

CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

ANDA 75-570

STATISTICAL REVIEW(S)

Statistical Review of ANDA
December 20, 1999

ANDA 75-570, AMMONIUM LACTATE LOTION 12%

Firm: CLAY PARK LABS, INC.
Reference Listed Drug: LAC-HYDRIN 12%, WESTWOOD-SQUIBB
PHARMACEUTICALS, INC.
Reviewed Material: Volumes 1.1 and 1.5
QMRS Receipt Date: Oct. 10, 1999
Statistical Reviewer: Yi Tsong, Ph.D.
OGD Requestor: Harvey Greenberg

I. Statistical Review of Protocol

I.1 Study design:

The study was a randomized, double-blind, placebo-controlled, comparative, 3-arm (12% Ammonium Lotion, Clay-Park Labs, Inc; Lac-Hydrin 12%, Westwood-Squibb; Vehicle, Clay-Park Labs, Inc.) parallel clinical trial. It was designed to evaluate the efficacy of the test treatment, to demonstrate the equivalence between test and reference treatments and to demonstrate the safety of the test treatment.

I.2. Efficacy variable:

The primary endpoint, the severity of Ichthyosis Vulgaris, was assessed using the 10-point Severity of Ichthyosis Vulgaris Scale (i.g., score: 1=Normal, 2=Barely perceptible, 3=Fine Scales, 4=Shallow Furrow With Fine Scales, 5=Furrow More Evident, 6=Furrow Very Evident, 7=Fissures and Furrow Present, 8=Predominate Fissures and Deep Furrows, 9=Extremely Deep Fissures With Pain). The overall score for each subject was calculated from the mean of all affected skin areas. Ichthyosis vulgaris assessment of each patient was taken at each of the 4 visits, with the first visit at pretreatment, 2nd visit at 14± 3 days, 3rd visit at 28 ±3 days and the 4th visit at 42 ±5 days. A score of 1 or less was considered to be a treatment success and any score greater than 1 was considered treatment failure. Treatment success or failure was determined at the 3rd visit, at 28 days.

The secondary endpoint, improvement of the skin area was evaluated using the 4-category Global Assessment Scale ("Normal", "Slight Improvement", "Moderate Improvement", "No Improvement or Worse").

Ichthyosis Vulgaris was also measured at Day 42 visit independently from Day 28. It was used solely for reference purpose.

Statistical Reviewer's comments:

Although Ichthyosis Vulgaris was evaluated at both Day 28 and Day 42, the Day 28 evaluation was used as primary efficacy variable.

I.3. Statistical comparisons

The statistical comparison for equivalence between test and reference treatments was done using the binary categorical scale of success or failure. The 90% Confidence Intervals of the difference of success rates between test and reference treatment had to be bounded within -20% and 20% to support equivalence. The efficacy of the active treatment versus Placebo was tested at two points, the end of treatment at Day 28 and two weeks after the treatments ended at Day 42. The primary endpoint was the success rate at day 28. The proportion of successfully treated subjects on the active treatments had to be significantly greater than that for the Placebo subjects at 0.05 level (one-sided).

Statistical Reviewer's Comments:

- a. *The efficacy of the Test and Reference treatments should be demonstrated against Placebo separately. The test of efficacy of the reference treatment is needed to verify the quality of the trial in order to assure that the equivalence test is valid.*
- b. *The equivalence comparison is valid only if the reference treatment is effective against placebo in the trial. By pooling the test and the reference groups into one active treatment, it is not possible to show both the test and the reference treatments were separately effective against Placebo.*
- c. *In general, tests for the test and the reference treatment efficacy against placebo need to be significant at 0.025 level as a one-sided test or at 0.05 level as a two-sided test. There was no special request or reasoning provided by sponsor to justify a different significance level.*

I.4 Sample Size Requirement

The sponsor calculated a sample size of 69 per group of both active drugged based on the two-sided type I error rate of 5%, 95% power, for equivalence testing with a delta of 1.5 on the 10-point scale. The comparison of active versus placebo required a sample of 22 in the placebo arm with the maximum variance, one-sided alpha of 5%, 80% power, and a delta of 1.5 on the 10 point scale. A 30% dropout rate was factored into the final sample size of 230:99 (test), 99(reference), and 32(placebo).

Statistical Reviewer's Comments:

There are four problems in the sample size calculation as proposed.

1. *As pointed out in statistical comparisons, the hierarchy of the comparisons should be demonstrating the efficacy of Test and Reference against placebo separately in order to make the equivalence testing between Test and Reference treatments valid. Hence the sample sizes calculation should be done to have enough subjects in Placebo for both Test vs. Placebo and Reference vs. Placebo.*
2. *Also pointed in statistical comparison, the efficacy test is required at $\alpha = 0.05$ for a two-sided test or $\alpha = 0.025$ for a one-sided test.*
3. *The sample size calculation was calculated for testing the difference between 2 treatments of means of a variable with a 10-point scale. However, the statistical comparison is to be done on the difference in success rates of two treatments.*
4. *It was also unclear what does the maximum variance mean in the sample size calculation for placebo comparison testing.*

In this review, the sample size requirements were re-calculated for the efficacy testing for 15% difference using chi-square test and for equivalence testing of 15% limit using two one-sided tests (Farrington and Manning's asymptotic test procedure (see Reference 1)). The required sample size would be the maximum of the two sizes. The sample size table is given below. The sample size required for the placebo group is between 100 and 170 depending on the assumed true response proportion of the placebo group. The sample size of each of the active treatment groups is between 112 and 159 depending on the true response proportion of the reference treatment group assuming that the response proportion of the reference treatment group ranges from 60% to 75%. It shows that the study was not designed with appropriate sample sizes to demonstrate efficacy of test and reference over placebo or equivalence of test and reference treatment of dichotomized outcome.

Sample sizes required for the tests of efficacy ($\alpha=5\%$ 2-sided test) and equivalence ($\alpha=5\%$ two 1-sided tests)

<i>Assumed placebo proportion</i>	<i>Efficacy test</i>	<i>Assumed Reference proportion</i>	<i>Equivalence test</i>	<i>Placebo Sample size</i>	<i>Active treatment sample size</i>
50%	170	50%		170	
55%	163	55%		163	
60%	152	60%	159	152	159
65%	138	65%	147	138	147
70%	121	70%	131	121	131
75%	100	75%	112	100	112

II. Statistical Review of Study Results

II.1 Study Populations

There were 188 subjects randomized and 164 of them completed the study. The numbers of subjects in each of the analyses are given in the following table.

