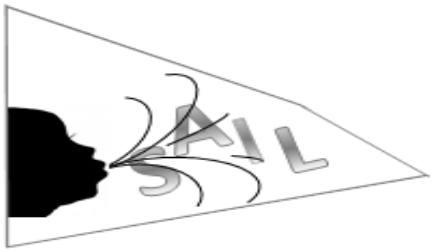


# THE SUSTAINED AERATION OF INFANT LUNGS (SAIL) STUDY:
## RANDOMIZED CLINICAL TRIAL

### A RANDOMIZED MULTICENTER CLINICAL TRIAL TO DETERMINE THE BEST OF TWO STRATEGIES TO OPTIMALLY AERATE THE PRETERM INFANT LUNG

**DATA ANALYSIS AND MONITORING PLAN (DAMP)**

Version 1.2, Monday, April 14, 2014

Sarah J. Ratcliffe, PhD
Russell Localio, PhD
Xiaoyan Han, MS

Department of Biostatistics and Epidemiology
University of Pennsylvania Perelman School of Medicine
Blockley Hall
423 Guardian Drive
Philadelphia, PA 19104

Sponsored by the
Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD)
National Institutes of Health (NIH)

## TABLE OF CONTENTS

## 1. PURPOSE

The purpose of this document is to describe the methods to be used for data analysis for the Sustained Aeration of Infant Lungs (SAIL) randomized clinical trial. The methods described will cover the standard accrual, safety, and data quality reporting for presentation at SAIL Steering Committee meetings, as well as the interim and final analyses of safety and efficacy data specified in the study protocol. This plan will go into specific detail for each analysis including, where appropriate,

The endpoints or data items to be evaluated
The population or subset upon which the evaluation is being performed
Rules for handling exceptional data
The descriptive and statistical methods to be applied to each endpoint or data item
The format to be used for presenting the results
The frequency of analyses and other monitoring
The parties to whom each report will be provided

A summary of the types and frequencies of reports produced, the personnel responsible for each report, and distribution lists are given in Appendix A. Additional information regarding procedures to be followed by the SAIL Data Safety and Monitoring Committee (DSMC) are provided in the DSMC section of the SAIL Protocol.

## 2. STUDY DESCRIPTION

### 2.1 Study Design

This study is a 2-arm randomized, controlled, multi-center clinical trial to determine which of two strategies at birth are best to optimally aerate the lung of preterm infants. Treatment allocation is blinded, though implementation of intervention is not blinded. Specifically we will determine in 600 infants of 23-26 weeks gestational age (GA) requiring respiratory support at birth which of two lung opening strategies – either a standard PEEP/CPAP of 5-7 cm H2O in the delivery room (DR), as compared to early lung recruitment using Sustained Inflation (SI) in the DR, will result in a lower rate of the combined endpoint of death or Bronchopulmonary Dysplasia (BPD) (using a standardized oxygen reduction test) at 36 weeks PMA. We will also compare which has the lower rate of other important secondary outcomes including rates of neurodevelopmental impairment at 22-26 months of corrected age in survivors. A full description of the study design is contained in the study protocol.

### 2.2 Study Objectives

Hypotheses:

1. Early lung recruitment with SI superimposed upon standard PEEP/CPAP in the DR will reduce the need for mechanical ventilation in the first seven days of life, and reduce need for surfactant use; and
2. A policy of DR SI on standard PEEP/CPAP recruitment will confer better outcomes at 36 weeks post-menstrual age (PMA) than standard PEEP/CPAP.

Primary Aim:

1. To determine in 600 infants born at 23-26 weeks GA requiring respiratory support at birth, which of two lung opening strategies – either a standard PEEP/CPAP of 5-7 cm H2O in the DR, compared to early lung recruitment using SI in the DR, results in a lower rate of the combined endpoint of death or BPD (using a standardized oxygen reduction test) at 36 weeks PMA.

Secondary Aims:

To compare the rates of other important secondary outcomes such as:

1. Detailed outcomes of potential importance in the first 10 days of life:
    a. Heart rate in the DR
    b. Detailed status on departure from the DR
    c. Use of inotropes on arrival in NICU
    d. Chest X-ray reports showing pneumothorax or new chest drains in the first 48 hours of life;
    e. Need for new chest drains after NICU admission
    f. Duration of any chest drain in-situ post-DR
    g. Oxygen profile over first 48 hours post DR using hourly FiO2 records
    h. Oxygen profile with highest FiO2 up to 48 hours
    i. Head US and/or MRI findings of intraventricular hemorrhage grades 3 and 4 by 48 hour and by day 10, if clinically available
    j. CXR appearance between days 7-10, if clinically obtained
2. Components of the primary outcome (i.e. death by 36 weeks PMA or BPD at 36 weeks PMA
3. Need for intubation in DR or by 24 hours of age
4. Pressure-volume characteristics in the DR (at several but not all sites)
5. Death or need for positive pressure ventilation at 7 days
6. Highest FiO2 and Area under the curve FiO2 for first week of life
7. Survival to discharge home without BPD, retinopathy of prematurity (grades 3 & 4), or significant brain abnormalities on head ultrasound
8. Pneumothorax and pulmonary interstitial emphysema (PIE)
9. Duration of respiratory support (ventilation, CPAP, supplemental oxygen)
10. Retinopathy of prematurity (ROP) stage 3 or greater requiring treatment
11. Death before discharge
12. Use of postnatal steroids for treatment of BPD
13. Length of hospital stay
14. Neurodevelopmental and respiratory outcome at 22-26 months corrected GA

## 2.3 Study Endpoints

The primary endpoint for this study is both an efficacy and safety endpoint. Improved death / BPD rates suggest efficacy, while worse rates suggest a safety problem. As such, all comparisons are inherently two-sided. Interim and final analyses present the classic spectrum of decisions: superiority, non-inferiority, and inferiority.

