CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

050803Orig1s000

STATISTICAL REVIEW(S)

US 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
New Drug Application
Clinical Studies

NDA/Serial Number:	50-803/SN000
Drug Name:	Veltin™ (clindamycin, tretinoin) Gel, 1%, 0.025%
Indication(s):	Acne vulgaris
Applicant:	Stiefel Laboratories, Inc.
Dates:	Submitted: 10/16/2009
	PDUFA: 07/16/2010
Review Priority:	Complete Response Resubmission
Biometrics Division:	Division of Biometrics III
Statistics Reviewer:	Mat Soukup, Ph.D.
Concurring Reviewer:	Mohamed Alosh, Ph.D.
Medical Division:	Division of Dermatology and Dental Products
Clinical Team:	Reviewer: Gary Chiang, M.D. (DDDP)
	Lead: David Kettl, M.D. (DDDP)
Project Manager:	Cristina Petruccelli, MPH (DDDP)
Keywords:	acne, superiority, combination drug, resubmission

Contents

1 Executive Summary **5**

1.1 Conclusions and Recommendations . . . 5

1.2 Brief Overview of Clinical Studies . . . 6

1.3 Statistical Issues and Findings . . . 6

2 Introduction **9**

2.1 Product Description . . . 9

2.2 Regulatory History . . . 9

2.2.1 Compete Response: 6/10/2005 . . . 9

2.2.2 Post-NA Meeting: 08/17/2005 . . . 9

2.2.3 Special Protocol Assessment: 11/17/2005 . . . 10

2.2.4 Post-SPA Guidance Meeting: 12/18/2007 . . . 11

2.2.5 Protocol Review (SN064): Stamp Date of 01/17/2008 . . . 11

2.3 Clinical Trial Overview . . . 12

2.4 Data Sources . . . 12

3 Statistical Evaluation **13**

3.1 Evaluation of Efficacy - Study W0265-03 . . . 13

3.1.1 Study Design . . . 13

3.1.2 Endpoints . . . 15

3.1.2.1 Co-Primary Efficacy Criteria . . . 15

3.1.2.2 Secondary Efficacy Criteria . . . 16

3.1.3 Statistical Methodology . . . 17

3.1.3.1 Populations . . . 17

3.1.3.2 Statistical Analysis . . . 17

3.1.3.2.1 Investigator Global Assessment . . . 18

3.1.3.2.2 Lesion Counts . . . 18

3.1.4 Patient Disposition and Baseline Characteristics . . . 19

3.1.4.1 Patient Disposition . . . 19

3.1.4.2 Baseline Demographic Factors . . . 19

3.1.4.3 Baseline Prognostic Factors . . . 21

3.1.5 Populations Analyzed . . . 21

3.1.6 Primary Endpoint Results (ITT-LOCF) . . . 22

3.1.6.1 Investigator Global Assessment . . . 22

3.1.6.2 Change in Lesion Counts . . . 23

3.1.6.2.1 Total Lesion Counts . . . 23

3.1.6.2.2 Inflammatory Lesion Counts . . . 24
3.1.6.2.3 Non-Inflammatory Lesion Counts . . . 25
3.1.7 Primary Endpoint Results (PP-LOCF) . . . 25
3.1.7.1 Investigator Global Assessment . . . 26
3.1.7.2 Change in Lesion Counts . . . 26
3.1.8 Additional Sensitivity Analyses of Primary Endpoints . . . 27
3.1.8.1 Missing Data Sensitivity Analysis . . . 27
3.1.8.1.1 Investigator Global Assessment . . . 27
3.1.8.1.2 Change in Lesion Counts . . . 28
3.1.9 Secondary Endpoint Results . . . 29
3.2 Consistency of Efficacy Findings - All Phase 3 Trials . . . 30
3.2.1 Investigator Global Results - All Phase 3 Trials . . . 30
3.2.2 Lesion Counts - All Phase 3 Trials . . . 31
3.2.2.1 Total Lesion Counts . . . 33
3.2.2.2 Inflammatory Lesion Counts . . . 33
3.2.2.3 Non-Inflammatory Lesion Counts . . . 34
3.3 Evaluation of Safety . . . 35
3.3.1 MedDRA Tabulation . . . 35
3.3.1.1 All Reported Adverse Events . . . 35
3.3.1.2 Related Adverse Events . . . 37
3.3.2 Serious Adverse Events . . . 38
3.3.2.1 Local Skin Reactions . . . 38

4 Findings in Special/Subgroup Populations 39
4.1 Gender, Race, and Age . . . 40
4.1.1 Age . . . 40
4.1.2 Gender . . . 40
4.1.3 Race . . . 41
4.2 Other Special/Subgroup Populations . . . 42
4.2.1 Efficacy by Site . . . 42

5 Summary and Conclusions 43
5.1 Statistical Issues and Collective Evidence . . . 43
5.2 Conclusions and Recommendations . . . 46

Appendix 48
A.1 Supplementary Information for Study W0265-03 . . . 48
A.1.1 Efficacy Tables by Subgroups . . . 48

A.1.1.1 Investigator Global Assessment 48
A.1.1.2 Change in Total Lesion Counts 48
A.1.1.3 Change in Inflammatory Lesion Counts 49
A.1.1.4 Change in Non-Inflammatory Lesion Counts 50

SIGNATURES/DISTRIBUTION LIST **50**

1 Executive Summary

1.1 Conclusions and Recommendations

Veltin™ Gel is a combination drug product comprising clindamycin, 1% and tretinoin, 0.025% in a topical gel for the treatment of acne vulgaris. An earlier formulation of the drug product, Velac Gel, [redacted] (b) (4) was found to have a positive carcinogenicity signal during the initial filing of the NDA. On June 10, 2005 the sponsor was issued a not approval letter on the basis of this finding. The sponsor reformulated the product [redacted] (b) (4).

On the basis of the newly formulated product the sponsor conducted an additional Phase 3 trial (Study W0265-03) which included the combination drug product, each monad, and the vehicle gel. The protocol-defined primary efficacy endpoints were:

- The absolute change in lesion counts (total, inflammatory, non-inflammatory) from baseline to Week 12 (end of treatment); and
- The proportion of subjects who have a minimum two grade improvement in their IGA score from baseline to Week 12 (end of treatment).

In Study W0265-03, Veltin™ Gel was statistically superior to each monad and vehicle for the percent of IGA successes and the absolute change in total lesion counts. For the comparison of inflammatory lesions, Veltin™ Gel was superior to Tretinoin Gel and vehicle, however was not significantly superior to Clindamycin Gel ($p = 0.1797$). For non-inflammatory lesion counts, Veltin™ Gel was superior to Clindamycin Gel and vehicle, but did not reach statistical significance for the comparison to Tretinoin Gel ($p = 0.2541$).

To assess the consistency in efficacy findings of the newly formulated drug product, Veltin™ Gel, to that of the old formulation, Veltin Gel, the efficacy of Study W0265-03 was compared descriptively to the efficacy in Studies VLC.C.304 and VLC.C.305 for all co-primary endpoints. The following summarizes the efficacy for each of the co-primary endpoints across trials/formulations.

IGA Success : Treatment effects of the combination product versus the monads were highest in Study VLC.C.304 which were similar to those observed in Study W0265-03. Treatment effects of the combination product versus vehicle were similar for all studies.

Lesion Counts : A linear model was used to derive estimates of treatment effects for lesion counts in each study. Trends were similar across trials, though the magnitude of the treatment effect was smallest in Study W0265-03 for all lesion counts.

While the magnitude of the treatment effects for lesion counts were smallest in Study W0265-03, the observed treatment effects for the IGA endpoint for Study W0265-03 were similar to

that of Study VLC.C.304 which had the largest treatment effects. Overall, efficacy trends of all co-primary endpoints were consistent across all Phase 3 trials.

In the safety assessment of local skin reactions, on average, the irritability of Veltin™ Gel was similar to Tretinoin Gel. While the mean irritation was not severe in nature, the mean irritation peaked at week 2 with a gradual decrease towards baseline levels by week 12. Adverse events, other than local skin reactions, were not discordant in subjects treated with Veltin™ versus the monads and vehicle.

1.2 Brief Overview of Clinical Studies

In the clinical development of the combination product, two adequate and well-controlled twelve week, randomized, double-blind, and parallel group studies (VLC.C.304 and VLC.C.305) were conducted to assess the safety and efficacy of the original formulation (Velac Gel)[1]. Each of the two trials were intended to show the contribution of each component of the combination product as well as the superiority over vehicle. A single Phase 3 trial, Study W0265-03, was conducted with the reformulated product with the objective of establishing the safety and efficacy of Veltin™ Gel over each monad and vehicle. Studies VLC.C.304 and VLC.C.305 enrolled a total of 2219 subjects (634 randomized to the combination product) whereas Study W0265-03 enrolled a total of 1649 subjects (476 randomized to Veltin™ Gel).

1.3 Statistical Issues and Findings

Following the not approval (NA) action of June 10, 2005, the sponsor and Agency had several discussions about the future development of a reformulated combination drug product (see, Section 2.2 for a discussion of the regulatory history following the NA action). It was agreed that a single four arm trial with the objective of establishing the superiority of Veltin™ over each monad and vehicle would be sufficient for an efficacy claim if the efficacy results of the reformulated product, Veltin™, are similar to those observed with the original formulation.

During the protocol review of Study W0265-03, several comments which impact the evaluation of efficacy were not incorporated into the protocol as the comments do not appear to have been sent to the sponsor. These are summarized as follows.

Multiplicity adjustments : To justify the inclusion of multiple actives for a combination product for acne, the combination product should demonstrate efficacy over each monad for at least one lesion type, as well as the global. At a minimum, the following should be demonstrated:

[1]The review of these two studies was signed on April 26, 2005.

- CT Gel should be superior to vehicle for change in inflammatory and non-inflammatory lesions and success on the IGA;
- CT Gel should be superior to clindamycin for success on the IGA and for one pre-specified lesion type or the protocol should specify a multiplicity adjusted method for demonstrating superiority for at least one lesion type;
- CT Gel should be superior to tretinoin for success on the IGA and for one pre-specified lesion type or the protocol should specify a multiplicity adjusted method for demonstrating superiority for at least one lesion type.

IGA success definition : Improving from an IGA score of 5 to an IGA score of 3 would not be sufficient for success. If it is appropriate to enroll subjects with an IGA score of 5 at baseline, these subjects should improve to a score of 2 or better at Week 12.

As the above comments were not communicated to the sponsor this review utilized a multiplicity adjustment which was based on results from the completed Phase 3 trials of the original formulation. In addition, a stringent definition of IGA success was utilized in which subjects needed a week 12 score of 0 or 1 with at least a two grade reduction to be considered a treatment success.

In the evaluation of efficacy, the following criteria were used in *this review* to determine the efficacy of Veltin™ Gel for Study W0265-03.

IGA : Success = week 12 score of 0 or 1 with at least a two grade improvement. The objective for assessing IGA success was to establish the

- superiority of Veltin™ Gel over each monad, Clindamycin Gel and Tretinoin Gel, and Vehicle Gel.

Lesion Counts : The absolute change from baseline to week 12. The objective for assessing lesion counts was to establish the

- superiority of Veltin™ Gel over Clindamycin Gel for total and non-inflammatory lesions,
- superiority of Veltin™ Gel over Tretinoin Gel for total and inflammatory lesions, and
- superiority of Veltin™ Gel over Vehicle Gel for total, inflammatory, and non-inflammatory lesions.

Testing was performed at the two-sided $\alpha = 0.05$ significance level. Analysis of the IGA success rates was performed using CMH stratified on analysis centers. Comparisons of the absolute changes was conducted using ANCOVA with terms for analysis centers and treatment and baseline counts as the covariate.

In Study W0265-03, Veltin™ Gel was superior to each monad and vehicle for IGA success. Veltin™ Gel was also superior to each monad and vehicle for total lesions. For inflammatory lesions, Veltin™ Gel was superior to Tretinoin Gel and vehicle, but not superior to Clindamycin Gel. For non-inflammatory lesions, Veltin™ Gel was superior to Clindamycin Gel and vehicle, but not for Tretinoin Gel (Table 24). Thus, Study W0265-03 met all efficacy objectives.

Table 1: **Efficacy Summary (ITT/LOCF); Study W0265-03**

	Veltin™ Gel ($N = 476$)	Clindamycin Gel ($N = 467$)	Tretinoin Gel ($N = 464$)	Vehicle Gel ($N = 242$)
IGA Success: 0 or 1 with Two Grade Imp.				
Success (%)	158 (33.2)	112 (24.0)	105 (22.6)	43 (17.8)
p-value†	-	0.0018	< 0.001	< 0.001
Absolute Change Total Lesions				
Mean Change (SD)	38.7 (26.8)	34.0 (25.2)	36.0 (28.3)	28.1 (27.7)
p-value‡	-	0.0028	0.037	< 0.001
Absolute Change Inflammatory Lesions				
Mean Change (SD)	15.5 (10.3)	14.5 (9.4)	13.9 (11.1)	11.1 (11.7)
p-value‡	-	0.1797	0.0022	< 0.001
Absolute Change Non-Inflammatory Lesions				
Mean Change (SD)	23.2 (20.4)	19.5 (19.7)	22.1 (21.7)	17.0 (20.6)
p-value‡	-	0.0011	0.2541	< 0.001

† CMH stratified by 'analysis center' (Source: reviewer's analysis).

‡ p-value is based on an ANCOVA models with terms for treatment and 'analysis center' with covariate for the baseline lesion count (Source: reviewer's analysis).

To assess the consistency of efficacy findings across all Phase 3 trials; analyses were conducted to compare the efficacy of the reformulated product with that of original formulation. Treatment effects for the IGA endpoint were similar in Study VLC.C.304 and W0265-03 which were higher than in Study VLC.C.305. For assessing changes in lesion counts, the treatment effects in Study W0265-03 were smaller than in both Studies VLC.C.304 and VLC.C.305 (results of these analyses are presented in Section 3.2). Overall, efficacy trends of all co-primary endpoints were consistent across all Phase 3 trials.

In the safety assessment of local skin reactions for Study W0265-03, on average, the irritability of Veltin™ Gel was similar to Tretinoin Gel. While the mean irritation was not severe in nature, the mean irritation peaked at week 2 with a gradual decrease towards baseline levels by week 12. Adverse events, other than local skin reactions, were not discordant in subjects treated with Veltin™ Gel versus the monads and vehicle. Note that safety results are only provided for Study W0265-03 due to the observed carcinogenicity signal of the original formulation.

2 Introduction

2.1 Product Description

Veltin™ Gel is a combination product developed by Stiefel for the treatment of acne vulgaris. Veltin™ Gel consists of the two active ingredients, clindamycin, 1% and tretinoin, 0.025% in a gel vehicle. Treatment with Veltin™ Gel was studied by applying drug product once daily in the evening for 12 weeks.

2.2 Regulatory History

The following description of the regulatory history originates *after* the submission of the original NDA which was submitted on 8/26/2004.

2.2.1 Compete Response: 6/10/2005

On June 10, 2005 the sponsor was issued a non-approval (NA) for their NDA submitted on 8/26/2004. The response for the action is stated below.

> "A positive carcinogenicity signal was detected in a Tg.AC mouse dermal carcinogenicity study in the vehicle arm as well as in the arms containing clindamycin phosphate. This suggests that the vehicle or a component of the vehicle of the drug product may be a tumor promoter or may be carcinogenic. This is an unacceptable safety finding for a topical drug product intended for the chronic treatment of a nonlife threatening or non-severely debilitating disease for which many alternative therapies exist. Therefore, it was determined that the drug is unsafe for use under the conditions prescribed, recommended, or suggested in its proposed labeling or the results do not show that the drug product is safe for use under those conditions."

While the application was not approved, the statistical evaluation of the efficacy of the combination drug product (Velac Gel) showed the statistical superiority of the combination product over each monad and its vehicle (statistical review signed on April 26, 2005).

2.2.2 Post-NA Meeting: 08/17/2005

The Division and sponsor held a Post-NA meeting on 08/17/2005 to discuss the necessary clinical studies needed to support a change in formulation of their combination product. Note that the reformulated product removed [redacted] (b) (4) not listed in the FDAs Inactive Ingredients Guide.

To demonstrate that the new formulation maintains the efficacy demonstrated in the original formulation, the Division recommended that the sponsor conduct a 4-arm study comparing

the new formulation (referred to as CT gel), the original Velac formulation, and clindamycin and tretinoin in the new vehicle. The Division recommended that CT gel should be non-inferior to Velac gel and superior to clindamycin and tretinoin gels for each of the co-primary endpoints. At the meeting, the sponsor proposed conducting a 3-arm study demonstrating the superiority of CT gel to its vehicle, and the non-inferiority of CT gel to Velac gel. The sponsor proposed establishing superiority over vehicle using both lesion counts and the IGA scale, and non-inferiority using only lesion counts (2 out of 3 percent reductions) as the primary endpoint.

2.2.3 Special Protocol Assessment: 11/17/2005

Following the complete response action of 08/26/2004 and the Post-NA Meeting on 08/17/2005, the sponsor submitted a Phase 3 protocol for Special Protocol Assessment (SPA) on 10/03/2005. The title of the protocol was "A Phase III Multi-center, Randomized, Double-Blind, Active and Vehicle-Controlled Study of the Safety and Efficacy of clindamycin-tretinoin (CT) Gel versus Velac Gel versus Vehicle Gel for the Treatment of Acne Vulgaris". The proposed objective of this trial was to establish:

- CT gel was non-inferior to Velac Gel using a 10% NI margin, and
- CT gel was superior to vehicle.

In the evaluation of the SPA, the Division provided comment that the use of a NI trial with an agreeable NI margin may be prohibitive as it would require a large number of subjects in order to establish the non-inferiority of CT Gel to Velac Gel for both lesion counts and the IGA. Consequently, it was recommended that the sponsor consider conducting a four-arm trial which was of similar design as the completed Phase 3 trials assessing the efficacy of Velac Gel in which CT Gel be superior to each component and vehicle or conducting a NI trial with a margin of 7%. If efficacy results of the CT Gel formulation are similar to those seen in the Phase 3 trials of Velac Gel, this would be sufficient to support the new formulation.

In addition to the above comment on the design of the Phase 3 trial for assessing CT Gel, the following comments were also provided to the sponsor.

1. The protocol should also analyze absolute change in lesion counts as well as include a secondary analysis of the IGA score in which a success is defined as a Week 12 score of 0 or 1 with at least a two grade reduction.
2. If the sponsor pursues a trial with a NI comparison, both the ITT and PP populations would be considered primary for the NI comparisons.
3. The protocol should specify a pooling strategy for centers with a small recruitment size.

2.2.4 Post-SPA Guidance Meeting: 12/18/2007

For the Post-SPA Guidance Meeting held on 12/18/2007 the sponsor proposed to conduct a 3-arm study: Velac Gel (old formulation), CT Gel (new formulation), and new vehicle. The study was designed to demonstrate that CT Gel is non-inferior to Velac Gel for the four endpoints (percent reduction in inflammatory, non-inflammatory, and total lesion counts and global success) using a 17% non-inferiority margin and to also demonstrate that CT Gel is superior to vehicle for all four endpoints. Prior to the meeting the Division informed the sponsor that a smaller NI margin of 7% would be preferable as stated in SPA letter of 11/17/2005. During the meeting the sponsor noted that they might find it more feasible from sample size considerations to conduct a 4-arm superiority study rather than a 3-arm non-inferiority study with a non-inferiority margin of 7%.

2.2.5 Protocol Review (SN064): Stamp Date of 01/17/2008

Following the advice from the Post-SPA Guidance Meeting, the sponsor submitted a revised Phase 3 protocol in SN064 for a four-arm superiority trial. In the review of the protocol, the Division provided the following comments. Note that a more detailed description of the protocol is provided in Section 3.1.1 of this review.

1. Since the sponsor has proposed updating the protocol to current recommendations and options from the draft Acne Guidance (using absolute change in lesions and using at least 2 grades reduction on the IGA instead of percent change and 0 or 1 on the IGA), the proposal to evaluate '2 out of 3' lesion types should also be updated. To justify the inclusion of multiple actives for a combination product for acne, the combination product should demonstrate efficacy over each monad for at least one lesion type, as well as the global. At a minimum, the following should be demonstrated:

    - CT Gel should be superior to vehicle for change in inflammatory and non-inflammatory lesions and success on the IGA;
    - CT Gel should be superior to clindamycin for success on the IGA and for one prespecified lesion type or the protocol should specify a multiplicity adjusted method for demonstrating superiority for at least one lesion type;
    - CT Gel should be superior to tretinoin for success on the IGA and for one prespecified lesion type or the protocol should specify a multiplicity adjusted method for demonstrating superiority for at least one lesion type.

2. Improving from an IGA score of 5 to an IGA score of 3 would not be sufficient for success. If it is appropriate to enroll subjects with an IGA score of 5 at baseline, these subjects should improve to a score of 2 or better at Week 12.

3. The sponsor should justify the rationale for testing the success on the IGA for all treatment arms first, and then if significant test the combination arm versus the other arms.

4. The protocol should include additional sensitivity analyses regarding the handling of missing data to ensure that the results are not driven by the method of imputation.

Reviewer Comment: On 03/07/2008 the statistical review was signed off. However, based on the application history in DARRTS, it does not appear that the comments in the statistical review of this revised protocol were sent to the sponsor. Consequently, the sponsor was not aware of the comments and corrections to the protocol were not made.

2.3 Clinical Trial Overview

The sponsor's development program of the original formulation, Velac Gel, consisted of two Phase 3 trials, Studies VLC.C.304 and VLC.C.305, which included treatment arms for each monad and a vehicle. Using the newly formulated combination product, Veltin™ Gel, the sponsor completed a Phase 3 trial, Study W0265-03, comparing Veltin™ Gel to each monad and vehicle using the newly formulated vehicle. Studies VLC.C.304 and VLC.C.305 were reviewed in the original submission which was signed off on April 26, 2005. These same two studies will be incorporated into the current review for an integrated analysis of efficacy to assess the consistency of the efficacy response of the new formulation (Veltin™ Gel) and the old formulation (Velac Gel). Table 2 provides a summary of the Phase 3 trials using the old formulation as well as the current formulation.

2.4 Data Sources

For the resubmission the sponsor did not include the efficacy data from Studies VLC.C.304 and VLC.C.305. As the Agency was interested in exploring the consistency in efficacy with the new formulation in comparison to the old formulation, data sets for all Phase 3 trials were requested in which the data sets were structured in similar ways. The sponsor submitted a data set which contained efficacy information for all three Phase 3 trial on January 21, 2010. However, based upon this submitted data set, the reviewer was not able to reproduce the efficacy results provided by the sponsor in the study report. Subsequent to this inspection of the data, the Agency and sponsor held a teleconference to discuss the issues related to the data sets and the lack of reproducibility of their efficacy results. The sponsor agreed to submit a revised data set which accounted for subjects who missed planned visits, had visits outside the visit window, as well as including both raw and imputed data values for cases of missing data. This revised data set and the implications it had on efficacy results was received on March 18, 2010. The

Table 2: **Efficacy and Safety Studies Overview**

Study	Development Objective	Drug Products	Number Subjects	Date†
VLC.C.304 (Study 304)	Phase 3 Superiority	Velac Gel	309	12/16/2002 – 01/20/2004
		Clindamycin Gel	311	
		Tretinoin Gel	310	
		Vehicle Gel	153	
VLC.C.305 (Study 305)	Phase 3 Superiority	Velac Gel	325	12/16/2002 – 01/21/2004
		Clindamycin Gel	324	
		Tretinoin Gel	325	
		Vehicle Gel	162	
VLC.C.305 (Study 305)	Phase 3 Superiority	Veltin™	476	04/01/2008 – 04/07/2009
		Clindamycin Gel	467	
		Tretinoin Gel	464	
		Vehicle Gel	242	

† Dates correspond to the start and end of the study.

data received in this resubmission are those used in the following review to form the basis of the efficacy conclusion. The revised data sets are located in the electronic document room under the listing `//FDSWA150/NONECTD/N50803/N_000/2010-03-18`.

The safety review of the newly formulated drug product will be based upon study W0265-03 only. This data set can be found in the electronic document room under the listing `//FDSWA150/NONECTD/N50803/N_000/2009-11-10/W0265-03/SDTM/data`. Safety results from Studies VLC.C.304 and VLC.C.305 are presented in the original statistical review signed on 4/26/2005.

3 Statistical Evaluation

The statistical evaluation is divided into several sections. The first section, Section 3.1, provides a thorough examination of the efficacy data for Study W0265-03. This is followed by Section 3.2 which is an integrated summary of efficacy. Lastly, Section 3.3 provides a safety summary of the safety data collected in Study W0265-03.

3.1 Evaluation of Efficacy - Study W0265-03

3.1.1 Study Design

Protocol W026503 is a four-arm (CT Gel, Clindamycin Gel, Tretinoin Gel, Vehicle Gel) trial to evaluate the efficacy and safety of CT Gel gel in the treatment of acne. The trial enrolled a total

of 1649 subjects of which 476 were randomized to CT Gel, 467 were randomized to Clindamycin Gel, 464 were randomized to Tretinoin Gel, and 242 were randomized to Vehicle Gel. Subjects were enrolled from 32 investigative centers with the average enrollment of 51.5 subjects per center. Centers were located in the United States (27 centers), Canada (3 centers), and Belize (2 centers). The first subject was enrolled on April 1, 2008 and the last subject completed the trial on April 7, 2009.

At baseline subjects enrolled with a minimum IGA score of 2 and a maximum score of 4 (note that IGA scale is a 6 point scale numbered from 0 to 5)[2]. The protocol defined inclusion criteria listed the baseline inflammatory lesion count to be between 17 and 40 lesion and the non-inflammatory lesion count to be between 20 and 150 lesions (the actual minimum and maximum lesion counts at baseline are provided in table 3 for each treatment group). Note that several subjects enrolled with inflammatory lesions counts that fell outside of these bounds; 1 subject randomized to CT Gel, Clindamycin Gel, and Tretinoin Gel had a baseline inflammatory count less than 17 lesions, 3 subjects randomized to CT Gel had a baseline inflammatory lesion count greater than 40, 2 subjects randomized to Clindamycin Gel had a baseline inflammatory lesion count greater than 40, 3 subjects randomized to Tretinoin Gel had a baseline inflammatory lesion count greater than 40, and 2 subjects randomized to Vehicle Gel had a baseline inflammatory lesion count greater than 40.

Table 3: **Baseline Lesions Count Range (Study W0265-03)**†

	CT Gel ($N = 476$)	Clindamycin Gel ($N = 467$)	Tretinoin Gel ($N = 464$)	Vehicle Gel ($N = 242$)
Inflammatory Lesions	16, 60	1, 42	17, 47	16, 48
Non-Inflammatory Lesions	20, 146	20, 141	20, 148	20, 126

† Values represent the minimum and maximum lesion counts recorded at baseline.

Source: Reviewer analysis.

The study treatment period was 12 weeks with five planned visits. Planned visits occurred at baseline (day 1), and on weeks 2, 4, 8, and 12 (end of treatment). Subjects were instructed to apply a sufficient amount of study treatment to the entire face once daily (in the evening) for 12 weeks. In addition to applying treatment to the face, subjects were allowed to apply treatment to the neck, upper chest, and/or upper back though these locations were not included in the efficacy evaluations.

[2]Note that one subject, ID = 027-1968 randomized to CT Gel did not have a baseline IGA score recorded.

3.1.2 Endpoints

The six point static Investigators Global Assessment (IGA) is outlined in the following table:

Value	Description
0	Normal, clear skin with no evidence of acne vulgaris.
1	Skin almost clear; rare non-inflammatory lesions present, with rare non-inflamed papules (papules must be resolving and may be hyperpigmented, though not pink-red) requiring no further treatment in the Investigator's opinion.
2	Some non-inflammatory lesions are present, with few inflammatory lesions (papules/pustules only, no nodulo-cystic lesions)
3	Non-inflammatory lesions predominate, with multiple inflammatory lesions evident; several to many comedones and papules/pustules, and there may or may not be 1 small nodulo-cystic lesion
4	Inflammatory lesions are more apparent; many comedones and papules/pustules, there may or may not be a few nodulo-cystic lesions
5	Highly inflammatory lesions predominate; variable number of comdedones, many papules/pustules and nodulo-cystic lesions

The count of lesions was to be performed by the investigator/designee at each study visit using the following guidelines:

1. If at all possible, the same qualified efficacy assessor should perform all of the lesion counts for a given subject as they progress through the study to ensure consistency.

2. Include only the face in these assessments: the face is defined as the hairline edge to the mandibular line.

3. Do not count non-inflammatory lesions (comedones) on the nose.

3.1.2.1 Co-Primary Efficacy Criteria The following are the *protocol defined* co-primary efficacy endpoints.

- The absolute change in lesion counts (total, inflammatory, non-inflammatory) from baseline to Week 12 (end of treatment).

- The proportion of subjects who have a minimum two grade improvement in their IGA score from baseline to Week 12 (end of treatment).

Superiority of CT Gel to Clindamycin Gel, Tretinoin Gel, and Vehicle Gel will be demonstrated based on absolute change in 2 out 3 lesion counts (total, inflammatory, non-inflammatory)

from baseline to Week 12, as well as on the proportion of subjects who have a minimum 2-grade improvement in their IGA score from baseline to Week 12.

Reviewer Comment: Note that based on the most recent submission of the protocol (SN064: stamp date 01/17/2008) the Agency recommended the following for demonstrating the efficacy of CT Gel.

