Neurodevelopmental outcome after standard dose sevoflurane versus low-dose sevoflurane/dexmedetomidine/remifentanil anaesthesia in young children

The TREX Trial

Statistical Analysis Plan for the neurodevelopmental outcomes at 3 years

Version 1.0

03 September 2026

Clinicaltrials.gov ID: NCT03089905

Contents

1. ADMINISTRATIVE INFORMATION 4
1.1 Document Version History 4
1.2 Approvals 4
2. STUDY SYNOPSIS 5
2.1 Aim(s) 5
2.2 Study Objectives 5
2.3 Study Population 6
2.4 Intervention 7
2.5 Outcomes 7
2.6 Randomisation and Blinding 10
2.7 Sample Size Estimation 10
2.8 Protocol violations 10
3. GENERAL STATISTICAL METHODOLOGY 12
3.1 Objectives of this Analysis Plan 12
3.2 Analysis Software 12
3.3 Data Verification 12
3.4 Definition of Analysis Populations 12
3.5 Adjustment for Multiplicity 12
3.6 Interim Analyses 13
3.7 Handling of Missing Data 13
3.8 Estimands 14
4. DESCRIPTIVE STATISTICS 15
4.1 Recruitment and Follow-up 15
4.2 Baseline Characteristics 15
4.3 Protocol Violations 16
4.4 Compliance 17
4.5 Concomitant Therapies 17
5. Primary Outcome 18
5.1 Primary Analysis 18
5.2 Sensitivity analyses 18
5.3 Supplementary analysis 19
6. Secondary Outcomes 19
6.1 Key Secondary Outcomes 19

6.1.1 Primary analysis 19
6.1.2 Sensitivity analyses 20
6.1.3 Supplementary analysis 21
6.2. Non-key Secondary Outcomes 22
6.2.1 Primary analysis 22
7. Subgroup analyses 26
8. Exploratory Analyses 26
9. Safety Outcomes 26

List of abbreviations

ABAS: Adaptive Behavior Assessment System, 3rd Ed 0-5 years

BRIEF-P: Behavior Rating Inventory of Executive Function-Preschool

CELF-P2: Clinical Evaluation of Language Fundamentals- Preschool, Version 2

CBCL: Child Behavior Checklist

CI: Confidence Interval

EMI: Emergent Metacognition Index

FI: Flexibility Index

GA: General Anaesthesia

GLM: Generalised Linear Model

ICE: Intercurrent event

IPW: Inverse Probability Weighting

IQ: Intelligence Quotient

ISCI: Inhibitory Self-Control Index

ITT: Intention to treat

MI: Multiple Imputation

mITT: Modified Intention to Treat

n: Number

NEPSY: A Developmental NEuroPSYchological Assessment– Second Edition

PP: Per protocol

SAP: Statistical Analysis Plan

SD: Standard Deviation

SSIS: Social Skills Improvement System

TIVA: Total intravenous anaesthesia

WPPSI-IV: Wechsler Preschool and Primary School Intelligence Scale- Fourth Edition

1. ADMINISTRATIVE INFORMATION

Protocol: version 1.9; 01 December 2020

ClinicalTrials.gov register Identifier: NCT03089905

1.1 Document Version History

Version Date	Version	Author	Change Description	Reason/Comment
03 September 2026	1.0	A Davidson X Wang K Lee L Crowe S Sheppard	Initial release.	Not applicable.

1.2 Approvals

The undersigned have reviewed this plan and approve it as final. They find it to be consistent with the requirements of the protocol as it applies to their respective areas. They also find it to be compliant with ICH-E9 principles and confirm that this analysis plan was developed in a completely blinded manner (i.e. without knowledge of the effect of the intervention being assessed)

Name	Role on Study	Affiliation	Signature	Date

2. STUDY SYNOPSIS

TREX is a Phase III randomised, active controlled, parallel arm, assessor blinded, multicentre, superiority trial comparing

1. **Low-dose** - low-dose sevoflurane/ dexmedetomidine/remifentanil anaesthesia, and

2. **Standard-dose** - standard dose sevoflurane anaesthesia.

At trial design, we planned to randomise 450 infants aged less than 2 years of age scheduled for anaesthesia expected to last 2 hours or longer 1:1 to either the low-dose or standard dose arm.

The trial was expected to run for 7 years from the start of participant screening to the last participant completing the study. Recruitment was expected to take 4 years with a subsequent follow up of to 3 years. Participants were recruited from 20 sites. Recruitment took considerably longer than expected, partly due to COVID and, in some US sites, ongoing substantial disruptions to clinical research post COVID.

2.1 Aim(s)

The aim of this trial is to determine if low-dose sevoflurane/dexmedetomidine/remifentanil anaesthesia is associated with superior neurodevelopmental outcome compared to standard dose sevoflurane anaesthesia in children less than 2 years of age having anaesthesia expected to last 2 hours or longer.

