Statistical Analysis Plan
"Randomised trial comparing CPAP machines with reusable vs disposable circuits"

Introduction

1. This document sets out the statistical analysis plan for a randomised trial conducted at Ramaiah Medical College hospital, Bangalore, Karnataka, India, between June 2017 and July 2018 (inclusive). At each point, the plan sets out the text in the original plan and highlights any variations from that plan, as updated on 31 August 2019.

2. **Limitations**: There were two important limitations to the study analysis as envisaged in the protocol:

   a. The achieved sample size was n=51. This compares to the 80 randomised infants planned for as a feasible target in the setting, despite a 5-month extension to recruitment. The planned sample size was only capable of estimating a major difference (a difference of 0.70 standard deviations between the means) in the primary outcome. The achieved sample size therefore substantially reduces the study power.

   b. Many of the outcomes, including the primary outcome, were predicated on the assumption that staff would be targeting oxygen saturations of 90-95%, as measured through pulse oximetry (SpO_2); it transpired that this was rarely the case. SpO_2 within the 90-95% target range was achieved in 54 of 224 readings (24.1%) overall, by reading:

      - 14 of 51 (27.5%) of readings at baseline;
      - 9 of 50 (18.0%) of readings at 6 hours;
      - 9 of 49 (18.4%) of readings at 12 hours;
      - 15 of 41 (36.6%) of readings at 24 hours;
      - 4 of 24 (16.7%) of readings at 48 hours; and
      - 3 of 9 (33.3%) of readings at 72 hours.

      Exclusion of other outcomes SpO_2 was outside the target readings would have markedly reduced the sample size available, so analysis was continued as planned, including all records, with a note that the analysis may be contaminated. In the case of the primary outcome, for example, the analysis interprets the amount of oxygen required as being reflective of that needed to achieve targeted saturations; when the SpO_2 is above the target range, it is difficult to ascribe meaning to the oxygen level that was provided.

Primary outcome

3. **Descriptive analysis as specified in the protocol**:

   Primary and selected secondary outcomes - descriptive: The raw values of the primary outcome (FiO_2) and five secondary outcomes (SpO_2, respiratory rate, pH, PaO_2, $PaCO_2$) will be graphed to allow visual comparison of differences at baseline and over time. [Protocol, para 5.2]

4. **Descriptive analysis as revised**: No change.

5. **Planned analysis as described in the protocol**:

   Primary outcome: With longitudinal measurement, the interpretation of later results can be impacted by differential losses before 72 hours (death or intubation and ventilation). Analysis will therefore commence with a Kaplan-Meier curve comparing 'loss' (death or intubation and ventilation), and an associated Cox regression model to compare the rate of loss in the
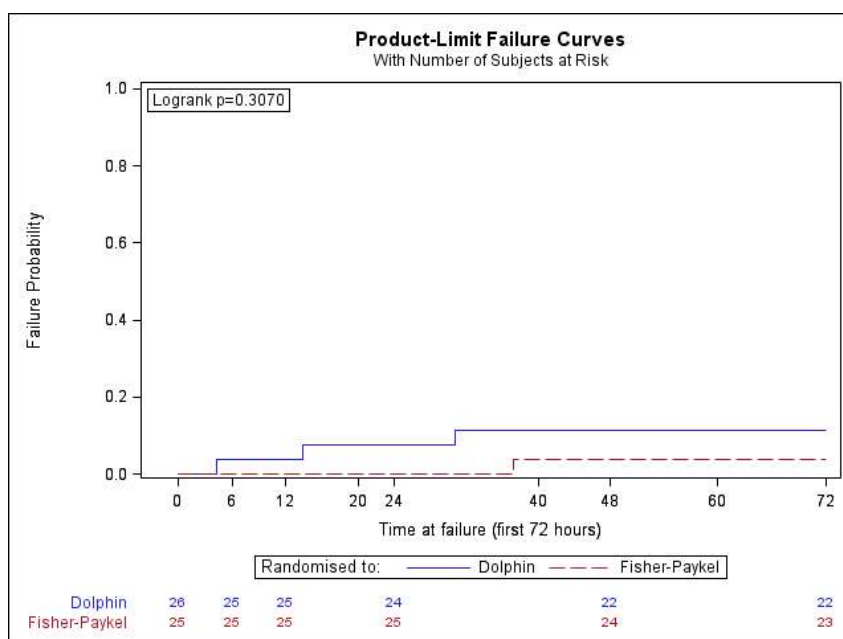
two treatment arms. The results of this survival analysis will provide the context within which the longitudinal results will be interpreted. If there are differential losses over time, the least affected time point will be 6 hours after treatment commencement, and this time point will be promoted to being the primary outcome. If losses over time are not differential, the planned analysis is described below.

The preferred omnibus analysis is a longitudinal mixed effects model (using PROC MIXED in SAS 9.4), controlling for baseline through a random 'intercept', in line with international recommendations for controls on baseline.[30] Individual analyses at each time point (6, 12, 24, 48, and 72 hours) will be by multiple regression, again adjusting for baseline as a covariate.[30] If multiple regression model requirements for normality and homoscedasticity are not met, analysis will be by t-test which is known to be robust to violations of the normality assumption, and will be based on the change in variable from baseline value, providing some control for different baselines, but without the additional control offered by multivariate analysis.