Table II.1.1 Study Populations

	Test	Reference	Placebo	Total
Enrolled and randomized	85	77	26	188
Withdrawals	10	8	6	24
Completed	75	69	20	164
Treatment period intent-to-treat population (patients completed treatment period and had at least 1 valid assessment performed at pre-treatment and post-treatment)	75	69	21	165
Treatment period with the endpoint analysis population (patients received treatment)	81	73	25	179
Post-treatment efficacy analysis population (Patients had the 4 th visit)	75	69	20	164
Per protocol population (Patients complied fully with the protocol, see medical review for protocol violations)	75	65	19	159
Safety analysis with Exams/vital signs	84	75	26	185
Adverse event analysis	85	77	26	188

Statistical Reviewer's Comments:

For the per protocol population, the medical reviewer re-adjusted the population in three scenarios. In the first scenario, the population would exclude the patients who did not meet the sponsor's criteria for the definition of per protocol. It is entitled Study Termination Summary in the following table. In the second scenario, the population would exclude all patients who did not complete the study according to the protocol. In the third scenario, patients who were out of the visit window for the bioequivalence primary endpoint would be excluded. The patients excluded in each population were given in medical reviewer's report. The sample sizes of the per protocol population according to the three scenarios are given below.

Table II.1.2 Medical reviewer's per protocol populations for equivalence assessment

Per Protocol Population	Test	Reference	Placebo	Total
According to Study Termination Summary	65	61	17	143
All patients who completed the study according to protocol	59	51	17	127
All patients within the visit window for the equivalence study	62	57	17	136

II.2 Efficacy Assessment

Primary Endpoint - Using the mean score (of all areas) of 1 or less as the definition of success, the success proportions of the Test, Reference and Placebo ITT groups are given in the Table II.2.1. The ITT analysis was carried out with two versions of the ITT population. The p-value of normal approximation test for $H_0: p_{Ref} = p_{Placebo}$ vs. $H_a: p_{Ref} > p_{Placebo}$ had p-value = 0.345 for the ITT population defined for the treatment period, p-value = 0.394 for ITT population defined for the treatment period with endpoint analysis, and p-value = 0.227 for per protocol population for treatment period. The Test-Ref group were pooled for testing $H_0: p_{Ref \& Test} = p_{Placebo}$ vs. $H_a: p_{Ref \& Ref} > p_{Placebo}$. There was no statistical difference between the pooled treatment group and Placebo. The corresponding p-values using the three populations were 0.348, 0.397 and 0.236 respectively.

Table II.2.1 Proportion of success for the treatment period

Treatment	Total	Success	Proportion (%)	Efficacy Comparison p-value
Clay-Park	75	50	66.7	Test +Ref. vs. Placebo p=0.348
Lac-Hydrin	69	49	71.0	Ref vs. Placebo p=0.345
Placebo	21	13	61.9	

Proportion of success for the treatment period with endpoint analysis approach

Treatment	Total	Success	Proportion (%)	Efficacy Comparison p-value
Clay-Park	81	51	63.0	Test + Ref. vs. Placebo p=0.397
Lac-Hydrin	73	50	68.5	Ref. vs. Placebo p=0.394
Placebo	25	15	60.0	

Proportion of success for the per protocol population for the treatment period

Treatment	Total	Success	Proportion (%)	Efficacy Comparison p-value
Clay-Park	75	50	66.7	Test + Ref. vs. Placebo p=0.236
Lac-Hydrin	65	47	72.3	Ref. vs. Placebo p=0.227
Placebo	19	11	57.9	

There was no statistically significant difference found between the pooled test and reference groups and Placebo in additional post-hoc analyses using logistic regression and repeated measurement analysis. The efficacy of the test product also failed to be demonstrated in some additional ad hoc analysis using a more stringently defined success (success only if the overall mean score <1 at 28th day).

Statistical Reviewer's Comments:

The efficacy of the test product was to be demonstrated by showing superiority of Test to Placebo. The result was to be confirmed by also demonstrating that Reference was superior to Placebo in the current patient population. The superiority of pooled data of Test and Reference over Placebo (as performed by sponsor) would not support the efficacy claim for the test product. The following is the statistical reviewer's re-analysis of the efficacy claim of the test product.

The itchthyosis vulgaris score was originally an ordinal categorical variable. The mean scores of the three treatment groups were close at each visit (see the following table and figure). A simple normal approximation test for efficacy failed to show that there was any statistically significant difference between the mean scores of placebo and either of the active treatments at each visit before adjusting for multiple testing.

The mean itchthyosis vulgaris score of the three treatment groups at the 4 visits

Visit Day	Mean Score			Std Error		
	Clay-Park	Lac-Hydrin	Placebo	Clay-Park	Lac-Hydrin	Placebo
0	3.61	3.55	3.64	0.0705	0.076354	0.11963
14	1.42	1.28	1.61	0.121866	0.150721	0.258
28	0.83	0.82	0.99	0.101614	0.130017	0.196396
42	1.72	1.56	2.06	0.123553	0.134832	0.2549

When considering the dichotomized outcome at the 3rd visit as proposed in the protocol, the reviewer's analyses based on maximum likelihood based asymptotic test (2-sided 5% test) of three populations were given in the following three tables. It confirms the sponsor's results using the simple normal approximation test. Based on these results, neither the Clay-Park reference product, or the Lac-Hydrin test product was superior to the placebo treatment.

A "Page 6" was not present in the document room record of this statistical review.

A pagination error may have occurred.

IGA at 2nd visit (14th day)

Treatment (N)	Normal	Slight	Moderate	No/Worse	p-value (chi-square for equal distribution in 3 treatments)
Clay-Park (75)	32 (41%)	13 (16%)	33 (42%)	1 (1%)	p=0.35
Lac-Hydrin (69)	35 (49%)	12 (17%)	23 (32%)	1 (1%)	
Placebo (21)	8 (32%)	5 (20%)	10 (40%)	2 (8%)	

IGA at 3rd visit (28th day)

Treatment (N)	Normal	Slight	Moderate	No/Worse	p-value (chi-square for equal distribution in 3 treatments)
Clay-Park (75)	49 (65%)	7 (9%)	18 (24%)	1 (1%)	p=0.39
Lac-Hydrin (69)	48 (70%)	10 (14%)	10 (14%)	1 (1%)	
Placebo (21)	11 (52%)	2 (10%)	8 (38%)	0 (0%)	

IGA at 4th visit (42nd day)

Treatment (N)	Normal	Slight	Moderate	No/Worse	p-value (chi-square for equal distribution in 3 treatments)
Clay-Park (75)	24 (32%)	22 (29%)	24 (32%)	5 (7%)	p=0.83
Lac-Hydrin (69)	27 (39%)	18 (26%)	20 (29%)	4 (6%)	
Placebo (20)	5 (25%)	8 (40%)	5 (25%)	2 (10%)	

Statistical Reviewer's Comments:

For the IGA scores, the reviewer compared the test and reference products with the placebo separately using two sample chi-square test for distribution difference between the active treatment and the placebo and using ordinal statistics, $\Delta = \text{Prob}(\text{treatment} > \text{Placebo}) - \text{Prob}(\text{Placebo} > \text{treatment})$ (ordinal-nominal association, Delta, reference: Agresti: Ordinal Data Analysis, Wiley & Sons, 1982). This test is used to test against $H_0: \text{Prob}(\text{treatment} > \text{Placebo}) = \text{Prob}(\text{Placebo} > \text{treatment})$. The treatment (test or reference product) is effective if the null hypothesis is rejected and the estimated $\Delta < 0$. The results are summarized in the following table.