### *2.3.1 Primary Efficacy Endpoint*

To compare the rate of Death or BPD between the two intervention arms, with BPD being defined using a standardized oxygen reduction test. The primary endpoint will be recorded at 36 weeks PMA.

### *2.3.2 Secondary Efficacy Endpoints*

The secondary efficacy endpoints (and associated secondary aim) are defined below. Endpoints considered adverse events / safety endpoints are given in Section 2.3.3. The timing of evaluation of each endpoint is given in bolded font.

**Delivery Room (DR):**

1. Heart rate in the DR (SA1a)
   Categorical variable with 3 levels: <60, 60-100, >100
2. Type of respiratory support (CPAP, PPV) on departure from DR (SA1b)
3. Fraction of Inspired Oxygen ($FiO_2$) on departure from DR (SA1b)
4. Pressure-volume characteristics in the DR (at several but not all sites) (SA4)
5. Need for intubation in DR (SA3)

**First 48 hours of life:**

6. Use of inotropes (SA1c)
   Any use within first 48 hours of life
7. Pneumothorax (SA1d, SA8)
   Source documentation needed for adjudicated yes/no variable
8. Need for new chest drains (SA1de)
   Calculated as an increase in the number of chest drains recorded hourly.
9. Oxygen requirement of $Fi0_2 \geq 40\%$ for 2 hours or more (SA1g)
10. Highest $FiO_2$ level recorded during the first 48 hours post DR (SA1h)
11. Area under the hourly $FiO_2$ curve (SA6)
12. Need for intubation in DR or during first 48 hours of life (SA3)

**First 7 days of life:**

13. Death or need for positive pressure ventilation at 7 days (SA5)

**First 10 days of life:**

14. Need for new chest drains (SA1de)
    Calculated as an increase in the number of chest drains recorded hourly for the first 48 hours, and then daily up to 10 days of life.

15. Duration of any chest drain (SA1f)

    Calculated as the amount of time the number of chest drains is >0. Since this will be measured hourly for 48 hours and then daily, the finest level possible will be calculated for each infant and then the data will be examined to determine the scale for analysis.

16. Highest $FiO_2$ level recorded from 48 hours to 10 days of life (SA6)
17. Air leak (SA8)

    Defined as radiographic evidence of pneumothorax, pulmonary interstitial emphysema (PIE) or pneumopericardium.

**36 weeks PMA:**

18. Death (SA2)
19. BPD defined using a standardized oxygen reduction test (SA2)

**Discharge:**

20. Retinopathy of prematurity (ROP) stage 3 or greater requiring treatment (SA10)
21. Death in hospital (SA11)
22. Survival to discharge home without BPD, retinopathy of prematurity (grades 3 & 4), or significant brain abnormalities on head ultrasound (SA7)
23. Length of hospital stay in days (SA13)
24. Use of postnatal steroids for treatment of BPD (SA12)
25. Duration of respiratory support (ventilation, CPAP, supplemental oxygen) in days (SA9)

**22-26 Months corrected gestational age:**

26. Neurodevelopmental outcomes (SA14)
27. Respiratory outcomes (SA14)

### *2.3.3 Safety*

The primary safety endpoint is:

- Death or BPD at 36 weeks PMA.

The secondary safety endpoints are:

- Death within 48 hours of delivery.
- Oxygen requirement of Fi02 ≥ 40% for 2 hours or more within the first 48 hours post delivery.
- Rate of pneumothorax, pulmonary interstitial emphysema (PIE), and pneumopericardium within the first 10 days of life. These will determined by radiographic evidence and supplemented by data on a) any chest tube in-situ post DR and b) need for new chest tube after arrival in NICU.
- Grade 3 or 4 IVH within the first 10 days of life. Head ultrasound findings will be used to determine incidence of IVH.

- Any other serious adverse events that have been adjudicated as potentially relating to the intervention. (See Protocol section 9.2.2)

Adverse events will be compared between intervention arms. Adverse event data, including all serious adverse events, will also be descriptively summarized.

## 3. REGULAR MONITORING OF RECRUITMENT, DATA QUALITY, AND SAFETY

Monitoring of the study will be performed by the DCC. The DCC consists of personnel from three groups: faculty biostatisticians, staff biostatisticians from the Biostatistics Analysis Center (BAC) and Clinical Data Management (CDM) staff from the CRCU.

| Person | Group | Role |
|---|---|---|
| Sarah Ratcliffe | Faculty | MPI / DCC director |
| Russ Localio | Faculty | Co-I |
| | | |
| Rosemary Madigan | CRCU | Project Manager |
| Denise Cifelli | CRCU | Project Manager |
| Steve Durborow | CRCU | Database Manager |
| Chris Helker | CRCU | Database Manager |
| Melissa Fernando | CRCU | Data manager |
| Maria Blanco | CRCU | Data manager |
| | | |
| Sherry Han | BAC | Staff Statistician |
| Min Du | BAC | Programmer |

Four types of standardized reports will be produced on a regular basis: participant recruitment and follow-up, data quality and monitoring reports, demographics, and adverse events. All of these reports will be done combining the intervention arms and presented both overall and by clinical site as appropriate. Selected examples of these reports will be provided when the database development has been completed.

