CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

125291

STATISTICAL REVIEW(S)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

Statistical Review and Evaluation

CLINICAL STUDIES

NDA/Serial Number:	BLA STN125291/0
Drug Name:	Lumizyme (recombinant human acid alpha-glucosidase (rhGAA), 2000L)
Indication(s):	Treatment of patients with late-onset Pompe disease
Applicant:	Genzyme
Date(s):	Stamp date: 5/30/08
	Advisory Committee Meeting: 10/21/08
	PDUFA: 2/28/09
Review Priority:	Six months, with a three month extension
Biometrics Division:	Division of Biometrics 3
Statistical Reviewer:	Lisa A. Kammerman, Ph.D. LAK 2/9/09
Concurring Reviewers:	Mike Welch, Ph.D. MW 2/9/09
Medical Division:	Division of Gastroenterology Products, HFD-180
Clinical Team:	Lynne Yao, MD; Joanna Ku, MD
Project Manager:	Wes Ishihara, LT USPHS

Keywords: clinical studies, BLA review, adaptive design, minimization

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

From a statistical perspective, this application is unusual in several ways. While the study was ongoing, the study was changed from a fixed design to a design with an adaptive strategy. This necessitated a change to the endpoint, which became the linear rate of change over the duration of the study. For each subject, this rate of change was estimated by a slope. Thus, instead of comparing the six minute walk test (6MWT) at 52 weeks, the primary analysis became a comparison of the slopes.

To evaluate the slopes, the primary analysis method was a linear mixed effects model, where the patient-level intercepts and slopes were random effects and other effects (e.g., treatment, baseline strata) were fixed effects. The statistical analysis plan specified a model-dependent estimator of the variance-covariance matrix.

The statistical analysis plan also specified analysis of covariance (ANCOVA) models and re-randomization tests as supportive analyses. The re-randomization tests were included to address the minimization algorithm that was used to allocate subjects in a 2:1 ratio to either 2000L or placebo.

When the data were analyzed, diagnostic tests determined the 6MWT departed from the assumption of linearity and the assumption of normality. Moreover, the applicant asserts the observed non-linearity compromises the estimate of the model-dependent variance-covariance matrix.

As a result, the applicant proposed the use of a "sandwich" estimator of the variance-covariance matrix. While I agree the "sandwich" estimator is more appropriate, its use leaves some unanswered questions. If the model assumptions are correct, the model-based estimator of the variance-covariance matrix is more efficient than a "sandwich" estimator. Had the results been statistically significant, I wonder whether the applicant would have explored the use of the "sandwich" estimator and argued for its use if the results had been statistically non-significant.

The applicant also included results from analyses that were not pre-specified: generalized estimating equation (GEE) models and non-parametric assessments of the data. These tests gave statistically significant results for the 6MWT. However, I question the use of models that were not prespecified.

Because of the violations of the assumptions underlying the linear mixed effects model and the changes to the model after the data were unblinded, I believe the results of the ANCOVA should be emphasized. The ANCOVA is consistent with the clinical question of interest, which is whether the change from baseline to the last observation differs between 2000L-treated subjects and placebo-treated subjects.

Further complicating the interpretation of the study results is the scheme used to allocate subjects to 2000L and placebo. Instead of a blocked randomization, which is typically used, the study used a minimization algorithm in order to maintain a 2:1 (2000L:placebo) ratio within study sites and within strata defined by baseline values for the 6MWT and forced vital capacity (FVC).

Re-randomization tests are the appropriate approach for assessing statistical significance when a minimization algorithm used. Usually, the result from a re-randomization test is consistent with the result from the classical test. However, that is not the case in this submission. For the ANCOVA of the 6MWT, the p-value changes from 0.035 to 0.06; and for the LME model with the sandwich estimator, the p-value changes from 0.046 to 0.150. I discount the applicant's argument that subjects can be assumed to have arrived in a random order, an assumption which leads to statistically significant results for the 6MWT. Unfortunately, the allocation probabilities were not retained by the applicant's contract research organization and I was unable to explore whether the probabilities of assignment were somehow related to the differences between the standard test results and the re-randomization test results. The results for FVC are statistically significant regardless of the test used.

Although the p-value of 0.06 for the re-randomization test, which I believe is the appropriate test, corresponding to the ANCOVA for the 6MWT is not statistically significant at the traditional alpha level of 0.05, I believe the orphan status of the indication needs to be entertained when deciding on the efficacy of this product for the treatment of adult patients with Pompe disease.

1.2 Brief Overview of Clinical Studies

In 2006, Myozyme was approved for use in patients with Pompe disease (GAA deficiency). Comparisons of data from a single study of patients with infantile-onset disease with data from a historical control group and with the natural history of the disease were the basis for the approval. The approved label indicates use of Myozyme in patients with other forms of Pompe disease has not been adequately studied to assure safety and efficacy.

The studies submitted with the initial BLA used product lots manufactured from a 160 L process, which is the process currently approved. The applicant's request for approval of a 2000 L manufacturing process was denied due to concerns of lack of comparability with the 160 L product.

With this BLA, the applicant is requesting approval of recombinant human acid alpha-glucosidase (rhGAA) produced from the 2000 L manufacturing process. Because FDA product reviewers determined the product produced by the 2000 L process differs from the product produced by the 160 L process, the product from the 2000 L process has its own trade name – Lumizyme.

The evidence in support of the indication comes from a single study (AGLU02704):

> "A randomized, double-blind, multicenter, multinational, placebo-controlled study of the safety, efficacy, and pharmacokinetics of Myozyme, recombinant human acid alpha-glucosidase (rhGAA) treatment in patients with late-onset Pompe disease."

The overall study objective was to evaluate the safety, efficacy and pharmacokinetics of 2000L product in patients with late-onset Pompe disease as compared to placebo. The study was conducted at 8 study sites: 5 in the United States and 3 in Europe. After 6 months of treatment, subjects could transfer to one of 22 local investigational sites that were closer to the subjects' homes. Subjects returned to the primary investigational site every 3 months for assessment.

Originally, the primary endpoint was the six-minute walk test (6MWT) evaluated at 52 weeks. The endpoint was changed to a linear slope of 6MWT versus time, over 78 weeks. The duration of the study was increased to 78 weeks as a result of an adaptive design strategy and interim analysis. The endpoint was changed to a linear slope to accommodate the change from a study of fixed duration to a study with an adaptive strategy.

1.3 Statistical Issues and Findings

Statistical and design issues important to the understanding and interpretation of the results from Study AGLU02704 (LOTS) are the following:

- Change from a fixed design to the use of an adaptive strategy
- Change in the endpoint from 6MWT (FVC) at 52 weeks to average change in 6MWT (FVC) as measured by a linear slope
- Change in statistical analysis methods after data were shown to violate assumptions of prespecified model
- Use of a minimization algorithm to allocate subjects
- Results of re-randomization analyses

2. INTRODUCTION

2.1 Overview

In 2006, Myozyme was approved for use in patients with Pompe disease (GAA deficiency). Comparisons of data from a single study of patients with infantile-onset disease with data from a historical control group and with the natural history of the disease were the basis for the approval. The approved label indicates use of Myozyme in patients with other forms of Pompe disease has not been adequately studied to assure safety and efficacy.