When using chi-square test to analysis the binary categorized data, the results show that there was no statistically significant difference between either of the two treatments and the placebo at any visit.

When using Delta statistic to analyze the ordinal categorical data, the results show that there was no statistically significant difference between either of the active treatments and the placebo in distribution of IGA score at any visit except visit 2. At visit 2, the 95% confidence interval of $\text{Prob}(\text{Lac-Hydrin} > \text{Placebo}) - \text{Prob}(\text{Placebo} > \text{Lac-Hydrin})$ is (0.05, 0.36). It indicates that Lac-Hydrin is more effective than placebo at visit 2. Which means that patients treated with Lac-Hydrin were more likely to receive better (i.e. higher) IGA score than those treated with the placebo based on the response at visit 2.

In summary, the efficacy of test product over placebo was not supported by either the sponsor's analysis using categorical data or the statistical reviewer's analysis using either binary categorical data or ordinal categorical data.

Two group comparison between each treatment and Placebo in IGA score

Comparison	p-value of Chi-square test	95 % CI of $\text{Prob}(\text{treatment} > \text{Placebo}) - \text{Prob}(\text{Placebo} > \text{treatment})$
visit =2		
Clay-Park vs. Placebo	0.321	(-0.02, 0.23)
Lac-Hydrin vs. Placebo	0.239	(0.05, 0.36)
visit =3		
Clay-Park vs. Placebo	0.588	(-0.01, 0.42)
Lac-Hydrin vs. Placebo	0.121	(-0.05, 0.33)
visit =4		
Clay-Park vs. Placebo	0.726	(-0.04, 0.26)
Lac-Hydrin vs. Placebo	0.497	(-0.17, 0.03)

III. Assessment of Therapeutic Equivalence

The equivalence endpoint defined in the protocol was the proportion of successfully treated subjects as assessed by the Ichthyosis Vulgaris Severity Score at Day 28. The sponsor's analysis showed equivalence of Test and Reference in ITT population and per-protocol population as shown in the following Tables.

Table III.1 Proportion of success for the treatment period

Treatment	Total	Success	Proportion (%)	Equivalence 90% CI
Clay-Park	75	50	66.7	(-0.18, 0.10) Complete Treatment period
Lac-Hydrin	69	49	71.0	

Proportion of success for the treatment period with endpoint analysis approach

Treatment	Total	Success	Proportion (%)	Equivalence 90% CI
Clay-Park	81	51	63.0	(-0.19, 0.08)
Lac-Hydrin	73	50	68.5	

Proportion of success for the per protocol population for the treatment period

Treatment	Total	Success	Proportion (%)	Equivalence 90% CI
Clay-Park	75	50	66.7	(-0.20, 0.09)
Lac-Hydrin	65	47	72.3	

Statistical Reviewer's Comments:

1. It was pointed out by Farrington and Manning (ref #1) that the decision based on regular asymptotic confidence interval may lead to significance decision different from the two 1-sided tests and it is recommended the usage of test results and test-based CI. The results of reviewer's re-analysis using two 1-sided tests are given in the following table using the same populations given by sponsor. The results are consistent with the sponsor's analysis which supports the equivalence of the test and reference products.

P-value of two 1-sided test based on maximum likelihood estimate of standard error restricted under null hypothesis

Null Hypothesis	Treatment Period ITT	Treatment Period with Endpoint Analysis	Sponsor's PP
$H_0: T - R \leq -20\%$	$p=0.019$	$p=0.027$	$p=0.029$
$H_0: T - R \geq 20\%$	$p=0.0008$	$p=0.0004$	$p=0.0006$
90% CI	(-0.17, 0.08)	(-0.18, 0.07)	(-0.18, 0.07)
<i>With Continuity Correction</i>			
$H_0: T - R \leq -20\%$	$p=0.03$	$p=0.039$	$p=0.045$
$H_0: T - R \geq 20\%$	$p=0.002$	$p=0.0008$	$p=0.001$
90% CI	(-0.18, 0.10)	(-0.19, 0.08)	(-0.196, 0.09)

2. The medical reviewer reviewed the inclusion/exclusion criteria for the per protocol population and re-classified the subjects for the per protocol population (see medical review). The results of the comparison were re-calculated with Wall's statistic adjusted with continuity correction and using Farrington and Manning's approach. The results fail to support the equivalence of the test and the reference products in all three revised per protocol population by failing to reject $H_0: T - R \leq -20\%$ using four tests with and without continuity correction. The lower limits of the 90% confidence intervals of the difference in success rate are lower than -20% in all three populations with all three equivalence tests, with and without continuity correction.

P-value of two 1-sided tests of Wall's statistic with continuity correction and with maximum likelihood estimate of standard error restricted under null hypothesis

Null Hypothesis	Day 28 PP Group	Post Treatment PP Group	Day 28 PP Group Excluding out of window subjects
<i>Proportion of success</i>			
Test	43/68	24/68	38/62
Reference	43/61	27/61	43/57
Difference	-7.26%	-8.97%	-14.15%

<i>Wald's Test with Continuity Correction</i>			
$H_0: T - R \leq -20\%$	$p=0.084$	$p=0.132$	$p=0.308$
$H_0: T - R \geq 20\%$	$p=0.001$	$p=0.0006$	$p=0.00008$
90% CI	(-0.224, 0.079)	(-0.247, 0.067)	(-0.297, 0.0137)
<i>Farrington and Manning approach</i>			
$H_0: T - R \leq -20\%$	$p=0.058$	$p=0.097$	$p=0.241$
$H_0: T - R \geq 20\%$	$p=0.0005$	$p=0.0003$	$p=0.00003$
90% CI	(-0.21, 0.06)	(-0.229, 0.049)	(-0.28, -0.0005)
<i>Farrington and Manning approach With Continuity Correction</i>			
$H_0: T - R \leq -20\%$	$p=0.084$	$p=0.132$	$p=0.308$
$H_0: T - R \geq 20\%$	$p=0.001$	$p=0.006$	$p=0.00008$
90% CI	(-0.222, 0.079)	(-0.245, 0.065)	(-0.295, 0.016)

IV. Safety Summary

Thirty-eight patients had ADE's that were determined to be definitely or probably treatment related. Of these, 21 were treated with the Clay-Park product, 13 were treated with Lac Hydrin and 4 treated with the placebo. The percentages of patients with ADE's were not statistically significant between any two treatments.

