CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

22-395

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 ADDENDUM

NDA Number:	22,395 / Serial 000
Drug Name:	Capcaisin Dermal Patch 8% (Qutenza™)
Indication:	Prolonged reduction of neuropathic pain with postherpetic neuralgia (PHN)
Applicant:	NeurogesX, Inc. 215 Bridgepointe Parkway, Suite 200 San Mateo, CA 94404-5067 Study conducted at: (b) (4)
Date:	Submitted October 13, 2008
Review Priority:	Standard
Biometrics Division:	Division 6
Statistical Reviewer:	Steve Thomson
Concurring Reviewer:	Team Leader: Karl Lin, Ph. D.
Medical Division:	Anethesia, Analgesia, and Rheumatology Products
Pharmacologist:	Lawrence Leshin, Ph.D. Adam Wasserman, Ph.D.
Project Manager:	Tanya Clayton
Keywords:	Carcinogenicity, Trend test

Table of Contents

1. EXECUTIVE SUMMARY

The Sponsor's report indicates that the objective of this study was to assess the carcinogenic potential of trans-Capsaicin when administered weekly via topical application to the dorsal skin of Tg.AC mice for 26 weeks. The FDA statistical analysis report for this study was completed on 10 April 2009. Later, the Sponsor sent corrected data for several data sets used in the analysis. The exact changes in the data sets provided by the Sponsor are described in Section 2.2 below. As discussed in that section, these changes required no revision of the original statistical analyses as presented in the orginal statistical analysis report. However, during the evaluation of this new data, an error in one item in the computed incidence tables in the FDA report was discovered. This affected only the reported incidence of the item, not the results of statistical tests.

1.1. Conclusions and Recommendations

This submission summarized the results of a study to assess the oncogenic potential of trans-Capsaicin when administered weekly via topical application to the dorsal skin of mice for 26 weeks. The Sponsor reported that male and female Model TGAC-T (hemizygous), FVB/NTac-Tg(v-Ha-ras)TG.ACled mice were assigned to 7 treatment groups per gender, 25 mice/sex/group. Mice in group 1 were treated with the vehicle only. Mice in groups 2-4 were described as receiving the dose formulations containing the vehicle control, diethylene glycol monoethyl ether (DGME), and the test drug (trans-Capsaicin in DGME) at drug levels levels of 0.64, 1.28, and 2.56 mg/mouse/ week, also labeled as the low, medium, and high dose groups respectively. Group 6 animals, the positive control, were administered Tetradecanoylphorbol-13-Acetate (TPA, in DGME). Group 5 animals received lidocaine only. Group 7 animals were untreated. Animals in all treatment groups were dosed once per week except for the positive control group, group 6, who were dosed twice per week.

Analyses of mortality, including tests of homogeneity and trend in survival over dose groups, were described in the orginal FDA statistical analysis report. Kaplan-Meier estimated survival curves across dose groups for each gender in each study were also displayed in Figures A.1.1-A.1.2 of Appendix 1. These curves were supported by tests of homogeneity and trend in survival over dose groups. To summarize, in the three Capcaisin treatment groups and its vehicle, none of the Wilcoxon and log rank tests of homogeneity in survival and in time to detection of first tumor groups were statistically significant (all $p \geq 0.1937$). As noted in the report, absence of proof is not proof of absence, but here the consistency of results seems to be fairly strong evidence of no differences in patterns of survival between these four treatment groups.

Tumorigenicity analysis in Tg.AC mice is traditionally based on papilloma counts, particularly at the site of application (SOA) and non site of application (NSOA). The Sponsor originally supplied data on the number of animals in the various treatment groups with any tumor, including post mortem tumors. This data were analyzed using so-called poly-k tests,

which modify the original Cochran-Armitage test of dose related trend in an event to adjust for differences in mortality. Due to a programming error, several of the reported incidences in the "Any Skin" category in the original FDA statistical analysis report were inflated. This error did not apply to the statistical tests used, only to the incidences in the summary table.

The following Table 1 displays both the incidence, including the corrected incidence in the "Any Skin" category, and those tests of trend and pairwise comparisons that were statistically significant. For each dose group, the tumor incidence was the number of animals where histopathological analysis detected a tumor. The column labeled "Trend" provides the observed p-value of the tests of trend over the vehicle controls, low, medium, and high dose groups, i.e., the Capcaisin Groups 1-4. The columns labeled "HvsV", "MvsV", and "LvsV" provide the p-values of the corresponding pairwise tests of tumor incidence in each of the High, Medium, and Low dose groups versus the vehicle group (i.e., groups 2-4 versus group 1). The columns labeled "VvsN", "LvsN", "MvsN", and "HvsN" provide the p-values of the corresponding pairwise tests of tumor incidence in each of the vehicle, low, medium, and high dose groups versus the no treatment group (i.e., groups 1-4 versus group 7).

Table 1. Potentially Statistically Significant Neoplasms Based on the Number of Animals with Tumor.

Organ	Tumor	Incidence						P-values							
		Veh	Low	Mid	Hi	Lido	TPA	None	Trend	VvsH	VvsM	VvsL	VvsN	LvsN	MvsN	HvsN
Males																
Treated Skin	B-Papilloma, Squamous Cell															
		1	2	6	3	1	21	0	.1615	.2890	.0491	.5000	.5000	.2449	.0082	.1092
Females																
Any	Papilloma															
		2	6	8	13	8	19	8	.0002	.0004	.0353	.0780	.9573	.5124	.5400	.0776
Skin/SubQ, Other	B-Papilloma, Squamous Cell															
		2	2	4	8	3	8	5	.0061	.0301	.3326	.6388	.7742	.7353	.4439	.2045
Stomach, Nongl	B-Papilloma, Squamous Cell															
		0	4	4	6	4	2	3	.0156	.0094	.0543	.0383	.8752	.4255	.4513	.1901

As reported in the original FDA analysis, in males, the pairwise comparison for the number of animals (6) with papilloma in treated skin in the medium dose group is statistically significantly higher than the no treatment group (p=0.0082) and barely higher in the vehicle treatment group (p=0.0491). In group1-group 4 females there is a clear trend in any papillomas over dose (p=0.0002), while the test of differences between vehicle and the high dose group and the medium dose group is also statistically significant (p=0.0004 and p=0.0353, respectively). In Skin/SubQ the test of trend in benign papillomas over dose is statistically significant (p=0.0061), as is the pairwise test between vehicle and the high dose group (p=0.0301). In the Stomach, Nongl the test of trend in benign papillomas over dose is also statistically significant (p=0.0156), as are the pairwise tests between vehicle and the high and low dose groups (p=0.0094 and p=0.0383, respectively). The pairwise test between

vehicle and the medium dose group is nearly statistically significant (p=0.0543). Note that the corresponding pairwise tests with the no-treatment group were not quite statistically significant at the usual 0.05 level (all $p \geq 0.0776$). This is clearly due to the large number of papillomas in the no treatment group. Particularly with the relatively small sample sizes in this study (25 animals) it seems that the evidence for a trend over several treatment groups is much stronger than the evidence in a single treatment group like the untreated group. However, there is no consistency in results across genders.

Appendix 1 of this report, presents the corrected "Any Skin" entries in Appendix 3 of the original report.

1.2. Brief Overview of the Studies

This submission consisted of one 26 week Tg.AC mouse study:

(b) (4) **26-Week Dermal Oncogenicity Study with trans-Capsaicin in Tg.AC Hemizygous Mice (FVB/N)**

1.3. Statistical Issues and Findings

Please see corresponding section in the original report.

1.3.2. Statistical Findings

Please see summary in Section 1.1 above and in the original report.

2. INTRODUCTION

2.1. Overview

Results from *trans*-Capsaicin administered weekly via topical application to the dorsal skin of Tg.AC mice for 26 weeks were submitted.

2.2. Data Sources

Eight SAS transport files, were originally provided by the Sponsor and placed in the CDER electronic data room (edr). Each of these contained a SAS data set with the same prefix but with the extension "sas7bdat."

food.xpt	mass.xpt	signs.xpt	macro.xpt
weights.xpt	micro.xpt	tumor.xpt	mortal.xpt

Only the mortal.xpt, mass.xpt, and tumor.xpt data set were used in the FDA statistical analysis. Later a papill.xpt data set was added.

Early in the summer, the Sponsor submitted new papill, tumor, and mortal data sets. The mortality data seemed to be unchanged. The major changes in the papilloma data set were

apparently in 4 of 4901 records. Specifically, 1) one animal in group 2 that should have indicated the NSOA papilloma count in the old data was not indicated as an NSOA site. However this animal had no papillomas in the study so its impact would be minimal. 2) There was a single entry for an animal that was coded as being in the wrong group (group 7 versus group 6), but that was for an organ site that was not analyzed. 3) Assuming the new data is correct one group 7 animal had data for a couple of the sub rows switched. However, only weekly, per animal, papilloma totals were analyzed so again that should have absolutely no impact. It appears that only one record in the tumor data set was changed. 4) One animal whose time of papilloma incidence was recorded as week 27 in the original data was coded as week 23 in the new data. Because the animal did not die at that time, this would have no impact on the poly-k analysis of any tumors, and thus should have no impact on test results. For these reasons, the poly-k and papilloma analyses reported in the FDA analysis of the original study were allowed to stand. Finally, 5) the new mortality data completely matched that reported in the original data.

However, during the analysis summarized above, a programming error was noted. The "Any Skin" category for papillomas was added later to the analysis, by pooling records for papilloma counts at several locations. The incidence counts reported in the tables were computed separately from the incidence counts used for the poly-k tests on tumors, and the latter do seem to be correct. In particular, the reported incidence counts in the "AnySkin" category were only corrected for the number of papillomas at each location, but not for possible presence of different locations. Again, the data used for the actual tests were corrected for this possibility so the p-values for all tests should be correct for the data as originally provided by the sponsor. This error only applied to the "Any Skin" organ category. In males, where the original FDA analysis cited incidence counts (for groups 1-7, respectively): 9, 4, 14, 6, 10, 36, 9, the corrected values are: 9, 4, 12, 4, 9, 24, 7. In females this organ category had reported incidence counts as: 2, 7, 10, 16, 8, 29, 9, which should have been 2, 6, 8, 13, 8, 19, 8. So almost all the differences are in the TPA group, group 6. Our interest focuses on groups 1-4, and group 7. To reiterate, the p-values for the actual tests did seem to use the appropriate incidence counts, and the only error appears to be in the reported incidence for the "Any Skin" category for papillomas. This applies to the entries for "Any Skin" females in the essentially identical text tables 2, 7, and appendix A.2.1, plus the "Any Skin" entry for males in A.2.2 and females in A.2.3 of the FDA statistical analysis report. Since it did not affect the p-values it should have no effect on conclusions.

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

NA

3.2. Evaluation of Safety

3.2.1. (b) (4) 26-Week Dermal Oncogenicity Study with trans-Capsaicin in Tg.AC Hemizygous Mice (FVB/N)

STUDY DURATION: 26 Weeks (Although data extend to 27 weeks)
STARTING DOSING DATE: 16 May 2005
LAST DOSING DATE: 23 November 2005
STARTING TERMINAL SACRIFICE: 23 November 2005
RAT STRAIN: Tg.AC Hemizygous Mice
ROUTE: Dermal

3.2.1.1. Sponsor's Results and Conclusions

Please see corresponding section in the original report.

The Sponsor provided a new statistical analysis report. For mortality data the Sponsor's new statistical review uses a number of pairwise tests, whereas the FDA analysis tends towards overall tests of homogeneity. One might argue that which you choose is largely a matter of taste. The Sponsor's approach is perhaps simpler, but involves many more tests. Thus, theoretically you then have many more opportunities to see a significant effect when there actually is no such effect. But, in practice, both analyses seem to lead to similar conclusions. For the papilloma counts the Sponsor's analyst uses time to first tumor event analysis, with a separate analysis of the number of tumors. The analysis performed in the FDA report simultaneously models time of incidence and papilloma counts, and is arguably superior. But again, the new Sponsor analysis and the original FDA analysis seem to be reasonably consistent. So the Sponsor's new analyses are not reviewed further.

3.2.1.2. FDA Reviewer's Results

Please see the overview in Section 1.1 above and the corresponding section in the original report.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

NA

5. SUMMARY AND CONCLUSIONS

5.1. Statistical Issues and Collective Evidence

Please see Section 1.3 in the original FDA report.

5.2. Conclusions and Recommendations

Please see section 1.1 above.

APPENDIX:

Appendix 1. Updated FDA Poly-k Tumorigenicity Analysis

Tables A.2.1 through A.2.3 in the original FDA report display the number of animals with neoplasms by organ and tumor combination, and the results of tests of trend over dose and the results of pairwise comparisons with the vehicle control (Group 1) and the no dose control (Group 7). For each dose group, the tumor incidence was supposed to be the number of animals where a tumor was detected. The column labled "Trend" provided the observed p-value of the tests of trend over the vehicle controls, low, medium, and high dose groups, i.e. Groups 1-4. The columns labeled "HvsV", "MvsV", and "LvsV" provided the p-values of the corresponding pairwise tests of tumor incidence in each of the High, Medium, and Low dose groups versus the vehicle group (i.e. groups 2-4 versus group 1). The columns labeled "VvsN", "LvsN", "MvsN", and "HvsN" provided the p-values of the corresponding pairwise tests of tumor incidence in each of the vehicle, low, medium, and high dose groups versus the no treatment vehicle group (i.e., groups 1-4 versus group 7). Incidence in the TPA group, group 6, is used to verify the sensitivity of the mice. As with the TPA group, for the lidocane group (Group 5) only the incidence of animals with tumor is reported.

Due to the error described in Section 2.2 the several of the incidences in the "Any Skin" category of Tables A.3.1-A.3.3 of the original report were inflated. The corrected incidences are included in Table A.1.1 below:

Table A.1.1 Tests of Trend and Pairwise Differences in Neoplasms in Males

Organ	Tumor													
	Incidence							P-values						
	Veh	Low	Mid	Hi	Lido	TPA	None	Trend	VvsH	VvsM	VvsL	VvsN	LvsN	MvsN	HvsN
Males:														
Any Skin						Papilloma								
	9	4	12	4	9	24	7	.7969	.8870	.2836	.9018	.3812	.7519	.0989	.7275
Females:														
Any Skin						Papilloma								
	2	6	8	13	8	19	8	.0002	.0004	.0353	.0780	.9573	.5124	.5400	.0776

Again, results and interpretation of all reported p-values remain unchanged.

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-22395	ORIG-1	NEUROGESX INC	Qutenza

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

STEVEN F THOMSON
11/13/2009

KARL K LIN
11/13/2009
Concur with review

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	22-395 / 00
Drug Name:	Qutenza ® (8% capsaicin patch)
Indication(s):	The prolonged reduction of neuropathic pain associated with postherpetic neuralgia (PHN)
Applicant:	Neurogesx
Date(s):	Letter date: 10/13/08
	PDUFA date: 8/14/09
Review Priority:	Standard
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	Kate Meaker, M.S.
Concurring Reviewers:	Dionne Price, Ph.D.
	Thomas Permutt, Ph.D.
Medical Division:	Division of Anesthesia, Analgesia, and Rheumatology Products (DAARP)
Clinical Team:	Medical Officer: Neville Gibbs, M.D.
	Medical Team Leader: Robert Shibuya, M.D.
Project Manager:	Tanya Clayton

Keywords: Clinical studies; NDA review

1. EXECUTIVE SUMMARY OF STATISTICAL FINDINGS

1.1 Conclusions and Recommendations

The applicant submitted results from two Phase 3 clinical studies intended to assess the efficacy of Qutenza ® (8% capsaicin patch) for the prolonged reduction of neuropathic pain associated with postherpetic neuralgia (PHN). In both studies, the results indicate that the Qutenza patch was statistically significantly superior to the control ($p<0.05$) in reducing pain. Based on the results of the two studies, there is sufficient evidence to support the efficacy of Qutenza ® (8% capsaicin patch) for the prolonged reduction of neuropathic pain associated with postherpetic neuralgia (PHN).

1.2 Brief Overview of Clinical Studies

The applicant conducted a two prospectively planned, randomized, double-blind, active-controlled clinical studies to assess the efficacy of Qutenza for the prolonged reduction of neuropathic pain associated with postherpetic neuralgia (PHN). Both studies (C116 and C117) had the same design, patient population, efficacy measurements, and planned analyses.

The two treatment arms were Qutenza (8% concentration patch; 640 mcg/cm^2) or a low concentration control patch (3.2 mcg/cm^2). The control patch was needed to maintain blinding because some patients experience a burning sensation from the topical capsaicin at the application site.

In the clinical studies, the Qutenza patch was referred to as high-concentration capsaicin, 640 mcg/cm^2, or by the code NGX-4010. The control patch was referred to as low-concentration capsaicin, or 3.2 mcg/cm^2. After screening for eligibility, patients received a topical local anesthetic on the area to be treated, and then the blinded study treatment patch was applied for 60 minutes. When the patch was removed, the area was cleansed to remove any remaining study drug from the skin. Follow-up visits were planned for 4, 8, and 12 weeks after the day of study treatment.

In each study eligible patients were randomized equally to the two treatment groups. In study C116, patients were randomized within four strata defined by gender and risk of cardiovascular AE (high/low). The risk of cardiovascular AEs showed no imbalance in cardiac AEs. In study C117, randomization was stratified by gender.

Patients were adults, ages 18-90, with moderate to severe pain for PHN. Pain was measured on an NPRS scale, with 0 = No Pain and 10 = Worst Possible Pain. The question for the daily pain score was phrased as “the average pain over the last 24 hours” and was to be recorded in the evening. Baseline pain was the average of the daily pain scores during a 14-day screening period. Patients continued recording daily pain for 12 weeks after the treatment application.

The applicant's primary endpoint was the percent change in average pain from baseline to Weeks 2-8. The Division of Anesthesia, Analgesia, and Rheumatology Products (DAARP) medical reviewers preferred a primary endpoint defined as the percent change in average pain from baseline to Week 8. Secondary endpoints of interest to the medical officer, Dr. Gibbs, were the actual change in average pain from baseline to Week 8, and the proportion of patients who improved by 30% or 50% from baseline to Week 8.

The planned primary analysis was an ANCOVA model with terms for treatment, gender, and baseline pain as the covariate. The hypothesis was superiority of the Qutenza patch to the low-concentration control patch. In each study a single primary endpoint was pre-specified, with a single between group comparison, so no adjustment for multiplicity was needed.

In both studies, the results indicate that the Qutenza patch was statistically significantly superior to the control ($p<0.05$) in reducing pain. Based on the results of the two studies, there is sufficient evidence to support the efficacy of Qutenza ® (8% capsaicin patch) for the prolonged reduction of neuropathic pain associated with postherpetic neuralgia (PHN).

1.3 Statistical Issues and Findings

The primary endpoint defined in the protocols was not the endpoint preferred by the Division for studies of treatment of chronic pain. The applicant was given that advice in a protocol review, but did not revise the protocol to match the medical officer's request. For the review of these studies, I performed the planned analyses and the preferred analyses for both studies. In both cases the results for the primary analyses did meet statistical significance ($p<0.05$).

2. Introduction

2.1 Overview

The active ingredient in Qutenza is capsaicin. This is an unapproved drug which is currently marketed (in concentrations of between 0.025% and 0.075%) in topical ointments to relieve the pain of peripheral neuropathy such as post-herpetic neuralgia caused by shingles. It is also available in creams or large bandages for the temporary relief of minor aches and pains of muscles and joints associated with arthritis, simple backache, strains and sprains. Capsaicin is also sold as a dietary supplement.

On May 22, 2009, Qutenza received orphan drug designation for the management of neuropathic pain in patients with PHN.

The Phase 3 study design, patch application time, and use of the low-dose control patch were

agreed to at the End of Phase 2 meeting on March 6, 2003. In a protocol review for study C117 (letter to sponsor dated September 10, 2007), DAARP medical reviewers advised the sponsor that they did not agree with the primary endpoint of change from baseline in average pain during Weeks 2-12. The explanation was "For a drug intended to treat a chronic condition, we are interested in an evaluation of durability of effect. Therefore, analyses of efficacy at the end of treatment (in this case 12 weeks) are of primary interest." The protocol was revised to indicate that Week 8 was considered the duration for efficacy, and the sponsor's primary endpoint was redefined as the change from baseline in average pain during Weeks 2-8. This did not adequately address the DAARP preferred analysis at the end of treatment, i.e. Week 8.

