CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

20-941

STATISTICAL REVIEW(S)

APR 22 1999

Statistical comments on data re-analysis by Lidakol

NDA:	20-941 /1S
Applicant:	Lidak Pharmaceuticals (Avanir)
Name of Drug:	Lidakol Top Cream 10%
Route of Administration:	Topical
Documents Reviewed:	Lidakol meeting briefing package (dated Feb 26,1999) Letter from Professor R. Thisted to Dr. Wilkin (Dated Mar. 19, 1999; faxed by Mr. James Berg of Avanir on mar. 24, 1999)
Indication:	Treatment of oral/facial herpes simplex
Medical Officer:	Martin Okun, M.D.(HFD-540)

Among the arguments made by the statistical consultants for Lidakol (Avanir), two points seem to be most persuasive to clinicians:

1. The new model (i.e. proportional odds ratio regression, the method used in their post hoc analysis) they are using, is a natural extension of the Wilcoxon test (specified in the protocol).
2. Since the new method was not computationally possible at the time when the study was designed, and became available only recently, they are justified in applying this method to the data from the two studies in NDA 20-941.

Our response at the meeting with the sponsor was:

1. Please explain in detail about the "natural" extension claim.
2. This is a post hoc analysis, and is similar to "shoot the arrow first, then draw the target".

In a letter to Dr. Wilkin, dated March 19, 1999, Professor Thisted (statistical consultant to Lidakol/Avanir) stated:

"Exercise 5.9 through 5.11 [of the book Generalized linear models. McCullagh and Nelder (1989)], deal with two important results (the proofs of which are left to the graduate student reading the book): First, the statistical test for a difference between two treatments using the *proportional odds* model is equivalent to the Wilcoxon test. Second, the statistical test for a difference between two treatments using the *proportional hazards* model is equivalent to the log-rank test."

"That is, the log-rank test is equivalent to the proportional hazards model, and proportional hazards regression generalizes this model to include covariates without imposing additional assumptions. By the identical argument, the Wilcoxon test is equivalent to the proportional odds test, and proportional odds regression generalizes this model to include covariates **without imposing additional assumptions**".

"In the 06/07 protocol, the Wilcoxon test was specifically compared to the logrank test, and the Wilcoxon was selected because of its power profile. Since the Wilcoxon test is

identified as the appropriate two-sample test, the proportional odds regression is the appropriate method to control for covariates. **This is the essence of the argument**."

This seems to be a very strong argument for the "natural extension" claim. Our response will consist of two parts:

1) *Professor Thisted's argument is fatally flawed. The essence of his argument for the "natural extension" is wrong. The Wilcoxon-Gehan test does not impose assumptions on the population distributions. The proportional odds ratio regression model does, and the assumptions are not trivial.*

First, let me explain what "exercise 5.9 through 5.11" is about.

Usually, a statistical test is performed on two random samples, say S_x and S_y, drawn from two population distributions X and Y, the purpose being to see if the two distributions are equal. If the test results in a p-value less than a certain value α (usually, α =0.05), then we say that the test rejects the null hypothesis, and we accept the alternative (i.e. X and Y are different). Otherwise we accept the null hypothesis (i.e., X and Y are equal).

Among the statistical tests, there are non-parametric tests (such as Wilconxon-Gehan test), which do not assume any relationship between the two population distributions X and Y. There are parametric and semi-parametric model tests, which impose assumptions on the distributions X and Y.

The proportional odds ratio regression model, which was used by Lidakol in the re-analysis of the Lidakol data, assumes that the population distributions X and Y (from which two random samples S_x and S_y are drawn for testing) satisfy the proportional odds ratio model. (Contrary to the statement made by Professor Thisted that instead of using the Wilcoxon-Gehan test [which does not need any assumptions on X and Y], one can use this model "***without imposing additional assumptions***").

What does it mean when one assumes that two population distributions X and Y satisfy the proportional odds ratio model?

Suppose X and Y both assume values j=1,2,...,k. (For the Lidakol data, k=10 days × 24 hours =240 hours for time to healing). Let

$$\gamma_j(X)=\text{Prob}(X\leq j)$$

be the cumulative probability of X being up to and including the value j. Then

$$\text{odds}(X\leq j) =\gamma_j(X)/(1-\gamma_j(X))$$

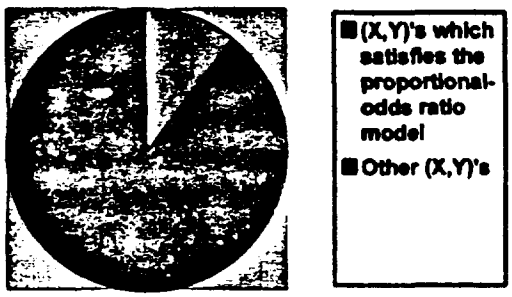
is the odds of X being up to and including the value j. $\gamma_j(Y)$ and odds($Y\leq j$) are defined similarly. If

$$\text{odds}(X \le j) / \text{odds}(Y \le j) = c = \text{constant} > 0,$$

and this constant does not depend on j, *then the pair (X,Y) is said to satisfy the proportional-odds ratio model .*

Because of this definition, the assumption that two population distributions X and Y satisfy the proportional odds ratio model is not a trivial one.

Let A={all pairs of population distributions (X,Y) which satisfy the proportional-odds ratio model}. Ω={All pairs of population distributions (X,Y) }. Then A is a very small fraction of Ω.



In the above picture, the blue area represents the set A, and the red area and the blue area together represents Ω. We point out that the proportional-odds ratio model test is appropriate only for pairs in the blue region but not for pairs in the red region. The Wilcoxon test can be used for pairs in both red and blue regions. For pairs (X,Y) in the blue region, three tests can be performed:

1. The Wilcoxon test;
2. The proportional-odds ratio model test without covariates;
3. The proportional-odds ratio model test with covariates.

Obviously, test 3 is an extension of test 2. "Exercise 5.9 through 5.11" show that tests 1 and 2 are equivalent for pairs in the blue region. Since test 3 is an extension of test 2, it is also an extension of test 1 (for pairs in the blue region). Professor Thisted failed to mention that this is true only for pairs in the blue region. The statement in his March 19, 1999 letter implied that test 3 is also an extension of test 1 on the red region. But in fact, both test 2 and test 3 (the proportional-odds ratio model) are not appropriate for pairs in the red region. Do we know if the pair (X, Y) (i.e. the pair (Lidakol treatment group, placebo group)) satisfy the proportional-odds ratio model? We don't. To suggest that they do, is to <u>impose</u> an assumption. Since Wilcoxon-Gehan test does not impose any assumption on the population distributions, this is contrary to the statement made by Professor Thisted ("the Wilcoxon test is equivalent to the proportional odds test, and proportional odds regression generalizes this model to include covariates <u>**without imposing additional assumptions**</u>"). So the proportional-odds ratio model may not be appropriate for the data in studies 06, 07, not to mention being the natural extension of the Wilcoxon test (which can be used on any pair, regardless of its distribution). Now we see that the ESSENCE of professor Thisted's argument is ***<u>false</u>***.

2) *"If you torture the data long enough, it will confess". We'll demonstrate how the Lidakol data can be made to confess, and explain why post hoc analysis should not be accepted.*

We will show how one can torture the Lidakol data to make it confess. We'll elaborate on why one should not "shoot first, then draw the target" (i.e., perform post hoc analysis). The implication of doing it is that the shooter will be declared a good shooter even if he/she may not be one. In statistical jargon, a type I error is committed. In clinical trials, the post hoc analysis does not follow a plan, which was set *a priori*. The analyst looks into the data (shot has been fired), and then chooses a model (and covariates) he/she thinks to be the best fit (draw a target with the arrow in the center). In the new model used by Lidakol, two new covariates, "stage" and "history", were included. These two covariates were not mentioned in the protocol, and were added to the new model because they provided a "better" fit ("better" means that it produces a smaller p-value). I asked Dr. Okun if there are some other baseline factors which could possibly influence the "time to complete healing". Dr. Okun provided me with a list of possible baseline covariates from information sponsor collected at baseline from subjects enrolled in 96-06 and 96-07: gender, race, lesion location (upper lip, lower lip, corner of mouth, nasolabial junction, cheek, or chin), Presence of other sensations (itching/tingling/burning) at first sign/symptom of episode, Erythema at first sign/symptom of episode, Presence of other sensations (itching/tingling/burning) at initial assessment, Erythema at initial assessment; Average Episode Duration/ Duration of the most Recent Episode . From a literature review on the natural history of herpes simplex (Spruance, S.L. "Herpes Simplex Labialis" in Clinical Management of Herpes Viruses, IOS Press, 1995), the following biologically plausible baseline covariates were: Number of episodes in the past 12 months, Presence of prodrome (i.e., symptoms presaging an episode) at episode onset, Age, Year since first herpes episode, Time that elapses between initial signs/symptoms and first application of study cream. It is noteworthy that "history" is patient's recollection of the average duration of herpes episodes in the past. This data is subject to recall bias and is of questionable accuracy. A review of the literature reveals no information on the intra-subject variability in duration of herpes episodes. Sponsor in other studies (e.g., 92-02) assumes that inter-subject variability equals intra-subject variability. So a natural question is, why were "stage", "history" included in the Lidakol new model, but not other factors? The answer is actually quite straightforward: because the data from studies 06 and 07 demonstrated that the new method with "stage" and "history" in the model is a "good fit". In other words, this target is a good one. A *data driven* model selection process (in which a method and covariates were selected from a range of possibilities) has occurred here. This process is, in essence, equivalent to multi-comparison, but without any adjustment. This increases the chance of a treatment effect being declared when there is none. ***In statistical jargon, post hoc analysis inflates the type I error***. The amount of the inflation depends on the true distribution of (X,Y), and the model selection process (i.e., the multiplicity of the tests). For the Lidakol data, a simulation with the following design can give us an estimate of the type I error resulted from the post hoc analysis.

In this simulation, we mix the patients from the Lidakol and the vehicle groups together. So we have only one population (distribution). Then we draw two random samples from this pooled distribution. So there is no "TRUE" treatment difference between the two samples. If we perform only one pre-specified statistical test on these two samples, then the chance of falsely declaring a treatment difference is 0.05 (the type I error, if we choose α=0.05). If we do the same for both trials 06 and 07, then the type error should be 0.05×0.05=0.0025 (since the tests on samples from 06 and 07 are independent of each other). However, if we go through a model selection process (i.e., we conduct multiple tests), and select a model which produces the smallest p-value for the treatment effect, then we'll increase the chance of falsely declaring a treatment effect dramatically.

Step 1: For the data from study 06, draw two random samples $S_{06,1}$ and $S_{06,2}$ from the totality of all the patients in the study (i.e., both treatment and placebo together), with sample size n=185 (the sample size for each group in studies 06 and 07 were about 185).

For the data from study 07, draw two random samples $S_{07,1}$ and $S_{07,2}$ from the totality of all the patients in the study (i.e., both treatment and placebo together), with sample size n=185.

Step 2: Perform the following tests on $S_{06,1}$ and $S_{06,2}$; then on $S_{07,1}$ and $S_{07,2}$.

1. Wilcoxon test on models which include covariates:
   a. treatment, center;

2. Proportional hazards regression models which include covariates:

   a. treatment, center;
   b. treatment, center, plus any combination of {Stage, History, Gender, Race, Lesion location, Presence of other sensations at first sign/symptom of episode, Erythema at initial assessment, Average Episode Duration/ Duration of the most Recent Episode, Number of episodes in the past 12 months, Presence of prodrome, Age, Year since first herpes episode, Time that elapses between initial signs/symptoms and first application of study cream}

Note: Since the number of factors in the { } bracket is 13, the number of possibilities contained in b is 2^{13}-1=8192-1. So a total of 8192 models will be tested using the proportional hazards regression model. The same is true for the following Proportional-odds ratio regression model and the Log-logistic regression model.

