CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

50-819

STATISTICAL REVIEW(S)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Pharmacoepidemiology and Statistical Science
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CARCINOGENICITY STUDY

NDA Number:	50,819 / Serial 000
Drug Name:	[redacted] Gel, 1% clindamycin phosphate 2.5% benzoyl peroxide b(4)
Indication(s):	Topical treatment for acne vulgaris.
Applicant:	Dow Pharmaceutical Sciences, Inc.
Date(s):	Submitted 05/07/08 Reports submitted 12/21/07 Data submitted 02/20/08
Review Priority:	Standard
Biometrics Division:	Division 6
Statistical Reviewer:	Steve Thomson
Concurring Reviewer:	Team Leader: Karl Lin, Ph. D.
Medical Division:	Dermatological and Dental Drug Products
Toxicologist:	Reviewer: Jiaqin Yao, Ph.D. Team Leader: Barbara Hill, Ph.D.
Project Manager:	Tamika White
Keywords:	Carcinogenicity, Cox regression, Kaplan-Meier product limit, Survival analysis, Trend test

b(4)

Table of Contents

b(4)

1. EXECUTIVE SUMMARY

According to the reports provided by the Sponsor, this submission was intended to assess the carcinogenic potential of ——— ———, Gel (Benzoyl Peroxide 2.5% and Clindamycin Phosphate 1%) through the similar ——— Gel (Benzoyl Peroxide 5% and Clindamycin Phosphate 1%) when administered orally to rats (gavage) and dermally as an ointment to mice for periods of up to 24 months. The sponsor was Dow Pharmaceutical Sciences, Inc. Both studies were conducted in 1993-1995. The rat study was conducted by ' ——— ——— The mice study was conducted by ' ——— ——— The descriptions of the studies given below are taken from the corresponding Final Reports.

b(4)

1.1. Conclusions and Recommendations

The submission summarizes the results of both an oral rat study and a dermal mouse study of the carcinogenic potential of mixture Gel (Benzoyl Peroxide 5% and Clindamycin Phosphate 1%) when administered for up to two years. In both studies there were nine treatment groups per gender, numbered as groups 1-9, with doseages as follows:

Table 1. Study Dose Groups

Rats - Oral Gavage

Treatment Group	1	2	3	4	5	6	7	8	9
Dose level (ml/kg/day)	3	3	0.3	0.9	3	0.9	3	0.9	3
Benzoyl Peroxide (% conc.)	0	0	5	5	5	5	5	0	0
Clindamycin (% conc.)	0	0	1	1	1	0	0	1	1
Label	Cntrl1	Cntrl2	ClBP3	ClBP4	ClBP5	BP6	BP7	Cl8	Cl9

Mice - Dermal Application

Treatment Group	1	2	3	4	5	6	7	8	9
Dose level (cc/kg/day)	15	15	0.9	2.7	15	2.7	15	2.7	15
Benzoyl Peroxide (% conc.)	0	0	5	5	5	5	5	0	0
Clindamycin (% conc.)	0	0	1	1	1	0	0	1	1
Label	Cntrl1	Cntrl2	ClBP3	ClBP4	ClBP5	BP6	BP7	Cl8	Cl9

For statistical analyses the control treatment groups 1 and 2 were pooled, but are kept separate when reporting incidence of events.

In both genders, in rats, each of the main treatment groups, including controls, had 62 animals per group while in mice there were 60 animals per treatment group. Animals in both studies were housed singly. Tests of trend require an ordered structure on the doses. In each study note there seem to be at least three such easily observed structures, one holding Clindamycin at 0 and increasing Benzoyl Peroxide, namely pooled groups 1 & 2, and treatment groups 6 and 7, another with increasing Clinamycin and holding Benzoyl Peroxide at 0, namely

pooled groups 1 & 2, and groups 8 and 9, and finally, one where both drugs increase proportionally, namely pooled groups 1 & 2, with treatment groups 3, 4, and 5. All tests cited below use pooled control groups, and thus, for testing purposes there are 8 treatment groups with one control group.

The statistical significances of the tests of differences in survival across treatment groups are given below in Table 2. The tests for homogeneity are tests that survival is equal across treatment groups, namely the logrank and the Wilcoxon test. Note that the Wilcoxon test places more weight on earlier events than does the logrank test. The test of trend and departure from trend are Wald statistics based on proportional hazards models.