2.2 Data Sources

All data was supplied by the applicant to the CDER electronic data room (edr) in SAS transport format. All necessary documentation, formats, and links were provided as well. The data and final study report for the electronic submission were archived under the network path location \\CDSESUB1\EVSPROD\NDA022395\022395.ENX

3. Statistical Evaluation

3.1 Evaluation of Efficacy

Study C116 (conducted 5/05 to 8/06)

Design

Study C116 was a randomized, double-blind, parallel arm, multi-center study. The objective was to compare the efficacy of a single 60-minute application of Qutenza 8% capsaicin patch to a low concentration control patch in patients with postherpetic neuralgia (PHN). Efficacy was assessed as the reduction in average daily pain at the treatment site.

During the screening period, patients recorded average daily pain on the NPRS scale with 0 = No Pain and 10 = Worst Pain Possible. An average score of 3 to 9 during screening was required for eligibility in the study. Patients were also required to not be using any topical creams and, if on pain medication, to be on a stable dose prior to baseline assessments.

After eligibility was determined, patients were randomized equally to the two treatment arms. A total of 402 patients were enrolled, with 206 randomized to receive Qutenza and 196 randomized to receive the low-concentration control patch. Randomization was stratified by gender and cardiovascular risk (high/low).

On the day of study drug application, a topical local anesthetic was applied to the area to be treated prior to the application of the study drug patch. After 60 minutes, the patch was removed

and the area was cleaned to remove any remaining study drug. Safety and tolerability assessments were made throughout the procedure. Pain measurements were recorded on the day of study patch application, but efficacy was not assessed at the time of patch application.

Patients recorded the average daily pain on the NPRS scale for 12 weeks after the patch application. Due to the potential residual pain from the capsaicin treatment itself, the daily pain assessments for the neuropathic pain during the first 7 days after treatment were not included in the efficacy analyses. If a patient had no pain observations after Day 7, the baseline average pain was carried forward (BOCF). Any missing observations beyond Day 8 were imputed using Last Observation Carried Forward (LOCF). This imputation approach is not typically used for pain assessments, but in this case there was no additional treatment, so there was less concern about impacting the pain assessments. In addition, a responder analysis was done in which any patient who discontinued was defined as a non-responder, to address the potential impact of imputing data for patients who discontinued due to a bad outcome. The analyses used the Intent-to-Treat (ITT) patient population which included all randomized patients.

The protocol planned for a single primary efficacy endpoint: the percent change from baseline in the "average pain for the past 24 hours" of NPRS score during Weeks 2-8. Secondary endpoints planned were the absolute change from baseline in the "average pain for the past 24 hours" of NPRS score during Weeks 2-8, and the proportion of subjects reaching 30% and 50% reduction in average pain from baseline during Weeks 2-8. For each of these endpoints, the medical officer, Dr. Gibbs, requested that a landmark timepoint of Week 8 be used instead of the Weeks 2-8 timeframe used by the applicant.

The protocol planned to analyze the pain endpoints using an ANCOVA model stratified by gender with models terms for treatment and baseline pain. The applicant described additional terms in the model regarding pain before and after the topical anesthesia prior to application of the study drug patch as "candidate covariates" in the protocol (Section 9.2). There terms were included in the analyses provided in the application. Dr. Gibbs requested the results for the ANCOVA model without those additional terms.

Patient Disposition

Patients were adult males and females with neuropathic pain due to PHN. A total of 402 patients were randomized to the study, 206 to the Qutenza treatment group and 196 to the low dose control group. There were few dropouts, with no notable differences between the treatment groups in terms of disposition. Table 1 shows the distribution of patient discontinuations by reason.

Table 1: Patient Disposition (Study C116)

	Low concentration 3.2 mcg/cm^2 [CONTROL]	Qutenza ® High concentration 640 mcg/cm^2
Randomized	196 (100%)	206 (100%)
Discontinued	18 (9%)	19 (9%)
Adverse event	0	1 (<1%)
Lack of efficacy	9 (5%)	10 (5%)
Noncompliance	1 (1%)	1 (<1%)
Lost to Follow-up	2 (1%)	3 (1%)
Other	6 (3%)	4 (2%)
Completed	178 (91%)	187 (91%)
Intent-to-Treat (ITT) All randomized who received study drug and had at least 3 NPRS values during baseline screening	196 (100%)	206 (100%)

Sources: Clinical Study Report Figure 1 and Table 14.1.1.

Baseline Demographics

The two treatment groups were well balanced with respect to relevant demographic and baseline characteristics as shown in Table 2.

Table 2. Patient Demographics for Intent-to-Treat (ITT) Population (Study C116)

	Low concentration 3.2 mcg/cm^2 [CONTROL] n=196	Qutenza ® High concentration 640 mcg/cm^2 n=206
Age (years)		
Mean (SD)	71 (12)	72 (12)
Age categories:		
<50 yrs	10 (5%)	7 (3%)
50-64 yrs	42 (21%)	41 (20%)
≥65 yrs	144 (73%)	158 (77%)
Gender		
Female	105 (53%)	107 (52%)
Male	92 (47%)	98 (48%)
Race		
White	181 (92%)	188 (92%)
Black	7 (4%)	6 (3%)
Asian	2 (1%)	5 (2%)
Other	7 (4%)	6 (3%)
Duration of PHN Pain (years)		
Mean (SD)	3.7 (4.9)	4.1 (4.3)
Min, Max	0.4, 28.7	0.5, 25.4
Size of Treatment Area (cm^2)		
Mean (SD)	349 (219)	330 (219)
Min, Max	22, 1020	9, 1000

Sources: Clinical Study Report Tables 5 and 6

Efficacy Results

Table 3 presents the results for the analyses of the average daily pain scores. The applicant's analyses use a timeframe of Weeks 2-8, which was calculated as the average of all daily pain scores from days 8 through 56. The Division requested a landmark analysis using just the Week 8 results, calculated as the average pain scores for days 50-56. Dr. Gibbs also requested that I provide the results for the actual change in pain.

In Study 116, the results for the applicant's analyses, which averaged pain scores over weeks 2 through 8, were nearly the same as the results using the DAARP landmark timepoint of Week 8. This suggests that the pain scores were consistent across weeks 2 through 8. The applicant presented results (Section 11.4.1.3.3; Figure 2) which support this conclusion.

The other difference between the analyses is in the term used in the ANCOVA models, as noted in the footnotes. The applicant included terms for the pain scores before and after the topical local anesthetic cream (LMX4® lidocaine 4%) was applied prior to the study treatment patch. The inclusion of these terms did not change the conclusions.

Figure 1 and Table 4 show the results of the continuous responders analysis for the change in average pain from baseline to Week 8. Both show that the reduction in pain was greater for Qutenza group than for the control group.

The results of all the analyses are consistent and indicate that the Qutenza group was statistically significantly superior to the control group in the reduction of pain through Week 8 after treatment with the study patch.

Table 3. Efficacy Results (Study C116)

Change in Average Pain from Baseline		Low concentration 3.2 mcg/cm^2 [CONTROL] n=196	Qutenza ® High concentration 640 mcg/cm^2 n=206
Applicant's Primary Analysis:* Percent Change from Baseline to Average of Weeks 2 through 8	LSMeans (SE) Diff. p-value vs. control	-19.9 (2.0)	-29.6 (2.0) 9.7 0.001
Applicant's Secondary Analysis:* Actual Change from Baseline to Average of Weeks 2 through 8	LSMeans (SE) Diff. p-value vs. control	-1.2 (0.1)	-1.7 (0.1) 0.5 0.002
DAARP Preferred Analysis:** Percent Change from Baseline to Week 8	LSMeans (SE) Diff. p-value vs. control	-19.2 (2.3)	-29.9 (2.3) 10.7 0.001
DAARP Alternative Analysis:** Actual Change from Baseline to Week 8	LSMeans (SE) Diff. p-value vs. control	-1.1 (0.1)	-1.7 (0.1) 0.6 0.002

* P-value from ANCOVA model stratified by gender with terms for treatment + baseline pain + pre-LMX4® (lidocaine 4% cream) pain score + percent change in pain score after LMX4®
** P-value from ANCOVA model stratified by gender with terms for treatment + baseline pain
Source: Clinical Study Report Table 7 and SAS datasets

Figure 1: Continuous Responder Curves of Reduction in Pain Intensity (Study C116)

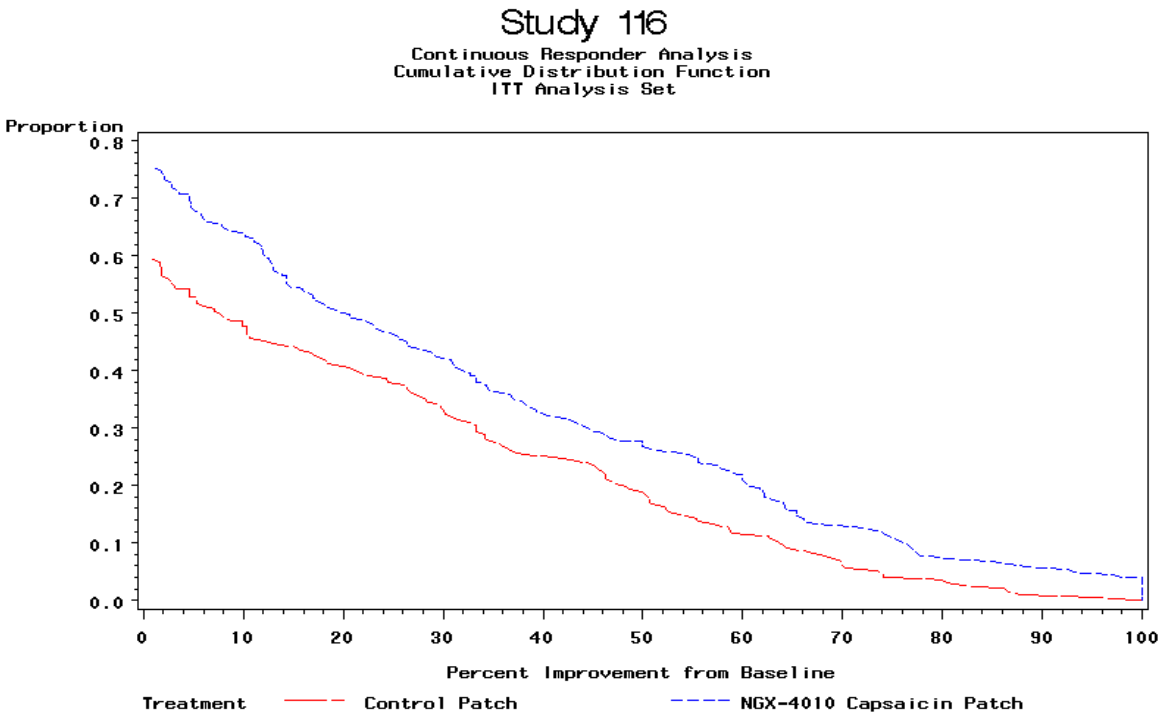


Table 4 shows the percentage of patients who achieved specific levels of improvement in pain intensity by Week 8. This represents the same analysis as Figure 1, using categories instead of the continuous responders curves. In Study 116, there were a higher percentage of patients in the Control group who did not report any improvement in pain intensity than patients in the Qutenza group. This is represented by the separation of the curves at the point of 0% improvement. The separation is maintained beyond the cut-off for at least 50% improvement. These results support the findings in the primary efficacy analysis, and provide evidence that treatment with Qutenza improved pain more than treatment with the control patch.

Table 4: Efficacy Results – Responder Analysis at Week 8 (Study C116)

		Percent change in pain intensity from baseline to Week 8										
Treatment Group	Discontinued	No Imprv. (≤ 0)	>0%	≥10%	≥20%	≥30%	≥40%	≥50%	≥60%	≥70%	≥80%	≥90%
Control N=196	18 9%	60 31%	118 60%	94 48%	81 41%	66 34%	50 26%	38 19%	23 12%	13 7%	7 4%	2 1%
Qutenza N=206	18 9%	33 16%	155 75%	132 64%	103 50%	87 42%	67 33%	57 28%	45 22%	27 13%	16 8%	12 6%

Sources: SAS datasets

Study C117 (conducted 3/06 to 7/07)

Design

Study C117 had a very similar design to Study C116, with four notable differences:

- A requirement that patients had "at least 6 months of pain since shingles vesicle crusting" was added to the eligibility criteria.
- Randomization of patients was stratified by gender, but not cardiovascular risk.
- The planned ANCOVA model was stratified by gender, with terms for treatment and baseline pain (no additional terms were planned or included).
- The Intent-to-Treat population for the efficacy analyses did not include all randomized patients. Two patients in the Qutenza group discontinued prior to application of the study drug (See table 5).

Patient Disposition

Patients were adults, ages 18-90, diagnoses with PHN and at least 6 moths of pain since shingles vesicle crusting. A total of 418 patients were randomized to the study, 214 to the Qutenza treatment group and 204 to the low dose control group. Table 5 shows the distribution of patient discontinuations and reasons. The two groups were similar in terms of their disposition.

Table 5: Patient Disposition (Study C117)

	Low concentration 3.2 mcg/cm^2 [CONTROL]	Qutenza ® High concentration 640 mcg/cm^2
Randomized	204 (100%)	214 (100%)
Discontinued	18 (9%)	20 (9%)
Adverse event	3 (1%)	3 (1%)
Lack of efficacy	5 (2%)	1 (<1%)
Noncompliance	4 (2%)	3 (1%)
Lost to Follow-up	5 (2%)	4 (2%)
Death	0	1 (<1%)
Other	1 (1%)	8 (4%)
Completed	186 (91%)	194 (91%)
Intent-to-Treat (ITT) All randomized who received study drug and had at least 3 NPRS values during baseline screening	204 (100%)	212 (99%) 2 patients withdrew prior to study drug application

Sources: Clinical Study Report Figure 1 and Table 14.1.1.

Baseline Demographics

The two treatment groups were well balanced with respect to relevant demographic and baseline characteristics as shown in Table 6.

Table 6: Patient Demographics for Intent-to-Treat (ITT) Population (Study C117)

	Low concentration 3.2 mcg/cm^2 [CONTROL] n=204	Qutenza ® High concentration 640 mcg/cm^2 n=212
Age (years)		
Mean (SD)	70 (13)	70 (12)
Age categories:		
<50 yrs	17 (8%)	16 (8%)
50-64 yrs	39 (19%)	48 (23%)
≥65 yrs	148 (73%)	148 (70%)
Gender		
Female	107 (52%)	119 (56%)
Male	97 (48%)	93 (44%)
Race		
White	191 (94%)	197 (93%)
Black	8 (4%)	6 (3%)
Asian	2 (1%)	3 (1%)
Other	3 (1%)	6 (3%)
Duration of PHN Pain (years)		
Mean (SD)	3.3 (3.7)	3.1 (3.6)
Min, Max	0.3, 26.6	0.1, 26.4
Size of Painful Area at Screening (cm^2)		
Mean (SD)	326 (215)	330 (226)
Min, Max	15, 980	13, 1008

Sources: Clinical Study Report Tables 5 and 6

Efficacy Results

In the protocol for study C117, the applicant planned an ANCOVA model stratified by gender with terms for treatment and baseline pain covariate. This is the same model Dr. Gibbs preferred. I provided the same analyses for this study as for Study C116 (see Table 7). The continuous responder analyses are provided in Figure 2 and Table 8.

In this study, the two analyses for the percent change from baseline (to Weeks 2-8 or to Week 8) both indicate that the Qutenza group was statistically significantly superior to the low-concentration control group. The secondary endpoint of interest was the actual change. For that endpoint, the analysis of change from baseline to Week 8 indicated Qutenza was not significantly different from the control. This result suggests that the difference between the groups at Week 8 was not as large as the average difference over the period from Week 2 through 8. The applicant provided the result by week in Section 11.4.1.3.1 and Figure 2 of the study report which supports this finding. The non-significant test result for the actual change from baseline to Week 8 does not contradict the other results. Additionally, the group means and differences between the groups were in a consistent direction in both studies.

Table 7: Efficacy Results (Study 117)

Change in Average Pain from Baseline		Low concentration 3.2 mcg/cm^2 [CONTROL] n=196	Qutenza ® High concentration 640 mcg/cm^2
Applicant's Primary Analysis:*			
Percent Change from Baseline to Average of Weeks 2 through 8	LSMeans (SE)	-24.4 (2.1)	-32.0 (2.1)
	Diff.		7.6
	p-value vs. control		0.011
Applicant's Secondary Analysis:*			
Actual Change from Baseline to Average of Weeks 2 through 8	LSMeans (SE)	-1.3 (0.1)	-1.7 (0.1)
	Diff.		0.4
	p-value vs. control		0.034
DAARP Preferred Analysis:*			
Percent Change from Baseline to Week 8	LSMeans (SE)	-26.3 (2.4)	-32.9 (2.3)
	Diff.		6.6
	p-value vs. control		0.046
DAARP Alternative Analysis:*			
Actual Change from Baseline to Week 8	LSMeans (SE)	-1.4 (0.1)	-1.7 (0.1)
	Diff.		0.3
	p-value vs. control		0.125

* P-value from ANCOVA model with terms for treatment + gender + baseline pain score (as planned in protocol)
Source: Clinical Study Report Table 6 and SAS datasets

Figure 2: Continuous Responder Curves of Reduction in Pain Intensity (Study C117)

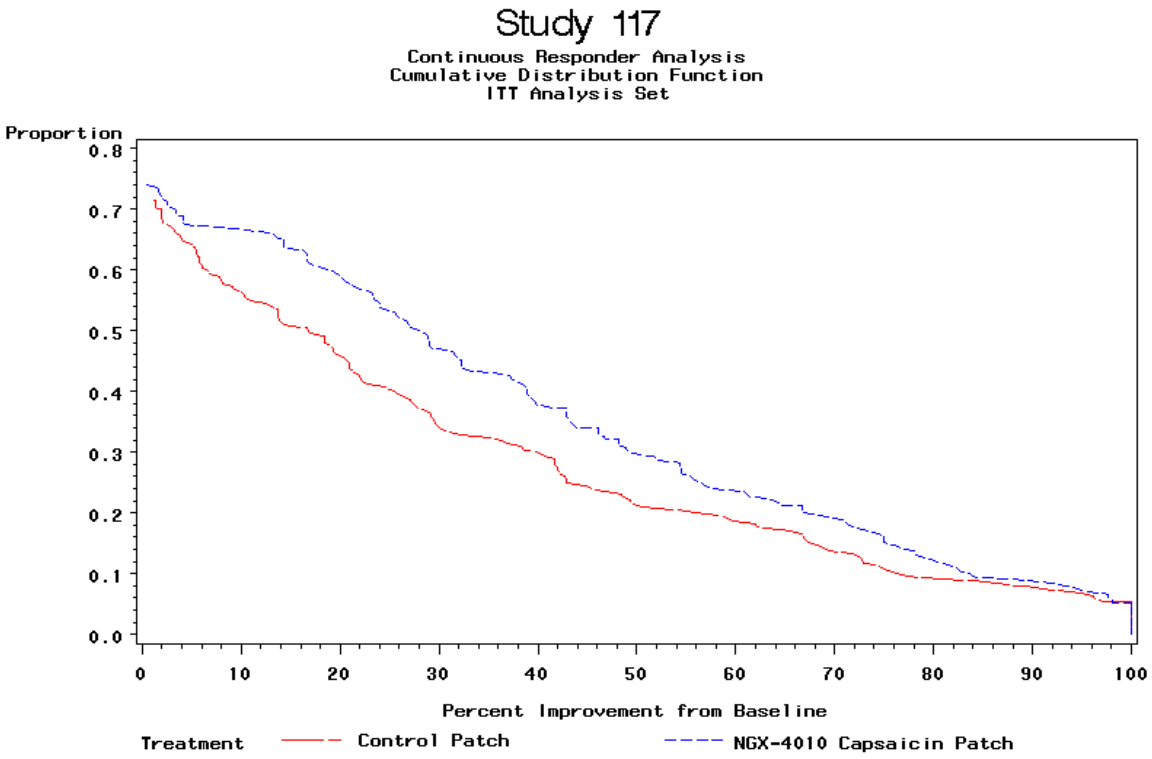


Table 8 shows the percentage of patients who achieved specific levels of improvement in pain intensity by Week 8. This represents the same analysis as Figure 2, but with categories instead of the continuous responders curves. As shown in the curves, the two treatment groups were similar in terms of the percentage of patients who discontinued or did not improve. The two curves separate at the ≥10% cut-off point, indicating that patients in the Qutenza group who improved reported greater improvement than those in the control group. These results support the primary analysis and provide evidence to support Qutenza for this indication.