	Test	Reference	Placebo
Number	21/85	13/77	4/26
Percent	24.7%	16.9%	15.4%
P-value of Chi-square test	Test vs. Reference	0.222	
	Test vs. Placebo	0.319	
	Test vs Reference	0.895	

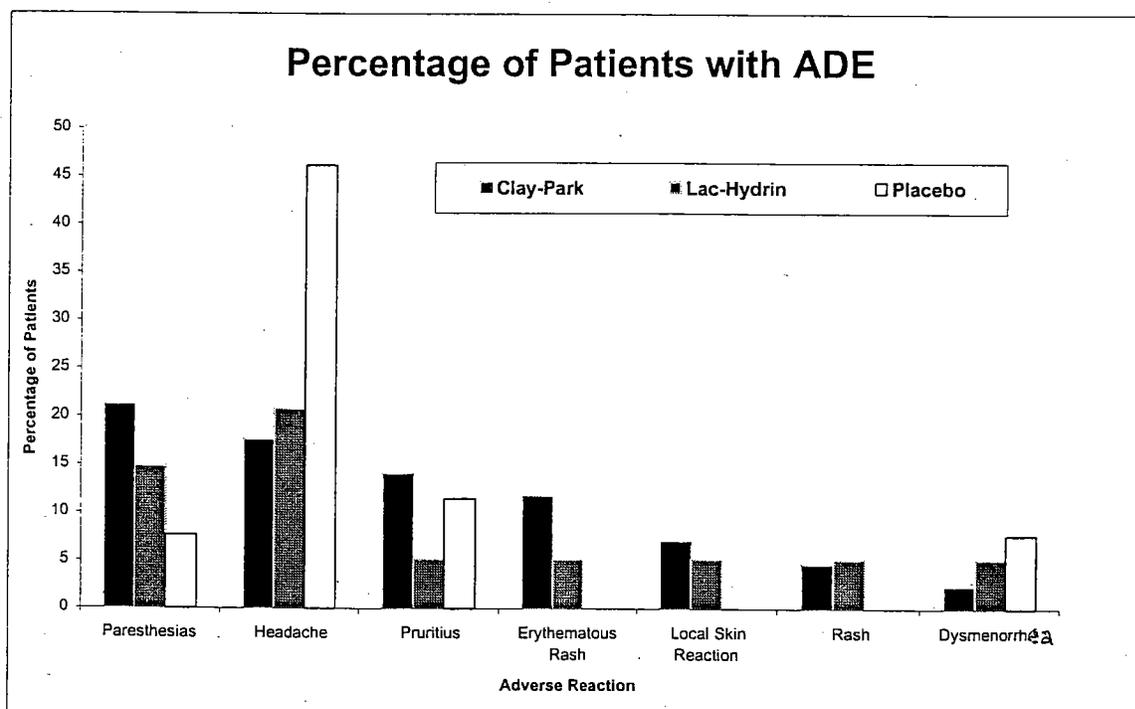
The total number of ADE's observed was 442, of which 439 were probably treatment related and occurred in 121 subjects. The were distributed as follows:

	Test	Reference	Placebo
Number of ADEs	207	167	65
# patients with ADE	56	44	21
% patients with ADE	65.9%	57.1%	80.8%

The most common ADE's in patients treated with the Clay-Park product were headache, paresthesias, pruritus, erythematous rash, skin reaction, rash, dysmenorrhea, contact dermatitis and dyspepsia. The frequencies of these ADEs were given in the following table and figure.

ADE	Test			Reference			Placebo		
	#	# Pts	%	#	# Pts	%	#	# Pts	%
Paresthesias	33	18	21.2	12	11	14.8	2	2	7.7
Headache	23	15	17.6	34	16	20.8	23	12	46.2
Pruritus	27	12	14.1	10	4	5.2	3	3	11.5
Erythematous Rash	14	10	11.8	8	4	5.2	0	0	0
Local skin reaction	10	6	7.1	4	4	5.2	0	0	0
Rash	5	4	4.7	4	4	5.2	0	0	0
Dysmenorrhea	3	2	2.4	6	4	5.2	5	2	7.7
Contact dermatitis	1	1	1.2	2	2	2.6	3	3	11.5
Dyspepsia	1	1	1.2	0	0	0	2	2	7.7

Figure 2, Profile of Percentage of Patients with ADE's in Three Treatment Groups



The majority of ADE's reported were of mild or moderate intensity. They were distributed in three treatments with similar proportions, although the placebo had slightly greater percentage of severe events.

The relationship of the ADE's with the study drug was also presented in sponsor's report. The proportions of treatment related ADE's were similar among the three treatment groups.

Statistical Reviewer's Comments:

There were no statistically significant differences in percent of patients with treatment related ADEs among the three groups. There was some difference in ADE profiles (percent of patients with the 6 most frequent ADE's) between the test and reference products. Both the test and reference groups had lower percentage of patients with headache than the placebo group. The test group had a slightly higher percentage of patients with paresthesias, pruritus and eryematous rash than the reference group.

Statistical Reviewer's Conclusion

This study failed to demonstrate the efficacy of either Clay-Park Inc.'s or Westwood-Squibb's Lac-Hydrin Ammonium Lactate lotion over Placebo in both the sponsor's analyses and the reviewer's analyses.

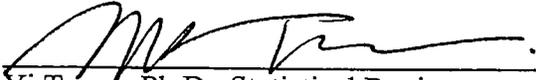
Equivalence was shown between Clay-Park Inc. and Westwood-Squibb's Lac-Hydrin lotion in the sponsor's analyses. However, in the per protocol populations defined after the medical reviewer's re-examination of the subject's status, the data failed to demonstrate equivalence in both the medical reviewer's and statistical reviewer's analyses.

Overall, the data failed to demonstrate the efficacy of the test and the reference treatments over Placebo. Without the support of treatment efficacy, any statistical support of therapeutical equivalence may not be valid. In addition, the data failed to support the therapeutical equivalence between Clay-Park Inc.'s and Westwood-Squibb's Lac-Hydrin products in the per protocol population re-determined by the medical reviewer.