2.2 Study Objectives

2.2.1 Primary objective

The primary objective of this trial is to determine if, in children less than 2 years of age having anaesthesia expected to last 2 hours or longer, low-dose sevoflurane/dexmedetomidine/remifentanil anaesthesia is superior to standard dose sevoflurane anaesthesia in terms of the global cognitive function as assessed by the full scale IQ score of the Wechsler Preschool and Primary School Intelligence Scale- Fourth Edition (WPPSI-IV) assessed at 3 years of age.

2.2.2 Secondary objectives

Secondary objectives include determining if, in children less than 2 years of age having anaesthesia expected to last 2 hours or longer, low-dose sevoflurane/dexmedetomidine/remifentanil anaesthesia is superior to standard dose sevoflurane anaesthesia in terms of:

a) a range of other neurodevelopmental tests performed at 3 years of age including subscales of general cognitive functioning, language, executive function, memory, adaptive behaviour, clinical behaviour and social skills.

b) diagnosis of any neurodevelopmental disorder at 3 years of age.

Other short term secondary outcomes are detailed in a separate statistical analysis plan and have already been analysed and published. [Rita Saynhalath et al. Short-term Outcomes in Infants after General Anesthesia with Low-dose Sevoflurane/Dexmedetomidine/Remifentanil *versus* Standard-dose Sevoflurane (the TREX Trial) *Anesthesiology* 2024: 141; 1075-1085]

Prioritisation of secondary outcomes

Since the trial was designed, a possible phenotype has emerged for neurodevelopmental delay post anaesthesia exposure in infancy, with several robust observational studies showing an association between anaesthesia exposure and executive function and behaviour [Caleb Ing et al. Prospectively assessed neurodevelopmental outcomes in studies of anaesthetic neurotoxicity in children: a systematic review and meta-analysis. *British Journal of Anaesthesia* 2021: 126;433-444]. In this trial executive function and behaviour are measured in the Child Behavior Checklist (CBCL) and the Behavior Rating Inventory of Executive Function-Preschool (BRIEF-P). Thus, before database lock and analysis these outcomes have been defined a priori as Key Secondary Outcomes of particular interest.

2.3 Study Population

Number of Participants

450 male or female children less than 2 years of age were enrolled.

Eligibility Criteria

Participants were assigned to a randomised study treatment only if they meet all of the inclusion criteria and none of the exclusion criteria.

Inclusion criteria

Each participant was required to meet all of the following criteria to be enrolled in the study:

- Younger than 2 years (chronological age)
- Scheduled for anaesthesia that is expected to last at least 2 hours (and/or total operating room time is scheduled to be at least 2.5 hours)
- Has a legally acceptable representative capable of understanding the informed consent document and providing consent on the participant's behalf.

Exclusion criteria

- Known neurologic, chromosomal or congenital anomaly which is likely to be associated with poor neurobehavioural outcome
- Existing diagnosis of behavioural or neurodevelopmental disability
- Prematurity (defined as < 36 weeks gestational age at birth)
- Birth weight less than 2 kg.
- Congenital cardiac disease requiring surgery
- Intracranial neurosurgery and intracranial craniofacial surgery (isolated cleft lip is not an exclusion)
- Previous cumulative exposure to general anaesthesia exceeding 2 hours
- Planned future cumulative exposure to anaesthesia exceeding 2 hours before the age of 3 years.
- Any specific contraindication to any aspect of the protocol
- Previous adverse reaction to any anaesthetic
- Circumstances likely to make long term follow-up impossible

- Living in a household where the primary language spoken at home is not a language in which we can administer the WPPSI-IV
- Planned postoperative sedation with any agent except opioids (e.g. benzodiazepines, dexmedetomidine, ketamine, barbiturates, propofol, clonidine, chloral hydrate, and other non-opioid sedatives). For example if such sedation is planned for post-operative ventilation

2.4 Intervention

Low-dose - After induction the child received a loading dose of 1mcg/kg of dexmedetomidine over 10 minutes followed by an infusion of dexmedetomidine at 1 mcg/kg/hr, and a loading dose of remifentanil 1 mcg/kg over 2 minutes followed by an infusion starting at 0.1 mcg/kg/min or greater. The loading doses were started immediately after induction (or as soon after induction as practicable) and completed within 10 minutes of induction. Dexmedetomidine was discontinued 10 minutes before the end of surgery and remifentanil after last stitch/dressing was applied.

After induction and after dexmedetomidine loading was completed, the child received sevoflurane aiming for an end tidal concentration of 0.6 -0.8% or less. Sevoflurane was discontinued after last stitch and/or dressing applied. Morphine or other long-lasting opioids were administered 15 minutes prior to end of the case at the discretion of the anaesthetist.

