**A randomised controlled trial comparing palate surgery at six months versus twelve months of age (the TOPS trial): a Statistical Analysis Plan**

**Running head:** Analysis plan timing for cleft palate repair trial

**Keywords:** randomised controlled trial, complex intervention; surgical intervention; trial design; trial analysis; statistical analysis plan; palatal surgery; unilateral cleft palate; Sommerlad technique; velopharyngeal function.

**Manuscript information**

Number of figures: 1

Number of colour figures: 0

Number of tables: 6

Number of words: 3887

Supplementary materials: 1

**Authors**

Mrs Elizabeth J Conroy (EJC) MSc. (0000-0003-4858-727X) ejconroy@liverpool.ac.uk 1 \*

Ms Rachael Cooper (RC) MSc. (0000-0002-4741-6818) rachael.cooper@liverpool.ac.uk 1

Prof. William Shaw (WS) (0000-0002-3208-6000) bill.shaw@manchester.ac.uk 2

Assoc. Prof. Christina Persson (CP) PhD. (0000-0002-7860-1573)christina.persson@neuro.gu.se 2, 3

Prof. Elisabeth Willadsen (EW) PhD. (0000-0001-9013-8764) elisabethw@hum.ku.dk 2, 4

Prof. Kevin J Munro (KM) PhD. (0000-0001-6543-9098) kevin.munro@manchester.ac.uk 5, 6

Prof. Paula R Williamson (PRW) PhD. (0000-0001-9802-6636) p.r.williamson@liverpool.ac.uk 1

Dr Gunvor Semb (GS) gunvor.semb@manchester.ac.uk 2

Tanya Walsh (TW) tanya.walsh@manchester.ac.uk 2

Prof. Carrol Gamble (CG) PhD. (0000-0002-3021-1955) carrolp@liverpool.ac.uk 1

On behalf of the TOPS trial management group.

1 Liverpool Clinical Trials Centre, University of Liverpool, a member of Liverpool Health Partners, Liverpool, UK.

2 School of Medical Sciences, Division of Dentistry, University of Manchester, UK.

3 Institute of Neuroscience and Physiology, Speech and Language Pathology Unit, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.

4 Department of Nordic Studies and Linguistics, University of Copenhagen, Copenhagen, Denmark.

5 Manchester Centre for Audiology and Deafness, School of Health Sciences, The University of Manchester, Manchester, Greater Manchester, UK.

6 Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK.

**\*Details of corresponding author and person to be contacted for reprint requests**

*Name:* Elizabeth J Conroy

*Postal Address:* Liverpool Clinical Trials Centre, University of Liverpool, Institute of Child Health, Alder Hey Children’s NHS Foundation Trust, Liverpool, L12 2AP

*Tel:* +44 151 795 8791

*Email address:* ejconroy@liverpool.ac.uk

**ABSTRACT**

**Background**

Cleft palate is among the most common birth abnormalities. The success of primary surgery in the early months of life is crucial for successful feeding, hearing, dental development and facial growth. Over recent decades, age at palatal surgery in infancy has reduced. The Timing Of Primary Surgery for cleft palate (TOPS) trial aims to determine whether, in infants with cleft palate, it is better to perform primary surgery at age six or twelve months (corrected for gestational age).

**Methods/design**

The TOPS trial is an international, two-arm, parallel group, randomised controlled trial. The primary outcome is insufficient velopharyngeal function at five years of age. Secondary outcomes, measured at twelve months, three years and five years of age, include growth, safety of the procedure, dentofacial development, speech development, hearing level and middle ear function. The analysis approaches for primary and secondary outcomes are described here, as are the descriptive statistics which will be reported. The TOPS protocol has been published previously.

**Discussion**

This paper provides details of the planned statistical analyses for the TOPS trial and will reduce the risks of outcome reporting bias and data driven results.

**Trial registration**

ClinicalTrials.gov Identifier: NCT00993551. Registered on 9th October 2009.

**MANUSCRIPT**

**Background**

Clefts of the lip and/or palate are among the most common birth anomalies, occurring with an incidence of 1 in 600 births. [*Shaw, 2019*] The timing of palatal surgery has been a controversial issue since the 1930s. [*Peterson-Flazone, 2006*] Traditionally, rationale for delaying hard palate surgery was partly based on the belief that postponing the trauma of palatal closure may reduce maxillary growth disturbance. However, there is little evidence that facial skeletal growth in individuals with isolated cleft palate is substantially affected by different surgical protocols, though maxillary arch form, especially transversely, may be affected. [*Hellquist, 1978; Friede, 1993; Nystrom, 1994; Friede, 2000*]

Over recent decades, the age at which palatal surgery is carried out has reduced. This has led to one-stage palatal closure within twelve months of age at cleft units in Europe and the USA. Protagonists of early closure of the palatal cleft have proposed that since speech is a learnt behaviour, the sooner an intact anatomy is created, the better. [*Dorf, 1982; Chapman, 2008; Chapman, 2011; Jones, 2016*] As yet, however, there is no evidence that early surgery would lead to better speech development.

The Timing of Primary Surgery for cleft palate (TOPS) trial is an international, two-arm, parallel group, randomised controlled trial designed to determine whether, in infants with isolated cleft palate, it is better to perform primary surgery at age six or twelve months (corrected for gestational age). This research will investigate the effect of the timing of surgery by assessing and comparing speech development outcomes measured across twelve months, three years and five years of age. In addition, secondary outcomes include growth, perioperative complications, dentofacial development, hearing level and middle ear function. The protocol paper for the TOPS trial has been published previously [*Shaw et al, 2019*]; the aim of this paper is to report in detail the statistical analysis plan. This paper has been prepared according to the published guidelines on the content of statistical analysis plans. [*Gamble, 2017*]

**Methods and design**

***Trial design***

TOPS is an international, multi-centre trial using a parallel arm design aiming to detect whether surgery at six months is superior to surgery at twelve months. Infants with a diagnosis of cleft palate are randomised to receive primary surgery for cleft palate using a standardised technique (the Sommerlad technique [*Sommerlad,* 2003]) at either six months or twelve months (corrected for gestational age). Eligible patients are randomised on a 1:1 basis using minimisation routine, incorporating a random element to reduce predictability, to balance the two groups by surgeon (n=24) and size of cleft (soft palate only vs. soft and hard palate). The nature of the intervention prevented this trial from being blind to participants or their carers. However, speech outcomes, at ages twelve months, three and five years, will be rated blind to the randomly allocated group by independent assessment of speech recordings taken at visit. The primary outcome is assessed at age five years with secondary outcomes assessed 48 hours and 30 days post surgery and at age twelve months, three years and five years. Full details of the trial design, study population and study procedures have been published previously. [*Shaw, 2019*]

The trial is registered with the ClinicalTrials.gov Identifier: NCT00993551 (registered: 9th October 2009).

