Overview

This report contains a brief summary of the statistical protocol as described by "consulting firm", as well as a number of critiques of and recommended additions to the statistical methods. The recommended changes include a change in the primary endpoint (Length of Stay), however, due to cost limitations, using Length of Stay (LOS) as the primary endpoint in this study may be the best option. Although LOS may not be the most clinically meaningful primary endpoint, we maintain that an argument can be made for LOS as being clinically relevant, if we can show that LOS is strongly indicative of the health of a patient.

ConsultingFirm Protocol Summary

The protocol provided by "consulting firm" on January 22nd 2017 describes the Statistical methods to be used in analyzing data from clinical trials. The experiment is a"Prospective, Double Blind, Randomized Multi-Center, Evaluation of Efficacy Safety and Tolerability of Nitric Oxide Given Intermittently via Inhalation to Subjects with Bronchiolitis".

The protocol outlines a sample size calculation that is based on the use of LOS (length of stay of each subject) as the endpoint. This sample size calculation is based on an unpaired two-sample t-test, using balanced groups and a pooled (common) variance. The populations that will be analyzed are described, including the entire population of test subjects (the safety population), the ITT (intent to treat) population, and the per protocol (PP) population. A description of how the categorical and continuous data will be tabulated and presented is provided, including which sample statistics will be included (95% CI, median, standard deviation, etc.).

Using LOS as the endpoint (response), the protocol describes the proposed efficacy analysis using two sample t-tests and Wilcoxon-Mann-Whitney Rank Sum tests to test for significant difference between the treatment group and the control. Using 'time to achieve a clinical score of 5' as the endpoint (response), the protocol once again proposes the use of two sample t-tests and Wilcoxon-Mann-Whitney Rank Sum tests to test for significant difference between the treatment group and the control. The protocol also proposes conducting a Survival Analysis using the

non-parametric Kaplan-Meier survival function curves, with 'time to achieve a clinical score of 5' as the endpoint (response).

The remainder of the protocol describes standard procedures for assessing safety. Coding of all adverse effects according to coding dictionaries is described, such as Treatment-Emergent Adverse Events, drug-related adverse events , as well as other observational endpoints including selected laboratory results.

Proposed Additions/Modifications

Choice of Primary Endpoint

The first recommendation concerns a change in the choice of primary endpoint. The FDA describes a direct endpoint as a "clinically meaningful endpoint that directly measure(s) how a patient feels, functions, or survives", and is customarily the basis for approval of new drugs. The "time to reach a clinical score of 5" can be considered to be a subjective direct endpoint, as it is clinically significant and directly measures how a patient feels and functions. An argument can be made for why Length of Stay (LOS) is a direct measure of patient improvement, and is therefore a good primary endpoint. This would require us to identify a how the health of a patient can be defined by the duration of their stay at the hospital.

Optimally, we recommend recalculating the sample sizes based on the use of "time to reach a clinical score of 5" as the primary endpoint. These calculations have been run for the mITT group, with $LOS > 24$, $LOS > 36$, as well as for all the subjects. Considering the dropout rate, as well as the proportion of subjects with $LOS > 24$ or $LOS > 36$, the following sample sizes have been determined using a two-sample t-test with a significance level of 0.05 and a power of 0.8. We recommend confirming our calculations.

It is understood that due to cost limitations, the upper limit to the sample size will likely be around 80. The use of LOS as the primary endpoint may be the only option considering the cost limitations of the study.

Table 1: Estimated Sample sizes with "time to reach a clinical score of 5" as the primary endpoint

Population	Sample Size (total patients enrolled in study)
mITT with LOS >24	104
mITT with LOS >36	102
all mITT subjects	290

Exploratory Analysis

The validity of the clinical trial relies on the random allocation of subjects to the control/treatment groups. This randomization should be checked across all covariates relating to the initial condition of the subject (for example the initial clinical score for each subject). This can be done graphically using box-plots. If there are significant differences between the values of explanatory variables between the treatment group and the control, the 'conditional change model' should be implemented: this means including the initial values of the specified explanatory variables as covariates in a linear model.

Imputation

Imputation is the process of replacing missing data with substituted values. The nature of how the data is missing can create bias in the data, which imputation aims to remedy. In addition, imputation aims to improve the efficiency of the analysis. There are a number of methods for doing so, many of which involve predicting the missing values using the remaining explanatory variables. A 'filled in' data set can be generated and additional analysis can be provided using this 'filled in' data.

The presence of missing values should be assessed in the exploratory analysis. If missing values are present at random, there is may be no need for imputation, but imputation methods can still be implemented without risking introduction of bias. In either case, when the data is missing at random or not, stochastic imputation is recommended if data is missing from only the response variable, and multiple imputation is recommended if data is missing from all variables.

