CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER: NDA 21134

STATISTICAL REVIEW(S)

FROM

STATISTICAL REVIEW AND EVALUATION

AUG 30 1999

NDA #: 21-134

Applicant: Parke-Davis Pharmaceuticals Limited

Drug Name: Nitrostat (Nitroglycerin) 0.6 mg tablets

Indication: Treatment of acute relief of an attack or prophylaxis of angina pectoris due to coronary artery disease

Document Reviewed: Cowmeadow M, Cramer C, Imus J, et al. A randomized, multicenter, double-blind, 3-way crossover-design trial to evaluate the effect of nitroglycerin compressed tablets and nitroglycerin molded tablets (Nitrostat®) vs placebo in patients with chronic stable angina known to respond to Nitrostat®

Protocol: 782-17

Volumes: 1-34

Rec'd Date: June 4, 1999

1. INTRODUCTION

The sponsors submit a NDA for a compressed tablet formulation of Nitrostat. The primary efficacy study is reported in Cowmeadow et al. Forty patients who passed a screening phase and a qualifying phase participated in the study. In the planning stage of the experiment, a sample size of 36 was deemed sufficient to provide a 90% chance of detecting a difference of 30% of the expected average response with no treatment. Each patient was given all three treatments (0.6 mg compressed nitroglycerin tablet, 0.6 mg molded nitroglycerin tablet, placebo) in random order. In each period, the patient was given a compressed tablet and a molded tablet, but at most one of the tablets had an active ingredient. In other words, if the patient was given a compressed nitroglycerin tablet in a period, then he was also given a molded placebo tablet at the same time. The number of patients randomized into each of the six different possible sequences ranged from 5 to 8. The patients walked on a treadmill until a terminal event was observed (onset of moderate angina, fatigue, headache, etc.). There was a two-hour rest period between each of the three treatment periods. The primary efficacy variable is time to moderate angina. The secondary efficacy variable is time to myocardial ischemia. Five centers were used in this study. The stated objectives are to show that both formulations of Nitrostat are superior to the placebo and that the new formulation of Nitrostat is equivalent to the marketed formulation. No criteria for showing equivalence of the two nitroglycerin tablets was explicitly stated in the protocol.

All of the analyses, graphs, and tables in this review are from the reviewer's own analysis. In some cases, but not all, the analyses are identical to those in the sponsor's report.

2. PLANNED ANALYSES

The protocol states exercise time to event (ETT) will be analyzed by ANOVA to compare each formulation of nitroglycerin tablet to the placebo. The model will include effects due to sequence, treatment, period, and patient nested within sequence. Linear contrasts will be utilized to compare the treatments. Survival analysis techniques will be used to support the primary analysis. Finally, McNemar's test and the Cochran-Mantel-Haenszel procedure will be used to compare incidence of each event between treatments. All tests will be two-sided and conducted at level 0.05.

In the IND submission review dated 11/24/1997 and the protocol review dated July 31, 1998, a reviewer made several important comments that are summarized here. The response is nonnegative. Hence, the ANOVA assumptions need to be checked (normality and homogeneity of variance) and a nonparametric analysis or unpooled

analysis may be necessary. Duration of exercise is not the same as time to moderate angina because there are other events which terminate the exercise. It is unclear how the sponsors define similarity of the compressed and molded formulations.

3. SUMMARY OF ANALYSES IN THE STUDY

Let X= the time until moderate angina for a randomly selected patient and Y= the time until the occurrence of some other terminal event (such as fatigue). The ultimate efficacy variable is X, but what we observe in this data set is ETT= min{X, Y} for 40 different patients together with a variable that indicates whether moderate angina occurred. The distributions of all of these variables depend on a set of covariates such as treatment, age of patient, etc. First, the sponsors model the mean ETT as a linear function of a subset of the covariates. In order to support this analysis, survival analysis techniques are used to try to detect if there is a difference in the distribution of X given different treatments. Specifically, the log-rank test is used to compare the survival curves for the different treatments.