***Objectives***

The primary objective is to determine whether surgery for cleft palate, using the Sommerlad technique, at age six months when compared to surgery at age twelve months improves velopharyngeal function at age five years. Secondary research objectives include whether timing of surgery improves speech development, safety of the procedure, hearing level, middle ear function, dentofacial development and growth.

***Outcomes***

*Primary outcome*

The primary outcome is defined as a dichotomous outcome of whether the child has been perceived by Speech and Language Therapists (SLTs), following independent review of speech recordings, to have insufficient velopharyngeal function at age five years or not. Velopharyngeal insufficiency is measured by Velopharyngeal Composite Score (VPC) sum, which is a sum of scores, based on three components: hypernasality, non oral errors and velopharyngeal insufficiency (VPI) symptoms. Each component is classified and each classification mapped on to a score, see *Table 1*. The sum of the three scores, see *Equation 1*, give the VPC sum on the scale 0-6. [*Lohmander, 2017*] Scores ≥4 on this scale will be considered insufficient.

Table 1: Calculating the VPC sum

|  |  |  |
| --- | --- | --- |
| Component | Classification | Score for component |
| Hypernasality | Within normal limits | 0 |
|  |  | Mild resonance | 1 |
|  |  | Moderate/severe resonance | 2 |
| Non oral errors | 0-2 errors | 0 |
|  |  | 3-5 errors | 1 |
|  |  | ≥6 errors | 2 |
| Velopharyngeal insufficiency symptoms | 0-2 symptoms | 0 |
|  |  | 3-5 symptoms | 1 |
|  |  | ≥6 symptoms | 2 |

Equation 1: Using the three component scores to calculate VPC sum

$$VPC sum=Hypernasality score+Active non oral errors score+VPI symptoms score$$

*Secondary outcomes*

Secondary outcome measures are defined in the following list. Outcomes 1 to 5 are a measure of speech development, which are classified by SLTs following independent review of speech recordings. Outcomes 6, 7, 8, and 10 are a measure of safety of the procedure, hearing level, middle ear function, and growth respectively and are measured at the relevant follow up visits. Outcome 9 is a measure of dentofacial development measured independently on a profile photograph and maxillary arch impression taken during the five year follow up visit.

*1. Velopharyngeal function at age five years:*

*a. Velopharyngeal composite score summary (VPC sum)*: a long ordinal outcome of individual score that contributes to the primary outcome, see *Table 1*. This score is measured on a scale of 0-6.

*b. Insufficient velopharyngeal function (VPC rate):* a dichotomous outcome of whether the child has “insufficient” VPC rate.

*2. Velopharyngeal function at age three years:*

*a. Insufficient velopharyngeal function (VPC rate):* a dichotomous outcome of whether the child has “insufficient” VPC rate.

*b. Velopharyngeal insufficiency symptoms:* a bounded continuous outcome, the proportion of times that a target consonant uttered has a velopharyngeal insufficiency symptom. Each child will attempt a minimum of 15 and a maximum of 30 predetermined target consonants (in words).

*3. Canonical babbling at age twelve months:*

*a. Canonical babbling present:* a dichotomous outcome of whether the child is “canonical” or “not canonical”.

*b. Canonical babbling ratio:* a bounded continuous outcome, the proportion of times that a syllable produced is “canonical”. Determined as the average proportion from the three SLTs undertaking independent review.

*c. Consonant inventory:* a continuous outcome of the number of unique consonants, identified by at least two of three SLTs undertaking independent review, uttered by a child.

*4. Articulation at age three years:* Each child is required to have attempted a minimum of 15 and a maximum of 30 predetermined target consonants (in words) for articulation assessment.

*a. Percent consonants correct (PCC):* a bounded continuous outcome, the proportion of times that a target consonant is uttered correct.

*b. Percent correct placement (PCP):* a bounded continuous outcome, the proportion of times that a target consonant has the correct place of articulation.

*c. Percent correct manner (PCM):* a bounded continuous outcome, the proportion of times that a target consonant has the correct manner of articulation.

*d. Non-oral consonant errors:* a bounded continuous outcome, the proportion of times that a target consonant is realised as a non-oral error.

*e. Oral consonant errors:* a bounded continuous outcome, the proportion of times that a target consonant is realised as an oral error.

*5. Articulation at age five years:* Each child is required to have attempted a minimum of 18 and a maximum of 36 predetermined target consonants (in words).

*a. PCC:* a bounded continuous outcome, the proportion of times that a target consonant is uttered correct.

*b. PCP:* a bounded continuous outcome, the proportion of times that a target consonant has the correct place of articulation.

*c. PCM:* a bounded continuous outcome, the proportion of times that a target consonant has the correct manner of articulation.

*d. Non-oral consonant errors:* a bounded continuous outcome, the proportion of times that a target consonant is realised as a non-oral error.

*e. Oral consonant errors:* a bounded continuous outcome, the proportion of times that a target consonant is realised as an oral error.

*6. Postoperative/long-term complications:*

*a. Dehiscence:* a dichotomous outcome of whether the child has a postoperative dehiscence, measured 48 hours and 30 days postoperatively.

*b. Infection:* a dichotomous outcome of whether the child has a postoperative infection, measured 48 hours and 30 days postoperatively.