3. Proportional-odds ratio regression models which include covariates:

   a. treatment, center;
   b. treatment, center, plus any combination of {Stage, History, Gender, Race, Lesion location, Presence of other sensations at first sign/symptom of episode, Erythema at initial assessment, Average Episode Duration/

Duration of the most Recent Episode, Number of episodes in the past 12 months, Presence of prodrome, Age, Year since first herpes episode, Time that elapses between initial signs/symptoms and first application of study cream}

4. Log-logistic regression models which include covariates:

   a. treatment, center;
   b. treatment, center, plus any combination of {Stage, History, Gender, Race, Lesion location, Presence of other sensations at first sign/symptom of episode, Erythema at initial assessment, Average Episode Duration/ Duration of the most Recent Episode, Number of episodes in the past 12 months, Presence of prodrome, Age, Year since first herpes episode, Time that elapses between initial signs/symptoms and first application of study cream}

Each test will produce one p-value p_{06} for the samples $S_{06,1}$ and $S_{06,2}$; and a p_{07} for $S_{07,1}$ and $S_{07,2}$. If max $(p_{06}, p_{07}) > \alpha=0.05$, we'll proceed to another test until we finish every test and every model in the above list. If max $(p_{06}, p_{07}) \leq \alpha=0.05$, then reject the null hypothesis, and declare a treatment effect. No further tests will be conducted on the samples $S_{06,1}$ and $S_{06,2}$, $S_{07,1}$ and $S_{07,2}$. The corresponding model will be declared the best fit, and the test method will be declared the most appropriate test. (To be fair, I'm sure that Professor Thisted had not exhausted all the 8192 possibilities for each of the three methods: Proportional hazards regression model, Proportional-odds ratio regression model, Log-logistic regression model [which is a total of $3\times 8192=24576$ possibilities], when he declared that the Proportional-odds ratio regression model with covariates treatment, center, stage, history is the most appropriate model. After all, one only needs one pair of p-values (p_{06}, p_{07}) with both components less than $\alpha=0.05$).

Repeat steps 1 and 2 for N=10000 times, and count the number of rejections of the null hypothesis. Then (# of rejections of the null hypothesis)/N gives an estimate of the type I error (of falsely declaring a treatment effect for both studies). One should not be surprised if this estimate turns out to be equal or close to ***ONE***!

(Next we can choose $\alpha=0.4, 0.3, 0.2, 0.1, \ldots$ in the above steps. Until the type I errors are less then $0.05 \times 0.05=0.0025$).

In summary,

- The proportional odds ratio regression model (the new method adopted by Lidakol/Avanir) is ***not*** a "generalization" of the generalized Wilcoxon-Gehan test **without imposing additional assumptions** as claimed by Professor Thisted in his March 19, 1999 letter to Dr. Wilkin. The Wilcoxon-Gehan test (the test originally specified in the protocol) does not impose assumptions on the population distribution. When the proportional odds ratio regression is used, an assumption is imposed: the distributions being tested satisfy the "proportional odds ratio" relationship. This assumption is an additional one, and is not trivial.

- Post hoc analysis is equivalent to multiple testing without adjustment, it can dramatically increase the type I error (i.e., the possibility of falsely declaring a treatment effect that does not exist).

All above said, we suggest that the sponsor conduct another trial and *pre-select* a statistical method, be it the proportional odds ratio regression model, the proportional hazards regression model, the log-logistic regression model, or the Wilcoxon-Gehan test, with *pre-specified* covariates in the model, for the primary analysis.

/S/ 04/22/99

Ping Gao, PhD.
Mathematical Statistician, DOB III

/S/ Apr 22, 99

Concur: Rajagopalan Srinivasan, Ph.D.
Team Leader, DOB III

HFD 540
NDA 20-941
HFD-540/Dr. Wilkin
HFD-540/Dr. Walker
HFD-540/Dr. Okun
HFD-540/Mr. White
HFD-725/Dr. Huque
HFD-725/Dr. Srinivasan
HFD-725/Dr. Gao
HFD-344/Dr. Carreras
HFD-725 Chron.

This document contains 7 pages.

JUL 19 1999

Statistical Issues in NDA 20-941
(addendum)

NDA:	20-941 /1S
Applicant:	Lidak Pharmaceuticals (Avanir)
Name of Drug:	Lidakol Top Cream 10%
Route of Administration:	Topical
Documents Reviewed:	NDA 20-941 Amendment (Dated June 25, 1999)
Indication:	Treatment of oral/facial herpes simplex
Medical Officer:	Martin Okun, M.D.(HFD-540)

On June 25, 1999, the sponsor, Lidakol/Avanir submitted an amendment which included "Clarification of Remaining Statistical Issues for the Approval of Lidakol", written by the statistical consultants of Avanir.

In this submission, the statistical consultants reiterated the claim that the proportional odds model is a logical extension of the Wilcoxon test, by citing exercises 5.9 through 5.11 in the book: "Generalized linear models". By McCullagh and Nelder (1989).

Reviewer's note: This claim is inaccurate. The proportional odds model extends the Wilcoxon test only when the underlying distributions satisfy the proportional odds relationship, which is what the exercises 5.9 through 5.11 demonstrated. In this case, the null hypothesis is that the odds ratio is one, the alternative hypothesis is that the odd ratio is not one. This extension does not hold when the underlying distributions does not have the proportional odds relationship (in which case the alternative hypothesis would be different than the above mentioned).

If two distributions satisfy the proportional odds relationship, then the true log odds curves plotted against time (or any changed time scale such as log time) should be parallel curves. Using this fact, the sponsor plotted the estimated log-odds curves as diagnostic plots. The diagnostic plots did not show parallel curves. The statistical consultants for Avanir argued that, although the curves are not parallel, they do look like typical estimated log-odds curves from simulated data from truly proportional-odds model distributions, which this reviewer readily agrees. However, this reviewer did simulate data from other non-proportional-odds model distributions, and found that the estimated log-odds curves also look similar. These show that the diagnostic plots are not discriminative, and can not reliably be used to check the assumption of proportional-odds relationship.

Can we reach a conclusion that the proportional odds model should not be used, because we can not verify the assumptions? No. Statistical models are only meant to be tools for statistical inferences, and they are approximations. To use statistical modeling, assumptions have to be made. The assumptions are sometimes checked, but the validity of the model checking varies for different models. Some times, no model checking is done at all. For example, the Cox proportional hazards model is widely used, but many

users do not perform any model checking. Hence, there is no reason to restrict the use of proportional odds model on grounds that valid model checking can not be performed.

In this reviewer's opinion, although the proportional odds model is not a logical extension of the Wilcoxon test in the sense as claimed by the statistical consultants for Lidakol/Avanir, it still can serve as a useful tool for statistical inference. However, the use of the model should be pre-specified, assumptions be made prior to unblinding of the data, and follow the same principles as those when other statistical models are used.

The post-hoc analysis of the Lidakol data using the proportional odds model regression, is not any different than any other post-hoc analyses on other data, using other methods. They are the same in nature. When post-hoc analysis is performed, type I error can not be controlled.

Avanir claimed that the Cox regression was pre-specified, together with the covariates center, stage (prodrome or erythema at baseline), history (historical mean duration >5 days or ≤ 5 days).

Looking back into the records, the following have been the development of the analysis plan for the Lidakol data:

- Original protocol (release date: 6/19/96): "Baseline characteristics considered to be clinically meaningful and prognostically important are identified as the following: (i) Historical mean episode duration (≤ days vs >5 days); (ii) Whether or not the patient experienced prodrome during this recurrence; and (iii) Stage at entry, i.e., prodrome or erythema (with or without prodrome). These variables will be treated as covariates in the Cox regression analysis." "Time-To-event statistical comparisons will be conducted using Cox's proportional hazards regression analysis, using a two-sided test.... The study will be declared positive if the primary variable, 'Time-to-healing', is statistically significant using a two-sided test at a significance level of 0.05."

- In a letter from Dr. Donna J. Freeman, Acting Director of the Division of Antiviral Drug products, to Dr. David H. Katz of Lidakol, on July 25, 1996, it was stated that :
  "5. Our preference is that the primary comparison be based upon mean duration (in hours). Other analyses would be viewed as supportive.
  6. It may be desirable to compare the overall in mean duration of healing using a permutation test stratifying by center.
  7. The use of Cox regression for the comparison of the treatments in terms of the duration of healing may not add greatly to the analysis described above."

- In amended protocol #2, released 2/20/97, the analysis plan was revised: "Time – to-event distributions will be estimated using Kaplan-Meier Product limit estimates and compared using a permutation-based test, i.e., generalized

Wilcoxon test. Cox's proportional hazards regression model may be used to adjust for covariates.... The study will be declared positive if the primary variable, 'time-to-healing' is statistically significant using a two-sided test at a significant level of 0.05."

- In the NDA (20-941) submission (submitted Dec. 22, 1997), the primary statistical analysis on the primary efficacy variable, "time-to-healing", was the generalized Wilcoxon test, stratified by center. No results from Cox's proportional hazards regression model were included in the NDA submission. It is noted, that the baseline covariate, "history" was not included in the efficacy data submitted by Lidakol.

Reviewer's comments: The primary analysis on the primary efficacy variable, as in the original protocol(release date: 6/19/96), was indeed the Cox's proportional hazards regression model adjusted for baseline covariates stage and history. In the amended protocol #2 (released 2/20/97), the primary analysis on the primary efficacy variable, was changed to generalized Wilcoxon test. The efficacy analysis results included in the NDA submission also shows that the intended primary analysis on the primary efficacy variable, was the generalized Wilcoxon test. Results of efficacy analysis based on the Cox's proportional hazards regression model was not mentioned in the NDA submission. Late results submitted by Lidakol (Lidakol meeting briefing package [dated Feb 26,1999. Table A, Page 29]) showed the results were not statistically significant.

Based on these facts, it can be reasonably concluded that the generalized Wilcoxon test was the intended primary analysis method. Cox's proportional hazards regression model with covariates center, stage and history was not the intended primary analysis method. Later submission by Lidakol showed that results from the Cox regression, with covariates center, stage and history, were not statistically significant, confirming the belief of the agency as stated in Dr. Freeman's letter (7/25/96) that "The use of Cox regression for the comparison of the treatments in terms of the duration of healing may not add greatly to the analysis described above".

In the June 25, 1999 submission, the sponsor submitted a stratified Cox regression modeling results. In this analysis, center was not used as a covariate. Instead, center was used as a strata. The Cox regression was performed within each center (i.e., mini Cox regression modeling), with stage and center as covariates. Then the results of the mini Cox regression were combined to produce a single p-value. In this way, a p-value of 0.04 was obtained for study 96-LID-07. However, the p-value for study 96-LID-06 was 0.054.

This reviewer wants to point out that this stratified Cox regression analysis had not been mentioned by Lidakol/Avanir before this June 25, 1999 submission. It is noted that with this stratified Cox regression analysis, a significant p-value can be obtained for only one (p=0.054 for 96-LID-06, p=0.04 for 96-LID-07, see Table 1A, page 5 in the June 25,199 submission) of the two studies. Also, a significant p-value was obtained for only one of the two studies using the pre-specified primary analysis method, the generalized Wilcoxon test (p=0.023 for 96-LID-06, p=0.159 for 96-LID-07).

The proportional odds regression model, is a tool for statistical inference, just as the Cox proportional hazards regression model is. When the underlying distributions are reasonably believed (assumed) to have a proportional odds relationship, and the null hypothesis that the odds ratio is one is tested against the alternative hypothesis that the odds ratio is not one, the proportional odds regression model can be used. In this case, and only in this case, the proportional odds regression model adjusted for covariates extends the Wilcoxon test (which can not adjust for covariates). The logical extension claim made by the statistical consultants implies that the proportional odds regression model extends the Wilcoxon test in all situations when the Wilcoxon test can be used, not just the special case mentioned above, therefore is not true. The use of the proportional odds regression model should follow the same principles as the Cox regression model, all assumptions and covariates should be pre-specified, as well as the null and the alternative hypotheses. The post-hoc analysis on the Lidakol data using the proportional odds can result in inflated type I error with unknown magnitude.

It is noted that from Table A, Page 29 in the Feb 26,1999 submission from Avanir, the addition of the covariates stage, history to different statistical models did reduce the p-values for each model. Hence, some further investigations into the impact of these two covariates may help to better understand the data. Therefore, the sponsor is suggested to do some additional analysis (see attached list).

Reviewer's conclusion: The generalized Wilcoxon test was the intended primary analysis method for the NDA 20-941 submission. The proportional odds regression model extends the Wilcoxon test only under the pre-specified assumption that the underlying distributions satisfy a proportional odds relationship, and the null hypothesis that the odds ratio is one is tested against the alternative hypothesis that the odds ratio is not one. The post-hoc analysis on the Lidakol data using the proportional odds may result in inflated type I error with unknown magnitude. To better understand the impact of the covariates stage and history, the sponsor is suggested to perform some additional analysis, as stated in the attached list.

/S/ 07/19/99
Ping Gao, Ph.D.
Mathematical Statistician, DOB III

/S/ July 19,'99
Concur: Rajagopalan Srinivasan, Ph.D.
Team Leader, DOB III

SEP 14 1999

original copy

Statistical Issues in NDA 20-941
(addendum)

NDA:	20-941 /1S
Applicant:	Lidak Pharmaceuticals (Avanir)
Name of Drug:	Lidakol Top Cream 10%
Route of Administration:	Topical
Documents Reviewed:	NDA 20-941 additional information (Dated Aug. 3, 1999)
Related Documents:	Briefing package for Lidakol (Feb. 26, 1999), NDA 20-941 Amendment (Dated June 25, 1999)
Indication:	Treatment of oral/facial herpes simplex
Medical Officer:	Martin Okun, M.D.(HFD-540)

On August 3, 1999, the sponsor, Lidakol/Avanir submitted the results of additional statistical analyses as requested in Dr. Delap's letter to Avanir, dated July 21, 1999.