The studies submitted with the initial BLA used product lots manufactured from a 160 L process, which is the process currently approved. The applicant's request for approval of a 2000 L manufacturing process was denied due to concerns of lack of comparability with the 160 L product. Since approval of Myozyme, the applicant has experienced shortages in Myozyme production.

With this BLA, the applicant is requesting approval of the 2000 L manufacturing process for the treatment of patients with late-onset Pompe disease.

The BLA contains clinical data from Study AGLU02704, which was initially planned to be 52-weeks, with no plans for an interim analysis. The six minute walk test (6MWT; distance in meters) and forced vital capacity (FVC) upright (% predicted) at 52-weeks were the co-primary endpoints. The 6MWT was to be examined first. If the treatment effect was statistically significant at 0.05, then FVC upright would be evaluated.

The treatment effect at 52 weeks for the 6MWT was to be assessed by a repeated measures linear model with covariates. The model was to have the following characteristics:

- Site is an explanatory variable
- Distance walked at baseline is a continuous covariate
- Time of assessment is an explanatory variable (categorical) in the model
- The model includes a term for the interaction between time and treatment, i.e., to assess evidence about whether the time effect is different for each treatment
- Response covariance is modeled by a compound symmetry structure
- The model contains a parameter for the treatment effect between Myozyme and placebo at Study Week 52, adjusted for baseline.

The parameter for the treatment effect at Study Week 52 would be tested to assess the significance of the treatment effect.

The model parameters would be estimated by a restricted maximum likelihood method. Assumptions regarding the normality of errors and the form of the covariance matrix would be assessed. If the data indicate that model assumption of normality of errors is likely to be violated, then non-parametric methods may be used to carry out the significance testing.

Assessment of FVC upright would be carried out in a similar manner.

In May 2006, the protocol was amended to an adaptive clinical trial design, and the primary endpoint was changed to a linear rate of change in distance walked estimated from a longitudinal model where response is modeled as a linear function of time of assessment. The rationale for these changes was to determine through an interim analysis the optimal duration of the study and to "compare the two treatments over the course of the study, rather than focusing on comparisons at 52 weeks."

The interim analysis was to be done when the last patient enrolled in the study completed Week 38 or when 75% of the total statistical information targeted at the design stage had

arrived. The plan was to terminate the study if efficacy had been demonstrated by crossing a prospectively defined stopping boundary or to continue the study for an additional 3, 6 or 9 months on the basis of calculations of conditional power.

Because the medical division believed it was important to obtain 52 weeks of data on all subjects, the protocol was amended again (August 2006) to eliminate the possibility of terminating the study at the interim analysis. The study could continue to 52 weeks, as originally planned, or extended an additional 3 or 6 months. The maximum length of time any patient could participate in the study was 78 weeks.

2.2 Data Sources

My review is based on the following documents submitted to the BLA by the applicant.

Clinical Study Report (BLA 125291/0000)
\\cbsap58\M\eCTD_Submissions\STN125291\0000\m5\53-clin-stud-rep\535-rep-effic-safety-stud\pompe\5351-stud-rep-contr\aglu02704\aglu02704-16-1-1.pdf

Clinical Study Report Errata 2 (BLA 125291/0026)
\\cbsap58\M\eCTD_Submissions\STN125291\0026\m5\53-clin-stud-rep\535-rep-effic-safety-stud\pompe\5351-stud-rep-contr\aglu02704\aglu02704-csr-errata2.pdf

My statistical review of original submission (BLA 125141/0000) -- Redacted version:
http://www.fda.gov/cder/foi/nda/2006/125141s0000_Myozyme_StatR.pdf

Datasets
\\cbsap58\M\eCTD_Submissions\STN125291\0000\m5\datasets\aglu02704

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

The evidence in support of the indication comes from a single study (AGLU02704):

"A randomized, double-blind, multicenter, multinational, placebo-controlled study of the safety, efficacy, and pharmacokinetics of Myozyme, recombinant human acid alpha-glucosidase (rhGAA) treatment in patients with late-onset Pompe disease."

The overall study objective was to evaluate the safety, efficacy and pharmacokinetics of 2000L product in patients with late-onset Pompe disease as compared to placebo. The study was conducted at 8 study sites: 5 in the United States and 3 in Europe. After 6 months of treatment, subjects could transfer to one of 22 local investigational sites that were closer to the subjects' homes. Subjects returned to the primary investigational site every 3 months for assessment.

3.1.1 Study Design and Endpoints

Subjects at least 8 years of age, who could ambulate 40 meters in a six minute walk test on 2 consecutive dates and who had a FVC between 30% and 80% predicted in the upright position were eligible for enrollment. Subjects requiring invasive ventilatory support or use of noninvasive ventilatory support while awake and in an upright position were excluded from the study.

Subjects received either 2000L product or placebo as an IV infusion at a dose of 20 mg/kg every other week. The duration of treatment was to be 52-weeks, but was extended to 78-weeks based on a recommendation from the independent statistical center who conducted the interim analysis.

The sample size calculations for the original fixed design were based on detecting, with 80% power, a treatment difference of 0.75 standard deviations using a two sample t-test with a significance level of 0.05. This translated into detecting a treatment effect of 53.5 meters in the 6MWT and an effect of 7.5% predicted FVC upright. With a 2:1 treatment allocation (2000L:placebo), a sample size of 63 (2000L, n=42; placebo n=21) would satisfy these criteria. The plan was to enroll at least 72 subjects (2000L, n=48; placebo n=24) to account for a 10% to 15% dropout rate.

A minimization algorithm was used to maintain a 2:1 (2000L:placebo) treatment balance within study site (n=8), within 6MWT baseline strata ($\leq$ 300 meters, > 300 meters) and within FVC baseline strata ($\leq$55% predicted, >55% predicted).

The minimization algorithm is reproduced in Figure 1.

Figure 1. Minimization algorithm used to allocate subjects in a 2:1 ratio (2000L : Placebo)

Minimization Parameters

1. Number of patients: 90
2. Number of treatments: 2
3. Allocation ration: 2:1 (A:B)
4. Number of sites: 8
5. Number of strata: 4
6. Threshold percentage: 40%
7. Assignment probability: 90%
8. Site weight: 1
9. Strata weight: 3

Definitions

1. $nsite_i(A)$ = number of subjects assigned to treatment A currently in Site i
2. $nsite_i(B)$ = number of subjects assigned to treatment B currently in Site i
3. $nstrata_i(A)$ = number of subjects assigned to treatment A currently in Strata i
4. $nstrata_i(B)$ = number of subjects assigned to treatment B currently in Strata i

Algorithm Steps

1. Threshold check: if $\{[\max(nsite_i(A), 2*nsite_i(B)] - [\min(nsite_i(A), 2*nsite_i(B)]\}/(nsite_i(A) + 2*nsite_i(B)) > 0.40$ then patient is automatically assigned to the treatment out of balance. This essentially overrides the minimization allocation. Note that the multiplier of 2 is for the allocation ratio 2:1.

2. Compute site minimization score for treatment A: $site(A) = 1*\{[\max(nsite_i(A)+1, 2*nsite_i(B)] - [\min(nsite_i(A)+1, 2*nsite_i(B)]\}/ (nsite_i(A) + 2*nsite_i(B) + 1)$.