*c. Evidence of fistula:* a dichotomous outcome of whether the child has a postoperative fistula, assessed as “Yes” or “Probably”, measured 30 days postoperatively and at three and five years of age*.*

*7. Hearing level:*

*a. At twelve months:*

*i. Abnormal Transient Otoacoustic Emission (TEOAE):* a dichotomous outcome of whether the child has abnormal TEOAE.

*ii. Abnormal sound field audiometry:* a dichotomous outcome of whether the child has abnormal sound field audiometry. Abnormal sound field audiometry is indicated by a measurement of >30 dB HL for at least one of four frequencies tested: 500 Hz, 1000 Hz, 2000 Hz or 4000 Hz.

*b. At three and five years:*

*i. Abnormal pure tone audiometry in at least one ear:* a dichotomous outcome of whether the child has abnormal pure tone audiometry. If testing by pure tone audiometry is not possible, sound field audiometry can be used in its place. Abnormal audiometry in at least one ear is indicated by a measurement of >20 dB HL using the pure tone method, >25 dB HL for sound field, for at least one of four frequencies tested: 500 Hz, 1000 Hz, 2000 Hz or 4000 Hz.

*ii. Abnormal pure tone audiometry in both ears:* a dichotomous outcome defined in the same way as Secondary Outcome 7bi, for patients who have both ears tested and both tested ears indicate abnormal audiometry.

*iii. Severity of better ear:* a short ordinal outcome of the severity of the better ear. If testing by pure tone audiometry is not possible, sound field audiometry can be used in its place. Each patient will be classified according to the average score in the better ear if pure tone, or both ears if sound field, across the four frequencies (500 Hz, 1000 Hz, 2000 Hz or 4000 Hz) to the categories in *Table 2* [*British Society of Audiometry, 2011*].

Table 2: Classifying severity in better ear

|  |  |
| --- | --- |
| Average dB HL | Severity |
| ≤20 dB HL | Normal |
| Between 21 and 40 dB HL | Mild |
| Between 41 and 70 dB HL | Moderate |
| Between 71 and 95 dB HL | Severe |
| >95 dB HL | Profound |

*8. Middle ear function:*

*a. Flat line Tympanogram in at least one ear:* a dichotomous outcome of whether the child has flat line tympanogram, assessed at age twelve months, three and five years. Children with either ear measured as “Type B” will be classified as having flat line tympanogram in at least one ear.

*b. Flat line Tympanogram in both ears:* a dichotomous outcome of whether the child has flat line tympanogram, assessed at age twelve months, three and five years. Children with both ears measured as “Type B” will be classified as having flat line tympanogram in both ears.

*9. Dentofacial development at age five years:*

*a. Soft tissue ANB angle:* a continuous outcome of the angle between soft tissue nasion (points A and B) measured using a profile photograph. *[Bearn, 2002]*

*b. Maxillary arch constriction score:* a bounded continuous outcome, measured using the Huddart/Bodenham scoring system, on a maxilliary and mandibular arches impression. A score can range from -24 to 8 and is measured in whole numbers. *[Martin, 2016; Gray, 2005]*

*10. Growth at twelve months:*

*a. Nude weight:* a continuous outcome, measured in grams and recorded to the nearest whole number.

*b. Crown to heel length:* a continuous outcome, measured in centimetres and recorded to one decimal place.

*c. Occipitofrontal circumference:* a continuous outcome, measured in centimetres and recorded to one decimal place.

***Sample size***

The sample size calculation was based on a test for proportions using a normal approximation: 292 participants per arm will allow a reduction in insufficient velopharyngeal function at five years from 40% to 29% to be detected with 80% power using a chi-square test (2 sided significance test at 0.05 level). The estimate of 40% was obtained using data from a pilot trial of 50 five year old participants, conducted during the planning period for the grant application. [*Lohmander, 2009*] To allow an approximate attrition of 10%, 648 participants will be recruited. Restating the power for 300 participants per arm will allow the same difference to be detected with 81% power using a chi-square test (2 sided significance test at 0.05 level). To consider the potential impact of variability around the value of 40%, 300 participants with valid data per group would provide 80% power to detect a reduction from 30% to 20% and 76% power to detect a reduction from 20% to 12%.

***Statistical analysis***

*General analysis principles*

Three analysis populations will be considered: the intention-to-treat (ITT), the per-protocol (PP) and the safety population.

The principle of ITT, as far as practically possible, will be the main strategy of the analysis adopted for the primary outcome and all the secondary outcomes. These analyses will be conducted on all randomised participants, in the group to which they were allocated, and for whom the outcome(s) of interest have been observed/measured. No imputations are planned.

A per protocol analysis, which will mirror the ITT population but exclude participants defined as having a major protocol deviation, will only be considered in the event of major protocol deviations in more than 10% of the ITT analysis population and apply to a secondary analysis of the primary outcome only. *Table S1* provides a list of the protocol deviations.

The safety dataset will classify participants who have surgery before 9 months of gestational corrected age as received 6 months surgery, and surgery at 9 months of gestational corrected age or beyond as received surgery at twelve months.

A p-value of 0.05 or less will be used to declare statistical significance for all analyses, p-values will be reported to two significant figures. Rather than adjust for multiplicity, relevant results from other studies already reported in the literature will be taken into account when interpreting the study. Percentages will be presented to one decimal place and continuous summary statistics will be given to a maximum of two decimal places.

All analyses will be performed using standard statistical software (SAS 9.4 or later). The finalised analysis datasets, programs and outputs will be archived following Good Clinical Practice guidelines and standard operating procedures at the Liverpool Clinical Trials Centre.

*Descriptive analyses*

The flow of participants through each stage of the trial, including the number of individuals screened, randomised, receiving treatment as allocated, and included in the primary analysis will be summarised using a CONSORT flow chart [*Schulz, 2010*] (*Figure 1*).

The baseline comparability of the two randomised groups in terms of minimisation factors, demographic characteristics and clinical genetics will be presented (*Table 3*).

The surgical comparability of the two randomised groups in terms of baseline surgery characteristics, intra operative events, early complications during the hospital stay, observations monitored 48 hours post-surgery and post-operative medication will be presented (*Table 4*).