### 3.1 Reporting of Participant Recruitment and Accrual

Sites will enter all recruitment activity into the REDCap database, from which the weekly recruitment and accrual report will be generated. This includes information on reasons for ineligibility of subjects. The following reports will be produced by Clinical Data Management (CDM) staff in the DCC on Tuesday mornings, both overall and by clinical center (see Appendix B, Table 3 for example):

- Number of infants screened
- Number of infants ineligible by reason
- Number of infants randomized
- Number of infants randomized versus established accrual targets
- Number of infants completing each post-delivery evaluation

Approximately every 3–4 months to correspond to full meetings of the SAIL Steering Committee, the staff biostatistician at BAC (Biostatistics Analysis Center) will produce a chart to display the cumulative number of infants successfully randomized at each site as compared to the targeted number, and a graph of overall cumulative accrual to date versus the targeted accrual based on the accrual goals. Reasons for withdrawal from follow-up and time to withdrawal will also be presented as data become available.

## 3.2 Data Quality and Monitoring Reports

CDM staff will produce the following reports to monitor overall data quality. These reports will be produced approximately every 3–4 months to correspond to full meetings of the SAIL Steering Committee. These reports will include, but are not limited to:

- Missing post-delivery data
- Missing forms
- Missing values
- Query rates
- Timely entry and verification

Additional monitoring reports that may be produced by CDM staff include incorrect opening of randomization envelopes, and query status updates.

## 3.3 Demographics and Other Baseline Data

The staff biostatistician will prepare reports summarizing the distributions of demographic and clinical characteristics (Appendix B, Tables 1-2), both overall and by clinical center. These will be prepared approximately every 3–4 months to correspond to full meetings of the SAIL Steering Committee.

## 3.4 Adverse Events

### *3.4.1 Serious Adverse Events*

The serious adverse event (SAE) definitions are given in Protocol Section 9.2.2. These will include, for example:

- Death within the first 48 hours post delivery
- Pneumothorax, pulmonary interstitial emphysema (PIE) and pneumopericardium determined by radiographic evidence within the first 10 days of life
- Grade 3 or 4 IVH determined by head ultrasound findings within the first 10 days of life

Based on the premature infant population studied, the SAE categories of events that result in congenital anomaly/birth defect or require intervention to prevent permanent impairment or damage (devices) do not apply when defining events.

Serious adverse events (SAEs), regardless of whether they are unexpected or considered to be associated with the study, will be communicated immediately upon discovery of the event, by either telephone or fax. Reporting requirements are provided in the protocol and additional details will be provided in the Manual of Procedures (MOP) for the trial.

A description of all serious adverse events to date, without any intervention identifiers, will be presented at each meeting of the Steering Committee. This list will also be provided to the DSMC. The DSMC will ultimately adjudicate whether the reported adverse event is in fact an

adverse event, and a serious AE, if the SAE could be attributed to the intervention, and unblind the intervention identifiers if they deem it necessary.

### *3.4.2 Other Adverse Events*

For Steering Committee Meetings, two types of summaries of adverse events will be produced. These data will also be provided to the DSMC.

1. A listing of all adverse events will be produced by the CDM staff in the DCC directly from the AE/SAE form in the REDCap database. This listing will include a description, date of onset, grade, and relationship to study intervention, and whether or not the event was serious.
2. The study biostatistician will produce summaries of the observed adverse events combined into categories, as needed, using SAS or another similar statistical tool. All the events, regardless of the relationship to the study intervention, will be included in these reports. These reports will combine data across all interventions and clinical sites. An example of the type of table which will be produced for safety analysis is shown in Appendix B.

## 4. INTERIM ANALYSIS

Since neonatal safety is a consideration in this study, we have chosen to use a group sequential design. Thus, in addition to the final analysis, there will be two interim statistical analyses during the course of this study. The purpose of the interim analyses will be to determine whether or not there is sufficient evidence of a difference between the treatment arms in the primary endpoint such that the trial should be discontinued prior to reaching the target accrual goal.

Interim analyses will be performed after approximately 1/3 (200 subjects) and 2/3 (400 subjects) of the total required patients have completed their primary outcome. An additional safety review will be undertaken after 1/6 (100 subjects) have completed their primary outcome. The results of the interim analyses will be presented to the SAIL Data Safety and Monitoring Committee (DSMC). The primary outcome for the interim analyses will be the comparison of death/BPD between the treatment arms. This comparison will be accomplished by means of a simple generalized estimating equation (GEE) model for death/BPD versus treatment (since adjustment for deliveries with multiple babies will be needed). An approximate O'Brien-Fleming boundary will be used at each interim look to calculate the nominal significance level to which interim p-values are compared (O'Brien PC, 1979). Using the O'Brien Fleming spending function, the three analyses (2 interim + final) should use the following incremental $\alpha$ values (0.0002, 0.012, and 0.038) in order to achieve an overall $\alpha$=0.05.

There are four comparisons to warrant early stopping: (1) Clear superiority: Superiority of the intervention is sufficiently large that the study should be discontinued because equipoise is lost and randomization should no longer proceed. (2) Futility: The evidence in favor of the intervention is such that at the completion of the study, there is inadequate power to detect an improvement and the study should be dropped. (3) Inferiority: The evidence shows that the intervention is worse than the control and the study needs to be stopped. (4) Safety: Secondary safety outcomes that are considered serious adverse events possible related to the study are clinically higher in the intervention group. (See also section 4.4 for additional safety outcomes).