> *"To justify the inclusion of multiple actives for a combination product for acne, the combination product should demonstrate efficacy over each monad for at least one lesion type, as well as the global. Rather than designing the study to find significance for '2 out of 3' lesion types for each comparison, the Agency recommends designing the study to demonstrate the following:*
>
> - *clindamycin/tretinoin [CT Gel] should be superior to vehicle for change in inflammatory and non-inflammatory lesions and success on the IGA*
> - *clindamycin/tretinoin [CT Gel] should be superior to clindamycin for success on the IGA and for one pre-specified lesion type or the protocol should specify a multiplicity adjusted method for demonstrating superiority for at least one lesion type*
> - *clindamycin/tretinoin [CT Gel] should be superior to tretinoin for success on the IGA and for one pre-specified lesion type (inflammatory or non-inflammatory) or the protocol should specify a multiplicity adjusted method for demonstrating superiority for at least one lesion type."*

Based upon the application history in DARRTS, it does not appear the above comments were sent to the sponsor.

3.1.2.2 Secondary Efficacy Criteria The following are the protocol defined secondary endpoints.

- Percent change in lesion counts (total, inflammatory, non-inflammatory) from baseline to Week 12.
- The proportion of subjects who have a subject's global assessment (SGA) score of 0 or 1 at Week 12.
- The proportion of subjects who have an IGA score of 0 or 1 at Week 12..

3.1.3 Statistical Methodology

The statistical methodologies described below correspond to those included in the protocol for Study W0265-03. Any deviations from protocol defined methods which will be incorporated into the statistical review are presented in *italics*. Note that efficacy results derived using methods defined in both the protocol and those by the reviewer will be presented in this review when different statistical approaches are used by the sponsor and reviewer for efficacy evaluation.

3.1.3.1 Populations The primary analysis population is defined as the intent-to-treat (ITT) population which includes all subjects who were randomized and dispensed study product. For supportive evidence, the per protocol (PP) population will be used for the analysis of the primary and secondary endpoints. The PP population excludes the following subjects.

- Subjects who miss more than 16 applications of the study product during the treatment period.
- Subjects who do not have efficacy evaluations at baseline and week 12.
- Subjects who used prohibited medications which are expected to interfere with the efficacy assessment.
- Subjects who were enrolled and did not meet the study inclusion and exclusion criteria.

3.1.3.2 Statistical Analysis All comparisons of CT Gel to Clindamycin Gel, Tretinoin Gel, and Vehicle Gel will be tested at the two-sided $\alpha = 0.05$ significance level. The study objective is to demonstrate the superiority of CT Gel over each component and its vehicle. The recruitment and enrollment goal of each center was to enroll a total of 70 subjects (10 subjects treated with Vehicle Gel and 20 subjects treated with either CT Gel, Clindamycin Gel, or Tretinoin Gel). Centers that enrolled fewer than 21 subjects are planned to be combined in the statistical analysis; such centers are referred to as 'analysis centers' in the body of the review.

Reviewer Comment: As it does not appear that the statistical comments made in SN064 were sent to the sponsor, the sponsor was not aware of the Agency's recommendations to address multiplicity for comparing lesion counts (see comments provided in Section 3.1.2.1 for a more detailed description of the Agency's recommendations). While the protocol failed to provide a pre-specification of the lesion type in which to compare CT Gel to each monad, based on the data provided in the previously conducted Phase 3 trials it would be reasonable to postulate the following comparisons on a specific lesion type would be most likely to demonstrate the efficacy of CT Gel over each monad.

- *CT Gel versus Clindamycin Gel: non-inflammatory lesions*

- *CT Gel versus Tretinoin Gel: inflammatory lesions*

Note that the data from the approved drug ZIANA which consists of the same two actives shows a similar trend as that observed in the previously conducted Phase 3 trials for CT Gel[1].

3.1.3.2.1 Investigator Global Assessment For the analysis on the percent of subjects with an IGA success, the protocol listed CMH stratified by "analysis center" as the primary method of statistical analysis to compare CT Gel to each of its monads and vehicle. A subject will be considered a success if the week 12 IGA score is at least 2 grades lower than the baseline assessment. The protocol listed primary method of data imputation is to impute all missing week 12 observations as failures with no proposed alternate imputations strategies to investigate the impact of imputing missing week 12 observations as failures.

Reviewer Comment: In the review of the protocol submitted in SN064, the sponsor was requested to propose sensitivity analyses to investigate the impact of the method of data imputation. However, these comments were not sent to the sponsor. As no sensitivity analyses were listed in the protocol and it is unclear if the proportion of subject discontinuation is balanced across treatment arms, the primary data imputation approach is LOCF with sensitivity analyses performed in which the week 12 IGA score will be imputed as a success in one analysis and as a failure in another. Additionally, as a secondary analysis, success on the IGA scale will be defined as a week 12 score of 0 or 1 with at least a two grade improvement from the baseline IGA score.

3.1.3.2.2 Lesion Counts The absolute change from baseline to week 12 in each of the lesion counts (total, inflammatory, non-inflammatory) will be analyzed using an ANOVA model with terms for treatment, 'analysis center', and treatment by 'analysis center' interaction. If the treatment by 'analysis center' interaction is not statistically significant at the 0.1 significance level, the interaction term will be dropped from the model. If the assumption of normality underlying the ANOVA model is not met (Shapiro-Wilks test at $\alpha = 0.05$), the ANOVA model will be fit on the rank transform of the absolute change with terms for treatment and 'analysis center'. LOCF is proposed as the method of data imputation with no additional imputation approaches proposed as sensitivity analyses.

Reviewer Comment: The ANOVA model proposed in the protocol does not account for the baseline lesion count nor was an alternate approached recommended by the Division in the review of the protocol. However, consistent with recent evaluations of acne products, the primary analysis method for absolute change in lesion counts will be ANCOVA with terms for treatment, 'analysis center' and the covariate baseline lesion count in this review. Note that while the protocol lists

ANOVA, the study report appears to also use an ANCOVA model as proposed by the reviewer with the dependent variable, absolute change in lesion count, being rank transformed.

Reviewer Comment: As a sensitivity analysis of the missing lesion count data the following two imputation approaches will be used in this review which are consistent with those used for assessing IGA scores.

- *Week 12 lesions counts will be imputed as the median lesion count for all IGA* **failures** *within the respective treatment group.*
- *Week 12 lesions counts will be imputed as the median lesion count for all IGA* **successes** *within the respective treatment group.*

Note that the choice of the median over the mean is that it is robust to nonsymmetric distributions and outlying observations.

Reviewer Comment: The protocol does not state whether the comparisons of CT Gel to each monad and vehicle use contrasts (i.e. the full covariance matrix) or if the data for the comparison of interest (e.g. CT Gel versus Clindamycin Gel) is used by itself. Analysis results reported based on reviewer analysis of the lesion counts is based on contrasts.

3.1.4 Patient Disposition and Baseline Characteristics

3.1.4.1 Patient Disposition In the one Phase 3 study a total of 203 out of 1649 randomized subjects (12.3%) discontinued from the trial (table 4). Overall, the rate of drop out in each trial was similar across treatment arms. However, both the CT Gel and Tretinoin Gel had higher rates of drop outs due to adverse events. The highest reported reason for dropout was 'Lost to follow-up' and 'Withdrawal by subject'.

For those that discontinued the trial and listed the reason as other, the rational for discontinuing include the following reasons: unable to attend all visits, did not satisfy entry criteria, did not use study drug, use of prohibited drug, pregnancy, incarcerated, and personal reasons. Note that these numbers are small and no apparent imbalance exists across the treatment arms.

3.1.4.2 Baseline Demographic Factors Of the 1649 subjects randomized in Study W0265-03, 58% of subjects were female; the average age was 20.4 years (median of 17 years); 73% were classified as white or Caucasian; and 83% were enrolled in the United States. No imbalances for any of the demographic characteristics were observed across the four treatment arms (table 5).

Table 4: Summary of Subject Completion/Discontinuation; Study W0265-03

	CT Gel ($N = 476$)	Clindamycin Gel ($N = 467$)	Tretinoin Gel ($N = 464$)	Vehicle Gel ($N = 242$)
Completed	414 (87.0)	416 (89.1)	405 (87.3)	211 (87.2)
Discontinued	62 (13.0)	51 (10.9)	59 (12.7)	31 (12.8)
Reason for Discontinuation				
Adverse event	6 (1.3)	0 (0.0)	5 (1.1)	0 (0.0)
Lack of efficacy	1 (0.2)	2 (0.4)	2 (0.4)	2 (0.8)
Lost to follow-up	26 (5.5)	19 (4.1)	17 (3.7)	11 (4.5)
Non-compliance with study drug	3 (0.6)	3 (0.6)	3 (0.6)	1 (0.4)
Other	7 (1.5)	9 (1.9)	6 (1.3)	4 (1.7)
Withdrawal by subject	19 (4.0)	18 (3.9)	26 (5.6)	13 (5.4)

Source: Study Report Table 4; reproduced by reviewer.

Table 5: **Baseline Demographic Factors by Treatment; Study W0265-03**

	CT Gel $N = 476$	Clindamycin Gel $N = 467$	Tretinoin Gel $N = 464$	Vehicle Gel $N = 242$
Age (years)	15 18 24	15 17 24	15 17 24	15 17 24
Sex : Male	43% (207)	43% (201)	43% (200)	37% (89)
Race :	0% (0)	0% (1)	0% (1)	0% (1)
American Indian or Alaska Native	1% (4)	1% (4)	1% (4)	1% (3)
Asian	2% (8)	2% (8)	2% (11)	2% (5)
Black or African American	20% (94)	18% (85)	22% (103)	21% (51)
Multiple	3% (12)	3% (16)	4% (19)	2% (5)
Native Hawaiian or Other Pacific Islander	0% (0)	1% (4)	0% (2)	0% (1)
White	75% (358)	75% (349)	70% (324)	73% (176)
Country : BLZ	7% (33)	6% (28)	7% (31)	7% (16)
CAN	11% (51)	10% (49)	10% (48)	11% (26)
USA	82% (392)	84% (390)	83% (385)	83% (200)

$a\ b\ c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables.
Numbers after percents are frequencies.

Source: Study Report Table 6 and Reviewer's analysis.

3.1.4.3 Baseline Prognostic Factors This exploratory analysis examines *baseline* characteristics of the clinical endpoints used in the evaluation of efficacy for acne vulgaris. Overall, the mean number of total lesions present at baseline was 70.9 of which the mean number of inflammatory lesions was 25.5 and non-inflammatory lesions 45.5. The majority of subjects had an IGA score of 3 (66.5% of subjects) and 52.5% had a subject's global score of 2. Distributions across the treatment arms for these prognostic factors were roughly balanced (table 6).

Table 6: **Baseline Prognostic Factors by Treatment; Study W0265-03**

	N	CT Gel $N = 476$	Clindamycin Gel $N = 467$	Tretinoin Gel $N = 464$	Vehicle Gel $N = 242$
Investigator Global	1647				
2		21% (100)	25% (117)	25% (114)	19% (47)
3		68% (324)	66% (306)	65% (303)	68% (164)
4		11% (50)	9% (44)	10% (47)	13% (31)
Subject's Global	1642				
0		0% (0)	1% (3)	0% (1)	0% (0)
1		12% (59)	14% (65)	11% (51)	9% (21)
2		49% (230)	52% (243)	56% (260)	55% (133)
3		34% (159)	28% (133)	27% (124)	31% (74)
4		5% (24)	5% (23)	5% (25)	6% (14)
Total Lesions	1648	51.00 62.00 81.50	49.00 61.00 81.00	51.00 64.00 83.25	51.00 67.00 85.75
Inflammatory Lesions	1648	20 24 30	20 24 30	21 24 30	20 24 30
Non-Inflammatory Lesions	1648	27.00 36.00 55.00	27.00 36.00 55.00	27.75 38.50 53.00	27.25 40.50 58.75

$a\ b\ c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables. N is the number of non–missing values. Numbers after percents are frequencies.

Source: Study Report Table 7 and Reviewer's analysis.

3.1.5 Populations Analyzed

The intent-to treat (ITT) population included all 1649 subjects who were randomized to treatment and who used study product (table 7). Note that 7 subjects were randomized to drug product but did not use study drug (2 subjects randomized to CT Gel, 2 subjects randomized to Clindamycin Gel, and 3 subjects randomized to Tretinoin Gel). The ITT population is the protocol-defined primary analysis population for the efficacy analyses. The per protocol (PP) population was a subset of the ITT population, which included those ITT subjects who met all major protocol criteria such as: all efficacy evaluations, did not use a prohibited medication, and did not satisfy enrollment criteria. A total of 1446 subjects (87.7%) of subjects were included in the PP analysis population.

Table 7: **Number of Subjects for each Analysis Population; Study W0265-03**

	CT Gel	Clindamycin Gel	Tretinoin Gel	Vehicle Gel
ITT Population	476	467	464	242
PP Population	402	404	389	206

Source: Study Report Table 3; reproduced by Reviewer.

3.1.6 Primary Endpoint Results (ITT-LOCF)

This section reports efficacy results for the ITT population in two sections. The first section provides efficacy results for the static IGA score in which the efficacy objective is to demonstrate that CT Gel is superior to each of its monads and vehicle. The second section presents efficacy results for each of the three lesion counts (total, inflammatory, non-inflammatory). Note that the method of data imputation in both sections is LOCF; use of alternate imputation strategies is provided in Section 3.1.8.1.

3.1.6.1 Investigator Global Assessment The protocol defines a success on the IGA scale as a week 12 score which is at least 2 grades improved from the baseline IGA score. As an alternate definition, and as a secondary analysis, one can define success as both a week 12 IGA score of 0 or 1 with at least a two grade reduction from baseline to week 12 which ensures that all subjects who reached an IGA score of 0 or 1 also had an improvement of the IGA score of 2 grades. Using CMH stratified by 'analysis center', CT Gel was compared to each monad and vehicle for both definitions of IGA success. Using both definitions of IGA success CT Gel was superior to each of its monads and vehicle based upon the two-sided $\alpha = 0.05$ level (table 8).

Table 8: **Investigator Global Results (ITT/LOCF); Study W0265-03**

	CT Gel ($N = 476$)	Clindamycin Gel ($N = 467$)	Tretinoin Gel ($N = 464$)	Vehicle Gel ($N = 242$)
Success: IGA Two Grade Improvement				
Success (%)	173 (36.3)	124 (26.6)	121 (26.1)	49 (20.2)
p-value†	-	0.0015	< 0.001	< 0.001
Success: IGA = 0 or 1 with Two Grade Imp.				
Success (%)	158 (33.2)	112 (24.0)	105 (22.6)	43 (17.8)
p-value†	-	0.0018	< 0.001	< 0.001

† CMH stratified by 'analysis center'.

Source: Reviewer's Analysis.

3.1.6.2 Change in Lesion Counts Recall that the Agency's recommendation to address multiplicity for comparing lesion counts was not addressed in the protocol (see comments provided in Section 3.1.2.1 for a more detailed description of the Agency's recommendations). While the protocol failed to provide a pre-specification of the lesion type in which to compare CT Gel to each monad, the following is postulated based upon the data observed with the original formulation:

- CT Gel will be superior to Clindamycin Gel for non-inflammatory lesions,
- CT Gel will be superior to Tretinoin Gel for inflammatory lesions, and
- CT Gel will be superior to Vehicle Gel for both inflammatory and non-inflammatory lesions.

In the sections that follow, efficacy results are provided for total, inflammatory, and non-inflammatory lesions.

3.1.6.2.1 Total Lesion Counts Figure 1 depicts a scatterplot for the total lesion counts at baseline (x-axis) and the change in total lesion counts (y-axis) for each treatment group. Within the figure, loess nonparametric regression lines[2] are provided. In this graphical summary of results across the range of the baseline total lesion count, CT Gel tends to have the largest change in lesion count (note that the sparse data for high baseline total counts impacts the loess regression such that it appears Tretinoin Gel has the greatest change in total lesion counts greater than 150 lesions).

An ANCOVA model fit the absolute change in total lesion counts with factors for treatment and 'analysis center' and the respective baseline total lesion count as a covariate. CT Gel was superior to each monad and vehicle on the basis of the absolute change in total lesion counts (table 9). Treatment effects were 4.7, 2.7, and 10.6 lesions when comparing CT Gel to Clindamycin Gel, Tretinoin Gel, and vehicle, respectively.

Figure 1: **Total Lesion Counts (ITT-LOCF)**

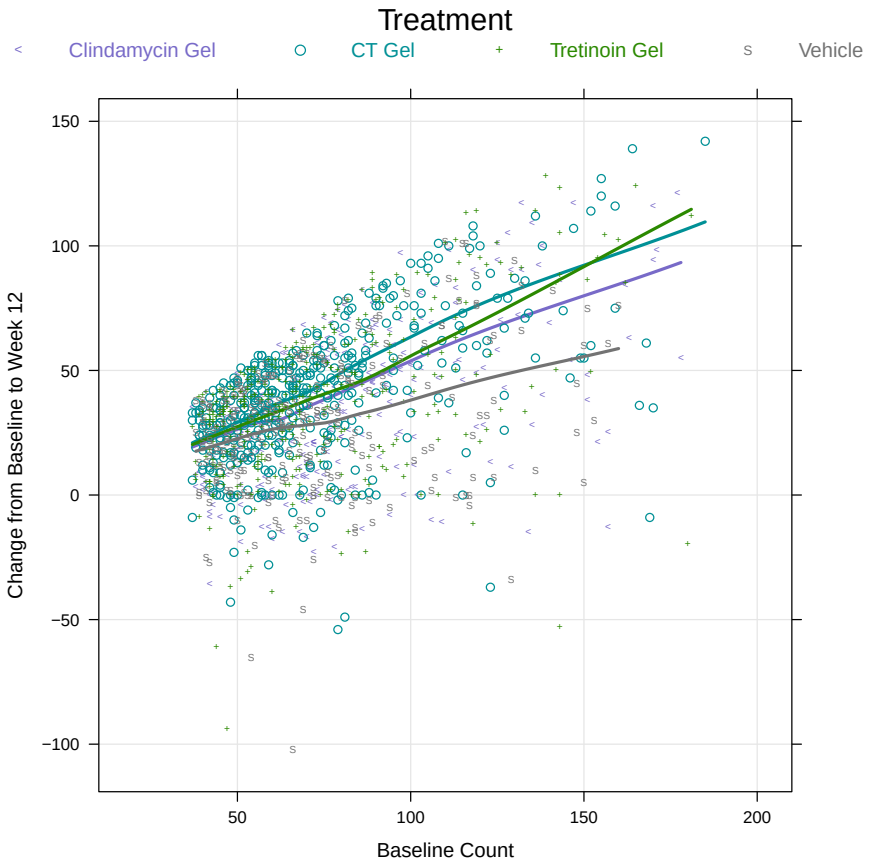


Table 9: **Total Lesion Counts (ITT/LOCF); Study W0265-03**

	CT Gel ($N = 476$)	Clindamycin Gel ($N = 467$)	Tretinoin Gel ($N = 464$)	Vehicle Gel ($N = 242$)
Mean Change (SD)	38.7 (26.8)	34.0 (25.2)	36.0 (28.3)	28.1 (27.7)
Percent Change (SD)	55.0 (30.5)	49.0 (30.1)	50.5 (34.8)	39.1 (36.5)
p-value†	-	0.0028	0.037	< 0.001
p-value‡	-	< 0.001	0.014	< 0.001

† p-value is based on an ANCOVA models with terms for treatment and 'analysis center' with covariate for the baseline lesion count (Source: reviewer's analysis).

‡ p-value is based on a ranked ANCOVA model as reported by the sponsor (Source: study report Table 01.03).

3.1.6.2.2 Inflammatory Lesion Counts An ANCOVA model fit the absolute change in inflammatory lesion counts with factors for treatment and 'analysis center' and the respective baseline inflammatory lesion count as a covariate. The observed treatment effects were 1.0, 1.6,

and 4.4 lesions when comparing CT Gel to Clindamycin Gel, Tretinoin Gel, and vehicle, respectively. The treatment effects comparing CT Gel to Tretinoin Gel and vehicle were significant using both the reviewer's analysis and the sponsor's analysis. The comparison of CT Gel to Clindamycin Gel did not reach the $\alpha = 0.05$ significance level for either the sponsor's or the reviewer's analysis.

Table 10: **Inflammatory Lesion Counts (ITT/LOCF); Study W0265-03**

	CT Gel ($N = 476$)	Clindamycin Gel ($N = 467$)	Tretinoin Gel ($N = 464$)	Vehicle Gel ($N = 242$)
Mean Change (SD)	15.5 (10.3)	14.5 (9.4)	13.9 (11.1)	11.1 (11.7)
Percent Change (SD)	60.4 (36.3)	56.5 (45.0)	54.5 (39.2)	43.3 (44.9)
p-value†	-	0.1797	0.0022	< 0.001
p-value‡	-	0.076	0.0003	< 0.001

† p-value is based on an ANCOVA models with terms for treatment and 'analysis center' with covariate for the baseline lesion count (Source: reviewer's analysis).

‡ p-value is based on a ranked ANCOVA model as reported by the sponsor (Source: study report Table 01.01).

3.1.6.2.3 Non-Inflammatory Lesion Counts An ANCOVA model fit the absolute change in non-inflammatory lesion counts with factors for treatment and 'analysis center' and the respective baseline inflammatory lesion count as a covariate. The observed treatment effects were 3.7, 1.1, and 6.2 lesions when comparing CT Gel to Clindamycin Gel, Tretinoin Gel, and vehicle, respectively. The treatment effects comparing CT Gel to Clindamycin Gel and vehicle were significant using both the reviewer's analysis and the sponsor's analysis. The comparison of CT Gel to Tretinoin Gel did not reach the $\alpha = 0.05$ significance level for either the sponsor's or the reviewer's analysis.

3.1.7 Primary Endpoint Results (PP-LOCF)

Recall from Table 7, a total of 1401 subjects are included in the per protocol (PP) analysis population, accounting for approximately 85% of subjects enrolled. The following efficacy results are presented for these 1401 subjects in the PP population.

Table 11: **Non-Inflammatory Lesion Counts (ITT/LOCF); Study W0265-03**

	CT Gel ($N = 476$)	Clindamycin Gel ($N = 467$)	Tretinoin Gel ($N = 464$)	Vehicle Gel ($N = 242$)
Mean Change (SD)	23.2 (20.4)	19.5 (19.7)	22.1 (21.7)	17.0 (20.6)
Percent Change (SD)	51.0 (33.0)	42.9 (36.6)	47.3 (43.5)	36.0 (39.1)
p-value†	-	0.0011	0.2541	< 0.001
p-value‡	-	0.003	0.531	< 0.001

† p-value is based on an ANCOVA models with terms for treatment and 'analysis center' with covariate for the baseline lesion count (Source: reviewer's analysis).

‡ p-value is based on a ranked ANCOVA model as reported by the sponsor (Source: study report Table 01.02).

3.1.7.1 Investigator Global Assessment Using the same definitions of IGA success as with the ITT populations, the following section depicts the efficacy results for the PP population. Trends in treatment effects observed in the ITT population are also observed in the PP population though point estimates for all treatment arms are slightly higher in the PP population (table 12). Consistent with the ITT population, CT Gel was superior to each of its monads and vehicle in the PP population for IGA.

Table 12: **Investigator Global Results (PP); Study W0265-03**

	CT Gel ($N = 402$)	Clindamycin Gel ($N = 404$)	Tretinoin Gel ($N = 389$)	Vehicle Gel ($N = 206$)
Success: IGA Two Grade Improvement				
Success (%)	166 (41.3)	116 (28.7)	115 (29.6)	47 (22.8)
p-value†	-	< 0.001	< 0.001	< 0.001
Success: IGA = 0 or 1 AND Two Grade Imp.				
Success (%)	153 (38.1)	104 (25.7)	102 (26.2)	41 (19.9)
p-value†	-	< 0.001	< 0.001	< 0.001

† CMH stratified by 'analysis center'.

Source: Reviewer's Analysis.

3.1.7.2 Change in Lesion Counts The analysis of the absolute change in lesions counts utilizes an ANCOVA model with terms for treatment and 'analysis center' and the baseline lesion count as the covariate. The change in lesion counts for all lesion types is presented in Table 13 along with the corresponding p-value. For the PP population, the observed treatment effects

were consistently higher for all comparisons of CT Gel to its monads and vehicle over those reported for the PP population. Despite the smaller sample size with the PP population, the comparison of CT Gel to Clindamycin Gel reaches the nominal significance level of $\alpha = 0.05$ for inflammatory lesions in the PP population whereas the p-value observed in the ITT population was 0.1797. All other comparisons of CT Gel to its monads and vehicle are consistent with the ITT population results.

Table 13: **Change in Lesion Counts (PP); Study W0265-03**

	CT Gel ($N = 402$)	Clindamycin Gel ($N = 404$)	Tretinoin Gel ($N = 389$)	Vehicle Gel ($N = 206$)
Total Lesions				
Mean Change (SD)	42.9 (25.1)	36.5 (24.9)	39.0 (27.5)	31.1 (27.1)
Percent Change (SD)	61.0 (25.7)	52.2 (28.9)	54.8 (33.6)	43.9 (35.2)
p-value†	-	< 0.001	0.0106	< 0.001
Inflammatory Lesions				
Mean Change (SD)	17.1 (9.3)	15.5 (9.1)	15.2 (10.7)	12.4 (11.4)
Percent Change (SD)	66.5 (30.8)	61.4 (31.6)	59.2 (37.5)	48.0 (43.6)
p-value†	-	0.0263	0.0006	< 0.001
Non-Inflammatory Lesions				
Mean Change (SD)	25.8 (19.9)	21.0 (19.8)	23.9 (21.2)	18.7 (20.3)
Percent Change (SD)	56.8 (29.3)	45.8 (35.9)	51.2 (43.4)	40.4 (38.0)
p-value†	-	< 0.001	0.1356	< 0.001

† p-value is based on an ANCOVA models with terms for treatment and 'analysis center' with covariate for the baseline lesion count (Source: reviewer's analysis).

3.1.8 Additional Sensitivity Analyses of Primary Endpoints

3.1.8.1 Missing Data Sensitivity Analysis The following sensitivity analyses are used to assess the impact of the primary method of data imputation (LOCF) on the co-primary endpoints. These analyses are based upon reviewer determinations as no sensitivity analyses were proposed in the protocol since this recommendation to conduct a sensitivity analyses for the method of data imputation were not conveyed to the sponsor. Note that the analysis population is ITT for the following sensitivity analyses.

3.1.8.1.1 Investigator Global Assessment Recall that no sensitivity analyses were included in the protocol for assessing the impact of missing data on the efficacy conclusion. In Section 3.1.6.1 efficacy results were presented using LOCF to impute missing data. In the current section two alternate imputation approaches are used: (a) impute all missing week 12

data as failures and (b) impute all missing week 12 data as successes. Note that the results presented in this section utilize an IGA definition of success as a week 12 IGA score of 0 or 1 with at least a 2 grade improvement from baseline. Using these alternate imputation approaches, CT Gel was superior to each monad and vehicle (table 14). Note that point estimates for the LOCF analyses are slightly higher than when imputing missing values as failures. Thus, results based upon these sensitivity analyses were consistent with the LOCF analysis of IGA.

Table 14: **Investigator Global Results (Missing Sensitivity); Study W0265-03**

	CT Gel ($N = 476$)	Clindamycin Gel ($N = 467$)	Tretinoin Gel ($N = 464$)	Vehicle Gel ($N = 242$)
Missing Imputed as Failures				
Success (%)	154 (32.4)	108 (23.1)	103 (22.2)	41 (16.9)
p-value†	-	0.0016	< 0.001	< 0.001
Missing Imputed as Success				
Success (%)	220 (46.2)	160 (34.3)	168 (36.2)	74 (30.6)
p-value†	-	< 0.001	< 0.001	< 0.001

† CMH stratified by 'analysis center'.

Source: Reviewer's Analysis defining success as a week 12 IGA score of 0 or 1 with at least a two grade improvement.

3.1.8.1.2 Change in Lesion Counts Recall that the protocol does not include a sensitivity analysis to assess the impact of the data imputation procedure (LOCF for the primary analysis). However, as described in Section 3.1.3.2.2 the review will conduct two imputation approaches for lesions counts which are consistent with those used for assessing IGA scores.

- Week 12 lesions counts will be imputed as the median lesion count for all IGA **failures** within the respective treatment group (denote this as "Median Failures").
- Week 12 lesions counts will be imputed as the median lesion count for all IGA **successes** within the respective treatment group (denote this as "Median Successes").

Point estimates for the mean absolute change and standard deviation for each lesion count are presented in Table 15 along with p-values based on the ANCOVA model when missing data are imputed as the median lesion count for the IGA *failures*. The point estimates are consistently higher for all treatment arms using such an imputation approach versus using LOCF. However, the comparison of *total* lesion counts between CT Gel and Tretinoin Gel does not reach the nominal $\alpha = 0.05$ significance level when imputing using "Median Failures" ($p = 0.0563$) whereas such a comparison did reach the nominal significance level when imputing

the missing data using LOCF. All other comparisons using the“ Median Failure” approach to data imputation are consistent with imputing missing data using LOCF.

