Study Title: Cooling to Help Injured Lungs (CHILL) Phase IIB Randomized Control Trial of Therapeutic Hypothermia in Patients with ARDS

NCT No. pending
Document: Statistical Plan 1.0
Date: July 30, 2020

Funding source: U.S. Department of Defense (W81XWH2010432)

CHILL Study PI: Jeffrey Hasday, MD

Contact phone number 410-916-5349

Email: Jhasday@som.umaryland.edu

CHILL Clinical Coordinating Center Director: Carl Shanholtz, MD
Contact phone number 410-598-4022
Email: cshanhol@som.umaryland.edu

CHILL Study DSMB Chair: James Shelhamer, MD
Contact phone number 240-204-2093
Email: shelhamer4001@gmail.com

Statistical Design and Power:

a. Sample Size and Stratification Design: We performed a power analysis to detect a 4-day increase in 28-day VFDs between the TH arm and the control arm. A 4-day increase was found in the PROSEVA trial of proning. We assumed a standard deviation (SD) of 10, the average SD among the ACURASYS, PROSEVA, and ROSE trials, resulting in the effect size to be detected equal to 4/10 = 0.40. Setting power = 0.90 and alpha level =0.05 (two-tailed), we calculate n = $2*[(1.96+1.28)^2]*(10)^2/(4)^2 = 131.2$. The sample size was then corrected to allow 10% one-way crossover assuming that subjects randomized to TH+NMB may not receive study treatment because of technical issues but that subjects randomized to standard treatment would not likely receive hypothermia treatment. The 10% is conservative (resulting in a larger total N) as the cross-over is typically ~3-4% in similar trials. This adjustment resulted in a sample size equal to $131.2/(1-.10)^2 = 162.0$ per group. Hence, the final total N = 324. Since proning has been shown to improve survival in ARDS patients when added to NMB [1], randomization will be stratified for baseline proning status (i.e. for both center and proning status) and decisions to start or stop proning will be based on protocolized study-wide criteria.

b. Data Analysis Plan: Primary and secondary analyses will be performed according to the principle of intention-to-treat. Per protocol analyses will also be performed as part of assessing the feasibility of a civilian clinical trial with a mortality outcome. The treatment group difference on the primary outcome, 28-Day VFD's following randomization will be tested with a mixed effects model with 28-Day VFD's as the dependent variable, treatment group as the primary independent variable, proning status (binary) as covariate, and treatment center as a random effect. The alpha level of the test will be 0.05. The random treatment center effect will account for site variability and intra-site correlation. As prior studies and our own pilot data indicate that 28-Day VFD's is a skewed variable with more zeros than would be expected in a normal distribution we will select a mixed effects model that is of adequate fit for this distribution and use robust Huber-White standard errors.

We will examine two interactions in secondary analyses. First, we will examine variability of treatment effect by treatment center by adding an interaction term between the treatment group indicator and site. Non-zero site variability will be tested by likelihood ratio test of improved fit comparing the model to the model without this random interaction term. If there is significant site variability we will estimate and report treatment effects by site (coded) as well. Second, we will test whether the treatment effect differs by proning status by adding the interaction term between treatment group indicator and proning status. If this interaction is significant we will estimate and report treatment effects separately by proning status.

Three interim analyses will be performed after ~25% (n=81), ~50% (n=162) and ~75% (n=243) of planned enrollment has either been followed-up for 28 days or has died prior to the 28th day. To inform the decision whether to stop the trial early due to superiority of the TH+NMB treatment arm (greater 28-Day VFD's), we will use the Lan-DeMets [2] alpha spending function with O'Brien-Fleming [3] boundaries, i.e. $\alpha(t^*) = 2 - 2\Phi\left(Z_{\alpha/2}/\sqrt{t^*}\right)$ where $t^* = n/N$ and n equals the interim sample size (i.e., 81, 162, 243). Whether to stop the study early due to futility at an interim analysis will be informed by a conditional power calculation [4, 5]. We approximated conditional power (alpha-level=0.05) for detecting a group difference in mean VFD's using the method of Lan and Wittes [5] for a range of differences. For example, at the 2nd interim analysis, conditional power was 0.85, 0.41, 0.07, and 0.00 when the difference in mean VFDs = 3, 2, 1, and 0, respectively. Based on these calculations, if the difference was 1 or less,

we would recommend that the trial be paused due to futility as the chances of finding a VFD difference that would be compelling would be small.

Several secondary analyses will be performed. First, group comparisons will be performed on baseline clinical and physiologic measurements with potential prognostic value. If there are one or more significant group differences on prognostic baseline variables, we will perform secondary analyses using the above mixed effects model to adjust for these variables. Second, group comparisons will be performed on secondary clinical outcomes and physiologic variables using mixed effects models, as above, including adjustment for proning status and including site as a random effect. Variable transformations such as ln(y) for right skew and y^2 for left skew will be used if needed to satisfy model assumptions. An interaction term will be added between treatment group indicator and baseline proning status and, if significant, separate results of these secondary analyses will be reported by baseline proning status. A mixed logistic regression model will be used to compare the proportion who die within 60 and 90 days. For outcomes with missing data that are not obviously informative, regression multiple imputation [6, 7] will be used. Third, a repeated measures mixed model (RMMM) ANOVA will be used to test group differences on physiologic and biomarker outcomes over days 0, 1, 2, 3, 4, and 7 (i.e., 0, 24, 48, 72, 96, and 168 hrs). To address the issue of truncation due to death and minimize resulting bias we will use a "pattern-mixture" model [6] approach to estimate group mean trajectories separately for different days of death. This approach will be implemented using the above RMMM stratified by day to death. We will test linear contrasts from the above fitted model to test the hypothesis that the level of cytokines (IL-6 and IL-8) are significantly different in the two study arms at the 48-hr time point and remain different out to the 168-hr time point. If there are significant group differences, we will follow this with analysis of group trajectories. For other types of missing data not due to death, if they occur more than infrequently, we will generally use multiple imputation.

References:

1. Guerin, C. and J. Mancebo, *Prone positioning and neuromuscular blocking agents are part of standard care in severe ARDS patients: yes.* Intensive Care Med, 2015. **41**(12): p. 2195-7.
2. Lan, K. and D. DeMets, *Discrete Sequential Boundaries for Clinical-Trials.* biometrika, 1983. **70**: p. 659-663.
3. O'Brien, P.C. and T.R. Fleming, *A multiple testing procedure for clinical trials.* Biometrics, 1979. **35**(3): p. 549-56.
4. Halperin, M., et al., *An aid to data monitoring in long-term clinical trials.* Control Clin Trials, 1982. **3**(4): p. 311-23.
5. Lan, K.K. and J. Wittes, *The B-value: a tool for monitoring data.* Biometrics, 1988. **44**(2): p. 579-85.
6. Kurland, B.F., et al., *Longitudinal Data with Follow-up Truncated by Death: Match the Analysis Method to Research Aims.* Stat Sci, 2009. **24**(2): p. 211.
7. Sterne, J.A., et al., *Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls.* BMJ, 2009. **338**: p. b2393.