Special comments:

- 1. The study was designed with sample size appropriate for both equivalence limit of 1.5 and efficacy power to detect 1.5 difference in ordinal scale of 10, between the test and the placebo treatment in severity of ichthyosis vulgaris scale. But as specified in the protocol, the analyses for both efficacy and equivalence were based on binary success rate instead, for which the sample sizes chosen were inadequate.*
- 2. The sponsor did not analyze the efficacy of the test drug versus the placebo. Instead, the sponsor provided a comparison between the pooled test and reference group and the Placebo group. This is not the correct analysis to show efficacy of the test and reference products separately.*

3. The sponsor's efficacy analysis included a few post-hoc analyses after failure of the planned efficacy analysis with the success rate defined in the protocol. The post-hoc change in definition was not acceptable without proper justification. In addition, this modified definition of success rate was not used in equivalence analysis.


Yi Tsong, Ph.D., Statistical Reviewer
Mathematical Statistician, QMR

Concur: Stella C. Machado 12/19/99
Stella Machado, Ph.D.
Director, QMR

cc:
HFD-615 Harvey Greenberg/Mary Fanning
HFD-705 Stella Machado/Yi Tsong
HFD-705 QMR Chron

O:\QMR\Chron99\ANDA 75-570.doc

Reference

1. Farrington, C. P. and Manning, G., 1990: "Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk", *Stat. in Medicine*, 9:1447-1454.

**APPEARS THIS WAY
ON ORIGINAL**

Statistical Review

ANDA 75-570

Drug Product: Ammonium Lactate Lotion, 12%

Sponsor: Clay Park Labs. Inc.

Reference Listed Drug (RLD): Lac-Hydrin® Lotion, 12%, Westwood-Squibb, NDA# 19155

Submission Date: March 29, 1999, July 14, 2003, October 15, 2003

Reviewer: Mohamed Moustapha, QMRS/OB/CDER

Requestor: Carol. Y. Kim, Pharm.D., OGD/CDER, 9/5/02

Remark: The data sets used in this analysis were supplied by the sponsor on a CD-ROM and received in the EDR on October, 15, 2003.

Objectives of the study

The objectives of the study were to demonstrate comparable safety and efficacy of Clay-Park Labs, Inc.'s Ammonium Lactate Lotion, 12% (Test product) and Lac-Hydrin® 12% (ammonium lactate) Lotion (Westwood-Squibb Pharmaceuticals Inc.; Reference product) in the treatment of subjects with moderate to severe Ichthyosis Vulgaris in order to establish bioequivalence, and to show the superiority of the active treatments over that of the Clay-Park Labs, Inc. Vehicle (Vehicle).

Study Design

This was a 3 arm parallel-Group, Double-Blind, Randomized, Placebo-Controlled, Multi-Center (25 sites) Study to Evaluate the Safety and Clinical Equivalence of a Generic Ammonium Lactate Lotion, 12% vs. Lac-Hydrin® 12% (Ammonium Lactate) Lotion in Subjects with moderate to severe Ichthyosis Vulgaris.

A total of 506 subjects were enrolled and randomly assigned to one of the three treatment groups in a ratio of 2:2:1 to Test, Reference, or Vehicle respectively.

At the enrollment visit, subjects diagnosed with moderate to severe Ichthyosis Vulgaris on both lower legs (test sites) were enrolled in the study. A moderate to severe Ichthyosis Vulgaris diagnostic was defined as a score of 6 or higher on the Overall Disease Severity Scale. Overall Disease Severity was measured on a scale of 0 to 9 (0 for normal skin and 9 for severe; see below). The Physician's Global Assessment (PGA) after the initial visit was also used to assess patient condition. A PGA score of 0 for a subject corresponded to a complete clearance and a score of 6 for a subject corresponded to one whose condition has worsened. The study was designed such that each subject had 5 visits, visit 1 (pretreatment Screening at up to 28 days prior to the first dose), visit 2 (Initial dosing at Week 0), visit 3 (Safety evaluation at Week 2), visit 4 (End of treatment at Week 4), and visit 5 (after treatment follow-up at Week 6).

Outcome Variables

Primary Endpoint:

The primary endpoint used to assess efficacy and equivalence is Clinical Success. Clinical Success was defined based on the Overall Disease Severity scale as a score of 2 or less on a scale of 0 to 9 at the end of treatment visit (Week 4).

The 10-point Overall Disease Severity (ODS) scale was defined as follows:

- 0 = Normal skin, no evidence of dryness.
- 1 = Barely perceptible scales.
- 2 = Perceptible scales with or without reticulation present.
- 3 = Mild. Fine, white adherent scales, reticulation present, skin slightly rough to touch.
- 4 = Shallow furrows with fine scales, skin rough to touch.
- 5 = Furrows more evident, more fine and larger scales.
- 6 = Moderate. Shallow furrows very evident, larger adherent scales plus occasional plaques.
- 7 = Fissures and furrows present, large scales plus plaques less than 0.5 mm in thickness.
- 8 = Predominant fissures and deep furrows, plaques 0.5 to 1.0 mm thick.
- 9 = Severe, Extremely deep fissures with pain, deep furrowed skin, inflammation and pigmented plaques greater than 1.0 mm.

Secondary Endpoints:

Per the OGD Medical reviewer's comments the following were considered secondary endpoints:

- The dichotomized (success or failure) Physicians Global Assessment (PGA) at the end of treatment and at the follow-up visit. The PGA was dichotomized to success or failure, where success is defined as a PGA score of 0 or 1.
- Clinical Success at the follow-up visit (Week 6).

The Physician's Global Assessment (PGA) scale was defined as follow:

- 0 = Completely cleared.
- 1 = Almost clear. Very significant clearing (about 90%); however, a slight degree of scaling and/or fissuring may be present.
- 2 = Marked improvement. Significant improvement (about 75%); however, some evidence of disease remains.
- 3 = Moderate improvement. Intermediate between slight and marked improvement, representing about 50% improvement.
- 4 = Slight improvement. Some improvement (up to 25%); however, significant amount of disease is still present.
- 5 = No change.
- 6 = Worse.

Statistical Analysis Methods

Efficacy Analysis

For the superiority of each active treatment over the Vehicle, the Clinical Success rates at the end of treatment (Week 4) and at the follow-up visit (Week 6) were used. Tests for superiority of the Clinical Success rate of each active treatment over that of the Vehicle were conducted using two-sided Fisher's exact tests at the 5% level of significance. The primary efficacy analysis was based on the Modified Intent-to-Treat (MITT) population.

Based on the OGD Medical reviewer's comments, additional analyses based on the following parameters were conducted:

- Clinical Success rates at Week 6 in the MITT population.
- The sponsor's PGA at the end of treatment and at follow-up visits in the MITT population.