Table 15: **Change in Lesion Counts (Median Failures); Study W0265-03**

	CT Gel ($N = 476$)	Clindamycin Gel ($N = 467$)	Tretinoin Gel ($N = 464$)	Vehicle Gel ($N = 242$)
Total Lesions				
Mean Change (SD)	42.1 (26.0)	36.0 (25.1)	40.0 (27.9)	30.9 (27.6)
Percent Change (SD)	58.9 (25.9)	51.1 (28.0)	54.7 (31.6)	42.1 (34.1)
p-value†	-	< 0.001	0.0563	< 0.001
Inflammatory Lesions				
Mean Change (SD)	16.6 (9.4)	15.3 (8.9)	15.3 (10.3)	12.2 (11.0)
Percent Change (SD)	64.4 (30.9)	59.1 (42.9)	59.1 (34.9)	47.0 (40.8)
p-value†	-	0.0324	0.0045	< 0.001
Non-Inflammatory Lesions				
Mean Change (SD)	25.8 (20.9)	20.9 (20.4)	25.0 (22.4)	19.0 (21.5)
Percent Change (SD)	55.2 (28.9)	44.5 (35.0)	51.5 (40.9)	38.4 (37.7)
p-value†	-	< 0.001	0.3000	< 0.001

† p-value is based on an ANCOVA models with terms for treatment and ‘analysis center’ with covariate for the baseline lesion count (Source: reviewer’s analysis).

Point estimates for the mean absolute change and standard deviation for each lesion count are presented in Table 16 along with p-values based on the ANCOVA model when missing data are imputed as the median lesion count for the IGA *successes*. As expected, the point estimates are consistently higher for all treatment arms using such an imputation approach versus imputing all missing data using “Median Failures”. Similar to the results when imputing missing data using “Median Failures”, the comparison of *total* lesion counts between CT Gel and Tretinoin Gel does not reach the nominal $\alpha = 0.05$ significance level when imputing using “Median Successes” ($p = 0.1076$). All other comparisons using the “Median Success” approach to data imputation are consistent with imputing missing data using LOCF.

3.1.9 Secondary Endpoint Results

No secondary endpoints have been proposed by the sponsor to be included in product labeling. As such, no further analyses are conducted on clinical endpoints assessed during the clinical trial. Note that within the review of the primary endpoints, point estimates of the mean percent change are provided for Study W0265-03.

Table 16: **Change in Lesion Counts (Median Successes); Study W0265-03**

	CT Gel ($N = 476$)	Clindamycin Gel ($N = 467$)	Tretinoin Gel ($N = 464$)	Vehicle Gel ($N = 242$)
Total Lesions				
Mean Change (SD)	44.3 (26.3)	37.9 (25.5)	42.5 (29.1)	34.3 (29.0)
Percent Change (SD)	62.5 (25.6)	54.3 (28.2)	58.4 (32.5)	47.4 (34.6)
p-value†	-	< 0.001	0.1076	< 0.001
Inflammatory Lesions				
Mean Change (SD)	17.5 (9.5)	16.0 (9.1)	16.1 (10.5)	13.5 (11.3)
Percent Change (SD)	68.6 (31.4)	61.9 (43.4)	62.5 (35.8)	52.3 (42.2)
p-value†	-	0.0087	0.0028	< 0.001
Non-Inflammatory Lesions				
Mean Change (SD)	26.9 (21.1)	22.0 (20.6)	26.2 (23.0)	20.9 (22.4)
Percent Change (SD)	58.3 (28.2)	47.8 (34.6)	54.6 (41.3)	43.6 (37.1)
p-value†	-	< 0.001	0.3898	< 0.001

† p-value is based on an ANCOVA models with terms for treatment and 'analysis center' with covariate for the baseline lesion count (Source: reviewer's analysis).

3.2 Consistency of Efficacy Findings - All Phase 3 Trials

The following sections provide efficacy results of all Phase 3 clinical trials to assess the consistency in efficacy findings. It should be noted that in the SPA review of the Phase 3 protocol for Study W0265-03, the Division stated that if the efficacy results of CT Gel Gel are similar to efficacy results observed in Studies VLC.C.304 and VLC.C.305, this would be sufficient to support the new formulation. A detailed review of Studies VLC.C.304 and VLC.C.305 is provided in the statistical review dated 4/26/2005. Results are presented for the ITT population with missing data imputed using LOCF. Definition of the endpoints is provided in each of the relevant sections.

Note that the sponsor did not provide a formal statistical analysis plan for comparing efficacy data from all Phase 3 trials nor was the efficacy data presented integrating data from all Phase 3 trials. Thus, the following presentation of efficacy results are based upon reviewer analysis only.

3.2.1 Investigator Global Results - All Phase 3 Trials

Efficacy results of all Phase 3 trials on the IGA scale utilize a definition of success as the percent of subjects with a week 12 IGA score of 0 or 1 with at least a two grade improvement. Note that this definition of success was not incorporated into the statistical review of Studies

VLC.C.304 and VLC.C.305 as success was defined in the protocol as an IGA score of 0 or 1 (details provided in the statistical review signed on 4/26/2005). Efficacy results below are based upon within study comparisons and do not include any statistical adjustments for cross-study comparisons.

Utilizing the above definition of IGA success, treatment effects comparing the combination gel to each of its monads and vehicle were similar for Study VLC.C.304 and W0265-03 (Table 17). However, based upon this definition of treatment success, the observed treatment effects were smaller in Study VLC.C.305 than in the other Phase 3 studies[3]. Also note that the estimates of success rate are highest in Study W0265-03. Overall, the trends in treatment effects were consistent across each of the three studies in which the smallest treatment effect occurred when comparing CT Gel to Clindamycin Gel. The largest treatment effect was observed when comparing CT Gel to Vehicle Gel.

Table 17: **Investigator Global Results (ITT/LOCF); All Trials**

	Combination Gel	Clindamycin Gel	Tretinoin Gel	Vehicle Gel
Study VLC.C.304				
X/N (%)	72/309 (23.3)	43/311 (13.8)	36/310 (11.6)	12/153 (7.8)
Trt. Effect ($\hat{\delta}$)	-	9.4	11.6	15.1
95% CI for $\hat{\delta}$	-	(3.3, 15.5)	(5.7, 17.5)	(8.7, 21.5)
Study VLC.C.305				
X/N (%)	95/325 (29.2)	81/324 (25.0)	79/325 (24.3)	18/162 (11.1)
Trt. Effect ($\hat{\delta}$)	-	4.2	4.9	17.8
95% CI for $\hat{\delta}$	-	(-2.6, 11)	(-1.9, 11.7)	(10.8, 24.7)
Study W0265-03				
X/N (%)	158/476 (33.2)	112/467 (24.0)	105/464 (22.6)	43/242 (17.8)
Trt. Effect ($\hat{\delta}$)	-	9.2	10.5	15.2
95% CI for $\hat{\delta}$	-	(3.4, 14.9)	(4.8, 16.2)	(8.8, 21.6)

Source: Reviewer's analysis. Confidence intervals are unadjusted and based upon Wilson estimates of the percent of successes.

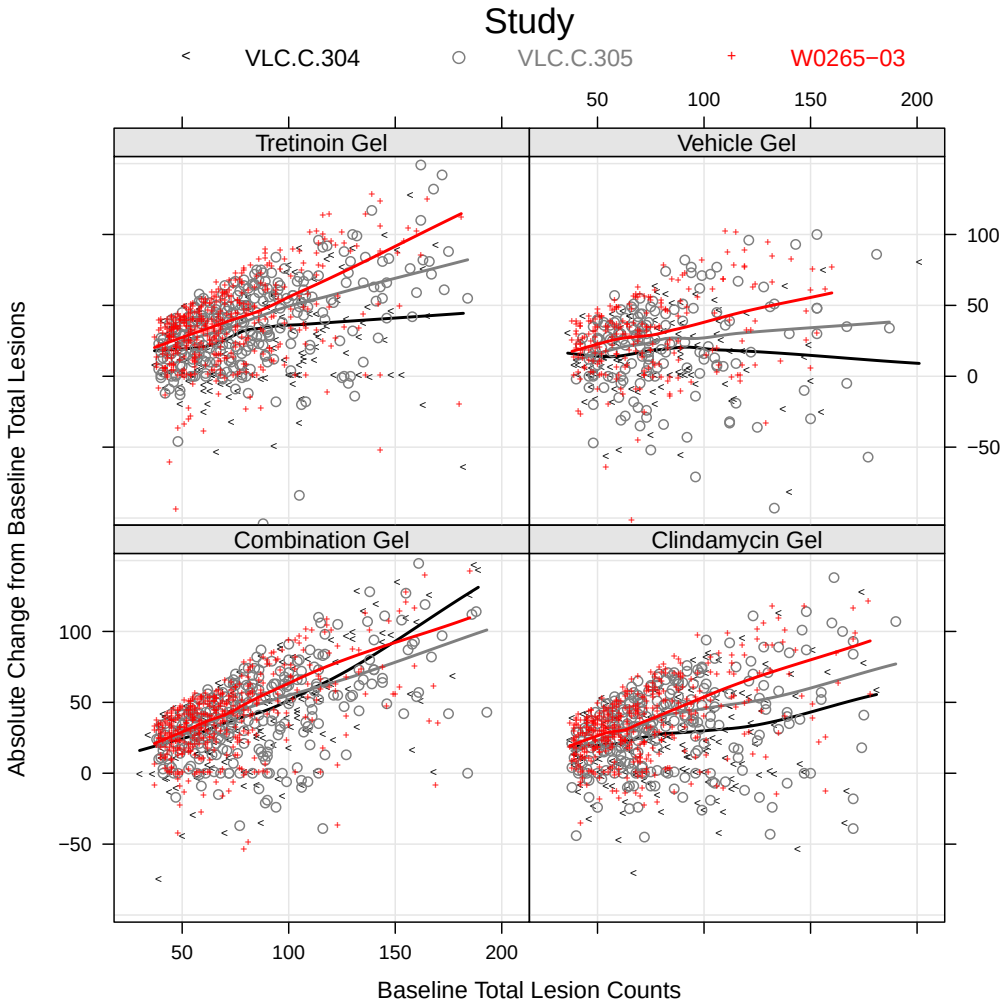
3.2.2 Lesion Counts - All Phase 3 Trials

To visualize the lesion count data across all studies, Figure 2 was constructed in which the baseline total lesion count is plotted against the end of treatment absolute change in total lesion counts where the data is paneled on the treatment and studies are shown as the grouping

[3]Using the above definition of IGA success, this reduced the number of successes for the combination product by 33 subjects versus defining IGA success as a score of 0 or 1

variable. Also provided in each panel is a loess nonparametric regression lines as a summary of the relationship between baseline and absolute change for each study. Similar to the ISE efficacy results using the IGA score, the absolute change for total lesion counts is roughly the greatest in Study W0265-03 for all treatment groups. A descriptive summary of inflammatory and non-inflammatory lesion types is provided below.

Figure 2: **Total Lesion Counts (ISE)**



In the subsequent sections, the absolute changes in lesion counts are assessed using a linear model which is constructed for each study. The model is of the form:

$$\Delta \text{ Lesion} = \beta_0 + \beta_1 \times \text{Base Lesion} + \beta_2 \times \text{Pooled Site ID} + \beta_3 \times \text{Treatment}. \quad (1)$$

Note that in the calculations of the estimates of β, only the comparison of CT Gel to a single comparator is used in which the reference drug is CT Gel. Thus, $\hat{\beta}_3$ can be interpreted as the mean change in lesion counts when comparing CT Gel to its comparator. In addition to the

estimate of β_3, a 95% confidence interval for $\hat{\beta}_3$ is provided. Also note that a term for pooled site is incorporated into the model and not site as this is the same approach used in the primary analysis of all Phase 3 studies.

3.2.2.1 Total Lesion Counts Using the model shown in Equation 1, estimates of β_3 were calculated comparing CT Gel to each monad and vehicle within each study. The estimates and the corresponding unadjusted 95% CI are provided in Table 18 for total lesion counts. While the trend shown in Figure 2 suggest the greatest change in total lesion counts for all treatment arms in Study W0265-03, based upon the model, treatment effects for total lesion counts were smallest in Study W0265-03.

Table 18: Absolute Change in Total Lesions (ITT/LOCF); All Trials

	Combo Vs. Clindamycin	Combo Vs. Tretinoin	Combo Vs. Vehicle
Study VLC.C.304			
$\hat{\beta}_3$	11.1	9.2	21.8
95% CI for $\hat{\beta}_3$	(7.0, 15.3)	(5.2, 13.3)	(16.2, 27.4)
Study VLC.C.305			
$\hat{\beta}_3$	7.5	4.7	21.9
95% CI for $\hat{\beta}_3$	(3.7, 11.2)	(1.1, 8.3)	(16.8, 27.1)
Study W0265-03			
$\hat{\beta}_3$	4.1	2.8	11.8
95% CI for $\hat{\beta}_3$	(1.5, 6.7)	(0.0, 5.6)	(8.4, 15.2)

Source: Reviewer's Analysis. Estimates and confidence intervals based on model as shown in equation 1.

3.2.2.2 Inflammatory Lesion Counts Using the model shown in Equation 1, estimates of β_3 were calculated comparing CT Gel to each monad and vehicle within each study. The estimates and the corresponding unadjusted 95% CI are provided in Table 19 for inflammatory lesion counts. Consistent with results observed for total lesions, Study W0265-03 had the smallest estimated treatment effect for all comparisons for inflammatory lesions. Overall the general trend for all Phase 3 studies is that the treatment effect is greatest comparing CT Gel to Vehicle with the smallest treatment effect occurring when comparing CT Gel to Clindamycin Gel.

Table 19: Absolute Change in Inflammatory Lesions (ITT/LOCF); All Trials

	Combo Vs. Clindamycin	Combo Vs. Tretinoin	Combo Vs. Vehicle
Study VLC.C.304			
$\hat{\beta}_3$	1.8	3.5	6.4
95% CI for $\hat{\beta}_3$	(0.3, 3.3)	(2.1, 5)	(4.5, 8.3)
Study VLC.C.305			
$\hat{\beta}_3$	1.8	2.1	6.0
95% CI for $\hat{\beta}_3$	(0.3, 3.3)	(0.7, 3.5)	(4.0, 8.1)
Study W0265-03			
$\hat{\beta}_3$	0.7	1.7	4.5
95% CI for $\hat{\beta}_3$	(-0.3, 1.8)	(0.6, 2.9)	(3.1, 6.0)

Source: Reviewer's Analysis. Estimates and confidence intervals based on model as shown in equation 1.

3.2.2.3 Non-Inflammatory Lesion Counts Using the model shown in Equation 1, estimates of β_3 were calculated comparing CT Gel to each monad and vehicle within each study. The estimates and the corresponding unadjusted 95% CI are provided in Table 20 for non-inflammatory lesion counts. Consistent with results observed for total lesions, Study W0265-03 had the smallest estimated treatment effect for all comparisons of non-inflammatory lesions. Overall the general trend for all Phase 3 studies is that the treatment effect is greatest comparing CT Gel to Vehicle with the smallest treatment effect occurring when comparing CT Gel to Tretinoin Gel.

Table 20: Absolute Change in Non-Inflammatory Lesions (ITT/LOCF); All Trials

	Combo Vs. Clindamycin	Combo Vs. Tretinoin	Combo Vs. Vehicle
Study VLC.C.304			
$\hat{\beta}_3$	9.3	5.7	15.4
95% CI for $\hat{\beta}_3$	(6.0, 12.6)	(2.5, 8.9)	(10.8, 19.9)
Study VLC.C.305			
$\hat{\beta}_3$	5.6	2.5	15.9
95% CI for $\hat{\beta}_3$	(2.7, 8.4)	(-0.1, 5.2)	(11.9, 19.9)
Study W0265-03			
$\hat{\beta}_3$	3.4	1.1	7.4
95% CI for $\hat{\beta}_3$	(1.4, 5.3)	(-0.9, 3.1)	(4.9, 9.8)

Source: Reviewer's Analysis. Estimates and confidence intervals based on model as shown in equation 1.

3.3 Evaluation of Safety

As the basis for the non-approval was due to the carcinogenicity signal in a Tg.AC mouse dermal carcinogenicity study, the product was reformulated for safety reasons. Therefore the following summaries of safety information are based only on Study W0265-03.

3.3.1 MedDRA Tabulation

Verbatim descriptions of adverse events were assigned to system organ classifications (SOC's) and preferred terms (PT's) using the MedDRA dictionary (version 11.0). The following summaries report all reported AE's during the trial, those related to the study drug, and those recorded as being serious in nature. In the tabulations, subjects are only counted once for each individual PT.

3.3.1.1 All Reported Adverse Events Regardless of whether the adverse event was listed as being related to study drug or not, Table 21 depicts the MedDRA preferred terms which were depicted in at least 2 subjects for a given treatment arm. PT's are reported within the respective SOC. The most frequently reported PT was nasopharyngitis which was reported in more than 5% of subjects. For the general disorders and administrative site conditions SOC, both CT Gel and Tretinoin Gel report several application site irritations whereas subjects treated with Clindamycin Gel and vehicle do not report these events.

Table 21: **Adverse Events by System Organ Class and Preferred Term (Study W0265-03)**

	CT Gel ($N = 476$)	Clindamycin Gel ($N = 467$)	Tretinoin Gel ($N = 464$)	Vehicle ($N = 242$)
Gastrointestinal disorders				
Abdominal pain upper	1 (0.2)	3 (0.6)	0 (0.0)	0 (0.0)
Diarrhoea	1 (0.2)	3 (0.6)	0 (0.0)	0 (0.0)
Nausea	1 (0.2)	3 (0.6)	4 (0.9)	2 (0.8)
Vomiting	1 (0.2)	2 (0.4)	4 (0.9)	1 (0.4)
Abdominal pain	0 (0.0)	2 (0.4)	0 (0.0)	0 (0.0)
Stomach discomfort	0 (0.0)	2 (0.4)	0 (0.0)	1 (0.4)
General disorders and administration site conditions				
Application site dryness	9 (1.9)	0 (0.0)	13 (2.8)	0 (0.0)
Application site dermatitis	6 (1.3)	0 (0.0)	6 (1.3)	1 (0.4)
Application site erythema	5 (1.1)	0 (0.0)	6 (1.3)	0 (0.0)
Application site irritation	5 (1.1)	0 (0.0)	13 (2.8)	0 (0.0)
Application site pruritus	2 (0.4)	0 (0.0)	2 (0.4)	0 (0.0)
Application site exfoliation	0 (0.0)	0 (0.0)	9 (1.9)	0 (0.0)
Application site pain	0 (0.0)	0 (0.0)	4 (0.9)	0 (0.0)
Infections and infestations				
Nasopharyngitis	27 (5.7)	35 (7.5)	27 (5.8)	19 (7.9)
Pharyngitis streptococcal	4 (0.8)	2 (0.4)	2 (0.4)	1 (0.4)
Sinusitis	4 (0.8)	1 (0.2)	4 (0.9)	0 (0.0)
Influenza	3 (0.6)	0 (0.0)	3 (0.6)	1 (0.4)
Rhinitis	3 (0.6)	1 (0.2)	2 (0.4)	3 (1.2)
Gastroenteritis viral	2 (0.4)	1 (0.2)	0 (0.0)	1 (0.4)
Bronchitis	1 (0.2)	1 (0.2)	2 (0.4)	1 (0.4)
Gastroenteritis	1 (0.2)	4 (0.9)	2 (0.4)	0 (0.0)
Upper respiratory tract infection	1 (0.2)	3 (0.6)	1 (0.2)	0 (0.0)
Injury, poisoning and procedural complications				
Sunburn	6 (1.3)	1 (0.2)	1 (0.2)	2 (0.8)
Procedural pain	1 (0.2)	4 (0.9)	1 (0.2)	0 (0.0)
Concussion	0 (0.0)	0 (0.0)	2 (0.4)	0 (0.0)
Musculoskeletal and connective tissue disorders				
Myalgia	3 (0.6)	2 (0.4)	1 (0.2)	1 (0.4)
Back pain	0 (0.0)	1 (0.2)	2 (0.4)	0 (0.0)
Nervous system disorders				
Headache	16 (3.4)	18 (3.9)	14 (3.0)	5 (2.1)
Psychiatric disorders				
Depression	0 (0.0)	2 (0.4)	1 (0.2)	0 (0.0)
Reproductive system and breast disorders				
Dysmenorrhoea	1 (0.2)	6 (1.3)	5 (1.1)	0 (0.0)
Respiratory, thoracic and mediastinal disorders				
Cough	5 (1.1)	6 (1.3)	3 (0.6)	1 (0.4)
Pharyngolaryngeal pain	3 (0.6)	8 (1.7)	4 (0.9)	2 (0.8)
Sinus congestion	2 (0.4)	0 (0.0)	0 (0.0)	0 (0.0)
Nasal congestion	1 (0.2)	3 (0.6)	2 (0.4)	2 (0.8)

Source: Reviewer's Analysis using AE.XPT

3.3.1.2 Related Adverse Events A total of 59 subjects reported adverse events that were listed as being related to study drug[4]. This is a subset of the PT's reported in Table 21. A tabulation of those events which were listed as related to study drug are reported in Table 22 where all events listed as related to study drug are reported. The majority of adverse events are from the general disorders and administration site conditions SOC; these predominate in subjects treated with CT Gel and Tretinoin Gel. Among these events, there appears to be a slightly higher incidence rate in subjects treated with Tretinoin Gel over those treated with CT Gel. A further exploration of local skin reactions is provided in Section 3.3.2.1.

Table 22: **Related Adverse Events by System Organ Class and Preferred Term (Study W0265-03)**

	CT Gel	Clindamycin Gel	Tretinoin Gel	Vehicle
Gastrointestinal disorders				
Lip dry	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)
Lip swelling	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)
Cheilitis	0 (0.0)	0 (0.0)	1 (0.2)	0 (0.0)
General disorders and administration site conditions				
Application site dryness	9 (1.9)	0 (0.0)	13 (2.8)	0 (0.0)
Application site dermatitis	5 (1.1)	0 (0.0)	6 (1.3)	1 (0.4)
Application site erythema	5 (1.1)	0 (0.0)	6 (1.3)	0 (0.0)
Application site irritation	4 (0.8)	0 (0.0)	13 (2.8)	0 (0.0)
Application site pruritus	2 (0.4)	0 (0.0)	2 (0.4)	0 (0.0)
Application site rash	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)
Application site exfoliation	0 (0.0)	0 (0.0)	9 (1.9)	0 (0.0)
Application site pain	0 (0.0)	0 (0.0)	4 (0.9)	0 (0.0)
Thirst	0 (0.0)	1 (0.2)	0 (0.0)	0 (0.0)
Injury, poisoning and procedural complications				
Sunburn	2 (0.4)	1 (0.2)	1 (0.2)	0 (0.0)
Nervous system disorders				
Headache	0 (0.0)	1 (0.2)	0 (0.0)	0 (0.0)
Skin and subcutaneous tissue disorders				
Skin hyperpigmentation	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)
Hyperhidrosis	0 (0.0)	1 (0.2)	0 (0.0)	0 (0.0)

Source: Study Report Table 21; Results reproduced by reviewer using AE.XPT.

[4]Related events were defined as being either: possibly related, probably related, or definitely related.

3.3.2 Serious Adverse Events

A total of six subjects reported an adverse event that was listed as serious. Table 23 depicts each subject and the reported adverse event along with timing of the event and whether the event was listed as being related to study drug or not as determined by the investigator. Overall, none of the serious adverse events were listed as being related to study drug with 5 of 6 occurring while on treatment.

Table 23: **Serious Adverse Events (Study W0265-03)**

Subject ID	Treatment	Event	AE Start†	AE End†	On Treatment	Related
001-1003	Tretinoin	Concussion	54	92	Yes	Not related
002-1259	Tretinoin	Right ovarian cyst	66	67	Yes	Note related
004-1695	Tretinoin	Suicide attempt	67	67	Yes	Not related
005-2205	CT Gel	Infectious Mononucleosis	52	65	Yes	Not related
010-1340	Clindamycin	Abdominal pain	14	16	Yes	Not related
026-1495	Vehicle	Acute bronchitis	81	-	No	Note related

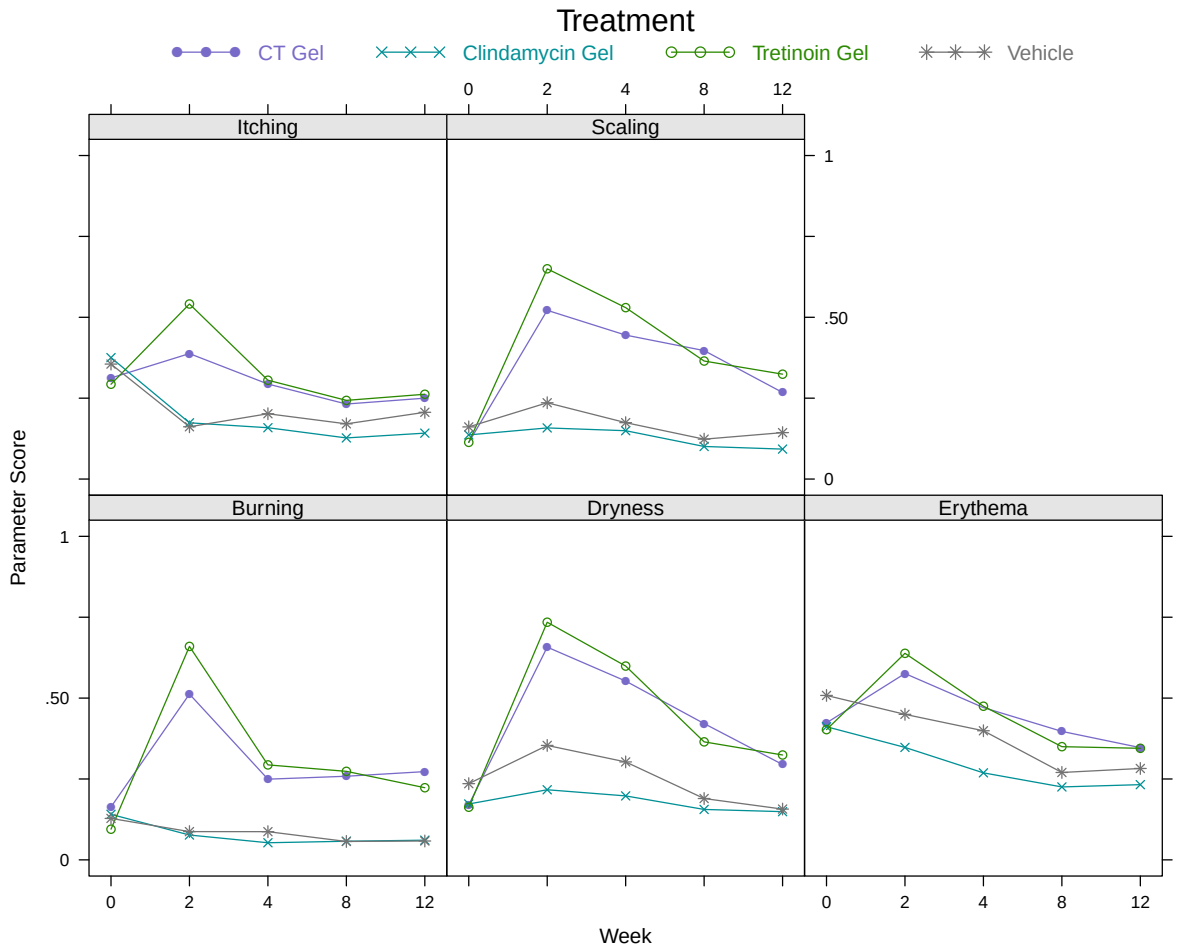
† Listed in days from study entry.

Source: Study Report Table 24 and Reviewer Analysis using AE.XPT

.

3.3.2.1 Local Skin Reactions Five local skin reactions: dryness, burning, erythema, itching, and scaling were actively assessed at each visit with scores of 0 to 5 with *no descriptors* for each level. The mean score over all subjects was calculated at each visit for each treatment group. Figure 3 depicts the mean profile over time for each study and local skin reaction grouped by treatment arm. For all local skin reactions the mean profile of CT Gel is similar to that of Tretinoin Gel which are both greater than that of Clindamycin Gel and vehicle. During the twelve weeks of treatment the peak of the mean for each local reaction is at week 2 with a gradual reduction thereafter. For CT Gel and Tretinoin Gel, the week 12 values are near that of their baseline values, though slightly higher.

Figure 3: **Local Skin Reactions (W0265-03)**



4 Findings in Special/Subgroup Populations

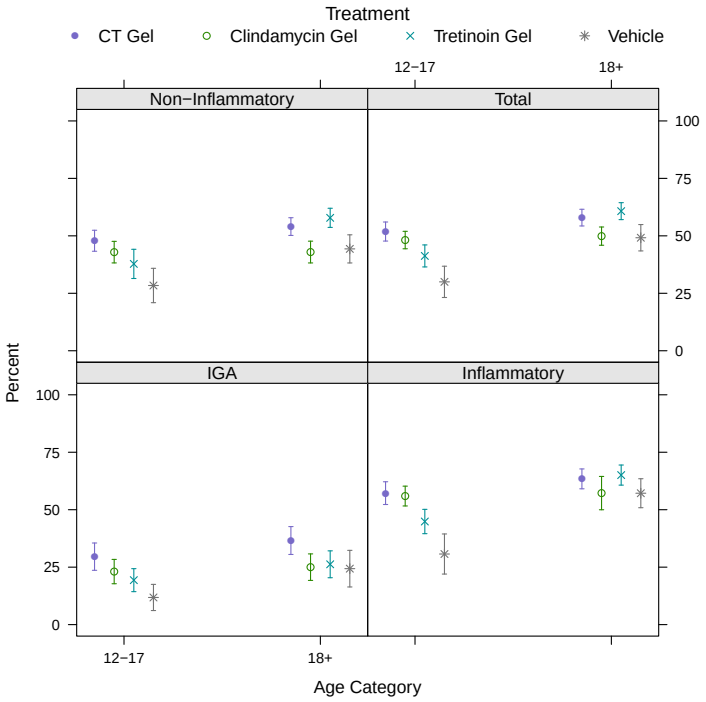
Section 4.1 provides a graphical assessment of efficacy by age, gender, and race for Study W0265-03. Rather than using the absolute change from baseline in the graphical depictions, the percent change is used as this would be on a similar scale as the percent of IGA successes (Week 12 IGA score of 0 or 1 with at least a two grade improvement from baseline). The analysis population is the ITT population with missing week 12 observations imputed using LOCF as in the primary analysis. Note that the protocol did not pre-specify any subgroup analysis which controlled the overall Type I error rate. For a tabular presentation of the data refer to the Appendix, Section A.1.1.