Standard-dose - After induction the child received sevoflurane aiming for an end tidal concentration of 2.5-3.0% or greater. Sevoflurane was discontinued after last stitch and/or dressing applied. Opioids (apart from remifentanil) were administered at the discretion of the anaesthetist. Dexmedetomidine was not permitted to be given in this cohort.

2.5 Outcomes

Neuropsychological tests

Table 1 contains a list of the outcome measures collected at the child's chronological age of 3 years ideally within a six-month window after their third birthday.

Table 1. Assessment Measures for Children aged 3 years at assessment

Cognitive Area Assessed	**Assessment Task**	**Scores**	**Variable type and range**	**Outcome number**
General Cognitive Function	Wechsler Preschool and Primary School Intelligence Scale – Fourth Edition (WPPSI-IV)	Full-Scale IQ **(primary outcome)**	[40- 160], continuous	1
Clinical Behaviour	Child Behavior Checklist for ages 1.5-5 years (CBCL)	Internalizing **(key secondary outcome)**	T-Scores [29- 100], continuous & binary (using 64 and above versus below 64)	2 & 2a
		Externalizing **(key secondary outcome)**	T-Scores [28- 100], continuous & binary (using 64 and above versus below 64)	3 & 3a
		Total Score **(key secondary outcome)**	T-Scores [28- 100], continuous & binary	4 & 4a

Cognitive Area Assessed	Assessment Task	Scores	Variable type and range	Outcome number
			(using 64 and above versus below 64)	
Executive Function	Behavior Rating of Executive Function- Preschool (BRIEF-P)	T-Scores Global Executive Composite **(key secondary outcome)**	T-Scores [33-115], continuous & binary (using 65 and above versus below 65)	5 & 5a
	BRIEF-P	Inhibitory Self-Control Index (ISCI)	T-Scores [35- 101], continuous & binary (using 65 and above versus below 65)	6 & 6a
		Flexibility Index (FI)	T-Scores [35- 98], continuous & binary (using 65 and above versus below 65)	7 & 7a
		Emergent Metacognition Index (EMI)	T-Scores [36- 111], continuous & binary (using 65 and above versus below 65)	8 & 8a
General Cognitive Function	WPPSI-IV	Verbal Comprehension Index	[45- 155], continuous	9
	WPPSI-IV	Visual Spatial Index	[45- 155], continuous	10
	WPPSI-IV	Working Memory Index	[45- 155], continuous	11
*Language **	Clinical Evaluation of Language Fundamentals- Preschool, Version 2 (CELF-P2) (non-Italian participants)	Expressive Language Score	[45- 155], continuous	12
		Receptive Language Score	[45- 155], continuous	13
		Core Language Score	[45- 155], continuous	14
	TVL (Italian participants)	Weighted average score	[0-10] continuous	15
Attention/Executive Function/impulse control	NEPSY-2: Statue Subtest	Subtest score	[1-19], continuous	16
Memory	NEPSY-2: Narrative memory	Subtest score	[1-19], continuous	17
Adaptive Behaviour	Adaptive Behavior Assessment System, 3rd Ed 0-5 years (ABAS-III)	General Adaptive Composite	[40-150], continuous	18

Cognitive Area Assessed	Assessment Task	Scores	Variable type and range	Outcome number
Social Skills	Social Skills Improvement System (SSIS)^	Social Skills Score	[55-145], continuous	19

*Note that there is no validated Italian version of the CELF-P2 so at Italian sites the TVL was conducted instead of the CELF-P2. This is an equivalent test but for rigor we will not combine the tests, so these languages outcomes will be reported separately for Italian and non-Italian participants.

^The SSIS was not conducted at Italian sites

Other secondary outcomes at 3 years of age

The following were also recorded at the 3 years follow up:

- Intervention for neurological or behavioural problems since enrolment
- Any diagnosis of neurodevelopmental delay (based on notes made by the study team after discussion with families or notes entered into the WPPSI form)

Italian neuropsychological tests

There is no, fully validated and normalised Italian version for the CELF-P2. This was discovered after the protocol was written and the trial had commenced but before testing began. Therefore, for Italian participants, an alternative robust validated and calibrated Italian test (the TVL) which is used clinically and in research to measure the same neuropsychological domain (language) was employed. The scores will be reported separately for the CELF-P2 and the TVL and not combined.

There is also no Italian version of the SSIS, so this test was not conducted at Italian sites.

2.6 Randomisation and Blinding

Randomisation 1:1 between the two treatment arms was stratified by site and age at exposure (less than 12 months and greater than 12 months) using block randomisation with variable block sizes. The treating anaesthetist was not blinded to treatment arm. The assessing neuropsychologist and parents were blinded to treatment arm.

