Clinical Trial Simulation to Compare Adaptive and Fixed Designs for a Phase 3 Clinical Trial of Nacystelyn®

(L-Lysine-N-acetyl-L-cysteinate) for Cystic Fibrosis

James A. Rogers * (1), Matthew M. Riggs (1), Monte Coffiner (2), William R. Gillespie (1), Robert Burford (3), Francis Vanderbist (2), Sophie De Niet (2), Marc R. Gastonguay (1)

(1) Metrum Research Group LLC, Tariffville, CT ; (2) Laboratoires S.M.B., Brussels, Belgium ; (3) Aptuit Consulting, Greenwich, CT.

Abstract

Background: *Nacystelyn® (L-Lysine-N-acetyl-L-cysteinate; "NAL") is being developed as a mucolytic and/or anti-inflammatory agent for use in Cystic Fibrosis. Objectives for a phase 3 clinical trial of NAL include dose selection as well as confirmation of efficacy and safety at the selected dose.*

Objectives: *Both conventional fixed designs and two-stage adaptive designs with dose selection were considered to address the phase 3 objectives. The objective of the research described in this poster was therefore to evaluate the relative merits of the different candidate designs with respect to operating characteristics related to power and quality of dose selection.*

Methods: *Models based on both existing Nacystelyn data and available literature data were developed to describe background hazard rates, plausible dose-response, covariate distributions, and attrition rates. A utility function was elicited in order to formalize the definition of true "best dose". The models were then used in conjunction with both clinical trial simulation and established statistical theory in order to evaluate candidate clinical trial designs and development strategies with respect to operating characteristics (statistical power, quality of dose selection, etc.), number of patients, and total trial duration.*

Results: *In general, the fixed and adaptive designs were essentially undifferentiated with respect to overall power, however the fixed design had a clear advantage with respect to the probability of rejecting the null hypothesis for the suitably defined best dose.*

Conclusions: *Based on estimated primary endpoint event rates and assumed recruitment rates, quality dose selection is unlikely to occur prior to full enrollment of a trial. A single fixed design including placebo and 3 active NAL dose groups (10, 20 and 40 mg BID) is therefore the most efficient strategy for selecting an appropriate dose and demonstrating efficacy at this dose.*

Methods

Models

Statistical models were developed based on existing NAL data and available literature data to describe:

- baseline hazard rates (estimated from Pulmozyme®(rhDNase) 6-month Phase III trial [1]) for time-to-first Pulmonary Exacerbation (PE, the primary efficacy endpoint) as a function of:
  - background treatment status (binary indicator for use of rhDNase 2.5 mg QD),
  - patient age (estimated from Goss and Burns [2]),
  - patient-level and study-level random effects.
- a plausible range of NAL efficacy dose-response scenarios was assessed in the context of a dose-response model for time-to-first respiratory infection (estimated from two NAL phase 2 studies), as a respiratory infection was considered to be a good physiological indication that a CF patient will proceed to develop a PE,
- likely age (covariate) distribution for given inclusion criteria,
- discontinuation rates (modeled as a constant hazard for the duration of the study) to approximate literature reports [1, 3, 4].

Candidate Trial Designs

The primary endpoint for all candidate designs was time to first pulmonary exacerbation (PE). In all cases the randomization was stratified to acheive 50% representation of rhDNase 2.5 mg QD background therapy. Additional features held constant for all candidate designs included: inclusion / exclusion criteria, treatment duration, recruitment rates, control of Type I error, and the definition of dose optimality. Precisely fair comparison with respect to sample size is difficult, since sample size for the candidate adaptive design is a random number whose distribution depends on unknown parameter values, however an attempt was made to maintain total sample size at approximately 900 patients for all designs considered.