Two alternative mixed models are fit in the study but the model specified in the protocol is not. There are five different sites (SITE); six different orders in which the 3 treatments can be given (RXGRP); three treatments (TRT); three visits (VISIT); and forty patients (PTNO). The models are summarized in Table 3.1. In fact, the mean response does seem to differ at different sites, so it may be useful to include this variable in the model. In addition to SITE, the reviewer would also include SEX and AGE in the model, if given the choice. In a crossover design such as this one, there is usually a concern about a carryover effect. However, in all the models that were studied, no carryover effect was evident. The lack of a carryover effect can be explained by the two-hour washout period between treadmill tests. The estimates of the treatment differences using the model in the protocol are given in Figure 3.1.

No discussion of the model assumptions or diagnostics was given in the study. Normal probability plots and boxplots of the residuals from the model specified in the protocol are given in Figure 3.2 and summary statistics for the residuals are given in Table 3.2. It appears that the residuals have a distribution with heavier tails than the normal distribution. The variances of the residuals appear to be close enough to assume the errors have homogeneous variance.

Table 3.1. Models used in the study and model specified in protocol.

	Fixed effects	Random effect
First model in study	SITE RXGRP TRT VISIT SITE*TRT	PTNO(RXGRP SITE)
Second model in study	SITE RXGRP TRT VISIT	PTNO(RXGRP SITE)
Model specified in protocol	RXGRP TRT VISIT	PTNO(RXGRP)

Figure 3.1a. Estimated treatment differences for ETT using the model specified in protocol.

Treatment difference	Estimate	Standard error	p-value
Market-New	0.039	0.171	0.820
Market-Placebo	0.897	0.170	0.0001
New-Placebo	0.858	0.170	0.0001

Figure 3.1b. Estimated treatment differences for ST Segment Depression of 1 mm using the model specified in protocol.

Treatment difference	Estimate	Standard error	p-value
Market-New	-0.083	0.290	0.775
Market-Placebo	0.833	0.283	0.005
New-Placebo	0.917	0.269	0.001

Figure 3.2. Normal probability plots and boxplots of residuals from model in protocol for ETT.

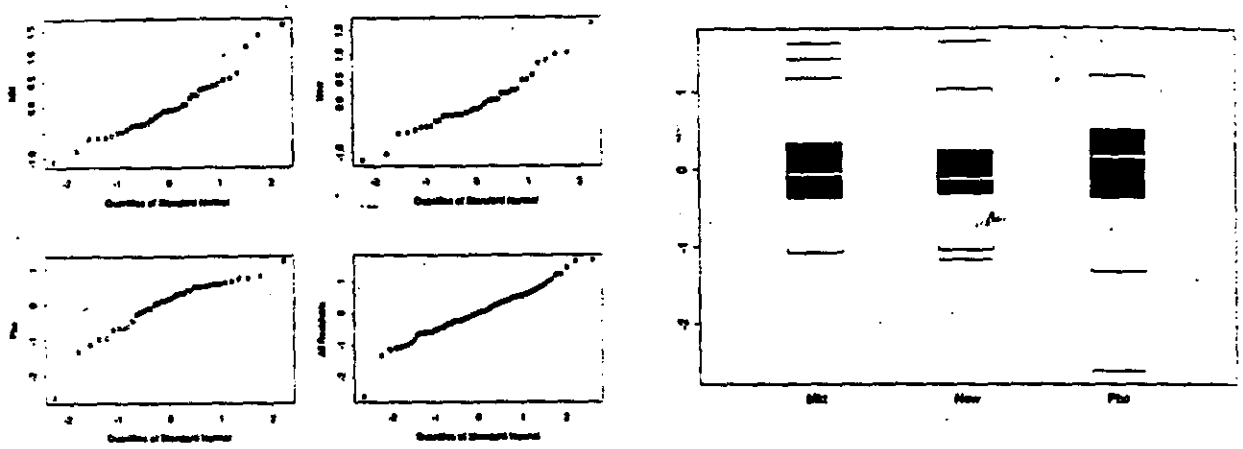

Table 3.2. Summary statistics of residuals from model in protocol. Number in parentheses indicates where this value ranks in a list of 100 values from random samples from the normal distribution with variance 0.382 with the same sample size, i.e. if the number in parentheses is close to 0 or 100, then the corresponding statistic is unusually small or large.