The results included the descriptive analysis for sub-populations defined by the levels of the covariates stage ("prodrome" or "erythema"), history (historical mean episode duration >5 days or ≤ 5days). The addition of these two baseline covariates did reduce the p-values in the Cox proportional hazards regression, log-rank regression, proportional odds regression analyses (see Feb. 26, 1999 and June 25, 1999 submissions). Therefore, the agency requested the sub-population analyses to check the treatment effects on the sub-populations. The requested descriptive analyses included the mean, median, n, variation, and the Kaplan-Meier curves. Generalized Gehan-Wilcoxon analyses results were also requested.

The Kaplan-Meier curves comparing the influence of the covariates history, stage demonstrated some separation between the healing time curves for patients with shorter and longer historical episode duration, and prodrome and erythema patients for the combined studies 96-06 and 96-07, indicating these two covariates may have impact on the healing times. However, as pointed out by the statistical consultants for Avanir, the sample sizes in the subsets of prodrome patients and patients with shorter historical episodes are small compared with erythema patients and patients with longer historical episodes. Since the curves from small population are subject to large variation, other quantitative descriptive statistics need to be investigated to help analyze the influence of these two covariates.

The results of the exploratory sub-set analyses, which include median, n, p-value from the Gehan-Wilcoxon test, are listed in Table 1. As pointed out by the statistical consultants for Avanir, the statistics based on means are difficult to interpret due to censoring, and are not included in the table.

Table 1 Results of sub-set analyses

Study	Subset		Treatment					
			Lidakol		Placebo			
	Historical episode duration	Stage	n	median	n	median	Diff. med	p-value
96-06			183	94.9	183	113.8	18.9	0.0235
	≤ 5 days		16	91.8	38	72.7	-19.1	0.136
	> 5 days		167	95.5	145	118.2	22.7	0.002
		Prodrome	40	76.5	50	74.0	-2.5	0.788
		Erythema	143	96.3	133	118.5	22.3	0.002
	≤ 5 days	Prodrome	3	51.3	11	54.6	3.3	0.356
	≤ 5 days	Erythema	13	94.9	27	94.2	-0.8	0.349
	> 5 days	Prodrome	37	93.6	39	94.5	0.9	0.631
	> 5 days	Erythema	130	100.5	106	122.3	21.8	0.007
96-07			187	102.3	184	118.2	15.9	0.1529
	≤ 5 days		31	70.0	36	101.4	31.4	0.043
	> 5 days		156	114.8	148	119.0	4.3	0.586
		Prodrome	31	49.4	30	87.6	38.2	0.072
		Erythema	156	116.7	154	122.0	5.3	0.717
	≤ 5 days	Prodrome	8	70.0	6	65.8	-4.2	0.606
	≤ 5 days	Erythema	23	69.5	30	114.6	45.1	0.019
	> 5 days	Prodrome	23	46.9	24	90.9	44.0	0.014
	> 5 days	Erythema	133	119.8	124	122.2	2.4	0.854
96-06 and 96-07 combined								
	≤ 5 days		47	82.4	74	93.8	11.3	0.490
	> 5 days		323	103.7	293	118.2	14.5	0.005
		Prodrome	71	68.5	80	77.1	8.6	0.456
		Erythema	299	112.1	287	119.3	7.2	0.009
	≤ 5 days	Prodrome	11	70.0	17	65.8	-4.2	0.981
	≤ 5 days	Erythema	36	86.9	57	99.4	12.5	0.264
	> 5 days	Prodrome	60	68.5	63	90.9	22.4	0.195
	> 5 days	Erythema	263	116.0	230	122.2	6.2	0.082

Table 1 provides some explanation for the reduction in p-values when the covariate "historical episode duration" and "stage" are added to the regression models. The influence of the covariates is assessed separately below.

Historical episode duration:

In study 96-06, for patients whose historical episode duration is less than or equal 5 days, the median time to heal for the Lidakol patients was longer than that for the placebo patients, with the difference being –19.1 hours; while for patients whose historical episode duration is longer than 5 days, the median time to heal for the Lidakol patients was shorter than that for the placebo patients, with the difference being 22.7 hours. This suggests that the patients with longer historical episode duration may benefit from the Lidakol treatment, but not the patients with shorter historical episode duration. Therefore, adding this covariate to the regression models may reduce the p-values for regression analyses on study 96-06. However, since the sample size of patients with shorter

historical episode (Lidakol 16, placebo 38) is small, the estimated influence of this covariate is subject to large variation, and thus difficult to interpret.

Figure 1 shows the difference in median time to heal between the Lidakol group and the placebo group for patients with shorter historical episode duration (≤ 5days) and with longer duration (> 5days) for study 96-06.

Figure 1 median difference between treatment groups according to patients historical episode duration for study 96-06

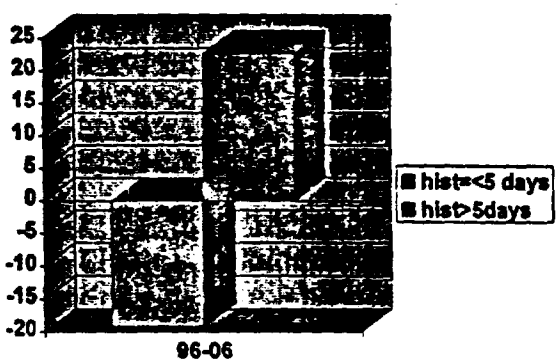


Figure 1 suggests that the patients with longer historical episode duration may benefit from the Lidakol treatment, but not the patients with shorter historical episode duration.

In study 96-07, for patients whose historical episode duration is less than 5 days, the median time to heal for the Lidakol patients was shorter than that for the placebo patients, with the difference being 31.4 hours; while for patients whose historical episode duration is longer than 5 days, the median time to heal for the Lidakol patients was also shorter than that for the placebo patients, but with a much smaller difference of 4.3 hours. This suggests that the patients with shorter historical episode duration may benefit from the Lidakol treatment, but the patients with longer historical episode duration may not benefit. Therefore, adding this covariate to the regression models may reduce the p-values for regression analyses on study 96-07. However, since the sample size of patients with shorter historical episode (Lidakol 31, placebo 36) is small, the estimated influence of this covariate is subject to large variation, and thus difficult to interpret.

Figure 2 shows the difference in median time to heal between the Lidakol group and the placebo group for patients with shorter historical episode duration (≤ 5days) and with longer duration (> 5days) for study 96-07.

Figure 2 median difference between treatment groups
according to patients historical episode duration for study 96-07

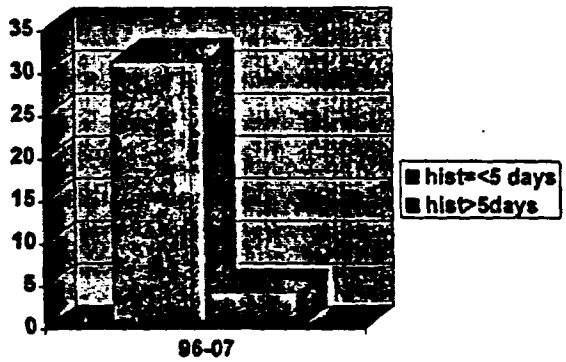


Figure 2 suggests that the patients with shorter historical episode duration may benefit from the Lidakol treatment, but the patients with longer historical episode duration may not benefit.

Hence, the reason that the covariate "historical episode duration" reduced the p-values from regression modeling in the study 96-06 is conflicting to that in 96-07: In study 96-06, treatment effect is *not* in favor of Lidakol in patients whose historical episode duration is less than 5 days, and in favor of Lidakol in patients whose historical episode duration is longer than 5 days. In study 96-07, treatment effect is in favor of Lidakol in patients whose historical episode duration is less than 5 days, and Lidakol showed a much smaller effect in patients whose historical episode duration is longer than 5 days.

It is noted that in both studies 96-06 and 96-07, the sample size of patients with shorter historical episode is small (In study 96-06: Lidakol group sample size is 16, placebo group size is 38; In study 96-07: Lidakol group sample size is 31, placebo group size is 36), which results in wide median difference prediction intervals, i.e., the estimated medians are subject to large variations. The conflicting suggestions on the influence of historical episode duration from the two studies may have been the results of the variations(in other words, random effects).

Therefore, the true influence of "historical episode duration" is difficult to interpret, and one can not rule out the possibility that the influence of the covariate "historical episode duration" on various regression modeling was due to random effect.

Figure 3 shows the difference in median time to heal between the Lidakol group and the placebo group for patients with shorter historical episode duration (≤ 5days) and with longer duration (> 5days) for the studies 96-06 and 96-07.

Figure 3 median difference between treatment groups according to patients historical episode duration for studies 96-06 and 96-07

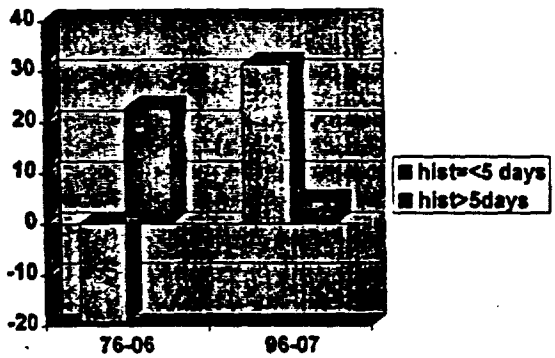


Figure 3 shows a conflict between the suggestions from study 96-06 and 96-07 as to who may benefit from Lidakol treatment, and who may not: Patients with shorter historical episode? Or patients with longer historical episode duration ? The large variation caused by the small sample size in patients with shorter historical episode makes it difficult to interpret the influence of this covariate.

Stage:

In study 96-06, for patients who initiated treatment in prodrome stage, the median time to heal for the Lidakol patients was longer than that for the placebo patients, with the difference being –2.5 hours; while for patients who initiated treatment in erythema stage, the median time to heal for the Lidakol patients was shorter than that for the placebo patients, with the difference being 22.3 hours. This suggests that the patients who initiated treatment in erythema stage may benefit from the Lidakol treatment, but not the patients who initiated treatment in prodrome stage. Therefore, adding this covariate to the regression models can reduce the variance, and reduce the p-values for regression analyses on study 96-06. However, since the sample size of prodrome patients (Lidakol 40, placebo 50) is small, the estimated influence of this covariate (regression coefficients) is subject to large variation, and thus difficult to interpret.

Figure 4 shows the difference in median time to heal between the Lidakol group and the placebo group for patients who initiated treatment at prodrome stage and at erythema stage for study 96-06.

Figure 4 median difference between treatment groups according to patients' stage at which treatment was initiated for study 96-06

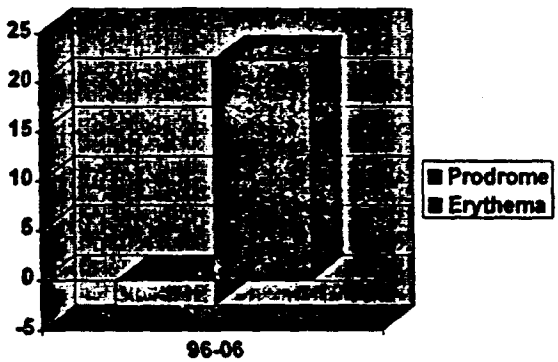


Figure 4 suggests that the patients who initiated treatment at erythema stage may benefit from the Lidakol treatment, but not the patients who initiated treatment at prodrome stage.

In study 96-07, for patients who initiated treatment in prodrome stage, the median time to heal for the Lidakol patients was shorter than that for the placebo patients, with the difference being 38.2 hours; while for patients who initiated treatment in erythema stage, the median time to heal for the Lidakol patients was also shorter than that for the placebo patients, but with a much smaller difference of 5.3 hours. This suggests that the patients who initiated treatment in prodrome stage may benefit from the Lidakol treatment, but the patients who initiated treatment in erythema stage may not. Therefore, adding this covariate to the regression models can reduce the variance, and reduce the p-values for regression analyses on study 96-07. However, since the sample size of prodrome patients (Lidakol 31, placebo 30) is small, the estimated influence of this covariate (regression coefficients) is subject to large variation, and thus difficult to interpret.

Figure 5 shows the difference in median time to heal between the Lidakol group and the placebo group for patients who initiated treatment at prodrome stage and at erythema stage for study 96-07.