Where performed, multivariate analysis will control for stratification; if univariate analysis is used, separate stratum-specific sub-analysis will be undertaken, and a stratified t-test will be performed. Additional covariates will be adjusted for in multivariate analysis, if the baseline comparison shows substantial group difference in significant prognostic factors (e.g., onset of labour is associated with the stimulation of surfactant production, so analysis will control for mode of delivery if there is a clinically relevant mal-distribution). [Protocol, para 5.2]

6. **Revisions to planned analysis**:

   a. As noted above (para 2b), all data analysed, not just data where SpO_2 was 90-95%

   b. Kaplan-Meier curve produced to assess treatment failure in the first 72 hours, with difference assessed by log-rank test. Three Dolphin and one Fisher-Paykel failure in first 72 hours; not statistically different by log-rank test, p=0.30). Two infants ceased CPAP and were discharged <72 hours (one left against medical advice, the other discharged home), and were excluded from the denominators from the point of discharge. No control for stratification in analysis, as no failures in the older gestational group.

c. While the difference between groups was not statistically significant in survival analysis, change at in FiO_2 from baseline to 6 hours was preferred as the primary endpoint, given the small sample size. As the distribution was strongly peaked, a non-distributional test (Wilcoxon) was preferred, controlling for stratification (van Elteren's extension).

Secondary outcomes

7. **Descriptive analysis as specified in the protocol**:

   Primary and selected secondary outcomes - descriptive: The raw values of the primary outcome (FiO_2) and five secondary outcomes (SpO_2, respiratory rate, pH, PaO_2, $PaCO_2$) will be graphed to allow visual comparison of differences at baseline and over time. [Protocol, para 5.2]

8. **Descriptive analysis as revised**: No change.

9. **Analyses as specified in the protocol**:

   Secondary outcomes: Five of the secondary outcomes (SpO_2, respiratory rate, pH, PaO_2, $PaCO_2$) will be analysed using the same procedures as for the primary outcome. It is anticipated that the three arterial blood gases (pH, PaO_2, $PaCO_2$) will not be performed on all infants, particularly infants that recover rapidly; if so, this analysis will be restricted to change from baseline at the 6-hour reading.

   Duration of CPAP (hours), for those not failing CPAP treatment, is the only continuous secondary outcome this is expected to be non-normally distributed and will therefore be assessed using Wilcoxon's rank sum test, with stratum-specific contributions aggregated using van Elteren's extension to the Wilcoxon test.[31]

   The remaining secondary outcomes are dichotomous and will be tabulated and compared using stratum-specific relative risks (RRs); the homogeneity of stratum-specific RRs will be assessed using Cochran's Q and the I^2 tests[32]_._ [Protocol, para 5.2]

10. **Revisions to planned analysis**:

    a. As with the primary endpoint, change from baseline at the 6-hour reading was adopted as the preferred analysis for SpO_2 (Secondary outcome #2), respiratory rate (Secondary outcome #3), pH, PaO_2, $PaCO_2$ (all reported as part of Secondary outcome #4) and duration of CPAP (Secondary outcome #8). Due to concerns about non-normality, van Elteren's extension to the Wilcoxon test was used to assess statistical significance, while controlling for stratification by gestational age group. No testing was performed for Secondary outcome #4 (pH, PaO_2, $PaCO_2$) as there were only five infants who had readings at both baseline and six hours of age.

    b. Dichotomous secondary outcomes were compared using the Cochran-Mantel-Haenszel test, with assessment for homogeneity using the Breslow-Day test. The exceptions to this were:

    - Where there were no events in one stratum (i.e., CPAP failure [Secondary outcome #5], Serious Adverse Event [Secondary outcome #9], Intubation and mechanical ventilation [Secondary outcome #15] and Death [Secondary outcome #17]), Fisher's Exact test was used on the full data; and

    - Where there were no events in either stratum (Damage to the nasal septum [Secondary outcome #10], Damage to the nares [Secondary outcome #11], Pneumothorax

diagnosed by X-Ray [Secondary outcome #12], Intra-ventricular haemorrhage (IVH), intra-cranial haemorrhage (ICH), or periventricular leukomalacia (PVL) as diagnosed by cranial ultrasound scan [Secondary outcome #13], FiO2 ≥ 60% to maintain SpO2 at 90-95% for one hour or more, during CPAP treatment [Secondary outcome #14] and, for infants born at 28+0 to 33+6 weeks' gestation, oxygen dependent at 36 weeks' gestation [Secondary outcome #16]), no statistical test was performed.

Other issues

11. **Multiple comparisons**: The principal control for multiple comparisons was the designation of a single primary endpoint.

    As an exploratory study, with a small planned sample size determined by time available for recruitment rather than a definitive assessment of a pre-determined clinically significant difference in endpoints, it was decided that no adjustments would be made for multiple comparisons of secondary endpoints, many of which were sentinel events. Rather any statistically significant results would be assessed within the context of the primary endpoint result, and with full acknowledgement of the risks of multiple comparisons.

ENDS