Table 2. Statistical Significances of Tests of Homogeneity and Trend in Survival

Tests over treatment groups (all with pooled controls)	Rats				Mice			
	Males		Females		Males		Females	
	LR/ Wald	Wil-coxon	LR/ Wald	Wil-coxon	LR/ Wald	Wil-coxon	LR/ Wald	Wil-coxon
Homogeneity over all groups (pooled vehicle)	0.0064	0.0723	0.7098	0.5857	0.5059	0.5474	0.6586	0.4693
Both BP & Cl (Pooled Cntrl1&2, BPCl3, BPCl4, BPCl5)								
Homogeneity	0.3968	0.4660	0.5373	0.5847	0.4966	0.4221	0.7830	0.6817
Trend over 4 groups	0.2911		0.3827		0.6445		0.3918	
Departure from trend	0.4166		0.5197		0.3359		0.8544	
Benzoyl Peroxide Only (Pooled Cntrl1&2, BP6, BP7)								
Homogeneity	0.0019	0.0281	0.2079	0.1270	0.6394	0.6515	0.2506	0.1376
Trend over 3 groups	0.0006		0.3088		0.6110		0.3514	
Departure from trend	0.8445		0.1651		0.4232		0.1666	
Clindamycin Only (Pooled Cntrl 1&2, Cl8, Cl9)								
Homogeneity	0.9282	0.8820	0.5897	0.4391	0.2823	0.3249	0.6471	0.7805
Trend over 3 groups	0.7018		0.7814		0.2312		0.7702	
Departure from trend	0.9917		0.3328		0.3140		0.3785	

Note that among both genders in mice and among female rats, either over all the treatment groups, or among the subgroups of treatments, there is no strong evidence of heterogeneity in survival (all 24 p-values for testing homogeneity are 0.1270 or larger). In male rats there is some evidence of heterogeneity over the eight treatment groups formed when pooling the controls, particularly later in the study (logrank $p = 0.0064$, Wilcoxon $p = 0.0723$). In the treatment subgroups in male rats there is no evidence of heterogeneity in survival in the clindamycin only group and the group with both clindamycin and benzoyl peroxide increasing proportionally (all four p-values ≥ 0.3980). However, in the benzoyl peroxide only group there is statistically significant evidence of heterogeneity in survival (logrank $p = 0.0019$, Wilcoxon $p = 0.0281$). The more powerful, but more restrictive, test of linear trend in dose among these benzoyl peroxide only groups is highly statistically significant (Wald $p = 0.0006$), reflecting the increasing mortality in dose apparent in Table 9 of this report or, somewhat more hidden, in figure A.1.1 in the Appendix 1. However, there is no evidence of heterogeneity among these

groups beyond the simple dose related trend (Wald p = 0.8445). Further details and plots of Kaplan-Meier estimates of survival are presented in Appendix 1.

The reports indicate complete histopathological examinations were done for all cited organs in all treatment groups in rats and most organs in mice. Testing the various neoplasms involved a large number of statistical tests, which in turn necessitates an adjustment in experiment-wise Type I error. Current FDA practice is based primarily on the Haseman-Lin-Rahman rules. These rules, originally designed for Peto tests, seem to be applicable to the poly-k tests (Lin, personal communication). In particular, to control the overall Type I error rate to roughly 10% for a standard two species, two sex study, one compares the unadjusted significance level to the appropriate bound below:

Haseman – Lin - Rahman Bounds: Comparison	Rare Tumor (Incidence ≤ 1%)	Common Tumor (Incidence > 1%)
Trend (over 3 or more groups)	0.025	0.005
Pairwise	0.05	0.01

So, for example, for a rare tumor (with incidence in the appropriate control groups < 1%, i.e., in these studies, one or no tumors), a pairwise test between the high dose group and control would be considered statistically significant if the computed significance level was at or less than 0.05.

However, it should be noted these that these multiplicity results assume the only relevant tests are single tests of trend in each gender in each of the two species, as well as comparisons between the high dose and controls in each gender in both species. When there are tests of trend among three different subgroups as here, and a large number of pairwise comparisons the overall experiment-wise Type I error can be expected to increase, likely well above the rough nominal 10% level. The amount of this inflation in Type I error is not known. (See sections 1.3.1.3 and 1.3.1.4).