2.7 Sample Size Estimation

The planned sample size was based on the primary objective of the trial: to determine if low-dose sevoflurane/dexmedetomidine/remifentanil anaesthesia was associated with superior neurodevelopmental outcome based on the global cognitive function as assessed by the full scale IQ score of the Wechsler Preschool and Primary School Intelligence Scale assessed at 3 years of age compared to standard dose sevoflurane anaesthesia when anaesthesia is expected to last 2 hours or longer. In a normal population the Wechsler composite cognitive score (full scale IQ) has a mean of 100 and a standard deviation (SD) of 15. A difference of one third of a standard deviation (5 points) would be regarded as being minimally clinically significant. 190 children in each arm were required to have a 90% power to detect a difference of 5 points on the Wechsler composite cognitive score based on a two-sided test with alpha=0.05 and SD=15. We planned to recruit a total of 450 children to allow for 15% loss to follow up based on a previous trial (the GAS trial) where we had a retention rate at 2 years of greater than 85%. [Andrew Davidson et al. Neurodevelopmental outcome at 2 years of age after general anaesthesia and awake-regional anaesthesia in infancy (GAS): an international multicentre, randomised controlled trial. *The Lancet* 2006: 387; 239-250]

2.8 Protocol violations

All protocol violations were regularly reviewed during the trial as per normal management of a trial. They were reviewed in a blinded way where possible.

Misrandomisations are not considered to be violations.

The violations will be grouped into those that have no biologically plausible impact on treatment effect, and those with a plausible impact.

No plausible impact on outcomes violations include:

- Timing of starting or stopping dexmedetomidine infusions not exactly as per protocol
- Fentanyl, morphine or extra remifentanil given not exactly as per protocol
- End Tidal CO_2 slightly higher than as in protocol

Plausible impact on outcomes protocol violations include:

- Giving propofol or other hypnotic anaesthetic drugs, including alpha 2 agonists, during the intervention (outside allowed in the protocol)
- Assessing a child outside the target window
- Including a child whose gestational age at birth is between 34 and 36 weeks

3. GENERAL STATISTICAL METHODOLOGY

3.1 Objectives of this Analysis Plan

This statistical analysis plan (SAP) covers the analysis of the objectives related to the neurodevelopmental outcomes measured at 3 years of age within the TREX trial, including the primary outcome of the trial.

3.2 Analysis Software

Data will be analysed using Stata (StataCorp. 2019. Stata Statistical Software: College Station, TX, USA: StataCorp LLC) version 18 or later.

3.3 Data Verification

Data will be entered directly by study sites into a REDCap database. Any values out of the reference range will be checked by the Data Coordinating Centre in Melbourne and validated by the site. Once all data entry and in-house verification are complete, an unblinded statistician within the Clinical Epidemiology and Biostatistics Unit at the Murdoch Children's Research Institute will conduct some additional specifically programmed data validation checks to ensure the outcomes entered are consistent and accurate.

3.4 Definition of Analysis Populations

The population of interest in this SAP is the modified intention to treat (mITT) population. This consists of all participants who were randomised in the trial analysed according to the arm in which they were randomised to, irrespective of the intervention received or whether they completed the study, excluding participants who:

- did not proceed to surgery due to cancellation, and hence no intervention was involved;
- withdrew consent before participant carers, anaesthetist or surgeon learning the treatment allocation;
- had a pre-existing exclusion criterion that was present at randomisation but identified only afterwards, meaning that they should not have been randomised. These will be classified as misrandomised. Note that enrolling premature children at 34-36 weeks gestation at birth will not result in exclusion the mITT population.

The above excluded participants will be included in the CONSORT flow chart of the final manuscript but will be excluded from all analyses.

Participants who were withdrawn after the carers, surgeons or anaesthetist knew the treatment allocation will be included in the mITT population. In this case, participants' unavailable outcomes will be treated as missing and handled according to missing data approach (see below).

Randomisations which were generated in error due errors in testing or initial challenges in recruitment practices, and where no subject existed, will not be included anywhere including not being in the consort diagram.

3.5 Adjustment for Multiplicity

In this analysis there are no plans to adjust for multiplicity since there is a single primary outcome. However, results will be interpreted with caution focussing on the magnitude of the effect and the 95% confidence interval (CI) rather than the p-value.

3.6 Interim Analyses

No interim analyses were performed on any outcome.

3.7 Handling of Missing Data

If the proportion of missing data is greater than 5% in the primary outcome then we will consider the causal mechanism behind the missing data and use this to guide the handling of missing data. [https://academic.oup.com/ije/article/52/4/1268/7034967] For example, if we determine that the data are recoverable and that there are variables that are potentially associated with missingness or the incomplete variables, then we will use multiple imputation to handle the missing data in the primary analysis, presenting the results from a complete case analysis as a sensitivity analysis. If there are less than 5% missing data in the primary outcome, then a complete case analysis will be conducted as the primary analysis.