Table 8: Efficacy Results – Responder Analysis at Week 8 (Study C117)

		Percent change in pain intensity from baseline to Week 8										
Treatment Group	Discontinued	No Imprv. (≤ 0)	>0%	≥10%	≥20%	≥30%	≥40%	≥50%	≥60%	≥70%	≥80%	≥90%
Control N=204	18 9%	39 19%	147 72%	116 57%	94 46%	70 34%	62 30%	44 22%	38 19%	28 14%	19 9%	16 8%
Qutenza N=212	18 8%	36 17%	158 75%	142 67%	126 59%	100 47%	81 38%	63 30%	51 24%	41 19%	26 12%	19 9%

Sources: SAS datasets

3.2 Evaluation of Safety

No safety analyses were requested Dr. Gibbs.

4. Findings in Special/Subgroup Populations

4.1 Gender, Race and Age

The applicant provided descriptive analyses by age groups, gender, and race, but reported the change from baseline to Weeks 2 - 8. Table 9 shows the same subgroup analyses for the change from baseline to Week 8. There were no notable differences in the percent change from baseline to Week 8 for the treatments across any of these subgroups.

Table 9: Subgroup Analyses

Primary Endpoint: Percent change in average pain from baseline to Week 8 Mean (SE)	Study 116		Study 117	
Capsaicin Patch Concentration:	Low concentration 3.2 mcg/cm^2 Control n=196	High concentration 640 mcg/cm^2 Qutenza n=206	Low concentration 3.2 mcg/cm^2 Control n=204	High concentration 640 mcg/cm^2 Qutenza n=212
Age groups				
18-64 years	n=52 -26 (5)	n=48 -40 (5)	n=56 -44 (5)	n=64 -39 (5)
≥65 years	n=144 -17 (3)	n=158 -27 (3)	n=148 -20 (3)	n=148 -31 (3)
Gender				
Female	n=105 -21 (3)	n=107 -34 (3)	n=107 -33 (4)	n=119 -37 (3)
Male	n=92 -17 (3)	n=98 -25 (3)	n=97 -19 (3)	n=93 -30 (3)
Race				
Caucasian	n=181 -18 (2)	n=188 -29 (2)	n=191 -27 (3)	n=197 -34 (3)
All others	n=15 -35 (9)	n=18 -37 (7)	n=13 -19 (8)	n=15 -27 (8)

Sources: SAS datasets

Sources: Clinical Study Report Table

4.2 Other Special/Subgroup Populations

No additional subgroup analyses were requested Dr. Gibbs.

5. Summary and Conclusions

5.1 Statistical Issues and Collective Evidence

The applicant provided two prospectively-planned, randomized, double-blind, active-control, parallel-arm studies to support the efficacy of Qutenza for the treatment of pain in patients with PHN. The primary endpoint defined in the protocols was not the endpoint preferred by the Division for studies of treatment of chronic pain. For the review of these studies, I performed the planned analyses and the preferred analyses for both studies. There were no other statistical issues.

In both studies, pain assessments at Week 8 after treatment with the study treatment patch showed that Qutenza was significantly statistically superior to the low-concentration control patch. The secondary endpoints and continuous responder analyses support this conclusion. There is sufficient evidence in these two studies to support the high concentration patch for this indication.

5.2 Label Issues

The applicant's proposed label reports the results from the analysis in the Clinical studies section. The descriptions of the study designs are appropriate. The applicant reports the results for the average of Weeks 2-8 as planned in the protocol, but the DAARP medical reviewers may prefer that the results of the Week 8 landmark analyses be presented instead, and I would agree.

I would prefer the following changes in the reporting of the study results:

1. The applicant reports that results for analyses of efficacy through Week 12. Week 8 was prespecified as the timepoint of interest for efficacy, so comparisons beyond that should not appear in the label.

2. The applicant also reports results for the proportion of patients with 30%, 50% or a 2-point reduction in pain. These were secondary endpoints in the protocol, not prespecified as being intended for label claims, and should not be included in the label.

3. The statements that "pain reduction occurred as early as Week 1 and persisted throughout the study" should not appear in the label. The studies were not designed to make those comparisons. Figures 3 and 4 in the applicant's proposed label show the percent change from baseline in average pain by week. These figures visually represent the same time-to-effect concept and ideally would be removed from the label. If these figures are not removed, the asterisks indicating p-values should be removed for the same

reason.

5.3 Conclusions and Recommendations

The applicant provided results from two studies planned to demonstrate the efficacy of Qutenza patch for the treatment of pain in patients with PHN. In both studies the analysis of the percent change in average daily pain from baseline to Week 8 indicated that the Qutenza patch was significantly statistically superior to the low-concentration control patch. The secondary endpoints and continuous responder analyses support this conclusion. There is sufficient evidence in these two studies to support the high concentration patch for this indication.

Signatures/Distribution List Page

Katherine B. Meaker, M.S.
Mathematical Statistician

Dionne Price, Ph.D.
Team Leader

Thomas Permutt, Ph.D.
Division Director

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

Katherine Meaker
7/1/2009 01:49:26 PM
BIOMETRICS

Dionne Price
7/1/2009 01:51:52 PM
BIOMETRICS
Concur in my role as Team Leader and Acting Division Director.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH
OFFICE OF TRANSLATIONAL SCIENCES
OFFICE OF BIOSTATISTICS

STATISTICAL REVIEW AND EVALUATION

Stability Studies

NDA/Serial Number:	22-395/000, 007, 008
Drug Name:	Qutenza (Capsaicin Patch 8%)
Indication:	For Relief of Neuropathic Pain Associated with Post-herpetic Neuralgia.
Applicant:	NeurogesX, Inc.
Date:	13 October, 2008
Classification:	Standard
Statistical Reviewer:	Roswitha Kelly, M.S., OTS/OB/DB6
Concurring Reviewer:	Yi Tsong, Ph.D., OTS/OB/DB6
Medical Division:	Division of Anesthesia, Analgesia, and Rheumatology Products
Chemistry Reviewer:	Theodor Carver, Ph.D., OPS/ONDQA/DPA1
Chemistry Branch Chief:	Danae Christodoulou, Ph. D., OPS/ONDAQ/DPA I
Project Manager:	Tanya Clayton, OND/ONDEII/ DAARP

Keywords: Capsaicin Patch 8%, drug product stability, patch, cleansing gel.

Distribution: NDA 22-395/Qutenza
OND/ONDEII/DAARP/T. Clayton
ONDQA/T. Carver, Ph.D./D. Christodoulou, Ph.D.
OTS/OB/DB6/R. Kelly/Y. Tsong, Ph.D./S. Machado, Ph.D.
OTS/OB/L. Patrician, M.S./R. Tiwari, Ph.D.

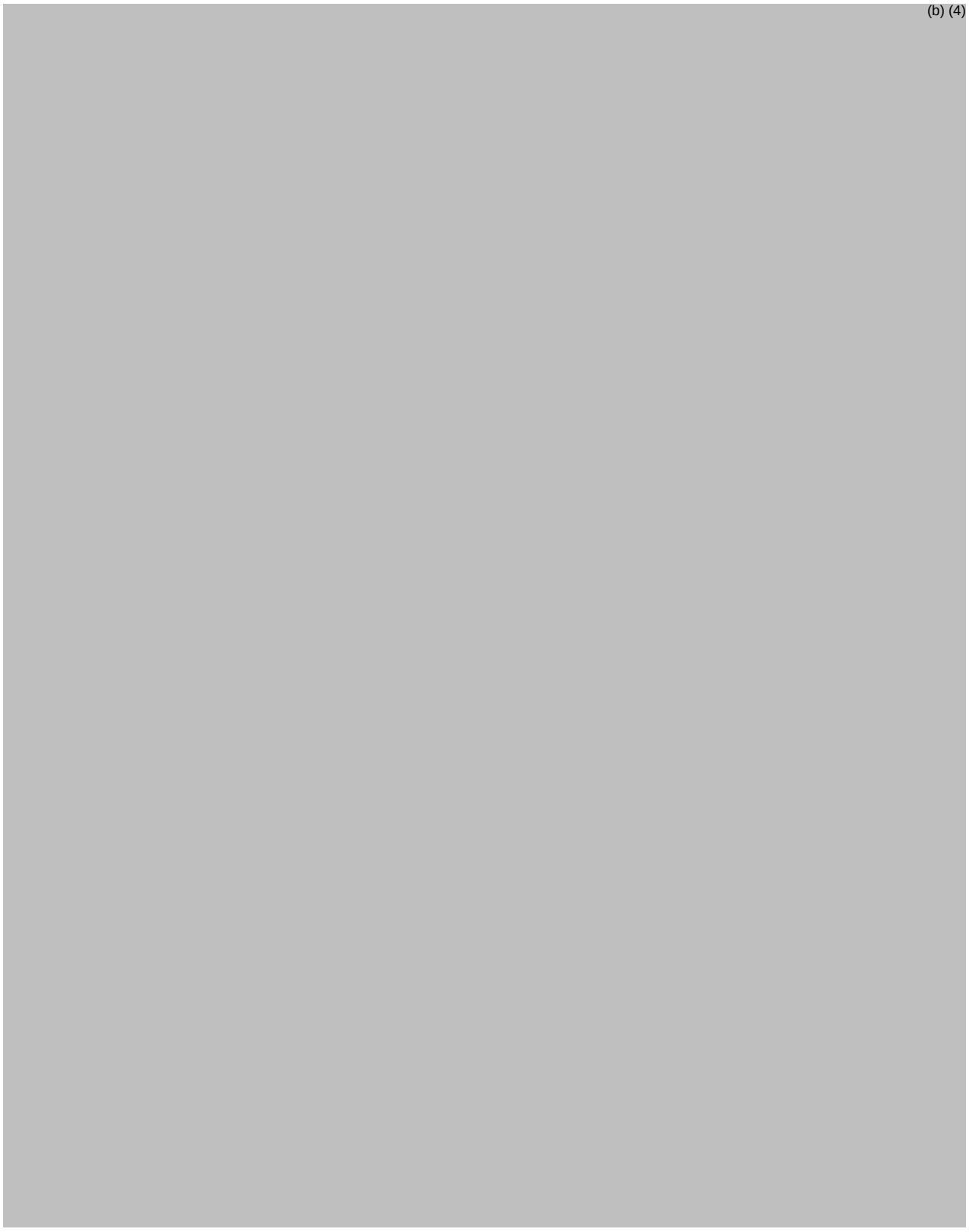
File directory: My Documents/NDA 22395_F.pdf

Table of Contents

Table of Tables

(b) (4)

Table of Figures




(b) (4)

1. EXECUTIVE SUMMARY

1.1. Conclusions and Recommendations

The Qutenza product consists of a dermal patch containing 8% capsaicin as well as a tube of cleansing gel to clean off any rem aining drug when the patch is rem oved. In consultation with the reviewing chemist, capsaicin assay, cis-capsaicin, DGME, adhesive force and dissolution were determ ined to be important stability attributes for the patch. Water content and viscosity were decided to be the im portant stability attributes fo r the cleansing gel.

The sponsor requested a (b) (4) shelf life without any statistical evaluation of the updated stability data. The reviewer found that the stability da ta from the patches as well as from the cleans ing gel were fairly stable, but not to the degree desired by the sponsor (Table 1 below). None of the groups of batc hes pooled com pletely indicating that either the indiv idual batch es d id not clos ely reprodu ce a common stability p rofile or th at the variability around each regression line was so tight that d ifferences between batch es became significant. Based on statistical extra polation of the patch stab ility data, the non-dissolution attributes supported a (b) (4) expiry for both th e commercial and the pilot lots. None of the commercial batches supported a (b) (4) expiry for the 0.5, 1.0, or 4.0 dissolution time points. One comm ercial batch estimated a shelf life as low as 35 months based on the 4.0 hour dissolution observations (the sponsor did not update the 7 and 23 hour dissolution tim e points). All dissolution da ta from the pilot batches supported a (b) (4) (b) (4) expiry. The cleansing gel cam e in three tube sizes of which the 50g one will become the commercial presentation. The wa ter content data from the 30g and the 50g tubes supported only a 46 m onth expiry. All viscosity data extrapolated to shelf lives longer than (b) (4).

As noted, none of the groups of batches for either patch or clea nsing gel pooled to a single regression line. Hence the estim ated shelf life for the whole product will be derived from the shortest sh elf life estim ated based on a single batch. From a purely statistical point of view one can allo w extrapolation wh en the assum ption of a continued similar stability p rofile is tenab le. In addition, th is assum ption will be v erified as m ore stability data becom e available and the shel f life is updated. Taking this view point the shortest extrapolated shelf life would be 35 months based on 14 m onth stability data for the 4.0 hour dissolution of a comm ercial lot of the patch produc t. The regulatory point of view allows for extrap olation of twice the available d ata but of not m ore than an additional 12 m onths. Following this appro ach the shortest shelf life for the product would be 20 m onths based on a 50g tube ba tch of the cleansing gel with only 10 month stability data (this is afte r m oving the nine m onth stabil ity data by one m onth lag tim e after the actual releas e from manufacture). The actua l granting of a shelf life lies within the purview of the reviewing chemist.

1.2. Brief Overview of Stability Studies

The sponsor submitted stability data from three pilot and three commercial lots of the capsaicin patch 8% as well as from a number of batches for each of the three tube sizes of the cleansing gel. With the first stability submission the sponsor provided regression analyses and estimated shelf lives. One commercial batch of the patch was excluded from the analyses because it had only 3 month of stability data. However there was a conflict between the sponsor's request of (b) (4) expiry and one cleansing gel batches estimating only a 25 months shelf life based on their own calculations.

The reviewer was made aware that there was a lag time of 0.5 to 6.0 months between the release time point and the time point a batch was put on stability. The sponsor was asked for clarification and to describe in particular in what form the material had been held. There were several exchanges of information between the sponsor and the Agency, some of which had updated but incomplete stability data. It was also found that the sponsor had chosen to stop dissolution determinations with hour 4, though originally there had been the additional time points of 7 and 23 hours. The chemistry and statistical reviewers decided to accept the last stability update (3/10 and 3/11/09) and use these data for estimating the expiries based on the commercial and pilot batches of the patch and the cleansing gel batches in various tube sizes.

1.3. Statistical Issues and Findings

The chemistry reviewer alerted the statistical reviewer that the sponsor's time zero of the electronically submitted stability data was actually the start of the stability program for a given batch and not the release time point. Depending on the batch, there was a lag time of 0.5 to 6.0 months between the release date and the start of the stability study. The sponsor had retested all batches at the start of the stability study and taken that time point as time zero in their analyses. As the patch and the cleansing gel were held in their final product form during this lag time, the reviewer added the actual release data and adjusted all stability time points by the lag time appropriate for each batch. As can be seen in the Appendix, the release information rarely affected the stability profile of a batch. According to the sponsor it is not uncommon that there is a notable time difference between the release of a lot from manufacture and the time the lot is put on stability.

The reviewer observed that some stability profiles appeared non-linear. However, these were isolated cases and did not seem to present a general profile appropriate for a given attribute. Hence the reviewer did not fit any non-linear models, as they would not be representative of an attribute or group of batches.

The sponsor had submitted statistical analyses and regression graphs with the original submission. They did not submit any updated analyses with the stability updates.

For none of the attributes analyzed, none of the stability data of groups of batch representing a type of patch product (pilot or commercial) or a tube size of the cleansing gel pooled to single regression line. Hence the shelf life estimate applicable to the

product will be based on a single batch. There is a wide range of available data (9 to 48 months or 10 to 53.5 months after insertion of actual release data and lag time) and these will influence the length of potential extrapolation beyond the actual data. When taking a pure statistical point of view, one could permit longer extrapolation because the shelf life will be updated at least yearly with additional data and re-evaluation. With this approach, the shortest shelf life would be 35 months which comes from a commercial patch batch (4.0 hour dissolution time) with 14 month stability data. However, the regulatory guideline is stricter and permits only the doubling of available data not to exceed an additional 12 months. With this approach, the shelf life for the product would be 20 months as one cleansing gel batch has 10 month data after adjusting the 9 month stability data by inserting the actual release information and the lag time.

2. INTRODUCTION

2.1. Overview

The Qutenza product consists of a dermal patch containing 8% capsaicin as well as a tube of cleansing gel to clean off any remaining drug when the patch is removed. In consultation with the reviewing chemist, capsaicin assay, cis-capsaicin, DGME, adhesive force and dissolution were determined to be important stability attributes for the patch. Water content and viscosity were decided to be the important stability attributes for the cleansing gel.

The sponsor submitted stability data from three pilot and three commercial lots of the capsaicin patch 8% as well as from a number of batches for each of the three tube sizes of the cleansing gel. With the first stability submission the sponsor provided regression analyses and estimated shelf lives. One commercial batch of the patch was excluded from the analyses because it had only 3 month of stability data. However there was a conflict between the sponsor's request of a (b) (4) expiry and one cleansing gel batches estimating only a 25 months shelf life based on their own calculations.

The reviewer was made aware that there was a lag time of 0.5 to 6.0 months between the release time point and the time point a batch was put on stability. The sponsor was asked for clarification and for a description of the form in which the material had been held. There were several exchanges of information between the sponsor and the Agency, some of which had updated but incomplete stability data. It was also found that the sponsor had chosen to stop dissolution determinations with hour 4, though originally there had been the additional dissolution time points of 7 and 23 hours. The chemistry and statistical reviewers decided to accept the last stability update (3/10/09 and 3/11/09) and use these data for estimating the expiries based on the commercial and pilot batches of the patch and the cleansing gel batches in the various tube sizes.

2.2. Data Sources

The sponsor provided statistical analyses and shelf life estimations in the original submission, but the stability data were not in electronic data files. Hence, the stability data were requested as SAS transport files. In their response to this request the sponsor included stability updates but missed data for certain attributes. In their 3/10/09 and 3/11/09 submissions the sponsor included another stability update which covered all attributes. However, the sponsor had decided to stop obtaining dissolution data after 4 hours. The data files from these recent stability submissions served as the basis for the reviewer's statistical analyses and shelf estimates discussed here.

There were stability data from three commercial and three pilot batches of the patch product. With the most recent stability update one commercial batch had as few as 12 month data which increased to 14 months after adding the release information and the lag time. The remaining commercial batches had 25.5 and 37 month stability data and the pilot batches had 37, 50, and 53.5 month data. These stability times reflect the added available release information and lag times and formed the basis for the reviewer's analyses.

The sponsor considered the dissolution results mainly a measure of quality of the patch product, not a measure of how much drug is available to the patient. As the patch is to be used for one hour only they saw no need to continue to collect dissolution data beyond 4 hours. However, the dissolution time point of 0.5 hour would seem important with respect to drug delivery to the patient and this dissolution time point is not available for several early stability time points of the pilot batches.

The cleansing gel had batches from the three tube sizes, 20g, 30g, and 50g. For the to-be-marketed tube size of 50g, the seven batches had been on stability between 10 and 49 months after the addition of release data and lag times.

3. STATISTICAL EVALUATION

3.1. Sponsor's Results

The sponsor submitted regression analyses and graphs for most batches and important attributes for the patch product and cleansing gel with the original submission. At that time one commercial patch batch had only 3 month stability data and was excluded from these analyses. A cleaning gel batch estimated only a 25 month expiry, but the sponsor stated that a (b) (4) shelf life was supported. In the ensuing stability updates, the sponsor confirmed this position but did not submit formal statistical analyses.

3.2. Reviewer's Results

In the original submission, the stability data were not available as electronic data files but as copies of laboratory forms. The sponsor did submit regression graphs and summary statistics for each batch, including a shelf life estimate and concluded that a (b) (4)

expiry was supported. However this was in conflict with their own calculations based on a gel batch, which estimated only a (b) (4) shelf life. When submitting the requested SAS transport files, the sponsor updated the stability data but did not repeat any statistical analyses on them. They still claim that a (b) (4) shelf life is supported.