Complete incidence tables of any tumor that had a trend or comparison that was statistically significant at the usual 0.05 level are presented in tables A.2.1 and A.2.2 in Appendix 2. The following table shows those comparisons in either species that satisfied the Haseman-Lin-Rahman rules for multiplicity cited above. Note that the tables below are moderately complex. In the table, for each organ, the following line displays the name of the neoplasm, followed by the incidence of that neoplasm. For each such organ-neoplasm combination, the next line provides the p-values of the various tests of trend and pair-wise comparisons to the pooled controls (groups 1 and 2). The p-values of the test of trend over the pooled controls and dose groups 3, 4, and 5 are labeled "pboth", while the p-values of the pair-wise comparisons of groups 3, 4, 5 appear under the columns "pCvs3", "pCvs4", and "pCvs5", respectively. The p-values of the test of trend over the pooled controls and the benzoyl peroxide treatment groups (with no clindamycin) is labeled "pBP", while the p-values of the corresponding pair-wise comparisons of groups 6 and 7 appear under the columns "pCvs6" and "pCvs7", respectively. The p-values of the test of trend over the pooled controls and the clindamycin treatment groups (with no benzoyl peroxide) is labeled "pBP", while the p-values of the corresponding pair-wise comparisons of groups 6 and 7 appear under the columns "pCvs8" and "pCvs9", respectively. These are summarized in the table below:

Labeling of Tests

pboth	pCvs3	pCvs4	pCvs5	pBP	pCvs6	pCvs7	pCl	pCvs8	pCvs9
Trend over Controls & BPC1 3-5	Pairwise difference between BPCl 3 & Controls	Pairwise difference between BPCl 4 & Controls	Pairwise difference between BPCl 5 & Controls	Trend over Controls, BP 6, & BP 7	Pairwise difference between BP 6 & Controls	Pairwise difference between BP 7 & Controls	Trend over Controls, C1 8, & Cl 9	Pairwise difference between Cl 8 & Controls	Pairwise difference between Cl 9 & Controls

Note that no tests of tumors in female rats satisfied the Haseman-Lin-Rahman criteria. All of the tumors cited below in male rats would be classified as rare (since incidence in controls was less than 1%), and hence are compared to the values in the middle column above. The pairwise tests comparing mesenchymal tumors in the kidney and lymphoma in the lymph node in the low dose group with both benzoyl peroxide and clindamycin (i.e. BPCl3) was statistically significantly different from the pooled controls (both $p = 0.049 < 0.05$). The pairwise tests comparing histiocytic sarcoma in each of the lung, mesenteric lymph node, mediastinal lymph node, mammary gland, skeletal muscle, and tail in the low dose of clindamycin alone (i.e., Cl8) to controls were all exactly statistically significant (all $p=0.05$). Similar results were observed in fibroadenoma in the mammary gland, papilloma on the tail, and follicular adenoma of the thyroid compared to the controls (i.e. in the tests of Cl8 versus pooled controls $p=0.05$). The pairwise test comparing systemic histiocytic sarcoma in the Cl8 cell to the pooled controls ($p=0.014 < 0.05$). The pairwise test comparing follicular adenoma of the thyroid in the low benzoyl peroxide group to the pooled controls was also statistically significant ($p=0.013 < 0.05$). Finally, the pairwise test comparing benign basal cell tumors of the skin in the low benzoyl peroxide group (i.e., BP6) to the pooled controls was also exactly statistically significant ($p = 0.05$). These do suggest that there is some difference associated with the low clindamycin dose (i.e., Cl8).

In male mice the paired comparison of the high dose in the benzoyl peroxide alone group (i.e.. BP7) to the pooled controls in liver hepatocellular adenoma was statistically significant (i.e., $p = 0.008 < 0.01$ for a common tumor). In female mice the paired comparison of the low dose of the both benzoyl peroxide and clindamycin (i.e. ClBP3) in focal hyperplasia of the pituitary was statistically significant (i.e., $p = 0.048 < 0.05$ for a rare tumor), as was the similar comparison of the low dose the both benzoyl peroxide and clindamycin in polyps of the uterus (i.e., $p = 0.005$ for a common tumor). Also in female mice the paired comparison between middle dose of the both benzoyl peroxide and clindamycin (i.e., ClBP4) and the pooled controls in terms of stromal sarcoma of the uterus was also exactly statistically significant (i.e., $p = 0.010$ for a common tumor). Note however, again, due to the large number of tests these are almost surely anticonservative, that is, the true overall error is likely to be considerably higher than the nominal roughly 10%.