The two fundamental classes of designs evaluated were fixed designs and two-stage adaptive group sequential designs (also known as "flexible designs") as described below:

Fixed Designs

Fixed randomization probabilities are used to allocate patients to one of the following: placebo, NAL 10 mg BID, NAL 20 mg BID, or NAL 40 mg BID using 1:1:1:1 allocation. The primary analysis is:

> Each dose level of NAL will be compared to placebo using the Wald test based on the Cox proportional hazards model [5]. Wald test p-values will be multiplicity adjusted using the fixed sequence step-down procedure [6], proceeding from the highest dose to the lowest.

Adaptive Designs

Randomization probabilities are fixed at 1:1:1:1 during the first recruitment stage (duration approximately 1 year), an interim analysis employing a Bayesian dose-response model is used to select a single dose from among the three, and stage 2 proceeds with 1:1 randomization to placebo and the selected dose. The primary (end-of-trial) analysis is then:

> A suitably defined *stage 1 analysis set* is used to compute stage 1 p-values for the Wald tests comparing each active dose to placebo, based on the Cox proportional hazards model and multiplicity adjusted using the fixed sequence step-down procedure [6]. A stage 2 p-value was to be similarly computed based on a suitably defined *stage 2 analysis set* for the selected dose (without multiplicity adjustment, because there is only one comparison in this stage). The stage 1 and stage 2 p-values for the selected dose are then combined using the inverse normal p-value combination function [7] and rejection of the null hypothesis for the selected dose is based on this combination p-value.

Definition of Dose Optimality and Dose Selection Methodology

As no target toxicities had been identified, the true "best" dose was operationally defined as the lowest dose (among those tested) with a true hazard reduction of 30% relative to placebo. In keeping with this definition, the following utility function u was defined:

$$u(d) = f(d) - p(d) \tag{1}$$

where f is the dose-response function, i.e. $f(d)$ gives the relative efficacy at dose d mg BID and p is a penalty function, defined as:

$$p(d) = \begin{cases} 0 & \text{if } d \le d^* \\ \alpha(d - d^*)/d^* & \text{otherwise.} \end{cases}$$

where $d^* = f^{-1}(0.3)$, i.e. the dose that would elicit exactly a 30% response.

For $\alpha > 0$, this definition of utility is consistent with a qualitative determination that the "best" dose is the minimum dose with at least a 30% effect (hazard reduction) relative to placebo. Based on visual inspection of figures such as Figure 1, and based also on exploratory simulation results, the tuning parameter α was set at the value 0.25 in order to meaningfully differentiate the 20 mg BID dose group from the 40 mg BID dose group under scenarios where 20 mg BID was the "best" dose.

The dose selection strategy for adaptive designs was then to:

- Use a Bayesian model fitting procedure to compute the estimated dose response function $\hat{f}$.
- Use $\hat{f}$ to compute the derived values $\hat{u}(d)$ for each experimental dose.
- Select the dose with the highest estimated utility as the dose to carry forward to the second stage of the trial.

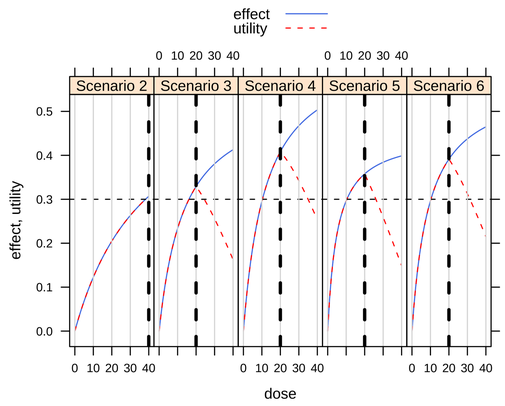

Figure 1: Dose response and utility under various scenarios. The vertical dashed line in each panel represents the "best dose" with respect to the utility function.

Operating Characteristics

The two operating characteristics identified to be of primary interest were overall power (probability of rejecting the null hypothesis for at least one dose) and the probability of rejecting the specific null hypothesis for the true "best" dose.