Binary and categorical data will be summarised by frequencies and percentages. Continuous data will be presented by means and standard deviations (SDs), or medians and inter-quartile range (IQR) if data are skewed. Tests of statistical significance will not be undertaken for baseline characteristics; rather the clinical importance of any imbalance will be noted. The amount missing in each case will be summarised.

*Lost to follow-up, withdrawals and missing data*

The timing of withdrawal in relation to surgery and scheduled visits, level of withdrawal, who made the decision and reason for withdrawal will be summarised both overall and for each randomised group. Frequencies will be presented along with percentages using the number of participants who withdrew as the denominator.

The number lost to follow up both overall and within each randomised group will be reported and the reasons where known will be documented. Any deaths and their causes will be reported separately.

Based on experience from the ScandCleft study the structure of the centralized cleft palate care system and trial specific systems in place will ensure the occurrence of missing data is likely to be low. Therefore the potential impact of any missing data is likely to be low. For all assessments and outcomes, participants with insufficient data to make their assessments will be expressed as a frequency and a percentage with the denominator being those who were randomised, treated and consented.

*Adherence*

Reasons for participants not receiving the randomised allocation will be summarised in a table. Adherence with follow up time points (30 days, twelve months, three and five years) will be summarised at the visit level (at least one scheduled assessment visit completed per time point) and assessment level, which will specify adherence to specific assessments. When applicable, whether or not assessments were made within the expected window (*Table 5*) will be presented. Summaries will be presented both overall and for each randomised group.

Table 5: Follow up visits: estimated visit dates and corresponding window limits

|  |  |  |
| --- | --- | --- |
| Time point | Estimated visit date, calculated from dateof birth | Window limits of estimated visit date |
| Earliest | Latest |
| Twelve months | 52 weeks 1 | *Six months arm:*- 4 weeks | + 48 weeks (equivalent to 23 months of age) 2 |
| *Twelve months arm:*- 2 weeks |
| Three years | 156 weeks | - 4 weeks | + 48 weeks (equivalent to 3 years 11 months of age) |
| Five years | 260 weeks | - 4 weeks | + 48 weeks (equivalent to 5 years 11 months of age) |
| **Notes:**1 Corrected for gestational age. Full term is defined as day 1 of the 40th week of pregnancy. 2 Speech assessments vary in their latest acceptable date: * Six months surgery patients should have their speech assessed by + 26 weeks (equivalent to 18 months of age)
* Twelve months surgery should have their speech assessed before surgery
 |

*Analysis of primary outcome*

The number of participants at age 5 years who have the primary outcome of insufficient velopharyngeal function (VPC sum ≥4) or not (VPC sum <4) will be summarised overall, for each randomised group and for each region (defined according to the location of the recruiting site: Brazil, Scandinavia, UK) by frequencies and percentages. The numbers of insufficient velopharyngeal function or not between the randomised groups will be compared using a chi-squared test, with the relative risk [*Morris, 1988*] and 95% confidence interval (95% CI) also reported.

In the circumstance that the expected insufficient VPC sum or not for each randomised group contains less than five participants, thereby raising concerns over the appropriateness of a chi-squared test, then a Fishers exact test will be used.

For inclusion in the primary analysis set, a participant must have an eligible five year speech recording for SLT assessment. An eligible speech recording requires the child to have attempted at least 18 of the 36 pre-specified target words for assessment of non oral and VPI symptom and 5 of the 9 pre-specified words for hypernasality. Participants with speech recordings that do not meet the inclusion criteria, have a recording that could not be assessed e.g. insufficient sound, or did not complete a speech recording, will be excluded from the analysis.

A sensitivity analysis will be performed to check the robustness of the results to the inclusion/exclusion set for speech recordings such that assessments made on:

* Non oral and VPI symptom recordings with less than 18 target words attempted are included;
* +Hypernasality recordings with less than 5 target words attempted are included;
* Audio recordings, where video recording not possible, for non oral and VPI symptom are excluded;
* Any recordings taken outside of the speech recording follow up window (*Table 5*) are excluded.

A multilevel logistic regression model (for insufficient VPC sum) adjusting for operating surgeon, size of cleft at baseline (soft palate only vs. soft and hard palate), randomised group and an intercept will be applied to check the robustness of the results to an unadjusted analysis approach. [*Kahan, 2013; Kahan, 2012; Hernandez, 2004, Pocock, 2002*].

An exploratory analysis will be undertaken where the primary endpoint of insufficient velopharyngeal function, defined as a score of 4-6 on VPC-sum, also includes patients who have a secondary surgery due to velopharyngeal insufficiency. This group will be compared to patients who have a score of 0 to 3 on the VPC-sum scale and have not received a secondary surgery due to velophyngeal insufficiency. This redefined binary endpoint will be compared using a chi-squared test, with the relative risk [*Morris, 1988*] and 95% CI also reported.

*Analysis of secondary outcomes*

Analysis approaches for each secondary outcome are dependent on the type of outcome, *Table 6* provides a summary.

Dichotomous outcomes

For outcomes of this type, see *Table 6*, the number of participants categorised as having, or not having, the outcome of interest will be summarised overall, for each randomised group by frequencies and percentages. The numbers with the outcome between the randomised groups will be compared using a chi-squared test, with the relative risk [*Morris, 1988*] and 95% CI also reported.

In the circumstance that the expected number of participants with and without the outcome for each randomised group contains less than five participants, thereby raising concerns over the appropriateness of a chi-squared test, then a Fishers exact test will be used.

No sensitivity analysis will be performed.

Short ordinal outcomes

For outcomes of this type, see *Table 6*, number of participants who are categorised into each classification of interest will be summarised overall, for each randomised group by frequencies and percentages. The numbers with the outcome between the randomised groups will be compared using a chi-squared test for trend.

In the circumstance that the expected number of participants in each classification for each randomised group contains less than five participants, thereby raising concerns over the appropriateness of a chi-squared test for trend, an alternative appropriate analysis approach will be used e.g. combining like groups or applying a proportional odds model.

No sensitivity analysis will be performed.

Long ordinal, bounded continuous and continuous outcomes

Outcomes of this type, see *Table 6*, will be summarised overall and for each randomised group by means and SDs, or medians and IQRs if data are skewed. Minimum and maximum values will also be presented.