Table 3. Potentially Statistically Significant Results in Mice and Male Rats

Organ/tumor	Incidence/p-values Ctrl1 Ctrl2 ClBP3 ClBP4 ClBP5 BP6 BP7 Cl8 Cl9 pboth pCvs3 pCvs4 pCvs5 pBP pCvs6 pCvs7 pCl pCvs8 pCvs9
Male Rats	
KIDNEY	
Mesenchymal Tumor	0 0 0 3 1 0 0 0 0 .139 . .049 .297 . . . . . .
LUNG (BRONCHI)	
HISTIOCYTIC SARCOMA [M]	0 0 1 0 2 1 1 2 0 .047 .316 . .094 .127 .212 .278 .497 .050 .
LYMPH NODE	
LYMPHOMA [M]	0 0 0 3 1 0 0 0 0 .139 . .049 .297 . . . . . .
LYMPH NODE (MESENTERIC)	
HISTIOCYTIC SARCOMA [M]	0 0 0 0 0 0 1 2 0 . . . . .238 . .278 .497 .050 .
LYMPH NODE - MEDIASTINAL	
HISTIOCYTIC SARCOMA [M]	0 0 0 0 0 0 1 2 1 . . . . .238 . .278 .241 .050 .366
MAMMARY GLAND	
Adenoma/Fibroadenoma/-carcinoma	0 0 1 2 1 0 0 3 0 .160 .316 .128 .297 . . . .355 .013 .
FIBROADENOMA [B]	0 0 1 1 1 0 0 2 0 .146 .316 .350 .297 . . . .497 .050 .
HISTIOCYTIC SARCOMA [M]	0 0 1 0 0 0 0 2 1 .629 .316 . . . . . .241 .050 .366
SKELETAL MUSCLE (GROSS LESION)	
HISTIOCYTIC SARCOMA [M]	0 0 0 0 0 0 0 2 1 . . . . .241 .050 .366
Skin	
BASAL CELL TUMOR, BENIGN	0 0 1 2 0 2 1 1 0 .584 .316 .128 . .146 .050 .278 .702 .212 .
Systemic	
HISTIOCYTIC SARCOMA [M]	1 0 2 0 2 1 1 4 1 .165 .242 .675 .219 .283 .374 .473 .479 .014 .592
TAIL (GROSS LESION)	
PAPILLOMA	0 0 0 0 0 0 0 2 0 . . . . .497 .050 .
THYROID	
ADENOMA, FOLLICULAR	0 0 1 2 2 3 0 2 2 .052 .316 .128 .094 .512 .013 . .105 .050 .139
Male Mice	
LIVER	
Hepatocellular Adenoma	12 6 11 11 10 15 19 9 6 .460 .342 .259 .453 .010 .069 .008 .225 .529 .301
Female Mice	
PITUITARY GLAND	
Focal Hyperplasia	0 1 4 2 1 0 1 2 0 .447 .048 .288 .567 .471 .691 .580 .424 .261 .680
UTERUS	
Polyp	5 7 0 6 4 7 5 3 5 .436 .005 .539 .310 .282 .374 .401 .504 .182 .532
Stromal Sarcoma	1 1 3 7 1 2 5 0 0 .259 .221 .010 .736 .032 .360 .053 .257 .443 .464

1.2. Brief Overview of the Studies

One mouse study and one rat study were submitted:

Study SLS 3242.36: 2 Year Oral (Gavage) Carcinogenicity Study of an Admixture Gel (Benzoyl Peroxide 5% and Clindamycin Phosphate 1%) and Its Components in Rats,

and,

Study 11484: Dermal Carcinogenicity Study of an Admixture Gel (' — - Benzoyl Peroxide 5% and Clindamycin Phosphate 1%) and Its Components in Mice, b(4)

These studies were designed to assess the potential carcinogenic effect of the mixture of Benzoyl Peroxide 5% and Clindamycin Phosphate 1% when administered orally to rats (gavage) and dermally as an ointment to mice. Both studies had nine treatment groups, including two nominally identical control groups, with a further three treatment groups with both benzoyl peroxide and clindamycin phosphate increasing proportionally as in increasing doses of 0.3, 0.9, and 3 in rats and 0.9, 2.7, and 15 in mice. Two further treatment groups in each of benzoyl peroxide and clindamycin were increasing as 0.9 and 3 in rats, and 2.7 and 15 in mice. In rats each treatment group had 62 animals, while in mice each had 60 animals. As discussed in section 2.2, a detailed examination of the rat data resulted in the discovery of a number of data of errors in the rat data.

