Premature Infants Receiving Milking or Delayed Cord Clamping: Randomized Controlled Multicenter Non-Inferiority Trial (PREMOD2)

Statistical Analysis Plan (v1)

Prepared by

Data Coordinating Center (UAB SOPH)

August 25, 2017

1. Introduction

Extremely premature infants can experience severe bleeding in the brain, or severe intraventricular hemorrhage (IVH) which usually occurs within 72 hours after birth. Approximately 65% of all severe IVH are in infants < 28 weeks gestation, though only 1-2% of overall births are at this gestation. This has significant public health implications, by causing increased death and long-term morbidities in this high risk population. The study intervention is poised to understand if we can reduce this burden, and the trial will provide data on a not yet established therapy. If the strategy is indeed successful, potential benefits for preterm infants are life-long. Our multicenter study will test the research hypothesis that umbilical cord milking is not inferior to delayed cord clamping in premature newborns.

2. Descriptive Statistics

Patient characteristics at randomization will be summarized by randomization group. The two study groups will be labeled as UCM (for Umbilical Cord Milking) and DCC (for Delayed Cord Clamping). For continuous variables, means/medians and standard deviations/interquartile ranges will be reported. To assess and/or identify covariates for which adjusted sensitivity analyses might be conducted, Student t-tests will be used to compare means between study groups. Where appropriate, medians and quartiles will be reported and the Wilcoxon rank sum test will be used as an alternative comparison procedure. Categorical measures will be presented as counts and percentages and will be compared using the χ^2 tests of association to identify potential group differences. For rare outcomes such that the χ^2 test of association is not appropriate, Fisher's exact test will be used. Balance overall is expected because of the large sample size and we expect approximately 5% to be different by chance, since we are not adjusting these baseline comparisons for multiple testing. Any group characteristics that are identified as statistically significantly different between the two groups at a 0.05 level of significance will be considered as covariates in multivariable models in subsequent analyses of the primary study outcome.

3. Primary Analysis

3.1. Primary Outcome: The primary endpoint to be used for efficacy evaluation is the rate of severe IVH (grade 3 or 4) and/or death. The occurrence of 1 or more of these items will be considered an occurrence of the primary study outcome.

3.2. Analysis Plan: The primary hypothesis to be tested is whether the UCM group results in a non-inferior event rate compared to the DCC group. The non-inferiority margin is set at 1% (0.01). The null hypothesis for the formal statistical test is that UCM is inferior to DCC with a 1% non-inferiority margin. That is, the difference in primary outcome rates, $p_{UCM} - p_{DCC}$, is 1% or higher. The alternative hypothesis is that UCM is not inferior. That is, the true difference in rates ($p_{UCM} - p_{DCC}$) is less than 1%. This will be tested based on a 1-sided confidence interval for the true difference in rates ($p_{UCM} - p_{DCC}$). If the upper bound of this confidence interval is completely below 1%, then non-inferiority of UCM will be established.

Although the difference in proportions is used for sample size calculations below, rates in the two study groups will be additionally evaluated in logistic regression models. These models will allow for control of covariates as well as investigation of effect modification. Potential covariates include gender, gestational age, maternal corticosteroid use, chorioamnionitis, preeclampsia, and small for gestational age.[1-4] As randomization is stratified by gestational age (GA), stratified analyses will be conducted in early preterm and late preterm infants. Differential consent practices at sites (antenatal vs. postnatal) may also skew subject acuity/gestation or maternal complications. Clinical site will be used as a stratifying factor to control for any confounding by site through residual, site-level treatment imbalance. Standard regression diagnostics will be used to assess model adequacy and to examine for potential outlying or

influential data points. As sensitivity analyses, generalized estimating equations (GEE) will be used to model outcomes while accounting for any clustering effects resulting from multiple gestations.

Both intent-to-treat and per-protocol analyses will be conducted. Since in a non-inferiority trial an intention to treat analysis biases away from the null, in the per protocol analysis, the covariate of the ordinal scale of the quality of the manipulation for UCM or DCC will be included. If non-inferiority is established by rejecting that the outcome event rate is worse by 1% or more in the UCM group, then superiority will be tested at the 5% level following FDA guidelines. For all superiority testing, the intention to treat analysis will be utilized with a per protocol analysis as a sensitivity analysis.

Prior studies offer no basis for assuming a priori interactions between treatment groups, strata 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 are not powered for with the sample size. We propose 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 establishing consistency and/or generating hypotheses for future studies.