**(1) Clear superiority** -- whether the experimental treatment is clearly inferior to standard therapy. For these calculations, we trade off the power to detect a difference and the size of that difference. Assuming $\alpha$=0.0002 for boundaries for the initial interim analysis when 200 children have been followed to completion, power is limited except to detect large reductions in outcomes of death and BPD. For example, power is approximately 0.81 to detect a reduction from 65% in the conventional therapy to 32.5% in the experimental therapy, a 50% relative reduction. For the second interim analysis, using $\alpha$=0.012, power with 400 children followed is about 0.85 to detect an absolute reduction from 65% to 40% in the risk of death plus BPD. Thus, early or premature stopping for superiority is unlikely, without dramatic improvement from experimental treatment.

**(2) Futility**. Early stopping based on futility of the primary outcome will not be considered independently of the secondary clinical and safety outcomes. In the event that the intervention arm has equivalent BPD/death rates to the standard care arm, it would still be clinically useful to know if the intervention improves any of the secondary outcomes (that are closer to the time of the intervention) or decreases the serious adverse event rate.

**(3) Inferiority**. Power is limited to identify inferiority of the experimental treatment, with 0.8 power to demonstrate a 15% point increase in the risk of outcome from 65% in the standard therapy to 80% in the intervention.

**(4) Safety**.

(a) Early stopping based on inferior safety and non-inferior efficacy must be based largely on descriptive data and close examination of adverse events. With 200 subjects per group at a second early stopping review, and assuming that the experimental therapy is actually no worse than conventional care, observed risk in the experimental group would have to be at most 0.5 (risk of death and BPD) to have 80% power to show non-inferior efficacy (with alpha = 0.012).

(b) To justify stopping for non-inferior efficacy and superior safety again will require a substantial observed improvement in the experimental arm at the second early stopping time.

In summary, given the projected number of patients to be enrolled, early stopping will be unlikely unless the observed effect of experimental care is clearly better or worse than standard care at the planned early stopping assessment times.

All of the standard reports described in the previous section will be provided to the DSMC with this analysis. Additional measures to be considered at the interim analysis include baseline clinical factors and adverse events. The study population for all of the analyses provided to the DSMC is the intent-to-treat population, including all participants and time points for which data have been entered and verified in the study database. A summary of the analyses to be included in the interim report is outlined below. Additional details of the statistical methods to be used are given in the next section regarding the final analysis.

### 4.1 Blinding

The study will be blinded in treatment allocation; patients, parents and clinicians will be unblinded once treatment is assigned; and then blinded analyses. All analyses provided to the DSMC will be provided by intervention, but with the individual interventions identified only as A and B. The same code will be used for each intervention throughout the report. A sealed envelope with the codes for these interventions will be provided to the chair of the DSMC, to allow the DSMC to unblind these results if needed for safety reasons. These codes are also maintained by a member of the DCC who is not involved in the group trials and are not made

available to any other DCC members, including the statisticians. None of the personnel in the DCC, including the statisticians, will have access to these codes until the study has been completed, all data are entered and verified in the database, and the DSMC has approved unblinding of the intervention results.

### 4.2 Baseline/Delivery Data

Selected baseline factors will be summarized and compared between intervention groups to evaluate the adequacy of randomization, the impact of any post-randomization dropout, and to identify any imbalances that may affect intervention comparisons (Appendix B, Table 1 and Table 2). Standard descriptive statistics will be used to describe baseline characteristics, both overall and within each intervention group. Baseline characteristics examined will include maternal and gestational age, race/ethnicity, Clinical Site, consenting procedure (antenatal vs. deferred consent), route of treatment (facemask vs. nasopharyngeal tube), and enrollment date (in 3 month increments). Baseline values of selected secondary outcomes will be included as part of that analysis (see below).

Summary statistics such as medians, ranges, minima and maxima, and percentages will be produced for all measured variables. Frequencies will be computed for all categorical and ordinal variables. Graphical methods including stem and leaf plots, histograms, scatterplots, and boxplots, will be used in order to understand aspects of data quality, examine assumptions (such as normality) underlying statistical models, identify potential influential points, and guide in the choice of transformations if warranted. The balance of baseline measures across the intervention groups will be compared by inspection and by standardized differences. Overall imbalance across several or all covariates will be examined using multivariable logistic regression with treatment assignment as the outcome. .

### 4.3 Subject Accounting and Withdrawals

A detailed summary of subject accounting over the course of the study will be prepared for each treatment group (Appendix B, Table 3 & 4). A study flow diagram, derived from the one recommended by CONSORT, will also be provided (Appendix B, Figure 1). Withdrawal rates over time and reasons for withdrawal will be summarized. For the follow-up evaluation at 22-26 month corrected GA, withdrawal rates will be compared between arms using standard methods for dichotomous measures.

### 4.4 Evaluation of Safety

In addition to the standard reports, statistical analyses will be performed to compare adverse event rates between treatments. In particular, the rate of pneumothorax, pulmonary interstitial emphysema (PIE), and/or other serious adverse events (that have been adjudicated as potentially relating to the intervention; see Protocol section 8.2.2). In order to minimize the risk, this safety outcome will be compared between treatment arms after 1/6 (100 subjects) have completed the primary outcome, as well as at the two planned interim analyses for the primary outcome (1/3 and 2/3), since we would not expect any statistically significant differences in the rate after 100 subjects (with O'Brien-Fleming alpha 0.000002). Instead, we will look for a clinically meaningful double the risk in the intervention versus standard care arms. (Appendix B, Table 5)

All other adverse events not explicitly mentioned in the Protocol will also be included, regardless of their presumed relationship to the study intervention. Adverse events will be collapsed into groups based on body part depending on numbers seen. The proportion of subjects

with any adverse event will be compared between groups using an exact Mantel-Haenszel test. Comparisons between interventions will be primarily descriptive

## 5. FINAL DATA ANALYSIS UPON COMPLETION OF TRIAL

Upon completion of the primary outcome at 36 weeks PMA, after all data has been entered in the database and query resolution is complete, the primary statistical analysis and description of the data will be performed. The final analysis for the 22-26 month cGA secondary outcomes will be completed separately, after that data has been entered in the database and cleaned. The staff biostatistician under the direction of the faculty biostatisticians will produce a final report outlining all analyses and interpretation of the results. This report will be used as the basis of the primary manuscript to be prepared for publication. Details of the analyses and statistical methods to be included in the final report are outlined below. All statistical tests described below will be conducted using a two-sided level of significance.

