Statistical Analysis Plan (SAP):

The GAS study
A multi-site RCT comparing regional and general anaesthesia for effects on neurodevelopmental outcome and apnoea in infants

Analysis of the five year follow up data

Author(s): Anneke Grobler

Version 1, 31 July 2018

Document Version History

Version Date	Version	Author	Change Description	Reason/Comment
28 March 2017	0.1	A Grobler	Initial release	Not applicable
31 July 2018	1	A Grobler	Incorporated reviews by team	

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ABAS-II	Adaptive Behavioural Assessment System – Second Edition
ADHD	Attention Deficit Hyperactivity Disorder
ASD	Autism spectrum disorder
Bayley-III	Bayley Scales of Infant & Toddler Development – Third Edition
BRIEF-P	Behavioural Rating Inventory of Executive Function- Preschool Version
CBCL	Child Behavior Checklist – Ages 1½ - 5 years, Caregiver
GMFSC	Gross motor function classification system
CMS	Children's Memory Scale
GA	general anaesthesia
GAC	Global Adaptive Composite
GEC	Global Executive Composite
GLM	generalised linear models
ITT	Intention to treat
MAR	missing at random
MICE	Multivariate Imputation by Chained Equation
MVN	multivariate normal
NEPSY-II	Developmental Neuropsychological Assessment, Second Edition
PIQ	Performance Intelligence Quotient
PP	Per protocol
PSQ	Processing Speed Quotient
PDD	Pervasive Developmental Disorder
RA	regional anaesthesia
SAP	statistical analysis plan
VIQ	Verbal Intelligence Quotient
WIAT-II	Wechsler Individual Achievement Test – Second Edition
WPPSI-III	Wechsler Preschool and Primary Scale of Intelligence - Third Edition

1. STUDY OBJECTIVES

Study outcomes are measured at three different times, namely early post surgery period, 2 years of age and 5 years of age. This statistical analysis plan will focus on the outcomes measured at 5 years chronological age.

1.1. PRIMARY OBJECTIVE

The primary aim of this prospective, observer blind, randomised, multi-site, controlled, clinical, equivalence trial is to determine whether different types of anaesthesia (regional vs general) given to infants undergoing inguinal hernia repair result in equivalent neurodevelopmental outcomes at 5 years of age.

1.2. SECONDARY OBJECTIVES AT 5 YEARS

The secondary objective is to compare a range of secondary neurodevelopmental measures between groups at 5 years of age.

2. BACKGROUND/INTRODUCTION

2.1. STUDY DESIGN

The study is a prospective, observer blind, multi-site, randomised, controlled, equivalence trial. There are a total of 28 sites from seven countries. Participants are randomised in a 1:1 ratio to receive either regional anaesthesia (RA) or general anaesthesia (GA) during surgery for hernia repair. Randomisation is stratified by gestational age at birth and by site, giving a total of 52 strata.

Participants are infants scheduled for unilateral or bilateral inguinal hernia repair at one of the participating sites, identified through review of planned theatre schedules. The duration of participation was five years. Infants were enrolled before surgery. Data were collected before and during anaesthesia and in the early post-operative period (up to 5 days following the surgery). Formal neurodevelopmental assessments occurred at 2 years corrected age and 5 years chronological age.

2.2. TREATMENT GROUPS AND INTERVENTION

The GA group received sevoflurane for induction and maintenance. The airway could be maintained with a face mask, laryngeal mask or endotracheal tube, with or without neuromuscular blocking agents. Analgesia can be supplied with a caudal and/or ilioinguinal nerve block with bupivacaine or levo-bupivacaine up to a maximum dose of 2.5 mg/kg.

The RA group received no sedative agents. The regional blockade may be with spinal block alone, spinal block with caudal block, spinal with ilioinguinal block or caudal alone. A maximum dose of 2.5 mg/kg of bupivacaine or levo-bupivacaine can be used.