4.1 Gender, Race, and Age

4.1.1 Age

The age of subjects was dichotomized into two categories: 12 to 17 years old and 18 years and older (median age was 17 years old). Figure 4 depicts efficacy results according to age category along with unadjusted 95% confidence intervals. In general, subjects who were 18 and older tended to have slightly higher efficacy than subjects 12 to 17 years old. Interestingly, for subjects 12 to 17 years old, the efficacy for Clindamycin Gel was higher than Tretinoin Gel for all co-primary endpoints.

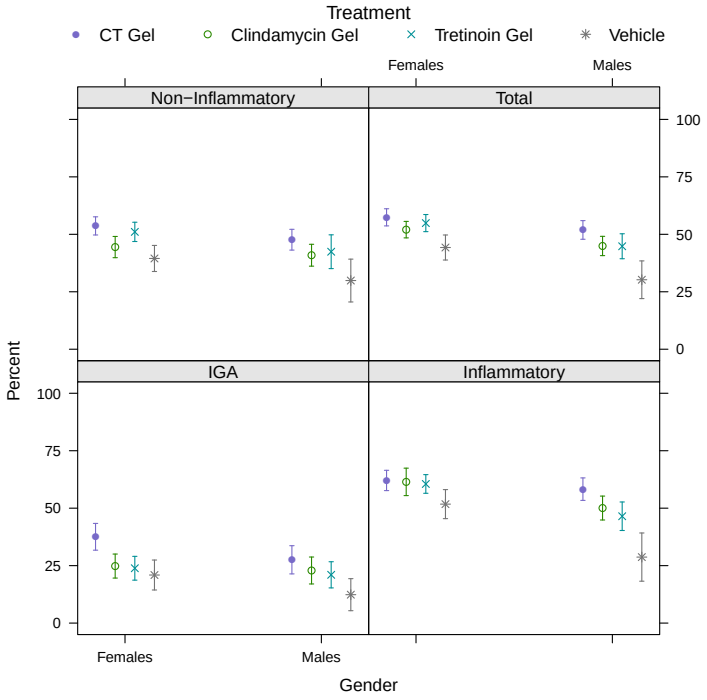
Figure 4: **Efficacy Results According to Age**



4.1.2 Gender

Figure 5 depicts efficacy results according to gender along with unadjusted 95% confidence intervals. For IGA successes, there are larger treatment effects for females treated with CT Gel versus its monads and vehicle than in males. For inflammatory lesions, females had similar mean percent changes when treated with CT Gel and the two actives, whereas males had a higher mean percent change when treated with CT Gel than when treated with the monads. Patterns are similar in males and females for both non-inflammatory and total lesion mean percent changes.

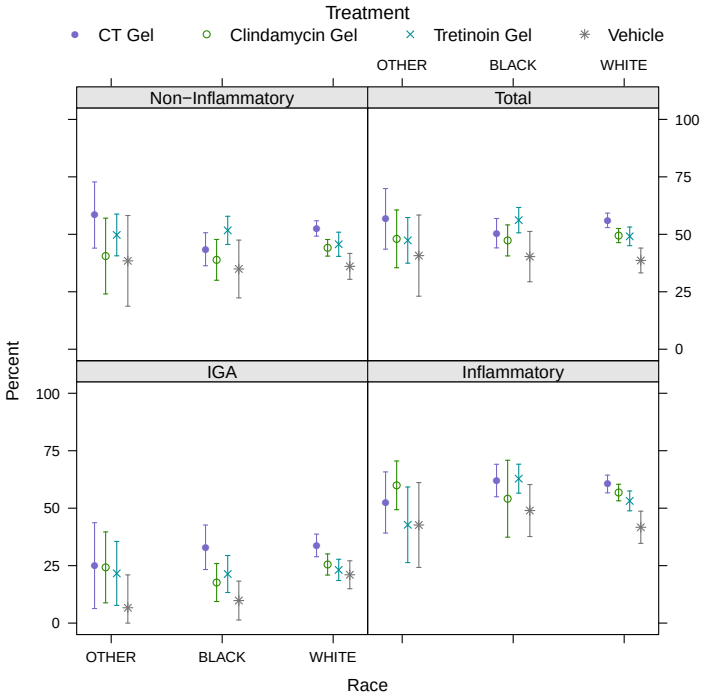
Figure 5: **Efficacy Results According to Gender**



4.1.3 Race

Race was dichotomized into three categories: White, Black, and Other where Other consisted of those with multiple races recorded, native American, or Native Hawaiian or Pacific Islander (table 5). Figure 6 depicts the means for each endpoint along with unadjusted 95% confidence intervals by race for each of the co-primary endpoints. Overall the efficacy results were quite consistent across subgroups which show the highest means in subjects treated with CT Gel.

Figure 6: **Efficacy Results According to Race**

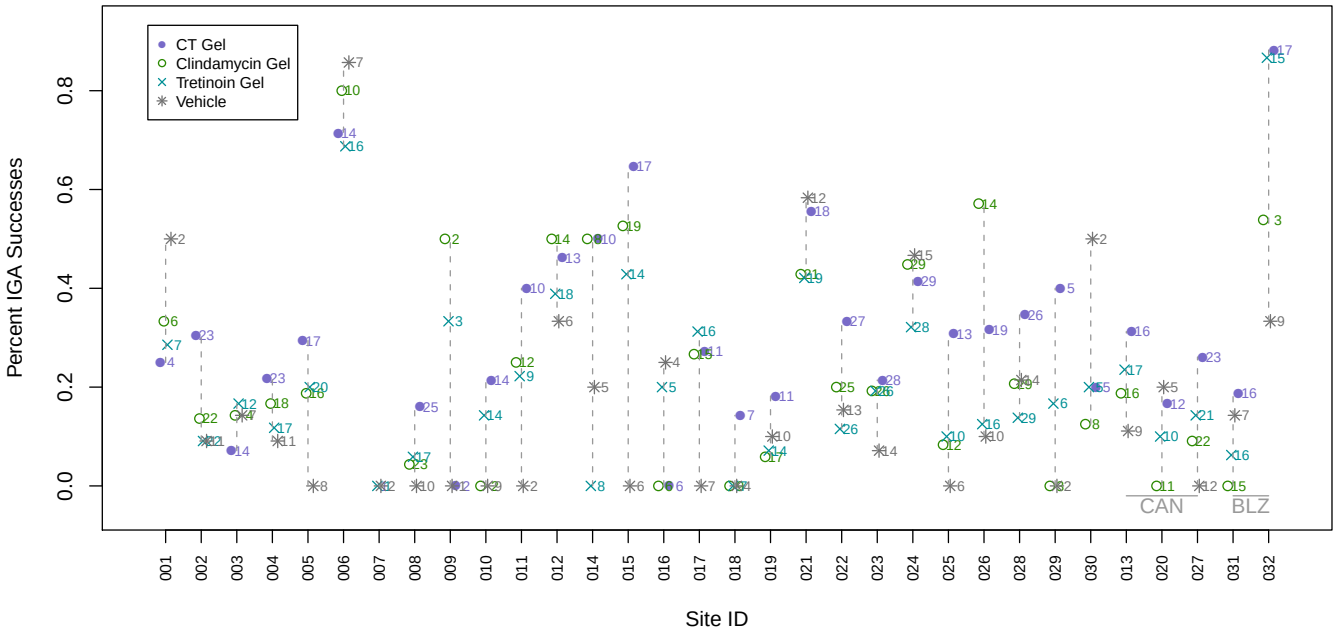


4.2 Other Special/Subgroup Populations

4.2.1 Efficacy by Site

Study W0265-03 was conducted in 32 centers in the U.S., Canada, and Belize. Using the primary analysis population (ITT-LOCF), Figure 7 depicts the mean percent of IGA successes (week 12 score of 0 or 1 with at least a two grade improvement) for each site along with the number of subjects enrolled for a given treatment arm within a site. Note that the three sites from Canada and the two sites from Belize are designated in Figure 7. The figure shows that efficacy results for the IGA endpoint varies across sites, and in some sites, the reported percent of successes for Vehicle is higher than the active treatment arms. No apparent reasons exist to explain the degree of variability observed.

Figure 7: Efficacy Results By Investigative Site (Study W0265-03)



5 Summary and Conclusions

5.1 Statistical Issues and Collective Evidence

Following the not approval (NA) action of June 10, 2005, the sponsor and Agency had several discussions about the future development of a reformulated combination drug product (see, Section 2.2 for a discussion of the regulatory history following the NA action). It was agreed that a single four arm trial with the objective of establishing the superiority of Veltin™ over each monad and vehicle would be sufficient for an efficacy claim if the efficacy results of the reformulated product, Veltin™, are similar to those observed with the original formulation.

During the protocol review of Study W0265-03, several comments which impact the evaluation of efficacy were not incorporated into the protocol as the comments do not appear to have been sent to the sponsor. These are summarized as follows.

Multiplicity adjustments : To justify the inclusion of multiple actives for a combination product for acne, the combination product should demonstrate efficacy over each monad for at least one lesion type, as well as the global. At a minimum, the following should be demonstrated:

- CT Gel should be superior to vehicle for change in inflammatory and non-inflammatory

lesions and success on the IGA;

- CT Gel should be superior to clindamycin for success on the IGA and for one pre-specified lesion type or the protocol should specify a multiplicity adjusted method for demonstrating superiority for at least one lesion type;
- CT Gel should be superior to tretinoin for success on the IGA and for one pre-specified lesion type or the protocol should specify a multiplicity adjusted method for demonstrating superiority for at least one lesion type.

IGA success definition : Improving from an IGA score of 5 to an IGA score of 3 would not be sufficient for success. If it is appropriate to enroll subjects with an IGA score of 5 at baseline, these subjects should improve to a score of 2 or better at Week 12.

As the above comments were not communicated to the sponsor this review utilized a multiplicity adjustment which was based on results from the completed Phase 3 trials of the original formulation. In addition, a stringent definition of IGA success was utilized in which subjects needed a week 12 score of 0 or 1 with at least a two grade reduction to be considered a treatment success.

In the evaluation of efficacy, the following criteria were used in this review to determine the efficacy of Veltin™ Gel for Study W0265-03.

IGA : Success = week 12 score of 0 or 1 with at least a two grade improvement. The objective for assessing IGA success was to establish the

- superiority of Veltin™ Gel over each monad, Clindamycin Gel and Tretinoin Gel, and Vehicle Gel.

Lesion Counts : The absolute change from baseline to week 12. The objective for assessing lesion counts was to establish the

- superiority of Veltin™ Gel over Clindamycin Gel for total and non-inflammatory lesions,
- superiority of Veltin™ Gel over Tretinoin Gel for total and inflammatory lesions, and
- superiority of Veltin™ Gel over Vehicle Gel for total, inflammatory, and non-inflammatory lesions.

Testing was performed at the two-sided $\alpha = 0.05$ significance level. Analysis of the IGA success rates was performed using CMH stratified on analysis centers. Comparisons of the absolute changes was conducted using ANCOVA with terms for analysis centers and treatment and baseline counts as the covariate.

In Study W0265-03, Veltin™ Gel was superior to each monad and vehicle for IGA success. Veltin™ Gel was also superior to each monad and vehicle for total lesions. For inflammatory lesions, Veltin™ Gel was superior to Tretinoin Gel and vehicle, but not superior to Clindamycin Gel. For non-inflammatory lesions, Veltin™ Gel was superior to Clindamycin Gel and vehicle, but not for Tretinoin Gel (Table 24).

Table 24: **Efficacy Summary (ITT/LOCF); Study W0265-03**

	Veltin™ Gel ($N = 476$)	Clindamycin Gel ($N = 467$)	Tretinoin Gel ($N = 464$)	Vehicle Gel ($N = 242$)
IGA Success: 0 or 1 with Two Grade Imp.				
Success (%)	158 (33.2)	112 (24.0)	105 (22.6)	43 (17.8)
p-value†	-	0.0018	< 0.001	< 0.001
Absolute Change Total Lesions				
Mean Change (SD)	38.7 (26.8)	34.0 (25.2)	36.0 (28.3)	28.1 (27.7)
p-value‡	-	0.0028	0.037	< 0.001
Absolute Change Inflammatory Lesions				
Mean Change (SD)	15.5 (10.3)	14.5 (9.4)	13.9 (11.1)	11.1 (11.7)
p-value‡	-	0.1797	0.0022	< 0.001
Absolute Change Non-Inflammatory Lesions				
Mean Change (SD)	23.2 (20.4)	19.5 (19.7)	22.1 (21.7)	17.0 (20.6)
p-value‡	-	0.0011	0.2541	< 0.001

† CMH stratified by 'analysis center' (Source: reviewer's analysis).

‡ p-value is based on an ANCOVA models with terms for treatment and 'analysis center' with covariate for the baseline lesion count (Source: reviewer's analysis).

To assess the consistency of efficacy findings across all Phase 3 trials; analyses were conducted to compare the efficacy of the reformulated product with that of original formulation. Treatment effects for the IGA endpoint were similar in Study VLC.C.304 and W0265-03 which were higher than in Study VLC.C.305. For assessing changes in lesion counts, the treatment effects in Study W0265-03 were smaller than in both Studies VLC.C.304 and VLC.C.305 (results of these analyses are presented in Section 3.2). Overall, efficacy trends of all co-primary endpoints were consistent across all Phase 3 trials.

In the safety assessment of local skin reactions for Study W0265-03, on average, the irritability of Veltin™ Gel was similar to Tretinoin Gel. While the mean irritation was not severe in nature, the mean irritation peaked at week 2 with a gradual decrease towards baseline levels by week 12. Adverse events, other than local skin reactions, were not discordant in subjects treated with Veltin™ Gel versus the monads and vehicle. Note that safety results are only provided for Study W0265-03 due to the observed carcinogenicity signal of the original formulation.

5.2 Conclusions and Recommendations

Veltin™ Gel is a combination drug product comprising clindamycin, 1% and tretinoin, 0.025% in a topical gel for the treatment of acne vulgaris. An earlier formulation of the drug product, Velac Gel, (b) (4) was found to have a positive carcinogenicity signal during the initial filing of the NDA. On June 10, 2005 the sponsor was issued a not approval letter on the basis of this finding. The sponsor reformulated the product (b) (4).

On the basis of the newly formulated product the sponsor conducted an additional Phase 3 trial (Study W0265-03) which included the combination drug product, each monad, and the vehicle gel. The protocol-defined primary efficacy endpoints were:

- The absolute change in lesion counts (total, inflammatory, non-inflammatory) from baseline to Week 12 (end of treatment); and
- The proportion of subjects who have a minimum two grade improvement in their IGA score from baseline to Week 12 (end of treatment).

In Study W0265-03, Veltin™ Gel was statistically superior to each monad and vehicle for the percent of IGA successes and the absolute change in total lesion counts. For the comparison of inflammatory lesions, Veltin™ Gel was superior to Tretinoin Gel and vehicle, however was not significantly superior to Clindamycin Gel ($p = 0.1797$). For non-inflammatory lesion counts, Veltin™ Gel was superior to Clindamycin Gel and vehicle, but did not reach statistical significance for the comparison to Tretinoin Gel ($p = 0.2541$).

To assess the consistency in efficacy findings of the newly formulated drug product, Veltin™ Gel, to that of the old formulation, Veltin Gel, the efficacy of Study W0265-03 was compared descriptively to the efficacy in Studies VLC.C.304 and VLC.C.305 for all co-primary endpoints. The following summarizes the efficacy for each of the co-primary endpoints across trials/formulations.

IGA Success : Treatment effects of the combination product versus the monads were highest in Study VLC.C.304 which were similar to those observed in Study W0265-03. Treatment effects of the combination product versus vehicle were similar for all studies.

Lesion Counts : A linear model was used to derive estimates of treatment effects for lesion counts in each study. Trends were similar across trials, though the magnitude of the treatment effect was smallest in Study W0265-03 for all lesion counts.

While the magnitude of the treatment effects for lesion counts were smallest in Study W0265-03, the observed treatment effects for the IGA endpoint for Study W0265-03 were similar to that of Study VLC.C.304 which had the largest treatment effects. Overall, efficacy trends of all co-primary endpoints were consistent across all Phase 3 trials.

In the safety assessment of local skin reactions, on average, the irritability of Veltin™ Gel was similar to Tretinoin Gel. While the mean irritation was not severe in nature, the mean irritation peaked at week 2 with a gradual decrease towards baseline levels by week 12. Adverse events, other than local skin reactions, were not discordant in subjects treated with Veltin™ versus the monads and vehicle.

References

[1] Ziana label approved on 11/07/2006. Retrieved February 18, 2010, from `http://www.accessdata.fda.gov/Scripts/cder/DrugsatFDA/index.cfm`.

[2] Cleveland, W.S. *The Elements of Graphing Data*. Hobart Press, Summit, New Jersey, 1985.

[3] Statistical Analysis and Graphics produced with R software. R Development Core Team (2007). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org`.

Appendix

A.1 Supplementary Information for Study W0265-03

A.1.1 Efficacy Tables by Subgroups

The following tables present point estimates of efficacy by age, gender, and race for Study W0265-03. In the reporting of lesion counts, the means and standard deviation are based upon absolute change from baseline whereas results presented in Section 4.1 were based upon the percent reduction from baseline. Note that no statistical comparisons are made as the results are presented for descriptive purposes only.

A.1.1.1 Investigator Global Assessment Consistent with the definition of IGA success used in the main body of the review, the following section defines IGA success as a week 12 score of 0 or 1 with at least a two grade improvement. Table 25 depicts efficacy results for the IGA successes (percent of successes and fraction of successes). The analysis population was the ITT population with missing data imputed using LOCF.

Table 25: **Investigator Global Results (ITT-LOCF) by Subgroup**

	CT Gel	Clindamycin Gel	Tretinoin Gel	Vehicle
Age Category				
12 - 17 years	29.6 $\frac{68}{230}$	23.1 $\frac{57}{247}$	19.3 $\frac{47}{243}$	11.8 $\frac{15}{127}$
18+ years	36.6 $\frac{90}{246}$	25.0 $\frac{55}{220}$	26.2 $\frac{58}{221}$	24.3 $\frac{28}{115}$
Gender				
Females	37.5 $\frac{101}{269}$	24.8 $\frac{66}{266}$	23.9 $\frac{63}{264}$	20.9 $\frac{32}{153}$
Males	27.5 $\frac{57}{207}$	22.9 $\frac{46}{201}$	21.0 $\frac{42}{200}$	12.4 $\frac{11}{89}$
Race				
Other	25.0 $\frac{6}{24}$	24.2 $\frac{8}{33}$	21.6 $\frac{8}{37}$	6.7 $\frac{1}{15}$
Black	33.0 $\frac{31}{94}$	17.6 $\frac{15}{85}$	21.4 $\frac{22}{103}$	9.8 $\frac{5}{51}$
White	33.8 $\frac{121}{358}$	25.5 $\frac{89}{349}$	23.1 $\frac{75}{324}$	21.0 $\frac{37}{176}$

Source: Reviewer's Analysis.

A.1.1.2 Change in Total Lesion Counts Table 26 depicts efficacy results for the endpoint: absolute change from baseline for total lesions in Study W0265-03. Estimates for the mean and standard deviation (shown in parentheses) are depicted within each table. The analysis population was the ITT population with missing data imputed using LOCF.

Table 26: **Change in Total Lesions (ITT-LOCF) by Subgroup**

	CT Gel	Clindamycin Gel	Tretinoin Gel	Vehicle
Age Category				
12 - 17 years	38.2 (28.9)	34.2 (24.9)	30.7 (28.9)	23.4 (29.5)
18+ years	39.2 (24.8)	33.7 (25.5)	41.8 (26.4)	33.3 (24.6)
Gender				
Females	39.8 (27.2)	35.8 (25.8)	38.3 (27.3)	32.3 (27.2)
Males	37.4 (26.4)	31.5 (24.2)	32.9 (29.3)	20.8 (27.1)
Race				
Other	38.7 (30.1)	37.5 (32)	40.7 (26.6)	30.5 (31.3)
Black	42.2 (26.1)	37.8 (29.7)	36 (27.2)	26.2 (21.6)
White	38.5 (26.0)	32.8 (22.7)	34.5 (28.8)	27.6 (27.1)

Source: Reviewer's Analysis.

A.1.1.3 Change in Inflammatory Lesion Counts Table 27 depict efficacy results for the endpoint: absolute change from baseline for inflammatory lesions in Study W0265-03. Estimates for the mean and standard deviation (shown in parentheses) are depicted within each table. The analysis population was the ITT population with missing data imputed using LOCF.

Table 27: **Change in Inflammatory Lesions (ITT-LOCF) by Subgroup**

	CT Gel	Clindamycin Gel	Tretinoin Gel	Vehicle
Age Category				
12 - 17 years	15.1 (10.7)	14.4 (9.7)	11.9 (11.8)	8.3 (13.1)
18+ years	15.9 (10.0)	14.6 (9.1)	16.2 (9.7)	14.2 (9.0)
Gender				
Females	15.5 (10.4)	15.7 (8.6)	15.0 (9.3)	13.2 (10.4)
Males	15.6 (10.3)	13.0 (10.2)	12.5 (12.9)	7.6 (13.0)
Race				
Other	15.7 (10.3)	15.5 (9.3)	16.0 (9.6)	12.2 (10.8)
Black	13.5 (8.8)	14.6 (8.3)	9.9 (15.1)	10.6 (8.2)
White	15.6 (10.5)	14.3 (9.5)	13.7 (10.8)	10.8 (12.2)

Source: Reviewer's Analysis.

A.1.1.4 Change in Non-Inflammatory Lesion Counts Table 28 depict efficacy results for the endpoint: absolute change from baseline for non-inflammatory lesions in Study W0265-03. Estimates for the mean and standard deviation (shown in parentheses) are depicted within each table. The analysis population was the ITT population with missing data imputed using LOCF.

Table 28: **Change in Non-Inflammatory Lesions (ITT-LOCF) by Subgroup**

	CT Gel	Clindamycin Gel	Tretinoin Gel	Vehicle
Age Category				
12 - 17 years	23.1 (22.4)	19.8 (19.7)	18.8 (21.9)	15.1 (22)
18+ years	23.3 (18.4)	19.1 (19.8)	25.6 (21.0)	19.0 (18.8)
Gender				
Females	24.3 (20.8)	20.2 (20.6)	23.3 (21.9)	19.2 (20.5)
Males	21.8 (19.9)	18.5 (18.5)	20.4 (21.4)	13.2 (20.3)
Race				
Other	23.0 (22.9)	22.0 (26.3)	24.7 (20.8)	18.2 (25.1)
Black	28.7 (21.5)	23.2 (23.9)	26.1 (19.2)	15.6 (16.3)
White	22.9 (19.7)	18.5 (17.3)	20.8 (22.2)	16.7 (19.5)

Source: Reviewer's Analysis.

Signatures/Distribution List

Primary Statistical Reviewer: Mat Soukup, Ph.D.
Date: June 3, 2010

Statistical Team Leader: Mohamed Alosh, Ph.D.

cc:
Archival NDA
DDDP/Walker
DDDP/Kettl
DDDP/Chiang
DDDP/Petruccelli
OBIO/Patrician
DB3/Wilson
DB3/Alosh
DB3/Soukup

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-50803	ORIG-1	STIEFEL A GSK CO	Veltin

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

/s/

MATTHEW J SOUKUP
06/03/2010

MOHAMED A ALOSH
06/03/2010
Concur with the review and conclusion.

Mid-Cycle Review (Biostatistics)

NDA Number: 50-803/resubmission
Applicant: Stiefel Laboratories
Drug Name: Veltin (clindamycin 1%, tretinoin 0.025%) Gel
Indication: Acne
PDUFA Date: 04/16/2010
Statistical Reviewer: Mat Soukup, Ph.D., DDDP

1 Major Findings/Approval Issues

Confirmation of sponsor's major conclusions or identification of statistical issues that might preclude approval.

1.1 Study Description

Phase 3 Study: Study W0265-03

A single four-arm study was conducted with the new formulation to demonstrate the superiority of Veltin (i.e. CT) Gel versus clindamycin gel, tretinoin gel, and vehicle gel. From a statistical perspective, the *design* of the study appears to be in agreement with the Division.

1.2 Primary Analysis

1.2.1 Primary Endpoint

The co-primary endpoints for the study are the following:

IGA : Success is defined as a two-grade improvement from baseline to week 12

Lesion Counts : Absolute change in 2 out of 3 lesions

1.2.2 Secondary Endpoints

The secondary endpoints are the percent change in lesion counts and the proportion of subjects with an subject global assessment of 0 or 1.

1.2.3 Analysis Plan Summary

Change in lesion counts will be analyzed with an ANOVA model with terms for treatment, center and treatment by center interaction (the interaction will be deleted if not significant at 0.1). If the assumption of normality is not met (via Shapiro-Wilks test at 0.05), then a rank

transform will be used. Success on the ISGA will be analyzed for the 4 treatment arms with a Cochran-Mantel-Haenszel test stratified on center. If the overall test is significant at 0.05 then the combination arm will be compared with each of the other arms at 0.05.

Missing values for the lesion count analyses will be imputed with last observation carried forward. Missing values for treatment success will be counted as failures. The ITT population is defined as all subjects randomized and dispensed study product. The per protocol population will exclude subjects who miss more than 16 applications, do not have efficacy evaluations at baseline and Week 12, use prohibited medications, or are enrolled in error with respect to inclusion or exclusion criteria.

1.2.4 Analysis Results

1.2.4.1 Investigator Global Assessment Defining success using multiple definitions, the following table presents the efficacy results.

Table 1: Investigator Global Results (ITT/LOCF); Study W0265-03

	CT Gel ($N = 476$)	Clindamycin Gel ($N = 467$)	Tretinoin Gel ($N = 464$)	Vehicle Gel ($N = 242$)
Success: IGA Two Grade Improvement				
Success (%)	173 (36.3)	124 (26.6)	121 (26.1)	49 (20.2)
p-value†	-	0.0014	<.001	<.001
Success: IGA = 0 or 1				
Success (%)	205 (43.1)	171 (36.6)	157 (33.8)	55 (22.7)
p-value†	-	0.0476	0.0011	<.001
Success: IGA = 0 or 1 AND Two Grade Imp.				
Success (%)	158 (33.2)	112 (24.0)	105 (22.6)	43 (17.8)
p-value†	-	0.0017	<.001	<.001

† CMH stratified by pooled site.

Source: Reviewer's Analysis.

1.2.4.2 Lesion Counts Not yet analyzed.

1.2.4.3 Secondary Endpoint Analysis Not yet assessed.

2 Disposition

Summarization of patient disposition for major studies (patients randomized, completed, dropouts, etc.).

Table 2: Summary of Subject Completion/Discontinuation

	CT Gel ($N = 476$)	Clindamycin Gel ($N = 467$)	Tretinoin Gel ($N = 464$)	Vehicle Gel ($N = 242$)
Completed	414 (87.0)	416 (89.1)	405 (87.3)	211 (87.2)
Discontinued	62 (13.0)	51 (10.9)	59 (12.7)	31 (12.8)
Reason for Discontinuation				
Adverse event	6 (1.3)	0 (0.0)	5 (1.1)	0 (0.0)
Lack of efficacy	1 (0.2)	2 (0.4)	2 (0.4)	2 (0.8)
Lost to follow-up	26 (5.5)	19 (4.1)	17 (3.7)	11 (4.5)
Non-compliance with study drug	3 (0.6)	3 (0.6)	3 (0.6)	1 (0.4)
Other	7 (1.5)	9 (1.9)	6 (1.3)	4 (1.7)
Withdrawal by subject	19 (4.0)	18 (3.9)	26 (5.6)	13 (5.4)

Source: Study Report Table 4; reproduced by reviewer.

3 Missing Data Investigation

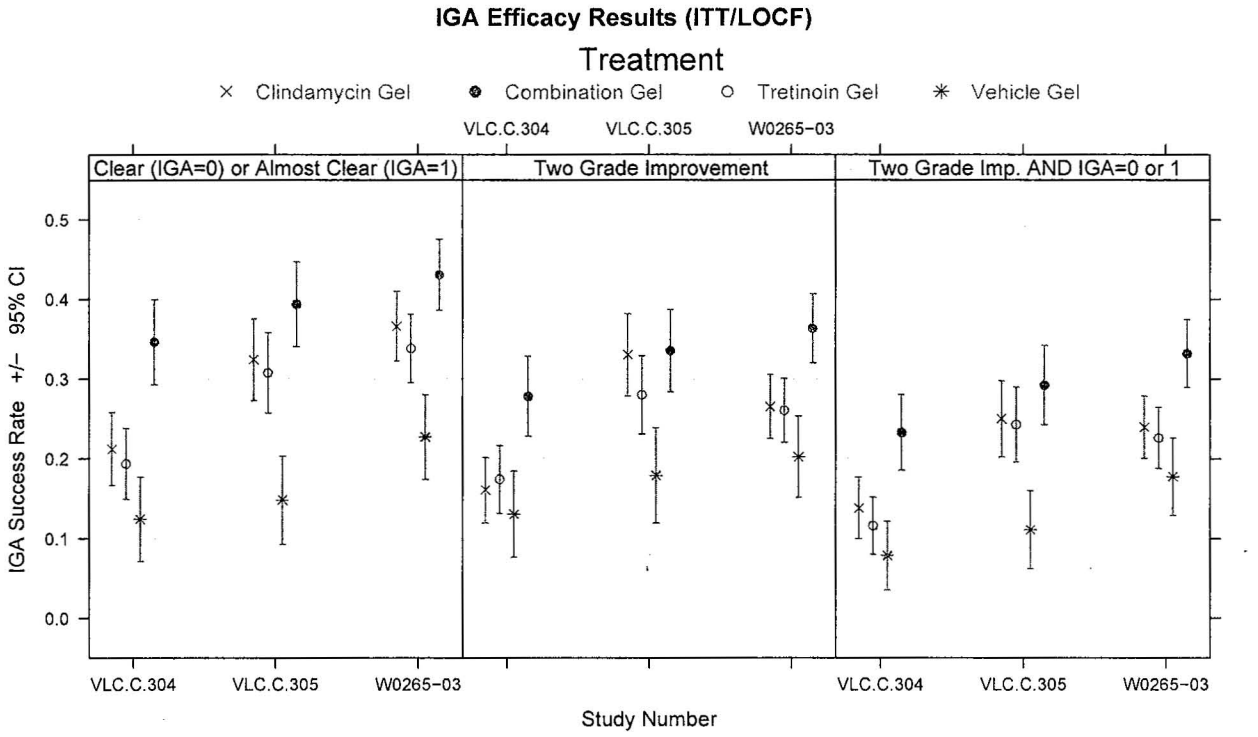
No assessment to date for lesion counts. Preliminary analysis reveals similar results when imputing missing IGA scores as failures as that presented above.