ANCOVA / Linear Regression Analysis

Currently, the issue with using a two-sample t-test or a Wilcoxon Rank Sign test is that they do not incorporate additional covariates that could potentially explain some of the variation in the data, thereby leading to a more precise assessment of the treatment effects. It is recommended that a linear regression analysis or ANCOVA be implemented to account for additional explanatory variables such as the initial clinical score for each subject, the age of the subjects, their particular type of infection, among other variables. In these cases, testing for statistical significance of the treatment would involve testing for a significant difference between treatment coefficients. This provides a better estimate of the effects of the treatment, as more of the variation in the data is explained by the additional variables.

Intent-to-Treat vs As-treated analysis

The current proposal is an 'intent to treat' analysis in that the condition of the subjects is assessed based on the intent to treat them or not (whether they are in the treatment arm or the control arm), rather than the amount of treatment they actually receive. An 'as-treated' analysis is recommended to supplement the results. In this clinical study, an 'as-treated' analysis would involve regressing the primary endpoint against the actual amount of Nitric Oxide inhalation received by each subject. Such an analysis may yield results from the entire population, rather than having to rely on only results from the population with $LOS > 24$ or $LOS > 36$.

Normality Assumption/ Non-parametric Alternatives

The assumption of normality of residuals may not be satisfied when considering only the primary endpoint, however, this assumption may be satisfied once other explanatory variables are considered. In the case where this assumption is still not satisfied after additional explanatory variables are considered, there remains the unconventional option of using non-parametric multivariate regression methods (such as Gaussian Processes).

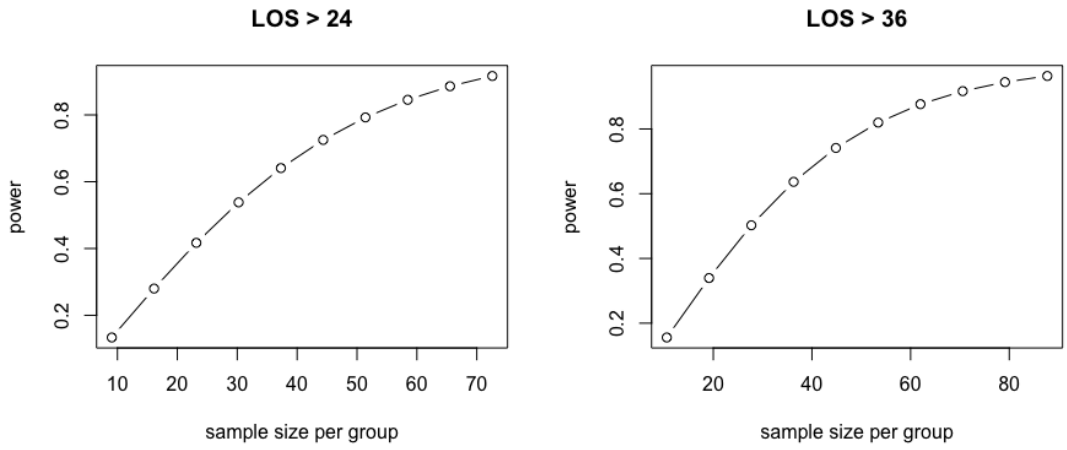

Figure 1: For all subjects in the mITT group with $LOS > 24$ and $LOS > 36$ hours, this plot illustrates how the probability of a false negative decreases with an increase in sample size (Power $= 1 -$ probability of false negative).

Using LOS as the Primary Endpoint

The statistical 'Power' of a test can be interpreted as the probability of not committing Type 2 Error. In the context of this clinical trial, a Type 2 Error would mean not being able to detect a statistically significant difference between the treatment and the control when truly there is a significant difference. Therefore, the higher the power, the lower the chance of identifying a false-negative. Given the use of Length of Stay (LOS) as the primary endpoint in this clinical trial, we have constructed the plots in Figure 1 to illustrate how the power of the test is compromised with reduced sample size.

Conclusion

Although a change in the primary endpoint results in the need for more subjects than this study can afford, a more in-depth analysis can accommodate for the limited sample size: Including additional explanatory variables via a regression analysis will help explain more of the variation in the data. Including variables that have to do with the initial conditions of subjects will help in explaining varia-

tion due to imbalances between the treatment group and the control. Imputation will help reduce the bias that is caused by missing values, and may increase the efficiency of the analysis. Finally, an 'as-treated' analysis may prevent us from having to focus on only a small subset of the subject population. It should also be noted that the recommended methods are not only allowed but recommended by the FDA, as per the documents linked below.

Sample FDA documents

1. "Clinical Investigator Training Course - Clinical Trial Endpoints"
2. "Guidance for Industry, Statistical Principles for Clinical Trials"

1. http://www.fda.gov/downloads/Training/ClinicalInvestigatorTrainingCourse/UCM283378.pdf
2. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm073137.pdf