The Last Observation Carried Forward (LOCF) approach was used to impute missing data for the MITT population.

Equivalence Analysis

The standard method in OGD to test for clinical equivalence for binary outcomes is based on the 90% confidence interval. The interval was calculated using Wald's method with Yates's continuity correction. Bioequivalence was established if this 90% confidence interval for the difference in the Clinical Success rates between the Test product and Reference product at the Week 4 visit was contained within the interval [-20%, 20%]. The primary equivalence analysis was based on the per-protocol (PP) population.

The null hypothesis to be tested was defined as follow:

$H_0: p_T - p_R < -.20$ or $p_T - p_R > .20$, versus $H_A: -.20 \leq p_T - p_R \leq .20$, Where:

p_T = Clinical Success rate of the Test treatment.

p_R = Clinical Success rate of Reference treatment.

Let n_T = sample size of Test treatment.

n_R = sample size of Reference treatment.

$$se = \left(\hat{p}_T(1 - \hat{p}_T)/n_T + \hat{p}_R(1 - \hat{p}_R)/n_R \right)^{1/2}$$

The 90% confidence interval for the difference in proportions between the Test and Reference products was calculated as follows, using Yates' correction:

$$L = (\hat{p}_T - \hat{p}_R) - 1.645 se - (1/n_T + 1/n_R)/2$$

$$U = (\hat{p}_T - \hat{p}_R) + 1.645 se + (1/n_T + 1/n_R)/2$$

We reject H_0 if $L \geq -.20$ and $U \leq .20$. Rejection of the null hypothesis H_0 supports the conclusion of equivalence of the two products.

It should be noted that the sponsor conducted efficacy analyses and equivalence analyses on both the MITT and the PP populations, but that standard practice in OGD is to assess efficacy only based on the MITT population and equivalence based only on the PP population.

Statistical Analysis Results

Demographic characteristics

A total of 506 patients were enrolled in the study, 178 were male and 328 were female. Eighty percent (406) were White, 14% (69) Black, 2% (12) Hispanic, 2% (9) Asian, and 2% (10) were described as others. The mean age of patients was 52, and the mean age was comparable among the treatment groups. Table 1 describes the sponsor's reported demographic characteristics. The racial, gender, and age distributions were comparable among the treatment groups.

Table 1- Demographic characteristics for enrolled subjects

Parameter	Ammonium Lactate Lotion, 12% (N=205)	Lac-Hydrin® Lotion 12% (N=199)	Vehicle (N=102)	p-value
Gender (n, %)				
Male	70 (34%)	72 (36%)	36 (35%)	0.9
Female	135 (66%)	127 (64%)	66 (65%)	
Race (n, %)				0.339 ¹
White	170 (83%)	158 (79%)	78 (76%)	
Black	25 (12%)	29 (15%)	15 (15%)	
Hispanic	4 (2%)	4 (2%)	4 (4%)	
Asian	2 (1%)	5 (3%)	2 (2%)	
Other	4 (2%)	3 (2%)	3 (3%)	
Age (Years)				0.740 ²
Mean ± SD	52.22 ± 14.45	52.20 ± 15.64	51.07 ± 13.57	
Min - Max	18.0 - 84.8	18.3 - 90.2	20.5 - 79.3	

¹ p-value for treatment comparisons from Cochran-Mantel-Haenszel test for general association, adjusted for site.

² p-value for treatment comparisons from a two-way analysis of variance with factors of treatment and site.

For the variable race, the p-value was calculated after combining the categories: Black, Hispanic, Asian, and Other.

ODS baseline characteristics

The sponsor's analysis in Table 3.7 (Page 98 of the report) listed that all patients enrolled in the study had an ODS score of 6 or higher at baseline visit (Week 0). However, from the SAS dataset which the sponsor submitted, 34 subjects had an ODS score lower than 6 (16 in the Test

product group, 14 in the Reference product group, and 4 in the Vehicle group). Therefore these subjects did not meet the entry criteria at baseline and with the OGD Medical reviewer's concurrence, were excluded from both the MITT and PP populations (Table 2 lists these subjects per identification number and per Investigational site). Two of these subjects (Patient 54 (8), and patient 197 (7) receiving the Test and Reference products, respectively) were among subjects that the sponsor excluded from the PP population due to minor protocol violations. However the OGD's Medical reviewer recommended their inclusion in the PP population. Nevertheless, due to their lower scores (Less than 6) at baseline these two subjects were excluded from both the MITT and PP populations based on entry criteria at baseline.

The following subjects did not meet the inclusion criteria at baseline as defined on page 1 of this review and were excluded from both the MITT and PP populations.

Table 2- Patients with ODS score lower than 6 at baseline

Numbers of patients	Test (T)	Reference (R)	Vehicle (V)	Total
ENROLLED	205	199	102	506
Did not meet the minimum ODS score at baseline	4 (1), 54 (8), 59 (8), 109 (16), 110 (16), 112 (16), 202 (10), 208 (10), 255 (14), 266 (18), 277 (8), 288 (20), 402 (12), 403 (12), 530 (11), 585 (13)	2 (1), 5 (1), 60 (8), 100 (15), 111 (16), 162 (5), 197 (7), 211 (10), 220 (10), 228 (17), 267 (18), 275 (8), 290 (20), 448 (2)	56 (8), 13 (10), 279 (8), 404 (12)	
Total Excluded from ITT and PP	16	14	4	34
ITT	189	185	98	472
PP	143	148	65	356

The sponsor stated (Page 39 of the report) that the nonparametric Friedman's test was used as alternative to the analysis of variance (ANOVA) when the assumptions of ANOVA were not satisfied. Moreover they stated that the test showed no significant differences in the ODS scores across treatment groups at the enrollment visit (p-value = 0.514). However the sponsor did not provide details on how the test was implemented. Furthermore it should be noted that Friedman's test is only appropriate for a randomized complete block design, where the number of subjects in each block (in this case site, varied from 0 to 36 subjects per site) must equal the number of treatments (3 treatment groups).

Table 3-a – ODS Scores at Week 0 for Intent-to-Treat Subjects (Sponsor’s analysis)

Week 0 Overall Disease Severity Score ¹	Ammonium Lactate Lotion, 12% (N=205)	Lac-Hydrin® Lotion 12% (N=199)	Vehicle (N=102)	p-value
Mean ± SD	6.60 ± 0.78	6.58 ± 0.73	6.73 ± 0.89	0.514 ²
Median	6.0	6.0	6.3	
Min - Max	6.0 - 9.0	6.0 - 9.0	6.0 - 9.0	

¹ Scores are the average of the right leg and the left leg. ² p-value from the Friedman's Chi-Square Test.