### 5.1 Examination of Baseline Characteristics

These will be reported as described above under the interim analysis.

### 5.2 Subject Accounting and Withdrawal Rates

These will be reported as described above under the interim analysis.

### 5.3 Analysis of Safety and Toxicity

These will be reported as described above under the interim analysis.

### 5.4 Analysis of Efficacy Endpoints

#### *5.4.1 Primary Endpoint*

In addition to the two-sample comparison (Appendix B, Table 5), adjusted analyses of the primary endpoint (death/BPD rate) may be performed. These adjusted analyses of the primary endpoint will rely on logistic regression and generalized estimating equation (GEE, to control for clustering within multiple births) methods to evaluate whether observed differences, if any, are attributable to imbalances in prognostic factors such as gender, gestational age, initial heart rate, maternal corticosteriods use, and small for gestational age (SGA). Standard regression diagnostics will be used to assess model adequacy and examine potential outlying or influential data points for these analyses. If there are clinical or demographic characteristic imbalances between the two treatment arms, a propensity score analysis (via stratification into 5 subgroups, and via inverse propensity treatment weights) will be used to assess the sensitivity of the results to any treatment allocation biases.

Prior studies offer no basis for assuming a priori interactions between treatment arms and subgroups defined by sex, race/ethnicity, gestational age, site or a combination of these groups, beyond that already controlled for in the randomization. For that reason, preplanned tests for interactions with treatment assignment are not warranted, and not powered for. We propose, however, to table all results by subgroups for descriptive purposes and to explore in secondary analyses possible subgroup differences by treatment group, solely for purposes of generating hypotheses for future studies. In particular, there is interest in exploring subgroups defined by:

- Consenting procedure which may vary by site (antenatal vs. deferred consent). While consenting procedure differences may result in some selection bias, this will be implicity adjusted for with the already planned clinical site stratification factor. Additionally, the sample size still results in 80% power at $\alpha$=0.038 for the primary outcome even if the selection bias results in a reduced control arm rate.
- Route of treatment (facemask vs. nasopharyngeal tube). While not expected to modify the intervention effectiveness, exploring this difference will be important for future study design and clinical practice.
- Enrollment date. Clinical practices unrelated to this intervention, as well as increased intervention experience, could result in changes in the clinical effectiveness of the intervention over time (positively or negatively). We will explore graphically any temporal trends in the intervention effect, and will allow/test for a time-varying treatment effect via a functional logistic regression model, if needed (for which the original power calculations are conservative) (Ratcliffe SJ, et al., 2002).

### *5.4.2 Secondary Endpoints*

Secondary outcomes will be analyzed using similar procedures to the primary outcome. Descriptive statistics for secondary outcome measures will be calculated. Outcomes will be compared between treatment groups using GEE methods (where needed to account for twins) with logistic regression (dichotomous outcomes), linear regression (continuous outcomes), or survival analysis (survival time outcomes), as appropriate. If indicated, these models will be expanded to include baseline covariates and potential interactions between baseline factors and intervention.

Standard data and regression diagnostics will be used to assess model adequacy and examine potential outlying or influential data points for all of the above analyses. If these analyses suggest significant skewness or other violations of parametric model assumptions, the comparisons above may be supplemented with non-parametric methods based on rank scores within the generalized Mantel-Haenszel testing framework, adjusting for clustering of multiple births.

### *5.4.3 Group Sequential Boundary for Primary Endpoint*

To avoid inflating the overall Type I error rate for the primary analysis of efficacy, an O'Brien-Fleming boundary will be used to calculate the nominal significance level to which the interim p-value for the primary endpoint is compared. The efficacy boundary is 0.038 at the final analysis.

## 6. INTENT-TO-TREAT ANALYSIS AND MISSING DATA

### 6.1 Treatment and Study Withdrawals

It is anticipated that there will be a small number of infants who, because of acute clinical deterioration are treated according to the preference of their medical team rather than by study protocol. However, their outcomes will be measured and analyzed according to original allocation by intention-to-treat principles.

It should be noted that infants who discontinue in the trial, particularly in the case of an adverse event, will not be considered withdrawals from the study unless their parent withdraws consent for further follow-up. The parents of such infants will be encouraged to continue in the study in order to provide complete follow-up information. The characteristics at time of randomization for those participants without complete follow-up will be examined; however, there will be limited statistical power to detect any but major differences between these infants and those with complete follow-up. As mentioned above, in order to assess the potential biases introduced by differential withdrawal among intervention arms, a comparison of withdrawal rates will be carried out using the exact conditional test (ECT) version of Mantel-Haenszel methods to adjust for within-center clustering.

Potential effects of withdrawals and other sources of missing data on the statistical analysis and interpretation of results are discussed further below.

### 6.2 Other Missing Data Considerations

In addition to data that is naturally missing from participants who withdraw from study, it is expected that there may be occasional missing values for other subjects. All attempts will be made to keep missing data to a minimum. Extensive procedures to minimize missing data during the data collection and entry process will be put in place, following standard operating procedures at the DCC. These procedures will be described further in the Manual of Procedures. In addition, since the infants are still hospitalized at 36 weeks PMA as part of routine clinical care, it has been our experience that this type of missing data is minimal. The number of subjects included with each individual analysis will be given with the results. In general, missing data will not be directly imputed in any way.

