP-treated and control subjects with 90% power and at a significance level of 0.05. The calculation assumes a moderate correlation of 0.5 between the baseline and follow-up measures. If the correlation is higher the study will be powered to detect a smaller difference.

Allowing for 20% withdrawal/loss-to-follow-up, a sample size of 30 children divided equally between the two groups will have 90% power to detect a difference of 2.3 points in cancellation between the two groups using an ANCOVA analysis to adjust for the baseline cancellation measure (or 80% power to detect a difference of 2.0).

A cohort of 30 adults will be similarly recruited to allow comparisons of the outcomes in older subjects, and a separate SAP will describe the analyses for this older cohort.

### 1.2.3 Randomisation and Blinding

POMS-2b follows a single-blind strategy in which the outcome assessments are performed by blinded researchers. In order to maintain this blind, the data management team will also remain blinded. The Chief Investigator, local principal investigators and coordinators will be unblinded. Central randomisation will be implemented in this study via Sealed Envelope. Randomisation will be minimised by age range, presence or absence of silent cerebral infarction on MRI, minimum oxygen saturation ≥ or <90%, and hydroxyurea use.

## 2. GENERAL ANALYSIS DEFINITIONS

The intervention period will be defined as beginning on the day of randomisation and ending 90 days after the 6-month outcome assessment. Age will be calculated based on the date of randomisation for cohort descriptions and based on the day of the evaluation for neuropsychological and other measurements and tests.

### 2.1 Visit Windows

The baseline period will include all visits and records up to and including the day of randomisation. Randomisation occurs after all other Visit 2 data are collected, so information collected on that date should not be influenced by randomisation assignment. When multiple records on or prior to randomisation are available, the latest one (closest to randomisation) should be defined as the baseline record. Records after the randomisation date may not be counted as baseline records.

The visit 9 (month 6) window will extend from day 152 (5 months) until day 272 (6 months + 90 days). The primary analyses will include only records within this window, but endpoint analysis including all available visit 9 records will also be performed for comparison.

## 2.2 Analysis Sets

The following analysis sets will be defined:

### 2.2.1 Efficacy Analysis Set

Intention to treat will be used to define the efficacy analysis set. Subjects will be analysed according to the group to which they were randomized, and all randomized subjects will be included.

### 2.2.2 Per-Protocol Analysis Sets

The per-protocol analysis set will be the subgroup of the efficacy analysis excluding subjects randomised to APAP who did not use APAP for at least 4 hours on the night prior to the visit 9 cognitive outcomes assessments. There may be both a short-term and longer-term impact of APAP on cognitive function, and the purpose of this analysis set is to determine whether the results are similar between this group and the Efficacy Analysis Set.

Adequate compliance is defined as use of the APAP for at least 4 hours a night and at least 16 nights per month. Select outcomes may be analysed including all placebo subjects and restricted to APAP subjects with adequate compliance. These analyses, though specified a priori, should be considered exploratory and are underpowered.

### 2.2.3 Safety Analysis Set

The safety analysis set will include all subjects who complied with randomised treatment at least once during the study period. Subjects who used APAP at least once during the study period will be analysed with the APAP group, and subjects who did not use APAP during the study will be analysed with the control group.

### 2.2.4 Missing Data

Missing data will not be imputed for the primary analysis. However, the value of missing scores necessary to change the inference from the primary outcome will be determined and reported. Only baseline and visit 9 cancellation will be collected, so longitudinal models that accommodate data that are missing at random cannot be used. Mixed models for repeated measures (MMRM) will be used to analyse parameters with more than 2

collection points. These models assume missing data are missing at random, and they are more statistically robust than imputation via last-observation-carried-forward.

Each PRO measure provides guidelines about the number of elements that can be missing for total scores and subscale scores to be valid. Missing values within PROs will be addressed according to the recommendations of the scoring manuals.

## 1. SUBJECT INFORMATION

Descriptive statistics (N, mean, SD, median, and range) will be calculated by treatment group for continuous variables and counts and percentages will be calculated by treatment group for categorical variables.

### 1.1 Baseline Anthropometric and Demographic Characteristics

Age, gender, post-code based deprivation index, body mass index, heart rate, oximetry, respiratory rate, and systolic and diastolic blood pressure will be summarised.