Figure 5 median difference between treatment groups according to patients' stage at which treatment was initiated for study 96-07

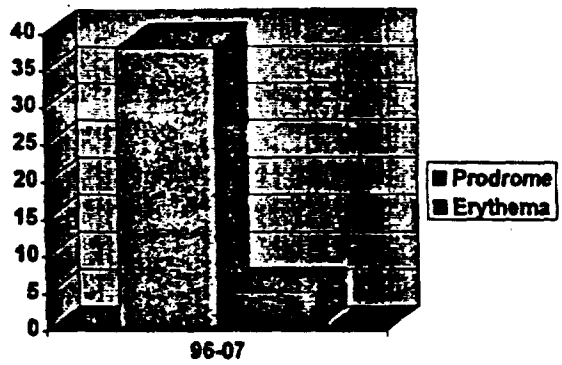


Figure 5 suggests that the patients who initiated treatment at erythema stage may benefit from the Lidakol treatment, but not the patients who initiated treatment at prodrome stage.

Hence, the reason that the covariate "stage" reduced the p-values from regression modeling in the study 96-06 is conflicting to that in 96-07: In study 96-06, treatment effect is *not* in favor of Lidakol in patients who initiated treatment in prodrome stage, and in favor of Lidakol in patients who initiated treatment in erythema stage. In study 96-07, treatment effect is in favor of Lidakol in patients who initiated treatment in prodrome stage, and Lidakol showed a much smaller effect in patients who initiated treatment in erythema stage.

It is noted that in both studies 96-06 and 96-07, the sample size of prodrome patients is small (In study 96-06: Lidakol group sample size is 40, placebo group size is 50; In study 96-07: Lidakol group sample size is 31, placebo group size is 30), which results in wide median difference prediction intervals, i.e., the estimated medians are subject to large variations. The conflicting suggestions on the influence of stage from the two studies may have been the results of the variations (in other words, random effects).

Therefore, the true influence of "stage" is difficult to interpret, and one can not rule out the possibility that the influence of the covariate "historical episode duration" on various regression modeling was due to random effect.

Figure 6 shows the difference in median time to heal between the Lidakol group and the placebo group for patients who initiated treatment at prodrome stage and at erythema stage for studies 96-06 and 96-07.

Figure 6 median difference between treatment groups according to patients' stage at which treatment was initiated for studies 96-06 and 96-07

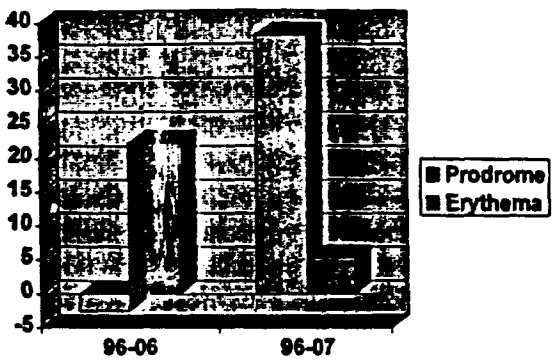


Figure 6 shows a conflict between the suggestions from study 96-06 and 96-07 as to who may benefit from Lidakol treatment, and who may not: Patients who initiated treatment at prodrome stage? Or patients who initiated treatment at erythema stage? The large variation caused by the small sample size in prodrome makes it difficult to interpret the influence of this covariate.

Reviewers comments: The conflicting suggestions implied in the two individual studies 96-06 and 96-07 about the influence of the covariates "historical episode duration" and "stage" points to the possibility that the reduction in p-values in the regression modeling by adding these two covariates to the models may have been the results of random effect.

One purpose of the sub-set analysis was to identify some sub-population of patients who may benefit from Lidakol treatment. As pointed out by the statistical consultants for Avanir, some sub-sets had too few patients to provide valid information. However, some sub-sets, although small, did have strong influence on the results from the overall population. There are also some sub-populations that had more than 100 patients per treatment group. Hence, in order to study the effect of the covariates and to identify sub-population of patients who benefit from Lidakol treatment, the results from these larger subsets and the influential small sub-sets are re-grouped in Table 2 for comparison between study 96-06 and 96-07.

Table 2 Results from larger or influential sub-sets

subset		Study	Treatment					
			Lidakol		Placebo			
Historical episode duration	stage		n	median	n	median	Diff. med	p-value
≤ 5 days		96-06	16	91.8	38	72.7	-19.1	0.136
		96-07	31	70.0	36	101.4	31.4*	0.043
> 5 days		96-06	167	95.5	145	118.2	22.7	0.002
		96-07	156	114.8	148	119.0	4.3**	0.586
	Prodrome	96-06	40	76.5	50	74.0	-2.5*	0.788
		96-07	31	49.4	30	87.6	38.2*	0.072
	Erythema	96-06	143	96.3	133	118.5	22.3	0.002
		96-07	156	116.7	154	122.0	5.3**	0.717
≤ 5 days	Erythema	96-06	13	94.9	27	94.2	-0.8	0.349
		96-07	23	69.5	30	114.6	45.1*	0.019
> 5 days	Erythema	96-06	130	100.5	106	122.3	21.8	0.007
		96-07	133	119.8	124	122.2	2.4**	0.854

*These subsets have large median difference in favor of Lidakol, and carried the results for study 96-07. However, these results are reversed in the corresponding subsets in study 96-06, showing a large variation due to the small sample size in the subsets.

** These subsets have relatively large sample size (n>100 per group). The treatment effect for Lidakol is minimal on these subsets in study 96-07.

In subsets with small sample size (n≤ 50 per treatment group) but with large median difference in at least one of the two studies, it is noted that:

In patients whose historical episode duration were less than 5 days, the median healing time for the Lidakol patients (n=31) in study 96-07 was 31.4 hours shorter than the placebo patients (n=36); however, the median healing time for the Lidakol patients (n=16) in study 96-06 was 19.1 hours longer than the placebo patients (n=38).

In patients who initiated treatment at prodrome stage, the median healing time for the Lidakol patients (n=31) in study 96-07 was 38.2 hours shorter than the placebo patients (n=30); however, the median healing time for the Lidakol patients (n=40) in study 96-06 was 2.5 hours longer than the placebo patients (n=50).

In patients who initiated treatment at erythema stage and whose historical episode duration was less than 5 days, the median healing time for the Lidakol patients (n=23) in study 96-07 was 45.1 hours shorter than the placebo patients (n=30); however, the median healing time for the Lidakol patients (n=13) in study 96-06 was 0.8 hours longer than the placebo patients (n=27).

It is noted that there are large median differences in the small subsets in study 96-07, in favor of Lidakol. However, in each corresponding small subset in 96-06, there is a reversal to the results in study 96-07. It is seen that these large variations are due to the small sample sizes in the subsets. The prospect of repeating the results of study 96-07 in the subsets does not seem very likely.

In subsets with relatively larger sample size (n>100 per treatment group), it is noted that:

In patients whose historical episode duration was longer than 5 days, the median healing time for the Lidakol patients (n=167) in study 96-06 was 22.7 hours shorter than the placebo patients (n=145); however, the median healing time for the Lidakol patients (n=156) in study 96-07 was only 4.3 hours shorter than the placebo patients (n=148).

In patients who initiated treatment at erythema stage, the median healing time for the Lidakol patients (n=143) in study 96-06 was 22.3 hours shorter than the placebo patients (n=133); however, the median healing time for the Lidakol patients (n=156) in study 96-07 was only 5.3 hours shorter than the placebo patients (n=154).

In patients who initiated treatment at erythema stage and whose historical episode duration was longer than 5 days, the median healing time for the Lidakol patients (n=130) in study 96-06 was 21.8 hours shorter than the placebo patients (n=106); however, the median healing time for the Lidakol patients (n=133) in study 96-07 was only 2.4 hours shorter than the placebo patients (n=124).

It is noted that in study 96-06, in all subsets with relatively large sample size (n>100 per treatment group), Lidakol was statistically superior to placebo. While in study 96-07, in all subsets with relatively large sample size (n>100 per treatment group), the difference in median time to healing between Lidakol and placebo was minimal, showing little treatment effect by Lidakol.

From the above results, it is difficult to pick out a sub-population of patients who might benefit from Lidakol treatment, since the efficacy of Lidakol was not consistently demonstrated in the two studies 96-06 and 96-07 for any of the sub-populations, whether large, or small.

Conclusions:

- The subset analysis showed that the influence of the covariates "historical episode duration" and "stage" on the p-values in regression modeling analyses (as shown in the June 25, 1999 submission by Avanir) is difficult to interpret. The conflicting suggestions implied in the two individual studies 96-06 and 96-07 about the influence of the covariates "historical episode duration" and "stage" point to the possibility that the reduction in p-values in the regression modeling by adding these two covariates to the models may have been the results of random effect. This further exposed the inappropriateness of the post-hoc analyses using the proportional odds ratio regression modeling incorporating these two covariates, performed by the statistical consultants for Avanir.
- The subset analyses did not provide consistent and convincing evidence that Lidakol is beneficial for some sub-population of patients, defined according to the patient's historical episode duration and stage at the initiation of treatment.

In study 96-07, the difference in treatment effect between Lidakol and placebo for the larger subsets (n>100 per treatment group) were minimal. The larger difference for the study as a whole was carried by some small sized subsets with very large difference in favor of Lidakol. These large differences in the small sized subsets are subject to large variation, as evidenced by reversals in similar subsets in study 96-06. The prospect of repeating the results of study 96-07 in the subsets does not seem very likely. In study 96-07, in all subsets with relatively large sample size (n>100 per treatment group), the difference in median time to healing between Lidakol and placebo was minimal, showing little treatment effect by Lidakol.

/S/ Sept 14, 99

For ...ing Gao, Ph.D.
Mathematical Statistician, DOB III

/S/

Sept 14, '99

Concur: Rajagopalan Srinivasan, Ph.D.
Team Leader, DOB III

HFD 540
NDA 20-941
HFD-540/Dr. Wilkin
HFD-105/Dr. DeLap
HFD-540/Dr. Walker
HFD-540/Dr. Okun
HFD-540/Mr. White
HFD-725/Dr. Huque
HFD-725/Dr. Srinivasan
HFD-725/Dr. Gao
HFD-700/Dr. O'Neil
HFD-700/Dr. Anello
HFD-720/Dr. Flyer
HFD-344/Dr. Carreras
Chron. (HFD-725)

This review contains 11 pages.

SEP 15 1998

STATISTICAL REVIEW AND EVALUATION

NDA: 20-941 /1S
Applicant: Lidak Pharmaceuticals
Name of Drug: Lidakol Top Cream 10%
Route of Administration: Topical
Documents Reviewed: NDA 20-941: Vol. 2.1, 2.62-2.112 (Total Vol. : 2.1-2.133) (submitted Dec. 22, 1997)
Indication: Treatment of oral/facial herpes simplex
Medical Officer: Martin Okun, M.D.(HFD-540)

Table of Contents:

Background : The active component in LIDAKOL, *n*-docosanol, is a broad-spectrum antiviral compound active *in vitro* and *in vivo* against lipid-enveloped viruses, including herpes simplex virus-1 (HSV-1) and HSV-2. The proposed indication for LIDAKOL is the treatment of recurrent oral-facial herpes simplex (fever blisters or cold sores) in adults.

The submission included two U.S. double blind, PEG placebo-controlled pivotal trials (**96-LID-06** and **96-LID-07**), to support the claim of efficacy. The results (no data) of four U.S. double blind stearic acid "placebo"-controlled clinical trials (94-LID-04, 95-LID-10, 92-LID-04, and 92-LID-02), one active-treatment controlled trial (94-LID-01), were submitted to provide additional evidence for the claim of effectiveness in this indication.

Study Design : **96-LID-06** and **96-LID-07** were both pivotal, randomized, double blind, parallel study.

Table 1. Study designs

Study	Study Design	Dose/duration	No. of subjects Enrolled		Enrolled Subjects
			LIDAKOL	Placebo	sex, age
96-LID-06	Clinic-initiated, double blind, PEG placebo controlled in patients with recurrent oral facial herpes simplex	Study drug 5 times a day until episode aborted, complete healing, or a maximum of 10 days	183	183	male =110
					female = 256
					age: 18-80
96-LID-07	Clinic-initiated, double blind, PEG placebo controlled in patients with recurrent oral facial herpes simplex	Study drug 5 times a day until episode aborted, complete healing, or a maximum of 10 days	187	184	male=103
					female =268
					age: 18-77

Efficacy Analysis

Primary Efficacy Variables: The primary efficacy parameter was the time from therapy initiation to complete resolution of all local signs/symptoms (censored at Day 10) for all patients, including those with classical episodes and those with aborted episodes.

Secondary Efficacy Variables : include

- the time from treatment initiation to complete cessation (duration) of pain and/or burning, itching, or tingling
- the time from treatment initiation to complete cessation (duration) of burning, itching, or tingling
- the time from treatment initiation to complete cessation (duration) of pain
- the time from first experience of pain to first reduction of pain
- the time from treatment initiation to complete healing of lesions which progressed to the vesicular or later stages (i.e. classical episodes)
- the time from treatment initiation to cessation of vesicular stage, of ulcer/soft crust stage, and of hard crust stage
- the percentage of cases that were aborted episodes (i.e., did not progress to the vesicular stage)

Results:

Study 96-LID-06

The patient deposition for Study 96-LID-06 is listed in Table 2.