	Variance	Skewness	Kurtosis
Mkt	0.329 (35)	0.898 (99)	3.959 (94)
New	0.314 (24)	0.677 (99)	3.864 (93)
Pbo	0.523 (95)	-1.357 (1)	5.543 (99)
All residuals	0.382	-0.309 (11)	5.132 (100)

The protocol states that survival analysis techniques will be used to supplement the ANOVA analysis, but is not more specific about how this will be done. The log-rank test showed a significant difference between both treatments and the placebo, but no significant difference between the two treatments. When there is a big difference between two populations- as there is between both treatments and the placebo- the log-rank test finds a significant difference. Since the log-ranks test is an omnibus test, it is designed to detect any difference in the survival curves.

Therefore, it cannot be expected to have much power for a specific alternative. Moreover, the sample size in this study was chosen to detect very large differences between nitroglycerin and a placebo, not the difference there may be between the molded and compressed formulations. In order to investigate more subtle differences between populations, it makes sense to model the dependence between the response and the covariates using Cox's proportional hazards model, for instance. Some of the variation in the response may be explained by these covariates. A relatively simple model includes the covariates SITE, AGE, SEX, and HRTSD (standing heart rate prior to exercise). The estimated hazard functions for each treatment group (after adjusting for these covariates) and the log-log survivor functions are presented in Figure 3.3. Under the assumptions of the proportional hazard model, the hazard functions should be constant multiples of each other and the log-log survivor functions should be parallel. Parameter estimates for the proportional hazards model are presented in Figure 3.4.

Figure 3.3a. Graph of smoothed hazard functions for 3 treatment groups. Within each treatment, a different baseline hazard function is used in producing these figures. Covariates in model: SITE1, SITE4, AGE, SEX, HRSTD.