The reviewing chemist alerted the reviewer to the fact that the sponsor's time zero in their stability data and original analyses was the time the batches were put on stability which was between 0.5 and 6.0 months after the batches were released. Upon request for clarification, the sponsor explained that the products were held in final form and that a lag time between release from manufacture and start of stability was not unusual. In addition, the sponsor had decided that dissolution times after the 4 hour time point were no longer important and had stopped to obtain the 7 and 23 hour dissolution time points.

The reviewer used the updated stability data files sent with the 3/10/09 and 3/11/09 submissions and added the release data and lag times to each batch. The patch product had data from three commercial and three pilot batches. As the data from the commercial batches did not pool, the youngest commercial batch's results became stability limiting. According to ICH guidelines, a batch with 14 months data could maximally extrapolate for another 12 months, if supported by statistical analysis. The pilot batches and the other commercial batches would support longer shelf lives, mainly due to the fact that they had been on stability for longer times.

Table 1 gives the shortest shelf life estimate based on statistical extrapolation, i.e. not restricted by ICH guidelines, for each attribute based on a relevant group of batches. The Appendix presents the statistical analyses and the shelf life estimates and regression graphs for each batch individually. As noted before, none of the groups of batches pooled to a single regression line, but occasionally a common slope model was achieved.

Capsaicin content of the **patch product** showed a fair amount of variability around the regression lines which is also reflected in the fairly small R^2. The slope estimates pooled for the commercial and the pilot batches and the extrapolated shelf life estimates depended on the intercept points but were well beyond the desired (b) (4)

Cis-capsaicin for the commercial batches had basically a slope of zero. As noted before, a non-linear pattern of one batch was not deemed worthy of further investigation. This attribute showed somewhat more variability in slope and intercept for the pilot batches, which was reflected in very low R^2. Still, the extrapolated shelf life estimates went well beyond the desired expiry.

Adhesive force and DGME of the patch displayed obvious linearity for both the commercial product and for the pilot batches. Both types of batches extrapolated well beyond the desired expiry for either attribute.

(b) (4)

The 0.5 dissolution time for the pilot batches had missing release data and did not start on stability till 6-24 m onths later. Based on th e f ew stability data points no conclusion should be m ade. The observed th e stability pattern was very flat and did not raise any concerns. The 1.0 hour dissolution observations for the pilot ba tches showed increases as the batches aged and a fair am ount of vari ability around the regre ssion lines. Again the release information of one batch was m uch lower than the rem aining observations but all shelf life estim ates were well beyond th e requested (b) (4) For the 4.0 hour dissolution tim e point, the variability around th e regression lines increased substantially as did the slope estim ates. Still, the estim ated shelf lives were well beyond the desired one.

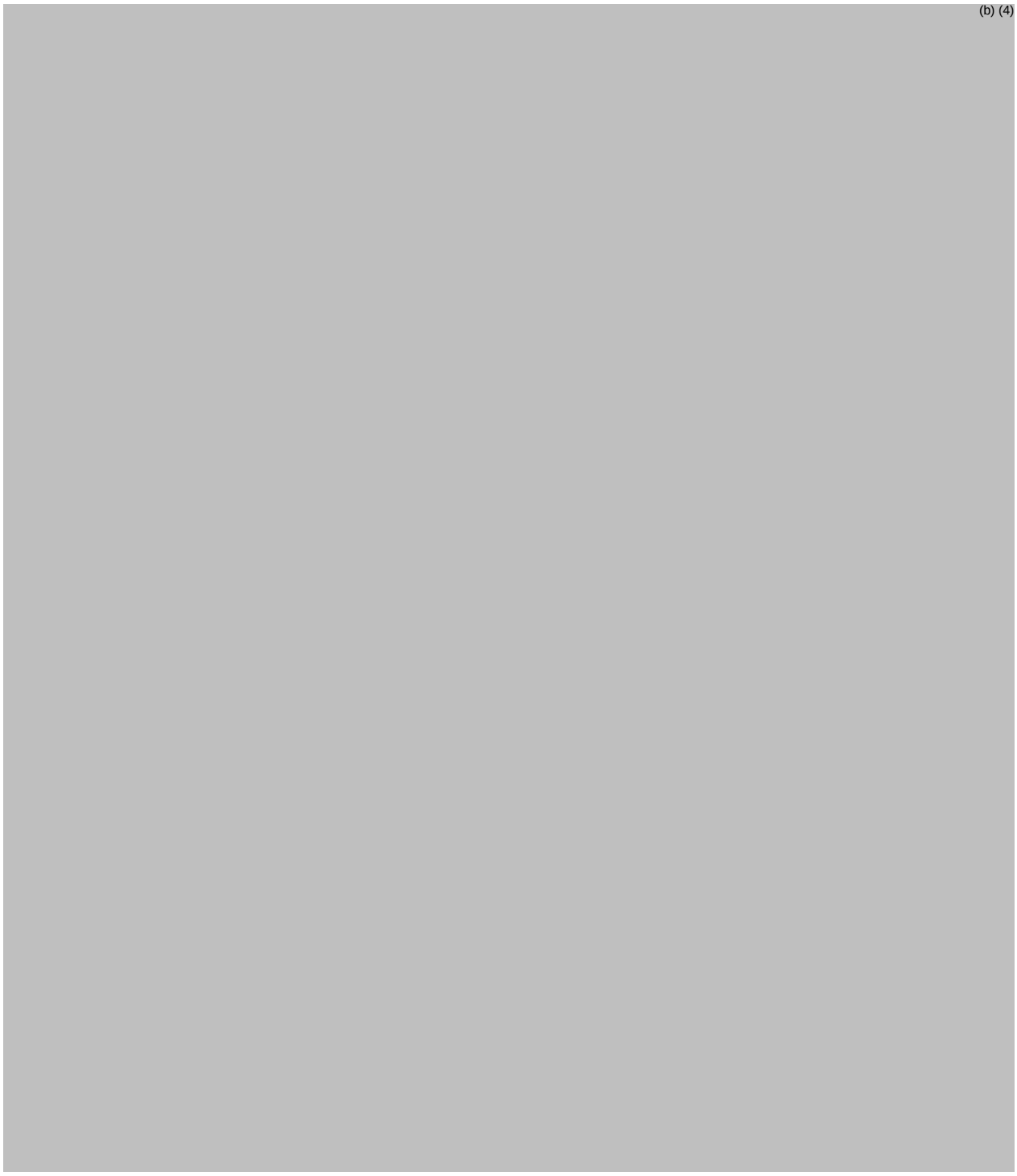
For the patch product the stability data supported an extrapolat ed shelf life of (b) (4) when consid ering statistical extrapolation on ly and the f act that f urther updated stability data will co nfirm or negate the cu rrent es timate. A strict regul atory viewpoint would allow only a 26 m onth expiry (14 m onths of actual stab ility data plus a 12 m onths extrapolation).

The water content of the **cleansing gel** in 20g tubes showed non-linear patterns for two of the three batches, one of which was pronoun ced (batch 25293). Line ar shelf estim ates were at least (b) (4) The 30g tube had onl y one batch on stability which presented a linear pattern and estimated a (b) (4) shelf life. It seems that one batch of the 50g tubes did not have water content data. All but one batch had positiv e slopes with generally linear patterns. One batch showed a very non-linear pattern which was due to low observations at m onth 1. The release data poin t was well within the line ar pattern of the stability data. The shortest shelf life estim ate was (b) (4) based on the batch with 10 month data.

All viscos ity data showed littl e variability regardless of batch or tube size. The shortest estimated shelf life was (b) (4) based on a batch with 10 months data.

For the cleansing gel the stabil ity data supported an extrapol ated shelf life of (b) (4) when considering statistical extrapolation onl y. However, a strict regulatory viewpoint would allow only a 20 month expiry.

Table 1: Shelf Life Estimates for Selected Attributes of Patch and Cleansing Gel Product



4. STATISTICAL ISSUES

According to the sponsor, the dissolution performance of the patch product does not characterize delivery of capsaicin to the patient but is intended as a measure of quality of the commercial drug product. Hence they stopped collecting the 7 and 23 hour dissolution time points. However, the 0.5 and 1 hour dissolution time point would be of special interest. The pilot batches have few early stability observations (irrespective of the lag times) for the 0.5 dissolution time point. Batches 7006204 and 7028014 do not start 0.5 dissolution determinations until 6 months on stability (8 and 11.5 months respectively after release), and batch 7006753 not until month 24 (month 25 after release). Hence 0.5 dissolution evaluations should rely only on the commercial batches.

As the data of the commercial batches of the patch product did not pool, the batch with 12 months stability data became stability limiting. If following the ICH guideline, a batch with 14 month data should be allowed a maximum extrapolation of only another 12 months, if supported by statistical analysis. The pilot batches and the other commercial batches would support longer shelf lives, mainly due to the fact that they had been on stability for longer times.

Similarly, one cleansing gel lot had only 10 months when measured from the time of release. Notwithstanding the statistical extrapolation to at least 46 months, a strict regulatory viewpoint would allow only a 20 month expiry.

Several of the capsaicin assay and cis-capsaicin stability data showed a non-linear profile during the early months on stability. The reviewer did not fit a non-linear model, because a general non-linear pattern could not be ascribed to all commercial or pilot batches. Also, the R^2 was not smaller for the lots with a non-linear pattern than for the lots with a linear pattern.

5. CONCLUSIONS

The sponsor requested a (b) (4) month shelf life despite a (b) (4) expiry based on a cleansing gel batch in the original submission and no statistical evaluation of the updated stability data. The reviewer found that the stability data from the patches as well as from the cleansing gel were fairly stable, but not to the degree desired by the sponsor (Table 1 above). It is noted that the stability times analyzed by the reviewer are between 0.5 and 6.0 months longer than the data submitted by the sponsor. The reviewer incorporated into the data sets the true release data and the lag time between release and start of stability. The additional data and the longer length of each study generally contributed to longer shelf life estimates than had these changes not been made.

None of the groups of batches pooled completely indicating either that the individual batches did not closely reproduce a common stability profile or that the variability around each regression line was so tight that differences between batches became significant. Since none of the groups of batches regressed to a common line the amount of available

stability data per batch and their potential extrapolation became important. Two pilot batches from the patch product and one batch of the cleansing gel (50g tube) had over 48 month stability data and no problem supporting a (b) (4) shelf life. There are another two patch batches and one cleansing gel batch (20g tube) which have over 36 month stability data and also supported an extrapolated (b) (4) expiry. Only six of the total of 17 (patch and cleansing gel) batches could support (b) (4) extrapolated) shelf life when following regulatory guidelines strictly.

From a purely statistical point of view longer extrapolations may be tenable, as the interim expiries will be re-estimated at yearly or more frequent intervals. From this perspective all non-dissolution attributes of the commercial and pilot patch lots as well as the dissolution data from the pilot lots supported a (b) (4) expiry. On the other hand none of the commercial batches supported a (b) (4) expiry for the 0.5, 1.0, or 4.0 dissolution time points. The shortest expiry was (b) (4) based on the 4.0 hour dissolution data from a batch with 14 month stability. Hence, (b) (4) seem to be the most lenient shelf life for the patch product.

The cleansing gel had three 20g tube batches, one 30g tube batch and seven 50g tube batches on stability. The 50g tubes will be marketed. The water content from the 30g and the 50g tubes did not support a (b) (4) expiry, but estimated a (b) (4) shelf life based on batches with 30 and 10 months stability, respectively. As with the patch product, only batches with either at least 36 or 48 month stability could support a (b) (4) (b) (4) expiry under ICH guidelines. There were only two such batches, one 20g tube (37 months) and on 50g tube (49 months). The other batches were between 10 and 30 months on stability. Again, none of the batches per tube pooled to a common regression line and hence the shortest estimated shelf life of an individual batch becomes applicable to all batches. Disregarding the regulatory restriction of no more than 12 months beyond the actual data, and assuming that the observed pattern of the stability profile will be maintained into the future, the most lenient shelf life estimate would be (b) (4). Two batches estimated this shelf life: the single 30g tube batch and one of the 50g tube batches with 10 months data. Again, if following ICH Q1E strictly, the lot with 10 month data would limit the shelf life estimate to 20 months.

From a purely statistical point of view, the shortest extrapolated shelf life of (b) (4) was obtained by a commercial patch lot for the 4.0 hour dissolution data. The batch had 14 months stability data. Because none of the batches pooled to a single regression line, and invoking the ICH Q1E guidelines, the extrapolated shelf lives would be limited to double the minimum of available data, which would be 20 months based on a 50g tube batch of the cleansing gel. The actual granting of a shelf life lies within the purview of the reviewing chemist.

42 pages of Statistics has been withheld in full as B4 CCI/TS

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

Roswitha Kelly
4/14/2009 03:00:19 PM
BIOMETRICS

Yi Tsong
4/14/2009 03:04:04 PM
BIOMETRICS

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:	22,395 / Serial 000
Drug Name:	Capcaisin Dermal Patch 8% (Qutenza™)
Indication:	Prolonged reduction of neuropathic pain with postherpetic neuralgia (PHN)
Applicant:	NeurogesX, Inc. 215 Bridgepointe Parkway, Suite 200 San Mateo, CA 94404-5067 Study conducted at: (b) (4) Vienna, Virginia 22182-1699
Date:	Submitted October 13, 2008
Review Priority:	Standard
Biometrics Division:	Division 6
Statistical Reviewer:	Steve Thomson
Concurring Reviewer:	Team Leader: Karl Lin, Ph. D.
Medical Division:	Anethesia, Analgesia, and Rheumatology Products
Pharmacologist:	Lawrence Leshin, Ph.D. Adam Wasserman, Ph.D.
Project Manager:	Tanya Clayton
Keywords:	Carcinogenicity, Cox regression, Kaplan-Meier product limit, Survival analysis, Trend test, Bayesian analysis

Table of Contents

1. EXECUTIVE SUMMARY

The Sponsor's report indicate that the objective of this study was to assess the carcinogenic potential of trans-Capsaicin when administered weekly via topical application to the dorsal skin of Tg.AC mice for 26 weeks.

It should be emphasized that the usual analysis for these transgenic mice is based on the weekly skin tumor counts. During the course of the study a technician recorded the number of observed "masses." A data set with these counts was submitted. The Sponsor also submitted a data set that purports to include weekly papilloam counts. However, as discussed in Section 2.2 below, this reviewer has some reservations about this data. Apparently the Sponsor also claimed that at the conclusion of the study a retrospective analysis by a toxicologist/ pathologist would be able to differentiate beween papillomas and other masses, and included a data set that indicated the number of animals with tumor. From this assessment at the end of the study it is clear that not all papillomas were identified as masses, nor were all masses identified as papillomas. Following some discussion with toxicologists it was decided to provide an analysis of mice with tumors (Appendix 2) as well as an analysis based on time to first tumor (Appendix 1), and one for the incidence of masses (Appendices 3 and 4). Finally there is an analysis based on the Sponsor supplied papilloma counts (Appendix 5).

1.1. Conclusions and Recommendations

This submission summarizes the results of a study to assess the oncogenic potential of trans-Capsaicin when administered weekly via topical application to the dorsal skin of mice for 26 weeks. The Sponsor reports that male and female Model TGAC-T (hemizygous), FVB/NTac-Tg(v-Ha-ras)TG.ACled mice were assigned to 7 treatment groups per gender, 25 mice/sex/group. Mice in group 1 were treated with the vehicle only. Mice in groups 2-4 are described as receiving the dose formulations containing the vehicle control, diethylene glycol monoethyl ether (DGME), and the test drug (trans-Capsaicin in DGME) at drug levels levels of 0.64, 1.28, and 2.56 mg/mouse/ week. These groups are also labeled as the low, medium, and high dose groups. Group 6, the positive control, were administered Tetradecanoylphorbol-13-Acetate (TPA, in DGME). Group 5 animals received lidocaine only. Group 7 animals were untreated. All treated groups were dosed once per week except for the positive control group, group 6, which was dosed twice per week.

The Sponsor reports that "A number of unscheduled deaths were observed in most groups during the first few months of the inlife phase of this study. At Week 12, the dosing schedule was changed to span the dosing over a 2-day period (Groups 2, 4, and the first weekly dose for Group 6 on Tuesday and Groups 1, 3, and 5 on Wednesday; Group 6 received a second dose on Friday). This change was made to better balance study room activities. Once this change in dosing procedure was implemented, the result was a notable decrease in the number of unscheduled deaths for the remainder of the study." (page 2 of report)

Kaplan-Meier estimated survival curves across dose groups for each gender in each study are displayed in Figures A.1.1-A.1.2 in Appendix 1. These curves are supported by tests of homogeneity and trend in survival over dose groups. The statistical significances of the tests of differences in survival across treatment groups using the log rank and the so-called Wilcoxon test are given in Table 1 below. The Wilcoxon test tends to put higher more weight on later events than does the log rank test. Simple summary life tables in mortality are presented later in the report (Tables 4 and 5).

Table 1. Statistical Significances of Tests of Homogeneity and Trend in Survival

	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Homogeneity over Groups 1-7	<0.0001	<0.0001	<0.0001	0.0004
Homogeneity over Groups 1-4	0.3139	0.3185	0.1937	0.2267
Trend over Groups 1-4	0.9543	0.9480	0.8411	0.9119

Note that in both genders, over all dose groups, the Wilcoxon and log rank tests of homogeneity in survival were statistically significant (one $p \leq 0.0004$, and the rest $p < 0.0001$). It is clear from the Kaplan-Meier curves in Appendix 1, as well as from Tables 6 and 7 below, that this lack of homogeneity is almost solely due to the large number of events in the active control, group 5, TPA. Only groups 1-4 involved Capcaisin and its vehicle, so a test of trend over dose is only interpretable in these four dose groups. These four groups are henceforth also labeled as the Capcaisin groups. In these dose groups, none of the Wilcoxon and log rank tests of homogeneity in survival and in time to detection of first tumor groups were statistically significant (all $p \geq 0.1937$). As is often noted, absence of proof is not proof of absence, but here the consistency of results seems to be fairly strong evidence of no differences in patterns of survival between these four treatment groups. Results are even stronger for lack of simple linear trend in these four groups (all $p \geq 0.8411$).

Tumorigenicity analysis in Tg.AC mice is traditionally based on papilloma counts, particularly at the site of application (SOA) and non site of application (NSOA). The Sponsor originally supplied data on the number of animals in the various treatment groups with any tumor, including post mortem tumors. The data were analyzed using so-called poly-k tests, which modify the original Cochran-Armitage test of dose related trend in an event to adjust for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). One problem with typical frequentist tumorgenicity analyses is that for each tumor-organ-gender-study combination there is one test of significance for each comparison of an actual treatment group to controls plus a test of overall trend. As discussed in section 1.3.1.4 below, this usually requires a multiplicity adjustment, typically the Haseman-Lin-Rahman rules. However with the small number of different types of tumors observed here, it is not clear if the multiplicity adjustment is necessary. The following Table 2 displays all tests of trend and pairwise comparisons that are statistically significant. For each dose group, the tumor incidence is the number of animals where histopathological analysis detected a tumor. The column labeled "Trend" provides the observed p-value of the tests of trend over the vehicle controls, low,

medium, and high dose groups, i.e., the Capcaisin Groups 1-4. The columns labeled "HvsV", "MvsV", and "LvsV" provide the p-values of the corresponding pairwise tests of tumor incidence in each of the High, Medium, and Low dose groups versus the vehicle group (i.e., groups 2-4 versus group 1). The columns labeled "VvsN", "LvsN", "MvsN", and "HvsN" provide the p-values of the corresponding pairwise tests of tumor incidence in each of the vehicle, low, medium, and high dose groups versus the no treatment group (i.e., groups 1-4 versus group 7).

Table 2. Potentially Statistically Significant Neoplasms Based on the Number of Animals with Tumor.