1.3. Statistical Issues and Findings

1.3.1. Statistical Issues

In this section, several issues, typical of statistical analyses of these studies, are considered. These issues include details of the survival analyses, tests on tumorigenicity, multiplicity of tests on neoplasms, and the validity of the designs.

1.3.1.1. Treatment Groups:

In plots and summary tables results from the two control groups are presented separately. However, in all statistical analyses results from the two controls are pooled, resulting 8 different treatment groups. Both studies started with 9 treatment groups, including two nominally identical control groups with treatments as indicated below:

Table 4. Study Dose Groups

Group	1	2	3	4	5	6	7	8	9
Rats Dose level (ml/kg/day)	3	3	0.3	0.9	3	0.9	3	0.9	3
Mice Dose level (ml/kg/day)	15	15	0.9	2.7	15	2.7	15	2.7	15
Benzoyl Peroxide %	0	0	5	5	5	5	5	0	0
Clindamycin %	0	0	1	1	1	0	0	1	1
FDA Group Label	Cntrl1	Cntrl2	ClBP3	ClBP4	ClBP5	BP6	BP7	Cl8	Cl9

Tests of trend in tumorigenicity (or in survival) require an ordered structure on the doses used. In each study note there seem to be at least three such easily observed structures, each starting from the pooled controls. One structure holds clindamycin at 0 and increases benzoyl peroxide, namely pooled groups 1 & 2, and groups 8 and 9, another holds benzoyl peroxide at 0 and increases clindamycin, namely pooled groups 1 & 2, and groups 6 and 7, and finally, a group where both drugs increase proportionally, namely pooled groups 1 & 2, and groups 3, 4, and 5. In plots and summary tables results from the two control groups are presented separately.

1.3.1.2. Survival Analysis:

Both the logrank and Wilcoxon tests were used to test homogeneity of survival among the treatment groups. Tests of dose related trend and homogeneity after adjusting for trend using Cox proportional hazards models were also performed. The number of such tests raises issues of multiple testing, but from the point of view of finding differences among treatment groups (i.e., reducing the probability of Type II error), this should be acceptable. Appendix 1 reviews the animal survival analyses in some detail. The Sponsor's conclusions are summarized in Sections 3.2.1.1 and 3.2.2.1.

1.3.1.3. Tests on Neoplasms:

In the past, the usual FDA tumorigenicity analyses was based on so-called Peto tests where the analysis of fatal tumors was based on the time of death, and for observable tumors on the time of detection. Non-fatal, incidental tumors found at the time of the animal's death were analyzed by the so-called prevalence method. Then these results were pooled for a final test. However, this method depends on the attribution of whether a tumor is fatal, incidental, or observable (mortality independent). In the report of the Society of Toxicological Pathology "town hall" meeting in June 2001 the poly-k modification of the Cochran-Armitage test of trend was generally recommended over use of these Peto tests. The poly-k tests (here k=3) do not require accurate specification of whether a tumor is fatal, incidental, or mortality independent.

Tests of trend require an ordered structure on the doses. In each study note there seem to be at least three such easily observed structures, one holding Clindamycin at 0 and increasing Benzoyl Peroxide, namely pooled control groups Cntrl1 & Cntrl2, and groups BP6 and BP7, another structure with increasing Clinamycin and holding Benzoyl Peroxide at 0, namely the pooled controls, and groups Cl8 and Cl9, and finally, one where both drugs increase proportionally, namely the pooled control groups Cntrl1 & Cntrl2, and groups BPCl3, BPCl4, and BPCl5.

1.3.1.4. Multiplicity of Tests on Neoplasms:

Testing the various neoplasms involved a large number of statistical tests, which in turn necessitated an adjustment in experiment-wise Type I error. Current FDA practice is based on the Haseman-Lin-Rahman rules. Namely, based on his extensive experience with such analyses, for pairwise Peto tests comparing control to the high dose group, Haseman (1983) claimed that for a roughly 0.10 (10%) overall false positive error rate, rare tumors should be tested at a 0.05 (5%) level, and common tumors (with a historical control incidence greater than 1%) at a 0.01