If used, multiple imputation (MI) will be conducted using chained equations, also known as fully conditional specification, imputing all incomplete variables in a single imputation procedure where possible. Within the chained equations algorithm, continuous variables will be imputed using linear regression, ordinal variables will be imputed using ordinal regression and binary variables using logistic regression. Baseline and intermediate variables plausibly associated with incomplete outcomes or their missingness will be included as auxiliary variables (listed and marked (+) in section 4.2 of this SAP) in the imputation model. Imputation will be carried out separately by treatment arm, to ensure that any treatment effects are maintained, using 50 imputed datasets.

Specific missing data scenarios

- Lost to follow up or withdrawal by family with no assessments, and no reason for withdrawal related to any plausible neurodevelopmental delay:
  All outcomes *will be treated as missing*
- Not followed up with all or some of the outcome measures due to the study team or family deciding being bilingual would have an impact on ability to the test:
  Incomplete or missing outcomes *will be treated as missing*
- No assessment due to a plausible neurodevelopmental delay:
  All outcomes *will be treated as missing. For the primary and key secondary outcomes, a sensitivity analysis will be conducted where the outcome will be set to 2SD below normal.*
- Assessments done but the outcome of interest was not completed due to any reason (eg lack of attention, tiredness, or distraction):
  Incomplete or missing outcomes *will be treated as missing.* For participants with partially completed assessments, if a total outcome score can be calculated based on the available items, it will be used in the analysis. Only participants where the total outcome cannot be calculated will be treated as missing data.

If there are difficulties in getting the MI model to convergence, then inverse probability weighting (IPW) will be considered as an alternative approach to handle the missing data. This approach includes only participants who have complete data in the analysis but reweights the data so that it accounts for characteristics with missing data. Weights will be calculated by treatment arm based on a logistic regression model predicting the missingness of the outcome of interest, using the set of variables listed and marked (+) in section 4.2 of this SAP.

3.8 Estimands

An estimand is a precise description of the treatment effect reflecting the clinical question posed by a given clinical trial objective. It has 5 attributes: population, treatment, variable of interest e.g. outcome, summary measure, and possible intercurrent events (defined as an event that can occur post-randomization and preclude or affect the interpretation of the variable of interest e.g. discontinuation of treatment). When defining an estimand, it must be made clear how intercurrent events will be handled in the analysis. Different approaches can be taken towards handling intercurrent events and are described below:

i) Hypothetical: a strategy which envisages a scenario in which the intercurrent event would not occur, e.g. if participants had not switched treatment or if death had not occurred

ii) Treatment policy: a strategy which seeks to understand the treatment effect on the variable regardless of the intercurrent event, i.e. an outcome is of interest whether or not the intercurrent event occurred prior to the outcome, e.g. the final outcome is of interest irrespective of whether the participant takes additional medication

iii) Composite: a strategy which considers the occurrence of the intercurrent event as informative about the participants outcome. Under this strategy the intercurrent event is included in the endpoint definition, e.g. classifying the use of rescue medication as failure, in addition to disease progression, in a time-to-event analysis

iv) Principal Stratification: a strategy wherein treatment effects are assessed in the stratum of participants who would have a specific status with respect to the intercurrent event e.g. examining the effect of treatment in participants who would not require rescue medication

v) While-on-treatment: a strategy which considers response to treatment prior to the occurrence of the intercurrent event to be of interest. For repeated measures, values up to the intercurrent event are of interest but not values after the intercurrent event. Generally, this strategy is only useful if the duration of treatment is not relevant either because it is not clinically relevant or because the rate of an event or outcome is constant over time e.g. the rate of adverse events, where one assumes a constant hazard.

Intercurrent events relevant to this analysis are:

1. Exposure to subsequent General Anaesthesia (GA)
2. Received an intervention for neurodevelopment
3. Protocol violation with plausible impact on treatment effect, as described in section 2.8.

4. DESCRIPTIVE STATISTICS

4.1 Recruitment and Follow-up

Recruitment and follow-up will be presented in a table and using a consort diagram where applicable.

The following will be reported by allocation arm unless otherwise specified:

- Number screened (overall)
- Number randomised
- Number in the mITT population
- Number excluded from the mITT population and reasons for exclusion
- Number of protocol violations listed by type (see above)
- Number who completed the study intervention
- Number who did not have any neurodevelopmental assessment and if applicable the reason why including loss to follow up, withdrawal by parents, patient too developmentally delayed, not followed up because of language not being applicable to the test
- Number who completed at least part of neurodevelopmental assessment
- Number where some secondary outcomes were completed but not the primary outcome and if applicable, reasons why including: loss to follow up, withdrawal by parents, patient too developmentally delayed, not followed up because of language not being applicable to the test, child unable to sustain attention
- Number with the primary outcome completed

4.2 Baseline Characteristics

Baseline characteristics and clinical data will be summarised using mean and standard deviations (SD) for continuous variables and numbers and percentages for categorical variables, presented separately for the low-dose and standard-dose arms as randomised (mITT population).

Variables marked with (+) will be included as auxiliary variables in the MI/IPW procedure if used to handle missing data.