Organ	Tumor	Incidence						P-values							
		Veh	Low	Mid	Hi	Lido	TPA	None	Trend	VvsH	VvsM	VvsL	VvsN	LvsN	MvsN	HvsN
Males																
Treated Skin	B-Papilloma, Squamous Cell															
		1	2	6	3	1	21	0	.1615	.2890	.0491	.5000	.5000	.2449	.0082	.1092
Females																
Any	Papilloma															
		2	7	10	16	8	29	9	.0002	.0004	.0353	.0780	.9573	.5124	.5400	.0776
Skin/SubQ, Other	B-Papilloma, Squamous Cell															
		2	2	4	8	3	8	5	.0061	.0301	.3326	.6388	.7742	.7353	.4439	.2045
Stomach, Nongl	B-Papilloma, Squamous Cell															
		0	4	4	6	4	2	3	.0156	.0094	.0543	.0383	.8752	.4255	.4513	.1901

In males the pairwise comparison for the number of animals (6) with papilloma in treated skin in the medium dose group is statistically significantly higher than the no treatment group (p=0.0082) and barely higher in the vehicle treatment group (p=0.0491). In group1-group 4 females there is a clear trend in any papillomas over dose (p=0.0002), while the test of differences between vehicle and the high dose group and the medium dose group is also statistically significant (p=0.0004 and p=0.0353, respectively). In Skin/SubQ the test of trend in benign papillomas over dose is statistically significant (p=0.0061), as is the pairwise test between vehicle and the high dose group (p=0.0301). In the Stomach, Nongl the test of trend in benign papillomas over dose is also statistically significant (p=0.0156), as are the pairwise tests between vehicle and the high and low dose groups (p=0.0094 and p=0.0383, respectively). The pairwise test between vehicle and the medium dose group is nearly statistically significant (p=0.0543). Note that the corresponding pairwise tests with the no-treatment group were not quite statistically significant at the usual 0.05 level (all $p \geq 0.0776$). This is clearly due to the large number of papillomas in the no treatment group. Particularly with the relatively small sample sizes in this study (25 animals) it seems that the evidence for a trend over several treatment groups is much stronger than the evidence in a single treatment group like the untreated group. However, there is no consistency in results across genders.

Kaplan-Meier curves and the results from tests for time to first detection of any tumor are included in Appendix 1, summarized with the following results:

Table 3. Statistical Significances of Tests of Homogeneity and Trend in Time to First Tumor Detection

	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Homogeneity over Groups 1-7	<0.0001	<0.0001	<0.0001	<0.0001
Homogeneity over Groups 1-4	0.0563	0.0645	0.0019	0.0038
Trend over Groups 1-4	0.4260	0.4229	0.0064	0.0024

Note that in both genders, over all dose groups, the Wilcoxon and log rank tests of homogeneity in time to detection of first tumor over all dose groups are highly statistically significant (all four $p < 0.0001$). From the Kaplan-Meier plots in figures A.1.3 and A.1.4 in Appendix 1, it appears that in male mice the time to first tumor seems roughly similar in the medium and high dose groups, both tending to be slightly higher than in the otherwise fairly similar vehicle and low dose groups. However the tests of differences in male mice are only borderline statistically significant (Log rank p=0.0563, Wilcoxon p=0.0645), with no evidence of a trend (both $p \geq 0.4229$). In female mice there is a strong evidence of lack of homogeneity over dose in time to detection of first tumor (both $p \leq 0.0038$), with good evidence of a decrease in time to detection in dose (both $p \leq 0.0064$).

During the study the animal technician identified observable masses over time, similar to the usual identification of papillomas. However, based on the toxicologists/pathologists retrospective identification in the animals with tumor data set reported in Table 2 above, it is apparent that very few masses are identified as actual papillomas, while a few papillomas were not identified as masses. The counts of masses are summarized in Appendix 3. As discussed in Appendix 4, initially a number of attempts were made to analyze Dunson's model (2000) for incidence of masses. However, this model did not fit. A model similar to Dunson's model, also discussed in Appendix 4, that does seem to fit the data is to use a Poisson model to directly analyze the number of masses or the maximum number of masses over the weeks. This model includes linear effects of week and dose, plus terms to reflect the differences between the Capcaisin groups 1-4 and the the no treatment group 7, and finally a random effect for the individual animal that is assumed to follow an autoregressive error structure over weeks. In female mice there is a statistically significant increasing effect due to dose (simple count of masses p=0.0177, maximum number of masses p=0.0310). Although the estimated dose effects in male mice are greater than zero, the tests that these parameters are 0 are not statistically significant (count of masses p=0.2077 and maximum number of masses p=0.2116). So there is statistically significant evidence of a positive dose effect on the number of masses in females but not in males.

Later, the Sponsor provided data that represent the Papilloma counts over time. As discussed in Section 2.2 below, there may be some question about this data. The same simplified Poisson model described in Appendix 4 was used to model the site of application (SOA) and non site of application (NSOA) papilloma counts. There was not sufficient data to

model the SOA papilloma counts in males. However, it is apparent from the SOA papilloma summary tables in Appendix 5 that there is no dose related trend in male mice. Again an autoregressive error structure within each individual mouse was assumed in female mice and male mice at the NSOA. Detailed results for this model are presented at the end of Appendix 5. As in the analysis of masses, primary interest focuses on the linear effect of dose. In male mice there is no evidence of a dose related trend (NSOA p=0.4429). However, in female mice there does seem to be evidence of such a positive trend (SOA p=0.0173, NSOA p<0.0001).

1.2. Brief Overview of the Studies

This submission consisted of one 26 week Tg.AC mouse study:

(b) (4) **26-Week Dermal Oncogenicity Study with trans-Capsaicin in Tg.AC Hemizygous Mice (FVB/N)**

1.3. Statistical Issues and Findings

1.3.1. Statistical Issues

In this section, several issues in the analysis of this data are considered. These issues include details of the survival analyses, tests on tumorigenicity, and the multiplicity of tests on neoplasms, and the validity of the designs.

1.3.1.1. Survival Analysis:

The analysis presented here is based on both the log rank test and the Wilcoxon test. The log rank test incorporates a weight that is equal across all events, while the Wilcoxon includes a weight that ranks events by their time, so that latter events are weighted higher than earlier events. Both tests were used to test both homogeneity of survival among the treatment groups and the effect of dose on trend in survival. 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 more detail. The results of the Sponsor's analyses are summarized in Sections 3.2.1.1. and 3.2.2.1.

1.3.1.2. Tumorigenicity Endpoints:

There was some discussion among the toxicologists about the appropriate endpoints for analysis. In most rodent studies, including those with genetically modified mice, the usual endpoints are the presence or absence of the specified tumors at any point during the study. For Tg.AC mice the usual endpoints are weekly papilloma counts. Further, in this particular study the animal technician recorded detectable masses. Later, the Sponsor added a data set containing papilloma counts. However, this reviewer has some concerns about this data (please see section 2.2). Although some results on papilloma counts are provided in Appendix 5, it may be appropriate to emphasize results on the presence or absence of tumors or the count of

detectable masses (please see Appendices 2-4 for details on these endpoints). In addition, Appendix 1 includes an analysis of time to first detection of a tumor.

1.3.1.3. Tests on Neoplasms:

Appendix 2 presents the results from the FDA poly-k analysis on the number of animals with the specified tumor. The poly-k test modifies the original Cochran-Armitage test of trend on response to adjust for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). Up until recently, the Division has usually recommended so-called Peto tests, which require accurate specification of cause of death. It was noted in the report of the Society of Toxicological Pathology "town hall" meeting in June 2001 that the poly-k modifications of the Cochran-Armitage tests of trend have been recommended over the Peto tests.

1.3.1.4. Multiplicity of Tests on Neoplasms:

In two year studies, testing the various neoplasms involves a large number of statistical tests, necessitating an adjustment in experiment-wise Type I error. In the usual FDA analysis of a submission with studies in two species the nominal significance levels of the resulting tests are then assessed using the Haseman-Lin-Rahman rules. For a roughly 0.10 (10%) overall false positive error rate, test of trend in rare tumors should be tested at a 0.025 level, and common tumors (with a historical control incidence greater than 1%) at a 0.005 level. The corresponding tests comparing the high dose group to control should be tested at a 0.05 level for rare tumors and 0.01 for common tumors. This approach is intended to balance both Type I error and Type II error (i.e., the error of concluding there is no evidence of a relation to tumorgenicity when there actually is such a relation). However it is not clear if these rules apply to studies in Tg.AC mice. Because of the relatively small number of tests involved here only the nominal significance levels of 0.05 are used.

1.3.2. Statistical Findings

Please see Section 1.1 above.

2. INTRODUCTION

2.1. Overview

Results from *trans*-Capsaicin administered weekly via topical application to the dorsal skin of Tg.AC mice for 26 weeks were submitted.

2.2. Data Sources

Eight SAS transport files, were originally provided by the Sponsor and placed in the CDER electronic data room (edr). These each contained a SAS data set with the same prefix but the extension "sas7bdat."

food.xpt mass.xpt signs.xpt macro.xpt
weights.xpt micro.xpt tumor.xpt mortal.xpt
Only the mortal.xpt, mass.xpt, and tumor.xpt data set are used in this analysis.

Much later the Sponsor added a transport file papill.xpt containing the SAS data set pap.sas7bdat. This was described as counts of "the number of skin papillomas by study week and body area." However, among males only one animal in the vehicle group, group 1, had non site of application (NSOA) counts of two papillomas and one animal in the TPA active control, group 6, had site of application (SOA) counts of two papillomas. Among female mice, only one mouse in the TPA group was credited with a single instance of two papillomas. So in almost all cases, the original per animal data were scored as either 0 or 1, including Group 6, the active control. This seems like a low number of papillomas, particularly in the active control group and may indicate data problems. From some comments on this study, it may be that papilloma counts were actually retrospective estimates. The Sponsor also included a breakdown of the observed incidence at various sites other than the site of application. While the vast majority of cases do match, in a few cases incidence in these sites does not match reported incidence in the supposed over all non sites of application. This also may be an indication of data problems. For these reasons this reviewer would recommend care in the interpretation of results from the papilloma count data set.

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

NA

3.2. Evaluation of Safety

3.2.1 (b) (4) Study 7215-150: 26-Week Dermal Oncogenicity Study with trans-Capsaicin in Tg.AC Hemizygous Mice (FVB/N)

STUDY DURATION: 26 Weeks (Although data extend to 27 weeks)
STARTING DOSING DATE: 16 May 2005
LAST DOSING DATE: 23 November 2005
STARTING TERMINAL SACRIFICE: 23 November 2005
RAT STRAIN: Tg.AC Hemizygous Mice
ROUTE: Dermal

3.2.1.1. Sponsor's Results and Conclusions

This section will present a summary of the Sponsor's analysis on survivability and tumorigenicity in Tg.AC mice.

Survival analysis:

The Sponsor states that "The numbers of males and females surviving to terminal sacrifice were lowest for those given the positive control article, with 19 of 25 males and 17 of 25 females either dead or moribund prior to terminal sacrifice . . . . Among mice given the vehicle control or test article, early deaths and moribund sacrifices ranged from 1 of 25 (vehicle control males) to 9 of 25 (0.64 mg/animal/week females). The cause of early death/moribundity was not determined from microscopic examination in every instance. However, among males and females given the positive control article, early death/moribundity was often considered related to large numbers of skin papillomas." (page 16 of report) The Sponsor claims that exposure to the test article was not a causative factor for early death/moribund sacrifice.

Tumorigenicity analysis:

The Sponsor summarizes this as follows: "Dermal application of *trans*-Capsaicin to male and female Model TGAC-T (hemizygous), FVB/NTac-Tg(v-Ha-ras)TG.ACled mice for 26-weeks resulted in no increased incidence of pre-neoplastic or neoplastic skin lesions (potential pre-neoplastic lesions consist of changes associated with hyperplasia and occasionally chronic-active inflammation). In contrast, over half the male and female mice exposed to the positive control article, Tetradecanoyl phorbol-13-acetate (TPA) had multiple skin papillomas; the majority of the positive control animals died early or were sacrificed in a moribund condition. Spontaneously-occurring neoplasms in the TGAC mice were not appreciably increased in *trans*-Capsaicin-treated animals." (page 9 of report)

"Therefore, under the conditions of the FDA-preapproved protocol design, *trans*- Capsaicin is considered to be negative in this Tg.AC mouse dermal transgenic study." (page 9 of report)

3.2.1.2. FDA Reviewer's Results

This section will present the Agency findings on survival and tumorigenicity in male and female rats.

Survival analysis:

The following tables (Table 4 for male rats, Table 5 for female rats) summarize the mortality results for the dose groups. The data were grouped for the specified time period, and present the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage cited is the percent that survived at the end of the interval.

Table 4. Summary of Male Rats Survival (dosed at mg/kg/day)

Period (Weeks)	1. Vehicle 0 mg/kg	2. Low 10 mg/kg	3. Medium 40 mg/kg	4. High 80 mg/kg	5. Lidocane mg/kg	6. TPA	7. No Treatment
1-10	0/25[1] 100%[2]	2/25 92%	2/25 92%	0/25 100%	1/25 96%	0/25[1] 100%	0/25 100%
11-15	1/25 96%	0/23 92%	0/23 92%	1/25 96%	1/24 92%	0/25 100%	0/25 100%
16-20	0/24 96%	2/23 84%	0/23 92%	1/24 94%	1/23 88%	10/25 60%	1/25 96%
21-25	0/24 96%	0/21 84%	0/23 92%	0/23 94%	0/22 88%	7/15 32%	0/24 96%
26-27	0/24 96%	1/21 80%	1/23 88%	0/23 94%	0/22 88%	2/8 24%	0/24 96%
Terminal	24	20	22	23	22	6	24

[1] number deaths / number at risk
[2] per cent survival to end of period.

In these tables the terminal period includes only those animals were sacrificed. Animals that died of other causes during the terminal period are included in the preceding, but overlapping time period.

Table 5. Summary of Female Rats Survival (dosed at mg/kg/day)

Period (Weeks)	1. Vehicle 0 mg/kg	2. Low 10 mg/kg	3. Medium 40 mg/kg	4. High 80 mg/kg	5. Lidocane mg/kg	6. TPA	7. No Treatment
1-10	1/25[1] 96%[2]	3/25 88%	2/25 92%	4/25 84%	3/25 88%	2/25[1] 92%	0/25 100%
11-15	1/24 92%	2/22 80%	0/23 92%	1/21 80%	2/22 80%	1/23 88%	0/25 100%
16-20	1/23 88%	0/20 80%	1/23 88%	0/20 80%	2/20 72%	6/22 64%	1/25 96%
21-25	1/22 84%	3/20 68%	0/22 88%	0/20 80%	0/18 72%	6/16 40%	2/24 88%
26-27	0/21 84%	1/17 64%	0/22 88%	0/20 80%	0/18 72%	2/10 32%	0/22 88%
Terminal	21	16	22	20	18	8	22

[1] number deaths / number at risk
[2] per cent survival to end of period.

The statistical significances of the tests of differences in survival across treatment groups using the log rank and the so-called Wilcoxon test are given in Table 6 below. One might note that the log rank tests puts equal weight on all events, while the Wilcoxon test weights them by the square of the time rank, and thus places more weight on later events than

does the log rank test. So the Wilcoxon test will be more sensitive to later separation of events than the log rank test.

Table 6. Statistical Significances of Tests of Homogeneity and Trend in Survival

	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Homogeneity over Groups 1-7	<0.0001	<0.0001	<0.0001	0.0004
Homogeneity over Groups 1-4	0.3139	0.3185	0.1937	0.2267
Trend over Groups 1-4	0.9543	0.9480	0.8411	0.9119

In both genders, over all dose groups, the Wilcoxon and log rank tests of homogeneity in survival over dose groups were statistically significant (one $p = 0.0004$, and the rest < 0.0001). It is clear from the Kaplan-Meier curves, figures A.1.1 and A.1.2 in Appendix 1, this evidence for the lack of homogeneity is primarily due to the large number of events in the active control, group 6, TPA. Only groups 1-4 involved Capcaisin and its vehicle. In these dose groups 1-4, none of the Wilcoxon and log rank tests of homogeneity in survival and in time to detection of first tumor were statistically significant (all $p \geq 0.1937$). As is often noted absence of proof is not proof of absence, but here the consistency of results seems to be fairly strong evidence of no differences between these four treatment groups. Results are even stronger for lack of trend in survival over the four dose groups (all four $p \geq 0.8411$).

Tumorigenicity analysis:

Table 7 below shows the tumors that had at least one mortality adjusted test on the presence or absence of the specified tumor with a nominal statistical significance of at least 0.05. More complete tables are presented in Appendix 2. For each dose group, the tumor incidence is the number of animals where a tumor was detected. The column labled “Trend” provides the observed p-value of the tests of trend over the vehicle controls, low, medium, and high dose groups, i.e. Groups 1-4. The columns labeled “HvsV”, “MvsV”, and “LvsV” provide the p-values of the corresponding pairwise tests of tumor incidence in each of the High, Medium, and Low dose groups versus the vehicle group (i.e. groups 2-4 versus group 1). The columns labeled “VvsN”, “LvsN”, “MvsN”, and “HvsN” provide the p-values of the corresponding pairwise tests of tumor incidence in each of the vehicle, low, medium, and high dose groups versus the no treatment vehicle group (i.e. groups 1-4 versus group 7). Note that the period ‘.’ denotes pairwise tests of dose groups with no tumors in either group.

Table 7. Potentially Statistically Significant Neoplasms

Organ	Tumor													
Incidence							P-values							
	Veh	Low	Mid	Hi	Lido	TPA	None	Trend	VvsH	VvsM	VvsL	VvsN	LvsN	MvsN	HvsN
Males														
Treated Skin						B-Papilloma, Squamous Cell								
	1	2	6	3	1	21	0	.1615	.2890	.0491	.5000	.5000	.2449	.0082	.1092
Females														
Any Skin						Papilloma								
	2	7	10	16	8	29	9	.0002	.0004	.0353	.0780	.9573	.5124	.5400	.0776
Skin/SubQ, Other						B-Papilloma, Squamous Cell								
	2	2	4	8	3	8	5	.0061	.0301	.3326	.6388	.7742	.7353	.4439	.2045
Stomach, Nongl						B-Papilloma, Squamous Cell								
	0	4	4	6	4	2	3	.0156	.0094	.0543	.0383	.8752	.4255	.4513	.1901

In males the pairwise comparison for the number of animals (6) with papilloma in treated skin in the medium dose group is statistically significantly higher than the no treatment group (p=0.0082) and barely higher in the vehicle treatment group (p=0.0491). In group 1-group 4 females there is a clear trend in papillomas over the four Capsaicin doses (with vehicle) dose (p=0.0002), while the test of differences between vehicle and the high dose group and the medium dose group is also statistically significant (p=0.0004 and p=0.0353, respectively). In Skin/SubQ the test of trend in benign papillomas over dose is statistically significant (p = 0.0061), as is the pairwise test between vehicle and the high dose group (p=0.0301). In the Stomach, Nongl the test of trend in benign papillomas over dose is also statistically significant (p=0.0156), as are the pairwise tests between vehicle and the high and low dose groups (p= 0.0094 and p=0.0383, respectively). The pairwise test between vehicle and the medium dose group is nearly statistically significant (p=0.0543). Note that the corresponding pairwise tests with the no-treatment group are not quite statistically significant at the usual 0.05 level (all $p \geq 0.0776$). This is clearly due to the large number of papillomas in the no treatment group. Particularly with the relatively small sample sizes in this study (25 animals), it seems that the evidence for a trend over several treatment groups is much stronger than the evidence in a single treatment group like the untreated group. Note that there is no consistency across genders, which may lead one to question this result. However, this is a decision requiring the expertise of the toxicologist.

Table 8. Statistical Significances of Tests of Homogeneity and Trend in Time to First Tumor Detection

	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Homogeneity over Groups 1-7	<0.0001	<0.0001	<0.0001	<0.0001
Homogeneity over Groups 1-4	0.0563	0.0645	0.0019	0.0038
Trend over Groups 1-4	0.4260	0.4229	0.0064	0.0024

Note that in both genders, over all dose groups, the Wilcoxon and log rank tests of homogeneity in time to detection of first tumor over all dose groups was highly statistically significant (all four $p < 0.0001$). From the Kaplan-Meier plots in figures A.1.3 and A.1.4 in Appendix 1, it appears that in male mice the time to first tumor seems roughly similar in the medium and high dose groups, both tending to be slightly higher than in the otherwise fairly similar vehicle and low dose groups. However the tests of differences in male mice are only borderline statistically significant (Log rank p=0.0563, Wilcoxon p=0.0645), with no evidence of a trend (both $p \geq 0.4229$). In female mice there is a strong evidence of lack of homogeneity over dose in time to detection of first tumor (both $p \leq 0.0038$), with good evidence of a decrease in time to detection in dose (both $p \leq 0.0064$).