2.3. STUDY POPULATION

The planned study population comprised 660 infants of postmenstrual age 60 weeks or less, requiring inguinal hernia repair under anaesthesia. Exclusion criteria include pre-existing recognised risk factors for adverse neurodevelopmental outcome or previous exposure to general anaesthesia; or being born at less than 26 weeks gestation.

2.4. SAMPLE SIZE

The sample size was based on the 5 year follow up neurodevelopmental outcome. The WPPSI-III full scale IQ score is a standardised score with a mean of 100 and standard deviation of 15. A difference of 5 points (1/3 of a standard deviation) will be taken as the largest difference that would be acceptable to demonstrate equivalence.

The planned sample size was chosen so that if the two methods of anaesthesia really were equivalent and we assumed an expected difference of only 1 standardised score point, there was a 90% chance that a 95% confidence interval will exclude a difference of more than 5. With these assumptions, the trial would need 598 infants in total. The planned sample size of 660 allowed for 10% loss to follow-up. The enrolment target was increased to enable replacement of participants who were excluded due to major protocol violations.

3. ANALYSIS POPULATIONS

3.1. MISRANDOMISATION AND HANDLING OF STRATIFICATION ERRORS

The following cases will be considered misrandomisations and will be considered not to be part of the study population and will be excluded from all statistical analyses:
- Randomisation occurred after surgery had taken place
- Consent withdrawn between prior to surgery
- Randomised due to technical error; no subject data collected
- The child found to fail to meet inclusion criteria or met exclusion criteria *prior to* surgery

Infants randomised under the incorrect stratum will be analysed under their correct stratum with no additional adjustment.

3.2. PROTOCOL VIOLATIONS

Major protocol violations include:
- surgery cancelled
- received any sevoflurane or other volatile anaesthetic agents if allocated to awake RA arm
- received any other sedative agent if allocated to awake RA arm

Minor protocol violations include:
- variations in local anaesthetic agent
- use of opioids and/or nitrous oxide if allocated to the GA arm
- variations in volatile anaesthetic agent if allocated to the GA arm
- assessed at an age outside the planned window at 5 years

3.3. STUDY POPULATIONS AND ANALYSES SETS

Per protocol population

The PP population includes all individuals that received anaesthesia treatment as randomised and also did not encounter any major protocol violation as defined previously. Misrandomised cases will be excluded from the PP population.

Intention to treat population

The intention to treat (ITT) population includes all individuals that were randomised to one of the anaesthesia arms. Misrandomised cases will be excluded from the ITT population. Patients in the ITT population will be analysed as randomised irrespective of the actual anaesthesia treatment received.

4. OUTCOME VARIABLES

4.1. PRIMARY OUTCOME AT AGE 5 YEARS

The primary outcome is the Full Scale IQ score of the Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III) at 5 years chronological age. Verbal, visuo-spatial and processing speed skills are incorporated into the Full Scale IQ score, which is indicative of general intellectual ability.

4.2. SECONDARY OUTCOMES AT AGE 5 YEARS

The results of the secondary outcomes will be interpreted with caution, and in the context of the results of the primary outcome and other similar secondary outcomes. Any conclusions drawn from the results of the secondary outcomes will be based on the totality of the evidence, with any indications in the direction of equivalence, benefit or harm, considered carefully.

1. Composite/domain scores of the WPPSI-III, specifically:
    1.1. Verbal IQ (VIQ; reflecting verbal intellectual skills)
    1.2. Performance IQ (PIQ; reflecting visuo-spatial intellectual skills)
    1.3. Processing Speed Quotient (PSQ; reflecting speed of cognitive processing)