Table 2 Patient disposition- 96-LID-06

Population	LIDAKOL	PLACEBO	TOTAL
Randomization/Safety Evaluable	185	185	370
ITT	183	183	366
Efficacy Evaluable	178	179	357

In Table 3, the results of the analysis of the primary efficacy variable, hours to complete healing, are shown. When the analysis is not stratified by sites (study centers), the "lifetest" procedure in SAS showed no significant difference in the distribution between the Lidakol group and the placebo group(log-rank test: 0.4607; (Gehan)-Wilcoxon-(Breslow): 0.2416; -2log (LR): 0.5105). When the analysis is stratified by sites (study centers), using the "lifetest" procedure in SAS (which does not provide the stratification by site analysis), combined with a SAS macro provided by the sponsor (the stratification by site analysis can be done by using the combination of this macro and the "lifetest" procedure), the Gehan-Wilcoxon-Breslow (stratified by sites) rank sum test showed a statistically significant p-value of 0.0235.

Table 3 Primary Efficacy Result - hours to complete healing
ITT Population- 96-LID-06

	Lidakol (N=183)	Placebo (N=183)	p-value	Test
Number (%) healed within 10 days	173 (95%)	170 (93%)		
Number (%) Censored Number (%) Discontinued early Number (%) Not healed by day 10	10 (5%) 3 (2%) 7 (4%)	13(7%) 5(3%) 8(4%)		
Hours to complete healing 25th percentile 50th percentile (median) 75th percentile Mean time to heal	 55.3 94.9 150.9 109	 65.5 113.8 161.5 115.8		
			0.4607	log-rank
			0.2416	Wilcoxon*
			0.5105	-2log(LR)
			0.0235	Gehan-Wilcoxon**

* (Gehan)-Wilcoxon-Breslow rank sum test in SAS "lifetest" procedure
** Stratified (by site) (Gehan)-Wilcoxon-Breslow rank sum test

Reviewer's Note: Although the Stratification-by-site analysis was the specified method in the protocol, the difference between the p-value from the stratified analysis (0.0235) and those form the non-stratified tests (0.4607, 0.2416, 0.5105) raises a question on the real implication of the p-value from the stratified analysis: Is the stratified analysis more powerful, or does it inflate the type I error? Investigation into each site was inconclusive.

Table 4 contains the results of the analysis of the primary efficacy variable, hours to complete healing, for the subjects who developed classical episode. None of the tests showed a statistically significant p-value (p>0.27.)

Table 4 Classical Oral-Facial Herpes Simplex Episodes
ITT Population- 96-LID-06

	Lidakol (N=112)	Placebo (N=128)	p-value	Test
Number (%) healed within 10 days	102 (91.1%)	115 (89.8%)		
Hours to complete healing 25th percentile 50th percentile (median) 75th percentile mean	 91.0 138.8 184.5 136.7	 103.6 138.3 190.0 142.3		
			0.8120	log-rank
			0.4438	Wilcoxon*
			0.7279	-2log(LR)
			0.2710	Gehan-Wilcoxon**

* (Gehan)-Wilcoxon-Breslow rank sum test in SAS "lifetest" procedure
** Stratified (by site) (Gehan)-Wilcoxon-Breslow rank sum test

In Table 5, the results of the analyses of time to cessation of discrete classical lesion stages are shown. None of the tests on hours to cessation of vesicle stage, hours to cessation of hard crust stage was statistically significant ($p>0.35$). None of the non-stratified (by site) test showed any statistically significant difference ($p>0.05$) in the hours to cessation of ulcer/soft crust stage. The stratified (by site) Gehan-Wilcoxon test had a p-value of 0.014.

Table 5 Time to cessation of Discrete classical lesion stages
ITT Population- 96-LID-06

	Lidakol	Placebo	p-value	Test
Hours to cessation of vesicle stage				
Number evaluated/total N	70/70	91/91	0.7024	log-rank
Median	49.4	49.9	0.9674	Wilcoxon*
mean (s.e.)	57.3 (3.26)	56.1 (2.54)	0.8916	-2log(LR)
			0.46713	Gehan-Wilcoxon**
Hours to cessation of ulcer/soft crust stage				
Number evaluated/total N	90/90	110/110	0.2232	log-rank
Median	76.5	89	0.0518	Wilcoxon*
mean (s.e.)	86.7 (4.69)	95.8 (3.8)	0.4849	-2log(LR)
			0.0141	Gehan-Wilcoxon**
Hours to cessation of hard crust stage				
Number evaluated/total N	93/100	110/120	0.8740	log-rank
Median	138.8	138.3	0.5064	Wilcoxon*
mean (s.e.)	138 (5.47)	143 (4.54)	0.7286	-2log(LR)
			0.3562	Gehan-Wilcoxon**

* (Gehan)-Wilcoxon-Breslow rank sum test in SAS "lifetest" procedure
** Stratified (by site) (Gehan)-Wilcoxon-Breslow rank sum test

Time to reduction/cessation of signs and symptoms was analyzed and the results are shown in Table 6. Non-stratified tests on hours to first reduction of pain score, hours to cessation of pain, hours to cessation of burning/itching/tingling, hours to cessation of pain and/or burning/itching/tingling did not show statistically significant difference between the Lidakol group and the placebo group ($p>0.05$), while the stratified (by site) Gehan-Wilcoxon test showed the differences were statistically significant ($p<0.05$) for all of the four end-points.

Table 6 Time to reduction/cessation of signs and symptoms
ITT Population- 96-LID-06

	Lidakol	Placebo	p-value	Test
Hours to first reduction of pain score				
Number evaluated/total N	102/102	106/106	0.2474	log-rank
Median	20	23.46	0.2269	Wilcoxon*
mean (s.e.)	28.43 (2.65)	32.3 (2.7)	0.3600	-2log(LR)
			0.0062	Gehan-Wilcoxon**
Hours to cessation of pain				
Number evaluated/total N	102/103	106/106	0.0572	log-rank
Median	48.3	53	0.0614	Wilcoxon*
mean (s.e.)	56.73 (4.13)	69.3 (4.7)	0.1529	-2log(LR)
			0.0125	Gehan-Wilcoxon**
Hours to cessation of burning/itching/tingling				
Number evaluated/total N	176/179	175/178	0.8489	log-rank
Median	48.67	54.2	0.1986	Wilcoxon*
mean (s.e.)	64.66 (3.49)	66.74 (3.12)	0.8432	-2log(LR)
			0.0403	Gehan-Wilcoxon**
Hours to cessation of pain and/or burning/itching/tingling				
Number evaluated/total N	179/183	176/179	0.4589	log-rank
Median	52.3	65.1	0.1450	Wilcoxon*
mean (s.e.)	69.64 (3.57)	74.81 (3.55)	0.5719	-2log(LR)
			0.0182	Gehan-Wilcoxon**

* (Gehan)-Wilcoxon-Breslow rank sum test in SAS "lifetest" procedure
** Stratified (by site) (Gehan)-Wilcoxon-Breslow rank sum test

Table 7 lists the number and percentage of patients with aborted episodes according to disease stage at baseline. The Lidakol group had a higher percentage of patients with aborted episodes for those with erythema at baseline.

Table 7 Number (%) of patients with aborted episodes by stage at baseline visit- ITT Population- 96-LID-06

Parameter	Lidakol	Placebo	p-value*
Patients with prodrome at baseline, N	40	50	0.559
Patients with aborted episodes, N (%)	22 (55%)	24 (48%)	
Patients with erythema at baseline, N	143	133	0.048
Patients with aborted episodes, N (%)	49 (34.3%)	31 (23.3%)	
Patients with prodrome or erythema at baseline, N	183	183	0.078
Patients with aborted episodes, N (%)	71 (38.8%)	55 (30.1%)	

*p values were from the Mantel-Haenszel chi-square test.

The difference in time to episode abortion for the two treatment groups was not statistically significant by the (non-stratified) Wilcoxon test, or the stratified (by site) Gehan-Wilcoxon test ($p>0.45$), as shown in Table 8.

Table 8 Time to episode abortion- ITT Population- 96-LID-06

	Lidakol (N=74)	Placebo (N=58)	p-value	Test
Number with aborted episodes Number censored	71 3	55 3		
Hours to episode abortion				
Median	54.6	51.6	0.0332	log-rank
mean (s.e.)	68.2 (5.25)	55.5 (2.88)	0.4633	Wilcoxon*
			0.2256	-2log(LR)
			0.5660	Gehan-Wilcoxon**

* (Gehan)-Wilcoxon-Breslow rank sum test in SAS "lifetest" procedure
** Stratified (by site) (Gehan)-Wilcoxon-Breslow rank sum test

Reviewer's Note: The stratification-by-site Gehan-Wilcoxon rank sum test showed that the median time to complete healing for the Lidakol treatment group was statistically shorter than that for the placebo group. However, the difference between the p-value from the stratified analysis (0.0235) and those from the non-stratified tests (log-rank: 0.4607, Gehan-Wilcoxon: 0.2416, -2log(LR): 0.5105) puts a question mark on the real implication of the p-value from the stratified analysis. An investigation into each site was inconclusive. The discrepancies of the p-values from the stratified (by site) Gehan-Wilcoxan and the non-stratified Gehan-Wilcoxan tests (Tables 3, 6) raises the question whether the stratified (by site) Gehan-Wilcoxan test is more powerful than the non-stratified Gehan-Wilcoxan test, or does it inflates the type I error?

Study 96-LID-07

Table 9 lists the patient deposition of Study 96-LID-07.

Table 9 Patient disposition- 96-LID-07

Population	LIDAKOL	PLACEBO	TOTAL
Randomization/Safety Evaluable	188	185	373
ITT	187	184	371
Efficacy Evaluable	184	177	361

In Table 10, the results of the analysis of the primary efficacy variable, hours to complete healing, are shown. None of the tests showed a statistically significant difference in treatment effect between the Lidakol group and the placebo group.

Table 10 Primary Efficacy Result - hours to complete healing
ITT Population- 96-LID-07

	Lidakol (N=187)	Placebo (N=184)	p-value	Test
Number (%) healed within 10 days	165(88%)	162 (88%)		
Number (%) Censored Number (%) Discontinued early Number (%) Not healed by day 10	22(12%) 2 (1%) 20 (11%)	22(12%) 8(4%) 14(8%)		
Hours to complete healing 25th percentile 50th percentile (median) 75th percentile Mean time to heal	 60.5 102.3 166.8 114.8	 68.5 118.2 189.0 125.1		
			0.3347	log-rank
			0.1927	Wilcoxon*
			0.5480	-2log(LR)
			0.1529	Gehan-Wilcoxon**

* (Gehan)-Wilcoxon-Breslow rank sum test in SAS "lifetest" procedure
** Stratified (by site) (Gehan)-Wilcoxon-Breslow rank sum test

In Table 11, the results of the analysis of time (hours) to complete healing for patients with classical episodes, are shown. When the analysis is not stratified by investigational sites (study centers), the "lifetest" procedure in SAS showed no significant difference in the distribution between the Lidakol group and the placebo group(log-rank test: 0.4844; (Gehan)-Wilcoxon-(Breslow): 0.1923; -2log (LR): 0.7914). With the stratified (by investigational sites (study centers)) analysis, using the "lifetest" procedure in SAS (which does not provide the stratification by site analysis), combined with a SAS macro provided by the sponsor (the stratification by site analysis can be done by using the combination of this macro and the "lifetest" procedure), the Gehan-Wilcoxon-Breslow (stratified by sites) rank sum test showed a statistically significant p-value of 0.0226.

Table 11 Classical Oral-Facial Herpes Simplex Episodes-ITT Population-96-LID-07

	Lidakol (N=111)	Placebo (N=114)	p-value	Test
Number (%) healed within 10 days	89(80.2%)	92 (80.7%)		
N (%) censored	22 (19.8%)	22(19.3%)		
Hours to complete healing 25th percentile 50th percentile (median) 75th percentile mean	 100.2 143.0 212.3 148.3	 116.3 165.0 214.9 160.2		
			0.4844	log-rank
			0.1923	Wilcoxon*
			0.7914	-2log(LR)
			0.0226	Gehan-Wilcoxon**

* (Gehan)-Wilcoxon-Breslow rank sum test in SAS "lifetest" procedure
** Stratified (by site) (Gehan)-Wilcoxon-Breslow rank sum test

In Table 12, the results of the analyses of time to cessation of discrete classical lesion stages are shown. None of the tests on hours to cessation of vesicle stage, hours to cessation of hard crust stage was statistically significant (p>0.13). Both the non-stratified and the stratified (by site) Gehan-Wilcoxon tests showed statistically significant difference in the hours to cessation of ulcer/soft crust stage (p=0.0244, 0.0066.)