### 6.3 Intent-to-Treat Analysis for the Primary Endpoint

An as randomized (formerly called “intent-to-treat”) analysis, in which all available data on all randomized participants are included, will be used for the primary endpoint comparison between interventions.

### 6.4 Potential Effects of Missing Data on Secondary Endpoints

The possibility of missing data can have serious consequences for the interpretation of statistical analyses. This is particularly problematic if the probability of missing data is related to the intervention, the outcome, or both. There is a large statistical literature on the potential effects of missing data and also some proposed methods for analysis, although these methods often require assumptions that cannot be tested within the trial dataset itself. Some of the issues regarding the potential effects of missing data on the evaluation of secondary endpoints in this trial are briefly addressed below.

For measures recorded at 22-26 months corrected GA, some missing data due to loss-to-follow-up is to be expected, though minimal. Data for those subjects without 22-26 months data will be omitted from these analyses. Although these analyses will be conducted, it is important that the results be interpreted in light of the fact that they will only include a subgroup of the population (“completers”), and that the subgroup is based on outcomes such that the benefit of the original randomization may no longer hold. In the event that the analysis of the secondary endpoints at 22-26 months becomes an important issue for clinical decision-making, methods which directly model the outcomes and the withdrawal process may be considered.

## 7. PRESENTATION AND FORMAT OF RESULTS

Examples of some of the tables described in the previous section are given in Appendix B. Additional tables and appropriate figures will be produced as needed. For general reporting, one decimal point for most efficacy and safety endpoints, p-values will be reported to two significant digits (following the ICJME conventions of x.xx for most, $p>0.2$, or $p<0.001$), and no rounding in any intermediate data or analysis steps will be used.

Data will be primarily managed in REDCap. SAS will be used for secondary data management and descriptive statistics. SAS datasets will be downloaded at least weekly from the study database for analysis. Graphics will be prepared using Stata v13 (or later) (Stata Corp, San Antonio TX, 2009) and the R programming package v 3.0 or later (The R Foundation for Statistical Computing, Vienna, Austria, 2011). Analyses will be implemented in SAS, Stata, or R as needed. These programs will be documented according to the Standard Operating Procedures of the Biostatistics and Epidemiology Consulting Center (BECC) and will be made available to sponsoring and regulatory agencies during study close out (See Manual of Procedures).

## 8. REVISIONS TO THE DAMP

An attempt has been made to anticipate possible data problems and to pre-specify handling conventions. However, it is recognized that this DAMP may not have covered all possible issues related to data analysis and reporting. Blinded data reviews will be performed, and data problems found through such reviews will be handled according to the principles outlined in this DAMP and will be properly documented. Data problems found during final analysis will be handled in the same manner and will be clearly noted in study reports. Changes to these procedures that occur during the course of this study, will be incorporated as amendments to the DAMP and included in the protocol, if appropriate.

## 9. APPENDIX A: SCHEDULE OF STUDY REPORTING

| Type of Report | Prepared By: | Provided To: | Frequency: |
|---|---|---|---|
| Serious Adverse Events (SAEs) | Sites, DCC | SC[1], DSMC, IRBs | Immediately |
| Patient Recruitment/Targets | CDM[2], DCC | SC | weekly |
| Data Quality, Timeliness | CDM | SC,<br>DSMC | weekly<br>quarterly 3-4 mos |
| Demographics (combined) | Biostat[3] | SC, DSMC | quarterly 3-4 mos |
| Interim analysis: | Biostat | DSMC | After 100, 200, 400 subjects complete 36 weeks PMA |
| Final analysis:<br>Safety and Efficacy | Biostat | DSMC | 100% accrual, follow-up after randomization until 36 weeks PMA. |

[1] SC denotes the SAIL Steering Committee, consisting of all site Investigators, the Data Coordinating Center (DCC), and NICHD representatives.
[2] CDM denotes clinical data management staff in the DCC.
[3] Biostat denotes staff biostatistician under the direction of the faculty biostatisticians in the DCC.

## 10. APPENDIX B: DRAFT STUDY REPORT TABLES

The following are examples of some of the tables that may be included in final report for this clinical trial. Selected tables may also be used for reporting the results of the interim analyses. Although the exact content and format may change, an attempt has been made to summarize the most important population, safety, and efficacy data.

**Table 1. Summary of Baseline Demographic Characteristics by Intervention Group**

| | A | B | Total | p-value |
|---|---|---|---|---|
| **Number of Subjects** | ~n | ~n | ~n | |
| **Clinical Site** | | | | |
| University of Pennsylvania | n (%) | n (%) | n (%) | n/a |
| University of Melbourne | n (%) | n (%) | n (%) | |
| etc. | n (%) | n (%) | n (%) | |
| **Consenting Procedure Used** | | | | |
| Antenatal | n (%) | n (%) | n (%) | 0.xx |
| Deferred | n (%) | n (%) | n (%) | |
| **Sex** | | | | |
| Male | n (%) | n (%) | n (%) | 0.xx |
| Female | n (%) | n (%) | n (%) | |
| **Race** | | | | |
| White/Caucasian | n (%) | n (%) | n (%) | |
| Black/African-American | n (%) | n (%) | n (%) | |
| etc. | n (%) | n (%) | n (%) | |
| **Gestational Age** | | | | |
| Median | xx.x | xx.x | xx.x | 0.xx |
| Range | xx to xx | xx to xx | xx to xx | |
| **Maternal Age** | | | | |
| Median | xx.x | xx.x | xx.x | 0.xx |
| Range | xx to xx | xx to xx | xx to xx | |
| **Maternal Gravidity** | | | | |
| Median | xx.x | xx.x | xx.x | 0.xx |
| Range | xx to xx | xx to xx | xx to xx | |