Adaptive designs were evaluated with reference to a presumed "best case" (i.e. a version of the adaptive design presumed to have the best possible chance of outperforming a fixed design with respect to key operating characteristics), wherein:

- Stage 1 enrollment was specified to continue until at least one of the following milestone had been reached:
  - 600 patients enrolled.
  - 80 pulmonary exacerbation events observed.
  - 12 months elapsed since first-patient-first-visit.
- At the end of the first stage, the pre-specified dose selection strategy described above is applied to select a single active dose which is carried forward with placebo into stage two.
- The number of subjects enrolled into the second stage of the trial is fixed at 300.

Operating characteristics were evaluated by Monte Carlo simulation (for the adaptive design) and by established statistical theory (for the fixed design), under a range of plausible alternative hypotheses. In all cases, parameter estimates from the models described above formed the basis for the computation / estimation of operating characteristics.

Computing

Clinical trial simulations were carried out in using R 2.6.2 [8] using eight-fold parallelization on Windows 2003 Server x64 Edition. The Rmpi package (http://www.stats.uwo.ca/faculty/yu/Rmpi) was used to facilitate parallelization from the R environment. For the Markov Chain Monte Carlo (MCMC) computations associated with Bayesian model fitting, calls were made from R to WinBUGS 1.4.3 [9] using the R2WinBUGS package (http://www.stat.columbia.edu/ gelman/bugsR/).

The simulation of each clinical trial required, on average, approximately one minute of CPU time. For two of the more critical reference scenarios, 1000 clinical trials were simulated, resulting in probability estimates with standard errors of, at worst, 1.6% (based on standard binomial calculations); for one of the reference scenarios, 400 clinical trials were simulated, leading to standard errors of, at worst, 2.5%; for one of the reference scenarios, 104 clinical trials were simulated, leading to standard errors of, at worst, 5%. These standard errors turned out to be suitably small to permit clear conclusions vis-à-vis the relative merits of a fixed versus adaptive design.

Results

Event Rates, Covariates, and Attrition

Model-based estimates used in the simulations included the following:

- PE event rate was estimated at approximately 30%–40% per year for younger patients ($\le$ 17 years) and 40%–50% for older patients ($>$ 17 years). Event rates were estimated to decrease by approximately 5 percentage points with the addition of rhDNase therapy.
- The age (covariate) distribution was modeled using a lognormal distribution with a geometric mean of 17 years (approximately 55% CV) and truncated to the range 12–55 (per inclusion / exclusion criteria).
- Drop-out was modeled using a constant (exponential) hazard, with estimates corresponding to approximately 12% dropout per year.

Dose-Response

Based on data from the phase 2 NAL-96-12-RD trial, an approximate 30% and 50% hazard reduction relative to placebo was estimated for 8 mg BID and 16 mg BID NAL treatments, respectively. As a respiratory infection may be considered a good physiological indication that CF patients would proceed to develop a PE, it was expected that this reduction in first respiratory infection rate will translate into a consistent reduction in PE. However, since the exact relative magnitude of such a reduction was unknown, a plausible range was to be investigated during the trial simulation/optimization (see Figure 2).

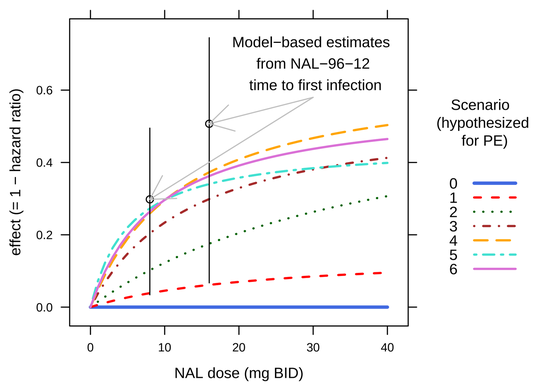

Figure 2: Dose-response scenarios for pulmonary exacerbation for which the trial designs were evaluated. The estimates and 95% credible intervals for the effects of 8 and 16 mg BID NAL on infection rates are superimposed for reference.