### 1.2 Baseline Disease Characteristics

Minimisation factors will be tabulated (age range, presence or absence of silent cerebral infarction on MRI, minimum oxygen saturation ≥ or <90%, and hydroxyurea use). In the case that there are discrepancies between the actual values of the minimisation factors and the values used by the randomisation system, the actual values will be used. Results of the baseline physical and neurological examination will also be tabulated.

### 1.3 Medical History

Medical history characteristics including number of hospital days, A&E visits, missed school days, MRC dyspnoea scale, cough, chest crisis, asthma diagnosis, strokes, seizures and the frequency of headaches and nighttime urination will be tabulated for each treatment group.

### 1.4 Disposition Information

The number of subjects who complete and discontinue the study will be tabulated. The reasons for and the timing of study termination will also be summarised.

### 1.5 APAP Exposure and Compliance

The proportion of possible on-study nights when APAP was used for at least 4 hours will be calculated for each subject in the APAP group and the cumulative distributions of proportion of nights with at least 4 hours of use will be produced. On-study nights will be defined as nights beginning on the date of randomisation and ending on the date of the visit 9 neuropsychological evaluation.

Adequate adherence will be defined as use of the APAP for at least 4 hours a night and at least 16 nights per month. The proportion of APAP subjects with adequate adherence and 95% confidence intervals will be produced.

### 1.6 Protocol Deviations

Protocol deviations (compliance, visit delays, missed visits etc.) will be summarised.

### 1.7 Prior and Concomitant Medications

Baseline drug use will be tabulated for individual drugs based on standardised drug terms. On-study drug changes (additions, discontinuations and dose changes) will be summarised.

## 2. EFFICACY

### 2.1 Data Handling Rules

Valid efficacy records are baseline records that occur on or prior to the date of randomisation, and valid endpoint records will include the latest assessment at least 150 and no more than 272 days from randomisation.

### 2.2 Primary Efficacy Endpoint: Cancellation

The primary analysis of change in cancellation scaled score will be based on change in cancellation scores from baseline to visit 9. Scaling is provided for 3-month intervals of age. The scaling transformation is non-linear in age, and

in a small study, randomisation may not prevent imbalances by treatment for subjects at "steep" or "flat" spots in the transformation from raw scores to scaled scores. In other words, the 6 month age change during the trial might make a big difference in the scaling factor for some subjects and a smaller difference for others; the magnitude of the change may not be balanced across randomised groups. Thus, scores for the primary endpoint will be scaled to baseline age.

### 2.2.1 Analysis Methods

The primary analysis of change in cancellation will be based on an ANCOVA model containing the baseline cancellation scaled score and minimisation factors as fixed covariates. A sensitivity analysis excluding the minimisation factors will also be performed.
The ANCOVA model will be parameterized as follows:

$$E[Y|x_1, x_2, x_3, x_4, x_5, x_6] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

Where Y is the change in cancellation from baseline to visit 9, $x_1$ is the baseline cancellation scaled score, $x_2$ is Baseline oximetry minimum overnight saturation <90% or ≥ 90% (i.e., $1_{[\min O_2\ Sat.<90]}$), $x_3$ is current prescription of hydroxyurea versus no prescription ($1_{[current\ hydroxyurea]}$), $x_4$ is infarction on brain MRI versus no infarction ($1_{[infarction]}$), $x_5$ is age 8-11 versus 12-15 ($1_{[age\ 8-11]}$), and x6 is randomisation to active treatment with APAP versus placebo ($1_{[APAP]}$). Correct values of minimisation factors will be used (in the event that values used in the minimisation were erroneous). Least-square mean change and 95% confidence intervals will be reported for each treatment group.

### 2.2.2 Supportive Efficacy Analyses and Methods

Two strategies will be used for supportive analyses. The first will be based on scaling the cancellation scores using the age at the time of the testing, and the second will be based on the raw cancellation scores with no scaling.

### 2.2.3 Subgroup Analyses for the Primary Endpoint

Pre-specified subgroups include each of the minimisation factors and the subgroup of subjects with adequate APAP compliance on the night prior to cognitive testing. Subgroups by geographic area (postcode deprivation index) will also be performed. An analysis using tertiles of % nights with adequate compliance (with indicators for tertile and control group as the reference group) and baseline cancellation score will also be used. If empirical tertiles do not represent clinically meaningful compliance differences, no models will be fit. Models including variables for above or below the median hemoglobin, reticulocytes, and $SpO_2$ will also be analysed. These subgroups will be evaluated using ANCOVA models, adjusted for baseline cancellation score.