During the study the animal technician identified observable masses over time, similar to the usual identification of papillomas. However, based on the toxicologists/pathologists retrospective identification in the animals with tumor data set, it is apparent that very few masses are identified as actual papillomas, while a few papillomas were not identified as masses. These counts are summarized in Appendix 3. As discussed in Appendix 4, initially a number of attempts were made to analyze Dunson's model (2000) for incidence of masses. However, this model did not fit. A model similar to Dunson's model that does seem to fit the data is to directly model the number of masses or the maximum number of masses over the weeks. This model includes the linear effect of week and dose, plus terms to reflect the differences between Capcaisin groups 1-4 and the the no treatment group 7, and finally a random effect for animal that is assumed to follow an autoregressive error structure. In female mice there is a clear increasing effect due to dose (number of masses p=0.0177, maximum number of masses p=0.0310). Although the estimated dose effects in male mice are greater than zero, the tests that these parameters are 0 are not statistically significant (simple count of masses p=0.2077 and maximum count p=0.2116). So there is evidence of a dose effect in females but not in males.

Later the Sponsor provided data that represent the Papilloma counts over time. As discussed inn Section 2.2 below, there may be some question about this data. The same simplified Possin model described in Appendix 4 was used to model the site of application (SOA) and non site of application (NSOA) papilloma counts. There was not sufficient data to model the SOA papilloma counts in males. However, it is apparent from the papilloma summary tables in Appendix 5 that there is no dose related trend in male mice at the SOA. Again an autoregressive error structure within each individual mouse was assumed. Detailed results are presented in Appendix 5. As in the analysis of masses, primary interest focuses on the linear effect of dose. In male mice there is no evidence of a dose related trend (NSOA p=0.4429). However, in female mice, there does seem to be evidence of such a positive trend on papilloma counts (SOA p=0.0173, NSOA $p<0.0001$).

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

NA

5. SUMMARY AND CONCLUSIONS

5.1. Statistical Issues and Collective Evidence

Please see Section 1.3 above.

5.2. Conclusions and Recommendations

Please see section 1.1 above.

APPENDICES:

Appendix 1. Survival Analyses

Simple summary life tables in mortality are presented in the report (Tables 8 and 9), above. Kaplan-Meier estimated survival curves across dose groups for each gender in each study are displayed in Figures A.1.1 - A.1.2 below. These plots include 95% confidence intervals around each curve (colored area around each curve). In addition, Figures A.1.3-A.1.4 provide similar curves for time of first detection of any tumor. The plots are also supported by tests of homogeneity and trend in survival over dose groups. The statistical significances of the tests of differences in survival across treatment groups using the log rank and the so-called Wilcoxon test are given in Table A.1.1 below. One might note that the log rank tests puts equal weight on all events, while the Wilcoxon test weights them by the square of the time rank, and thus places more weight on later events than does the log rank test. So the Wilcoxon test will be more sensitive to later separation of events than the log rank test.

Table A.1.1. Statistical Significances of Tests of Homogeneity and Trend in Survival

	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Homogeneity over Groups 1-7	<0.0001	<0.0001	<0.0001	0.0004
Homogeneity over Groups 1-4	0.3139	0.3185	0.1937	0.2267
Trend over Groups 1-4	0.9543	0.9480	0.8411	0.9119

The statistical significances of the tests of differences across treatment groups in time to first tumor using the log rank and the so-called Wilcoxon test are given in Table A.1.2 below.

Table A.1.2. Statistical Significances of Tests of Homogeneity and Trend in Time to First Tumor Detection

	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Homogeneity over Groups 1-7	<0.0001	<0.0001	<0.0001	<0.0001
Homogeneity over Groups 1-4	0.0563	0.0645	0.0019	0.0038
Trend over Groups 1-4	0.4260	0.4229	0.0064	0.0024

Note that in both genders, over all dose groups, the Wilcoxon and log rank tests of homogeneity in survival and in time to detection of first tumor groups were statistically significant (all $p \leq 0.0004$, and most < 0.0001). For survival, it is clear from the Kaplan-Meier curves below that this lack of homogeneity is primarily due to the large number of events in the active control, group 5, TPA. Only groups 1-4 involved Capcaisin and its vehicle. In these dose groups 1-4, none of the Wilcoxon and log rank tests of homogeneity in survival were statistically significant (all $p \geq 0.1937$). As is often noted absence of proof is not proof of absence, but here the consistency of results seems to be fairly strong evidence of no differences

between these four treatment groups. Results are even stronger for lack of trend in these four groups (all $p \geq 0.8411$).

However, when analyzing time to first tumor there are statistically significant differences among the four trans-Capsaicin and vehicle groups. From the Kaplan-Meier plots in figures A.1.3 and A.1.4, it appears that in male mice the time to first tumor seems roughly similar in the medium and high dose groups, both tending to be slightly higher than in the otherwise fairly similar vehicle and low dose groups. However the tests of differences in male mice are only borderline statistically significant (Log rank p=0.0563, Wilcoxon p=0.0645), with no evidence of a trend (both $p \geq 0.4229$). In female mice there is a strong evidence of lack of homogeneity over dose in time to detection of first tumor (both $p \leq 0.0038$), with good evidence of a decrease in time to detection in dose (both $p \leq 0.0064$).

Figure A.1.1 Kaplan-Meier Survival Curves for Male Mice

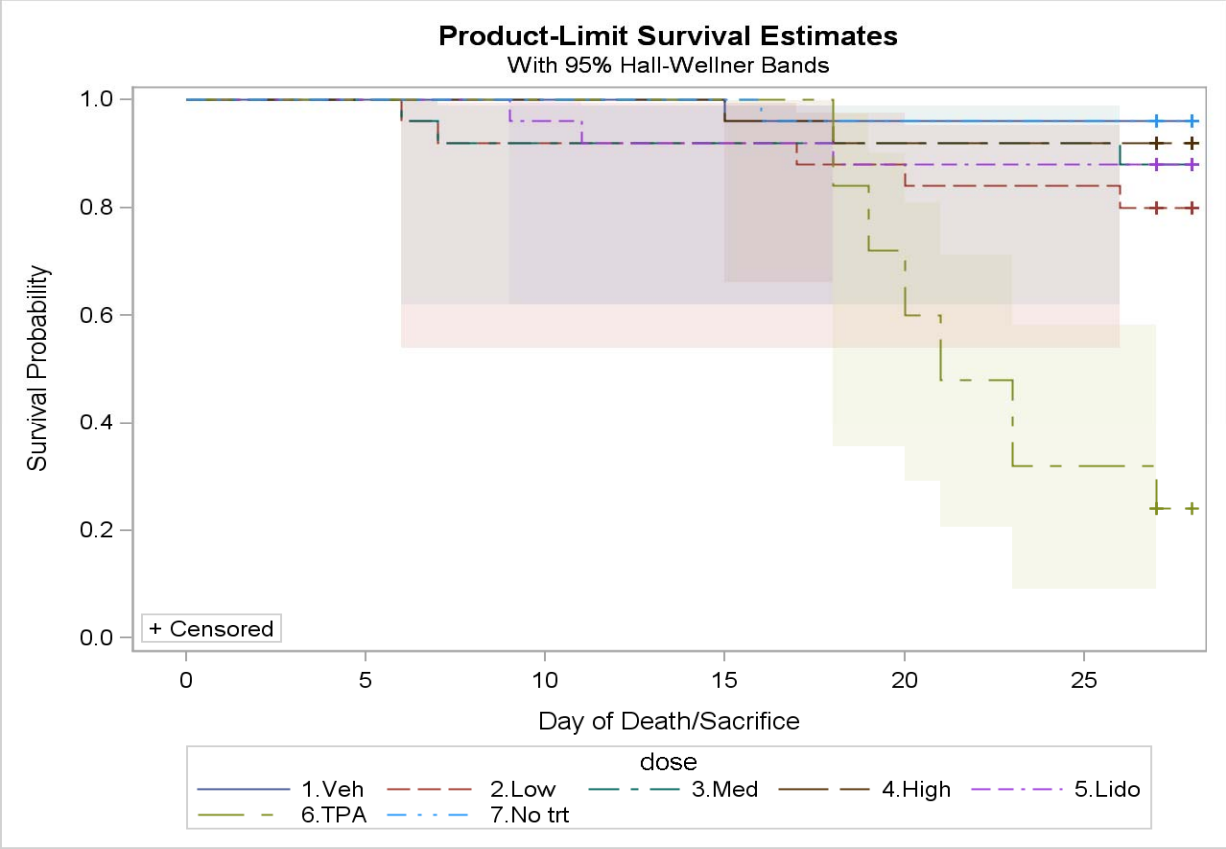


Figure A.1.2 Kaplan-Meier Survival Curves for Female Mice

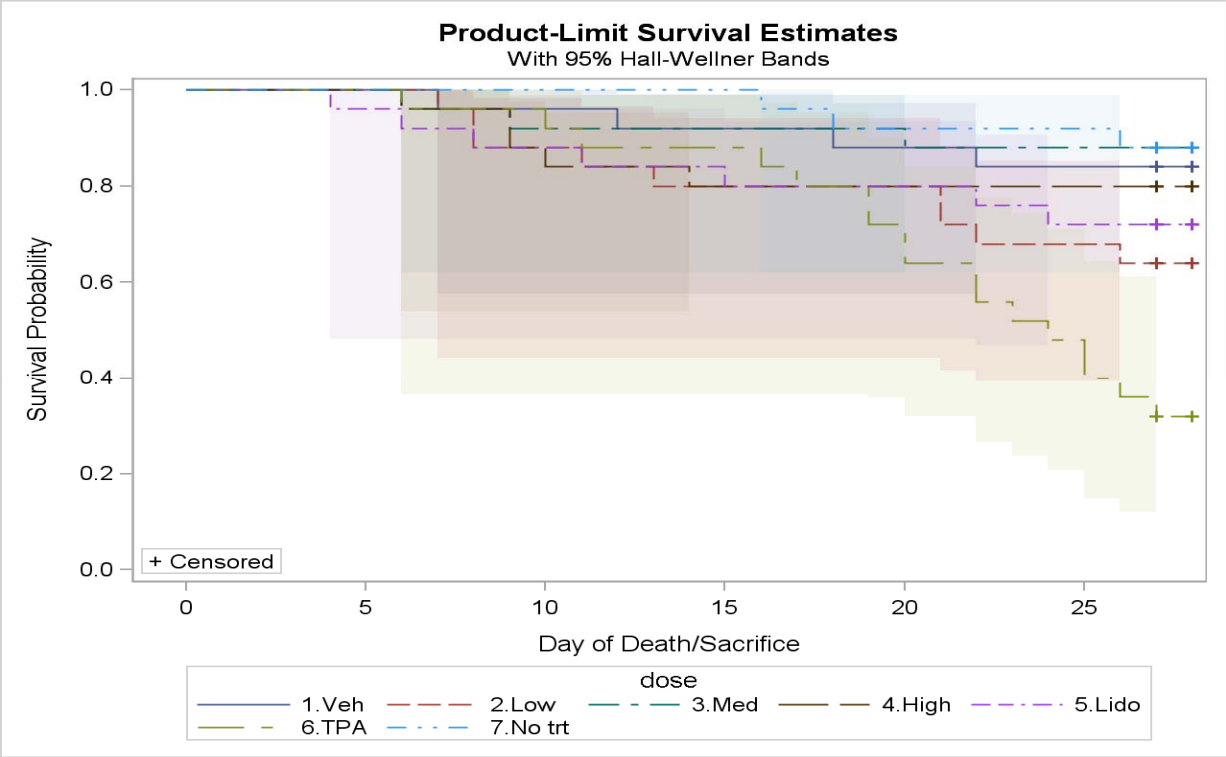


Figure A.1.3 Kaplan-Meier Curves for Time to Detection of First Tumor in Male Mice

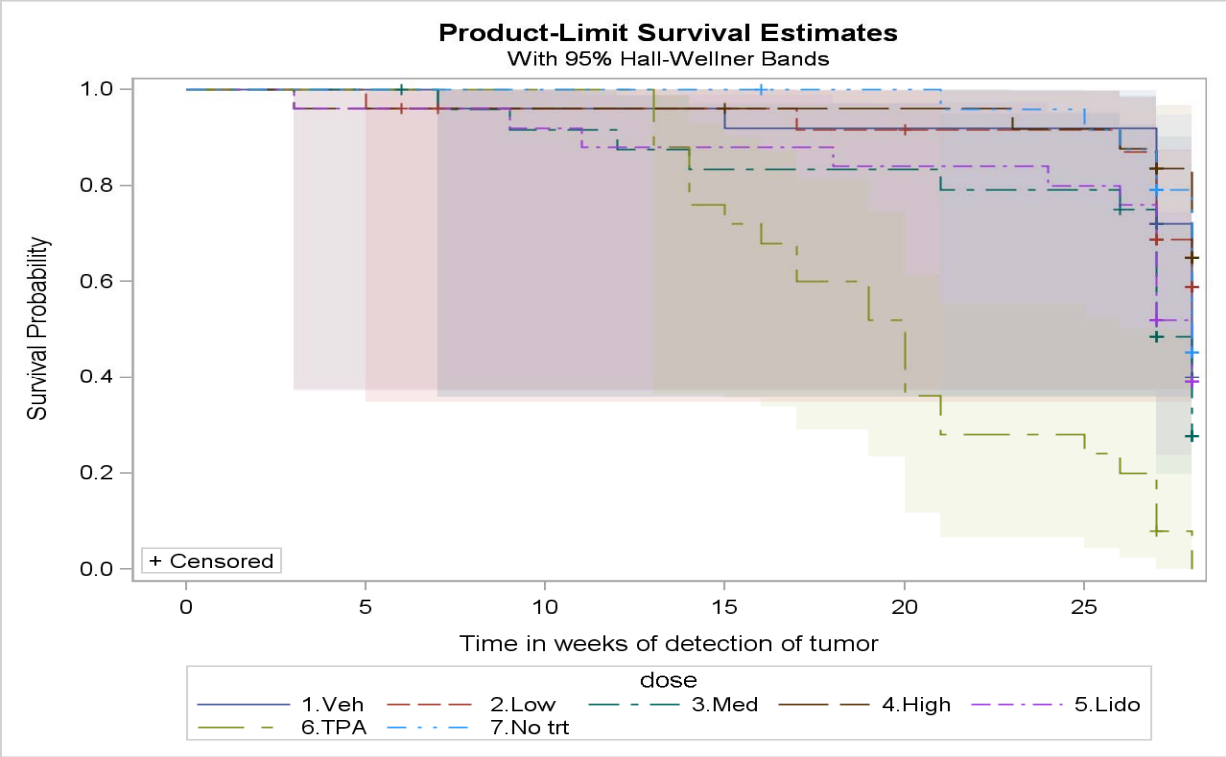
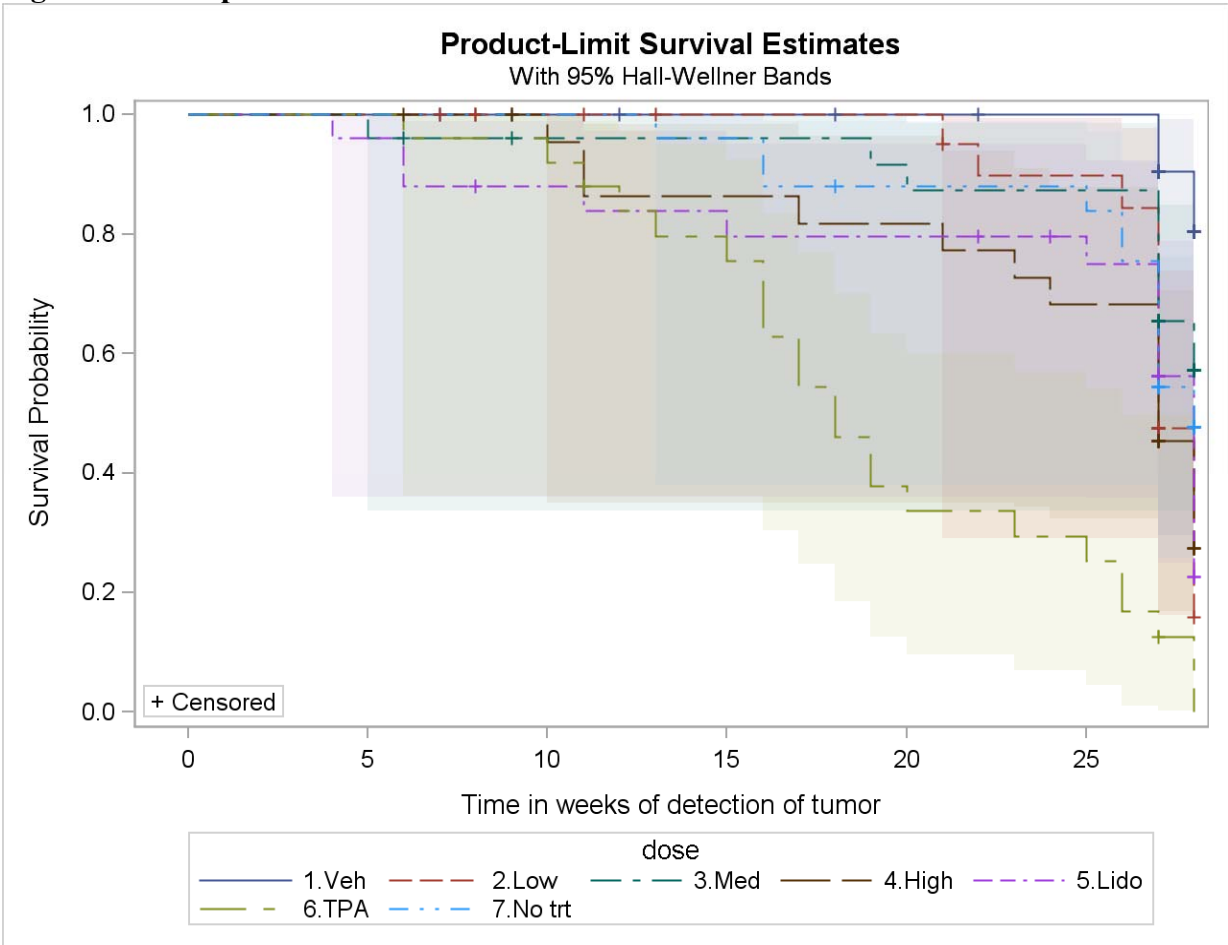


Figure A.1.4 Kaplan-Meier Curves for Time to Detection of First Tumor in Female Mice



Appendix 2. FDA Poly-k Tumorigenicity Analysis

Tables A.2.1 through A.2.3, below, display the number of animals with neoplasms by organ and tumor combination and the results of tests of trend over dose and the results of pairwise comparisons with the vehicle control (Group 1) and the no dose control (Group 7). For each dose group, the tumor incidence is the number of animals where a tumor was detected. The column labled "Trend" provides the observed p-value of the tests of trend over the vehicle controls, low, medium, and high dose groups, i.e. Groups 1-4. The columns labeled "HvsV", "MvsV", and "LvsV" provide the p-values of the corresponding pairwise tests of tumor incidence in each of the High, Medium, and Low dose groups versus the vehicle group (i.e. groups 2-4 versus group 1). The columns labeled "VvsN", "LvsN", "MvsN", and "HvsN" provide the p-values of the corresponding pairwise tests of tumor incidence in each of the vehicle, low, medium, and high dose groups versus the no treatment vehicle group (i.e., groups 1-4 versus group 7). Incidence in the TPA group, group 6, is used to verify the sensitivity of the mice. As with the TPA group, for the lidocane group (Group 5) only the incidence of animals with tumor is reported.

The poly-k (in these analyses k=3) tests modify the original Cochran-Armitage test to adjust for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). The p-values presented here are design based tests, and assume all marginal totals are fixed, a debatable assumption. Up until recently, when analyzing the number of animals with tumor, the Division has usually emphasized so-called Peto carcinogenicity tests, which require accurate specification of cause of death. It was noted in the report of the Society of Toxicological Pathology "town hall" meeting in June 2001 that the poly-k modification of the Cochran-Armitage tests of trend have been recommended over the Peto tests.