4 Additional Statistical Issues

Discussions held with medical officer regarding statistical issues of concern to him/her (such as secondary endpoints, important safety variables, etc.)

The following summary depicts the IGA efficacy results for all three Phase 3 studies.

Figure 1: **Efficacy Result Summary for IGA - All Phase 3 Studies**



5 Labeling Claims

Investigation of labeling claims and their purported basis.

A thorough investigation of the labeling claims has not been addressed to date.

6 Additional Requests

Identification of further requests to be made of the sponsor.

No additional requests are needed at this time.

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

STATISTICAL REVIEW AND EVALUATION

CARCINOGENICITY STUDY

NDA Number:	50,803 / Serial 000
Drug Name:	Veltin (applied for), Clindamycin 1% Tretinoin 0.25% gel
Indication(s):	Topical treatment of acne vulgaris.
Applicant:	Stiefel, a GSK Company.
Date:	Submitted 10/15/2009 (To reviewer 12/10/2009)';
Review Priority:	Standard
Biometrics Division:	Division 6
Statistical Reviewer:	Steve Thomson
Concurring Reviewer:	Team Leader: Karl Lin, Ph. D.
Medical Division:	Dermatological and Dental Drug Products
Toxicologist:	Reviewer: Merrill, Jill, Ph.D. Team Leader: Barbara Hill, Ph.D.
Project Manager:	Christina Attinello, M.P.H.
Keywords:	Carcinogenicity, Cox regression, Kaplan-Meier product limit, Survival analysis

Table of Contents

1. EXECUTIVE SUMMARY 3

1.1. CONCLUSIONS AND RECOMMENDATIONS 3
1.2. BRIEF OVERVIEW OF THE STUDIES 4
1.3. STATISTICAL ISSUES AND FINDINGS 5
1.3.1. Statistical Issues 5
1.3.2. Statistical Findings 8

2. INTRODUCTION 8

2.1. OVERVIEW 8
2.2. DATA SOURCES 8

3. STATISTICAL EVALUATION 8

3.1. EVALUATION OF EFFICACY 8
3.2. EVALUATION OF SAFETY 8

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS 13

5. SUMMARY AND CONCLUSIONS 13

5.1. STATISTICAL ISSUES AND COLLECTIVE EVIDENCE 13
5.2. CONCLUSIONS AND RECOMMENDATIONS 13

APPENDICES: 14

APPENDIX 1. SURVIVAL ANALYSIS 14
APPENDIX 2. FDA TUMORIGENICITY ANALYSIS 16
APPENDIX 3. ALTERNATIVE FDA TUMORIGENICITY ANALYSIS 19
APPENDIX 4. REFERENCES 22

1. EXECUTIVE SUMMARY

According to the reports provided by the Sponsor, this submission was intended to assess the carcinogenic potential and toxicokinetics of clindamycin with daily application to the skin of CD-1 mice for up to 104 weeks. For each gender, the study had three treatment groups, a sham treatment group, a clindamycin tretinoin gel group (denoted CTG – without the clindamycin and tretinoin), and a treatment group with clindamycin 1% in the CTG gel. The sponsor was Stiefel Laboratories, Inc., in Palo Alto, California. The study was conducted (b) (4)

1.1. Conclusions and Recommendations

The submission summarizes the results of a dermal mouse study of the carcinogenic potential of clindamycin 1% in clindamycin tretinoin gel group (CTG). There were three experimental groups per mouse gender, with study design as described in Table 1 below:

Table 1. Study Dose Groups

Group No.	Test Material	Clindamycin Dose (mg/kg/day)	Dose Volume (µL)	Clindamycin Conc. (mg/mL)	Treat-ment Area	No. of Animals	
						Carcino-genicity	Toxico-kinetic
0	Sham Control	0	0	0	0	60 (60) [1]	
1	CTG vehicle	0	80	0	2x2 cm	60	42 (42)[1]
2	1% clindamycin in CTG vehicle	32	80	10	2x2 cm	60	42

[1] Duplicate sets of carcinogenicity sham control animals (Group 1) and toxicokinetic CTG vehicle animals were housed in a separate room as "contamination controls".

The statistical significances of the tests of differences in survival across treatment groups are given below in Table 2. The tests for homogeneity are tests that survival is equal across treatment groups, namely the logrank and the Wilcoxon test. Note that the Wilcoxon test tends to place more weight on earlier events than does the logrank test, and thus is more sensitive to earlier differences in survival.

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

Comparison	Males		Females	
	Logrank	Wilcoxon	Logrank	Wilcoxon
Overall homogeneity	0.0001	0.0003	0.0492	0.0689
Pairwise between sham & vehicle	0.5738	0.4985	0.6884	0.7939
Pairwise between veltin & vehicle	0.0028	0.0061	0.0568	0.0629

Survival curves for both genders are given in Figures A.1.1 and A.1.2 in Appendix 1. From these Kaplan-Meier curves, in both genders in mice, the clindamycin treatment group has the highest mortality, although in female mice the difference is less evident than among males. This is consistent with the p-values above (Males: logrank p = 0.0001, Wilcoxon p = 0.0003,

Females: logrank p = 0.0492, Wilcoxon p = 0.0689). Neither gender shows any strong evidence of differences between the sham treatment and the vehicle, though the two survival curves are more closely intertwined in female mice (but all four $p \geq 0.4985$). This is also consistent with the fact that in each gender the results comparing the treatment group to the vehicle are similar to the results for tests comparing overall homogeneity.

The FDA analysis of tumorgenicity is summarized in Appendix 2. This is based on a poly-k analysis, where tumor onset rate is modeled as following a third degree polynomial (Bailer and Portier, 1988). For each organ-tumor combination, the results of pairwise tests of differences in incidence between the clindamycin group and the vehicle control, the clindamycin group and the sham treatment, and, finally, the difference beween the vehicle control and the sham treatment are provided. Note that dosing was terminated early in the clindamycin groups (Males: Week 93, Females Week 98), accompanied by early euthanasia in these groups (Males: Week 98, Females Week 103). This could be expected to reduce the number of observed tumors in the clindamycin groups and may make it more difficult to detect treatment differences (please see Section 1.3.1.1 for more discussion).

When following the frequentist statistical paradigm, most biostatisticians would recommend an adjustment for the multiple number of comparisons. However, in the primary analysis of tumor incidence, no pairwise differences among the sham, CTG, and clindamycin groups showed statistically significant differences at the usual 0.05 level.

An alternative analysis presented in Appendix 3 does suggest some differences in male animals. But it does this by confounding tumor incidence with toxicity, and thus also suffers from some problem of interpretability.

1.2. Brief Overview of the Studies

A single mouse study was submitted:

Study NPB00012: A Two-Year Carcinogenicity Study of Clindamycin Administered Dermally (Skin Painting) to Mice

Presumably this study was designed to assess the potential carcinogenic effect of the mixture of clindamycin 1% and tretinoin 0.25% gel. To achieve this goal, the Sponsor proposes to assess the carcinogenicity of clindamycin 1% in the CTG gel. The study had three treatment groups for each gender, each with 60 animals, a sham treatment group, a clindamycin tretinoin gel group (CTG – without the clindamycin and tretinoin), and a treatment group with clindamycin 1% in the CTG gel. An additional similar sham control group in a separate room was included as a "contamination control."

1.3. Statistical Issues and Findings

1.3.1. Statistical Issues

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

1.3.1.1. Early Termination in a Study With a Single Treatment Groups:

Dosing in male animals was stopped 11-12 weeks earlier in the clindamycin group than in the other two control groups, with a terminal sacrifice 5-6 weeks later. This means that tumors in this dose group that would have developed at the end of the study will be missed, reducing the tumor count in the high dose group. This is a particular problem with the Peto analysis used by the Sponsor. Effectively the poly-k adjustment reduces the impact of the lower numerator tumor count implied by early sacrifice by reducing the effective count of the animals at risk of tumor development, thus reducing the denominator count. This reduction is presumed to be appropriate for the roughly third order polynomial onset rate (Bailer and Portier, 1988). In practice, in most studies early termination does not present a real problem to either the Peto or poly-k method of analysis, since there will usually be three or more increasing treatment groups. Then the results in the next highest dose group without the early termination can be interpreted to adjust for the early termination in the higher dose groups. But with only a single treatment group, no such interpretation is possible. Dosing in female animals was stopped 6-7 weeks early with final sacrifice 1-2 weeks early. Such a small difference is not likely to be of much consequence.

1.3.1.2. Survival Analysis:

Both the logrank and Wilcoxon tests were used to test homogeneity of survival among the treatment groups. The log rank tests tends to put higher weight on later events while the Wicoxon tends to weight events more equally, and thus is more sensitive to early differences 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 some detail. The Sponsor's analyses are summarized in Section 3.2.1.1. Appendix 2 presents an experimental Bayesian approach that allows nonproportional hazards.

1.3.1.3. Tests on Neoplasms:

In the past, the usual FDA tumorigenicity analyses was based on so-called Peto tests where the analysis of fatal tumors was based on the time of death, and for observable tumors on the time of detection. Non-fatal, incidental tumors found at the time of the animal's death were analyzed by the so-called prevalence method. Then these results were pooled for a final test. This is the method of analysis used in the Sponsor's report However, this method depends on the attribution of whether a tumor is fatal, incidental, or observable (mortality independent). In the report of the Society of Toxicological Pathology "town hall" meeting in June 2001 the poly-k modification of the Cochran-Armitage test of trend was generally recommended over use of

these Peto tests. The poly-k tests (here k=3) do not require accurate specification of whether a tumor is fatal, incidental, or mortality independent.

There are two particular problems with the Sponsor's Peto analysis. The Sponsor used fixed time intervals for the incidental tumor analysis, with an exact permutation test for analyses with low incidence. The problem is that many of the intervals have no tumors in any treatment group, so effectively those animals that died in any time interval with no tumors in any group are dropped from the analysis. This increases the p-value of any test of pairwise differences, reducing the chance of detecting a statistically significant difference. A second problem is associated with the early termination in the clindamycin groups as discussed in section 1.3.1.1 above. The Sponsor states that the "statistical analysis of onset rates employed in this report adjusts for differces in survival rates which might otherwise bias the analysis and interpretation of results." (page 51, Volume 4). This reviewer would argue that while the Peto analysis does adjust for differences in survival, it does not completely adjust for the lack of time to develop incidental tumors in the groups chosen for early sacrifice.

Note that with only one treatment group and no natural ordering among the three treatment groups, the only tests that are appropriate are the pairwise tests between the treatment group and vehicle control, plus the pairwise tests between the vehicle control and the sham treatment.

1.3.1.4. Multiplicity of Tests on Neoplasms:

Testing the various neoplasms involved a large number of statistical tests, which in turn usually necessitates an adjustment in experiment-wise Type I error. For a study with only one species, the current, somewhat ad hoc rule, is that for an overall rough 10% type I error, pairwise comparisons should be declared statistically significant when the nominal significance levels are 0.01 for common tumors and 0.05 for rare tumors. Usually the observed incidence in the untreated control groups is used to decide if a tumor is rare (incidence ≤ 1%) or common (incidence > 1%). Again, however, in the primary poly-k analysis no pairwise comparisons were statistically significant at a 0.05 level, so the correction for multiplicity is moot.

1.3.1.5. Validity of the Designs:

When determining the validity of designs there are two key points:
1) adequate drug exposure
2) tumor challenge to the tested animals.

1) is related to whether or not sufficient animals survived long enough to be at risk of forming late-developing tumors and 2) is related to the Maximum Tolerated Dose (MTD), designed to achieve the greatest likelihood of tumorigenicity.

Lin and Ali (1994), quoting work by Haseman, have suggested that a survival rate of about 25 animals, out of 50 or more animals, between weeks 80-90 of a two-year study may be considered a sufficient number of survivors as well as one measure of adequate exposure. This criterion is satisfied in female mice and almost in male mice. Detailed survival curves are given

in Appendix 1. These seem to be consistent with the Sponsor's suggestion that the MTD was achieved.

Chu, Ceuto, and Ward (1981), citing earlier work by Sontag et al. (1976) recommend that the MTD "is taken as 'the highest dose that causes no more than a 10% weight decrement as compared to the appropriate control groups, and does not produce mortality, clinical signs of toxicity, or pathologic lesions (other than those that may be related to a neoplastic response) that would be predicted to shorten the animal's natural life span' " The apparent decrement in survival at the end of the study may be evidence that the MTD was exceeded, particularly in male mice. The Sponsor does indicate that there were statistically significant differences in weight gain and food consumption, but it was not generally associated with a decrement in the clindamycin group. The values in the following Tables 3 and 4 are transcribed from the Sponsor's reports, and display snapshots of the mean final weight and final mean food consumption.

Table 3. Animal Weight Data at Clindamycin Final Sacrifice (grams)

	Males Week 97			Females Week 101		
	Sham	Vehicle	Clindamycin	Sham	Vehicle	Clindamycin
Mean	42.8	44.4	42.0	36.6	39.6	40.7
N	37	33	17	29	28	17
% diff vs. Sham		3.8%	-1.7%		8.6%	6.6%

Table 4. Food Consumption Data at Clindamycin Final Sacrifice (grams/animal/day)

	Males Week 97-98			Females Week 101-102		
	Sham	Vehicle	Clindamycin	Sham	Vehicle	Clindamycin
Mean	7.4	7.8	8.2	8.6	8.6	9.2
N	37	31	15	27	26	16
% diff vs. Sham		5.0%	10.3%		-0.9%	7.0%

Note these results do not seem to suggest that the MTD was exceeded and, may even suggest that the MTD was not achieved.

Again from 2) above, excess mortality not associated with any tumor or sacrifice in the higher dose groups might suggest that the MTD was exceeded. One way to assess this possibility is to measure natural mortality not associated with any identified tumor or planned sacrifice. Note this seems to be a novel way to assess if the high dose is at the MTD. Table 5 below indicates the number of animals in each dose group that died of a natural death or moribund sacrifice, but did not show any tumors:

Table 5: Natural Death or Accident with No Identified Tumor

	Males			Females		
	Sham	Vehicle	Clindamycin	Sham	Vehicle	Clindamycin
Natural death w/o tumor	15	18	23	11	10	18
Other	45	42	37	49	50	42

The test statistics corresponding to the above correspond to tests of differences in dose groups in terms of the time to early non-tumor related, unplanned death. In male rats there is a statistically significant difference in these statistics (Males: Log Rank p=0.0367, Wilcoxon p=0.0458, Females: Log Rank p=0.0471, Wilcoxon p=0.0306). This seems to be evidence that there is excess mortality associated with toxicity in the clindamycin dose group.

The pharm/tox reviewers should use their expertise and other information such as clinical signs or severe histopathologic toxic effects that are attributable to the dosed animals in their final evaluation of the appropriateness of the doses used.

1.3.2. Statistical Findings

Please see Section 1.1 above.

2. INTRODUCTION

2.1. Overview

2.2. Data Sources

The Sponsor sent two SAS transport files, one for survival data mice and the other for tumorigenicity. Neither data set fit the requested format for submitted analysis. The SAS tumorigenicity data set, analdata.sas7bdat, included a variable mpeto taking values 0, 1, or 2. In this data set there are a large number observations for each animal, one for for each tumor. The mpeto value seems to indicate whether the specified tumor is absent, incidental, or fatal. With this assumption, and pulling values from the survdata.sas7bdat S a data set fitting the requested format was created.

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

NA

3.2. Evaluation of Safety

More detailed results on the study are presented below.

3.2.1. Study Study NPB00012: A Two-Year Carcinogenicity Study of Clindamycin Administered Dermally (Skin Painting) to Mice

STUDY DURATION: Males: Clindamycin group through Week 97
Females: Clindamycin Group through Week 102.
In both genders: Sham and Vehicle through Week 104.
DOSING STARTING DATE: November 14, 2006 (Study day 1).
DOSING TERMINATION DATE: Males: (Study day 645-Week 93)
Females: (Study day 683-Week 98)
EUTHANIZATION DATE: Group 3 Males: Week 98, Females: Week 103
Other groups: Week 105
TERMINATION OF IN-LIFE PHASE: November 13, 2008.
STUDY ENDING DATE (Final Report dated): July 20, 2009.
MOUSE STRAIN: Charles River CD-1 mice
ROUTE: Daily Skin Painting.

There was only a single experimental group in each gender, but with a daily sham dosing group and a vehicle group, with study design as described in Table 6 below:

Table 6. Study Dose Groups

Group No.	Test Material	Clindamycin Dose (mg/kg/day)	Dose Volume (µL)	Clindamycin Conc. (mg/mL)	Treat-ment Area	No. of Animals	
						Carcino-genicity	Toxico-kinetic
0	Sham Control	0	0	0	0	60 (60) [1]	
1	CTG vehicle	0	80	0	2x2 cm	60	42 (42)[1]
2	1% clindamycin in CTG vehicle	32	80	10	2x2 cm	60	42

[1] Duplicate sets of carcinogenicity sham control animals (Group 1) and toxicokinetic CTG vehicle animals were housed in a separate room as "contamination controls".

Animals were approximately eight weeks old at first dosing. According to the Sponsor, animals were "randomized to groups by a stratified randomization scheme designed to achieve similar group mean body weights." (page 27 of report) Animals were housed singly during the study. Food and water were available ad libitum.

The Sponsor indicates that the dorsal skin of each mouse was clipped, and reclipped as necessary. Prior to dosing each area was wiped with gauze soaked in de-ionized water. The body surface area to which the dose material was applied was approximately 2x2 cm, roughly 10% of the body surface area. The treatment area in the sham group was clipped and rubbd with an applicator or gloved finger.

According to the Sponsor, dose selection was based upon "A 90-day dermal tolerability study in CD-1 mice . . . was performed to assess the feasibility of using 1% clindamycin in CTG vehicle in a 2-year mouse dermal study. The maximum feasibility dose of 1% clindamycin in CTG vehicle was determined to be 80µL/day. Two doses of 1% clindamycin in CTG vehicle

(40 µL and 80 µL) were applied to the dorsal skin of CD-1 mice for up to 90 days. These dose volumes were equivalent to approximately 20 mg/kg/day and 40 mg/kg/day (based on a 20 gram mouse), respectively, of clindamycin phosphate. There were no compound-related mortalities or indications of systemic toxicity noted during 90 days of dosing in this study. Some mice administered 80 µL/day of 1% clindamycin in CTG vehicle had clinical signs of moderate to marked dermal erythema during the study. No other adverse signs were observed clinically or at gross necropsy in mice administered 40 or 80 µL/day. There were no treatment-related effects on weight gain or food consumption. Histologically, mice administered 40 or 80 µL/day of 1% clindamycin in CTG vehicle had minimal to moderate signs of dermal irritation characterized by epidermal hyperplasia, hyperkeratosis, and superficial dermal inflammation. The maximum tolerated dose in CD-1 mice was, therefore, determined to be 80 µL/day of 1% clindamycin in CTG vehicle (approximately 40 mg/kg clindamycin)." (page 28 of report)

The Sponsor states that detailed physical examinations were made on all animals each week. Mortality and morbidity checks were made twice daily. Both body weights and food consumption were recorded weekly for the first 13 weeks, and every 4 weeks thereafter. The test and control articles were administered dermally seven days per week.

3.2.1.1 Sponsor's Results and Conclusions

This section will present a summary of the Sponsor's analysis on survivability and tumorigencity.

Survival analysis:

A number of males and females from each study group were found dead or euthanized prior to study termination (i.e., unscheduled moribund euthanasia). The incidence of death and unscheduled euthanasia is presented in the following table:

Table 7. Sponsor's Summary of Survival:

		Deaths/At risk		
Sex	Interval Week	Control	CTG Vehicle	Veltrin
Male	1-50	3/60 (95)	3/60 (95)	7/60 (88)
	51-80	9/57 (80)	14/57 (72)	14/53 (65)
	81-EOS	20/48 (47)	16/43 (44)	23/38 (26)
	Term. Sac.	28	26	15
	p-value (C)		0.6624	<0.0001
	p-value (CTG)			0.0027
Female	1-50	0/60 (100)	1/60 (98)	4/60 (93)
	51-80	15/60 (75)	13/59 (77)	16/56 (67)
	81-EOS	21/45 (40)	25/46 (35)	25/40 (25)
	Term. Sac.	24	21	15
	p-value (C)		0.6884	0.0275
	p-value (CTG)			0.0568

(transcribed from page 74 of volume 10 of report)

Tumorigenicity analysis:

The Sponsor analyzed a number of tumors. Since the tumor incidence seems to follow the counts in Appendix 2, they are not reproduced in this report. Note the sponsor used a Peto analysis whereas the FDA analysis is based on a poly-k analysis. Table 8 below displays those organ-tumor combinations that have at least one potentially statistically significant comparison ($p \leq 0.10$ or close to 0.10). These are reproduced from the Tables 7.1 and 7.2 in the Statistical Analysis report (pages 87-94 of volume 10). The columns labeled "C vs V", "C vs S", and "V vs S" provide the significance levels of the pairwise Peto tests between the cindamycin and vehicle, clindamycin and sham, and vehicle and sham, respectively. The only tumors associated with a significance level of approximately 0.10 or less are reproduced here:

Table 8. Sponsor Derived Potentially Statistically Significant Neoplasms

		Incidence			p-values		
Organ	Tumor	Sham	Veh	Clin	C vs V	C vs S	V vs S
Males							
Harderian Gland	Adenoma+Adenocarcinoma	6	13	3	0.9856	0.4484	0.0909
	Adenoma	5	11	2	0.9839	0.5699	0.1085
Liver	Hepato. Adenoma+Carcinoma+ Hemangioma+Hemangiosarcoma+Cholangioma	16	18	19	0.2346	0.0793	0.4263
Females							
Hemolymphoreticular	Lymphoma	12	14	18	0.0483	0.0189	0.6872
Pituitary	Adenoma+Carcinoma	0	3	3	0.6206	0.1097	0.2495
	Adenoma	0	2	3	0.4617	0.1097	0.4985

To assess the effect of clindamycin, this reviewer would recommend the comparison to the vehicle. The only such pairwise comparison that would be statistically significant at the usual 0.5 level would be, just barely, hemolymphoreticular lymphoma in female mice. Since the incidence in the sham treatment group is above 1%, this would be considered as a common tumor. Thus, using Haseman's original suggestion of an ad hoc adjustment for multiplicity, at an overall rough 10% error rate, this would not be statistically significant (i.e. $p=0.0483 > 0.01$) . Note that no other comparison met this 0.01 criteria.3.2.1.2 FDA Reviewer's Results

3.2.1.2. FDA Reviewer's Results

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

Survival analysis:

The following Table 9 summarizes 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 survived at the end of the interval.

Table 9. Summary of Survival

	Males				Females		
Period (Weeks)	Sham	Vehicle	Clinda-mycin	Period (Weeks	Sham	Vehicle	Clinda-mycin
0-50	3/60[1] 95.0%[2]	3/60 95.0%	7/60 88.3%	0-50	0/60 100%	1/60 98.3%	4/60 93.3%
51-80	9/57 80.0%	14/57 71.7%	16/53 61.7%	51-80	15/60 75.0%	13/59 76.7%	16/56 66.7%
81-91	11/48 61.7%	12/43 51.7%	22/37 25.0%	81-91	14/45 51.7%	15/46 51.7%	19/40 35.0%
92-104	9/37 46.7%	5/31 43.3%	0	92-102	4/31 45.0%	7/31 40.0%	6/21 25.0%
Terminal	28	26	15	102-104	3/27 40.0%	3/24 35.0%	0
				Terminal	24	21	15

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

Table 10 below presents the result of tests on survival over the dose groups. The tests for homogeneity are tests that survival is equal across treatment groups, namely the logrank and the Wilcoxon test. Note that the Wilcoxon test tends to place more weight on earlier events than does the logrank test, and thus is more sensitive to earlier differences in survival.

Table 10. Statistical Significances of Tests of Homogeneity and Trend in Survival

	Males		Females	
Comparison	Logrank	Wilcoxon	Logrank	Wilcoxon
Overall homogeneity	0.0001	0.0003	0.0492	0.0689
Pairwise between sham & vehicle	0.5738	0.4985	0.6884	0.7939
Pairwise between veltin & vehicle	0.0028	0.0061	0.0568	0.0629

Survival curves for both genders are given in Figures A.1.1 and A.1.2 in Appendix 1. From these Kaplan-Meier curves, in both genders in mice, the clindamycin treatment group has the highest mortality, although in female mice the difference is less evident than among males. This is consistent with the p-values above (Males: logrank $p = 0.0001$, Wilcoxon $p = 0.0003$, Females: logrank $p = 0.0492$, Wilcoxon $p = 0.0689$). Neither gender shows any strong evidence of differences between the sham treatment and the vehicle, though the two survival curves are more closely intertwined in female mice (but all four $p \geq 0.4985$). This is also consistent with the fact that in each gender the results comparing the treatment group to the vehicle are similar to the results for tests comparing overall homogeneity.

Tumorigenicity analysis:

The FDA analysis of tumorgenicity is summarized in Appendix 2. For each organ tumor combination, the results of pairwise tests of differences in incidence between the

clindamycin group and the vehicle control, the clindamycin group and the sham treatment, and, finally, the difference beween the vehicle control and the sham treatment are provided. No pairwise test in either gender met the usual 0.05 level of significance. However, as noted several times above, dosing was terminated early in the clindamycin groups, followed by early sacrifice. This could be expected to reduce the number of observed tumors in the clindamycin groups and may make it more difficult to detect treatment differences (please see Section 1.3.1.3 for more discussion). An alternative secondary analysis that may make it too easy to detect treatment differences is presented in Appendix 3.

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 Analysis

The statistical significance of the tests of differences in survival across treatment groups are given below. The tests of equality in survival across treatmentgroups provided below are the logrank and the Wilcoxon test. Note that the Wilcoxon test tends to place more weight on earlier events than does the logrank test. When survival curves tend to be parallel, so that the proportional hazards assumption is tenable, the log rank test will be the most powerful.

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

	Males		Females	
Comparison	Logrank	Wilcoxon	Logrank	Wilcoxon
Overall homogeneity over all 3 groups	0.0001	0.0003	0.0492	0.0689
Pairwise between sham & vehicle	0.5738	0.4985	0.6884	0.7939
Pairwise between veltin & vehicle	0.0028	0.0061	0.0568	0.0629

Survival curves for both genders are given in Figures A.1.1 and A.1.2, below. Note that in both genders, the clindamycin treatment group has the highest mortality, although in female mice the difference is clearly less than among females. This is consistent with the p-values above (Males: logrank $p = 0.0001$, Wilcoxon $p = 0.0003$, Females: logrank $p = 0.0492$, Wilcoxon $p = 0.0689$). Neither gender shows any strong evidence of differences between the sham treatment and the vehicle, though the two survival curves are more closely intertwined in female mice (but all four $p \geq 0.4985$). This is also consistent with the fact that in each gender the results comparing the treatment group to the vehicle are similar to the results for tests comparing overall homogeneity (i.e., Males: logrank $p = 0.0028$, Wilcoxon $p = 0.0061$, Females: logrank $p = 0.0568$, Wilcoxon $p = 0.0629$).

The figures below display these Kaplan-Meier estimated survival curves for both genders in each rodent species. The colored bands signify 95% confidence band about each survival curve.