## 2.3 Secondary Endpoints and Analysis Methods

Secondary endpoints are described below. In general, continuous changes from baseline to visit 9 will be analysed with ANCOVA models similar to the primary outcome. Binary outcomes will be assessed using logistic regression models, including baseline values when appropriate, treatment assignment and the minimisation factors as fixed effects. Longitudinal outcomes will be analysed using mixed models for repeated measures with baseline, treatment, and minimisation factors included as fixed covariates. An unstructured correlation will be used, but compound symmetry may be used in the case that the model does not converged with the empirical correlation estimates.

### 2.3.1 Neuropsychology

Neuropsychological assessment will be conducted at baseline and visit 9 using the Wechsler Abbreviated Intelligence Scale (UK edition, WASI), Wechsler Intelligence Scale for Children (UK 4th edition; WICV-IV UK), Delis-Kaplan Executive Function System (Sorting & Tower tests; D-KEFS), Behavior Rating Inventory of Executive Function (BRIEF), Childrens Memory Scale (UK edition; CMS), and the Conners Continuous Performance Test (3rd Edition; CPT)

The following secondary outcomes will be analysed using ANCOVA as described in section 2.3:

- Executive function will be assessed by comparing the mean change in D-KEFS Tower total achievement score from baseline to 6 months;
- Verbal memory will be assessed by comparing the mean change in CMS word pairs delayed recall score from baseline to 6 months;
- Reaction time will be assessed by comparing the mean change in CPT hit reaction time from baseline to 6 months.

For all cognitive secondary endpoints, associations with haematological, oximetry and post-code based socio-economic measures will be explored in subsequent analyses for hypothesis generation.

### 2.3.2 Pain

#### *2.3.2.1 Pain burden*

This monthly questionnaire consists of 7 questions about frequency of pain with 5 ordered categorical response options (None, A Few, Some, Many, Everyday). At each visit, for each of the 7 items, the proportion of subjects with improvement, worsening and no change compared to baseline will be calculated across treatment groups and compared using the chi-squared test. Logistic regression, adjusted for stratification factors, will be used to calculate the odds ratio for improvement at visit 9 compared to baseline and for no worsening (improved + no change) at visit 9 compared to baseline.

#### *2.3.2.2 Pain outcomes*

The pain outcomes questionnaire was scheduled as 2, 14-day series. One series was prior to randomization (baseline series) and the other was at visit 9 (endpoint series). The questionnaire collects the following information each day:

- Presence or absence of pain at 65 sites across the body
- Adjectives to describe pain
- Worst pain (0 – 10 scale)
- Least pain (0 – 10 scale)
- Average pain (0 – 10 scale)
- Current pain (0 – 10 scale)

Feasibility of smartphone app data collection will be calculated as the percentage of subjects who completed 14 days of the diary at each of the collection points. Similar calculations will be performed for subjects who used

paper diaries. The target is for 85% of subjects to complete pain diary collection as instructed, and for 95% to have at least 7 sequential measurements at each time-point.

The baseline series will be defined as "valid" for analysis if there are at least 7 days of pain diaries that occur prior to randomization with no more than 48 hours of "gaps" in the measurements. The endpoint series will be valid if there are at least 7 days of pain diaries that occur after study day 152 with no more than 48 hours of "gaps" in the measurements. All recorded measurements prior to randomization and after day 152 will be used (e.g., if there are 15 pain diaries prior to randomization, all will be used). The differences from baseline to visit 9 in the within-subject average number of pain sites, average of the pain measurements (worst, least, average and current) and the difference in proportion of days with any pain will be compared using ANCOVA models as described in section 2.3.

Adjectives to describe pain will be classified as describing neuropathic worst, least, average, current) will be calculated for the baseline and endpoint series. Changes from the baseline mean to the endpoint mean will be analysed using similar models to the primary outcome, nociceptive and/or affective pain.

- Neuropathic pain adjectives: aching, stabbing, numb, shooting, pricking, burning, penetrating, radiating
- Nociceptive pain adjectives: sharp, squeezing, throbbing, gnawing, tender, dull, cramping
- Affective/evaluative: exhausting, nagging, unbearable, tiring, miserable, radiating, deep

The difference in the proportion of days from baseline to visit 9 reporting neuropathic, nociceptive, and affective pain will be compared across treatment groups.