Means will be compared between the two randomised groups using a t-test or by using a non-parametric equivalent. Testing for normality of data distributions will be based using a QQ plot by randomised group. 95% confidence intervals will be presented around the effect measure.

No sensitivity analysis will be performed.

*Inter and intra rater reliability*

Each of the outcomes assessed by independent assessors i.e. not at routine visit (Outcomes, 1-5 and 9), will each be reviewed by a minimum of one assessor and a maximum of three. Intra and inter assessments will be undertaken for a proportion of all outcomes to ensure reliability of the outcome measures. The number of assessors and proportion of inter and intra rater assessments is determined a priori and decided on a per outcome basis, led by members of the Trial Management Group who are specialists in the specific outcome field.

Agreement analysis will be exploratory and report agreement as frequencies and percentages or using Bland Altman agreement analysis as appropriate for the outcome type. [*Bland, 1999*]

*Additional analyses*

To support interpretation of the main trial outcomes, descriptive statistics will be summarised to report the results of: (i) the DENVER-II test at three years, (ii) additional speech therapy received outside of routine trial visits, (iii) reasons and the nature of any secondary surgeries received during the trial and (iv) nasometry at five years. Binary and categorical data will be summarised by frequencies and percentages. Continuous data will be presented by means and SDs, or medians and IQR if data are skewed.

Tests of statistical significance will not be undertaken for (i), (ii) and (iii); rather the clinical importance of any imbalance will be noted. The amount missing in each case will be summarised. (iv) Nasometry, at age five years, will be compared between the two randomised groups using a t-test or by using a non-parametric equivalent. Testing for normality of data distributions will be based using a QQ plot by randomised group. 95% confidence intervals will be presented around the effect measure.

*Safety evaluations*

Serious adverse events and unanticipated problems will be presented using descriptive statistics. Line listings of events will also be presented to provide further detail. Patients will be reported according to the safety dataset, with the number of events and patients in each safety group summarised. Tests of statistical significance will not be undertaken; rather the clinical importance of any imbalance will be noted.

**Discussion**

The TOPS trial will provide evidence to support whether surgery for cleft palate at age six months when compared to surgery at age twelve months improves velopharyngeal function at age 5 years. In addition, evidence regarding a wide range of pre-defined clinical secondary outcomes will be explored. This paper provides details of the planned statistical analyses of the trial. Publishing these plans prior to trial results will improve the scientific validity of the TOPS trial and reduce the risk of outcome reporting bias and data-driven results. [*Finfer, 2009*]

**Trial status**

The trial completed recruitment on 21st July 2015. In total 558 patients from 22 centres were recruited and the last patient is due to attend their last visit on 30th July 2020. The analysis of outcomes will be conducted thereafter.

**Abbreviations**

|  |  |
| --- | --- |
| 95% CI | 95% confidence interval |
| IQR | Inter-quartile range |
| ITT | Intention-to-treat |
| PCC | Percent consonants correct |
| PCM | Percent correct manner |
| PCP | Percent consonants placement |
| PP | Per protocol |
| SD | Standard deviation |
| SLT | Speech and Language Therapist |
| TOPS | Timing Of Primary Surgery for cleft palate |
| VPC | Velopharyngeal Composite Score |
| VPC rate | Insufficient velopharyngeal function |
| VPC sum | Velopharyngeal Composite Score summary |

**DECLARATIONS**

**Ethics approval and consent to participate**

The protocol was approved from the Multicentre Research Ethics Committee in the UK (MREC number: 09/H1306/119). Ethical approval has been sought in each participating country according to country specific procedures and from relevant ethics committees for each participating centre. The trial is registered with ClinicalTrials.gov (Identifier: NCT00993551).

**Consent for publication**

Not applicable.

**Availability of data and material**

Not applicable.

**Competing interests**

The authors have no competing interest to disclose.

**Funding**

This publication was made possible by Grants Number R21DE15128, U01DE018664 and U01DE018837 from the National Institute of Dental and Craniofacial Research (NIDCR). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIDCR. The funders had no role in the study design, data collection, analysis proposal, or writing this manuscript.

**Authors’ contributions**

EJC, RC, WS and CG drafted and finalised the statistical analysis plan. WS, CP, EW, and KM provided clinical advice with regards to the interpretation and derivation of the outcomes. EJC, CG and PRW wrote the statistical analysis sections in the trial protocol. WS is the chief investigator for this trial. WS, CP, EW, PRW, GS and CG contributed to the study design and developed the first version of the protocol. All authors read and approved the manuscript for publication.

**Acknowledgements**

The authors acknowledge our partners in Cleft Care centres in UK, Scandinavia and Brazil for participating in data collection, the Speech and Language Therapists for their contributions in calibration and rating of speech assessments during various stages of the TOPS trial, and oversight committee members for reviewing the Statistical Analysis Plan prior to sign off.

