Sample Size Analysis: Since the objective of this pilot study is to estimate the benefit of therapeutic hypothermia (TH in ARDS and the effect size on various outcome measures, the sample size (N=32) was determined based on time, resources, and predicted volumes of eligible patients. To approximate the precision of varying observed group differences on the primary (28-day ventilator-free days; VFDs) and a few key secondary outcomes using this sample size (N=32), we used outcome values from the full unmatched historical control sample (N=58) from our prior open-label trial with historical controls [38]. This assumption is more conservative than using the subset of matched controls. For VFDs in our prior study, we observed means of 10.38 VFDs in the cooled group and 5.54 VFDs in the controls with standard deviations of 11.16 and 7.96, respectively. Table 1 displays approximated 95% percent confidence intervals for differences of means for the primary and two secondary outcomes, 28-day ICU-free days (ICU-FDs) and oxygenation index (OI). In the proposed CHILL-pilot RCT we have reduced the need for invasive arterial gas measurements by substituting oxygenation saturation index (OSI) for OI. Based the high correlation between OI and OSI in ARDS patients [118], we have estimated precision of OSI measurements based on OI results in our prior study. The projected confidence intervals are provided to give a sense of the precision we will have given the number of subjects planned for CHILL-pilot RCT. Except for the mortality comparisons, these are t-distribution confidence intervals which rely on the robustness of t-distribution methods when there is deviation from normality, particularly for the 28-Day VFD and 28-Day ICU-FD outcomes. Our approach was to set the control group mean fixed at the estimate from the historical controls from the previous study and calculate 95% confidence intervals for varying potential CHILL group means and group differences. For VFDs and ICU-FDs we chose plausible rounded differences. For OI the range was based on increasing and decreasing the mean in half standard deviation increments. We took this approach as we viewed the estimates from the historical controls as substantially more stable (N=58) vs. the estimates from the cooled patients (N=8). For each outcome the result for the line with the footnoted values displays the estimated precision for the currently proposed trial if the means were the same as were found in the prior study. The precision estimates are adjusted (increasing their length) for an assumed 10% cross-over from the CHILL intervention to the control intervention. We assumed 10% of subjects randomized to TH may not receive study treatment because of technical issues but that no controls were likely to receive TH treatment. This is a conservative estimate based ARDSNet FACCT [56] and SAILS [119] trials, which had crossover rates of 2.3% and 3.4%, respectively.

Table 1. Precision of Estimates of Group Difference for N = 32 Randomized Patients (16 per group)

Outcome	Projected Means CHILL	Control	Δ	SE_Δ [a,b]	95% Confidence Interval Lower Limit	Upper Limit
28-Day VFD's	7.5	5.5	2.0	3.39	-5.0	9.0
	9.5	5.5	4.0	3.68	-3.6	11.6
	10.4[c]	5.5[c]	4.9[c]	3.80	-3.0	12.8
	13.5	5.5	8.0	4.27	-0.8	16.9
28-Day ICU-FD's	6.0	4.0	2.0	2.93	-4.6	8.6
	8.4[c]	4.0[c]	4.4[c]	3.19	-2.2	11.0
	10	4.0	6.0	3.38	-0.6	12.6
	12	4.0	8.0	3.61	1.4	14.6
Day 3 Oxygenation Index	15.5	17.1	-1.6	2.50	-6.8	3.7
	13.4	17.1	-3.7	2.50	-8.9	1.6
	9.2[c]	17.1[c]	-7.9[c]	2.50	-13.1	-2.6
	7.1	17.1	-10.0	2.50	-15.3	-4.8
	5.0	17.1	-12.1	2.50	-17.4	-6.9
Hospital Mortality	0.19[d]	0.50[d]	-0.31	0.18	-0.66	0.04
	0.25	0.53[c]	-0.25	0.18	-0.64	0.08
	0.31	0.50	-0.19	0.19	-0.56	0.18
	0.38	0.50	-0.12	0.19	-0.50	0.26

[a] Standard errors based on variance estimates from open-label trial
[b] Standard error calculation accounts for 10% cross-over from CHILL to treatment as usual (non-compliance). Dropout assumed to be negligible.
[c] Estimates from open-label trial with (8 CHILL and 58 Control patients).
[d] Proportion who died in the hospital.

Planned Statistical Analysis. Estimation of treatment effects with 95% confidence intervals will be the general approach to address both Aims 1 and 3, respectively, evaluating TH for its potential treatment effects and for estimating parameters (differences, variability) for planning a larger, more definitive clinical trial. Primary and secondary analyses will be performed according to the principle of intention to treat. For the primary outcome, 28-day VFDs, we will estimate the difference in group means with a bias-corrected 95% bootstrap confidence interval [120]. We will also estimate the difference of group medians by bootstrapping since 28-day VFDs has zero-inflation and thus the median has advantages as a measure of center. The same general bootstrap method will be used to make group comparisons on the secondary outcomes listed in Aim 1 such as differences in proportion of deaths and mean 28-Day ICU-FDs and day 3 OSI. Mixed model repeated measures (MMRM) ANOVA will be used to estimate group differences on physiologic and biomarker outcomes over days 0, 1, 2, 3, 4 and 7 as analyses using this model are unbiased if the missing mechanism is non-informative (e.g. due to research staff error). If there are significant group differences we will follow this with an analysis of group trajectories. Our primary approach to missing data is prevention. To this end we have included several control points to minimize incomplete data entry including (1) education and training, (2) automated pop-up messages in the electronic data capture system, (3) real-time data review, and (4) monthly reports of data quality to be reviewed by the CHILL Executive Committee. In addition, we have developed the following plan to handle missing data at the analysis stage.

If less ≤10% of observations are missing for an outcome variable we will restrict the analyze to the available outcome data. If an outcome is missing >10% of total observations we will use a multiple imputation procedure [121]. For repeated measures truncated by death, individual trajectories grouped by day of death will be explored graphically. For Aim 2 (safety), we will compare the proportion with each complication between treatment groups with a chi-square test, unless assumptions for this test are not met, in which case we will use Fisher's exact test. Randomization will be stratified by prone positioning status and exploratory subgroup analyses will used to compare treatment groups among those who are and are not prone positioned. Furthermore screening yields (% eligible out of all patients screened, % randomized out of patients screened and out of eligible patients) will be estimated with 95% confidence intervals as part of the feasibility assessment. We will also estimate two other screening rates: (1) number eligible divided by number screened, and (2) number randomized divided by number eligible.