**Table 2. Summary of Baseline Clinical Characteristics by Intervention Group**

| | A | | B | | Total | | p-value |
|---|---|---|---|---|---|---|---|
| **Number of Subjects** | ~n | | ~n | | ~n | | |
| **Exposure to Antenatal Steroids** | | | | | | | |
| No | n | (%) | n | (%) | n | (%) | 0.xx |
| Yes | n | (%) | n | (%) | n | (%) | |
| **Type of Antenatal Steroids** (only in subset with exposure) | | | | | | | |
| | n | (%) | n | (%) | n | (%) | 0.xx |
| | n | (%) | n | (%) | n | (%) | |
| **Exposure to Medications to Prolong Pregnancy** | | | | | | | |
| No | n | (%) | n | (%) | n | (%) | 0.xx |
| Yes | n | (%) | n | (%) | n | (%) | |
| **Placental Abruption** | | | | | | | |
| No | n | (%) | n | (%) | n | (%) | 0.xx |
| Yes | n | (%) | n | (%) | n | (%) | |
| **Chorioamnionitis** | | | | | | | |
| No | n | (%) | n | (%) | n | (%) | 0.xx |
| Yes | n | (%) | n | (%) | n | (%) | |
| **Membranes Rupture ≥ 24 hours before Delivery** | | | | | | | |
| No | n | (%) | n | (%) | n | (%) | 0.xx |
| Yes | n | (%) | n | (%) | n | (%) | |
| **Mode of Delivery** | | | | | | | |
| Vaginal | n | (%) | n | (%) | n | (%) | 0.xx |
| C-Section | n | (%) | n | (%) | n | (%) | |
| **Route of Intervention Administration** | | | | | | | |
| Facemask | n | (%) | n | (%) | n | (%) | 0.xx |
| Nasopharyngeal Tube | n | (%) | n | (%) | n | (%) | |

**Table 3: Screening and Eligibility**

| | Total |
|---|---|
| Total Infants Screened | ~n |
| Eligible | n (%) |
| Ineligible | n (%) |
| Reasons for exclusion / ineligible: | |
| Baby was active, crying, HR>100bpm | n (%) |
| Considered non-viable | n (%) |
| Major anomaly (ies) | n (%) |
| Mother unable to consent | n (%) |
| Parent refused consent | n (%) |
| Total Randomized | ~n |
| Completed Study | n (%) |
| Completed 36 week PMA assessment | n (%) |
| Completed first month of life | n (%) |
| Completed 48 hours post delivery | n (%) |

**Table 4: Subject Accounting by Intervention Arm**

| | A | B | Total |
|---|---|---|---|
| Total Randomized | ~n | ~n | ~n |
| Completed Study | n (%) | n (%) | n (%) |
| Still on Study | n (%) | n (%) | n (%) |
| Discontinued Study: | n (%) | n (%) | n (%) |
| Violation of protocol | n (%) | n (%) | n (%) |
| Withdrawal of consent | n (%) | n (%) | n (%) |
| Adverse event | n (%) | n (%) | n (%) |
| Other | n (%) | n (%) | n (%) |