3.3. Sample Size and Power: The initial pilot study of 154 newborns delivered by C/S was recently completed to determine the feasibility and efficacy of this study and revealed a 6 percent difference in the combined outcomes of severe IVH/death between newborns treated with UCM and DCC (4.1 vs. 10.1 percent, respectively). The pilot study was mainly conducted at SMBHWN. However, since severe IVH rates and death may vary from center to center, SMBHWN compared their data to the most recent Vermont Oxford Network (VON) data (over 900 NICUs). For 2015, SMBHWN center had a severe IVH or death rate of 16 percent, close to the 50^{th} percentile for the VON network. However, this includes very high risk-babies who would have been excluded from the trial (e.g., di-amniotic monochorionic twins, placental abruptions, hydrops, and congenital anomalies which have a higher IVH/mortality risk. This likely explains why the Phase 1 pilot PREMOD study had a lower composite number of severe IVH and/or death (10.1 percent vs. 4.1 percent, DCC and UCM respectively). We anticipate UCM and DCC subjects in this trial would have a similar incidence of this outcome.

The sample size for non-inferiority testing for infants born by C/S in each group is 502 (overall sample 1500), a two-group large-sample normal approximation test of proportions with a one-sided 0.05 significance level will have 90% power to reject the null hypothesis that the UCM is inferior to DCC (the difference in proportions, pUCM – pDCC, is 0.01 or higher, a 1% non-inferiority margin) in favor of the alternative hypothesis that the proportions in the two groups are not inferior, assuming that the expected difference in proportions is -0.04 and the proportion in the DCC group is 0.10. (Note using 0.101 yields 485, so we round the proportion difference down to be conservative).

Further, to show the sample size for C/S is adequate we examined the power to detect the difference between 0.10 for the DCC group and 0.04 for the UCM arm with 502 newborns per group. A two group χ^2 test with a 0.05 one-sided significance level will have 98% power to detect the difference between the DCC group proportion, p_{DCC}, of 0.10 and the UCM group proportion, p_{UCM}, of 0.04 (odds ratio of 0.375) when the sample size in each group is 502 and 75% power to show superiority, if the rate is 0.06. Both non-inferiority and superiority will use the same sample and will have the ability to test two hypotheses in a systematic manner for each aim.

3.4. Interim Monitoring: The study will be closely monitored for issues of data quality, study conduct, adherence to the prescribed treatment procedures, data quality, and adverse events. Reports including adverse events, serious adverse events, and primary outcomes will be distributed twice annually and reviewed by the DSMB. All reports will present information in masked format to prevent unblinding.

Interim analyses will be conducted at the discretion of the DSMB. As a non-inferiority trial, we are unlikely to demonstrate non-inferiority before reaching the target sample size. Early study termination would likely be the result of unexpected safety concerns and not efficacy. Early stopping based on inferior safety must be based largely on descriptive data and close examination of adverse

events. Given the projected number of subjects, early stopping is unlikely unless the observed effect of UCM is substantially worse than DCC or there are unexpected adverse results potentially seen. With 500 subjects per group at a second early stopping review (n=1000), and assuming that the UCM Group is actually no worse than DCC, observed risk in the experimental group would have to be at most 0.5 (risk of death and IVH) to have 80% power to show non- inferior efficacy (with $\alpha = 0.012$). 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. Additional analyses will be presented to the DSMB to ensure consistency over and above an appropriate p-value for termination.

4. Secondary Analyses

Secondary outcomes will be analyzed using similar procedures to the primary outcome. Comparisons between treatment groups will use logistic regression (dichotomous outcomes), linear regression (continuous outcomes), or survival analysis (survival time outcomes, such as time to discharge, etc.), as appropriate. Differential practices at sites (criteria for phototherapy or blood transfusions) may also skew these secondary outcomes. Therefore, the clinical site will be used as a stratifying factor to control for any confounding by site through residual, site-level treatment imbalance. Neurodevelopmental follow-up results will be assessed using ANCOVA models with covariates used for analyses of the primary outcome. BSID-III scaled and composite cognitive, language and motor scores may be compared.

References

1. Mercer JS, Erickson-Owens DA, Vohr BR, et al. Effects of Placental Transfusion on Neonatal and 18 Month Outcomes in Preterm Infants: A Randomized Controlled Trial. *The Journal of Pediatrics* 2016;168:50-5.e1.
2. Linder N, Haskin O, Levit O, et al. Risk factors for intraventricular hemorrhage in very low birth weight premature infants: a retrospective case-control study. Pediatrics 2003;111:e590-5.
3. Salafia CM, Minior VK, Rosenkrantz TS, et al. Maternal, placental, and neonatal associations with early germinal matrix/intraventricular hemorrhage in infants born before 32 weeks' gestation. *American journal of Perinatology* 1995;12:429-36.
4. Ment LR, Oh W, Philip AG, et al. Risk factors for early intraventricular hemorrhage in low birth weight infants. *The Journal of Pediatrics* 1992;121:776-83.