3. Compute site minimization score for treatment B: $site(B) = 1*\{[\max(nsite_i(A), 2*(nsite_i(B)+1)] - [\min(nsite_i(A), 2*(nsite_i(B)+1)]\}/ (nsite_i(A) + 2*nsite_i(B) + 1)$.

4. Compute stratum minimization score for treatment A: $strata(A) = 1*\{[\max(nstrata_i(A)+1, 2*nstrata_i(B)] - [\min(nstrata_i(A)+1, 2*nstrata_i(B)]\}/ (nstrata_i(A) + 2*nstrata_i(B) + 1)$.

5. Compute stratum minimization score for treatment B: $strata(B) = 1*\{[\max(nstrata_i(A), 2*(nstrata_i(B)+1)] - [\min(nstrata_i(A), 2*(nstrata_i(B)+1)]\}/ (nstrata_i(A) + 2*nstrata_i(B) + 1)$.
6. Compute total minimization score for treatment A: $tot(A) = site(A) + strata(A)$.

7. Compute total minimization score for treatment B: $tot(B) = site(B) + strata(B)$.

8. Assign treatment: if $tot(A)<tot(B)$ then assign treatment A with 90% probability. If $tot(B)<tot(A)$ then assign treatment B with 90% probability. If $tot(A)=tot(B)$ then assign either treatment with 50% probability.

Source: Section 16.1.7, Final Study Report

The primary assessments in the study were the 6-minute walk test (6MWT) and FVC (% predicted).

While the study was ongoing, the design was changed to an adaptive study. The rationale for these changes was to determine the optimal duration of the study and to "compare the two treatments over the course of the study, rather than focusing on comparisons at 52 weeks." To accommodate the adaptive design changes to the study, the primary endpoints became the linear rate of change in distance walked in a 6-minute walk test and the linear rate of change in FVC (% predicted). These rates of change were to be estimated by a linear model.

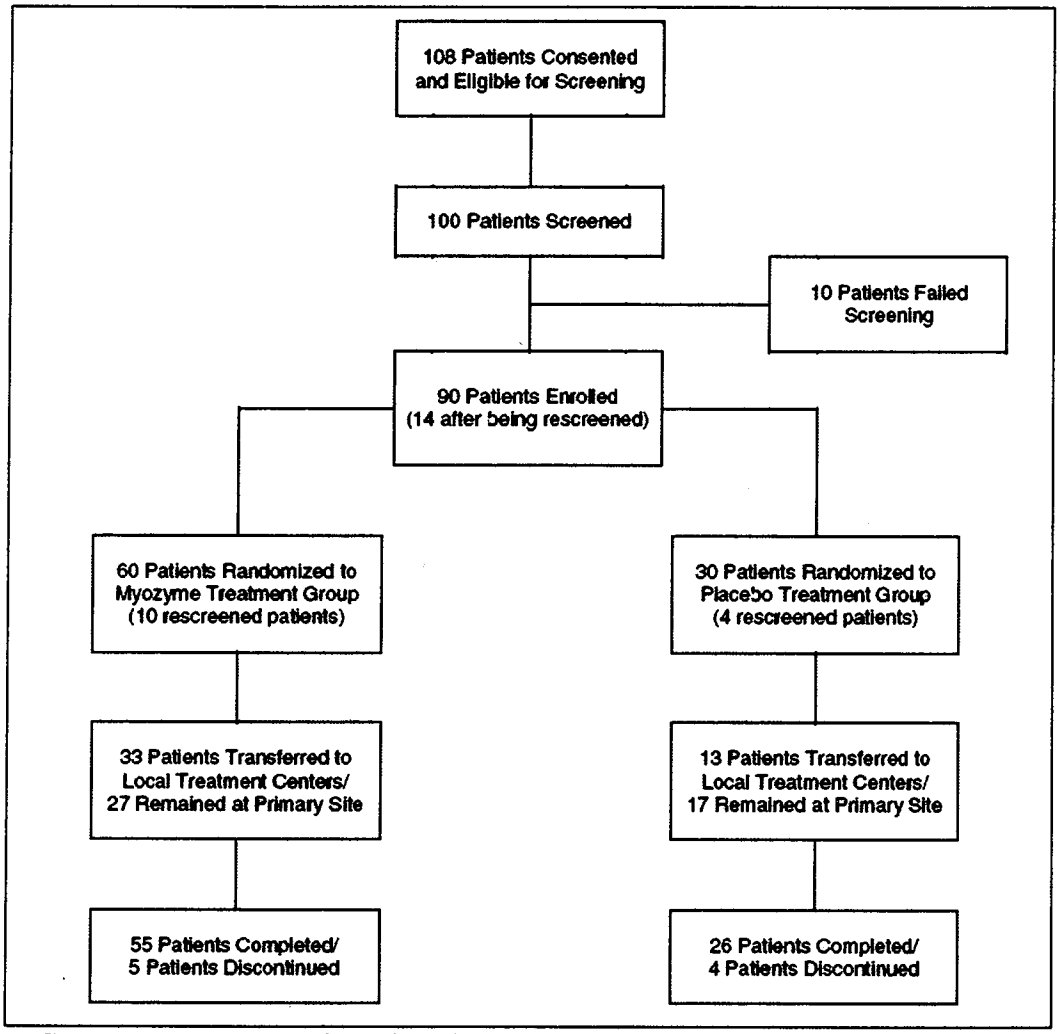
The interim analysis for the adaptive design was to be done when the last patient enrolled and continuing in the study completed Week 38 or when 75% of the total statistical information targeted at the design stage had been achieved. The plan was to continue the study, as planned, to 52 weeks or to continue the study for an additional 3 or 6 months on the basis of calculations of conditional power.

An external independent statistical center (ISC) performed an interim analysis of the 6MWT data when the last patient enrolled and continuing in the study had completed Week 38 of the study. Based on the results of the analysis, the ISC recommended extending the study from 52 weeks to 78 weeks.

3.1.2 Subject disposition, demographic and baseline characteristics

The study enrolled ninety subjects (2000L, n=60; Placebo, n=30). The disposition of subjects is shown in Figure 2. Nine subjects (2000L – 5; Placebo – 4) discontinued the study early.

Figure 2. Disposition of Subjects



Source: Figure 10-1, Clinical Study Report

Although the distribution of age for the 2000L-treated subjects is similar to that for placebo-treated subjects, the youngest age of subjects enrolled in the 2000L treatment arm was 16 years, compared with 10 years for the placebo treatment arm; see Table 1. These minimum enrollment ages are important because the applicant is seeking approval for all subjects 8 years of age or older. The oldest subject in either group was around 70 years of age.

Table 1 Distribution of age, by treatment group.

a. Quantiles

Percentile		2000L	Placebo
100.0%	maximum	70.000	68.417
99.5%		70.000	68.417
97.5%		69.519	68.417
90.0%		61.317	56.100
75.0%	quartile	53.354	50.917
50.0%	median	44.958	43.083
25.0%	quartile	37.188	38.125
10.0%		30.025	28.083
2.5%		16.748	10.083
0.5%		15.917	10.083
0.0%	minimum	15.917	10.083

b. Stem and leaf diagram of age

2000L

Stem	Leaf	Count
7	0	1
6	699	3
6	024	3
5	555679	6
5	000112222334	12
4	558889	6
4	0000223333444	13
3	78	2
3	001223334	9
2	68	2
2	1	1
1	68	2

1|6 represents 16

Placebo

Stem	Leaf	Count
6	8	1
6		
5	69	2
5	11124	5
4	555666	6
4	0012333	7
3	55889	5
3	3	1
2	8	1
2		
1	5	1
1	0	1

1|0 represents 10

The distribution of gender also differed numerically. Among subjects assigned to 2000L, 57% were male compared with 37% of subjects assigned to placebo. Over 90% of subjects were Caucasian, and the average duration of disease was about 9 years.