To adjust for the multiplicity of tests involved in a typical Peto analysis, the so-called Haseman-Lin-Rahman rules discussed in Section 1.3.1.3. are usually applied. However, as noted there it is not clear if those are applicable to this study. Hence only the nominal significance levels are cited.

Table A.2.1 shows the tumors that had at least one mortality adjusted test whose nominal statistical significance in either species was at least 0.05. Tables A.2.2 and A.2.3 show the overall results for male and female mice, respectively. Note that the period '.' denotes pairwise tests of dose groups with no tumors in either group.

Table A.2.1 Potentially Statistically Significant Neoplasms

Organ	Tumor	Incidence						P-values							
		Veh	Low	Mid	Hi	Lido	TPA	None	Trend	VvsH	VvsM	VvsL	VvsN	LvsN	MvsN	HvsN
Males																
Treated Skin	B-Papilloma, Squamous Cell															
		1	2	6	3	1	21	0	.1615	.2890	.0491	.5000	.5000	.2449	.0082	.1092
Females																
Any Skin	Papilloma															
		2	7	10	16	8	29	9	.0002	.0004	.0353	.0780	.9573	.5124	.5400	.0776
Skin/SubQ, Other	B-Papilloma, Squamous Cell															
		2	2	4	8	3	8	5	.0061	.0301	.3326	.6388	.7742	.7353	.4439	.2045
Stomach, Nongl	B-Papilloma, Squamous Cell															
		0	4	4	6	4	2	3	.0156	.0094	.0543	.0383	.8752	.4255	.4513	.1901

In males the pairwise comparison for the number of animals (6) with papilloma in treated skin in the medium dose group is statistically significantly higher than the no treatment group (p=0.0082) and barely higher in the vehicle treatment group (p=0.0491). In group 1-group 4 females there is a clear trend in any papillomas over dose (p=0.0002), while the test of differences between vehicle and the high dose group and the medium dose group were also statistically significant (p=0.0004 and p=0.0353, respectively). In Skin/SubQ the test of trend in females in benign papillomas over dose is statistically significant (p=0.0061), as is the pairwise test between vehicle and the high dose group (p=0.0301). In the Stomach, Nongl of females the test of trend in benign papillomas over dose is also statistically significant (p=0.0156), as are the pairwise tests between vehicle and the high and low dose groups (p=0.0094 and p=0.0383, respectively). The pairwise test between vehicle and the medium dose group is nearly statistically significant (p=0.0543).Note that the corresponding pairwise tests with the no-treatment group were not quite statistically significant at the usual 0.05 level (all $p \geq 0.0776$). This is clearly due to the large number of papillomas in the no treatment group. Particularly with the relatively small sample sizes in this study (25 animals) it seems that the evidence for a trend over several treatment groups is much stronger than the evidence in a single treatment group like the untreated group. Note that there is no consistency across genders, which may lead one to question this result. However, this is a decision requiring the expertise of the toxicologist.

Kaplan-Meier curves and the results from tests for time to first detection of any tumor are included in Appendix 1 and have outcomes virtually identical to the survival results summarized above.

Again, tables A.2.2 and A.2.3 below, show the overall results for male and female mice, respectively.

Table A.2.2 Tests of Trend and Pairwise Differences in Neoplasms in Males

Organ	Tumor	Incidence							P-values							
		Veh	Low	Mid	Hi	Lido	TPA	None	Trend	VvsH	VvsM	VvsL	VvsN	LvsN	MvsN	HvsN
Any Skin	Papilloma	9	4	14	6	10	36	9	.7969	.8870	.2836	.9018	.3812	.7519	.0989	.7275
Body, Whole/Cav	M-Leukemia, Erythrocytic	0	0	1	0	2	0	0	.4949	.	.5000	.	.	.	.4898	.
Body, Whole/Cav	M-Lymphosarcoma	0	1	1	0	1	0	0	.4898	.	.5000	.5000	.	.5000	.4792	.
Gl, Mandib Saliv	M-Carcinoma	0	0	0	0	1	0	0	.	.	.	.	.	.	.	.
Head	B-Fibroma	0	0	0	0	0	1	0	.	.	.	.	.	.	.	.
Head	B-Osteoma	1	0	0	0	0	0	0	.7475	.4898	.5000	.5000	.5000	.	.	.
Head	B-Papilloma, squamous	1	0	0	0	0	2	0	.7475	.4898	.5000	.5000	.5000	.	.	.
Head	M-Ameloblastoma	1	2	2	1	2	0	2	.4734	.7449	.5000	.5000	.5000	.6954	.6631	.4837
Head	M-Odontoma	0	1	0	0	0	0	0	.4949	.	.	.5000	.	.5000	.	.
Lung	B-Adenoma, Bronchiolar-Alveo	0	0	0	1	0	0	0	.2424	.4898	.	.	.	.	.	.4894
Skin/SubQ, Other	B-Acanthoma	0	0	1	0	0	0	0	.4949	.	.5000	.	.	.	.4792	.
Skin/SubQ, Other	B-Osteoma	1	0	0	0	0	0	0	.7475	.4898	.5000	.5000	.5000	.	.	.
Skin/SubQ, Other	B-Papilloma, Squamous Cell	1	0	4	1	3	8	6	.3663	.7449	.1743	.5000	.9509	.9889	.6099	.9453
Skin/SubQ, Other	M-Fibrosarcoma	0	0	1	1	0	0	0	.1869	.5000	.5000	.	.	.	.4792	.5000
Stomach, Nongl	B-Papilloma, Squamous Cell	6	2	4	2	6	5	3	.8659	.8636	.6374	.8766	.2317	.5000	.4513	.4791
Treated Skin	B-Papilloma, Squamous Cell	1	2	6	3	1	21	0	.1615	.2890	.0491	.5000	.5000	.2449	.0082	.1092
Treated Skin	M-Fibrosarcoma	0	1	0	0	0	0	0	.4949	.	.	.5000	.	.4783	.	.

Table A.2.3 Table A.2.2 Tests of Trend and Pairwise Differences in Neoplasms in Females

Organ	Tumor	Incidence						P-values							
		Veh	Low	Mid	Hi	Lido	TPA	None	Trend	VvsH	VvsM	VvsL	VvsN	LvsN	MvsN	HvsN
Any Skin	Papilloma	2	7	10	16	8	29	9	.0002	.0004	.0353	.0780	.9573	.5124	.5400	.0776
Body, Whole/Cav	M-Leukemia, Erythrocytic	0	1	0	0	2	4	1	.5000	.	.	.4634	.4898	.7225	.4792	.4792
Gl, Mandib Saliv	B-Adenoma	1	0	0	0	0	0	0	.7447	.4894	.5000	.4634	.4898	.	.	.
Gl, Mandib Saliv	M-Carcinoma	0	1	0	0	0	0	0	.5000	.	.	.4634	.	.4681	.	.
Gl, Zymbal's	B-Adenoma	0	0	0	1	0	0	0	.2447	.4894	.	.	.	.	.	.4792
Head	B-Osteoma	1	0	0	0	0	0	0	.7447	.4894	.5000	.4634	.4898	.	.	.
Head	B-Papilloma, squamous	0	0	0	0	0	2	0	.	.	.	.	.	.	.	.
Head	M-Ameloblastoma	0	3	1	1	1	1	0	.5068	.4894	.5000	.0909	.	.0950	.4792	.4792
Head	M-Fibrosarcoma	0	0	1	0	0	0	0	.5000	.	.5000	.	.	.	.4792	.
Head	M-Odontoma	0	1	1	1	2	0	3	.2893	.4894	.5000	.4634	.8752	.6454	.6631	.6631
Head	M-Osteosarcoma	0	2	0	0	0	0	0	.6209	.	.	.2085	.	.2137	.	.
Ovary	B-Cystadenoma	0	0	0	0	1	0	0	.	.	.	.	.	.	.	.
Ovary	B-Teratoma	0	0	0	0	0	0	1	.	.	.	.	.4898	.4681	.4792	.4792
Ovary	M-Malignant Granulosa/Theca	0	0	0	0	1	0	0	.	.	.	.	.	.	.	.
Ovary	M-Yolk Sac Carcinoma	0	0	0	0	0	1	0	.	.	.	.	.	.	.	.
Skin/SubQ, Other	B-Papilloma, Squamous Cell	2	2	4	8	3	8	5	.0061	.0301	.3326	.6388	.7742	.7353	.4439	.2045
Stomach, Nongl	B-Papilloma, Squamous Cell	0	4	4	6	4	2	3	.0156	.0094	.0543	.0383	.8752	.4255	.4513	.1901
Treated Skin	B-Papilloma, Squamous Cell	0	1	2	2	0	17	1	.1280	.2220	.2449	.4634	.4898	.7225	.5000	.4511
Untreated Skin	B-Papilloma, Squamous Cell	0	0	0	0	1	0	0	.	.	.	.	.	.	.	.

Appendix 3. FDA Analysis of Technician Identified Masses

The following tables summarizes the Sponsor supplied information on identified masses for each week. The rows "# risk" and "# w/ mass" give counts of the number of animals at risk (i.e. still alive at the beginning of the week) and the number with a tumor. The row for "Ovall Mean" is the mean number of tumors over all animals at risk. The row for "Nzro Mean" lists the mean number of tumors over all animals with at least one tumor. Thus the total number of tumors in the dose group is # w/mass times the Nzro Mean (or # risk times Ovall Mean).

Table A.3.1 Incidence of Masses in Male Mice

Dose Group	Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	# risk	25	25	25	25	25	25	25	25	25	25	25	25	25	25	24
	# w/mass	0	0	0	1	1	2	2	1	1	1	1	2	1	1	1
	Ovall Mean	0.00	0.00	0.00	0.04	0.04	0.08	0.08	0.04	0.04	0.04	0.04	0.08	0.04	0.04	0.04
	Nzro Mean	.	.	.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	# risk	25	25	25	25	25	24	23	23	23	23	23	23	23	23	23
	# w/mass	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
	Ovall Mean	0.00	0.00	0.00	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
	Nzro Mean	.	.	.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	# risk	25	25	25	25	25	24	23	23	23	23	23	23	23	23	23
	# w/mass	0	0	0	2	2	2	2	3	2	2	3	3	4	5	5
	Ovall Mean	0.00	0.00	0.00	0.08	0.08	0.08	0.09	0.13	0.09	0.09	0.13	0.13	0.17	0.22	0.22
	Nzro Mean	.	.	.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	# risk	25	25	25	25	25	25	25	25	25	25	25	25	25	25	24
	# w/mass	0	1	1	1	1	1	1	1	1	1	1	1	2	2	2
	Ovall Mean	0.00	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.08	0.08	0.08
	Nzro Mean	.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	# risk	25	25	25	25	25	25	25	25	24	24	23	23	23	23	23
	# w/mass	0	1	2	2	1	1	1	1	1	1	1	1	1	1	1
	Ovall Mean	0.00	0.04	0.08	0.08	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
	Nzro Mean	.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	# risk	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
	# w/mass	0	0	0	0	2	2	4	5	5	7	8	13	16	16	17
	Ovall Mean	0.00	0.00	0.00	0.00	0.08	0.08	0.16	0.20	0.20	0.28	0.36	1.40	2.40	3.24	3.44
	Nzro Mean	.	.	.	.	1.00	1.00	1.00	1.00	1.00	1.00	1.13	2.69	3.75	5.06	5.06
7	# risk	25	25	25	25	25	25	25	25	25	25	25	25	25	25	24
	# w/mass	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
	Ovall Mean	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.00	0.00	0.00	0.00	0.00
	Nzro Mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.	.	.	.	.

Table A.3.1 (cont.) Incidence of Masses in Male Mice

Dose Group	Week	16	17	18	19	20	21	22	23	24	25	26	27
1	# risk	24	24	24	24	24	24	24	24	24	24	24	20
	# w/mass	1	1	1	1	1	1	1	1	2	2	4	5
	Ovall Mean	0.04	0.04	0.04	0.04	0.04	0.13	0.13	0.13	0.21	0.21	0.21	0.35
	Nzro Mean	1.00	1.00	1.00	1.00	1.00	3.00	3.00	3.00	2.50	2.50	1.25	1.40
2	# risk	23	22	22	22	21	21	21	21	21	21	20	17
	# w/mass	1	2	2	2	2	3	4	4	4	4	4	3
	Ovall Mean	0.04	0.09	0.09	0.09	0.10	0.14	0.19	0.19	0.19	0.19	0.30	0.35
	Nzro Mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.50	2.00
3	# risk	23	23	23	23	23	23	23	23	23	22	22	21
	# w/mass	5	5	5	5	8	7	8	8	5	6	11	12
	Ovall Mean	0.22	0.22	0.22	0.22	0.35	0.30	0.39	0.43	0.30	0.36	0.64	0.86
	Nzro Mean	1.00	1.00	1.00	1.00	1.00	1.00	1.13	1.25	1.40	1.33	1.27	1.50
4	# risk	24	24	23	23	23	23	23	23	23	23	23	21
	# w/mass	2	1	0	0	1	1	2	2	2	4	5	8
	Ovall Mean	0.08	0.04	0.00	0.00	0.04	0.04	0.13	0.13	0.13	0.22	0.26	0.62
	Nzro Mean	1.00	1.00	.	.	1.00	1.00	1.50	1.50	1.50	1.25	1.20	1.63
5	# risk	23	23	22	22	22	22	22	22	22	22	22	19
	# w/mass	1	2	1	1	1	1	1	3	3	5	5	5
	Ovall Mean	0.04	0.09	0.05	0.05	0.05	0.05	0.05	0.14	0.14	0.23	0.27	0.47
	Nzro Mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.20	1.80
6	# risk	25	25	21	18	15	12	12	8	8	8	8	6
	# w/mass	19	21	17	14	12	9	9	5	5	6	6	5
	Ovall Mean	4.16	4.92	4.86	5.50	5.33	4.67	5.25	3.50	4.25	4.75	5.13	3.50
	Nzro Mean	5.47	5.86	6.00	7.07	6.67	6.22	7.00	5.60	6.80	6.33	6.83	4.20
7	# risk	24	24	24	24	24	24	24	24	24	24	24	22
	# w/mass	0	0	0	0	2	2	2	2	2	3	3	10
	Ovall Mean	0.00	0.00	0.00	0.00	0.08	0.08	0.08	0.08	0.08	0.13	0.13	0.64
	Nzro Mean	.	.	.	.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.40

Please recall that mice in group 1 were treated with the vehicle only, while mice in groups 2-4 are described as receiving the dose formulations containing the vehicle control, diethylene glycol monoethyl ether (DGME), and the test drug (trans-Capsaicin in DGME) at drug levels levels of 0.64, 1.28, and 2.56 mg/mouse/week. These groups are also labeled as the low, medium, and high dose groups. Group 6, the positive control, were administered Tetradecanoylphorbol-13-Acetate (TPA, in DGME). Group 5 animals received lidocaine only. Group 7 animals were untreated. All treated groups were dosed once/week except for the positive control group, group 6, dosed twice/week.

Table A.3.2 Incidence of Masses in Female Mice

Dose Group	Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	# risk	25	25	25	25	25	25	24	24	24	24	24	23	23	23	23
	# w/mass	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
Ovall Mean		0.00	0.00	0.00	0.00	0.04	0.04	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Nzro Mean		.	.	.	.	1.00	1.00	1.00	.	.	.	.	.	.	.	.
2	# risk	25	25	25	25	25	25	24	22	22	22	21	21	20	20	20
	# w/mass	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ovall Mean		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Nzro Mean		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3	# risk	25	25	25	25	25	24	24	23	23	23	23	23	23	23	23
	# w/mass	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
Ovall Mean		0.00	0.00	0.00	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Nzro Mean		.	.	.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	# risk	25	25	25	25	25	24	24	24	22	21	21	21	21	20	20
	# w/mass	0	0	0	0	0	0	0	0	0	3	3	3	3	2	1
Ovall Mean		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.14	0.14	0.14	0.10	0.05
Nzro Mean		.	.	.	.	.	.	.	.	.	1.00	1.00	1.00	1.00	1.00	1.00
5	# risk	25	25	25	24	24	23	23	22	22	22	21	21	21	20	20
	# w/mass	0	0	1	1	2	2	2	2	2	2	1	2	1	1	1
Ovall Mean		0.00	0.00	0.04	0.04	0.08	0.09	0.09	0.09	0.09	0.09	0.05	0.10	0.05	0.05	0.05
Nzro Mean		.	.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	# risk	25	25	25	25	25	24	24	24	24	23	22	22	22	22	22
	# w/mass	0	0	0	0	0	0	0	0	0	2	3	8	8	9	10
Ovall Mean		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.14	0.59	0.68	0.91	1.27
Nzro Mean		.	.	.	.	.	.	.	.	.	1.00	1.00	1.63	1.88	2.22	2.80
7	# risk	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
	# w/mass	0	0	0	0	0	0	0	0	0	0	0	1	2	2	2
Ovall Mean		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.08	0.08	0.12
Nzro Mean		.	.	.	.	.	.	.	.	.	.	.	1.00	1.00	1.00	1.50

Table A.3.2 (cont.) Incidence of Masses in Male Mice

Dose Group	Week	16	17	18	19	20	21	22	23	24	25	26	27
1	# risk	23	23	22	22	22	22	21	21	21	21	21	18
	# w/mass	1	2	2	2	2	3	3	3	3	5	6	8
	Ovall Mean	0.04	0.09	0.09	0.09	0.09	0.14	0.14	0.14	0.14	0.24	0.29	0.44
	Nzro Mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	# risk	20	20	20	20	20	18	17	17	17	17	16	13
	# w/mass	0	1	1	1	1	0	0	0	0	0	4	6
	Ovall Mean	0.00	0.05	0.05	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.25	0.54
	Nzro Mean	.	1.00	1.00	1.00	1.00	.	.	.	.	.	1.00	1.17
3	# risk	23	23	23	23	22	22	22	22	22	22	22	19
	# w/mass	1	1	2	3	2	2	2	2	2	2	3	6
	Ovall Mean	0.04	0.04	0.09	0.13	0.09	0.09	0.14	0.14	0.18	0.14	0.18	0.37
	Nzro Mean	1.00	1.00	1.00	1.00	1.00	1.00	1.50	1.50	2.00	1.50	1.33	1.17
4	# risk	20	20	20	20	20	20	20	20	20	20	20	17
	# w/mass	2	2	2	2	3	3	5	6	6	6	10	10
	Ovall Mean	0.10	0.10	0.15	0.15	0.20	0.20	0.30	0.45	0.55	0.50	0.70	1.00
	Nzro Mean	1.00	1.00	1.50	1.50	1.33	1.33	1.20	1.50	1.83	1.67	1.40	1.70
5	# risk	20	20	20	20	20	20	19	19	18	18	18	15
	# w/mass	1	1	1	1	2	5	5	5	5	5	5	5
	Ovall Mean	0.05	0.05	0.05	0.05	0.10	0.25	0.26	0.26	0.28	0.28	0.28	0.47
	Nzro Mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.40
6	# risk	21	20	20	17	16	16	14	13	12	10	9	8
	# w/mass	10	13	14	12	11	12	10	10	9	8	8	7
	Ovall Mean	1.71	2.55	3.10	3.12	3.19	3.44	4.71	4.54	4.92	4.70	4.89	4.88
	Nzro Mean	3.60	3.92	4.43	4.42	4.64	4.58	6.60	5.90	6.56	5.88	5.50	5.57
7	# risk	24	24	23	23	23	23	23	23	23	23	22	19
	# w/mass	2	2	2	2	2	2	2	2	4	5	7	7
	Ovall Mean	0.13	0.13	0.09	0.09	0.09	0.09	0.09	0.09	0.17	0.22	0.32	0.42
	Nzro Mean	1.50	1.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.14

Note that a number of animals had more than one assessment during Week 27. The data above reflect the maximum number of masses observed at this time point.