Figure A.1.1 Male Mice

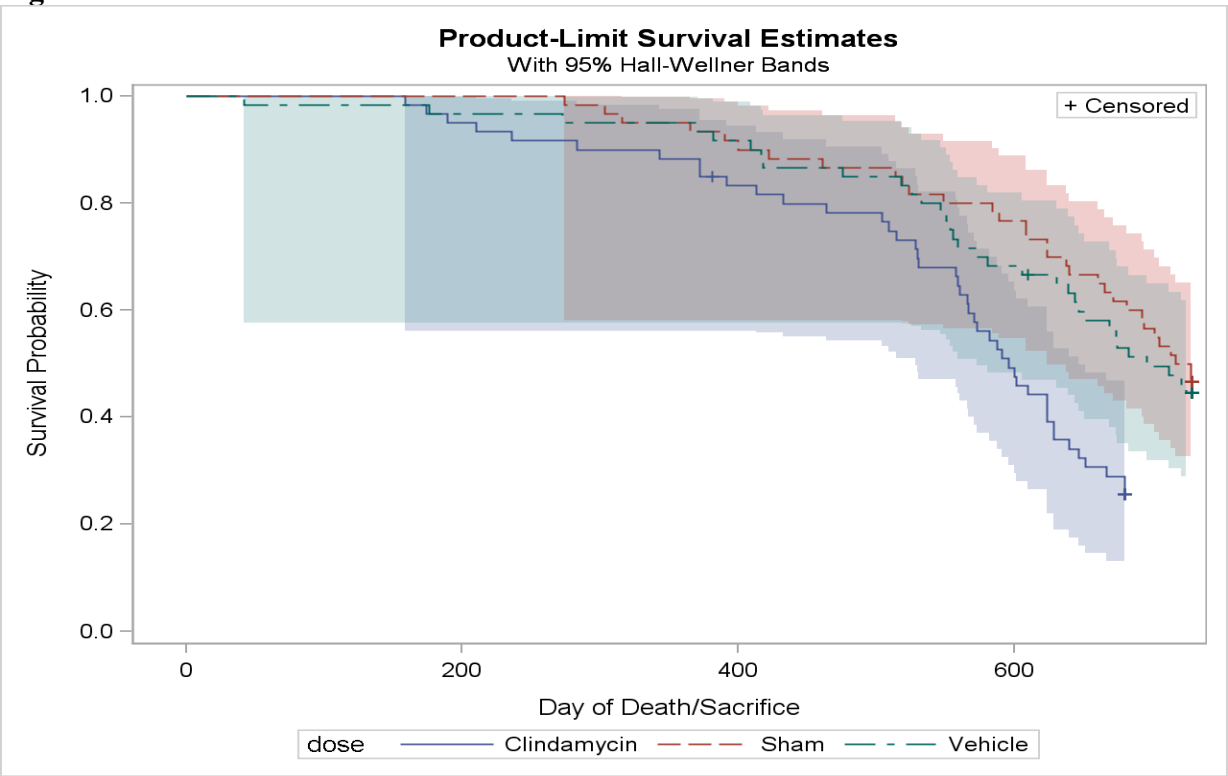
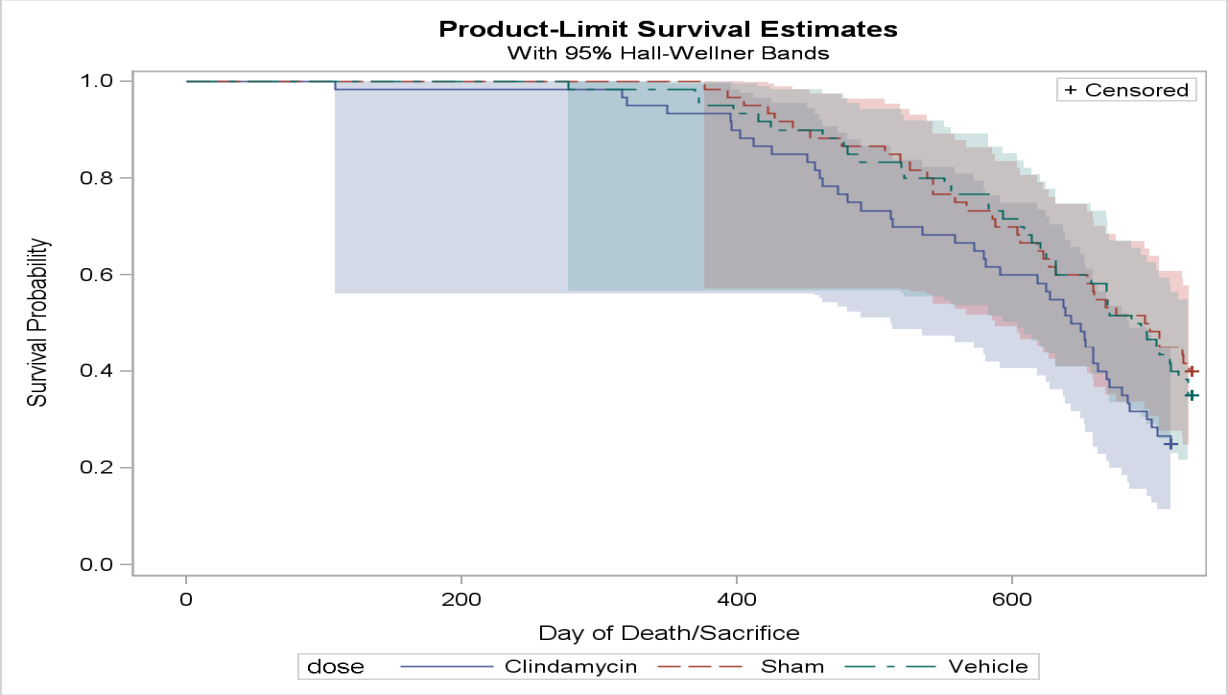


Figure A.1.2 Female Mice



Appendix 2. FDA Tumorigenicity Analysis

Table A.2.1 below displays the tumor incidence and the p-values using the poly-k adjustment to the Cochran-Armitage test of trend in dose. In the report of the Society of Toxicological Pathology "town hall" meeting in June 2001 the poly-k modification of the Cochran-Armitage test of trend was generally recommended over use of the Peto tests used by the Sponsor.

The incidence of each organ tumor combination in the sham treatment group, the vehicle group and the clindamycin group are listed, along with the p-values of the poly-k test that there is no pairwise difference between the clindamycin group and the vehicle control, the clindamycin group and the sham treatment, and, finally, the difference beween the vehicle control and the sham treatment. Following the Sponsor, the tumor combinations below are included in the following:

COMBINATION 1	Liver	Hepatocellular Adenoma	Hemangioma
		Hepatocellular Carcinoma	Hemangiosarcoma
		Cholagioma	
COMBINATION 7	Ovary	Cystadenoma	Luteoma
		Granulosa cell tumor	Mesenchymal tumor
		Hemangioma	
COMBINATION 9	Uterus	Adenocarcinoma	Hemangioma
		Endometrial stromal polyp	Leiomyoma
		Endometrial stromal sarcoma	

The other combinations defined by the Sponsor are specified by contractions in the tumor name. The endpoint actually used in the analysis is the time of detection when that is less than the time of death. Of course, most neoplasms were incidental and hence the time of detection is the same as the time of death.

As discussed in Section 1.3.1.1 above, actual dosing was stopped early in the clindamycin groups. Further, the conduct of the study provided more time for tumors to develop in the vehicle and sham treatment groups than in the clindamycin group. The extra time in females is short and is thus probably ignorable, but was a month and a half in male mice. Strictly speaking this means that the two control groups were treated differently from the clindamycin group, so that treatment differences are confounded with study conduct differences. Again, Bailer and Portier (1988) suggest that tumor onset rate is modeled as following a third degree polynomial, so the poly-k analysis adjusts for these conduct differences.

If one assumes that the groups are comparable, it is clear from Table A.2.1 that no pairwise comparisons between the three treatment groups were statistically significant at a nominal 0.05 level, let alone after any adjustment for the multiple number of tests.

Table A.2.1: Incidence and the Results of Pairwise Comparisons Between Groups

organ/ tumor	Incidence Sham	Incidence Veh	Incidence Clin	p-values Clin vs Veh	p-values Clin vs Sham	p-values Veh vs Sham
Male Mice						
HARDERIAN GLAND						
ADENOCARCINOMA	1	2	1	0.3696	0.6463	0.4831
ADENOMA	5	11	2	0.9703	0.6089	0.0693
Adenoma/Adenocarcinoma	6	13	3	0.9627	0.5478	0.0604
HEMOLYMPHORETICULAR						
HISTIOCYTIC SARCOMA	2	0	1	0.4324	0.3575	0.7240
LYMPHOMA	7	2	4	0.2120	0.4890	0.8988
LIVER						
COMBINATION 1	16	18	19	0.3167	0.1286	0.3257
HEMANGIOSARCOMA	3	6	2	0.7511	0.3218	0.1997
HEPATOCELLULAR ADENOMA	6	6	8	0.2496	0.1826	0.5384
HEPATOCELLULAR CARCINOMA	10	8	9	0.2957	0.4158	0.5418
LUNG						
BRONCHIOLO-ALVEOLAR ADENOCARCINOMA	7	4	4	0.4823	0.4821	0.6698
BRONCHIOLO-ALVEOLAR ADENOMA	10	10	8	0.5717	0.5292	0.5555
Bron.-Alv.adn.&carc./Carcinsarc	16	13	11	0.5246	0.5030	0.5924
SPLEEN						
HEMANGIOSARCOMA	3	0	2	0.1836	0.3218	0.8572
Hemangioma/Hemangiosarcoma	3	1	2	0.3979	0.3218	0.6480
Systemic						
HEMANGIOSARCOMA	6	6	4	0.4527	0.3798	0.5401
Hemangioma/Hemangiosarcoma	6	6	4	0.4527	0.3798	0.5401
TESTIS						
INTERSTITIAL CELL ADENOMA	3	4	2	0.5251	0.3218	0.4366
Interstit.+Sertoli Cell Adenoma	4	4	2	0.5251	0.4665	0.5779
Female Mice						
ADRENAL CORTEX						
ADENOMA	2	3	0	0.8548	0.7142	0.4887
CERVIX						
Endo. stromal polyp/sarcoma	1	2	0	0.7205	0.4625	0.4911
HARDERIAN GLAND						
ADENOCARCINOMA	2	1	0	0.4684	0.7080	0.4824
ADENOMA	5	3	2	0.4281	0.7059	0.6303
Adenoma/Adenocarcinoma	7	4	2	0.5877	0.8666	0.7123
HEMOLYMPHORETICULAR						
HISTIOCYTIC SARCOMA	7	5	2	0.7059	0.8509	0.5729
LYMPHOMA	12	14	18	0.1249	0.0643	0.4342
LIVER						
COMBINATION 1	2	2	4	0.2921	0.2819	0.6832
HEMANGIOSARCOMA	2	2	3	0.4408	0.4299	0.6832
LUNG						
BRONCHIOLO-ALVEOLAR ADENOCARCINOMA	2	3	3	0.5877	0.4281	0.5000
BRONCHIOLO-ALVEOLAR ADENOMA	11	9	8	0.6033	0.5872	0.6001
Bron.-Alv.adn.&carc./Carcinsarc	12	13	10	0.5590	0.4668	0.5000

Table A.2.1 (cont.): Incidence and the Results of Pairwise Comparisons Between Groups

organ/ tumor	Incidence Sham	Incidence Veh	Incidence Clin	p-values Clin vs Veh	p-values Clin vs Sham	p-values Veh vs Sham
Female Mice (cont.)						
MAMMARY GLAND						
ADENOACANTHOMA	2	0	1	0.4750	0.4445	0.7412
ADENOCARCINOMA	0	2	1	0.4345	0.4625	0.2529
Adenocanthoma/Adenocarcinoma	2	2	2	0.6346	0.6346	0.6919
OVARY						
COMBINATION 7	4	6	3	0.6641	0.4152	0.3842
CYSTADENOMA	0	3	1	0.6333	0.4625	0.1206
GRANULOSA CELL TUMOR	2	1	0	0.4684	0.7142	0.4911
HEMANGIOMA	0	1	2	0.4431	0.2108	0.5000
PITUITARY GLAND						
ADENOMA	0	2	3	0.4299	0.0989	0.2529
Adenoma/carcinoma	0	3	3	0.5894	0.0989	0.1249
Systemic						
HEMANGIOMA	0	1	2	0.4431	0.2108	0.5000
HEMANGIOSARCOMA	2	2	3	0.4408	0.4299	0.6832
Hemangioma/Hemangiosarcoma	2	3	5	0.2885	0.1602	0.4890
UTERUS						
COMBINATION 9	14	11	9	0.4748	0.6704	0.6070
ENDOMETRIAL STROMAL POLYP	9	8	7	0.3944	0.4793	0.4787
ENDOMETRIAL STROMAL SARCOMA	3	4	1	0.7700	0.6239	0.4870
LEIOMYOMA	1	0	2	0.2162	0.4431	0.4941
VAGINA						
ENDOMETRIAL STROMAL POLYP	2	0	0	.	0.7142	0.7471
Endo.str.polyp/Hemangiosarcoma	2	1	0	0.4684	0.7142	0.4911

Appendix 3. Alternative FDA Tumorigenicity Analysis

One problem with this study is that the two control groups have more opportunity to develop tumors, i.e. the conduct of the study is different for the two controls than with the single clindamycin group. Bailer and Portier (1988) suggest that the poly-k analysis in Appendix 2 appropriately adjusts for these conduct differences. An alternative approach would be to define the stopping point of the study so that the three treatment groups are treated the same, i.e. there would be no study conduct differences. One way to achieve this would be to base results only on those tumors detected prior to the time of the terminal sacrifice in the clindamycin group, i.e. Week 98 in male mice and Week 103 in female mice. That is, all tumors at the terminal sacrifice in the clindamycin group, and detected at or after the time of this terminal sacrifice in the other two groups are censored and ignored. Note this makes the groups comparable in terms of time to tumor, and thus has no conduct differences between the groups. However, most tumors are incidental, and only those animals whose tumors are detected prior to this sacrifice time are counted. This gives greater weight to those groups with higher mortality and thus confounds tumorigenicity with toxicity. However, it may be a useful secondary analysis.

The tables below display both the original tumor incidence and the censored tumor incidence, plus the results of the pairwise poly-k comparisons applied to the censored tumor data. As in Appendix 2, these are statistical tests that there are no pairwise differences between the clindamycin group and the vehicle control, the clindamycin group and the sham treatment, and, finally, the differences beween the vehicle control and the sham treatment.

Testing the various neoplasms involves a large number of statistical tests, which, at least following the frequentist statistical paradigm, requires an adjustment in experiment-wise Type I error. Current FDA practice is based on the Haseman-Lin-Rahman rules. In particular, to control the overall Type I error rate to roughly 10% for a single species, standard two-year study, pairwise tests between the high dose group and the control should be tested at a 0.01 level for common tumors and 0.05 for rare tumors. In this analysis, those tumord with incidence in the sham treatment group < 1% are considered to be rare, otherwise the tumor is classified as common.

In this tumor censored data the short early termination in female mice has little effect on conclusions, and, as with the uncensored data, no comparisons even achieve the nominal, multiplicity uncorrected 0.05 level. In male mice the much greater difference between the sacrifice times does have an effect. Note that the most relevant measure of the effect of clindamycin is the comparison to CTG vehicle. Adjusting for whether or not the tumor is classified as common, in male mice the only statistically significant comparison between the clindamycin group and the CTG vehicle is the pairwise comparison of COMBINATION 1 (i.e. pooled Hepatocellular Adenoma, Hepatocellular Carcinoma, Cholagioma, Hemangioma, and Hemangiosarcoma) in the liver, since ($p = 0.0071 < 0.01$). Although there are other such comparisons that achieve the 0.05 level, they do not exceed the multiplicity adjusted 0.01 limit. Similarly for COMBINATION 1 in the Liver, the comparison between the clindamycin group

and the sham treatment is statistically significant (p = 0.0003 < 0.01), as are the comparisons of hepatocellular adenoma and hepatocellular carcinoma (p = 0.0066 and 0.01 ≤ 0.01, respectively). No other comparisons achieved the multiplicity adjusted bounds,

Table A.3.1: Potentially Statistically Significant Results (any p ≤ 0.05)

organ/ tumor	Original Sham	Original Veh	Original Clin	Censored Sham	Censored Veh	Censored Clin	Clin vs Veh	Clin vs Sham	Veh vs Sham
Male Mice									
HARDERIAN GLAND									
Adenoma/Adenocarcinoma	6	13	3	0	6	0	0.9703	.	0.0133
LIVER									
COMBINATION 1	16	18	19	3	6	16	0.0071	0.0003	0.2232
HEPATOCELLULAR ADENOMA	6	6	8	0	1	6	0.0363	0.0066	0.4842
HEPATOCELLULAR CARCINOMA	10	8	9	2	4	9	0.0638	0.0100	0.3189
LUNG									
BRONCHIOLO-ALVEOLAR ADENOMA	10	10	8	0	2	5	0.1631	0.0159	0.2318
Bron.-Alv.adn.&carc./Carcinsarc	16	13	11	1	4	7	0.1853	0.0140	0.1617

Table A.3.2: Incidence and the Results of Pairwise Comparisons Between Groups Using Censored Tumor Data

organ/ tumor	Original Sham	Original Veh	Original Clin	Censored Sham	Censored Veh	Censored Clin	Clin vs Veh	Clin vs Sham	Veh vs Sham
Male Mice									
HARDERIAN GLAND									
ADENOCARCINOMA	1	2	1	0	2	0	0.6899	.	0.2371
ADENOMA	5	11	2	0	4	0	0.9039	.	0.0562
Adenoma/Adenocarcinoma	6	13	3	0	6	0	0.9703	.	0.0133
HEMOLYMPHORETICULAR									
HISTIOCYTIC SARCOMA	2	0	1	1	0	1	0.4524	0.6800	0.4792
LYMPHOMA	7	2	4	4	2	2	0.5989	0.5131	0.6317
LIVER									
COMBINATION 1	16	18	19	3	6	16	0.0071	0.0003	0.2232
HEMANGIOSARCOMA	3	6	2	1	2	1	0.4199	0.6800	0.4766
HEPATOCELLULAR ADENOMA	6	6	8	0	1	6	0.0363	0.0066	0.4842
HEPATOCELLULAR CARCINOMA	10	8	9	2	4	9	0.0638	0.0100	0.3189
LUNG									
BRONCHIOLO-ALVEOLAR ADENOCARCINOMA	7	4	4	1	2	2	0.6165	0.4046	0.4761
BRONCHIOLO-ALVEOLAR ADENOMA	10	10	8	0	2	5	0.1631	0.0159	0.2318
Bron.-Alv.adn.&carc./Carcinsarc	16	13	11	1	4	7	0.1853	0.0140	0.1617
SPLEEN									
HEMANGIOSARCOMA	3	0	2	1	0	2	0.2017	0.4046	0.4842
Hemangioma/Hemangiosarcoma	3	1	2	1	0	2	0.2017	0.4046	0.4842
Systemic									
HEMANGIOSARCOMA	6	6	4	2	2	3	0.3991	0.3708	0.6677
Hemangioma/Hemangiosarcoma	6	6	4	2	2	3	0.3991	0.3708	0.6677
TESTIS									
INTERSTITIAL CELL ADENOMA	3	4	2	1	1	2	0.4279	0.4046	0.7366
Interstit.+Sertoli Cell Adenoma	4	4	2	1	1	2	0.4279	0.4046	0.7366
Female Mice									
CERVIX									
Endo. stromal polyp/sarcoma	1	2	0	0	1	0	0.4578	.	0.5056
HARDERIAN GLAND									
ADENOCARCINOMA	2	1	0	1	0	0	.	0.4578	0.5000
ADENOMA	5	3	2	2	2	2	0.6360	0.6360	0.6918

Table A.3.2 (cont.): Incidence and the Results of Pairwise Comparisons Between Groups Using Censored Tumor Data

organ/ tumor	Original Sham	Original Veh	Original Clin	Censored Sham	Censored Veh	Censored Clin	Clin vs Veh	Clin vs Sham	Veh vs Sham
Female Mice									
CERVIX									
Endo. stromal polyp/sarcoma	1	2	0	0	1	0	0.4578	.	0.5056
HARDERIAN GLAND									
ADENOCARCINOMA	2	1	0	1	0	0	.	0.4578	0.5000
ADENOMA	5	3	2	2	2	2	0.6360	0.6360	0.6918
Adenoma/Adenocarcinoma	7	4	2	3	2	2	0.6360	0.4213	0.4895
HEMOLYMPHORETICULAR									
HISTIOCYTIC SARCOMA	7	5	2	6	5	2	0.7094	0.7888	0.4695
LYMPHOMA	12	14	18	8	8	12	0.1247	0.1472	0.4283
LIVER									
COMBINATION 1	2	2	4	1	1	3	0.2647	0.2647	0.7528
HEMANGIOSARCOMA	2	2	3	1	1	2	0.4459	0.4459	0.7528
LUNG									
BRONCHIOLO-ALVEOLAR ADENOCARCINOMA	2	3	3	1	3	2	0.4316	0.4543	0.3167
BRONCHIOLO-ALVEOLAR ADENOMA	11	9	8	3	3	6	0.1770	0.1862	0.3490
Bron.-Alv.adn.&carc./Carcinsarc	12	13	10	3	7	7	0.5008	0.1131	0.1667
MAMMARY GLAND									
ADENOACANTHOMA	2	0	1	2	0	1	0.4643	0.4459	0.7528
ADENOCARCINOMA	0	2	1	0	2	1	0.4376	0.4699	0.2584
Adenocanthoma/Adenocarcinoma	2	2	2	2	2	2	0.6372	0.6461	0.3166
OVARY									
COMBINATION 7	4	6	3	1	3	3	0.5799	0.2636	0.3253
CYSTADENOMA	0	3	1	0	1	1	0.7160	0.4699	0.5056
HEMANGIOMA	0	1	2	0	1	2	0.4459	0.2177	0.5056
PITUITARY GLAND									
ADENOMA	0	2	3	0	1	2	0.4459	0.2177	0.5056
Adenoma/carcinoma	0	3	3	0	2	2	0.6360	0.2177	0.2528
Systemic									
HEMANGIOMA	0	1	2	0	1	2	0.4459	0.2177	0.5056
HEMANGIOSARCOMA	2	2	3	1	1	2	0.4459	0.4459	0.7528
Hemangioma/Hemangiosarcoma	2	3	5	1	2	4	0.2834	0.1454	0.5000
UTERUS									
COMBINATION 9	14	11	9	7	4	4	0.5750	0.6523	0.7260
ENDOMETRIAL STROMAL POLYP	9	8	7	4	3	2	0.4316	0.5911	0.5000
ENDOMETRIAL STROMAL SARCOMA	3	4	1	3	1	1	0.7160	0.6360	0.6918
LEIOMYOMA	1	0	2	0	0	1	0.4643	0.4699	.

This analysis assumes that all animals are at risk. However, those animals with tumor detected after the sacrifice time of the high dose group are treated as having no tumor. One could argue that such an animal should not be in the risk set for that tumor. This would require a different risk set for each tumor and was not done. Again, when we assume all animals are at risk and censor at the Clindamycin dose censor time, we will tend to find more tumors in the Clindamycin group just due to the early deaths. So interpretation may be problematical.

Appendix 4. 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.

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

Ibrahim, J.G., Chen, M.-H., and Sinha, D. (2001), *Bayesian Survival Analysis*, Springer Verlag, New York.

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 Biopharma-ceutical Statistics*. 8(1), 1-15.

McConnell, E.E., Solleveld, H.A., Swnberg, 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.

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.

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-50803	ORIG-1	STIEFEL A GSK CO	Veltin

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
03/08/2010

KARL K LIN
03/09/2010
Concur with review

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

NEW DRUG APPLICATION

CLINICAL STUDIES

NDA/Serial Number:	50-803 / N000
Drug Name:	TRADENAME® Gel (Clindamycin 1% and Tretinoin 0.025%)
Indication(s):	Acne Vulgaris
Applicant:	Connetics Corporation
Dates:	Submitted: 08/26/2004 PDUFA: 06/25/2005
Review Priority:	Standard review
Biometrics Division:	Division of Biometrics III (HFD-725)
Statistics Reviewer:	Mat Soukup, Ph.D.
Concurring Reviewer:	Mohamed Al Osh, Ph.D.
Medical Division:	Division of Dermatologic and Dental Drug Products (HFD-540)
Clinical Team:	Bindi Nikhar, M.D.
Project Manager:	Margo Owens
Keywords:	Combination Drug, Interim Analysis

Table of Contents

STATISTICAL REVIEW AND EVALUATION 1

1 EXECUTIVE SUMMARY 4

1.1 Conclusions and Recommendations 4

1.2 Brief Overview of Clinical Studies 4

1.3 Statistical Issues and Findings 4

2 INTRODUCTION 6

2.1 Overview 6

2.2 Data Sources 6

3 STATISTICAL EVALUATION 7

3.1 Evaluation of Efficacy 7
3.1.1 Study Design and Endpoints 7
3.1.2 Patient Disposition, Demographic and Baseline Characteristics 8
3.1.2.1 Patient Disposition 9
3.1.2.2 Demographic Characteristics 10
3.1.2.3 Baseline Prognostic Characteristics 10
3.1.3 Statistical Methodology 11
3.1.4 Primary Endpoint Results 12
3.1.4.1 Percent reduction in Lesion Counts (ITT Population) 13
3.1.4.2 Success on Investigator Global Assessment (ITT Population) 14
3.1.5 Primary Endpoint Results (PP Population) 16
3.1.6 Secondary Endpoint Results 18
3.1.6.1 Absolute Change in Lesion Counts 18
3.1.6.2 Subject Global Assessment 19
3.1.6.3 Time to 50% Reduction in Total Lesion Counts 19
3.1.7 Interim Analysis Examination 22

3.2 Evaluation of Safety 27
3.2.1 Local Skin Reactions 27
3.2.2 Adverse Events 27

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS 29

4.1 Gender, Race, and Age 29

4.2 Other Special/Subgroup Populations 31

5 SUMMARY AND CONCLUSIONS 31

5.1 Statistical Issues and Collective Evidence 31

5.2 Conclusions and Recommendations 33

APPENDIX 34

A.1 Baseline Demographic Characteristics 34

A.2 Full ANOVA Results for Lesion Counts (ITT Population) 35

A.3 Efficacy Figures 37

A.4 Efficacy Results by Subgroup 39

A.5 References 42

SIGNATURES/DISTRIBUTION LIST 42

1 Executive Summary

1.1 Conclusions and Recommendations

TRADENAME® Gel (denoted in the review as TN® Gel as the proposed name Velac® Gel was not accepted by DMETS) is a new combination drug product consisting of the moieties Clindamycin 1% and Treninoin .025%. Connetics Corporation is seeking approval of this new combination drug product in the treatment of moderate to severe acne vulgaris. In two Phase 3 safety and efficacy studies, TN® Gel was found to be superior to each of its active components and vehicle. Determination of superiority of TN® Gel to its components and vehicle was assessed by winning on two of three lesion counts and on investigator global assessment (IGA). The local safety profile of TN® Gel is similar to that of Tretinoin with increasing levels of local skin reactions to drug product in the first two weeks of treatment followed by a reduction in skin irritation after two weeks of using the drug product. No serious adverse events or deaths attributable to drug product were reported by study investigators.

1.2 Brief Overview of Clinical Studies

Two Phase 3 safety and efficacy trials, Study VLC.C.304 and Study VLC.C.305, were conducted to compare TN® Gel to Clindamycin, Tretinoin, and Vehicle. Treatment was to be applied once daily in the evenings with a treatment duration of 12 weeks. The studies were conducted in 18 and 19 United States centers for Studies VLC.C.304 and VLC.C.305 recruiting a total of 1083 and 1136 subjects, respectively. Subjects entering the trials were required to have a minimum IGA score of 2, between 17 and 40 inflammatory lesions and between 20 and 150 non-inflammatory lesions. The objectives of the trial were to show superiority of TN® Gel to each of its components and its vehicle in 2 out of 3 lesion counts measured as the percent reduction from baseline and the percent of subjects that had IGA scores of 0 (clear) or 1 (almost clear) at the end of treatment.

1.3 Statistical Issues and Findings

Both Study VLC.C.304 and Study VLC.C.305 established superiority of TN® Gel to each of its components and its vehicle. The results of each study are reported in Table 1. Referring to the table, Study VLC.C.304 firmly establishes the superiority of TN® Gel on both the lesion counts and IGA as seen by the small p-values. Study VLC.C.305 was not as convincing as Study VLC.C.304 in establishing the superiority of TN® Gel to its components but still met study objectives. In Study VLC.C.305, TN® Gel was clearly superior to its components and vehicle on the basis of percent reduction in 2 of 3 lesion counts. However the percent of subjects achieving a 'clear' or 'almost 'clear' determination for Clindamycin and Tretinoin were quite high in this study compared to Study VLC.C.304. Thus, for the proportion of successes in IGA, a lower strength of evidence is observed in Study VLC.C.305 than compared to Study VLC.C.304.

Table 1: Efficacy Results Summary: Percent Reduction from Baseline for Lesion Counts and Percent Success according to Investigator Global Assessment (IGA)

	TN® Gel	Clindamycin	Tretinoin	Vehicle
Study VLC.C.304[1]				
Sample Size	309	311	310	153
Total Lesions	46.2%	33.8% (p < .0001)	35.6% (p < .0001)	20.0% (p < .0001)
Non-Inflammatory Lesions	41.5%	25.0% (p < .0001)	32.9% (p = .0002)	14.3% (p < .0001)
Inflammatory Lesions	52.3%	46.4% (p = .0529)	39.3% (p < .0001)	27.2% (p < .0001)
IGA	35%	21% (p = .0002)	19% (p < .0001)	12% (p < .0001)
Study VLC.C.305[2]				
Sample Size	325	324	324	162
Total Lesions	51.1%	42.5% (p = .0001)	44.9% (p = .0034)	26.2% (p < .0001)
Non-Inflammatory Lesions	48.8%	37.8% (p < .0001)	42.7% (p = .0121)	22.5% (p < .0001)
Inflammatory Lesions	54.5%	48.6% (p = .0306)	47.2% (p = .0041)	33.3% (p < .0001)
IGA	39%	32% (p = .0441)	31% (p = .0108)	15% (p < .0001)

[1] Tested at a significance level of α=.04806 due to interim analysis.
[2] Tested at a significance level of α=.05.