The following variables will be included:

Demographics

- Sex: male, female (+)
- Country: Aust, Italy, USA (+)
- Age at time of surgery (years) (+)
- Weight at time of surgery (kg)
- American Society of Anesthesiologists Physical Score Status: 1, 2, 3, 4, or 5
- Gestational age at birth (weeks) (+)
- Multiple pregnancy: Yes/No
- More than one language spoken at home: Yes/No
- Maternal age at birth: >21, 18-21, < 18 (+)
- Family structure: two caregivers living together, two caregivers separated but dual custody, single caregiver, other
- Relationship of caregiver one: mother, father, step-mother, step-father, other

- Education of caregiver one: Completed tertiary education, continuing or currently deferred form tertiary education, completed years 11 or 12, did not complete year 12
- Relationship of caregiver two: as above
- Education of caregiver two: as above
- Number of children < 16 years at home: n
- Child birth order: 1/2/3/4+
- Location of family: urban/rural
- Any previous anaesthesia: Yes/No. If yes how many: n

Anaesthesia and surgery details

- Duration of anaesthetic (minutes) (+)
- List of surgical procedures: as per short term outcome paper [Rita Saynhalath et al. Short-term Outcomes in Infants after General Anesthesia with Low-dose Sevoflurane/Dexmedetomidine/Remifentanil *versus* Standard-dose Sevoflurane (the TREX Trial) *Anesthesiology* 2024: 141; 1075-1085]
- Median end-tidal sevoflurane during exposure period
- Total sevoflurane exposure (MAC-hours) (+)

Follow up details

- Age at assessment (years) (+)
- Location of assessment: hospital/home/other
- Hospitalised since enrolment: Yes/No. If yes, how many hospitalisations: n
- Any anaesthetics since enrolment apart from the intervention: Yes/No. If yes, how many: n (+)
- Intervention for neurological or behavioural problems since enrolment: Yes/No. If yes list (+)
- Any diagnosis of neurodevelopmental delay
- Did the psychologist discover the intervention they were randomised to: Yes/No
- Did the parents discover the arm: Yes/No
- If WPPSI was performed outside the window 3yr and 6 months: Yes/No

4.3 Protocol Violations

Protocol violations will be summarised by treatment arm as:

- number and proportion of participants with 1 or more violations
- number and proportion of violations by type of violation
- number and proportion of violations which had: no plausible impact on treatment effect, plausible impact on treatment effect
- number and proportion of participants with violations that have no plausible impact on the outcome (and hence the treatment effect), and a plausible impact on the outcome. If a participant has multiple violations, they will be classified according to the category of the violation with the greater impact.

4.4 Compliance

Compliance will be summarised by treatment arm as the number and percentage of participants who were received their randomised allocation.

4.5 Concomitant Therapies

All concurrent medication not listed as on the protocol will be described as the number and percentage of participants who received each different type of concurrent medication.

5. Primary Outcome

5.1 Primary Analysis

Primary estimand for the primary outcome:

The primary estimand for the primary outcome (full scale IQ score of the WPPSI-IV) is shown in Table 2 below.

The missing outcome data will be handled as per section 3.7.

The full scale IQ score of the WPPSI-IV will be summarised as the mean and SD in each arm.

The mean difference between treatment arms will be estimated using a generalised linear model (GLM) which will employ a Gaussian distribution and an identity link with a random effect to account for the clustering by site, and a fixed effect for the stratification factor of age at randomisation (less than 12 months and greater than 12 months).

Table 2: Primary estimand for the primary outcome.

Main analysis					
Population	**Treatment**	**Outcome**	**Summary Measure**	**Potential intercurrent events**	**Strategy for intercurrent event**
mITT	low dose vs. standard dose anaesthesia	full scale IQ score of the WPPSI-IV assessed at 3 years of age	Mean difference (95% CI) between treatment arms, adjusted for stratification factors.	• Exposure to subsequent GA -- • Intervention for neurodevelopmental concern post intervention – • Protocol violation with a plausible impact on the outcome—	Treatment policy strategy for all ICEs

5.2 Sensitivity analyses

5.2.1 Sensitivity analysis 1

A sensitivity analysis will be conducted where children measured outside the predefined assessment window (6 months post their third birthday) will be set to missing using the same methodology as for the primary analysis.

5.2.2 Sensitivity analysis 2

In a second sensitivity analysis children who could not finish the assessment due to a neurodevelopmental disorder diagnosis will be set to 2 SD below normal. The analysis will be conducted using the same methodology as for the primary analysis.

5.2.2 Sensitivity analysis 3

If multiple imputation is used to handle missing data in the primary analysis, then a third sensitivity analysis will be conducted using a complete case analysis.

5.3 Supplementary analysis

Table 3 outlines a supplementary estimand of interest for the primary outcome.