To compare the ODS scores at baseline across treatment groups, we used pairwise comparisons (three comparisons) using the Cochran-Armitage to test for trend. In the three comparisons - Test vs. Reference, Test vs. Vehicle, and Reference vs. Vehicle - the two-sided tests yield the following p-values: 0.9384, 0.4129, and 0.3594. We performed the trend test at the 0.1 significance level with Bonferroni adjustment, that is $\alpha = 0.033$ (0.1/3.) The test is not statistically significant since all the p-values are greater than 0.033 and therefore there was no evidence of statistical difference in the ODS scores at baseline across treatment groups.

In addition to the Cochran-Armitage test, we also constructed 95% confidence intervals around the measure of association coefficients (such as: Kendall's tau-b, Stuart's tau-c, and Somers' D, Lambda, Pearson Correlation, and Spearman Correlation), all of these confidence intervals contain zero, leading to the conclusion that there was no meaningful differences in the ODS scores at baseline across treatment groups.

Table 3-b – ODS Scores at Week 0 for enrolled Subjects

Week 0 Overall Disease Severity Score ¹	Ammonium Lactate Lotion, 12% (N=205)	Lac-Hydrin® Lotion 12% (N=199)	Vehicle (N=102)	p-values
Mean ± SD	6.43 ± 0.95	6.42 ± 0.92	6.58 ± 0.99	0.9384, 0.4129, and 0.3594 ¹
Median	6.0	6.0	6.0	
Min - Max	3.0 - 9.0	3.0 - 9.0	3.0 - 9.0	

¹ p-values from the three pairwise comparisons. Each p-value should be compared to the value 0.033 in order to have an overall level of significance of $\alpha=0.10$ (Bonferroni adjustment for multiplicity: $\alpha/3 = 0.033$)

The following adjustments to the submitted datasets were recommended by the OGD's Medical reviewer and incorporated in this statistical review:

- a. The sponsor excluded 25 patients from the evaluable population analysis due to initiation of new medication during the 6 week study. Since these concurrent medication uses are not likely to alter the outcome of the study, it is not necessary to exclude them from the evaluable population. They were used mainly for the treatment of pain, seasonal allergy, high blood pressure, and infection not caused by *Ichthyosis Vulgaris*.
- b. The sponsor also excluded 20 patients from the evaluable population because they were not assessed by the same investigator. The study protocol was designed to have the same investigator perform the dermatological assessments throughout the study. However, in this case, the score that was used for a Clinical Success does not depend on change from baseline. Therefore, it is not necessary to exclude them from the evaluable population due to protocol deviations that are considered minimal. The dermatological assessments made by another dermatologist in the absence of the principal investigator are not likely to interfere with the outcome of the study.
- c. Patient #433 (20) was excluded by the sponsor due to loss to follow-up and the initiation of new medication for-muscle pain, which is not known to interfere with the study. Patient #572 (13) was excluded by the sponsor due to missing visit 3 data. Since visit 4 data are available for these two patients, they should be included in the evaluable population.

Therefore, the following patients should be included in the PP population at week 4.

Table 4 - Patients inclusion in the PP per the OGD's Medical reviewer recommendation

	Test	Reference	Vehicle
Excluded due to new medication use during 6 Week study which however are not known to affect the outcome of the study (pain killers, antibiotics, seasonal allergy drug products, blood pressure control medications)	445 (2), 323 (6), 330 (6), 276 (8), 79 (9), 580 (13), 256 (14), 114 (16), 232 (17), 131 (19), 283 (20), 433 (20), 392 (25)	27 (4), 43 (8), 47 (8), 55 (8), 104 (16), ^153 (23), 365 (24) -allergic rhinitis, ex. criteria per sponsor, 377 (24)	451 (11), 86 (15), 133 (19), 345 (22)
Assessment by different investigator (dermatologist) should not affect the outcome of the study.	223 (17), 227 (17), 230 (17), 550 (17), 373 (24), 378 (24)	174 (7), 224 (17), 226 (17), 231 (17), 234 (17), 541 (17), 551 (17), 554 (17), 128 (19),	233 (17), 237 (17), 553 (17), 372 (24)
Data at visit 4 available and not known to have other protocol violation	-	572 (13)	-

^^The sponsor claimed that patient #153 (R, site #23) used systemic corticosteroid during six Week study, but according to the case report form, this patient was not on systemic corticosteroid and received Avapro for the treatment of hypertension.

The final population to be analyzed to test for superiority of the Test and Reference products over the Vehicle as well as for the bioequivalence of the Test and Reference products consisted of 472 subjects in the MITT population and 356 patients in the PP population. Of these 472 MITT subjects, 189 subjects received the Test product, 185 subjects received the Reference

product, and 98 received the Vehicle. Of the 356 PP subjects, 143 subjects received the Test product, 148 subjects received the Reference product, and 65 received the Vehicle.

Efficacy and equivalence Analyses

We analyzed the data for efficacy and equivalence for Clinical Success rates, and PGA rates at Week 4 and Week 6. The PP was used to test for bioequivalence and the MITT population was analyzed to compare both the Test and the Reference Products to the Vehicle.

The primary efficacy endpoint is Clinical Success based on the ODS score at Week 4, where success was defined as an ODS score of 2 or less. For the secondary endpoint, the PGA was dichotomized into success or failure, where a subject was considered a success if his PGA score was 0 or 1. Table 5 and Table 6 summarize the efficacy and clinical equivalence analyses for the primary and secondary endpoints for both the PP and MITT populations.

Efficacy and equivalence analysis at the end of treatment visit (Week 4)

Our analysis for the MITT population showed that the Test and Reference products were both statistically significantly better than Vehicle for Clinical Success and for the dichotomized Physician's Global Assessment ($p \leq 0.004$; see Table 5.)

The Test and Reference products were found to be clinically equivalent for both Clinical Success (the primary endpoint) and the dichotomized PGA (secondary endpoint) in the PP population.

Table 5 - Summary of the efficacy and clinical equivalence analysis (Evaluation at Week 4)

Population	% of success (No. of success / Total No)			p-value ²		90% CI for Test vs. Ref. (%)	Is the 90% CI within (-20%, 20%)?
	Test	Reference	Vehicle	Test vs. Vehicle	Reference vs. Vehicle		
PP							
Clinical Success *	70% (100/143)	66% (98/148)	48% (31/65)			(-6.0,13.4)	YES
PGA	63% (90/143)	60% (88/148)	46% (30/65)			(-6.6,13.6)	YES
MITT							
Clinical Success	65% (122/189)	62% (115/185)	38% (37/98)	< 0.001	< 0.001		
PGA	58% (109/189)	54% (100/185)	36% (35/98)	< 0.001	0.0040		

2: The p-values were derived from the 2-sided Fisher's exact test. *: Primary endpoint for efficacy and equivalence tests.