Appendix 4. FDA Model for Technician Identified Masses

Dunson (2000) proposed a model for skin tumor counts in Tg.AC mice, which seems to be often accepted as a standard model for these mice. Following his notation, let Z_{ij} denote the number detectable masses on mouse i at time j. Define $M_{ij} = \max\{ Z_{i1}, \ldots, Z_{i,j-1}\}$, defined as the maximum tumor burden for mouse i before week j. He defines $Y_{ij} = M_{i,j+1} - M_{ij}$ and assumes the increase in counts Y_{ij} are distributed as Poisson(μ_{ij}) with

$$\mu_{ij} = \begin{cases} \exp(\beta_1 + (b_i + \gamma_1)t_j\, d_i) & where\ M_{ij} = 0 \\ \exp(\beta_2 + \gamma_2 d_i) & where M_{ij} > 0, \end{cases}$$

where b_i denotes a mouse-specific random susceptibility, assumed follow a normal distribution with mean 0 and variance σ^2. Other inputs involve t_j=j/T, where T is the duration of the study, and d_i is the dose on the log scale. Note that it may reasonable to define d_i as the logarithm of 1+dose, so that vehicle has d_i=0. He notes the parameters β_1, β_2 indicate the rate of appearance of spontaneous papillomas and γ_1, γ_2 for dose by exposure effect.

This model has a number of attractive features. It seems to be interpretable and the fact that it models the increase in papilloma counts should mitigate or perhaps even eliminate the autocorrelation over time within each animals counts. This model was attempted to be fit with the SAS procedure NLMIXED. However, despite much effort, it was not possible to get any of the several versions of this model to provide converged parameter estimates, including variances. Thus frequentist estimates were not interpretable. A Bayesian version of the model was also fit using the SAS procedure MCMC. Bayesian techniques inherently smooth out the posterior likelihood and can improve fit of a model. However, the history of the estimates showed extremely high autocorrelations over the Monte Carlo procedure, even at lags of 500+ iterations. This leads to extremely poor mixing of the stochastic process representing the parameters, and thus doubt if the process is stationary. Hence, even after 100,000 iterations, there is much doubt if ergodic theorems justifying the estimation of parameters are actually applicable. All these suggest that, in this particular case, Dunson's model does not apply.

A model similar to Dunson's model that does seem to fit the data is to directly model the number of masses, Z_{ij}, or the maximum number of masses, M_{ij}. The same model seems to be equally appropriate for both endpoints. In either case, we can model Z_{ij} or M_{ij} as Poisson(μ_{ij}) with:

$$\mu_{ij} = \begin{cases} \exp(\beta_1 + b_i + \gamma d_i + \delta t_j) & for\ groups\ 1-4 \\ \exp(\beta_2 + b_i + \delta t_j) & for\ group\ 7 \end{cases}$$

Similar to Dunson's model, b_i denotes a mouse-specific random effect on the counts, which is assumed to follow a normal distribution with mean 0 and variance σ^2. Other inputs involve a linear effect of t_j=j/T, with parameter δ, d_i as the linear effect dose with parameter γ. Note that since this model does not take the log of dose, unlike Dunson's model the actual effect of dose

in groups 1-4 is exponential, as is the effect of time. The intercept in the two sets of treatments are defined as β_1, β_2, respectively, the baseline rate of papillomas for the two sets of treatments. In Dunson's model the response variable is the increase in maximum tumor count. For Dunson's model it probably makes sense to assume that, conditional upon the subject effect b_i, the within mouse responses are independent. However, clearly the actual tumor count or the maximum cell count can be expected to increase so such an assumption would not be appropriate for these endpoints. The model does include a term for a linear effect of time, but even above this, we might expect counts closer in time to be more alike than counts further apart. This suggests an autoregressive error structure within each subject. Using SAS PROC GENMOD to estimate this structure we find the following estimates:

Table A.4.1 Parameter Estimates for Direct Count of Masses

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Male Mice						
beta1	-5.5040	0.8486	-7.1672	-3.8409	-6.49	<.0001
beta2	-5.3257	0.7181	-6.7332	-3.9182	-7.42	<.0001
dose	0.1311	0.1041	-0.0728	0.3351	1.26	0.2077
week	4.8470	0.8022	3.2748	6.4192	6.04	<.0001
beta1-beta2	-0.1783	0.3893	-0.9413	0.5846	0.21[2]	0.6468
Female Mice						
beta1	-6.6446	0.8413	-8.2935	-4.9957	-7.90	<.0001
beta2	-6.3886	0.8929	-8.1386	-4.6385	-7.15	<.0001
dose	0.2620	0.1104	0.0456	0.4785	2.37	0.0177
week	5.7072	0.8814	3.9796	7.4347	6.47	<.0001
beta1-beta2	-0.2560	0.3954	-1.0311	0.5190	0.42[2]	0.5174

[2]Wald Chi Square

Table A.4.2 Parameter Estimates for Maximum Count of Masses

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Male Mice						
beta1	-5.5261	0.8497	-7.1914	-3.8608	-6.50	<.0001
beta2	-5.1659	0.7577	-6.6510	-3.6808	-6.82	<.0001
dose	0.1263	0.1011	-0.0718	0.3243	1.25	0.2116
week	4.9730	0.7992	3.4066	6.5394	6.22	<.0001
beta1-beta2	-0.3602	0.3639	-1.0734	0.3530	0.98[2]	0.3223
Female Mice						
beta1	-6.1924	0.8427	-7.8440	-4.5408	-7.35	<.0001
beta2	-6.1139	0.8941	-7.8663	-4.3616	-6.84	<.0001
dose	0.2102	0.0975	0.0191	0.4012	2.16	0.0310
week	5.5047	0.8592	3.8207	7.1887	6.41	<.0001
beta1-beta2	-0.0785	0.3802	-0.8236	0.6667	0.04[2]	0.8364

[2]Wald Chi Square

Not surprisingly, even with an autoregressive error structure, with both endpoints in both genders there is a statistically significant effect of week during the study (all four $p<0.0001$). Beta1 and beta2 reflect the baseline probability of an event in the Capcaisin groups 1-4 or the no treatment group 7, respectively, and can be expected to be non-zero. Thus the high statistical significance with each endpoint in each gender (all eight $p<0.0001$) is only to be

expected. Note that in both endpoints for both genders the tests of differences in these parameters are not statistically significant (all four $p \geq 0.3223$). In female mice there is a clear increasing effect due to dose (p=0.0177 and p=0.0310). Although the estimated dose effects in male mice are greater than zero, the tests that these parameters are 0 are not statistically significant (p=0.2077 and p=0.2116).

Appendix 5. FDA Analysis of Sponsor Identified Papillomas

The data for the following tables come from a data set described as "the number of skin papillomas by study week and body area." However as discussed in Section 2.2 above, there may be problems with this data. Note that "SOA" denotes site of application while "NSOA" is supposed to be the total of non-site of application. However, among males only one animal in the vehicle group, group 1, had non site of application (NSOA) counts of two papillomas and one animal in the TPA active control, group 6, had site of application (SOA) counts of two papillomas. Among female mice, only one mouse in the TPA group was credited with a single instance of two papillomas. So in almost all cases, the original per animal data were scored as either 0 or 1, including Group 6, the active control. For the tables below the sum of the SOA and NSOA counts reported by the Sponsor are displayed. The row labeled "n" indicates the number of animals at risk.

Table A.5.1 Papilloma Incidence in Male Mice[1]

Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Group 1 Vehicle																											
NSOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
SOA	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	
n	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	24	24	24	24	24	24	24	24	24	24	24	
Group 2 Low Dose																											
NSOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SOA	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
n	25	25	25	25	25	25	24	23	23	23	23	23	23	23	23	23	22	22	22	21	21	21	21	21	21	21	20
Group 3 Medium Dose																											
NSOA	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	4
SOA	0	0	0	0	0	0	0	0	0	0	0	1	1	2	2	2	2	2	2	2	2	2	2	2	2	3	5
n	25	25	25	25	25	25	24	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	22
Group 4 High Dose																											
NSOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
SOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
n	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	24	24	24	23	23	23	23	23	23	23	23	23
Group 5 Lidocane																											
NSOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	3
SOA	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
n	25	25	25	25	25	25	25	25	25	24	24	23	23	23	23	23	23	23	22	22	22	22	22	22	22	22	22
Group 6 TPA Active control																											
NSOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	3	3	3	3	5	3	3	3	4
SOA	0	0	0	0	0	0	0	0	0	0	0	0	3	6	7	7	9	9	6	8	7	5	5	1	2	3	5
n	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	21	18	15	12	12	8	8	8	8
Group 7 No treatment																											
NSOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	2	3	4
SOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	24	24	24	24	24	24	24	24	24	24	24

[1] Sponsor claims these are the total number of tumors.

In male mice there clearly is no particular trend over dose (in groups 1-4) in the number of animals with papillomas, and any reasonable test of trend over dose in groups 1-4 would not be statistically significant. As discussed below, the results in females seem to differ.

Table A.5.2 Papilloma Incidence in Female Mice[1]

Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Group 1																											
NSOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
SOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n	25	24	24	24	24	24	24	23	23	23	23	23	22	22	22	22	22	22	22	22	22	22	21	21	21	21	21
Group 2																											
NSOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
SOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
n	25	25	25	25	25	25	25	24	22	22	22	21	21	20	20	20	20	20	20	20	20	18	17	17	17	17	16
Group 3																											
NSOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	4
SOA	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
n	25	25	25	25	25	25	24	24	24	23	23	23	23	23	23	23	23	23	23	23	22	22	22	22	22	22	22
Group 4																											
NSOA	0	0	0	0	0	0	0	0	0	0	2	2	2	2	1	1	2	2	2	2	3	3	3	4	5	5	7
SOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	2	2	2	2	2
n	25	25	25	25	25	25	24	24	24	22	21	21	21	21	20	20	20	20	20	20	20	20	20	20	20	20	20
Group 5																											
NSOA	0	0	0	0	0	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	3
SOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n	25	25	25	25	24	24	23	23	22	22	22	21	21	21	21	20	20	20	20	20	20	20	19	19	18	18	18
Group 6																											
NSOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	4	6	4	4	4	2	2	1	2	3
SOA	0	0	0	0	0	0	0	0	0	0	1	2	3	3	4	5	6	8	9	8	6	6	6	5	5	5	5
n	25	25	25	25	25	25	24	24	24	24	23	22	22	22	22	22	21	20	20	18	16	16	14	13	12	10	9
Group 7																											
NSOA	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	2	2	2	2	2	2	2	2	2	3	3	5
SOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
n	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	24	24	23	23	23	23	23	23	23	23	22

Unlike male mice, in female mice there is a clear evidence of trend in tumor incidence.

Again there were a number of unsuccessful attempts to apply Dunson's model to the SOA and NSOA data summarized above. The simpler Poisson model described in Appendix 4 was also fit. The clear lack of trend in males in the SOA data precluded adequate fit of the simpler Poisson model, since the lack of trend required more parameters than the simple trend model and it appears that that there are insufficient observations to adequately support estimation of variance. However the models for females and NSOA males do seem to adequately fit, leading to the following parameter estimates:

Table A.5.3 Parameter Estimates for Site of Application (SOA) Papillomas

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Male Mice						
As discussed above, there seems to be insufficient data to fit the Poisson model. However, it is apparent from Table A.5.1 that there is no statistically significant trend in dose.						
Female Mice						
beta1	-7.4728	1.7199	-10.8438	-4.1018	-4.34	<.0001
beta2	-7.0328	1.4768	-9.9273	-4.1382	-4.76	<.0001
dose	0.3768	0.1583	0.0664	0.6871	2.38	0.0173
week	4.0373	1.3105	1.4687	6.6058	3.08	0.0021
beta1-beta2	-0.4400	0.5617	-1.5410	0.6609	0.61[2]	0.4334

[2]Wald Chi Square

Table A.5.4 Parameter Estimates for Non-site of Application (NSOA) Papillomas

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Male Mice						
beta1	-9.4449	2.2725	-13.8989	-4.9909	-4.16	<.0001
beta2	-8.1904	1.8984	-11.9112	-4.4695	-4.31	<.0001
dose	0.1116	0.1454	-0.1734	0.3965	0.77	0.4429
week	6.4606	1.8350	2.8641	10.0572	3.52	0.0004
beta1-beta2	-1.2545	0.5494	-2.3313	-0.1777	5.21[2]	0.0224
Female Mice						
beta1	-8.2996	1.2143	-10.6796	-5.9196	-6.83	<.0001
beta2	-6.8668	1.2271	-9.2718	-4.4618	-5.60	<.0001
dose	0.5115	0.0362	0.4406	0.5824	14.14	<.0001
week	5.4941	1.1976	3.1467	7.8414	4.59	<.0001
beta1-beta2	-1.4328	0.1337	-1.6949	-1.1707	114.80[2]	<.0001

[2]Wald Chi Square

As discussed in Appendix 4, an autoregressive error structure within each individual mouse was assumed. As in the analysis of masses, primary interest focuses on the linear effect of dose. In male mice there is no evidence of a dose related trend (NSOA p=0.4429). However, in female mice there does seem to be evidence of such a positive trend in dose (SOA p=0.0173, NSOA p<0.0001). Again, beta1 and beta2 reflect the baseline probability of an event in groups 1-4 or group 7, respectively, and can be expected to be non-zero, but one would expect them to be nearly equal. Perhaps interestingly, here, among female mice, apparently due to the relatively large number of papillomas in the no treatment group 7, the difference with the baseline of the Capcaisin groups is statistically significant (SOA p=0.0224, NSOA p<0.0001). Since regressor effects do affect the intercept, this might suggest a protective effect in the vehicle. However, the number of papillomas is too small to interpret this as strong evidence.

Appendix 6. References

Bailer, A. and Portier, C. (1988), "Effects of Treatment-Induced Mortality on Tests for Carcinogenicity in Small Samples", *Biometrics*, **44**, 4, 417-431.

Bieler, G.S., and Williams, R.L. (1993), "Ratio Estimates, the Delta Method, and Quantal Response Tests for Increased Carcinogenicity", *Biometrics*, **49**, 4, 793-801.

Chu, K.C., Ceuto, C., and Ward, J.M. (1981), Factors in the Evaluation of 200 National Cancer Institute Carcinogen Bioassays, *Journal of Toxicology and Environmental Health,* **8**, 251-280.

Dunson, D.B., Haseman, J.K.,van Birgelen, A.P.J.M., Stasiewicz, S, and Tennant, R.W. (2000) Statistical Analysis of Skin Tumor Data from Tg.AC Mouse Bioassays, *Toxicology Sciences,* **55**, 293-302.

Haseman, J. K. (1983), A Reexamination of False-positive Rates for Carcinogenicity Studies, *Fundamental and Applied Toxicology,* 3, 334-339.

Kalbfleisch, J.D. (1978), Non-Parametric Bayesian Analysis of Survival Time Data, *Journal of the Royal Statistical Society, Series B (Methodological),* 40 (2), 214-221.

Kim, Y. and Kim, D. (2009), Bayesian Partial Likelihood Approach for Tied Observations, *Journal of Statistical Planning and Inferencce,* 139, 469-477.

Lin, K. K. and Ali, M.W. (2006), Statistical Review and Evaluation of Animal Tumorigenicity Studies, *Statistics in the Pharmaceutical Industry,Third Edition,* edited by C.R. Buncher and J.Y. Tsay, Marcel Dekker, Inc. New York.

Lin, K. K. and Rahman, M.A. (1998), Overall False Positive Rates in Tests for Linear Trend in Tumor Incidence in Animal Carcinogenicity Studies of New Drugs, *Journal of Biopharmaceutical Statistics*. 8(1), 1-15.

McConnell, E.E., Solleveld, H.A., Swenberg, J.A., and Boorman, G.A. (1986), Guidelines for Combining Neoplasms for Evaluation of Rodent Carcinogenesis Studies, *Journal of the National Cancer Institute*. 76, 283-289.

Peto, R., Pike, M.C., Day, N.E., Gray, R.G., Lee, P.N., Parrish, S., Peto, J., Richards, S., and Wahrendorf, J. (1980). Guidelines for sample sensitive significance tests for carcinogenic effects in long-term animal experiments, *IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans, supplement 2: Long term and Short term Screening Assays for Carcinogens: A Critical Appraisal*, International Agency for Research Against Cancer, 311-426.

Rahman, M.A. and Lin, K.K. (2008), A Comparison of False Positive Rates of Peto and Poly-3 Methods for Long Term Carcinogenicity Data Analysis Using Multiple Comparison Adjustment Method Suggested by Lin and Rahman, *Journal of Biopharmaceutical Statistics*. 18, 949-958.

STP Peto Working Group (2002), Statistical Methods for Carcinogenicity Studies, *Toxicologic Pathology*. 30 (3), 403-414.

U.S. Department of Health and Human Services (2001), *Guidance for Industry Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals (DRAFT GUIDANCE),* Center for Drug Evaluation and Research, Food and Drug Administration.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

Steven Thomson
4/10/2009 08:46:25 AM
BIOMETRICS

Karl Lin
4/13/2009 12:14:52 PM
BIOMETRICS
Concur with review

Screening of New NDA
Division of Biometrics II

Date: 11/30/08

NDA #: 22-395 Priority Classification: S

Trade Name: NGX-4010 Applicant : Neurogesx

Generic Name: Capsaicin patch 8% Date of Submission: 10/13/08

Indication: prolonged reduction of neuropathic pain associated with postherpetic neuralgia (PHN)

No. of Controlled Studies: 3

User Fee Goal Date: 8/13/09

Date of 45-Day Meeting: 12/1/08

Medical Officer: Neville Gibbs, M.D.

Project Manager: Tanya Clayton

Screened by: Kate Meaker, M.S.

Statistical sections: Sections 2.5, 2.7, 5.3.5

Anticipated Review Completion Date: 5/13/09

Comments:

1. Applicant reports both mean change from baseline and % change from baseline as efficacy endpoints. It is not clear which is intended as primary. I need to confirm with MO which is preferred primary efficacy endpoint.

2. The application is fileable.

CHECKLIST

Item	**Check (NA if not applicable)**
Index sufficient to locate necessary reports, tables, etc.	**Yes**
Original protocols & subsequent amendments available in the NDA	**Yes**
Designs utilized appropriate for the indications requested	**Yes**
Endpoints and methods of analysis spelled out in the protocols	**Yes**
Interim analyses (if present) planned in the protocol and appropriate adjustments in significance level made	**Yes**
Appropriate references included for novel statistical methodology (if present)	**NA**
Sufficient data listings and intermediate analysis tables to permit statistical review	**Yes**
Data from primary studies submitted to edr	**Yes**
Intent-to-treat analysis	**Yes**
Effects of dropouts on primary analyses investigated	**Yes (sensitivity an. using alt. imputations)**
Safety and efficacy for gender, racial, and geriatric subgroups investigated	**Yes (age and gender; race not done - explained in ISE- few non-Cauc.)**

BRIEF SUMMARY OF CONTROLLED CLINICAL TRIALS

Study Number (Dates Conducted)	Number of Centers (Locations)	Total Sample Size	Type of Control	Design Du	ration of Treatment
C117 (3/06 – 7/07)	61 (US; Canada)	NGX-4010 [640 mcg/cm^2] (n=212) Low dose Control [3.2 mcg/cm^2] (n=204)	Low concentration capsaicin control patch (3.2 mcg/cm^2)	Randomized, Double-blind, Multicenter, Active-control, Parallel arm	Single 60-minute application; 12-week efficacy assessments and follow-up
C116 (5/05 – 8/06)	52 (US)	NGX-4010 [640 mcg/cm^2] (n=206) Low dose Control [3.2 mcg/cm^2] (n=196)	Low concentration capsaicin control patch (3.2 mcg/cm^2)	Randomized, Double-blind, Multicenter, Active-control, Parallel arm	Single 60-minute application; 12-week efficacy assessments and follow-up
C110 (9/03 – 7/04)	20 (US)	NGX-4010 [640 mcg/cm^2] (n=102) Low dose Control [3.2 mcg/cm^2] (n=53)	Low concentration capsaicin control patch (3.2 mcg/cm^2)	Randomized, Double-blind, Multicenter, Active-control, Parallel arm	Single 60-minute application; 12-week efficacy assessments and follow-up

_____

Kat herine B. Meaker
Mathematical Statistician

Concur: Dionne Price Ph.D.
Team Leader

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

Katherine Meaker
12/3/2008 11:17:22 AM
BIOMETRICS

Dionne Price
12/3/2008 11:20:40 AM
BIOMETRICS