Table 3: Supplementary Estimand of interest for the primary outcome

Supplementary analysis					
Population	**Treatment**	**Outcome**	**Summary Measure**	**Potential intercurrent events**	**Strategy for intercurrent event**
mITT	low dose vs. standard dose anaesthesia	full scale IQ score of the WPPSI-IV assessed at 3 years of age	Mean difference (95% CI) between treatment arms, adjusted for stratification factors.	• Exposure to subsequent GA • Intervention for neurodevelopmental concern post intervention • Protocol violation with a plausible impact on the outcome	Hypothetical strategy for all intercurrent events - observed values will be set to missing in subjects who experienced the intercurrent event.

This supplementary estimand will be estimated as described for the primary estimand but with observed values set to missing in subjects who experienced the intercurrent event and imputed using multiple imputation as described above.

6. Secondary Outcomes

6.1 Key Secondary Outcomes

6.1.1 Primary analysis

The key secondary outcomes in this study are measures below:

- CBCL (Internalizing, Externalizing and Total Score T-scores)
- BRIEF (Global Executive Composite T-scores)

Key secondary outcomes will be summarised as means and standard deviations in each treatment arm.

The estimands of interest for the key secondary outcomes are described in Table 4.

The missing outcome data will be handled as per section 3.7.

Table 4: Estimands of interest for the key secondary outcomes

Secondary Estimands						
	Population	**Treatment**	**Outcome**	**Summary Measure**	**Potential intercurrent events**	**Strategy for intercurrent event**
2	mITT	low dose vs. standard dose anaesthesia	Internalizing of CBCL assessed at 3 years of age	Mean difference (95% CI) between treatment arms, adjusted for stratification factors, and binary outcomes as per table 1	• Exposure to subsequent GA • Intervention for neurodevelopmental concern post intervention • Protocol violation with a plausible impact on the outcome	Treatment Policy for all intercurrent events - all observed values will be used, regardless of whether or not the subject experienced the intercurrent event.
3			Externalizing of CBCL assessed at 3 years of age			
4			Total Score of CBCL assessed at 3 years of age			
5			T-Scores Global Executive Composite of BRIEF-P assessed at 3 years of age			

The mean difference between treatment arms for continuous outcomes will be estimated using a generalised linear model (GLM) which will employ a Gaussian distribution and an identity link with a random effect to account for the clustering by site, and a fixed effect for the stratification factor of age at randomisation (less than 12 months and greater than 12 months).

6.1.2 Sensitivity analyses

6.1.2.1 Sensitivity analysis 1

For the key secondary outcomes, a sensitivity analysis will be conducted where children measured outside the predefined assessment window (6 months post their third birthday) will be set to missing using the same methodology as for the primary analysis.

6.1.2.2Sensitivity analysis 2

In a second sensitivity analysis of key secondary outcomes, children who could not complete the assessment due to a diagnosed neurodevelopmental disorder will be set to 2 SD below normal. The analysis will be conducted using the same methodology as for the primary analysis.

6.1.2.3 Sensitivity analysis 3

If multiple imputation is used to handle missing data in the primary analysis, then a third sensitivity analysis will be conducted for key secondary outcomes using a complete case analysis.

6.1.3 Supplementary analysis

Table 5 outlines a supplementary estimand of interest for the key secondary outcomes.

Table 5: Estimands of interest for the supplementary analysis for the key secondary outcomes– hypothetical strategy for all 3 ICEs as per section 5.4 above

Supplementary Analysis						
	Population	**Treatment**	**Outcome**	**Summary Measure**	**Potential intercurrent events**	**Strategy for intercurrent event**
2	mITT	low dose vs. standard dose anaesthesia	Internalizing of CBCL assessed at 3 years of age	Mean difference (95% CI) between treatment arms, adjusted for stratification factors	• Exposure to subsequent GA • Intervention for neurodevelopmental concern post intervention • Protocol violation with a plausible impact on the outcome	Hypothetical strategy for all intercurrent events - all observed values will be set to missing and then imputed in those subjects who experienced the intercurrent event.
3			Externalizing of CBCL assessed at 3 years of age			
4			Total Score of CBCL assessed at 3 years of age			
5			T-Scores Global Executive Composite of BRIEF-P assessed at 3 years of age			

This supplementary estimand will be estimated as described for the primary estimand of these outcomes but with observed values set to missing in subjects who experienced the intercurrent event and imputed using multiple imputation as described above.

6.2. Non-key Secondary Outcomes

6.2.1 Primary analysis

Secondary outcomes that are binary will be summarised as the number and proportion in each treatment arm and secondary outcomes that are continuous will be summarised as means and standard deviations in each treatment arm.

The estimands of interest for the non-key secondary outcomes are described in Table 6.

Missing data for all non-key secondary outcomes will be handled as per section 3.7.