Sample Sizes

Simulation of the adaptive design using the specified rules for transitioning from stage 1 to stage 2 resulted in sample sizes described in Table 1. Given this distribution of potential sample sizes for the adaptive design, a fixed design using 900 patients was considered to be a fair comparator.

Stage	Median / Maximum Sample Size				
	Placebo	10 mg BID	20 mg BID	40 mg BID	Total
1	146 / 157	146 / 157	146 / 157	146 / 157	583 / 627
2	150 / 150	0 / 0	150 / 150	0 / 0	300 / 300
Total	296 / 307	146 / 157	296 / 307	146 / 157	883 / 927

Table 1: Approximate median and maximum sample sizes for each stage and treatment arm for the "best case" adaptive design under scenario 4, assuming that 20 mg BID is selected at the interim. The "Total" column was estimated directly from simulation and the approximate sample sizes in each arm were derived from this (with truncation to whole number values).

Operating Characteristics

Fixed Design

Operating characteristics for the fixed design were approximated using standard asymptotics in conjunction with the methodology of Genz [10, 11] (to compute asymptotic joint probabilities for Wald statistics) and are provided in Table 2.

Scenario		Asymptotic Approximation of Probability (%)	
	Best Dose	Best is Significant	Either is Significant
2	40 mg BID	59	59
3	20 mg BID	62	87
4	20 mg BID	85	98
5	20 mg BID	67	84

Table 2: Asymptotic approximations to probabilities for the fixed design with 900 patients.

Adaptive Design

Preliminary simulation of the adaptive design demonstrated that the proposed dose selection strategy resulted in too frequent selection of the two doses at the boundary (10 mg BID and 40 mg BID) even when the middle dose (20 mg BID) was the true best. In order to stabilize the dose selection procedure, it was therefore necessary to disallow selection of the 10 mg BID dose.

Simulation-based estimates of operating characteristics for the adaptive design are provided in Table 3.

Scenario			Simulation-based Estimate of Probability (%)		
	Best Dose	N Sim	Choose Best	Best is Significant	Either is Significant
2	40 mg BID	104	76	46	62
3	20 mg BID	1000	42	34	84
4	20 mg BID	1000	54	52	97
5	20 mg BID	400	53	47	87

Table 3: Estimated probabilities for the described adaptive design. The column "N Sim" indicates the number of clinical trials that were simulated in order to obtain the probability estimates . The column "Choose Best" provides the probability of selecting the "best" dose at the interim, while the column "Best is Significant" provides the probability that the "best" dose is both selected at the interim *and* is subsequently significant in the primary analysis. The last column, "Either is Significant", provides the probability of rejecting the null hypothesis for either the 20 mg BID dose or the 40 mg BID dose, which may be thought of as the "power" of the design (since, given, the modified version of the dose selection algorithm, selection of the 10 mg BID dose is not possible).

In general, the fixed and adaptive designs were essentially undifferentiated with respect to overall power, however the fixed design had a clear advantage with respect to the probability of rejecting the null hypothesis for the best dose. Conditional on selecting the "best dose", the probability of rejecting the null hypothesis for that dose was generally sufficiently high (between approximately 80% and 95% under scenarios of primary interest). However, the probability of selecting the best dose at the interim was unacceptably low (between approximately 40% and 50% under the scenarios of primary interest).

Implications for a Separate Phase 2 Dose-finding Trial

Based on the number of patient-days associated with stage 1 of the "best case" adaptive design, the information available at the interim analysis would be comparable to the information available from a 6-month phase 2 study with 532 patients or a 12-month phase 2 study with 266 patients. Thus, even if a sizable phase 2 study were used to prune doses for phase 3, the risk of poor dose selection would be unnecessarily high: the diversified approach of taking all three candidate doses directly into phase 3 results in a greater probability of rejecting the null hypothesis for the "best dose."