2. Selected Developmental Neuropsychological Assessment, second edition (NEPSY-II) subtests to assess other cognitive skills consisting of:
   2.1 Attention/executive function – span – Sentence Repetition scaled score
   2.2 Attention/executive function – sustained – Auditory Attention combined scaled score
   2.3 Attention/executive function – inhibitory control– Inhibition combined scaled score
   2.4 Attention/executive function – inhibitory control – Statue scaled score
   2.5 Verbal/language skills – Word Generation scaled score
   2.6 Verbal/language skills – Speeded Naming combined scaled score
   2.7 Memory & learning - Memory for Names and Memory for Names Delay (combined) scaled score
   2.8 Social perception - Affect Recognition scaled score
   2.9 Social perception - Theory of Mind scaled score
   2.10 Sensorimotor skills - Fingertip Tapping consisting of
      2.10.1 Repetitions scaled score
      2.10.2 Sequences scaled score
   2.11 Visuomotor integration - Design Copy Process Total scaled score

3. The Wechsler Individual Achievement Test – Second Edition Abbreviated (WIAT- II Abbreviated) to assess the academic skills of the child specifically
   3.1 Word Reading standard score
   3.2 Numerical Operations standard score
   3.3 Spelling standard score

4. The Children's Memory Scale (CMS) to assess
   4.1 Attention/executive function – working memory – Numbers Total scaled score
   4.2 Memory and learning – Word Lists I (learning) scaled score
   4.3 Memory and learning – Word Lists II (delayed) scaled score

5. Parent/caregiver rated scales to assess the behaviour of the children using
   5.1 The Global Executive Composite (GEC, T score) of the Behaviour Rating Inventory of Executive Function – Preschool Version Parent Form (BRIEF-P) to measure behavioural executive abilities
   5.2 The Global Adaptive Composite (GAC, composite score) of the Adaptive Behavior Assessment System – Second Edition (ABAS-II) to measure the child's adaptive behavior
   5.3 The Child Behaviour Checklist Caregiver Questionnaire (CBCL) to measure behavioural problems consisting of
      5.3.1 The Total Problems Score (T score)
      5.3.2 CBCL internalising problems T score
      5.3.3 CBCL externalising problems T score

6. Developmental issues
   6.1 Any developmental issue (any of the below)
   6.2 Child has speech or language issues / interventions*
   6.3 Child has psychomotor issues / interventions*
   6.4 Child has global developmental delay*

7. Behavioural disorders
   7.1 Any behavioural disorders (any of the below or oppositional defiant disorder)
   7.2 Child has been diagnosed with Attention Deficit Hyperactivity Disorder (ADHD)*
   7.3 Child has been diagnosed with Autism Spectrum Disorder (ASD) *

8. Physical disability
   8.1 Child has a hearing abnormality
      8.1.1 Child has a hearing aid*
   8.2 Child has a visual defect in either eye
      8.2.1 Child is legally blind*
   8.3 Child has cerebral palsy

9. Parents' awareness of group allocation*
10. Awareness of group allocation by Psychologist / Paediatrician*

* These variables will be summarised using descriptive statistics by treatment arm only. No treatment effect or confidence intervals will be calculated.

4.3. OTHER SECONDARY OBJECTIVES (ALREADY REPORTED)

The trial also had the following secondary outcomes that were assessed at time points other than 5 years:
1. Compare a range of secondary neurodevelopmental measures between groups at 2 years (corrected). These included:
   - the cognitive, motor and language scales of the Bayley Scales of Infant & Toddler Development-III (Bayley-III)
   - social-emotional scale and adaptive behavior scale from the Bayley-III
   - the MacArthur-Bates Communicative Development Inventory

2. A paediatric assessment including a neurological examination to determine the presence of cerebral palsy conducted by a paediatrician blinded to the type of anaesthetic used.
3. Describe the frequency and characteristics of apnoea in the postoperative period after both regional and general anaesthesia for inguinal hernia repair in infants and determine other factors associated with increased risk of apnoea.

These secondary endpoints have already been analysed when the 2-year database was locked and has been published previously (Davidson, 2016; Davidson, 2015).