The sponsor conducted an interim analysis to determine if sample sizes would be sufficient to achieve study objectives or to stop the trial if all primary endpoints were significant at α=.005 (O'Brien Fleming adjustment). The interim analysis was carried out by an in-house manager and issues about retaining blinding of study findings resulted as the original sample size was decided upon by an executive committee who chose to increase the sample size by approximately 196 subjects. This decision was made despite the reported conclusion that the in-house manager advised the executive committee that no sample size increases were needed to meet study objectives. The Agency requested further information of the sponsor about the conduct and data from the interim. From a review of these materials and by examining data as it accumulated in the trials, no apparent bias in efficacy claims could be found. Refer to Section 3.1.7 for further details.

The safety examination showed the local safety of TN® Gel is quite similar to that of Tretinoin. Both of these drug products seemed to elicit an initial local skin reaction to the drug product up until about 2 weeks after which the reactions tended to decrease. Both Clindamycin and Vehicle showed little change in local skin reactions and actually moved toward some improvement over baseline measurements. No serious adverse events that could be attributed to treatment were reported by study investigators for any treatment arm.

2 Introduction

2.1 Overview

TN® Gel is a combination drug product consisting of two individually approved active ingredients, 1% Clindamycin and .025% Tretinoin for the treatment of moderate to severe acne vulgaris. Connetics Corporation acquired the drug product from Yamanouchi Europe, BV and slightly modified the formulation (b) (4) However, this is the only modification to the formulation of TN® Gel (Source: sponsor's submission). The clinical program for TN® Gel in the NDA submitted by the sponsor, Connetics Corporation, consists of two phase 1/2 studies, six European Phase 3 safety and efficacy studies, and two American Phase 3 safety and efficacy studies shown in Table 2 below. The six studies in Table 2 with listings appearing as 9X-CTC-0X are the Phase 3 European Studies conducted by Yamanouchi Europe, BV. Studies VLC.C.304 and VLC.C.305 are the Phase 3 American studies reviewed thoroughly for safety and efficacy in this review.

Table 2: Sponsor's Clinical Study Overview

Study	Objective	Drug Product(s)	Number of Subjects	Treatment Duration
VLC.C.101	Evaluate irritation and sensitization	TN® Gel TN® Gel	243 Total	21 days and 48-72 hrs
VLC.C.201	Evaluate Bioavailabilty	TN® Gel	15	5 days
95-CTC-01	Evaluate Safety and Efficacy	TN® Gel Dalacin® T Lotion	20/group	12 weeks
96-CTC-01	Evaluate Safety and Efficacy	TN® Gel Dalacin® T lotion	209 subjects	12 weeks
91-CTC-01	Evaluate Safety and Efficacy	TN® Gel Clindamycin	81/group	12 weeks
91-CTC-02	Evaluate Safety and Efficacy	TN® Gel Tretinoin	157 Total (77 and 80)	12 weeks
91-CTC-04	Evaluate Safety and Efficacy	TN® Gel Aberela® Gel	161 Total (81 and 80)	12 weeks
92-CTC-01	Evaluate Safety and Efficacy	TN® Gel Aberela® Gel	40 Total (20/group)	12 weeks
VLC.C.304	Evaluate Safety and Efficacy	TN® Gel Clindamycin Tretinoin Vehicle	1083 Total (309, 311, 310 and 153)	12 weeks
VLC.C.305	Evaluate Safety and Efficacy	TN® Gel Clindamycin Tretinoin Vehicle	1136 Total (325, 324, 325 and 162)	12 weeks

2.2 Data Sources

This review is primarily concerned with the two studies VLC.C.304 and VLC.C.305. The data sets for this application are archived in the electronic data room under the listing

\\CDSESUB1\N50-803\N_000\2004-08 23\crt\datasets. The sponsor provided data sets for Studies VLC.C.101, VLC.C.201, VLC.C.304, and VLC.C.305.

3 Statistical Evaluation

3.1 Evaluation of Efficacy

3.1.1 Study Design and Endpoints

Studies VLC.C.304 and VLC.C.305 are multi-center, randomized, double-blind, active- and vehicle-controlled studies to evaluate the safety and efficacy of TN® Gel in treating moderate to severe acne vulgaris. This is a four arm study as TN® Gel is composed of two moieties, 1% Clindamycin and .025% Tretinoin. Efficacy is measured after the twelfth week of treatment with follow-ups at weeks two, four, and eight. The objective of the study is to establish the superiority of TN® Gel to Clindamycin, Tretinoin, and Vehicle on the basis of the two co-primary endpoints

1. Percent Reduction from baseline for two out of three lesion counts (Total, Non-Inflammatory, or Inflammatory).
2. Percent with Investigator Global Assessments (IGA) of 'clear' or 'almost clear'.

Enrollment eligibility requires subjects to have between 17 and 40 inflammatory lesions and between 20 and 150 non-inflammatory lesions among other criteria with the intention of recruiting subjects with moderate to severe acne vulgaris as measured by IGA (i.e. recruitment allows for IGA scores greater than or equal to 2). The initial sample size estimate was 847 where treatment allocation is 2:2:2:1 for TN® Gel, Clindamycin, Tretinoin and Vehicle, respectively. The protocol for Study VLC.C.304 planned an interim analysis after approximately 424 subjects completed 12 weeks of treatment to determine any sample size increases or to stop if all objectives were significant at α=.005 (O'Brien Fleming adjustment). The sponsor states that at the time of the interim analysis, the data suggested the initial sample size estimates were sufficient to power the study with 80% power. However, the application states "an executive decision" was made to increase sample size by 196 subjects per study. The sponsor stated this decision "was not based on the results of the interim analysis, but was made to increase the probability of success and to ensure robustness of the outcome." Aspects and influence of this interim analysis will be explored further in Section 3.1.7.

The first primary endpoint of winning on two of three lesion counts is based upon the percent reduction from baseline. To meet study objectives TN® Gel must be significantly greater than Clindamycin, Tretinoin and Vehicle on two of three lesion counts.
The second primary endpoint, Investigator's Global Assessment (IGA), is a six-value, discrete, static variable defined below in Table 3. IGA is dichotomized to a success or failure, where a success is defined for subjects that are 'clear' or 'almost clear' at the end of treatment (week 12). To win on IGA, TN® Gel must have a significantly greater proportion of successes at the end of treatment than Clindamycin, Tretinoin and Vehicle.

Significance is measured at α=.04806 for Study VLC.C.304 (O'Brien Fleming adjustment for interim analysis) and α=.05 for Study VLC.C.305.

Table 3: Definition of Investigator Global Assessment

Score	Description
0	Normal, clear skin with no evidence of acne vulgaris
1	Skin almost clear: rare non-inflammatory lesions present with rare non-inflammatory papules (papules must be resolving and may be hyper-pigmented though not pink-red) requiring no further treatment in the investigator's opinion.
2	Some non-inflammatory lesions are present with few inflammatory lesions (papules/pustules only, no nodulo-cystic lesions).
3	Non-inflammatory lesions predominate with multiple inflammatory lesion evident: several to many comedones and papules/pustules and there may or may not be one small nodulo-cystic lesion.
4	Inflammatory lesions are more apparent: many comedones and papules/pustules. There may or may not be a few nodulo-cystic lesions.
5	Highly inflammatory lesions predominate: variable number of comedones, many papules/pustules and nodule-cystic lesions.

Source: Module 5 Section 5.3.5.1.1 Page 35 Table 3.

Three secondary endpoints are proposed which include absolute reduction in lesion counts (total, inflammatory, and non-inflammatory), proportion of successes on subject's global assessment (SGA) defined below in Table 4, and time to 50% reduction in total lesion count.

Table 4: Definition of Subject Global Assessment (SGA)

Score	Description
0	My face is basically free of acne with only an occasional blackhead and/or whitehead.
1	My face has several blackheads and/or whiteheads and small pimples, but there are no tender deep-seated bumps or cysts.
2	My face has several to many blackheads and/or whiteheads and small- to medium-sized pimples and may have one deep-seated bump or cyst.
3	My face has many blackheads and/or whiteheads, many medium- to large-sized pimples and perhaps a few deep-seated bumps or cysts.
4	My face has blackheads and/or whiteheads and several to many medium- to large-sized pimples and deep-seated bumps or cysts dominate.

Source: Module 5 Section 5.3.5.1.1 Page 36 Table 4.

3.1.2 Patient Disposition, Demographic and Baseline Characteristics

Section 3.1.2.1 looks at numbers of patients in the analysis populations along with reasons for dropouts. Section 3.1.2.2 explores the demographic characteristics of age, race and gender for the four treatment arms. Lastly, Section 3.1.2.3 explores the baseline levels of the prognostic characteristics for the four treatment arms which could affect efficacy and safety claims if large discrepancies between treatment arms appear.

3.1.2.1 Patient Disposition

Described below are the two analysis populations used in the analysis, the intent-to-treat (ITT) population and the per-protocol (PP) population for each study. The ITT population includes all subjects that are randomized to a treatment arm. The PP population excludes all subjects that missed more than 16 doses, used prohibited medications, did not pass exclusion criteria, did not have efficacy measurements at week 12, or other deviations such as received incorrect dosing, randomized error, etc.

VLC.C.304

The size for each of the analysis populations is given below in Table 5. The breakdown of the excluded subjects in the PP populations for each treatment is included in Table 6. Since information about dropouts is unknown, results will be compared on both study populations to check for consistency in efficacy claims.

Table 5: Analysis Populations for Study VLC.C.304

	TN® Gel	Clindamycin	Tretinoin	Vehicle	Total
ITT population	309	311	310	153	1083
PP population	257 (83%)	260 (84%)	227 (73%)	115 (75%)	859 (79%)

Source: Module 5 Section 5.3.5.1.1 Page 59 Table 8.

Table 6: Listing of Excluded Subjects in PP Population for Study VLC.C.304

	TN® Gel	Clindamycin	Tretinoin	Vehicle	Total
Number Excluded from ITT Analysis Population	52	51	83	38	224
Missed more than 16 Doses	10	6	16	2	34
Failed Exclusion Criteria	2	0	1	0	3
Other Protocol Violations	3	5	3	4	15
No Assessment at Week 12[1]	37	40	63	32	172
Adverse Event	7	0	9	0	16
Non-Compliance	1	2	2	1	6
Disease Progression	0	2	1	1	4
Request to Withdraw	9	7	12	6	34
Subject Death	0	0	0	0	0
Lost to Follow-Up	20	28	38	24	90
Became Pregnant	0	1	1	0	2

Source: Data Sets \\CDSESUB1\N50803\N_000\2004-08-23\crt\vlc.c.304\COMP.xpt and \\CDSESUB1\N50802\N_000\2004-08-2003\crt\vlc.c.304\TERM.xpt.

[1] Reason for no week 12 assessment is given in the lower portion of the table.

VLC.C.305

The size for each analysis population and the reason for per-protocol exclusion is given in Table 7 and Table 8 for Study VLC.C.305. No major subject disposition discrepancies between the two studies are apparent from the information listed in these two tables.

Table 7: Analysis Populations for Study VLC.C.305

	TN® Gel	Clindamycin	Tretinoin	Vehicle	Total
ITT population	325	324	325	162	1136
PP population	258 (79%)	274 (85%)	272 (84%)	134 (83%)	938 (83%)

Source: Module 5 Section 5.3.5.1.2 Page 55 Table 8.

Table 8: Listing of Excluded Subjects in PP Population for Study VLC.C.305

	TN® Gel	Clindamycin	Tretinoin	Vehicle	Total
Number Excluded from ITT Analysis Population	67	50	53	28	198
Missed more than 16 Doses	17	9	9	2	37
Failed Exclusion Criteria	0	0	0	0	0
Other Protocol Violations	6	5	5	2	18
No Assessment at Week 12[1]	44	36	39	24	143
Adverse Event	4	0	8	0	12
Non-Compliance	5	3	2	4	14
Disease Progression	1	1	1	4	7
Request to Withdraw	14	9	10	6	39
Subject Death	0	0	0	0	0
Lost to Follow-Up	19	23	18	10	70
Became Pregnant	1	0	0	0	1

Source: Source: Data Sets \\CDSESUB1\N50803\N_000\2004-08-23\crt\vlc.c.305\COMP.xpt and \\CDSESUB1\N50802\N_000\2004-08-2003\crt\vlc.c.305\TERM.xpt.
[1] Reason for no week 12 assessment is given in the lower portion of the table.

3.1.2.2 Demographic Characteristics

An examination of the baseline demographic characteristics for age, gender and race revealed the subgroups were quite homogenous for both studies. In Study VLC.C.305 an age effect was highly significant, but it is unlikely it would have any impact on study conclusions. Detailed listings of the demographic characteristics for each study can be found in Table A.1.1 and Table A.1.2 in the Appendix.

3.1.2.3 Baseline Prognostic Characteristics

This exploratory analysis performed by the reviewer examines various characteristics that may be related to disease severity. Ideally, we want to see that randomization to treatment for the four treatment arms resulted in similar baseline values of the prognostic characteristics as large disparities could have an effect on safety and efficacy claims. The prognostic characteristics examined include

- Burning: 6-point scale with 0 = normal to 5 = continuous discomfort.
- Dryness: 6-point scale with 0 = normal to 5 = prominent and dense.
- Erythema: 6-point scale with 0 = normal to 5 = prominent and dense.
- Itching: 6-point scale with 0 = normal to 5 = continuous discomfort.
- Scaling: 6-point scale with 0 = normal to 5 = prominent and dense.
- IGA: 6-point scale with 0 = clear to 5 = highly inflammatory lesions predominate.

Also examined are the baseline values of lesion counts including total lesion count, inflammatory lesion count, and non-inflammatory lesion count.

No major differences in the baseline proportion of subjects within a given treatment arm for the associated value of the prognostic characteristic was found in either Study VLC.C.304 or Study VLC.C.305. Testing of whether any differences between treatment arms were significant was done using a Chi-square test which yielded no statistically

significant differences (p > .10). Thus, the randomization resulted in well-balanced treatment arms for baseline values of the prognostic factors.

Baseline lesion counts were also examined for the four treatment arms for each study. Using a Kruskal-Wallis test, no significant differences were found between treatment arms for any of the baseline lesion counts for either study (p > .10). Thus, Study VLC.C.304 and Study VLC.C.305 showed no major differences in prognostic characteristics that might have an impact on safety and efficacy claims.

3.1.3 Statistical Methodology

As stated in the protocol, the sponsor pooled the centers by geographic location when a center did not recruit at least 70 subjects. This resulted in using 13 pooled centers for each study. The pooled centers ranged in size from 70 to 105 and 80 to 105 in Studies VLC.C.304 and VLC.C.305, respectively. Missing data was handled using last-observation-carried-forward (LOCF) for all analyses.

The protocol for Study VLC.C.304 planned an interim analysis after approximately 424 subjects (half of the anticipated 827 subjects) had completed 12 weeks of treatment. Using O'Brien Fleming's methods, the study can be terminated early if all primary objectives are significant at α=.005. Otherwise, the purpose of the interim analysis is to determine if the estimated sample sizes are sufficient to reach study objectives. In order to account for the interim analysis, a significance level of α=.04806 is used for efficacy claims in Study VLC.C.304. Very little detail was provided in the protocol about details of the interim analysis and issues about blinding arose based upon details provided in the submission of the NDA. A thorough description of the interim analysis and possible impacts on efficacy is discussed in more detail in Section 3.1.7. Note that Study VLC.C.305 starts after Study VLC.C.304 so any sample size changes made in Study VLC.C.304 will also be reflected in Study VLC.C.305.

The protocol specified that percent reduction in lesion counts (total, inflammatory, and non-inflammatory) from baseline to week 12 will be analyzed using ANOVA with terms for treatment, pooled center, and pooled center by treatment interaction. If the interaction term was not significant (α=.10) then the interaction term was removed from the model. Since it is unlikely that the distribution of percent change is normal, a Shapiro-Wilk test for normality will be conducted. When the test rejects normality of the data (α = .05), a rank transform of percent change replaces percent change in the ANOVA model. In this review of percent reduction in lesion counts, results of the ANOVA model will be based on the *untransformed* response. This is based on the fact that the ANOVA model is fairly robust to the normality assumption with such large sample sizes. If large discrepancies between the untransformed and transformed response exist, these will be reported. Results of both the untransformed and transformed response will be provided in the Appendix. The protocol stated that if the normality assumption was violated, the primary analysis will be performed on the transformed response.

IGA scores are dichotomized into success and failure. A success consists of an IGA score of 0 or 1 which correspond to 'clear' and 'almost clear', respectively. The protocol specified that the dichotomized IGA will be analyzed using a Cochran-Mantel-Haenszel (CMH) test stratified by pooled center. The stratification term for pooled center is tested using the Breslow-Day test for homogeneity at significance level α =.10.

The secondary endpoints of absolute reduction in lesion counts are analyzed using the same ANOVA strategy as percent reduction as specified in the protocol. SGA values of 0 or 1 at week 12 were counted as successes. The CMH was then used to test for the percent of subjects with successful SGA scores at week 12 in the same fashion as testing for IGA as stated in the protocol.

This reviewer's analysis of the time to 50% reduction in total lesion counts is analyzed using a Cox Proportional Hazards (CPH) Model with a term for treatment stratified by pooled center. In this model, time to event is recorded as the time at which a subject has a 50% reduction in his/her total lesion count. Subjects that complete the study but did not reach 50% reduction are recorded as right-censored 12 week non-events. For subjects that dropout without achieving 50% reduction in total lesion counts, the time of dropout is recorded and censored for being a non-event. If a subject achieved 50% reduction in total lesion counts and dropped out, the time of 50% reduction is recorded and not censored. The score statistic is used to test for differences in TN® Gel and its components and vehicle. Since times are discrete with only a few distinct values, ties are handled using the discrete likelihood method (method = discrete in SAS© PHREG). The protocol specified that the time to 50% reduction would be analyzed "using Kaplan_meyer[sic] methodology." However, no other details about the analysis were provided in the protocol. The sponsor's submission proposes a log-rank test which is equivalent to the score statistic in a CPH Model comparing TN® Gel to Clindamycin, Tretinoin, and Vehicle *without* stratifying by pooled center. In the submission, it is not clear how the data was actually analyzed as inconsistencies in the reporting of number of events exist in the sponsor's submission. Results based on the reviewer's analysis and the sponsor's analysis is provided in this review, along with the data reported in the sponsor's submission.

3.1.4 Primary Endpoint Results

To establish efficacy and meet study objectives stated in the protocol, the sponsor needs to establish superiority of TN® Gel to each of its components and vehicle on percent reduction from baseline for two of three lesion counts and establish superiority of TN® Gel on Investigator's Global Assessment (IGA) using the ITT analysis population (results using the PP population are provided in Section 3.1.5). Results from each pivotal study for each of these co-primary endpoints are discussed in the following sections. The analysis of the co-primary endpoints follows the details given in Section 3.1.3. Detailed tabled results of the statistical analysis can be found in the Appendix Section A.2. Graphical depictions of each primary endpoint across time for each treatment arm are provided in the Appendix Section A.3.

3.1.4.1 Percent reduction in Lesion Counts (ITT Population)

Testing for treatment difference in the precent reduction of total lesion counts showed TN® Gel to be superior to each of its components and vehicle for both studies. The statistical results based on ANOVA are provided in Table 9 for Studies VLC.C.304 and VLC.C.305. More detailed statistical testing results are given in the Appendix Section A.2. From the tables, TN® Gel is superior to each of its components and vehicle for total lesion counts with p-values below α = .01 for every comparison which is below the pre-defined α-level of .04806 and .05 to declare superiority of TN® Gel for studies VLC.C.304 and VLC.C.305, respectively.

Table 9. Percent Reduction in Total Lesion Counts (Sponsor's Results)

	TN® Gel	**Clindamycin**	**Tretinoin**	**Vehicle**
Study VLC.C.304				
N	309	311	310	153
Mean	46.2%	33.8%	35.6%	20.0%
SD	35.3%	33.2%	30.8%	38.7%
p-value[1,*]		p < .0001	p < .0001	p < .0001
Study VLC.C.305				
N	325	324	325	162
Mean	51.1%	42.5%	44.9%	26.2%
SD	28.4%	31.7%	31.0%	38.5%
p-value[2,*]		p = .0001	p = .0034	p < .0001

[1] p-value is based on ANOVA and tested at α = .04806 level due to interim analysis adjustment.
[2] p-value is based on ANOVA and tested at α = .05 level.
[*] p-values are based on the *untransformed* response.
Source: Module 5 Section 5.3.5.1.1 Page 71 Table 16 and Section 5.3.5.1.2 Page 67 Table 16.

Testing for treatment differences in the percent reduction in non-inflammatory lesion counts showed TN® Gel to be superior to each of its components and vehicle for both studies. The ANOVA results are given in Table 10 on the following page for Studies VLC.C.304 and VLC.C.305. From the ANOVA model, p-values for comparing TN® Gel to Clindamycin and Vehicle are below .0001 for both pivotal studies. In comparing TN® Gel to Tretinoin, p-values are slightly larger but still significant for both studies; p=.0065 and p=.0121 for studies VLC.C.304 and VLC.C.305, respectively.

Testing for treatment differences in percent reduction of inflammatory lesion counts found all comparisons of TN® Gel to its components and vehicle to be significant with the exception of comparing TN® Gel to Clindamycin in Study VLC.C.304. The ANOVA results are given in Table 11 on the following page for Studies VLC.C.304 and VLC.C.305. From the table, the statistical superiority of TN® Gel to Vehicle and Tretinoin is quite strong for both pivotal studies (p < .005). When looking at percent change in inflammatory lesion counts, TN® Gel is not significantly greater than Clindamycin in Study VLC.C.304 (p = .0503), but does achieve statistical significance for Study VLC.C.305 (p = .0306).

Table 10. Percent Reduction in Non-Inflammatory Lesion Counts (Sponsor's Results)

	TN® Gel	Clindamycin	Tretinoin	Vehicle
Study VLC.C.304				
N	309	311	310	153
Mean	41.5%	25.0%	32.9%	14.3%
SD	43.5%	41.3%	36.0%	49.3%
p-value[1,*]		p < .0001	p = .0065	p < .0001
Study VLC.C.305				
N	325	324	325	162
Mean	48.8%	37.8%	42.7%	22.5%
SD	31.3%	39.7%	37.1%	45.3%
p-value[2,*]		p < .0001	p = .0121	p < .0001

[1] p-value is based on ANOVA and tested at α = .04806 level due to interim analysis adjustment.
[2] p-value is based on ANOVA and tested at α = .05 level.
[*] p-values are based on the *untransformed* response.
Source: Module 5 Section 5.3.5.1.1 Page 71 Table 16 and Section 5.3.5.1.2 Page 67 Table 16.

Table 11. Percent Reduction in Inflammatory Lesion Counts (Sponsor's Results)

	TN® Gel	Clindamycin	Tretinoin	Vehicle
Study VLC.C.304				
N	309	311	310	153
Mean	52.3%	46.4%	39.3%	27.2%
SD	36.2%	38.7%	37.4%	44.6%
p-value[1,*]		p = .0503	p < .0001	p < .0001
Study VLC.C.305				
N	325	324	325	162
Mean	54.5%	48.6%	47.2%	33.3%
SD	35.5%	37.2%	33.9%	46.9%
p-value[2,*]		p = .0306	p = .0041	p < .0001

[1] p-value is based on ANOVA and tested at α = .04806 level due to interim analysis adjustment.
[2] p-value is based on ANOVA and tested at α = .05 level.
[*] p-values are based on the *untransformed* response.
Source: Module 5 Section 5.3.5.1.1 Page 71 Table 16 and Section 5.3.5.1.2 Page 67 Table 16.

To be declared superior to all other treatments, TN® Gel must be significantly greater than all other treatments on the basis of percent reduction from baseline for two of the three lesion counts tested at the α =.04806 and α = .05 significance level, for Studies VLC.C.304 and VLC.C.305, respectively. The only time TN® Gel was not significantly greater than a comparator was in Study VLC.C.304 for mean percent reduction in inflammatory lesion counts. Thus, TN® Gel has met the requirements to win on two of three lesion counts for each of its components and vehicle for both studies. As a result, on the basis of percent reduction in two of three lesion counts and on the basis of the two pivotal studies, Study VLC.C.304 and Study VLC.C.305, it can be declared that TN® Gel is superior to its components and vehicle.

3.1.4.2 Success on Investigator Global Assessment (ITT Population)

In the analysis of percent success on IGA, TN® Gel was superior to each of its components and vehicle for both pivotal studies. A success is defined for subjects that receive an IGA score of 0 (clear) or 1 (almost clear) at week 12. To test the superiority of

TN® Gel to the other three treatment arms, a Cochran-Mantel-Haenszel (CMH) test was carried out with adjustments for pooled center. The results of the CMH test are given in Table 12. Based on the CMH tests, TN® Gel is superior to each of its components and vehicle in Study VLC.C.304 and Study VLC.C.305. Hence, both Studies VLC.C.304 and VLC.C.305 met the pre-specified objectives for IGA.

Consistency of the results for IGA success was verified using the Breslow-Day test of homogeneity of the odds ratios at the $\alpha = .10$ level. The only time the Breslow-day test of homogeneity was rejected was in the comparison of TN® Gel to Vehicle in Study VLC.C.305 ($p = .0274$). All other tests for homogeneity did not have any statistically significant results in both studies. In Study VLC.C.305, pooled center 2 shows a higher success rate for Vehicle than TN® Gel and this appears to be the cause of the statistical significance. However, the success rates for either TN® Gel or Vehicle are not extremely different in terms of raw percentages from success rates of other pooled centers decreasing any concerns about the lack of homogeneity of the odds ratios.

Table 12. Percent of Subjects with IGA Success (Sponsor's Results)

	TN® Gel	**Clindamycin**	**Tretinoin**	**Vehicle**
Study VLC.C.304				
Number of Subjects	309	311	310	153
Success	107 (34.6%)	66 (21.2%)	60 (19.4%)	19 (12.4%)
p-value[1]		p = .0002	p < .0001	p < .0001
Study VLC.C.305				
Number of Subjects	325	324	325	162
Percent Success	128 (39.4%)	105 (32.4%)	100 (30.8%)	24 (14.8%)
p-value[2]		p = .0441	p = .0108	p < .0001

[1] p-value is based upon a CMH test adjusting for pooled site at $\alpha = .04806$ level (interim adjustment).
[2] p-value is based upon a CMH test adjusting for pooled site at the $\alpha = .05$ level.
Source: Module 5 Section 5.3.5.1.1 Page 77 Table 18 and Module 5 Section 5.3.5.1.2 Page 72 Table 18.

Table 13 on the following page shows the success rate on the basis of IGA for the two studies broken down by the baseline IGA score and treatment arm. Here it can be seen that when baseline disease severity increased, the success rate of treatment decreased with the exception of Study VLC.C.305 which had high success rates when baseline was 5, but the rates are based on a very small number of subjects. Only 18 subjects were entered into Study VLC.C.305 with baseline scores of 5 and interestingly, all these subjects came from a single center. This was center number 110 with lead investigator Dr. Dan K. Chalker. In general, center 110 has quite high success rates in comparison to the other centers. Based upon an in depth review of whether any center may have driven efficacy claims for either study did not show center 110 or any other center to be overly influential in efficacy results.

Overall, the analysis of the primary endpoints showed that both pivotal studies met pre-specified protocol objectives in declaring TN® Gel superior to each of its components and vehicle. Both studies showed strong efficacy claims for percent reduction in two out of three lesion counts. Study VLC.C.304 also had strong efficacy claims as measured by IGA. Study VLC.C.305 was not as strong as Study VLC.C.304 in IGA results, but it still met its study objective.

Table 13: Percent of Subjects with IGA Success Broken Down by Baseline IGA Score.

Success/Total (Percent Success)	**TN® Gel**	**Clindamycin**	**Tretinoin**	**Vehicle**
Study VLC.C.304				
Base = 2	37/79 (46.8%)	26/77 (33.8%)	24/74 (36.5%)	7/34 (20.6%)
Base = 3	64/187 (34.2%)	35/193 (18.1%)	35/191 (18.3%)	11/98 (11.2%)
Base = 4	6/43 (14.0%)	5/40 (12.5%)	1/44 (2.3%)	1/21 (4.8%)
Base = 5	0/0 (0%)	0/1 (0%)	0/1 (0%)	0/0 (0%)
Total	107/309 (34.6%)	66/311 (21.2%)	60/310 (19.4%)	19/153 (12.4%)
Study VLC.C.305				
Base = 2	37/66 (56.1%)	24/55 (43.6%)	21/60 (35.0%)	8/28 (28.6%)
Base = 3	73/201 (36.3%)	64/189 (33.9%)	69/204 (33.8%)	15/96 (15.6%)
Base = 4	17/56 (30.3%)	14/72 (19.4%)	8/56 (14.3%)	0/35 (0%)
Base = 5	1/2 (50.0%)	3/8 (37.5%)	2/5 (40.0%)	1/3 (33.3%)
Total	128/325 (39.4%)	105/324 (32.4%)	100/325 (30.8%)	24/162 (14.8%)

Source: Reviewer's analysis using data sets \\CDSESUB1\N50802\N_000\2004-08-2003\crt\vlc.c.304\A_EFF.xpt. and \\CDSESUB1\N50802\N_000\2004-08-2003\crt\vlc.c.305\A_EFF.xpt.

3.1.5 Primary Endpoint Results (PP Population)

Recall from Section 3.1.2.1, the per-protocol (PP) analysis population excludes subjects that missed more than 16 doses, used prohibited medications, did not pass exclusion criteria, did not have efficacy measurements at week 12, or any other deviation. The results from the PP analysis population are consistent with those reported for the ITT analysis population showing superiority of TN® Gel to each of its components and vehicle on the basis of each primary endpoint for both studies. Tabled results are provided on the following page.