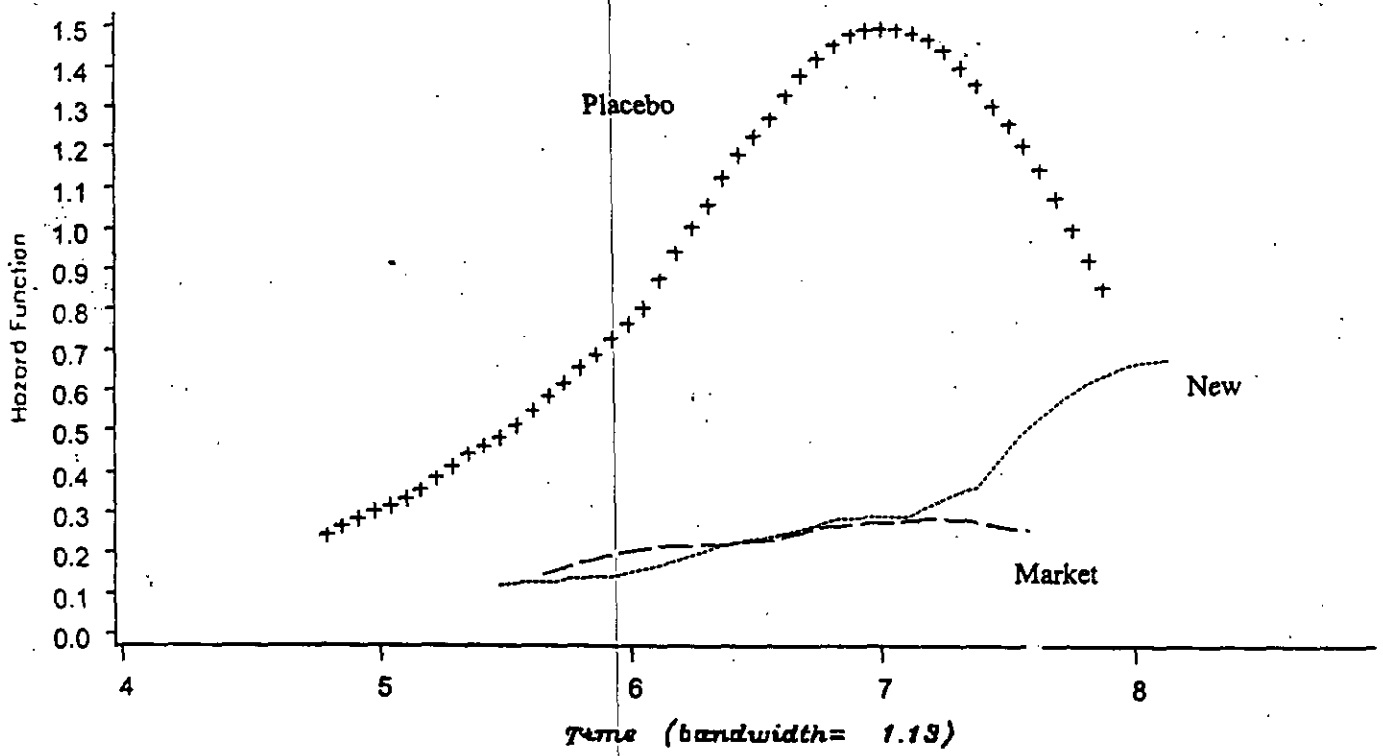

Figure 3.3b. Graph of log(-log survival) plots for 3 treatment groups.

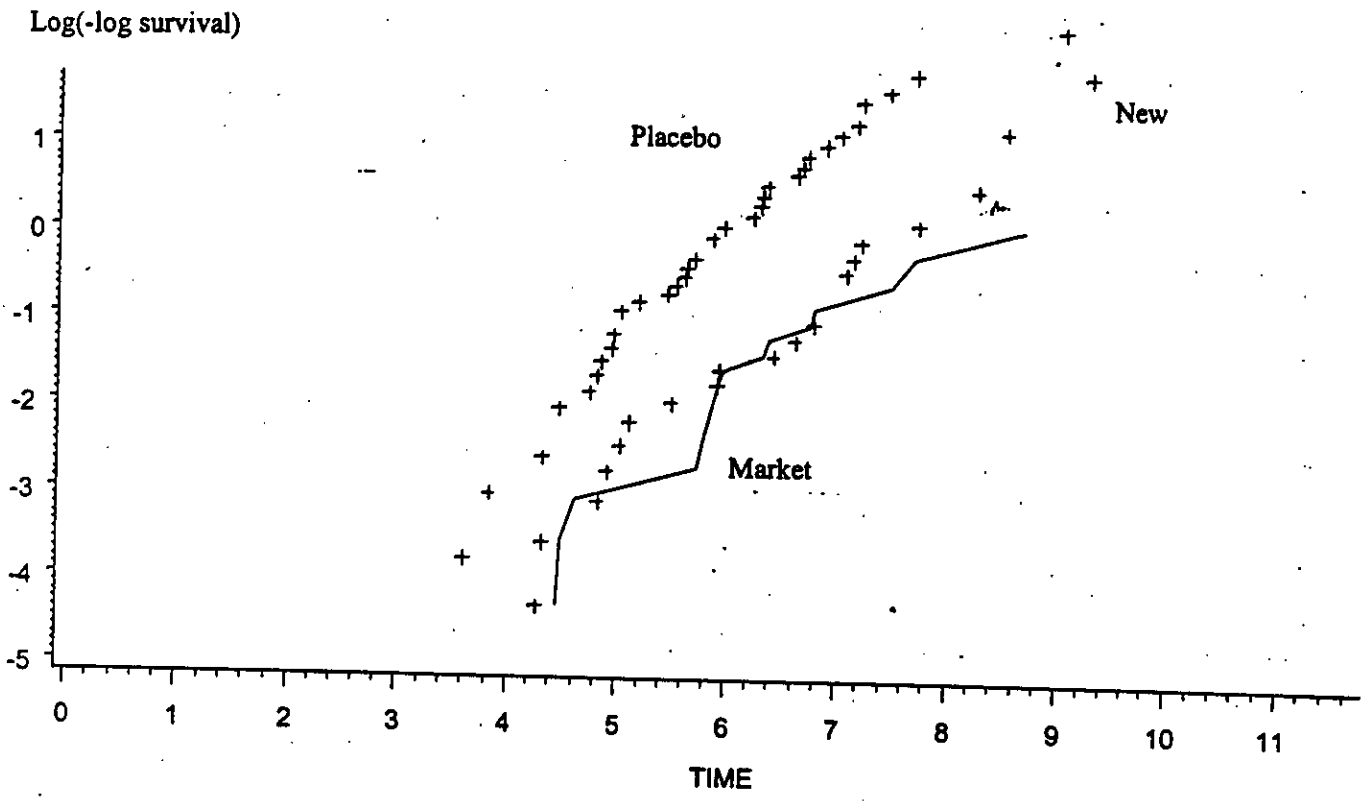

Figure 3.4. Proportional hazards model regression estimates.