Conclusion

In general, there are two primary benefit/risk considerations governing the relative merits of a fixed design versus a design with adaptive dose selection:

Benefit: The selected dose and placebo are studied with more patients than they would be under a single fixed design. Therefore, conditional on selecting the "best" dose, there is a higher probability of obtaining a significant result for this dose (relative to a fixed design, under any alternative hypothesis where this dose has an effect).

Risk: The "best" dose may not be selected, and, conditional on not selecting the "best" dose, there is of course zero probability of obtaining a significant result for that dose. By contrast, a single fixed design runs to completion with all doses, including the "best" dose (which then may or may not be significant in the final analysis).

Thus, the success of an adaptive design depends crucially on the ability to correctly select a dose on the basis of interim data. This in turn depends on the rate of endpoint events (PE rate in our case) relative to the rate of enrollment. As it turns out in this case, the event rate is too low relative to the enrollment rate to enable adequate dose selection that can be implemented in time to make a difference. Were there a qualified biomarker that could be used for earlier and appropriate selection of the best dose in CF patients, an adaptive design might be more beneficial.
For the purpose of selecting an appropriate dose of NAL and demonstrating efficacy at this dose, a single fixed design including placebo and three active NAL dose groups (10, 20 and 40 mg QD) is more efficient than either a two-stage adaptive design or a sequential development strategy involving a separate phase 2 study for dose selection.

References

[1] Fuchs, H.J., Borowitz, D.S., Christiansen, D.H., Morris, E.M., Nash, M.L., Ramsey, B.W., Rosenstein, B.J., Smith, A.L. and Wohl, M.E. Effect of aerosolized recombinant human dnase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. the pulmozyme study group. *N Engl J Med* **331** (1994):637–42.

[2] Goss, C.H. and Burns, J.L. Exacerbations in cystic fibrosis. 1: Epidemiology and pathogenesis. *Thorax* **62** (2007):360–7.

[3] Saiman, L., Marshall, B.C., Mayer-Hamblett, N., Burns, J.L., Quittner, A.L., Cibene, D.A., Coquillette, S., Fieberg, A.Y., Accurso, F.J. and Campbell, P.W.r. Azithromycin in patients with cystic fibrosis chronically infected with pseudomonas aeruginosa: a randomized controlled trial. *JAMA* **290** (2003 Oct 1):1749–1756.

[4] Clement, A., Tamalet, A., Leroux, E., Ravilly, S., Fauroux, B. and Jais, J.P. Long term effects of azithromycin in patients with cystic fibrosis: A double blind, placebo controlled trial. *Thorax* **61** (2006 Oct):895–902.

[5] Therneau, T.M. and Grambsch, P.M. Modeling survival data: Extending the cox model (2000).

[6] Hsu, J.C. and Berger, R.L. Stepwise confidence intervals without multiplicity adjustment for dose-response and toxicity studies. *Journal of the American Statistical Association* (1999).

[7] Posch, M., Koenig, F., Branson, M., Brannath, W., Dunger-Baldauf, C. and Bauer, P. Testing and estimation in flexible group sequential designs with adaptive treatment selection. *Stat Med* **24** (2005):3697–714.

[8] Team, R.D.C. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2008). ISBN: 3-900051-07-0. URL `http://www.R-project.org`

[9] Lunn, D., Thomas, A., Best, N. and Spiegelhalter, D. Winbugs – a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* **10** (2000):325–337.

[10] Genz, A. Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics* (1992):141–150.

[11] Genz, A. Comparison of methods for the computation of multivariate normal probabilities. *Computing Sciences and Statistics* (1993):400–405.

 Copies of this poster are available at www.metrumrg.com/publications

American Conference on Pharmacometrics, Mashantucket, CT, October 4–7, 2009