3.1.3 Statistical methodologies

3.1.3.1 Original Statistical Analysis Plan, dated 9/27/05

The original statistical analysis plan specified two co-primary variables:

- Number of meters walked in 6 minutes at Week 52, adjusted for baseline number of meters walked
- FVC upright (% predicted) at Week 52, adjusted for baseline FVC upright (% predicted)

The treatment effect at 52 weeks for the 6MWT was to be assessed by a repeated measures linear model with covariates. The model was to have the following characteristics:

- Site is an explanatory variable
- Distance walked at baseline is a continuous covariate
- Time of assessment is an explanatory variable (categorical) in the model
- The model includes a term for the interaction between time and treatment, i.e., to assess evidence about whether the time effect is different for each treatment
- Response covariance is modeled by a compound symmetry structure
- The model contains a parameter for the treatment effect between Myozyme and placebo at Study Week 52, adjusted for baseline.

The parameter for the treatment effect at Study Week 52 would be tested to assess the significance of the treatment effect.

The model parameters would be estimated by a restricted maximum likelihood method (REML). Assumptions regarding the normality of errors and the form of the covariance matrix would be assessed. If the data indicated that model assumption of normality of errors were likely to be violated, then non-parametric methods may be used to carry out the significance testing.

Assessment of FVC upright would be carried out in a similar manner.

3.1.3.2 Statistical Analysis Plan Amendment, dated 9/29/06

In May 2006, the protocol was amended to an adaptive clinical trial design, and the primary endpoint was changed to a linear rate of change in distance walked estimated from a longitudinal model where response was to be modeled as a linear function of time of assessment. The rationale for these changes was to determine through interim analysis the optimal duration of the study and to "compare the two treatments over the course of the study, rather than focusing on comparisons at 52 weeks."

The statistical analysis plan contained methods and plans for the implementation of the adaptive design.

The adaptive design required changes to the primary endpoints. The co-primary efficacy variables were changed to slopes (average monthly increase):

- Average monthly increase in 6MWT.
- Average monthly increase in FVC upright (% predicted).

The interim analysis for the adaptive design was to be done when the last patient enrolled in the study completed Week 38 or when 75% of the total statistical information targeted at the design stage had been achieved. The study could continue to 52 weeks, as originally planned, or extended an additional 3 or 6 months. The maximum length of time any patient could participate in the study was 78 weeks.

The changes in the average monthly increases in 6MWT (and FVC upright) among subjects assigned to 2000 L product compared with changes in average monthly increases in 6MWT (and FVC upright) among subjects assigned to placebo were to be analyzed using a linear mixed effects (LME) model:

- Independent variables are site, treatment, time, and treatment-by-time interaction
- Outcome vector contains the observed measurements of 6MWT (or FVC upright) collected at baseline and at study visits
- Model to be fit using restricted maximum likelihood estimation
- Model to use unstructured variance-covariance matrix for the random effects (i.e., intercept and slope)

The model would be used to estimate the rate of change for each subject. Model assumptions would be assessed.

Differences between the average monthly increase between subjects assigned to 2000L and to placebo would be tested by a Wald statistic. The Haybittle-Peto alpha-spending function would adjust for the interim analysis, resulting in a final nominal alpha of 0.04999 for declaring statistical significance.

Supportive analyses were to include a re-randomization analysis, which was to consist of running the minimization algorithm used for the treatment assignments 10,000 times. The LME model would be applied and the p-values for the test of the treatment effect recorded. The resulting empirical distribution for the p-values would be compared with the observed p-value.

An analysis of covariance model (ANCOVA) was an additional supportive analysis. The ANCOVA would model the change from baseline and, separately, relative change from baseline to the last observed assessment. The model would include the baseline strata used in the minimization algorithm, the baseline observation, and treatment indicator.

3.1.3.3 Changes to the statistical methods after data were unblinded

After the pre-specified LME model was fit to the data, diagnostics were done to determine the appropriateness of the LME model. These diagnostics were pre-specified in the SAP. Examination of the 6MWT data indicates significant departures from linearity and violations of the assumption of normality. Moreover, the applicant asserts the observed non-linearity also compromises the estimate of the model-dependent variance-covariance matrix.

Because of these findings, the applicant chose to use a sandwich estimator of the variance-covariance matrix. In addition, the applicant decided to test the difference in the monthly rate of change using GEE models and the repeated measures Wei-Lachin test, a nonparametric procedure. The applicant claims that GEE models relax the assumptions of linearity and a correctly specified variance-covariance matrix. Their GEE model uses a

compound symmetric working correlation matrix. ANCOVA models were used in place of LME models to estimate treatment effects in subgroups.

The FVC efficacy analysis did not violate any of the assumptions of the LME model. However, GEE, Wei-Lachin and ANCOVA models were used to analyze FVC to maintain consistency with the analyses of the 6MWT.

3.1.4 Results and conclusions

3.1.4.1 Interim analysis

An external independent statistical center (ISC) performed an interim analysis of the 6MWT data when the last patient had completed Week 38 of the study. The analysis used the linear mixed effects model specified in the statistical analysis plan. Based on the amount of information accrued, the ISC recommend extending the study from 52 weeks to 78 weeks in order to increase the number of measurements per subject, allowing for sufficient information to attain 90% power to detect a difference between 2000L and placebo of 3.75 meters/month.

A representative of the ISC presented the results of the interim analysis to the Data and Safety Monitoring Board. After deliberating on this information the DSMB recommended to the applicant that the study be extended for an additional 26 weeks.

3.1.4.2 Results for 6MWT

Figure 3 shows the means and standard errors of the total distance walked in the 6MWT at each study visit.

The results from the LME model with a sandwich estimator (i.e. robust variance estimation) and from the ANCOVA indicate that the improvement among subjects treated with the 2000L product is significantly better than among subjects treated with placebo (Table 2).