Efficacy and equivalence analysis at the follow-up visit (Week 6)

The Test and Reference products were both statistically significantly better than Vehicle for Clinical Success and for the dichotomized PGA in the MITT population at the follow-up visit (Week 6; see Table 6).

The Test and Reference products were found to be clinically equivalent for Clinical Success and the dichotomized PGA in the PP population at the follow-up visit (Week 6).

Table 6 - Summary of the efficacy and clinical equivalence analysis (Evaluation at Week 6)

Population	% of success (No. of success / Total No)			p-value ²		90% CI for Test vs. Ref. (%)	Is the 90% CI within (-20%, 20%)?
	Test ¹	Reference ¹	Vehicle ¹	Test vs. Vehicle	Reference vs. Vehicle		
PP							
Clinical Success	42% (60/143)	41% (60/148)	17% (11/65)			(-8.8,11.6)	YES
PGA	38% (54/143)	34% (50/148)	18% (12/65)			(-5.9, 13.9)	YES
MITT							
Clinical Success	38% (72/189)	38% (71/185)	13% (13/98)	< 0.001	< 0.001		
PGA	35% (66/189)	32% (60/185)	14% (14/98)	< 0.001	< 0.001		

2: The p-values were from the 2-sided Fisher's exact test.

Comments on the Sponsor's Analyses

The sponsor compared the Clinical Success rates of the Test and Reference groups for the Per Protocol population at Week 4. The Clinical Success rates of the active treatments versus the Vehicle group in the Intent-to-Treat Population at Weeks 4 and 6 were also tabulated and are shown in Table A.2 in the Appendix. The sponsor's summary of physician's global assessment for the ITT population at Weeks 4 and 6 is shown in Table A.1 in the Appendix.

According to the sponsor's analysis, the Test and Reference products were both significantly superior to the Vehicle (p < 0.001) for the Clinical Success rate at Weeks 4 and 6. In addition based on the Clinical Success rate at Week 4 (primary endpoint); the sponsor stated that the equivalence test met the 90% CI criteria (within -.20, +.20).

However, the sponsor excluded 46 patients due to minor protocol deviations. Since they are considered minimal and not likely to impact the outcome of the study, the OGD's Medical reviewer concludes that it is not necessary to exclude them from the evaluable population.

Therefore, this reviewer included these patients in the evaluable population. Moreover, this reviewer with OGD Medical reviewer concurrence excluded 34 subjects who did not meet the Overall Disease Severity score at baseline (that is a score of 6 or higher).

Conclusion

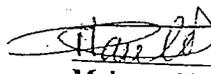
Efficacy:

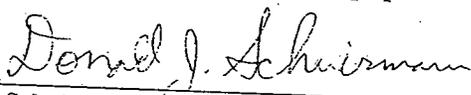
Our analysis showed that the Test and Reference products were both statistically significantly better than Vehicle for Clinical Success rate at Week 4 in the ITT population (primary endpoint).

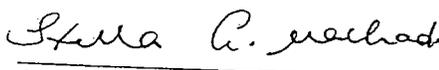
Secondary analyses based on Clinical Success at the follow-up visit (Week 6) to compare the Test and Reference products to the Vehicle also showed the superiority of both active products over the Vehicle in the ITT population.

Supportive analysis based on the dichotomized Physician's Global Assessment (Secondary endpoint) at Weeks 4 and 6 showed that the Test and Reference products were both significantly better than Vehicle for the ITT population.

Equivalence: The Test and Reference products were found to be clinically equivalent for both Clinical Success and dichotomized PGA at both visits (Weeks 4 and 6) in the PP population.


5/20/2004
Mohamed Moustapha
Mathematical Statistician, QMR


5/20/04
Donald J. Schuirmann
Expert Mathematical Statistician, QMR


5/20/04
Stella G. Machado, Ph.D.
Director, QMR

cc:
Original ANDA 75-570

HFD-600
Dena Hixon
Carol Kim
Krista Scardina

HFD-705
Stella Machado
Donald J. Schuirmann
Mohamed Moustapha
HFD-705 QMR Chron

This Review Includes 11 Pages, 05/20/2004

Appendix

Table A.1 Bioequivalence Comparisons (per sponsor)

	Ammonium Lactate Lotion, 12%	Lac-Hydrin® 12% Lotion	Vehicle	90% C.I. for Bioequivalence of Ammonium Lactate Lotion, 12% to Lac-Hydrin® 12% Lotion	p-values	
					Ammonium Lactate Lotion, 12% vs Vehicle	Lac-Hydrin® 12% Lotion vs Vehicle
Per-Protocol Subjects (n,%)						
Week 4	(N=136)	(N=141)	(N=59)			
Success	98 (72%)	101 (72%)	27 (46%)	-9.19% to 10.04% ¹		
Failure	38 (28%)	40 (28%)	32 (54%)			
Intent-to-Treat Subjects (n,%)						
Week 4	(N=205)	(N=199)	(N=102)			
Success	141 (69%)	136 (68%)	39 (38%)		<0.001 ²	<0.001 ²
Failure	64 (31%)	63 (32%)	63 (62%)			
Week 6	(N=205)	(N=199)	(N=102)			
Success	82 (40%)	85 (43%)	15 (15%)		<0.001 ²	<0.001 ²
Failure	123 (60%)	114 (57%)	87 (85%)			

¹Confidence intervals calculated using Wald's method with Yate's continuity correction.
²p-values from Z-test with Yate's continuity correction.

Table A.2 Summary of Physician's Global Assessment at Week 4 and Week 6 for the ITT Subjects

Parameter	Ammonium Lactate Lotion, 12% (N=205)	Lac-Hydrin® 12% Lotion (N=199)	Vehicle (N=102)
Week 4 Physician's Global Assessment ¹			
Mean ± SD	1.44 ± 1.22	1.51 ± 1.26	2.47 ± 1.57
Median	1.0	1.0	3.0
Min - Max	0.0 - 5.0	0.0 - 5.0	0.0 - 6.0
Week 4 Frequency Distribution (n,%) ¹			
0.0	46 (22.4%)	48 (24.1%)	12 (11.8%)
Week 6 Physician's Global Assessment ¹			
Mean ± SD	2.29 ± 1.34	2.35 ± 1.36	3.42 ± 1.54
Median	2.0	2.0	4.0
Min - Max	0.0 - 6.0	0.0 - 6.0	0.0 - 6.0
Week 6 Frequency Distribution (n,%) ¹			
0.0	12 (5.9%)	11 (5.5%)	2 (2.0%)

¹Scores are the average of the right leg and the left leg.