### 2.3.3 Quality of Life

*2.3.3.1 EuroQol five dimensions questionnaire – 5 level (EQ-5D-5L)*

EuroQol five dimensions questionnaire (EQ-5D-5L) index values will be derived using the UK value set[i]. Mean change from baseline to visit 9

will be analysed for both the EQ-5D-5L index and the visual analog scale (VAS) using a similar model to the primary outcome.

#### *2.3.3.2 Child Health Utility (CHU-9D)*

The Child Health Utility 9D consists of nine dimensions (worried, sad, pain, tired, annoyed, schoolwork/homework, sleep, daily routine and activities), each represented by a single item with five response options. The scoring assigns a value to each response option, for example for the worried item: I don't feel worried today (1), I feel a little bit worried today (2), I feel a bit worried today (3), I feel quite worried today (4), I feel very worried today (5). The score from these items can be mapped to a value index[ii]. Baseline score and change from baseline will be analysed using similar models to the primary outcome.

#### *2.3.3.3 Epworth Sleepiness Scale (ESSC)*

The Epworth Sleepiness Scale[iii] consists of 7 items with 4-level categorical responses. It is designed to capture 'daytime sleepiness', and asks how likely patients are to dose or sleep in specific situations. Points are assigned for each item response as follows:

Would never doze or sleep (0 points)
Slight chance of dozing or sleeping (1 points)
Moderate chance of dozing or sleeping (2 points)
High chance of dozing or sleeping (3 points)

The total score is summed over the 7 items. Scoring can be interpreted as follows:

0-5 Lower Normal Daytime Sleepiness
6-10 Higher Normal Daytime Sleepiness
11-12 Mild Excessive Daytime Sleepiness
13-15 Moderate Excessive Daytime Sleepiness
16-24 Severe Excessive Daytime Sleepiness

Mean change from baseline to visit 9 will be compared using ANCOVA models as described in section 2.3.

#### *2.3.3.4 PedsQL Sickle Cell Disease Model*

PedsQL includes 9 subscales (Pain and hurt subscale, Pain impact subscale, Pain management subscale, Worry I subscale, Worry II subscale, Emotions subscale, Treatment subscale, Communication I subscale, Communication II subscale) and a total score. The scoring manual[iv] describes how to convert raw items into these scores. Essentially, items are reversed scored and linearly transformed to a 0-100 scale as follows: 0=100, 1=75, 2=50, 3=25, 4=0. Subscales are calculated as the mean of non-missing items as long as at least 50% of items in the subscale are non-missing. Similarly, the total score is the average over all items as long as at least 50% of the total items are non-missing. Mean change in total score from baseline to visit 9 will be compared using ANCOVA as described in section 2.3.

### 2.3.4 Daytime Oximetry

Daytime oxygen saturation was collected at 4 visits, and will be assessed by comparing the mean change in $SpO_2$ from baseline to visit 9 using MMRM model as described in section 2.3.

### 2.3.5 Brain MRI

The MRI protocol for quantitative analysis includes:

- 3D T1-weighted sequence for volumetric analysis and segmentation.
- Multishell diffusion weighted sequence for voxel-wise white matter analysis and whole-brain tractography. Metrics will be derived by fitting various models (i.e. diffusion tensor imaging; DTI, neurite orientation dispersion and density imaging; NODDI, and spherical mean technique; SMT).
- Arterial Spin Labelling (ASL) sequence for non-invasive measurement of perfusion.

MRI results will be assessed in two ways.

1. Qualitative change in radiological status will be recorded (i.e. change from normal study, new or enlarged SCI, change in vasculopathy, and/or any other clinically significant change from baseline; see safety).
2. Quantitative imaging analyses will be conducted as follows;

In secondary analyses:

- Diffusion-weighted imaging will be assessed by comparing the change in change in mean radial diffusivity from baseline to 6 months using ANCOVA as described in section 2.3.
- Structural imaging will be assessed by comparing the mean change in hippocampal volume from baseline to 6 months, also using ANCOVA models.

In exploratory analyses:

- A longitudinal TBSS design will be implemented to show change in white matter metrics between treatment groups and correlations with change in haematological or cognitive outcomes;
- Global and regional cerebral blood flow maps will be calculated to show differences between treatment groups and correlations with change in haematological or cognitive outcomes;
- Volumetric segmentations of surface area, cortical thickness and grey matter volume will be calculated at each time point and differences will be compared between treatment groups as well as correlations with change in haematological or cognitive outcomes. Analyses where possible will be described in terms of changes and analysed using ANCOVA models adjusted for the trial minimisation factors.