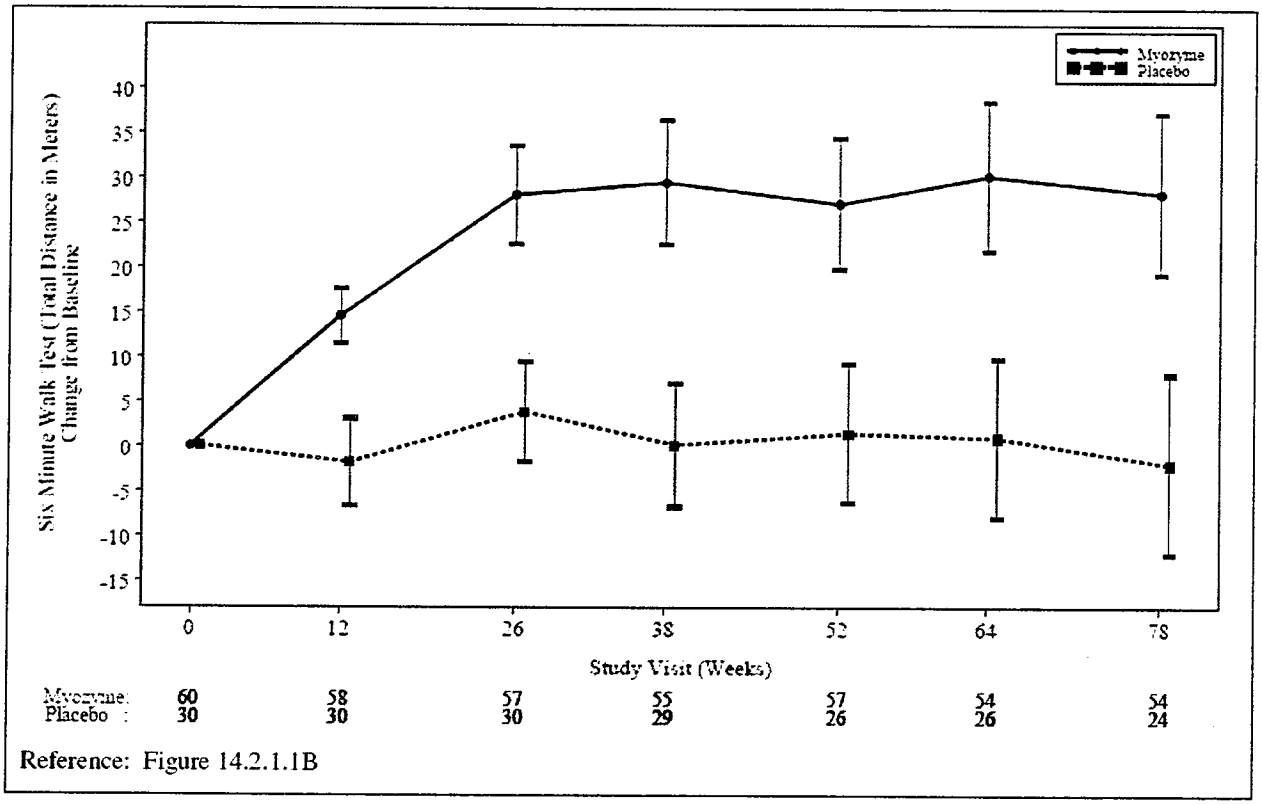
The p-values from the re-randomization analyses are statistically non-significant at a nominal level of 0.05: LME model with robust variance estimation (p=0.15) and ANCOVA (p=0.06); see Table 3 and Table 4.

The applicant asserts the re-randomization test is difficult to interpret because the discrete nature of the allocation algorithm restricts the space of possible treatment allocations. Moreover, they assert the re-randomization test is an inefficient test of the treatment effect because the reference distribution of the test statistic from the 10,000 re-randomizations is skewed and not centered at zero; see Figure 4.

Although I agree the space of possible treatment allocations is limited, primarily because of the 2:1 allocation, this limitation does not inviolate the re-randomization test. The limitations

are due to the use of the minimization algorithm. The skewness of the reference distribution is likely due to the unequal allocation. Further, the distribution shows that the t-score is not normally distributed. This feature calls into question the use of the p-values from the classical tests, which assume the t-scores are asymptotically normal. This lack of normality would argue in favor of the re-randomization test.

Figure 3. Mean (+/- one standard error of the mean) Change from baseline over time in six-minute walk test: total distance walked



Reference: Figure 14.2.1.1B

Source: Figure 11-1, Clinical Study Report

Table 2 Change in distance walked in six-minute walk test

	Myozyme N = 60	Placebo N = 30	Difference	P value
Estimates/Tests of Monthly Change in Distance Walked (Repeated Measures Analysis)				
GEE, meters/month (95% CI)	1.37 (0.42, 2.33)	-0.13 (-1.12, 0.85)	1.51 (0.12, 2.89)	0.0326
LME with model-based variance estimation, meters/month (95% CI)	1.18 (0.34, 2.03)	-0.06 (-1.26, 1.14)	1.24 (-0.21, 2.70)	0.0931
LME with robust variance estimation, meters/month (95% CI)	1.18 (0.26, 2.11)	-0.06 (-0.90, 0.78)	1.24 (0.02, 2.47)	0.0464
Wei-Lachin test	--	--	--	0.0133
Estimates/Tests of Change in Distance Walked From Baseline to Last Observation				
ANCOVA, meters (95% CI)	25.13 (10.07, 40.19)	-2.99 (-24.16, 18.18)	28.12 (2.07, 54.17)	0.0347
Wilcoxon-Mann-Whitney test				0.0283

Reference: 14.2.1.1.1, 14.2.1.2.A, 14.2.1.3.1

Source: Table 11-3, Clinical Study Report

Table 3 Re-randomization p-values

Endpoint	Model	CSR reported p-value	Correct p-value
6MWT	LME with robust variance estimation	0.1275	0.1500
	GEE	0.0895	0.1040
FVC upright	LME with model-based variance estimation	0.0038	0.0130

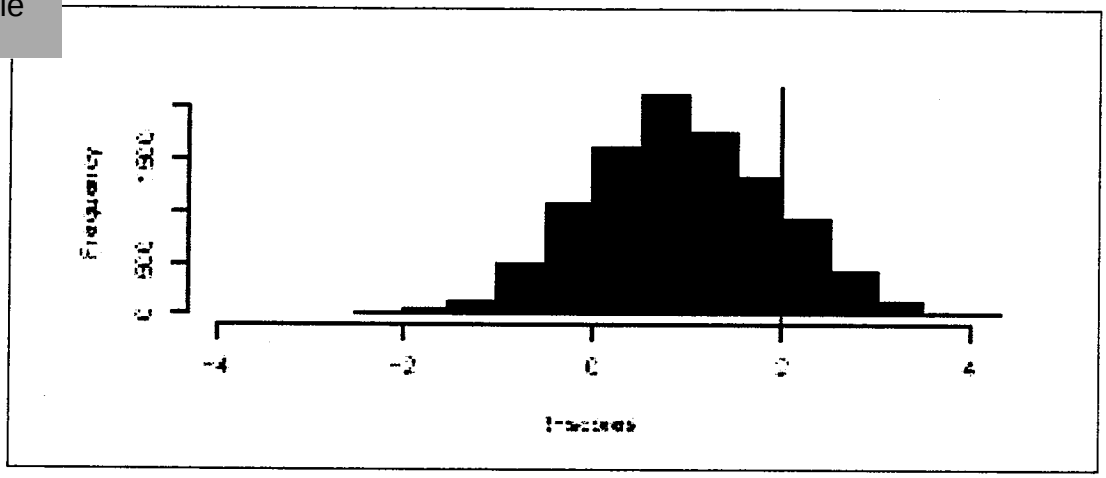
Source: Table 1, Errata 2, Clinical Study Report

Table 4 Re-randomization p-values for the 6MWT

Model	Re-randomization p-value
LME with robust variance estimation, meters/month	.15
ANCOVA	.06

Source: Table 1, Errata 2, Clinical Study Report; e-mail communication

Figure 4. Histogram of test statistics for LME with robust variance estimation



Source: Figure 1, Errata 2, Clinical Study Report

The re-randomization test results presented here assume subjects arrived in a fixed patient arrival sequence. The applicant argues that there is no reason to believe the arrival is not random. They did a re-randomization test assuming a random patient arrival sequence, resulting in a statistically significant p-value. However, I do not believe this approach is appropriate. The reasons for why patients arrive in the order that they do is unknowable. Moreover, some anecdotal reports suggest that subjects who enroll later in trials are healthier than those who enroll early trials.