Table 12 Time to cessation of Discrete classical lesion stages
ITT Population- 96-LID-07

	Lidakol	Placebo	p-value	Test
Hours to cessation of vesicle stage				
Number evaluated/total N	79/79	75/78	0.1383	log-rank
Median	50.9	53.5	0.3903	Wilcoxon*
mean (s.e.)	60.6 (3.57)	69.8 (5.1)	0.2858	-2log(LR)
			0.3318	Gehan-Wilcoxon**
Hours to cessation of ulcer/soft crust stage				
Number evaluated/total N	89/92	83/89	0.0688	log-rank
Median	92.7	100.8	0.0244	Wilcoxon*
mean (s.e.)	103.6 (5.37)	120 (5.58)	0.2385	-2log(LR)
			0.0066	Gehan-Wilcoxon**
Hours to cessation of hard crust stage				
Number evaluated/total N	72/87	82/96	0.8923	log-rank
Median	146	145.3	0.6091	Wilcoxon*
mean (s.e.)	152.4 (5.9)	158.5 (5.74)	0.9768	-2log(LR)
			0.2226	Gehan-Wilcoxon**

* (Gehan)-Wilcoxon-Breslow rank sum test in SAS "lifetest" procedure
** Stratified (by site) (Gehan)-Wilcoxon-Breslow rank sum test

Time to reduction/cessation of signs and symptoms was analyzed and the results are shown in Table 13. None of the tests on hours to first reduction of pain score, hours to cessation of pain, was significant (p>0.3). Both the non-stratified Wilcoxon and the stratified (by site) Gehan-Wilcoxon tests on hours to cessation of burning/itching/tingling, hours to cessation of pain and/or burning/itching/tingling showed the differences were statistically significant (p<0.05).

Table 13 Time to reduction/cessation of signs and symptoms- ITT Population- 96-LID-07

	Lidakol	Placebo	p-value	Test
Hours to first reduction of pain score				
Number evaluated/total N	125/125	125/128	0.5011	log-rank
Median	22.3	24	0.6078	Wilcoxon*
mean (s.e.)	27.58 (1.90)	29.87 (2.38)	0.3076	-2log(LR)
			0.4461	Gehan-Wilcoxon**
Hours to cessation of pain				
Number evaluated/total N	123/125	123/128	0.4123	log-rank
Median	46.2	45.5	0.8524	Wilcoxon*
mean (s.e.)	57.84 (3.95)	62.43 (4.75)	0.4622	-2log(LR)
			0.6746	Gehan-Wilcoxon**
Hours to cessation of burning/itching/tingling				
Number evaluated/total N	178/181	174/181	0.0219	log-rank
Median	46.8	64.3	0.0025	Wilcoxon*
mean (s.e.)	60.65 (3.80)	75.58 (4.14)	0.0484	-2log(LR)
			0.0054	Gehan-Wilcoxon**
Hours to cessation of pain and/or burning/itching/tingling				
Number evaluated/total N	177/187	173/183	0.0639	log-rank
Median	52.9	65.8	0.0255	Wilcoxon*
mean (s.e.)	67.65 (3.88)	79.86 (4.26)	0.1151	-2log(LR)
			0.0312	Gehan-Wilcoxon**

* (Gehan)-Wilcoxon-Breslow rank sum test in SAS "lifetest" procedure

** Stratified (by site) (Gehan)-Wilcoxon-Breslow rank sum test

Table 14 lists the number and percentage of patients with aborted episodes according to disease stage at baseline. The differences were not statistically significant.

Table 14 Number (%) of patients with aborted episodes by stage at baseline visit- ITT Population- 96-LID-07

Parameter	Lidakol	Placebo	p-value*
Patients with prodrome at baseline, N	31	30	0.595
Patients with aborted episodes, N (%)	23 (74.2%)	18 (60%)	
Patients with erythema at baseline, N	156	154	0.895
Patients with aborted episodes, N (%)	53 (34%)	52 (33.8%)	
Patients with prodrome or erythema at baseline, N	187	184	0.602
Patients with aborted episodes, N (%)	76 (40.6%)	70 (38%)	

*p values were from the Mantel-Haenszel chi-square test.

The difference in time to episode abortion for the two treatment groups was not statistically significant by the (non-stratified) Wilcoxon test and stratified (by site) Gehan-Wilcoxon test ($p>0.78$), as shown in Table 15.

Table 15 Time to episode abortion- ITT Population- 96-LID-07

	Lidakol (N=78)	Placebo (N=72)	p-value	Test
Number with aborted episodes	76	70		
Number censored	2	2		
Hours to episode abortion				
Median	59.5	66.5	0.9243	log-rank
mean (s.e.)	67.7 (4.02)	70 (4.32)	0.7886	Wilcoxon*
			0.9595	-2log(LR)
			0.8811	Gehan-Wilcoxon**

* (Gehan)-Wilcoxon-Breslow rank sum test in SAS "lifetest" procedure
** Stratified (by site) (Gehan)-Wilcoxon-Breslow rank sum test

Reviewer's Note: This trial did not demonstrate a statistically significant difference in the median time to complete healing between the Lidakol group and the placebo group.

Integrated safety:

Table 16 lists the frequencies of reported adverse events by body system and treatment groups. The frequencies of the adverse events in the Lidakol and placebo groups were not statistically significantly different (p>0.08).

Table 16 Number of Subjects Reporting All-Causalities Adverse Events (all studies)

	Treatment group				
	Lidakol (N=373)		Placebo (N=370)		p-value *
Body system	N	%	N	%	
Body as Whole	20	5.4%	21	5.7%	0.852
Cardiovascular System	0	0%	2	0.5%	0.155
Digestive System	4	1.1%	3	0.8%	0.712
Hemic and Lymphatic System	1	0.3%	2	0.5%	0.558
Metabolic and Nutritional	0	0%	1	0.3%	0.315
Musculoskeletal System	0	0%	2	0.5%	0.155
Nervous System	0	0%	1	0.3%	0.315
Respiratory System	4	1.1%	2	0.5%	0.418
Skin and Appendages	6	1.6%	7	1.9%	0.769
Sensory System	3	0.8%	0	0%	0.084
Urinogenital System	3	0.8%	2	0.5%	0.660

*p values were from the Mantel-Haenszel chi-square test

Table 17 lists the frequencies of reported treatment-related adverse events by body system and treatment groups. The differences in the reporting of the adverse events were not statistically significant (p>0.3) for the treatment groups.

Table 17 Number of Subjects Reporting Treatment-Related Adverse Events (all studies)

	Treatment group				
	Lidakol (N=373)		Placebo (N=370)		p-value *
Body system	N	%	N	%	
Cardiovascular System	0	0%	1	0.3%	0.315
Skin and Appendages	1	0.3%	3	0.8%	0.312

* p values were from the Mantel-Haenszel chi-square test

Reviewer's comments : The differences in reported adverse events between the placebo group and the Lidakol groups were not statistically significant.

Reviewer's Summary and Conclusion (which may be conveyed to the sponsor) : The sponsor submitted two well controlled pivotal trials (Studies 96-LID-06 and 96-LID-07) to compare Lidakol and placebo in the treatment of oral/facial herpes simplex, and demonstrated that Lidakol is statistically superior to placebo in Study 96-LID-06, but failed to show statistical superiority of Lidakol over placebo in Study 96-LID-07. The analyses by this reviewer showed sharp disparity between the p-values obtained from the non-stratified (by study site) Gehan-Wilcoxan and those from the stratified (by study site) Gehan-Wilcoxan rank sum test. An investigation into each site was inconclusive . The results of the analyses raise the question whether the stratified (by site) Gehan-Wilcoxon rank sum test (the main statistical method used in the analyses of efficacy in this NDA) inflates the type I error.

/S/ 9/15/95
Ping Gao, Ph.D.
Mathematical Statistician, DOB IV

/S/ Sept 15, '98
Concur: Rajagopalan Srinivasan, Ph.D.
Team Leader, DOB IV

HFD 540
NDA 20-941
HFD-540/Dr. Wilkin
HFD-540/Dr. Walker
HFD-540/Dr. Okun
HFD-540/Mr. White
HFD-725/Dr. Huque
HFD-725/Dr. Srinivasan
HFD-725/Dr. Gao
HFD-344/Dr. Carreras
Chron.

OCT 2 9 1999

STATISTICAL REVIEW AND EVALUATION

Date: October 29,1999
SPONSOR: Avanir Pharmaceuticals
DRUG: Lidakol/*n*-docosanol 10% cream
Proposed Indication: recurrent oral-facial herpes simplex

Document Reviewed:
Statistical Addendum of March 29, 1999 for the Study 92-LID-02 "A randomized, double-blind study comparing *n*-docosanol 10% cream and placebo cream in patients with recurrent herpes labialis."

Extent of this Review: Study 92-LID-02 has two components, the first phase and the crossover phase. This review addresses only the first phase where all but 2 randomized patients were included in the evaluation of the effectiveness of lidakol as compared to vehicle (stearic acid based). The crossover phase had only 22 patients treated out of the total 65 patients randomized.

Study Design of Study 92-LID-02

This European two-center study was designed as a double blind, randomized parallel group comparative trial (first phase) with a crossover extension phase. One center was located in Netherlands and the other in Belgium. All selected patients were randomly allocated to one of the two treatments. Patients were to self-medicate from the time of the first signs or symptoms. The treatment was to continue until healing occurred for a maximum of 10 days. The primary endpoint was time-to-healing. In the crossover phase patients were treated with the opposite study medication to that allocated in the first phase of the study.

Patient Population

For the first phase, data was reported on 65 patients, 16 for the Belgium center and 49 for the Dutch center. The demographic characteristics were about balanced for the first phase with the following exceptions. The mean ages for the lidakol and placebo groups were 32.1 and 37.3 years, respectively, with p=0.04. For the Belgium center, more patients were randomized to the lidakol group than to placebo (11 patients for lidakol and 5 to placebo). For the Netherlands center, the imbalance is reversed (21 patients allocated to lidakol and 28 to placebo).

Reviewer's Comments: This imbalance in patient allocation by randomization is somewhat unusual. The submission did not include randomization details or randomization chart for evaluation of this imbalance. Therefore, the original randomization needs to be checked against the actual assignments of patients at these centers. One also needs to make sure that patients at these centers were enrolled according to sequential patient numbers established by the pre-established

randomization, and that, the study centers were blinded to the pre-established randomization chart when they were enrolling patients.

Patient Disposition and Episodes

Table A (attached) gives a summary of patient disposition and episodes. As seen in this table, 2 patients were excluded as dropouts to make the evaluable data set for the first phase of 63 patients. Of these 63 patients in the first phase, 55 patients treated one episode and 8 patients treated 2 episodes. Thus, a total of 71 episodes were treated in the first phase.

A total of 22 patients were treated in the crossover phase, 13 by lidakol and 14 by placebo. Of these 22 patients, 17 treated one episode and 5 two episodes. Thus a total of 27 episodes were treated in the crossover phase.

Analysis of the Primary Endpoint (Original Results)

The primary endpoint for efficacy was defined as the time between initiation of treatment and occurrence of complete healing and is called *healing time*. It was supposed that early initiation of treatment, i.e., prodrome or erythema stage, would often end in a markedly reduced healing time (so called *abortion*). For this reason each treatment was classified as either 'early' or 'late', depending on the information about the onset of medication as provided by the patient.

All patients were instructed to start the treatment as soon as they became aware of the first signs of a recurrence. It was found, however, that only a small number of patient-treatments (20%) were classified as early. This indicated that a substantial proportion of patients either did not experience the prodrome/erythema stage, or this was of a very short duration, resulting in a papule or vesicle as the first clear signs of a new episode.

The results of treatment of the first phase using only first episode data were as in Table 1 (below)

Table 1 – Primary Endpoint Results (Days) for the First Phase

	Lidakol			Vehicle		
Treatment	Mean	SD	n	mean	SD	n
Early	2.5	2.4	10	6.8	4.2	4
Late	6.8	3.2	21	7.3	2.7	28
All Treatments	5.4	3.6	31	7.3	2.8	32

Comparisons of the means in the above table were done by the analysis of variance (ANOVA) for a hierarchical system of contrasts. This gave the following results.

	p-value
Overall - Between treatment means	0.0003
Contrast: vehicle early vs. late	0.72
Contrast: lidakol late vs. vehicle (early + late)	0.82
Contrast lidakol early vs. the rest (vehicle early +vehicle late + lidakol late)	0.0001

The overall mean reduction of healing time for the last comparison was 4.6 days with 95% confidence interval of (2.6, 6.6 days)

The proportion of early treatments with an abortive course of the episode were

Lidakol	8/10	with 95% confidence interval 44 – 97%
Vehicle	1/4	with 95% confidence interval 1 - 81%

Table B (attached) shows [redacted] analysis results by center for the first phase. The Netherlands center, which had majority of the total patients enrolled, showed results similar to those for Table 1 (above).