4.4. OTHER PARAMETERS

Baseline variables (as reported in table 1 of 2 year outcome paper published in Lancet)

Patient and family data at 5 years of chronological age
- Family demographics as per 2 year paper.
- Age at assessment
- Location of assessment
- Hospitalisation since hernia repair
- Number of anaesthetics since hernia repair
- Head injury that involved the loss of consciousness
- Any chronic illness
- List chronic illness
- Any prescribed medication for two months or longer
- Any seizures
- Height
- Weight
- Head circumference
- Arm circumference
- Any interventions for neurodevelopmental issues (speech therapy, physiotherapy, occupational therapy, psychology, other)
- Attends kindergarten, early learning or preparatory school on regular basis
- Duration of follow-up in study

5. STATISTICAL METHODOLOGY

5.1. DEMOGRAPHY AND BASELINE

Descriptive summary statistics of baseline variables as well as 5-year follow up data will be presented for both the ITT and PP populations by intervention group. For categorical data, frequencies and percentages will be provided. For continuous data, available sample size, mean, standard deviation, quartiles, minimum and maximum values will be provided.

5.2. ANALYSES OF PRIMARY OUTCOME VARIABLE

Descriptive summary statistics for the primary outcome variable at 5 years chronological age will be reported.

Linear regression will be used to analyse the Full Scale IQ score of the WPPSI-III at 5 years. The model will include the factor variables anaesthesia arm (factor levels: GA and RA), gestational age at birth (factor levels: "182 to 209 days", "210 to 258 days" and ">= 259 days") as fixed effects and site as a random effect. If the model does not converge (e.g. because of small numbers in some sites) consideration will be given to replacing site in the model with country as a fixed effect.

The analysis will be conducted for the PP and the ITT populations. The analyses will be repeated for the subset of participants with follow-up data at 5 years as well as using multiple imputation to handle the cases lost to follow-up under the MAR assumption. The primary analysis will be based on the PP population and the multiple imputation analysis method because this is an equivalence study. A PP analysis is likely to give a more conservative estimate (in the direction of non-equivalence) of the true causal effect.

The model coefficient for the anaesthesia treatment effect (conditional mean difference) will be computed along with a two-sided 95% confidence interval. Equivalence will be accepted if this confidence interval excludes values less than -5 points and values greater than +5 points.

Loss to follow-up will be investigated by providing a table with summary statistics by treatment arm for participants who attended and did not attend the 5 year visit. Variables included in this table will include key baseline variables as well as the composite cognitive, language, motor and social-emotional score of the Bayley scales of infant and toddler development measured at the 2-year time point.

5.3.   ANALYSES OF SECONDARY OUTCOMES

All continuous secondary outcomes at 5 years will be analysed in the same manner as described in Section 5.2 for the primary outcome. Some outcomes may exhibit skewed distributions. However, as long as the skewness is not severe, comparison of mean responses on their original scale is reasonable, more readily interpretable, and consistent with the way in which clinicians think about these measures in practice. If skewness is severe appropriate transformations of the data will be done prior to fitting these models.

All binary neurodevelopmental secondary outcomes at 5 years will be analysed using generalised linear models (GLM) with binomial link function in order to enable estimation of risk ratios. In case of non-convergence of the model parameters one of the following solutions will be pursued (in this order): site will be replaced with country in the models; binary logistic regression models will be fitted and marginal risk ratios with bootstrap 95% confidence intervals will be computed; or Poisson link functions will be used. The margins command in Stata will also be used to estimate risk differences with 95% confidence intervals. In analogy to the linear regression model, the respective GLM models will include the factor variables anaesthesia arm (factor levels: GA and RA), gestational age (factor levels: 182 to 209 days, 210 to 258 days and >= 259 days) as fixed effects and site as a random effect (or country as a fixed effect).

6.   STATISTICAL ISSUES

All analyses will be done in Stata (version 14.2 or higher) and all confidence intervals and p-values will be 2-sided.

6.1.   HANDLING OF MISSING DATA

Statistical analysis approaches and imputation techniques can never compensate for or exactly reproduce missing data. Therefore, no analytical strategy exists that can minimise all bias potentially introduced by missing data. However, consistencies or inconsistencies in the results based on different strategies to handle missing data can provide support for the interpretation of the study findings and either strengthen or qualify the study conclusions.