Exploratory analyses suggest that patient arrival was not random. For each treatment group, I constructed a cumulative distribution of entry into the study. In a completely randomized study, we would expect the distributions to be identical. However, this is not the case for this study. Early in the study and again later in the study, subjects were enrolled into the 2000L treatment group at a faster rate than subjects who enrolled in the placebo treatment group; see Figure 5.

A second exploratory analysis shows the largest study site (#26, The Netherlands) enrolled all of its subjects in the 2nd half of the study, again suggesting the order of subjects enrolling

the study can not be ignored; Figure 6. The US study sites (#29 and 47) enrolled their subjects in the first half of the study.

Figure 5. Cumulative distribution of entry into study, by treatment group

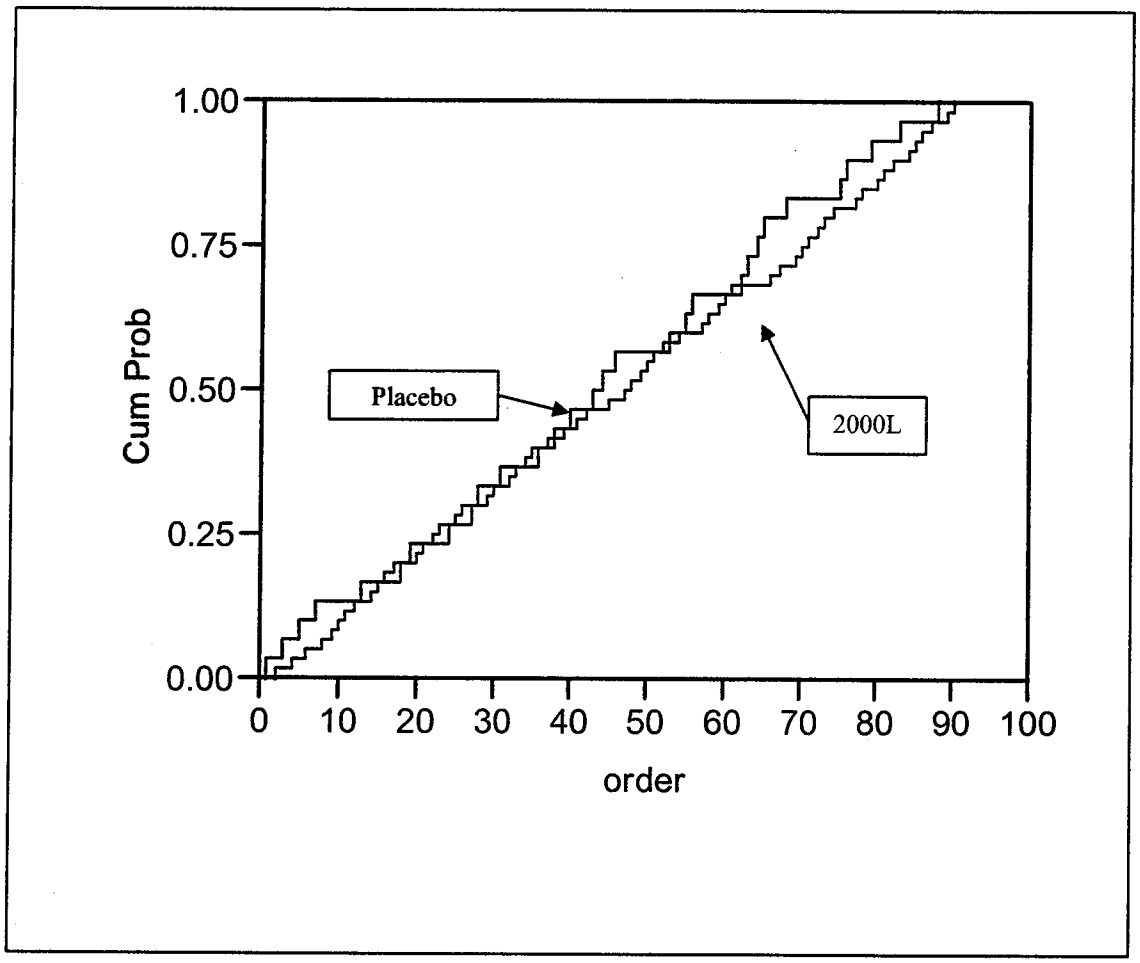
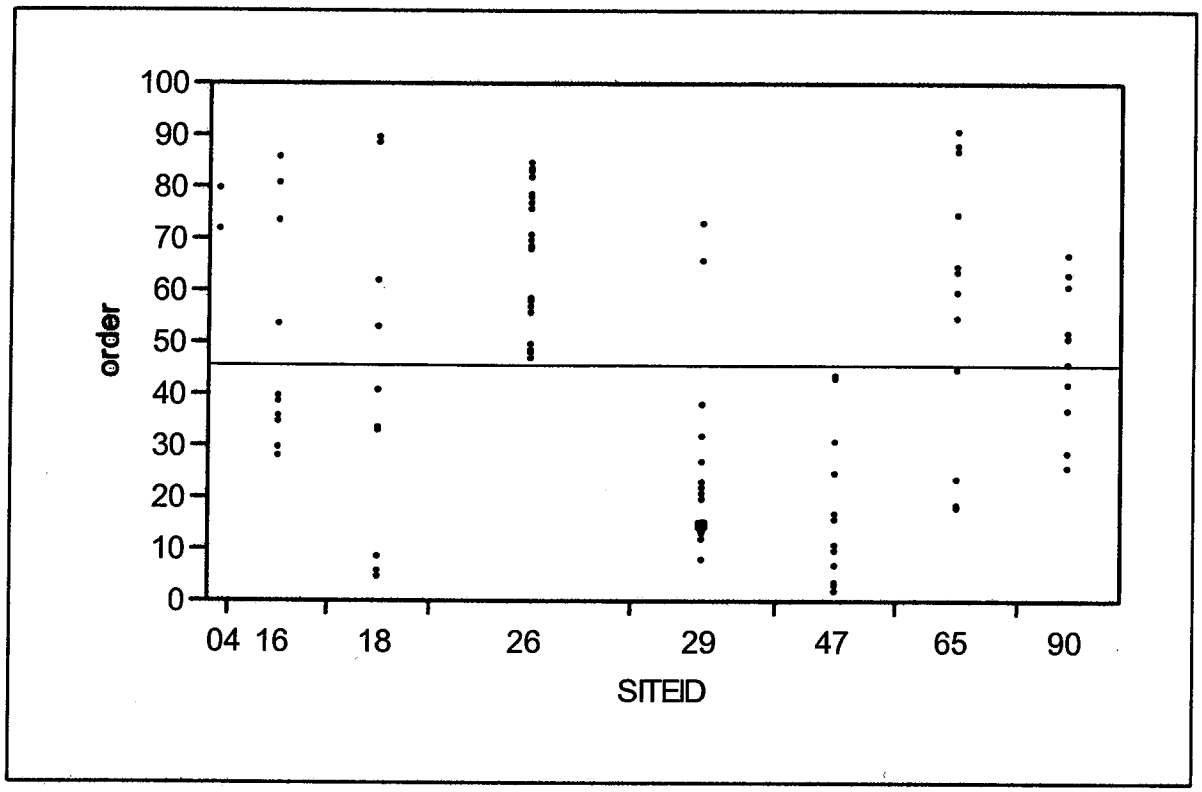


Figure 6. Order of arrival by study site.