New Analyses of the Primary Endpoint (By R.A. Thisted, Ph.D.)

The original statistical analysis by [redacted] compared early docosanol (lidakol) treatment to all other treatment moderates. Dr. Thisted argued that a more appropriate approach using ANOVA would be to compare only early stage-treatments (n=14) with docosanol (n=10) and vehicle (n=4) to one another. When this was done for early patients only, the comparison of healing times (docosanol vs. vehicle) gave a p-value of p=.034 in favor of docosanol. Dr. Thisted made this comparison also using the Generalized Wilcoxon test and came out with a p-value of 0.014 in favor of docosanol by 46 hours (1.9 days). However, when data was summarized as healed within 3-days and healed in more than 3-days, then the head to head comparison for early stage-treatments docosanol (n=10) vs. vehicle (n=4) gave p=0.095 by the Chi-Square test.

Dr. Thisted also proved head to head comparison for docosanol (n=31) vs. vehicle (n=32) considering all 63 patients of the total 65 randomized on using Generalized Wilcoxon test. He showed that patients treated with docosanol had a median time to healing of 5.44 days and vehicle 7.25 days. The difference of 43 hours (1.8 days) was significant in favor of docosanol with p=0.0012. The following is the summary of results provided by Dr. Thisted. Details of his analyses are in attachments marked pages 7 and 8.

Early stage treatments
docosanol (n=10) vs. vehicle (n-4), gain of 46 hours in favor of

ANOVA	p=0.034
Generalized Wilcoxon	p=0.014
Chi-Square	p=0.095

Overall regardless of stage early or late

Generalized Wilcoxon	p=0.0012

Reviewer's Comments:

- [redacted] *comparison [Contrast: lidakol early vs. the rest (vehicle early +vehicle late + lidakol late) with p= 0.0001] mixes the control group with vehicle and active treatment. This analysis is likely to be biased given the small sample size and mixing of the randomization groups.*

- *Dr. Thisted had tried appropriate conventional comparisons. However, his results for the early-stage is based on a small number of patients only (n=10 patients for the lidakol and n=4 for the vehicle). It is difficult for this reviewer to be confident with these results for this subgroup based on such small sample sizes for the active and control groups, and given the fact that there is imbalance in sample sizes between the two groups. In addition, under multiplicity of tests done by both [redacted] and Dr. Thisted (at least 6 multiple tests were done) the p-values provided by Dr. Thisted for the early stage subgroup in favor of lidakol are not convincing.*

- *The overall comparison of lidakol (n=31) vs. vehicle (n=32), regardless of stage, early or late, using Generalized Wilcoxon test with p=0.0012, is convincing and statistically appropriate and stands the adjustment for multiplicity of tests. However, this conclusion assumes that the quality of data including conduct of the trial with respect to randomization and blinding were appropriate.*

- *For the Belgium center, more patients were randomized to the lidakol group than to placebo (11 patients allocated to lidakol and 5 to placebo). For the Netherlands center, the imbalance is reversed (21 patients allocated to lidakol and 28 to placebo). This imbalance in patient allocation by randomization is somewhat unusual and therefore, the quality of the data at these centers needs to be audited.*

Overall Conclusion for Study 92-LID-02

Assuming that the quality of data, randomization and blinding were appropriate, the overall comparison of lidakol (n=31) vs. vehicle (n=32), regardless of stage early or late, using Generalized Wilcoxon test with p=0.0012, is convincing in favor of lidakol.

/S/

R. Srinivasan, Ph.D. 10/29/99

M. Huque, Ph. D. /S/ 10/29/99

TABLE A

APPENDIX 2

OVERVIEW OF NUMBERS OF PATIENTS AND EPISODES

Treatment:	Lidakol		Vehicle		
Centre:	Belg	Neth	Belg	Neth	TOTAL
Number of patients	11	21	5	28	65
Drop-out	1	-	-	1	2
Evaluable patients	10	21	5	27	63
First treatment phase:					
Patients with 1 episode treated	9	18	5	23	55
Patients with 2 episodes treated	1	3	0	4	8
Total number of episodes treated	11	24	5	31	71
Cross-over phase:					
Number of patients	0	11	1	10	22
Patients with 1 episode treated	-	9	1	7	17
Patients with 2 episodes treated	-	2	0	3	5
Total number of episodes treated	0	13	1	13	27
Overall number of episodes treated	**11**	**37**	**6**	**44**	98

TABLE B

APPENDIX 5-1

COMPARISON OF HEALING DAYS BY CENTRE

Healing days of first episodes

	Belgium						Netherlands					
	Lidakol			Vehicle			Lidakol			Vehicle		
	mean	SD	(n)	mean	SD	(n)	mean	SD	(n)	mean	SD	(n)
Early treatment	6.7	1.8	(2)	8.2	-	(1)	1.4	0.7	(8)	6.3	5.0	(3)
Late treatment	8.6	2.4	(8)	7.1	3.4	(4)	5.7	3.2	(13)	7.4	2.7	(24)
All treatments	8.2	2.3	(10)	7.3	3.0	(5)	4.1	3.3	(21)	7.3	2.9	(27)

Results of analysis of variance (GLM):

	Belgium	Netherlands
Treatment groups	p = 0.72 NS	p = 0.0001 significant
Contrast vehicle early vs late	p = 0.57 NS	p = 0.53 NS
Contrast Lidakol late vs vehicle	p = 0.60 NS	p = 0.34 NS
Contrast Lidakol early vs rest	p = 0.57 NS	p = 0.0001 significant

Statistical Report 92-LID-02
27 September 1993

2. All Treatment Results

With 63 patients participating in the study, a total of 98 episodes were evaluated. Using ANOVA, no significant difference was revealed between late docosanol treatment and early plus late placebo treatment. However, early treatment with docosanol resulted in shorter healing times compared with late docosanol combined with early and late placebo treatment ($p = 0.0002$). These data are summarized in Table 3.

Table 3. Time-to-healing: All Treatments (Days)

	Docosanol Cream			Placebo Cream		
	Mean	SD	(n)	Mean	SD	(n)
Early Treatment	3.4	3.0	(13)	6.7	3.9	(7)
Late Treatment	6.5	2.7	(35)	7.4	2.7	(43)
All Treatments	5.7	3.1	(48)	7.3	2.9	(50)

V. EFFICACY RESULTS WITH ADDITIONAL ANALYSES (R. A. Thisted, Ph.D.)

A. Revised ANOVA and Chi-Square Analyses

The original statistical analysis by [redacted] compared early docosanol treatment to all other treatment modalities (including late-stage docosanol treatment). A more appropriate approach using ANOVA would be to compare only the early-stage treatments with docosanol and placebo to one another. When this was done for early treatment patients only, it was demonstrated that docosanol treated patients had shorter healing times than placebo-treated patients ($p = 0.034$).

Although the [redacted] statistical report stated that the number of patients healing within three days for early treatment was, "too low to make a relevant statistical comparison," such a comparison is, in fact, straightforward. This difference does not reach customary levels of statistical significance (Fisher's 1-sided exact $p = 0.095$). The chi-square data are presented in Table 4.

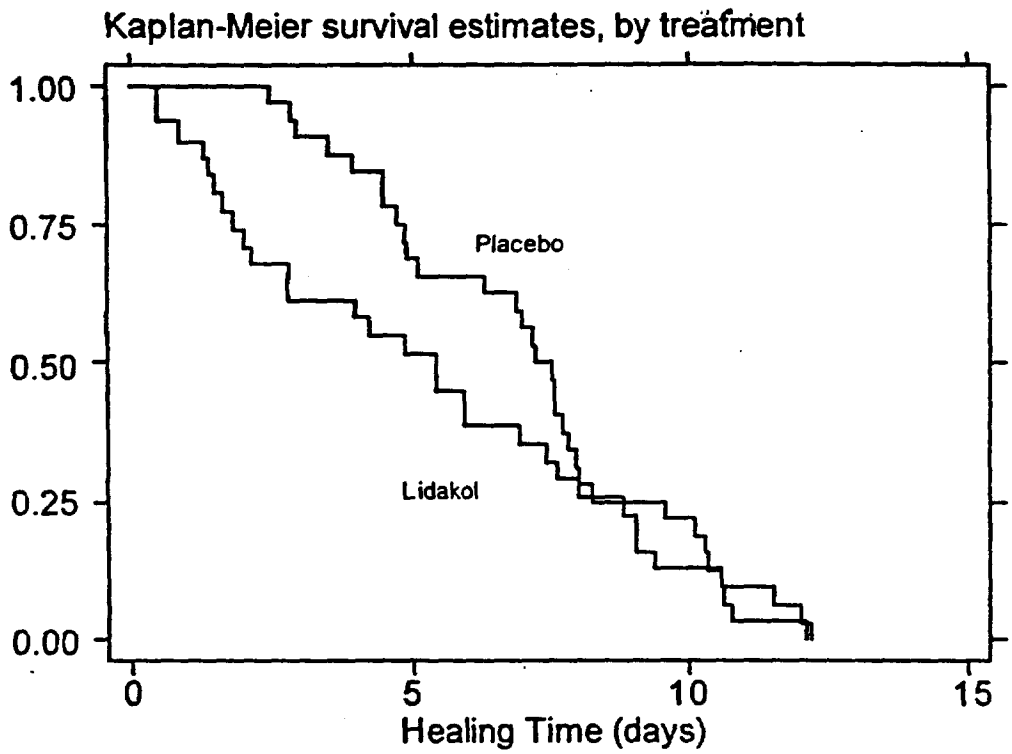
Table 4. Chi-Square Data for Early Treatment (Number of Patients)

	Healed in ≤ 3 days	Healed in > 3 days	Total
Docosanol	8	2	10
Placebo	1	3	4
Total	9	5	14

B. Survival Analysis

Using all of the study subjects (n = 63) from 92-LID-02, regardless of stage at first application, patients treated with docosanol had a median time-to-healing of 5.44 days, 43 hours (1.8 days) shorter than that of patients treated with placebo (7.25 days). This difference in distribution of healing times between docosanol and placebo groups was statistically significant (p = 0.0012) using the Gehan-Wilcoxon test, stratified by study center. The Kaplan-Meier curve for healing times in the two groups is displayed in Figure 1 below.

Figure 1



If only subjects who began treatment in the early stage of the disease (prodrome or erythema) are considered (n = 14), the difference in median time to healing favors docosanol-treated patients by 46 hours (1.9 days). Even though this comparison is based on a small number of patients, the difference in healing time distributions is statistically significant (p = 0.014).

A total of 22 patients participated in the cross-over extension of this study (these patients were treated through two episodes — the first on the randomized treatment; the second on the alternative). For these 22 intra-patient comparisons, the time-to-healing while on docosanol was reduced by a median of 3.4 days compared with their time-to- healing on placebo (p = 0.0025; Wilcoxon signed rank test). Altogether 18 of the 22 subjects had reduced healing times on docosanol.

Statistical Review Addendum for NDA 20-941
(2/23/2000)

NDA: 20-941 /1S
Applicant: Lidak Pharmaceuticals (Avanir)
Name of Drug: Lidakol Top Cream 10%
Route of Administration: Topical
Documents Reviewed: NDA 20-941 additional information (Dated Aug. 3, 1999)
Indication: Treatment of oral/facial herpes simplex
Medical Officer: Martin Okun, M.D.(HFD-540)

Background

Sponsor's February 26, 1999, and June 26, 1999 documents claimed that the baseline covariates stage ("prodrome" or "erythema"), history (historical mean episode duration >5 days or ≤ 5days) have impact in reducing p-values in the proportional odds regression, the Cox proportional hazards regression, log-logistic regression analyses. The following Table 1 summarizes these p-values. In the original NDA submission, before the NA letter to the sponsor, the main analysis for the efficacy claim was by the Generalized Wilcoxon (GW) method unadjusted for the covariates. This analysis gave a p-value of 0.023 in favor of lidakol for Study #06 and a p-value of 0.1529 for Study #07.