**REFERENCES**

|  |  |
| --- | --- |
| [1] | Shaw W, Semb G, Lohmander A, et al. Timing Of Primary Surgery for cleft palate (TOPS): protocol for a randomised trial of palate surgery at 6 months versus 12 months of age. BMJ Open 2019;9:e029780. doi:10.1136/ bmjopen-2019-029780.  |
| [2] | Peterson-Falzone S. Optimal age for palatoplasty to faclitate normal speech development: What is the evidence? In: Berkowitz S, ed. Cleft Lip and Palate. Berlin: Springer, 2006:691-700. |
| [3] | Hellquist R, Ponten B, Skoog T. The influence of cleft length and palatoplasty on the dental arch and the deciduous occlusion in cases of clefts of the secondary palate. Scandinavian journal of plastic and reconstructive surgery 1978;12(1):45-54  |
| [4] | Friede H, Persson EC, Lilja J, et al. Maxillary dental arch and occlusion in patients with repaired clefts of the secondary palate. Influence of push back palatal surgery. Scandinavian journal of plastic and reconstructive surgery and hand surgery / Nordisk plastikkirurgisk forening [and] Nordisk klubb for handkirurgi 1993;27(4):297-305  |
| [5] | Nystrom M, Ranta R. Effect of timing and method of closure of isolated cleft palate on development of dental arches from 3 to 6 years of age. European journal of orthodontics 1994;16(5):377-83  |
| [6] | Friede H, Enocson L, Moller M, et al. Maxillary dental arch and occlusion in repaired clefts of the secondary palate: influence of surgical closure with minimal denudation of bone. Scandinavian journal of plastic and reconstructive surgery and hand surgery / Nordisk plastikkirurgisk forening [and] Nordisk klubb for handkirurgi 2000;34(3):213-8  |
| [7] | Dorf DS, Curtin JW. Early cleft palate repair and speech outcome. Plastic and reconstructive surgery 1982;70(1):74-81  |
| [8] | Chapman KL, Hardin-Jones MA, Goldstein JA, et al. Timing of palatal surgery and speech outcome. The Cleft palate-craniofacial journal : official publication of the American Cleft Palate-Craniofacial Association 2008;45(3):297-308 doi: 10.1597/06-244[published Online First: Epub Date]|. |
| [9] | Chapman KL, Willadsen E. The Development of Speech in Children with Cleft Palate. Cleft Palate Speech: Assessment and Intervention 2011:23-40 doi: Book\_Doi 10.1002/9781118785065[published Online First: Epub Date]|. |
| [10] | Jones DL. Timing of palatoplasty and speech. In: Losee JE, ed. Comprehensive Cleft Care: Taylor & Francis Group, LLC, 2016:521-26.[1 |
| [11] | Gamble C, Krishan A, Stocken D, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337–2343. doi:10.1001/jama.2017.18556 |
| [12] | Sommerlad BC. A technique for cleft palate repair. Plastic and reconstructive surgery 2003;112(6):1542-8 doi: 10.1097/01.PRS.0000085599.84458.D2[published Online First: Epub Date]|. |
| [13] | Lohmander A, Persson C, Willadsen E, et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 4. Speech outcomes in 5-year-olds - velopharyngeal competency and hypernasality. Journal of plastic surgery and hand surgery 2017;51(1):27-37 doi: 10.1080/2000656X.2016.1254645[published Online First: Epub Date]|. |
| [14] | British Society of Audiology. Recommended Procedure: Pure-tone air conduction and bone-conduction threshold audiometry with and without masking (2011) |
| [15] | Bearn D, Sandy JR and Shaw, WC. Photogrammetric assessment of the soft tissue profile in unilateral cleft lip and palate. Cleft palate-Craniofacial J. 2002; 39, 597-603. |
| [16] | Martin CB, Ma X, McIntyre GT, et al. The validity and reliability of an automated method of scoring dental arch relationships in unilateral cleft lip and palate using the modified Huddart-Bodenham scoring system. Eur J Orthod 2016;38:353–8. |
| [17] | Gray D and Mossey PA. Evaluation of a modified Huddart/Bodenham system for assessment of maxillary constriction. European Journal of Orthodontics 2005; 27, 507-11. |
| [18] | Lohmander A, Willadsen E, Persson C, et al. Methodology for speech assessment in the Scandcleft project—an international randomized clinical trials on palatal surgery: experiences from a pilot study. Cleft Palate Craniofac J 2009;46:347-62.  |
| [19] | Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement updated guildelines for reporting parallel group randomized trials. BMJ 2010;340:c332. |
| [20] | Morris JA, Gardner MJ. Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. Stat Med 1988; 296: 1313-1316. |
| [21] | Kahan BC and Morris TP: Assessing potential sources of clustering in individually randomised trials. BMC Med Res Methodol. 2013, 13: 58-10.1186/1471-2288-13-58. |
| [22] | Kahan BC and Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. Stat Med. 2012, 31: 328-340. 10.1002/sim.4431. |
| [23] | Hernandez AV, Steyerberg EW, Habbema JD: Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. J Clin Epidemiol. 2004, 57: 454-460. 10.1016/j.jclinepi.2003.09.014. |
| [24] | Pocock SJ, Assmann SE, Enos LE, Kasten LE: Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med. 2002, 21: 2917-2930. 10.1002/sim.1296. |
| [25] | Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8:135–60. 10.1191/096228099673819272. |
| [26] | Finfer S, Bellomo R. Why publish statistical analysis plans. Crit Care Resusc. 2009;11(1):5–6. |

**ADDITIONAL TABLES:** Larger than A4

Table 3: Baseline characteristics and clinical genetics

|  | Six months surgery | Twelve months surgery | Overall |
| --- | --- | --- | --- |
| Number of patients | xx | xx | xx |
| Baseline characteristics |  |  |  |
| Gender  |  |  |  |
| NMale*; n (%)*Female*; n (%)**Not known; n*  | XxXx (xx.x)Xx (xx.x) Xx  | XxXx (xx.x)Xx (xx.x)Xx  | XxXx (xx.x)Xx (xx.x)Xx  |
| Gestational age (weeks) |  |  |  |
| *N**Mean (sd)**Median (IQR)**(Min, max)**Not known; n* | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx |
| Size of Cleft  |  |  |  |
| *N**Soft palate only; n (%)**Soft and hard palate; n (%)**Not known; n* | XxXx (xx.x)Xx (xx.x)Xx | XxXx (xx.x)Xx (xx.x)Xx | XxXx (xx.x)Xx (xx.x)Xx |
| Clinical genetics  |
| Ethnicity*White; n (%)**Black; n (%)**Asian; n (%)**Chinese; n (%)**Mixed; n (%)**Other; n (%)**Not stated; n (%)**Not known; n* | Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx | Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx | Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx |
| Weight at examination (grams)*N**Mean (sd)**Median (IQR)**(Min, max)**Not known; n* | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx |
| Length at examination (cm)*N**Mean (sd)**Median (IQR)**(Min, max)**Not known; n* | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx |
| Occipitofrontal circumference (cm)*N**Mean (sd)**Median (IQR)**(Min, max)**Not known; n* | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx |
| Interpretation of DENVER II*Normal; n (%)**Suspect; n (%)**Un-testable; n (%)**Not known; n* | Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx | Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx | Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx |
| Diagnosis *Known syndrome; n (%)**Unknown syndrome; n (%)**Severe developmental delay; n (%)**Uncertain; n (%)**Non-syndromic; n (%)**Not known; n* | Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx  | Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx  | Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx  |