The minimization allocation procedure may have had some unintended consequences that could have affected the 6MWT efficacy results. In particular, the distribution of gender was not consistent across the two treatment groups. As mentioned previously, among subjects assigned to 2000L, 57% were male compared with 37% of subjects assigned to placebo. Differences in the distribution of gender also persisted among the study sites. At the largest study site (#26), although half of the 20 subjects were male and half were female, there was a numerical imbalance within treatment groups: among those assigned to 2000L, 9 of 13 were male; among those assigned to placebo, 1 of 7 was male.

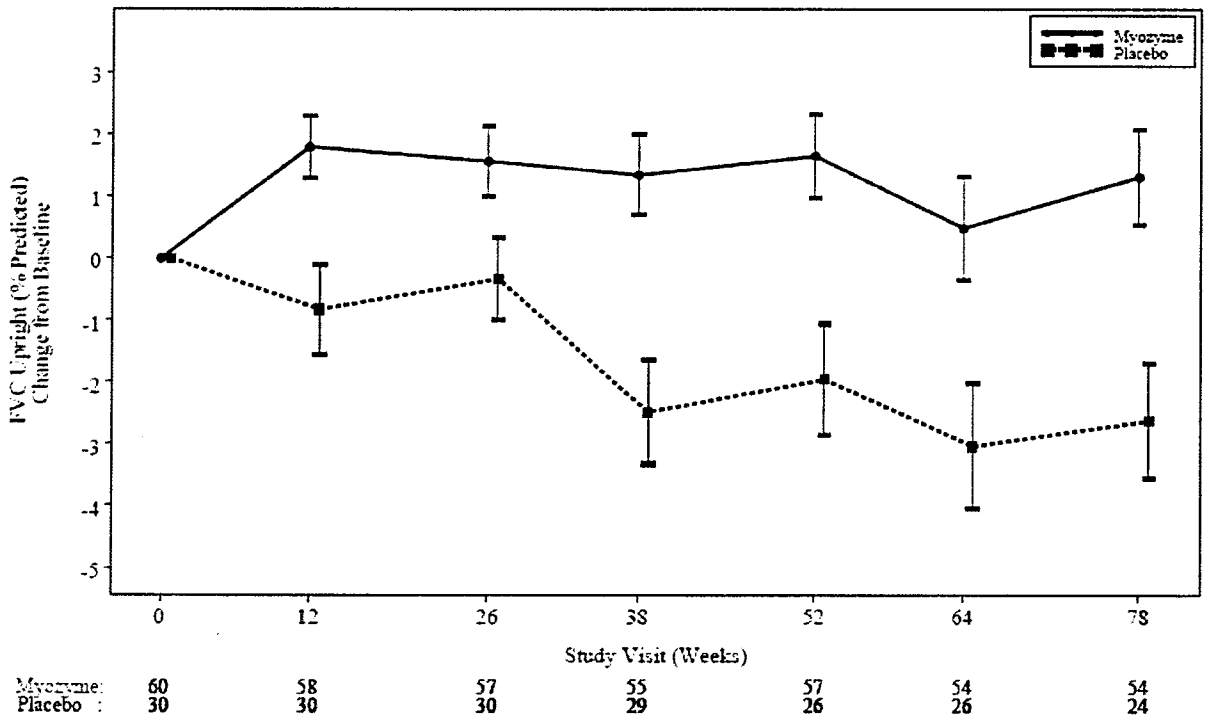
Because muscle mass differs between males and females, the potential for improvement in the 6MWT may differ as well. Whether these gender imbalances led to the observed differences in p-values between the re-randomization tests and the classical approaches is not known.

3.1.4.3 Results for FVC

According to the statistical analysis plan, FVC could be assessed if and only if the results for the 6MWT are statistically significant. I present the results here for the sake of completion.

The changes from baseline in FVC at each study visit are depicted in Figure 7. The results of the statistical analyses show that the differences between 2000L and placebo are statistically significant, regardless of the statistical test used; see Table 5 and Table 6.

Figure 7. Mean (+/- one standard error of the mean) Change from baseline over time in FVC upright (% predicted)



Reference: Figure 14.2.1.2B

Source: Figure 11-3, Clinical Study Report

Table 5 Change in FVC upright (% predicted)

	Myozyme N = 60	Placebo N = 30	Difference	P value
Estimates/Tests of Monthly Change in % Predicted FVC (Repeated Measures Analyses)				
LME, % predicted (95% CI)	0.03 (-0.05, 0.10)	-0.16 (-0.27, -0.05)	0.18 (0.05, 0.31)	0.0084
LME, with robust variance estimation % predicted (95% CI)	0.03 (-0.05, 0.10)	-0.16 (-0.25, -0.06)	0.18 (0.06, 0.30)	0.0041
GEE, % predicted (95% CI)	0.03 (-0.05, 0.11)	-0.17 (-0.26, -0.07)	0.20 (0.07, 0.32)	0.0019
Wei-Lachin test				0.0009
Estimates/Tests of Change in % Predicted FVC From Baseline to Last Observation				
ANCOVA—Mean Change, % Predicted (95% CI)	1.20 (-0.16, 2.57)	-2.20 (-4.12, -0.28)	3.40 (1.03, 5.77)	0.0055
Nonparametric Inference—Median Change, % Predicted (95% CI)	0.00 (-1.00, 3.00)	-3.00 (-5.00, 0.00)		
Wilcox-Mann-Whitney test				0.0026
ANCOVA—Mean Relative Change, % of % predicted (95% CI)	1.94 (-0.62, 4.50)	-3.79 (-7.40, -0.19)		

Reference: 14.2.1.1.1, 14.2.1.2.A, 14.2.1.18

Source: Table 11-7, Clinical Study Report

Table 6 Re-randomization p-values for FVC

Model	Re-randomization p-value
LME with robust variance estimation, meters/month	.004
ANCOVA	.004

Source: Table 1, Errata 2, Clinical Study Report; e-mail communication

3.1.4.4 Clinical meaningfulness of 6MWT and FVC results

The applicant conducted analyses to assess the "meaningfulness" of the 6MWT and FVC results. Using various thresholds to define a clinically important change – either improvement or deterioration – the proportions of subjects meeting the definitions were summarized; see Table 7.

Using methods developed by Guyatt, the treatment effect of 2000L versus placebo was estimated for each threshold level. The results were consistent across the 6MWT thresholds, and across the 10% and 5% thresholds for FVC.

Table 7 Proportions of subjects improving and declining, using a range of threshold values

Change Threshold	Patients Improving		Patients Declining		Myozyme versus Placebo Treatment Effect[1]
	Myozyme N = 60	Placebo N = 30	Myozyme N = 60	Placebo N = 30	
6MWT					
54m	23.7%	13.3%	5.1%	13.3%	16.2%
37m	28.8%	16.7%	8.5%	20.0%	19.3%
30m	30.5%	20.0%	11.9%	20.0%	14.9%
FVC					
15%	11.9%	0.0%	6.8%	6.7%	11.0%
10%	20.3%	6.7%	8.5%	26.7%	27.0%
5%	40.7%	20.0%	20.3%	46.7%	32.1%

[1] Calculated using the methods of Guyatt et al (1998).