Table 1: P-values in favor of lidakol for the lidakol vs. placebo comparison for the time-to-healing endpoint (Information extracted from the Sponsor's Table A (attached) of 2/26/99, 7/26/99)

Study	Covariates included	Generalized-Gehan Wilcoxon	Proportional Odds Regression (POR)	Proportional Hazard Regression (Cox-Regression) (PHR)	Log-Logistic Regression (LLR)
#06	Center	0.023	0.116	0.216	0.142
	Center, Stage	NA	0.063	0.173	0.056
	Center, Stage, History	NA	0.010	0.092 (0.054)*	0.024
#07	Center	0.1529	0.137	0.284	0.173
	Center, Stage	NA	0.036	0.131	0.149
	Center, Stage, History	NA	0.006	0.059 (0.040)*	0.125

*Sponsor's new analyses: mini Cox-regression fitted to each center and then pooled
June 25, 1999 submission

In order to interpret the above covariate-adjusted results, and to check the consistency of the results within each study and across the two studies (Studies #06 and #07), the FDA reviewers on July 21, 1999, requested some descriptive statistics (mean, median, n, variation, and the Kaplan-Meier curves) and simple analysis (GW test) at the covariate

levels: "prodrome", "erythema" for *stage;* historical mean episode duration >5 days, ≤ 5 days) for *history*. On August 3, 1999, the sponsor provided these results. The following Table 2 summarizes results in terms of medians for Study #06 and Study #07. For each study, Table 2 first shows results for historical episode duration at its two levels (≤ 5 days, > 5 days), then for stage at levels (Prodrome, Erythema), and then simultaneously for the two covariates at each of the 4 levels. The "treatment difference medians" if positive reflects numerical difference in favor of lidakol.

Table 2: Descriptive Statistic Results at the covariate levels: "prodrome", "erythema" for *Baseline Stage;* historical mean episode duration >5 days, ≤ 5 days for *Baseline History*

			Treatment				Treatment difference medians	2-sided pvalue (GW test)
Study	Baseline Covariate Levels		Lidakol		Placebo			
	Historical episode duration	stage	n	median	n	median		
96-06	All patients		183	94.9	183	113.8	18.9	0.0235
	≤ 5 days		16	91.8	38	72.7	-19.1	0.136
	> 5 days		167	95.5	145	118.2	22.7	0.002
		Prodrome	40	76.5	50	74.0	-2.5	0.788
		Erythema	143	96.3	133	118.5	22.3	0.002
	≤ 5 days	Prodrome	3	51.3	11	54.6	3.3	0.356
	≤ 5 days	Erythema	13	94.9	27	94.2	-0.8	0.349
	> 5 days	Prodrome	37	93.6	39	94.5	0.9	0.631
	> 5 days	Erythema	130	100.5	106	122.3	21.8	0.007
96-07	All patients		187	102.3	184	118.2	15.9	0.1529
	≤ 5 days		31	70.0	36	101.4	31.4	0.043
	> 5 days		156	114.8	148	119.0	4.3	0.586
		Prodrome	31	49.4	30	87.6	38.2	0.072
		Erythema	156	116.7	154	122.0	5.3	0.717
	≤ 5 days	Prodrome	8	70.0	6	65.8	-4.2	0.606
	≤ 5 days	Erythema	23	69.5	30	114.6	45.1	0.019
	> 5 days	Prodrome	23	46.9	24	90.9	44.0	0.014
	> 5 days	Erythema	133	119.8	124	122.2	2.4	0.854

In the August 3 submission, the sponsor also provided Kaplan-Meier curves and descriptive statistics in terms of mean. They are omitted here. Kaplan-Meier curves had similar interpretation as those in terms of medians, and means were not as informative as medians for the time-to-healing data. The sponsor also provided pooled result for the two studies #06 and #07. As discussed in the following, the pooled results were not interpretable, and therefore, are omitted in this review.

Comments

For both studies (Study #06 and #07) larger subgroups of patients were: Baseline Historical Episode Duration > 5 days; Basleine Stage=Erythema; Baseline Historical Episode Duration > 5 days and Baseiine Stage=Erythema. Sample sizes for these subgroups (in order) are 312, 276, and 236

patients for Study #06 and 304, 310 and 257 patients for Study #07. Thus, these subgroups are of fairly large sample sizes compared to other subgroups (see Table 2).

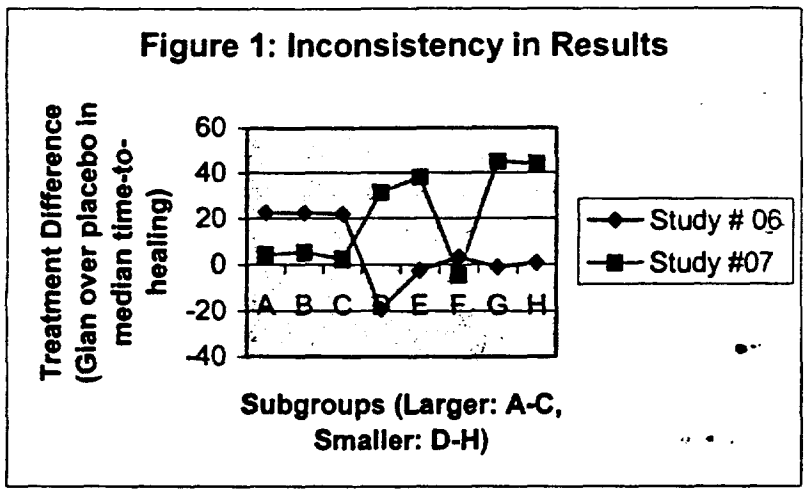
For these larger subgroups, Study #06 shows consistent results with a gain of 21 to 22 hours in median time-to-healing in favor of lidakol, p-values ranging from 0.002 to 0.007. However, for these larger subgroups, Study #07 shows practically no treatment difference, point estimates ranged from 2 to 5 hours and p-values form 0.6 to 0.9 (see Tables 3 and 4). Thus, exhibiting inconsistency in the results across the two studies (Studies #06 and #07). In addition, the results were reversed for smaller subgroups (see Figure 1 below), i.e., Study #07 generally showing large effects, and Study #06 generally showing a null effect. In the subgroup Historical Episode Duration ≤ 5 days, Study #06 shows benefit numerically in favor of placebo by 19 hours.

Table 3: Larger Vs Smaller Subgroup Results for Study #06

Study #06	Number of patients (lidakol, placebo)	Treatment Difference (median)	2-sided p (GW test)
Larger Subgroups:			
A. Hist. ED > 5 days	(167, 145)	22.5 hours	0.002
B. Erythema	(143, 133)	22.3	0.002
C. Hist. ED > 5 days, Erythema	(130, 106)	21.8	0.007
Smaller Subgroups:			
D. Hist. ED ≤ 5 days	(16, 38)	-19.1	0.136 (favoring placebo)
E. Prodrome	(40, 50)	-2.5	0.788
F. Hist. ED ≤ 5 days, Prodrome	(3, 11)	3.3	0.356
G. Hist. ED ≤ 5 days, Erythema	(13, 27)	-0.8	0.349
H. Hist. ED > 5 days, Prodrome	(37,39)	0.9	0.631

Table 4: Larger Vs Smaller Subgroup Results for Study #07

Study #07	Number of patients (lidakol, placebo)	Treatment Difference (median)	2-sided p (GW test)
Larger Subgroups:			
A. Hist. ED > 5 days	(156, 148)	4.3 hours	0.586
B. Erythema	(156, 154)	5.3	0.717
C. Hist. ED > 5 days, Erythema	(133, 124)	2.4	0.854
Smaller Subgroups:			
D. Hist. ED ≤ 5 days	(31, 36)	31.4	0.043
E. Prodrome	(31, 30)	38.2	0.072
F. Hist. ED ≤ 5 days, Prodrome	(8, 6)	-4.2	0.606
G. Hist. ED ≤ 5 days, Erythema	(23, 30)	45.1	0.019
H. Hist. ED > 5 days, Prodrome	(23,24)	44.0	0.014



Note: Points on the graph are connected for graphical convenience only

Statistical analysis of the time-to-healing data by the Cox regressions analysis method shows treatment by covariate interaction, and the nature of interaction reversed in the two studies. The coefficient for the interaction term treatment* baseline-stage is negative for Sudy #06 and positive for Study #07 (see Table 5). Thus, exhibiting inconsistency across the two studies in how the covariates explain the treatment effect.

Table 5: Treatment by Covariate Interactions

Interaction for Baseline-Stage:
Model = treatment baseline-stage treatment* baseline-stage Study #06: coefficient for interaction = -0.737, p=0.004 Study #07: coefficient for interaction = 0.626, p=0.033
Interaction for Baseline-Stage in the Presence of History:
Model = treatment baseline-stage baseline-history treatment* baseline-stage Study #06: coefficient for interaction = -0.715, p=0.005 Study #07: coefficient for interaction = 0.522, p=0.078

Conclusion

With regard to the two covariates baseline-stage (prodrome or erythema) and historical mean episode duration (>5 days or ≤ 5days), majority of patients for Studies #06 and #07 fell in the subgroups of baseline historical episode duration of >5 days, and erythema stage. For these subgroups, Study #06 showed effectiveness in favor of lidakol, but Study #07 failed to replicate these results.

ine two studies show evidence of treatment by covariate interaction as discussed above. The nature of interaction in Study #07 is reversed as compared to that for Study #06, leading to inconsistency in results. Therefore, the benefit of lidokol over placebo observed in Study #06 was not confirmed in Study #07.

/S/ 2/24/00

M. F. Huque, Ph.D.
Supervisory Mathematical Statistician
Division of Biometrics III/OB/CDER/FDA

/S/

Concur: Charles Anello,
Deputy office Director,
Office of Biostatistics, CDER, FDA

[This review contains 5 pages of text and one page of attachment]

HFD 540
NDA 20-941
HFD-540/Dr. Wilkin
HFD-105/Dr. DeLap
HFD-540/Dr. Walker
HFD-540/Dr. Okun
HFD-540/Mr. White
HFD-725/Dr. Huque
HFD-725/Dr. Al-Osh
HFD-700/Dr. O'Neil
HFD-700/Dr. Anello
HFD-725/Dr. Huque
Chron. (HFD-725)

Table A. Comparative Analyses of Time-to-Healing

		Generalized Wilcoxon		Proportional Odds Regression		Proportional Hazards Regression			Log-Logistic Regression		
Prospective Status	In Protocol: Power Profile: Level of Assumptions: Covariate Adjustment: Quality of Fit: Availability in 1996:	primary analysis yes non-parametric no not relevant yes		not identified yes semi-parametric yes best no		may be done no semi-parametric yes moderate yes			not identified yes parametric yes moderate yes		
96-06/07 Study	Terms in Model	P-value	Estimate	P-value	Estimate	P-value	Estimate	90% CI	P-value	Estimate	90% CI
	Rx, center	0.0076	17.42 hr	0.043	1.300	0.102	1.136	(0.999, 1.298)	0.044	1.100	(1.017, 1.189)
	Rx, center , stage	NA	NA	0.014	1.375	0.047	1.169	(1.026, 1.331)	0.013	1.119	(1.038, 1.207)
	Rx, center, stage, history	NA	NA	0.004	1.461	0.010	1.226	(1.076, 1.398)	0.0039	1.142	(1.058, 1.230)
96-06 Sub-study	Terms in Model	P-value	Estimate	P-value	Estimate	P-value	Estimate	90% CI	P-value	Estimate	90% CI
	Rx, center	0.023	18.83 hr	0.116	1.135	0.216	1.145	(0.956, 1.373)	0.142	1.098	(0.989, 1.221)
	Rx, center, stage	NA	NA	0.063	1.215	0.173	1.162	(0.968, 1.394)	0.056	1.128	(1.017, 1.251)
	Rx, center, stage, history	NA	NA	0.010	1.177	0.092	1.207	(1.005, 1.451)	0.024	1.156	(1.040, 1.283)
96-07 Sub-study	Terms in Model	P-value	Estimate	P-value	Estimate	P-value	Estimate	90% CI	P-value	Estimate	90% CI
	Rx, center	0.1529	15.84 hr	0.137	1.166	0.284	1.127	(0.938, 1.356)	0.173	1.100	(0.980, 1.236)
	Rx, center, stage	NA	NA	0.036	1.149	0.131	1.186	(0.985, 1.428)	0.149	1.101	(0.987, 1.123)
	Rx, center, stage, history	NA	NA	0.006	1.231	0.059	1.237	(1.027, 1.490)	0.125	1.106	(0.993, 1.231)

Efficacy endpoints are measured as the difference between LIDAKOL and placebo in median healing time for the Generalized Wilcoxon analysis: as the odds ratio for the Proportional Odds model or for the Log-Logistic model (ratios greater than 1 favor LIDAKOL), and as the hazard ratio for the Proportional Hazards model (ratios greater than 1 favor LIDAKOL). Significance tests of no efficacy are based on the permutation distribution for the Wilcoxon analysis and on the Likelihood ratio test for the other analyses. The consistency of the two sub-studies is seen in the nearly equal estimates of effect in each of the alternative analyses. Notice also that the significance of the efficacy results is strengthened by adjustment for covariates within each sub-study as well as in the overall study. [NA = not available]