Table 4: Surgery characteristics and observations

|  | Six months surgery | Twelve months surgery | Overall |
| --- | --- | --- | --- |
| Number of patients | xx | xx | xx |
| Baseline surgery characteristics |
| Grading of cleft at surgery*N**Grade1; n (%)* *Grade 2; n (%)**Grade 3; n (%)**Grade 4; n (%)* *Not known; n* | XxXx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx | XxXx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx | XxXx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)Xx |
| Shape of cleft*N**“U” shaped; n (%)* *“V” shaped; n (%)**Not known; n*  | XxXx (xx.x)Xx (xx.x)xx | XxXx (xx.x)Xx (xx.x)xx | XxXx (xx.x)Xx (xx.x)xx |
| Dimensions of Cleft Palate (mm) |  |  |  |
| Soft tissue width at posterior hard plate*N**Mean (sd)**Median (IQR)* *(Min, max)**Not known; n* | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx |
| Bony width at posterior hard palate *N**Mean (sd)**Median (IQR)* *(Min, max)**Not known; n* | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx |
| Width at base of uvula*N**Mean (sd)**Median (IQR)* *(Min, max)**Not known; n* | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx |
| Length of soft palate (distal base of uvula-hard palate)*N**Mean (sd)**Median (IQR)* *(Min, max)**Not known; n* | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx | Xxxx.x (xx.x)xx.x (xx.x)(xx.x, xx.x)xx |
| Intra operative events |
| Blood transfusion during surgery*N**Yes; n (%)**No; n (%)**Not known; n* | XxXx (xx.x)Xx (xx.x)xx | XxXx (xx.x)Xx (xx.x)xx | XxXx (xx.x)Xx (xx.x)xx |
| Anaesthetic complications*N**Yes; n (%)* *No; n (%)* *Not known; n* | XxXx (xx.x)Xx (xx.x)xx | XxXx (xx.x)Xx (xx.x)Xx | XxXx (xx.x)Xx (xx.x)xx |
| Bleeding*N**Yes; n (%)**No; n (%)**Not known; n* | XxXx (xx.x)Xx (xx.x)Xx | XxXx (xx.x)Xx (xx.x)Xx | XxXx (xx.x)Xx (xx.x)xx |
| Early complications during the hospital stay |
| Postoperative airway problems*N**Yes; n (%)* *No; n (%)**Not known; n*  | xxXx (xx.x)Xx (xx.x)Xx | xxXx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)xx |
| Postoperative blood loss *N**Yes; n (%)* *No; n (%)**Not known; n* | xxXx (xx.x)Xx (xx.x)Xx | xxXx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)xx |
| Anti-coagulants given*N**Yes; n (%)* *No; n (%)**Not known; n* | xxXx (xx.x)Xx (xx.x)Xx | xxXx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)xx |
| Dehiscence *N**Yes; n (%)* *No; n (%)**Not known; n* | xxXx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)xx |
| Readmission to operating room*N**Yes; n (%)* *No; n (%)**Not known; n* | xxXx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)xx |
| Observations monitored in first 48 hours post-surgery |
| Oxygen saturation levels*N**Clinically significant abnormality; n (%)**No clinically significant abnormality; n (%**No observation; n (%)**Not known; n* | xxXx (xx.x)Xx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)Xx (xx.x)xx |
| Carbon dioxide and oxygen*N**Clinically significant abnormality; n (%)**No clinically significant abnormality; n (%**No observation; n (%)**Not known; n* | xxXx (xx.x)Xx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)Xx (xx.x)xx |
| Arterial blood gasses*N**Clinically significant abnormality; n (%)**No clinically significant abnormality; n (%**No observation; n (%)**Not known; n* | xxXx (xx.x)Xx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)Xx (xx.x)xx |
| Heart rate *N**Clinically significant abnormality and treatment required; n (%)**Clinically significant abnormality and no treatment required; n (%)**No clinically significant abnormality; n (%**No observation; n (%)**Not known; n* | xxXx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)xx |
| Blood pressure*N**Clinically significant abnormality and treatment required; n (%)**Clinically significant abnormality and no treatment required; n (%)**No clinically significant abnormality; n (%**No observation; n (%)**Not known; n* | xxXx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)xx |
| Respiration*N**Clinically significant abnormality; n (%)**No clinically significant abnormality; n (%**No observation; n (%)**Not known; n* | xxXx (xx.x)Xx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)Xx (xx.x)xx |
| Body temperature*N**Clinically significant abnormality and treatment required; n (%)**Clinically significant abnormality and no treatment required; n (%)**No clinically significant abnormality; n (%**No observation; n (%)**Not known; n* | xxXx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)Xx (xx.x)Xx (xx.x)xx |
| Post-operative medication |
| Pain relief medication*N**Yes; n (%)**No; n (%)**Not known; n* | xxXx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)xx |
| Anti-inflammatory medication *N**Yes; n (%)* *No; n (%)**Not known; n* | xxXx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)xx |
| Antibiotic medication (other than prophylactic antibiotics)*N**Yes; n (%)* *No; n (%)**Not known; n*  | xxXx (xx.x)Xx (xx.x)Xx | xxXx (xx.x)Xx (xx.x)xx | xxXx (xx.x)Xx (xx.x)xx |