**Flow of Subjects Through Study Phases**

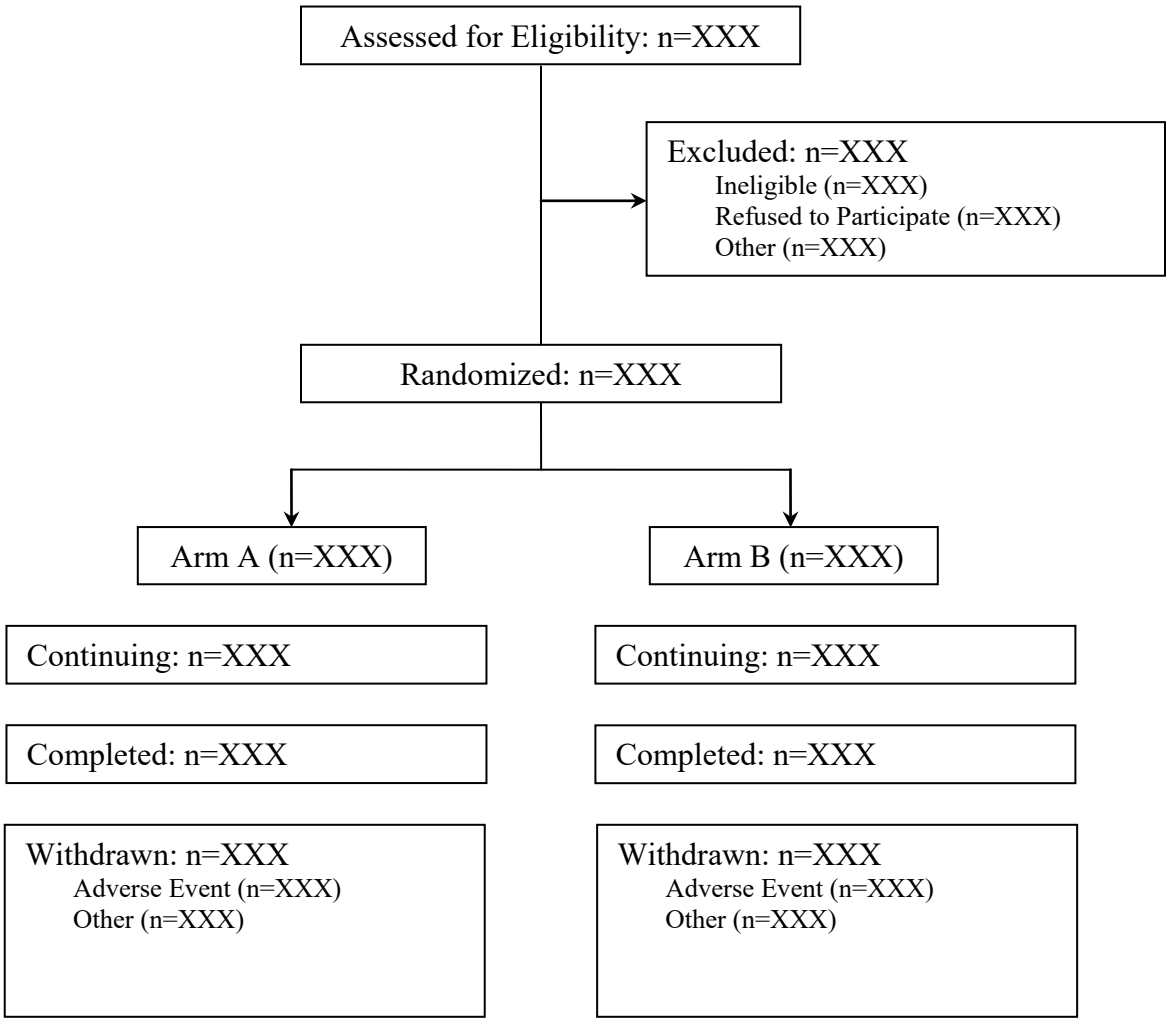


**Figure 1. Flow of Subjects Through Study Phases**

**Table 5. Cumulative Serious Adverse Events by Intervention Arm**

| | A | | B | | Total | | p-value |
|---|---|---|---|---|---|---|---|
| **Number of Subjects** | ~n | | ~n | | ~n | | |
| **Within the first 48 hours post delivery** | | | | | | | |
| Death | n | (%) | n | (%) | n | (%) | 0.xx |
| Oxygen Requirement (Fi02 ≥ 40% for 2 hours or more) | n | (%) | n | (%) | n | (%) | 0.xx |
| **Within first 10 days of life:** | | | | | | | |
| Pneumothorax | n | (%) | n | (%) | n | (%) | 0.xx |
| Pulmonary Interstitial Eemphysema | n | (%) | n | (%) | n | (%) | 0.xx |
| Pneumopericardium | n | (%) | n | (%) | n | (%) | 0.xx |
| Grade 3 or 4 IVH | n | (%) | n | (%) | n | (%) | 0.xx |
| **Within first month of life:** | | | | | | | |
| BPD | n | (%) | n | (%) | n | (%) | 0.xx |
| **Other:** | | | | | | | |
| XXX | n | (%) | n | (%) | n | (%) | 0.xx |

**Table 6. Assessment of SI Intervention**

| | Total |
|---|---|
| **Number of SI maneuvers performed** | |
| 1 | % |
| 2 | % |
| **2 SIs performed by Clinical Site** | |
| University of Pennsylvania | % |
| University of Melbourne | % |
| etc. | % |
| **Post SI intervention Respiratory Mode** | |
| CPAP | % |
| PPV | % |

NB: Percentages only in SI arm to maintain blinding until study completion.

**Table 7. Primary Endpoint: Death or BPD at 36 weeks PMA**

| | A | | B | | Total | |
|---|---|---|---|---|---|---|
| **Number of Subjects Randomized** | ~n | | ~n | | ~n | |
| **Death/BPD Rate (p=0.xx)** | n | (%) | n | (%) | n | (%) |
| Death | n | (%) | n | (%) | n | (%) |
| BPD | n | (%) | n | (%) | n | (%) |

## 11. APPENDIX C: STANDARD MONITORING REPORTS

The following are examples of the standard reports on patient recruitment, data quality, and demographics that will be prepared for the regular full Steering Committee meetings. These reports will also be sent to the DSMC as noted in Appendix A. ***For details, please refer to the MOP (under development).***

**SAIL Accrual and Data Reports Summary**

**Generated (MM/DD/YYYY)**

Notation: Tables 5 – 12 use only data that has been entered and verified. It may not include all infants that have completed a given time period assessment.

**List of Reports**

**Table 1: Number of Participants Randomized**

**Table 2: Participant Randomization and Withdrawal Report**

**Figure 1: Overall SAIL Randomization Profile**

**Figure 2: SAIL Randomization Profiles by Clinical Center**

**Table 3: Number of Participants With Completed Data** – The Randomization column represents the delivery room phase. For follow-up assessments, this report is based on data verified in the database as of the date shown. Therefore, it may not include all patients who have data at the stated gestational age.

**Table 4: Completion Rates Among Active Study Participants by Clinical Center**

**Table 5: Age Distribution - Gestational and Maternal**

**Table 6: Race Distribution**

**Table 7: Gender Distribution**

**Table 8: Missing Forms (Number of Missing Forms, by Type of Form)**

**Table 9: Timely Entry and Verification (Mean Number of Days from Date of Birth/Randomization to First Entry in Database)**

**Table 10: Missing Values (Missing Fields Per Completed Forms, by Type of Form)**

**Table 11: Query Rates (Fields Queried Per Completed Forms, by Type of Form)**

**Table 12: Adverse Event Reporting**