## 2.4 Other Endpoints

### 2.4.1 Cardiac Investigations

#### *2.4.1.1 Echocardiography*

No formal analysis of echocardiography variables will be performed.

#### *2.4.1.2 Cardiac MRI*

No formal analysis of cardiac MRI variables will be performed.

#### *2.4.1.3 6 minute walk*

No formal analysis of the 6 minute walk test will be performed.

### 2.4.2 Sleep

#### *2.4.2.1 Sleep Habits*

The sleep habits screener has no formal total score, but collects information about usual sleep/wake schedule, napping, sleep environment, sleep symptoms, snoring, morning waking and daytime symptoms, previous airway surgeries, and family sleep history. Baseline and visit 9 values will be tabulated by treatment, and changes in symptoms from baseline to visit 9 will be summarised. Sleep habits information may be used as precision variables for exploratory analyses of the impact of APAP.

#### *2.4.2.2 Epworth Sleepiness*

The Epworth Sleepiness Scale[v] (ESS) consists of 8 items with 4 response options. The scoring algorithm assigns a numeric score to each response option as follows: No chance of dozing (0), Slight chance of dozing (1), Moderate chance of dozing (2), High chance of dozing (3). The points are then summed across the 8 items to arrive at a total score. The interpretation of the total score is:

> 0-7: It is unlikely that you are abnormally sleepy.
>
> 8-9:You have an average amount of daytime sleepiness.
>
> 10-15:You may be excessively sleepy depending on the situation. You may want to consider seeking medical attention.
>
> 16-24:You are excessively sleepy and should consider seeking medical attention.

The change from baseline to visit 9 will be analysed using a similar model to the primary endpoint.

### 2.4.3 Biomarkers

Exploratory analyses will be conducted of based on biomarkers of

1. Renal function, including ADMA, SDMA and NGAL, as a hypoxia biomarker
2. Erythropoiesis including soluble transferrin receptor and erythropoietin levels.
3. Inflammation

Analyses of these examinations may not be available prior to database lock, in which case the results should be considered exploratory.

### 2.5 Multiplicity Adjustment

No adjustments will be made for multiple outcomes, and all analyses beyond the primary ANCOVA model for the primary outcome should be interpreted with caution and in the context of hypothesis generation due to the potential for Type I error inflation.

## 3. SAFETY

### 3.1 Adverse Events

Adverse events reported between randomisation date and day 272 (study end + 90 days) will be considered treatment-emergent and included in study reporting. The verbatim terms for reported adverse events will be standardized, and events will be tabulated for each group. In addition, the incidence of any adverse event, any serious adverse event, and any unexpected adverse event (see protocol section 6.11.3.5.4 Expectedness for a list of expected adverse events). Number of hospitalisations, number of hospital days, number of A&E visits by treatment will also be tabulated.

#### 3.1.1 Anaemia

Incidence of anaemia will be tabulated and compared across groups using Fisher's Exact test. Normal haemoglobin is >65 g/dL; a fall of >20g/l from baseline means that there is significant acute anaemia. If more than 25% of subjects experience anemia, Kaplan-Meier curves for time to the first incidence of anemia may be produced.

#### 3.1.2 Evidence of Bone Marrow Suppression

Evidence of bone marrow suppression will defined as change in reticulocytes and will be tabulated and compared across groups using Fisher's Exact test. Normal absolute reticulocyte count is 10-100 $x10^9$; a fall $<10x10^9$ is clinically significant.

### 3.2 Clinical Laboratory Tests

Change over time in laboratory measures will be analysed using MMRM as described in section 2.3. In particular, anaemia severity will be assessed by comparing the mean change in haemoglobin, bone marrow production will be assessed by comparing the mean change in reticulocyte count over time. In addition, subjects reaching the following clinical thresholds will be tabulated in each group:

- Normal Lactate Dehydrogenase is <800 IU/L; an increase of >1.5x from baseline is clinically significant
- Normal total Billirubin is <200 µmol/L; an increase of 1.5x from baseline is clinically significant
- Normal Creatinine is <104 µmol/L; Increase of >1.5 from baseline is clinically significant.