Table 6: Estimands of interest for non-key secondary outcomes

Secondary Estimands						
	Population	**Treatment**	**Outcome**	**Summary Measure**	**Potential intercurrent events**	**Strategy for intercurrent event**
6	mITT	low dose vs. standard dose anaesthesia	Inhibitory Self-Control Index (ISCI) of BRIEF-P assessed at 3 years of age	Mean difference (95% CI) between treatment arms, adjusted for stratification factors	• Exposure to subsequent GA • Intervention for neurodevelopmental concern post intervention • Protocol violation with a plausible effect on the outcome	Treatment Policy for all intercurrent events - all observed values will be used, regardless of whether or not the subject experienced the intercurrent event.
7			Flexibility Index (FI) of BRIEF-P assessed at 3 years of age			
8			Emergent Metacognition Index (EMI) of BRIEF-P assessed at 3 years of age			
9			Verbal Comprehension Index of the WPPSI-IV assessed at 3 years of age			
10			Visual Spatial Index of the WPPSI-IV assessed at 3 years of age			
11			Working Memory Index of the WPPSI-IV assessed at 3 years of age			
12			Expressive Language Score of CELF-P2 assessed at 3 years of age			
13			Receptive Language Score of CELF-P2 assessed at 3 years of age			
14			Core Language Score of CELF-P2 assessed at 3 years of age			
15			Weighted average score of TVL assessed at 3 years of age			

16			Subtest score of NEPSY-2 Statue Subtest assessed at 3 years of age			
17			Subtest score of NEPSY-2 Narrative Memory Subtest assessed at 3 years of age			
18			General Adaptive Composite of ABAS-III assessed at 3 years of age			
19			Social Skills Score of SSIS assessed at 3 years of age			
2a			Internalizing of CBCL assessed at 3 years of age --Binary	Risk difference (95% CI) between treatment arms, adjusted for stratification factors		
3a			Externalizing of CBCL assessed at 3 years of age--Binary			
4a			Total Score of CBCL assessed at 3 years of age--Binary			
5a			T-Scores Global Executive Composite of BRIEF-P assessed at 3 years of age—Binary			
6a			Inhibitory Self-Control Index (ISCI) of BRIEF-P assessed at 3 years of age—Binary			
7a			Flexibility Index (FI) of BRIEF-P assessed at 3 years of age—Binary			
8a			Emergent Metacognition Index (EMI) of BRIEF-P assessed at 3 years of age—Binary			

20	mITT	low dose vs. standard dose anaesthesia	Any intervention for any diagnosed or suspected neurodevelopmental disorder	Risk difference (95% CI) between treatment arms, adjusted for stratification factors	• Exposure to subsequent GA • Protocol violation with a plausible effect on the outcome	Treatment Policy for all intercurrent events - all observed values will be used, regardless of whether or not the subject experienced the intercurrent event.
21			Any diagnosis of neurodevelopmental delay			

The mean difference between treatment arms for continuous outcomes will be estimated using a GLM which will employ a Gaussian distribution and an identity link with a random effect to account for the clustering by site, and a fixed effect for the stratification factor of age at randomisation (less than 12 months and greater than 12 months).

The risk difference between treatment arms for binary outcomes will be estimated using a GLM which will employ a Gaussian distribution (to avert convergence difficulties with low prevalence outcomes) and an identity link with a random effect to account for the clustering by site, and a fixed effect for the stratification factor of age at randomisation (less than 12 months and greater than 12 months).

7. Subgroup analyses

The following subgroup analyses will be performed for primary and key secondary outcomes to examine the evidence for differences in the effect of the intervention between the subgroups. Since we have not powered the trial to consider subgroups, the results will be considered exploratory only. Subgroup analyses will be undertaken using the same analytic approach described above (sections 5.1 and 6.1.1), with the addition of a main effect of subgroup and a treatment by subgroup interaction term to the analysis model. Results will be presented as the treatment effect estimate within each subgroup and their 95% CIs, together with the p-value for the intervention-by-subgroup interaction, as a measure of the strength of the evidence for an interaction.

Subgroup analysis 1 – By sex

- Male vs.
- Female

Subgroup analysis 2 – By age at randomisation

- less than 12 months vs.
- greater than 12 months

Subgroup analysis 3 – Country

- Australia vs.
- USA vs.
- Italy

8. Exploratory Analyses

An exploratory analysis will be undertaken to determine whether the treatment effect between the two intervention arms on the primary and key secondary outcomes varies (linearly) with the duration of anaesthetic. In order to assess this, we will use the same analytic approach described in Sections 5.1 and 6.1.1, with addition of a main effect of duration of anaesthesia and a treatment-by-duration interaction term to the analysis model. The results will be presented as the fitted regression line between the outcome and duration of anaesthesia presented separately by treatment group, along with the point estimate of the interaction term and its 95% confidence interval and p-value.

9. Safety Outcomes

Adverse Events, and Serious Adverse Events will be listed by treatment arm and nature of the event.