Table 6: Endpoint type for secondary outcomes

| Outcome type | Outcome |
| --- | --- |
| Dichotomous | 1b: Insufficient velopharyngeal function (VPC rate) at five years  |
|  | 2a: Insufficient velopharyngeal function (VPC rate) at three years |
|  | 3a: Canonical babbling present at twelve months |
|  | 6a: Postoperative dehiscence |
|  | 6b: Postoperative infection |
|  | 6c: Evidence of fistula |
|  | 7ai: Abnormal Transient Otoacoustic Emission (TEOAE) at twelve months  |
|  | 7aii: Abnormal sound field audiometry at twelve months |
|  | 7bi\_1: Abnormal pure tone audiometry in at least one ear at three years  |
|  | 7bi\_2: Abnormal pure tone audiometry in at least one ear at five years |
|  | 7bii\_1: Abnormal pure tone audiometry in both ears at three years  |
|  | 7bii\_2: Abnormal pure tone audiometry in both ears at five years |
|  | 8a\_1: Flat line tympanogram in at least one ear at twelve months |
|  | 8a\_2: Flat line tympanogram in at least one ear at three years |
|  | 8a\_3: Flat line tympanogram in at least one ear at five years |
|  | 8b\_1: Flat line tympanogram in both ears at twelve months |
|  | 8b\_2: Flat line tympanogram in both ears at three years |
|  | 8b\_3: Flat line tympanogram in both ears at five years |
| Short ordinal | 7biii\_1: Severity of better ear at three years  |
|  | 7biii\_2: Severity of better ear at five years |
| Long ordinal | 1a: Velopharyngeal composite score summary (VPC sum) at five years |
| Bounded continuous | 2b: Velopharyngeal insufficiency symptoms at three years |
|  | 3b: Canonical babbling ratio at twelve months |
|  | 4a: Percent consonants correct (PCC) at three years |
|  | 4b: Percent correct placement (PCP) at three years |
|  | 4c: Percent correct manner (PCM) at three years |
|  | 4d: Non-oral consonant errors at three years |
|  | 4e: Oral consonant errors at three years |
|  | 5a: Percent consonants correct (PCC) at five years |
|  | 5b: Percent correct placement (PCP) at five years |
|  | 5c: Percent correct manner (PCM) at five years |
|  | 5d: Non-oral consonant errors at five years |
|  | 5e: Oral consonant errors at five years |
|  | 9b: Maxillary arch constriction score at five years |
| Continuous | 3c: Consonant inventory at twelve months |
|  | 9a: Soft tissue ANB angle at five years |
|  | 10a: Nude weight at twelve months |
|  | 10b: Crown to heel length at twelve months |
|  | 10c: Occipitofrontal circumference at twelve months |

**FIGURES**

Figure 1: CONSORT flow diagram for participants in trial up to final assessment



**SUPPLEMENTARY MATERIAL**

Table S1: Definition of protocol deviations

| Protocol specification | Potential deviation(s) | Impact | Justification |
| --- | --- | --- | --- |
| Inclusion criteria |
| Written informed proxy consent  | Consent not obtained | **Major** | Violation of participant rights |
| One parent/carer a native language speaker  | Neither parent or carer a native language speaker  | **Major** | Statistical impactPrimary outcome (insufficient velopharyngeal function) may be biased to poorer outcome in these patients. Primary analysis will be biased if violation is imbalanced between treatment groups.Safety impactNone |
| Exclusion criteria |  |  |  |
| Infants with syndromic cleft palate or severe developmental delay | Identification post randomisation  | **Major** | Statistical impactPrimary outcome (insufficient velopharyngeal function) may be biased to poorer outcome in these patients or may not be evaluable resulting in missing data.Primary analysis will be biased if occurrence of violation is imbalanced between treatment groups or a reduction in power caused by missing data.Safety impactNone |
| Congenital sensorineural hearing loss or structural middle ear anomalies | Identification post randomisation  | **Major** | Statistical impactPrimary outcome (insufficient velopharyngeal function) may be biased to poorer outcome in these patients. Primary analysis will be biased if occurrence of violation is imbalanced between treatment groups.Safety impactNone |
| One stage closure with the Sommerlad technique would be inappropriate | One stage closure with the Sommerlad technique not used at study surgery | **Major** | Statistical impactThis will introduce greater variation in definition of the study treatment. This may bias the primary outcome. Unknown whether it will increase variation in the primary outcome.Safety impactNo greater than standard medical care where the most appropriate surgical technique will be used. |
| Submucous cleft palate  | Submucous cleft palate is a different type of cleft palate condition than the study targeted isolated overt cleft palate condition | **Major** | Statistical impactInterpretations of results are intended only for isolated overt cleft palate medical condition. The inclusion of other types of cleft conditions in study summaries will add uncertainty in interpretation for the intended isolated overt cleft palate population. In addition, the presence of other cleft types may increase heterogeneity in the primary outcomes and therefore potential to decrease statistical power. Safety impactNone |
| Treatment: timing of surgery  |  |  |
| Surgery at six months or twelve months should be within visit windows  | Surgery outside visit window | **Major** | Statistical impactPatients would no longer be receiving the study intervention defined in protocol.Safety impactMajor impact if a child was operated on earlier than assigned timing of surgery. In particular, surgery earlier than 6 months is expected to carry the highest risk of peri-operative complications.Minor or no impact if surgery was conducted later than assigned timing of surgery. |
| Withdrawal before surgery due to parent/guardian decision | (a) Refusal to keep to assigned timing of surgery (b) Withdrawal with no follow up | **Major** | Statistical impact(a) Bias to treatment comparison as choice of treatment is non-randomised (b) Missing data if not followed upSafety impactNone |
| Study assessments |  |  |  |
| Scheduled assessment visit at 30 days post-surgery: late complications  | Missing or follow up assessment deviates from 30 day post-surgery date | **Minor** | Statistical impactNo risk of bias to primary outcome (insufficient velopharyngeal function at age three and five years)Safety impactNo impact. No visit window is specified in the protocol, but the number of day’s deviation from 30 days will be described. |
| Scheduled assessment visit at twelve months  | Missing or visit occurs outside allowed window either side of visit date, or for twelve month patients occurs after surgery | **Minor** | Statistical impactNo risk of bias to primary outcome (insufficient velopharyngeal function at age three and five years).However, there will be impact to secondary speech outcomes if speech visits are conducted in the twelve month group after the surgery has taken place the child is likely to be uncooperative or not perform as well as the surgery will be healing.Safety impactMinor for six month patients as surgery already performed.Minor for twelve month patients as timing of scheduled assessments are scheduled around the surgery date and not vice versa, in order to keep to the study timing of surgery. |
| Scheduled assessment visit at three years and five years | Missing or visit occurs outside allowed window either side of visit date. | **Major** | Statistical impactData from both three and five years required for primary outcome.Loss of power and potential for bias.Safety impactMinor impact as surgery already performed. |