Normal urinary albumin: creatinine ratio in children will be summarised based on mean change over time.

Hepatic function will be assessed by comparing the mean change in total bilirubin over time, renal function will be assessed by comparing the mean change in (i) albumin: creatinine ratio and (ii) creatinine over time, and hemolysis will be assessed by comparing the mean change in (i) lactate dehydrogenase (ii) bilirubin and (iii) reticulocytes over time.

### 3.3 Brain MRI/MRA

Any additional strokes, new silent infarcts or change in vasculopathy will be tabulated by treatment arm.

## 4. RESOURCE USE

The main aspect of the health economic analysis within the Phase II trial will be to monitor resource use and costs with a view to identify the main cost drivers and inform the design of the economic measures within the subsequent definitive trial. Detailed information on all resources required for both arms, including family born costs will be collected. Therefore the study will adopt a societal perspective, including costs from the health provider's (NHS) perspective. Resource use data will be collected using a modified version of the CSRI (Client Service Receipt Inventory) questionnaire. The modified CSRI questionnaire

adapted to meet the needs of the population in this study will be used to collect detailed information from all patients and parents of patients (children) indicating contacts with frontline professionals or specialised services at primary care settings as well as inpatient admissions and Accident and Emergency or Day Unit attendances. It will also be used to identify family borne costs, e.g. for analgesic medication, travel to hospital, food bought in hospital, utility bills, help with childcare or housework etc. and productivity losses and/or out of school days for children. Secondary care resource use will be collected from hospital records. Resources identified will be costed using appropriate local and national unit cost data to provide estimates of the cost incurred during the intervention.

The preference-based generic Health Related Quality of Life (HRQoL) measures CHU-9D for children and EQ-5D for adults will be used to enable estimation of Quality Adjusted Life Years (QALYs) for both arms. The sensitivity of these measures will be compared to other disease specific outcomes within the trial. Missing data will be examined as to whether cases with missing data are similar to those with full economic data in a similar manner to the main trial analysis.

Descriptive statistics will be presented for main cost outcomes and HRQoL measures in terms of QALYs, at each occasion inferences will be made and conclusions will be drawn allowing identification of the best way of measuring and collecting data for the full cost-effectiveness analysis within the phase III trial.

Resource use will be monitored within the Phase II trial with a view to identify the main cost drivers and inform the design of economic measures within the subsequent definitive trial. We will use the CHU-9D in children and the Euroqol EQ-5D in adults, before randomisation and 3 and 6 months after randomisation, to identify changes and determine whether scores vary with other outcome measures. The resources required to provide treatment will be recorded in terms of therapy resources, equipment and travel etc. We will also investigate whether, as a result of the intervention, use of other resources are affected, e.g. requirement for transfusion, opioids, other pain relief and hospital admissions. Resources identified will be costed using appropriate local and national cost data to provide a preliminary estimate of the cost of overnight respiratory support compared to standard care. Descriptive statistics will be presented, for quality of life measures in QALY terms and main cost drivers, at each occasion inferences

will be made on identifying the most appropriate way of measuring and collecting data for the economic analysis of SCD in adults and children for application within any Phase III trial. Results from this study will be used on identifying the most appropriate way of measuring and collecting data for the cost per QALY analysis of SCD in adults and children for application within the definitive trial.

[i] Devlin, N., Shah, K., Feng, Y., Mulhern, B. and van Hout, B., 2016. Valuing Health-Related Quality of Life: An EQ-5D-5L Value Set for England. OHE Research Paper 16/01. London: Office of Health Economics

[ii] The Development of a Preference Based Paediatric Health Related Quality of Life Measure for use in Economic Evaluation. Katherine Stevens. PhD Thesis 2008. The University of Sheffield

[iii] Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep. 1991 Dec;14(6):540-5.

[iv] Scaling and Scoring of the Pediatric Quality of Life Inventory™ (PedsQL)

[v] Johns MW. A new method for measuring daytime sleepiness: The Epworth Sleepiness Scale. Sleep 1991; 14(6):540-5

## Signatures

| | Print Name | Sign | Date |
|---|---|---|---|
| CCTU Statistician: | SIMON SKENE | Simon Skene. | 31-JUL-2017 |
| Trial Statistician: | April Skea | April Skea | 31 July 2017 |
| Trial CI: | Fenella Kirkham | Fenella Kirkham | 31 July 2017 |