Variable	Estimate	Standard error	p-value	Risk ratio
PBOIND	1.501	0.322	0.0001	4.487
NEWIND	0.221	0.341	0.517	1.248
SITE1IND	0.748	0.345	0.030	2.115
SITE3IND	-0.569	0.390	0.145	0.566
SITE4IND	0.755	0.453	0.096	2.129
SITE5IND	-1.405	1.042	0.177	0.245
AGE	0.061	0.016	0.0001	1.063
SEXIND	-1.802	0.348	0.0001	0.165
HRSTD	-0.032	0.008	0.0001	0.968

The parameter estimates in Figure 3.4 can be interpreted as telling us how likely it is that someone with a fixed set of covariates will stop exercising in the next minute. Suppose a 59 year old and a 60 year old patient have both been on the treadmill for 5 minutes and the patients have same sex, etc. Then, the 60-year-old person is 1.063 times as likely to stop in the next minute as the 59-year-old is. A patient who took the new formulation of Nitrostat is 1.248 times as likely to stop in the next minute as a patient who took the old formulation is. The risk ratio is found by exponentiating the parameter estimate in column 2.

4. EXPLORATORY DATA ANALYSIS

One analysis of the primary efficacy variable is the comparison of means of the ETT's using the mixed model. This analysis ignores the fact that the observations are censored. In defense of this analysis, most of the times for the placebo are actually observed and the data are right censored, so the actual significance (if analyzed correctly) is bound to be even stronger than they report for comparing a treatment to the placebo. This cannot be argued for the comparison of the two treatment groups. The analysis of the survival curves is presented to partially address the censoring. The analysis in the study is not satisfactory because the analysis ignores the fact that the observations are

paired. This criticism holds even if the covariates are included as discussed in the previous section. We don't have independent observations on the two treatments because they are observed on the same patients. The first analysis incorporates this structure, while the second does not.

The reviewer thinks that the data clearly show that both the marketed formulation and the new compressed tablet formulation of nitrostat are superior to the placebo. However, the data cast some doubt on the hypothesis that the marketed and the new formulation of nitrostat are equivalent. Hence, this section will only focus on this comparison.

One approach is to try to mimic the idea of a t-test by forming a set of values of the variable X which are consistent with the observed ETT. The empirical mass function for X is derived from the Kaplan-Meier estimates of the survival functions. The mean of this empirical mass function is automatically printed in the SAS output. The observed difference between the two empirical means is 8.588-7.830 = 0.759.

In order to estimate the variance of this estimator, we can draw bootstrap samples from the pooled estimate of the density of X (all 80 observed values of ETT when a patient was given either nitroglycerin treatment are pooled together). Then, we independently draw censoring times from the estimated density of times to terminal event other than moderate angina. Note that if observation i is censored with respect to time to moderate angina, then it is not censored with respect to time to other terminal event and vice versa. This is the nonparametric MLE of the density of X even when the observations are dependent (assuming there are two observations from the same distribution for each patient). So, for each patient, we simulate a time to moderate angina and a time to other terminal event and then define the simulated ETT as the minimum of these two. This creates the bootstrap samples that are used to estimate the variance of the estimator.

The estimated variance of the survival curve based estimate of the mean for uncensored data is 0.282. Hence, the estimated variance of the difference of two such means (if they are independent, these are not) is 0.564 and the standard deviation of the difference is 0.751. If $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ are all iid with variance σ^2, then the variance $\bar{X} - \bar{Y}$ is $2\sigma^2/n$. However, if the pairwise differences X_i-Y_i are iid with variance δ^2, then the variance of $\bar{X} - \bar{Y}$ is δ^2/n. So, we can expect that the variance of the difference in empirical means is, in fact, much smaller than 0.564. Even if the observed difference is only 1 standard deviation above the mean, the problem is that the difference

is in the wrong direction, i.e. the new drug has a lower mean than the old drug.