Source: Table 11-10, Clinical Study Report

Another approach to interpreting the treatment effect is to construct cumulative distributions of change from baseline and to examine the differences between the treatment groups over the entire observed range of changes from baseline.

The cumulative distributions for the 6MWT show a consistent treatment difference between the 2000L and placebo treatment groups over the range of -100 meters to +100 meters; see Figure 8. The only subjects who increased more than 100 meters were treated with 2000L. The cumulative distributions for FVC (% predicted) shows a consistent increase of about 4 units among 2000L-treated subjects as compared with placebo-treated subjects; see Figure 8 and Figure 9.

Figure 8. 6MWT: Cumulative distribution of change from baseline, by treatment group

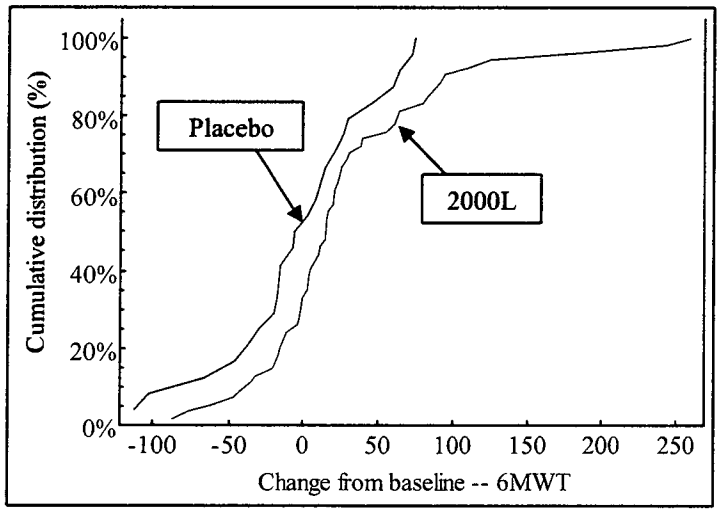
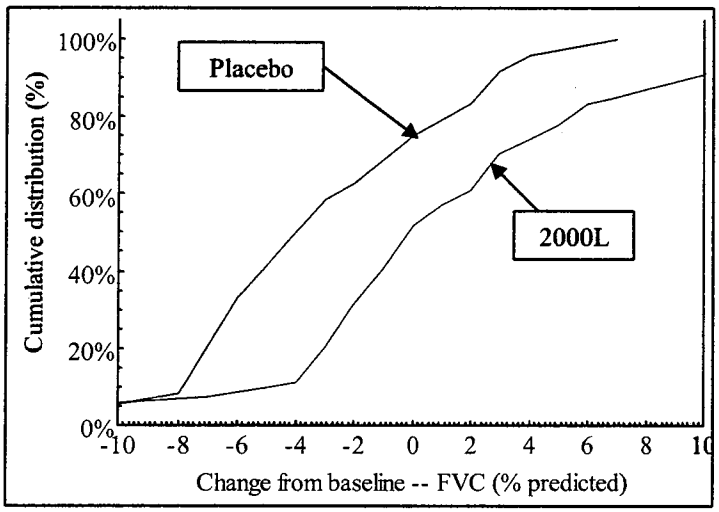


Figure 9. FVC (% predicted): Cumulative distribution of change from baseline, by treatment group



3.2 Evaluation of Safety

See medical officer's review.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

Overall, males assigned to 2000L appeared to have larger changes from baseline in the 6MWT than did females. Exploratory analyses, however, suggest the treatment effect is larger, numerically, for females than for males. These results are exploratory only and must be interpreted while keeping in mind that the minimization algorithm may have led to imbalances in gender between the treatment groups.

The small numbers of subjects in other subgroups precluded further analyses.

4.2 Other Special/Subgroup Populations

See medical officer's review.

5. SUMMARY AND CONCLUSIONS

From a statistical perspective, this application is unusual in several ways. While the study was ongoing, the study was changed from a fixed design to a design with an adaptive strategy. This necessitated a change to the endpoint, which became the linear rate of change over the duration of the study. For each subject, this rate of change was estimated by a slope. Thus, instead of comparing the 6MWT at 52 weeks, the primary analysis became a comparison of the slopes.

To evaluate the slopes, the primary analysis method was a linear mixed effects model, where the patient-level intercepts and slopes were random effects and other effects (e.g., treatment, baseline strata) were fixed effects. The statistical analysis plan specified a model-dependent estimator of the variance-covariance matrix.

The statistical analysis plan also specified ANCOVA models and re-randomization tests as supportive analyses. The re-randomization tests were included to address the minimization algorithm that was used to allocate subjects in a 2:1 ratio to either 2000L or placebo.

When the data were analyzed, diagnostic tests determined the 6MWT departed from the assumption of linearity and the assumption of normality. Moreover, the applicant asserts the observed non-linearity compromises the estimate of the model-dependent variance-covariance matrix.

As a result, the applicant proposed the use of a "sandwich" estimator of the variance-covariance matrix. While I agree the "sandwich" estimator is more appropriate, its use leaves some unanswered questions. If the model assumptions are correct, the model-based estimator of the variance-covariance matrix is more efficient than a "sandwich" estimator. Had the results been statistically significant, I wonder whether the applicant would have explored the use of the "sandwich" estimator and argued for its use if the results had been statistically non-significant.

The applicant also included results from analyses that were not pre-specified: generalized estimating equation (GEE) models and non-parametric assessments of the data. These tests gave statistically significant results for the 6MWT. However, I question the use of models that were not prespecified.

Because of the violations of the assumptions underlying the linear mixed effects model and the changes to the model after the data were unblinded, I believe the results of the ANCOVA should be emphasized. The ANCOVA is consistent with the clinical question of interest,

which is whether the change from baseline to the last observation differs between 2000L-treated subjects and placebo-treated subjects.

Further complicating the interpretation of the study results is the scheme used to allocate subjects to 2000L and placebo. Instead of a blocked randomization, which is typically used, the study used a minimization allocation in order to maintain a 2:1 (2000L:placebo) ratio within study sites and within strata defined by baseline values for the 6MWT and FVC.

Re-randomization tests are the appropriate approach for assessing statistical significance when a minimization algorithm used. Usually, the result from a re-randomization test is consistent with the result from the classical test. However, that is not the case in this submission. For the ANCOVA of the 6MWT, the p-value changes from 0.035 to 0.06; and for the LME model with the sandwich estimator, the p-value changes from 0.046 to 0.150. I discount the applicant's argument that subjects can be assumed to have arrived in a random order, an assumption which leads to statistically significant results for the 6MWT. Unfortunately, the allocation probabilities were not retained by the applicant's contract research organization and I was unable to explore whether the probabilities of assignment were somehow related to the differences between the standard test results and the re-randomization test results. The results for FVC are statistically significant regardless of the test used.

Although the p-value of 0.06 for the re-randomization test, which I believe is the appropriate test, corresponding to the ANCOVA for the 6MWT is not statistically significant at the traditional alpha level of 0.05, I believe the orphan status of the indication needs to be entertained when deciding on the efficacy of this product for the treatment of adult patients with Pompe disease.