Allowing for missing values using multiple imputation:
Assuming validity of the MAR assumption, fully conditional modelling will be used to multiply impute missing values based on variable-specific prediction models. In particular, multivariate imputation chained equations (MICE) will be used to perform imputation of missing values. Depending on the type of outcome, different link functions and model specifications will be employed to utilise the built-in imputation: predictive mean matching under a Gaussian model for continuous data, logistic regression for binary data, polytomous logistic regression for unordered categorical data and proportional odds models for ordered categorical data.

If MICE models do not converge missing values will be filled in using multivariate normal regression (MVN). These two methods have been shown to give broadly equivalent results.

A total of 20 imputed data sets will be generated based on the MICE algorithm. For each of the respective data sets, the outcome-specific effect estimate and standard error will be computed and pooled using Rubin's rules (Van Buuren, 2007).

In the analysis of the 5-year data of the GAS study, the variables used in the multiple imputation models will include baseline, post-randomisation, 2 year cognitive variables and 5 year outcome variables. The following prespecified variables will be used as predictor variables within the imputation approach (since most of these variables also have missingness, they will also be imputed where necessary):

Baseline: anaesthesia group, country, site (if possible), sex, gestational age at birth (categorised), birth weight, mother received antenatal steroids, mother's education, maternal age < 21

Surgery: need for fluid bolus for hypotension, duration of surgery, significant postoperative apnoea, age at surgery

2 years: composite cognitive, language, motor and social-emotional score of the Bayley scales of infant and toddler development, any additional anaesthetic exposures since the inguinal herniorraphy, any interventions for neurodevelopmental problems, any other neurological abnormality

5 years: Full Scale IQ score, verbal IQ, performance IQ and processing speed quotient of the WPPSI; any chronic illness, any additional anaesthetic exposures since the inguinal herniorraphy, total length of any readmission to hospital, cerebral palsy, any interventions for neurodevelopmental problems, any other neurological abnormality.

With additional missing data at 5 years convergence problems, especially with some binary variables with low prevalence, may occur when doing multiple imputations. If these convergence problems are experienced some of the variables specified will be excluded from the imputation models. None of the variables required to ensure congeniality with the analysis models will be removed from the imputation models.

6.2.   SENSITIVITY ANALYSES

Some sensitivity analyses regarding missing data were described in Section 6.1. A complete case analysis as well as an analysis using multiple imputation will be done.

6.3.   SUBGROUP ANALYSES

Exploratory subgroup analyses will be performed for the primary outcome variable with respect to the age at surgery (> 70 days vs <= 70 days), duration of surgery (< 2 hours vs >= 2 hours) and country. Stratified analyses will be done by repeating the same analysis as described in Section 5.2 for each of the above mentioned factors.

The assessment of treatment-by-country interaction will be made by fitting a model for the primary outcome variable. The model will include treatment arm, country, and gestational age as fixed effects as well as the interaction term country by treatment. The p-value for the country by treatment interaction will be presented.

6.4.   EXPLORATORY DATA ANALYSES

Any post-hoc or unplanned analyses not specified in this SAP will be clearly identified as such in the report and manuscripts for publication.

7.   REFERENCES

Van Buuren, S. (2007) Multiple imputation of discrete and continuous data by fully conditional specification. Statistical Methods in Medical Research, 16, 3, 219–242

Davidson, A.J. et al. (2016) 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*, Volume 387 , Issue 10015 , 239 – 250

Davidson A.J, Morton N.S, Arnup S.J, et al. (2015) Apnea after awake-regional and general anesthesia in infants: The General Anesthesia compared to Spinal anesthesia (GAS) study: comparing apnea and neurodevelopmental outcomes, a randomized controlled trial. *Anesthesiology*. 123(1):38-54. doi:10.1097/ALN.0000000000000709.