A 95% confidence interval for the difference between the times to moderate angina using the two formulations of Nitrostat is (-0.71, 2.23). The standard procedure for proving bio-equivalence is to conclude that the two treatments are equivalent if this confidence interval lies completely within tolerance limits defined by 20% of the reference treatment mean. Since the mean time using the marketed formulation is 8.58, these tolerance limits are from -1.72 to +1.72. However, the 95% confidence interval does not lie within these tolerance limits in this case.

For the censored data, the bootstrap average mean is 6.84 and the variance is 0.0625. The actual observed sample means were 6.85 (MKT) and 6.83 (NEW). It is interesting to note that the mean based on the Kaplan-Meier estimate of the survival function is biased downward for the bootstrap samples. On average, the bootstrap means differed from the true mean by –0.146. However, since our statistic is the difference of two means, the bias cancels and no adjustment is needed.

A different approach would model the vector (X^{Pbo}, X^{Mkt}, X^{New}) as a random vector whose distribution depends on the patient. It is convenient to use the accelerated failure time model. For each individual, we will model $\log(X) = \mu + \sigma \epsilon$ where μ is a constant that depends on the patient and the treatment, σ is a constant, and ϵ is an error which has a two-parameter extreme value distribution. Hence, X has a Weibull distribution. One advantage of using the Weibull distribution is that the resulting model can also be interpreted as a proportional hazards model. Unfortunately, three patients were censored with all three treatments. The result of this is that the maximization algorithm which is used to estimate the parameters in the model will not converge. However, if we include the data from the qualifying stage (which were obtained under similar circumstances), we are guaranteed to have at least one uncensored time for each patient. So, the final model contains the covariates PTID (patient ID), NITRO (indicates if patient was treated with nitrostat before exercise), NEWNIT (indicates if patient was treated with new formulation of nitrostat), NOTRT (indicates if patient was in the qualifying stage). The SAS output for the response time to moderate angina appears in Figure 4.1. The coefficient of NEWNIT is -0.0521242 with a standard error of 0.048242. Although it is not significantly different from 0, it is not convincingly close to 0, either. The interpretation of this coefficient is that the time to onset of moderate angina decreases by about 5%, holding everything else constant. This percentage is calculated from the formula $100\,(1-e^{-0.0521242})\%$. In other words, this data indicates that a patient that uses the new nitrostat tablet will experience moderate angina at a time which is 5% less than if the patient had used the old

formulation. In the preceding analysis where the Kaplan-Meier estimates were used without including any covariates, we found a decrease in time of about 8.8%.

Figure 4.1. Estimates of parameters for accelerated lifetime model for ETT. There is no parameter for patient because it is a categorical variable, i.e. each patient has his own individual coefficient.

Variable	Estimate	Standard error	p-value
Intercept	2.010	0.084	0.0001
NITRO (Nitrostat)	0.252	0.044	0.0001
NOTRT (Baseline)	0.044	0.029	0.1339
NEWNIT (New formulation)	-0.052	0.048	0.2799
PTID (Patient)			0.0001

5. SUMMARY

With respect to the placebo, the new formulation has been shown to extend the time to moderate angina and the time to ST segment depression of 1 mm. Likewise, the old formulation of nitrostat has been shown to be superior to the placebo. The new formulation has not been shown to be equivalent to the old formulation. The data suggest that the marketed formulation may, in fact, be superior to the compressed tablet.

/S/

John Lawrence, Ph.D.
Mathematical Statistician

This review consists of 12 pages of text, tables, and figures.

Concur: Dr. Hung /S/ 8/26/99
Dr. Chi 8/30/99

cc: NDA #21-134
HFD-110/Dr. Lipicky
HFD-110/Dr. Williams
HFD-110/Mr. Fromme
HFD-710/Dr. Barton
HFD-710/Dr. Chi
HFD-710/Dr. Mahjoob
HFD-710/Dr. Hung
HFD-710/chron

LAWRENCEJ/594-5375/NITROSTA.doc/08/26/99