Table 14: Percent Reduction in Lesion Counts for Study VLC.C.304 (PP)

	TN® Gel	Clindamycin	Tretinoin	Vehicle
	N=257	N=260	N=227	N=115
Total Lesions				
Mean	50.3%	38.1%	41.5%	20.6%
SD	33.9%	33.5%	30.4%	40.7%
p-value[1]		p < .0001	p = .0027	p < .0001
Non-Inflammatory Lesions				
Mean	46.0%	29.3%	37.9%	14.7%
SD	42.3%	42.5%	37.3%	52.9%
p-value[1]		p < .0001	p = .0339	p < .0001
Inflammatory Lesions				
Mean	55.8%	50.6%	46.5%	27.4%
SD	34.9%	38.1%	36.0%	48.0%
p-value[1]		p = .0815	p = .0018	p < .0001

[1] p-value is based upon ANOVA results excluding the interaction term tested at α = .04806.
Source: Module 5 Section 5.3.5.1.1 Page 74 Table 17.

Table 15: Percent Reduction in Lesion Counts for Study VLC.C.305 (PP)

	TN® Gel	Clindamycin	Tretinoin	Vehicle
	N=258	N=274	N=272	N=134
Total Lesions				
Mean	57.2%	47.4%	49.8%	32.5%
SD	23.3%	29.6%	25.8%	36.9%
p-value[1]		p < .0001	p < .0001	p < .0001
Non-Inflammatory Lesions				
Mean	54.1%	43.0%	47.7%	28.7%
SD	28.5%	37.1%	30.9%	43.6%
p-value[1]		p < .0001	p = .0022	p < .0001
Inflammatory Lesions				
Mean	62.0%	53.4%	51.7%	38.9%
SD	26.4%	35.4%	31.6%	44.6%
p-value[1]		p = .0012	p = .0018	p < .0001

[1] p-value is based upon ANOVA results excluding the interaction term tested at α = .05.
Source: Module 5 Section 5.3.5.1.2 Page 70 Table 17.

Table 16: Percent of Subjects with IGA Success (PP)

	TN® Gel	Clindamycin	Tretinoin	Vehicle
Study VLC.C.304				
Number of Subjects	257	260	227	115
Success	99 (38.5%)	65 (25.0%)	56 (24.7%)	16 (13.9%)
p-value[1]		p = .0009	p = .0009	p < .0001
Study VLC.C.305				
Number of Subjects	258	274	272	134
Percent Success	117 (45.3%)	101 (36.9%)	90 (33.1%)	23 (17.2%)
p-value[2]		p = .0432	p = .0031	p < .0001

[1] p-value is based upon a CMH test adjusting for pooled site at α = .04806 level (interim adjustment).
[2] p-value is based upon a CMH test adjusting for pooled site at the α = .05 level.
Source: Module 5 Section 5.3.5.1.1 Page 78 Table 19 and Section 5.3.5.1.2 page 74 Table 19.

3.1.6 Secondary Endpoint Results

Three secondary endpoints were listed in the statistical analysis plan. These are:

1. Absolute change in lesion counts (total, inflammatory, non-inflammatory) from baseline to week 12.
2. Percent with Subject Global Evaluation (SGA) scores of 0 or 1 at week 12.
3. Time to 50% reduction in total lesion counts.

Analysis of the secondary endpoints is limited to the ITT population only.

3.1.6.1 Absolute Change in Lesion Counts

The analysis of the absolute reduction in lesion counts (baseline – week 12) found TN® Gel to be superior to each of its components and vehicle for ALL lesion counts and for both studies. The results of this secondary analysis are shown in Table 17 and Table 18 on the following page. The conclusions based upon this analysis are consistent with those found for the percent reduction in lesion counts. All p-values from each test comparing TN® Gel to Clindamycin, Tretinoin and Vehicle are all below a significance level of α=.05 for both studies.

Table 17: Absolute Reduction in Lesion Counts for Study VLC.C.304 (Sponsor's Results)

	TN® Gel	**Clindamycin**	**Tretinoin**	**Vehicle**
	N=309	N=311	N=310	N=153
Total Lesions				
Mean	35.0	23.7	25.6	12.9
SD	31.5	25.9	25.2	29.5
p-value[1]		p = .0004	p < .0001	p < .0001
Non-Inflammatory Lesions				
Mean	21.9	12.4	16.1	6.1
SD	26.8	21.2	20.0	24.3
p-value[1]		p < .0001	p = .0010	p < .0001
Inflammatory Lesions				
Mean	13.1	9.8	9.5	6.8
SD	11.3	10.0	9.5	11.0
p-value[1]		p = .0234	p < .0001	p < .0001

[1] p-value is based upon ANOVA using a rank transform of the response.
Source: Module 5 Section 5.3.5.1.1 Page 80 Table 20.

Table 18: Absolute Reduction in Lesion Counts for Study VLC.C.305 (Sponsor's Results)

	TN® Gel	Clindamycin	Tretinoin	Vehicle
	N=325	N=324	N=325	N=162
Total Lesions				
Mean	41.5	32.6	35.5	19.6
SD	28.6	28	29.0	34.6
p-value[1]		p < .0001	p = .0036	p < .0001
Non-Inflammatory Lesions				
Mean	26.8	19.7	23.0	10.9
SD	22.7	22.7	24.4	27.3
p-value[1]		p < .0001	p = .0102	p < .0001
Inflammatory Lesions				
Mean	14.7	12.9	12.5	8.6
SD	10.3	10.8	10.0	13.9
p-value[4]		p = .0141	p = .0003	p < .0001

[1] p-value is based upon ANOVA using a rank transform of the response.
Source: Module 5 Section 5.3.5.1.2 Page 76 Table 20.

3.1.6.2 Subject Global Assessment

The analysis of the percent of subjects achieving success on the basis of the subject global assessment (SGA) found TN® Gel to be superior to each of its components and vehicle in both pivotal studies. A success occurs if subjects rate themselves with an SGA=0 or 1 as defined in Section 3.1.1. A CMH test was used to test for differences between TN® Gel and its components and vehicle stratifying by the pooled center. Table 19 depicts the results for Studies VLC.C.304 and VLC.C.305, respectively. Testing for the superiority of TN® Gel to Clindamycin, Tretinoin and Vehicle shows significance at an α =.05 level.

Table 19: Percent of Subjects with SGA Success

	TN® Gel	Clindamycin	Tretinoin	Vehicle
Study VLC.C.304				
Number of Subjects	309	311	310	153
Success	184 (59.7%)	159 (51.3%)	143 (46.1%)	57 (37.3%)
p-value[1]		p = .0330	p = .0006	p < .0001
Study VLC.C.305				
Number of Subjects	325	324	325	162
Percent Success	206 (63.4%)	177 (54.6%)	182 (56.0%)	71 (43.8%)
p-value[1]		p = .0188	p = .0478	p < .0001

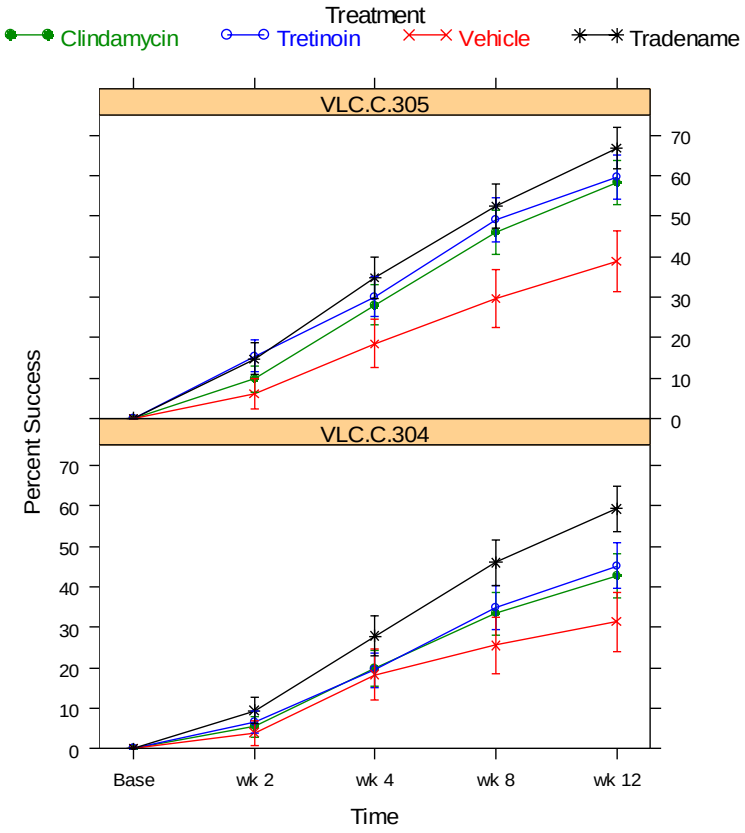
[1] p-value is based upon a CMH test adjusting for pooled center.
Source: Module 5 Section 5.3.5.1.1 Page 84 Table 22 and Module 5 Section 5.3.5.1.2 Page 80 Table 22.

3.1.6.3 Time to 50% Reduction in Total Lesion Counts

The reviewer's analysis of the time to 50% reduction in total lesion counts showed TN® Gel to be superior to its components and vehicle in both studies with the exception that TN® Gel failed to show a significant difference from Tretinoin in Study VLC.C.305

(p=.10). If we define a success when a subject reaches a 50% reduction in total lesion counts, Figure 1 depicts the percent of successes across time for each of the treatment groups. The banded curves in the figure correspond to a 95% confidence interval of the raw point estimate unadjusted for pooled center. From the figure it is quite apparent that a higher percentage of subjects reached 50% reduction in total lesion counts in the TN® Gel treatment arm than in the other treatment arms for Study VLC.C.304. However, TN® Gel does not appear to show such an increased percentage of successes over Tretinoin in Study VLC.C.305.

Figure 1. Time to 50% Reduction in Lesion Counts



To analyze the data, a time to event analysis was carried out, where an event is defined as reaching a 50% reduction in total lesion counts. This reviewer's analysis used a Cox Proportional Hazards model with a term for treatment and stratified by pooled center. A score test is used to test for treatment effects and ties are handled using the discrete likelihood method. The analysis was also carried out without stratification by pooled center which follows the sponsor's analysis (not defined in the protocol). From Table 20 and Table 21 we can see that the time to 50% reduction at week 12 for TN® Gel is significantly greater than the time to 50% reduction in Clindamycin, Tretinoin, and Vehicle for Study VLC.C.304 ($p < .05$). In Study VLC.C.304 over half of the subjects achieved a 50% reduction in total lesion counts when using TN® Gel, whereas the other three treatments had fewer than half of the subjects with at least a 50% reduction in total lesion counts. Contrasting this claim with Study VLC.C.305, it can be seen that all active

treatment arms show more than half of the subjects with 50% reduction in total lesion counts. As a result, in testing for statistical significance of TN® Gel to its components and vehicle, TN® Gel is not statistically superior to Tretinoin in terms of time to 50% reduction in total lesion counts at week 12 (p =.1009) in Study VLC.C.305. In the two tables it can be seen that the sponsor's results are inconsistent with this review. In Section 5.3.5.1.1 (Page 87) and Section 5.3.5.1.2 (Page 82) of Studies VLC.C.304 and VLC.C.305, respectively, Table 24 of the sponsor's submission reports exactly the same number of events as those in the table which are *not* in parentheses and located on the left side of the cells. However, in the sponsor's analysis, the number of events changes and these are the quantities reported in parenthesis on the right side of the cells in the tables below. These numbers can be found in the sponsor's submission in Section 5.3.5.1.1 pages 276-278 for Study VLC.C.304 and Section 5.3.5.1.2 pages 257-259 for Study VLC.C.305. It is unclear where these numbers came from, but it is still worth reporting that on the basis of the sponsor's results and those of the reviewer, TN® Gel failed to show a statistically significant difference from Tretinoin in Study VLC.C.305 when testing at the α = .05 level.

Table 20: Time to 50% Reduction for Study VLC.C. 304

	TN® Gel		**Clindamycin**		**Tretinoin**		**Vehicle**	
	N=309		N=311		N=310		N=152	
50% Reduction: Number Events (%)								
Week 2	29	(29)	17	(17)	20	(20)	6	(6)
	9.4%	(9.4%)	5.5%	(5.5%)	6.5%	(6.5%)	3.9%	(3.9%)
Week 4	86	(100)	62	(80)	60	(88)	28	(38)
	27.8%	(32.4%)	19.3%	(25.7%)	19.4%	(28.4%)	18.3%	(24.8%)
Week 8	142	(160)	104	(131)	108	(152)	39	(54)
	46.0%	(51.8%)	33.4%	(42.1%)	34.8%	(49.0%)	25.5%	(35.3%)
Week 12	183	(211)	133	(167)	140	(191)	48	(67)
	59.2%	(68.3%)	42.8%	(53.7%)	45.2%	(61.6%)	31.4%	(43.8%)
p-value[1]			< .0001		.0042		< .0001	
p-value[2]			< .0001		.0058		< .0001	
p-value[3]			< .0001		.0044		< .0001	

Results in parentheses (right side of cell) correspond to unverifiable results in the sponsor's application.

[1] p-value is from a Cox PH model for time to 50% reduction in total lesion counts using discrete likelihood methods for ties (method = discrete in SAS PHREG) stratified by pooled center.

[2] p-value is from a Cox PH model for time to 50% reduction in total lesion counts using discrete likelihood methods for ties (method = discrete in SAS PHREG) *NOT* stratified by pooled center.

[3] p-value reported in sponsor's submission: Figures 14.2.1 – 14.2.3 in Section 5.3.5.1.1 Pages 276-278.

Table 21: Time to 50% Reduction for Study VLC.C.305

	TN® Gel		Clindamycin		Tretinoin		Vehicle	
	N=325		N=324		N=325		N=162	
50% Reduction: Number Events (%)								
Week 2	48	(48)	32	(32)	50	(50)	10	(10)
	14.8%	(14.8%)	9.9%	(9.9%)	15.4%	(15.4%)	6.2%	(6.2%)
Week 4	113	(133)	91	(103)	98	(116)	30	(36)
	34.8%	(40.9%)	28.1%	(31.8%)	30.2%	(35.7%)	18.5%	(22.2%)
Week 8	171	(197)	149	(170)	160	(183)	48	(61)
	52.6%	(60.6%)	46.0%	(52.5%)	49.2%	(56.3%)	29.6%	(37.7%)
Week 12	217	(253)	189	(219)	194	(224)	63	(84)
	66.8%	(77.8%)	58.3%	(67.6%)	60.0%	(68.9%)	38.9%	(51.9%)
p-value[1]			.0019		.1009		< .0001	
p-value[2]			.0051		.0608		< .0001	
p-value[3]			.0045		.0539		< .0001	

Results in parentheses (right side of cell) correspond to unverifiable results in the sponsor's application.
[1] p-value is from a Cox PH model for time to 50% reduction in total lesion counts using discrete likelihood methods for ties (method = discrete in SAS PHREG) stratified by pooled center.
[2] p-value is from a Cox PH model for time to 50% reduction in total lesion counts using discrete likelihood methods for ties (method = discrete in SAS PHREG) *NOT* stratified by pooled center.
[3] p-value reported in sponsor's submission: Figures 14.2.1 – 14.2.3 in Section 5.3.5.1.2 Pages 257-259.
Source: Reviewer's analysis and Module 5 Section 5.3.5.1.2 Page 82 Table 24.

The results from the analysis of the time to 50% reduction in the total lesion counts do not establish that TN® Gel is superior to Tretinoin when testing at an α=.05 significance level. In the proposed labeling of TN® Gel the sponsor states, (b) (4) [redacted] When examining median times to 50% reduction, using the sponsor's analysis results (values denoted in parentheses in Tables 20 and 21) over half the subjects using the monads Clindamycin or Tretinoin in addition to TN® Gel reach 50% reduction in their total lesion counts at week 8 in Study VLC.C.305. If relying solely on the results found by the reviewer and also provided in the sponsor's submission, at 8 weeks only 46.0% of subjects had 50% reduction in lesion counts in Study VLC.C.304 suggesting the median time is not 8 weeks. Hence, the results (b) (4) [redacted] appear to rely on the combination of the trials and are not reproduced in both studies. In the sponsor's analysis, by combining the data and without stratifying by pooled center, all comparisons to TN® Gel are statistically significant (p < .005), but this result also seems to be driven by Study VLC.C.304 and the increase in sample size. (b) (4) [redacted]

3.1.7 Interim Analysis Examination

This section of the review examines the impact of the choice to increase sample sizes for the studies which was made after the results of an interim analysis concluded that no

sample adjustments need to be made to meet study objectives. The original protocol planned to recruit 847 subjects for Study VLC.C.304. The sponsor stated that the sample size estimate was based upon estimates of treatment effects from "the results of data analysis and published data" that also allowed for a 20% dropout rate and to have 80% power. The sponsor's stated objectives of the interim analysis were

1. Determine if sample sizes were adequate in comparing drugs to Clindamycin, Tretinoin, and Vehicle for Percent Reduction in Lesion Counts to maintain testing at an α =. 04806 level with 80% power.
2. Determine if sample sizes were adequate in comparing TN® Gel to Vehicle for IGA to maintain testing at an α = .04806 level with 80% power.
3. Stop early if all study objectives were significant at the α = .005 level (O'Brien Fleming adjustment).

In the Phase 3 protocol submitted to the Agency very little detail was provided about the conduct of the interim analysis. Upon receiving the NDA, several issues related to blinding of the results arose which may have impacted efficacy results. The first major issue is the fact that an in-house manager of Connetics conducted the interim analysis and was made aware of treatment effects to which he could have stopped the trial early or increase sample size based upon estimates of the treatment effects. To ensure study blinding, the Agency would have preferred that the interim analysis be planned to be carried out by a statistician or other person who is not affiliated with the sponsor. This is usually the case as set by the charter of the DSMB. Also, it should be noted that the O'Brien-Fleming adjustments proposed by the sponsor control for the Type I error associated with the multiple looks at the data for a sequential study design with fixed sample size. However, it does not control for the Type I error associated with adaptive designs in which sample size adjustments are based upon midcourse looks at the data, as demonstrated in Cui *et al.* (1999). Thus, the O'Brien-Fleming adjustments proposed by the sponsor might not preserve the Type I error if sample sizes are increased based upon interim data estimates of treatment effects. Some possible techniques that may be properly applied to allow for stopping the trial early or adjusting sample sizes can be found in Cui *et al.* (1999), Mehta and Tsiatis (2001) and Chen *et al.* (2004).

As mentioned, the interim analysis was carried out by an in-house manager of Connetics Corporation after approximately 424 subjects completed treatment. This manager was to sign a non-disclosure agreement and perform the analysis at an undisclosed location. The manager concluded that the original sample size estimate of 847 subjects was sufficient to detect statistical significance at an α = .04806 level with 80% power for objectives 1 and 2 above. This conclusion was then relayed to the Executive Committee who chose to increase sample size by 196 subjects in order to "increase the probability of the positive trial outcome and to ensure robustness of the results" without allegedly knowing any specific details of the interim analysis.

As we have seen, based upon the larger sample sizes, TN® Gel is superior to Clindamycin, Tretinoin, and Vehicle on the basis of percent reduction in two of three lesion counts and percent successes on IGA. In this section Studies VLC.C.304 and VLC.C.305 are analyzed to look at possible effects of the increased sample size. First,

Study VLC.C.304 is analyzed for the first 424 subjects that entered the study. A mock interim analysis is then conducted to see if the results agree with those submitted to the Agency on 12/09/2004 as requested by the Agency on 11/05/2004. Then both pivotal studies will be analyzed with 848 subjects (the original planned sample size) to see if results are consistent with the increased sample sizes.

Table 22 below shows the results submitted by the sponsor. The original sample size estimates were 242 for the active arms and 121 for the Vehicle arm. Note that comparisons of TN® Gel to Clindamycin and Tretinoin was not tested or considered in the sample size calculations for IGA. Also note that the original sample estimates were not sufficient in comparing TN® Gel to Clindamycin and Tretinoin for inflammatory lesions, but this was only one of three lesion counts that did not achieve its objective. In the sponsor's interim analysis protocol, sample size estimates were to allow for a 10% to 20% dropout rate. Including dropout rates and based upon the results of percent change for lesion counts in comparing TN® Gel to Clindamycin, the sample size estimates should be increased to somewhere in the range of 935 (assume 10% dropout) to 1019 (assume 20% dropout). Thus, from the data submitted by the sponsor, it is not clear the reasoning behind the manager's report to the Executive Committee that no sample size adjustments need to be made.

Table 22: Results of the Interim Analysis as Submitted by the Sponsor

	TN® Gel	**Clindamycin**	**Tretinoin**	**Vehicle**
	N=123	N=120	N=123	N=61
Total Lesions				
Mean (SD) *N* to maintain 80% power	44.2% (29.3%)	35.1% (35.2%) 242*	36.1% (29.2%) 211*	23.7% (39.1%) 45*
Non-Inflammatory Lesions				
Mean (SD) *N* to maintain 80% power	40.7% (35.8%)	24.7% (46.9%) 138*	31.1% (32.3%) 222*	17.9% (55.3%) 71*
Inflammatory Lesions				
Mean (SD) *N* to maintain 80% power	48.0% (35.9%)	48.2% (35.0%) x	41.3% (34.8%) 460	30.2% (39.0%) 58*
IGA				
Percent Success *N* to maintain 80% power	32.5%	N/A N/A	N/A N/A	16.4% 87*

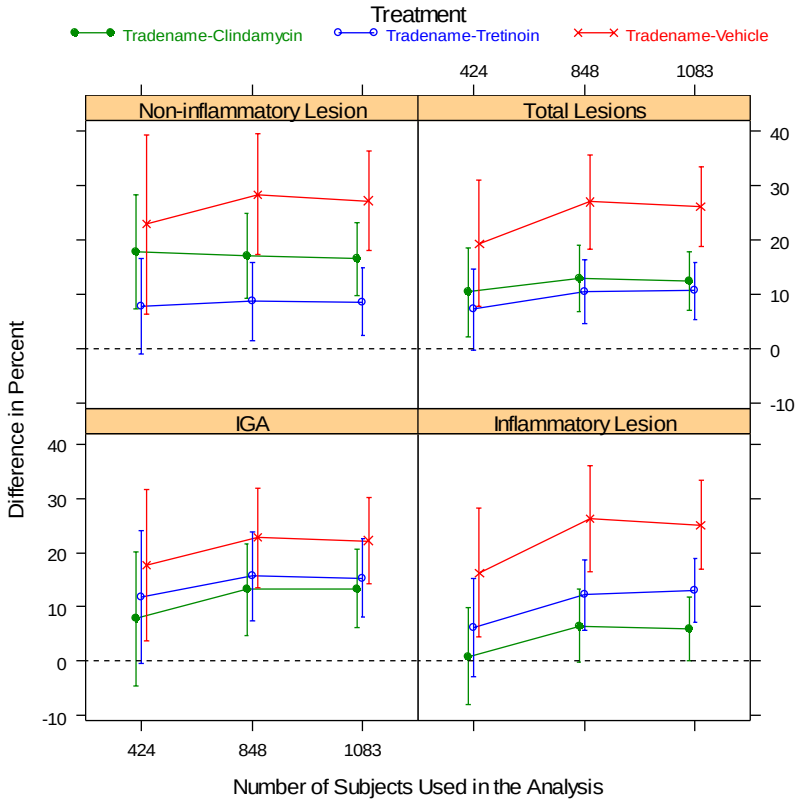
* Denotes cases where the required *N* to maintain 80% power and declare significance of TN® Gel to comparator at α=.04806 was less than or equal to the original sample size estimate (242 active, 121 Vehicle).
x denotes a case where the treatment effect of TN® Gel is below a comparator so no power calculations are performed.
N/A denotes cases that were not analyzed by the sponsor.
Source: Reviewer's analysis using \\CDSESUB1\N50802\N_000\2004-08-2003\crt\vlc.c.304\A_EFF.xpt.

Since the conclusion that the "proposed sample size was sufficient to achieve study objectives" does not seem to be supported, the behavior of the data as it accumulates is examined. In the following examination, the choice of subjects is based upon the date recorded for a subject's second visit. For example, the 424 subjects with the earliest recorded dates on the second visit will be assumed to be the subjects used in the interim analysis. Note that this doesn't coincide completely with the reported data from the sponsor, but they appear to be very similar. The choice of subsetting the data based upon

the date of later visits could be done, but the subsets would remain quite consistent with subsetting on the basis of the second visit date so second visit date is used to select 424 subjects. As with the efficacy analysis, missing data will be calculated by LOCF.

Figure 2 depicts the behavior of the Study VLC.C.304 data as it accumulates. The banded lines correspond to 95% confidence intervals of the raw estimate of the mean difference. In order for Study VLC.C.304 to be declared a success, it must win on IGA (lower left-hand panel) and also on 2 of 3 lesion counts (the remaining panels). A difference greater than zero implies TN® Gel is better than the component or vehicle (Note: this is without controlling for pooled center). The first banded lines correspond to what the data looked like at the interim for this examination that included 424 subjects chosen by date of entry. Note that this analysis data set is different from that actually used by the sponsor. If this data was used at the interim and the study would stop if $p < .005$ (O'Brien-Fleming adjustment), several conclusions can be made. First, on the basis of IGA TN® Gel does not reach a statistically significant difference from Clindamycin ($p = .1824$). Second, TN® Gel does not reach a statistically significant difference from Clindamycin for inflammatory lesions ($p = .8515$) and total lesions ($p=.0132$), but does achieve statistical significance for non-inflammatory lesion counts ($p = .0010$). Third, TN® Gel is not statistically superior to Tretinoin on the basis of total lesions ($p = .0573$), non-inflammatory lesions ($p = .0820$), inflammatory lesions ($p = .1892$), or IGA ($p=.0341$). Lastly, TN® Gel achieves statistical significance in the comparison to Vehicle for all lesion counts ($p < .005$) but not for IGA ($p = .0223$).

Figure 2: Behavior of Efficacy Data as Subjects Accrue in Study VLC.C.304

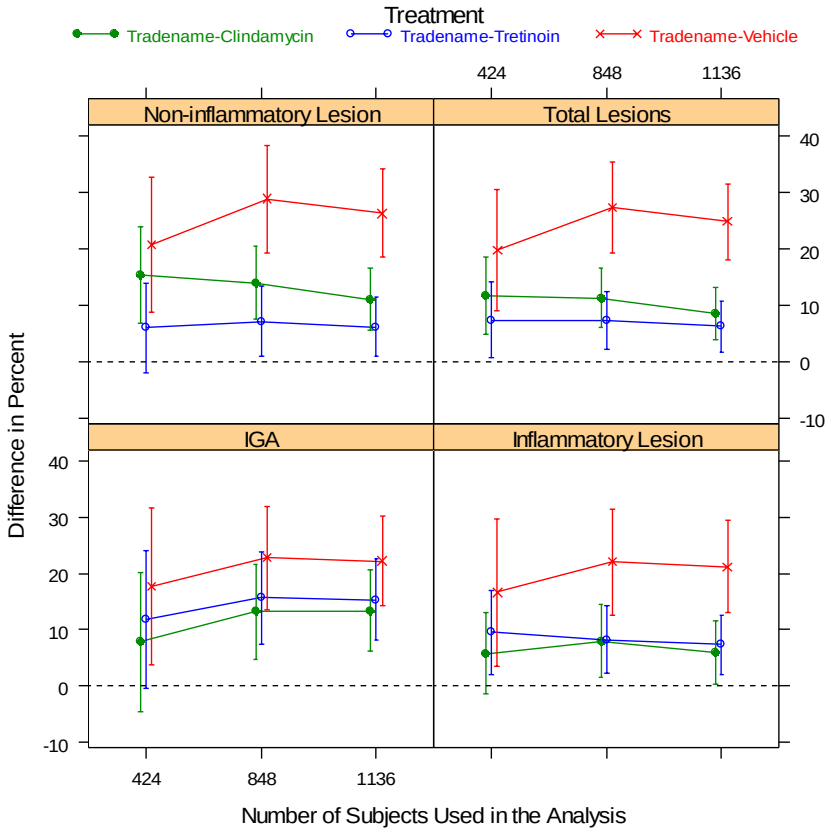


The second group of banded lines in Figure 2 corresponds to results for 848 subjects which is near the original planned sample size. Looking at the data up to this point, TN® Gel is superior to all comparators on the basis of IGA ($p < .002$) and two of three lesion counts ($p < .05$ for at least two of three lesions). Thus, if Study VLC.C.304 was terminated after 848 patients, all study objectives would be met if testing at α=.04806 level.

The third group of banded lines depicts the information of the 1083 subjects from Study VLC.C.304. As mentioned in Section 3, all study objectives were met when using these 1083 subjects. It is interesting to note that overall mean levels of the difference tended to decrease when increasing sample size from 848 to 1083. This tends to suggest that the influence of increasing sample size did not bias the efficacy claims towards a favorable conclusion for the sponsor in Study VLC.C.304.

This same type of exploration was also performed for Study VLC.C.305 and results are shown in Figure 3. It is of interest to note that if Study VLC.C.305 was stopped after 848 subjects had accrued, all objectives would have easily been met as can be seen in the figure as none of the lines cross the zero-threshold. However, when increasing the sample size to 1136 and looking at IGA, the difference between TN® Gel and Clindamycin decreases which results in a larger p-value than when using 848 subjects ($p = .0441$ vs. $p = .00060$).

Figure 3: Behavior of Efficacy Data as Subjects Accrue in Study VLC.C.305



The overall conclusion reached from this exploration is that the decision to increase sample size after the interim did not increase the likelihood of success which would thereby yield favorably biased conclusions for the sponsor. Thus, with the information provided by the sponsor and from the data of the two pivotal studies, it appears as if the conclusions of efficacy are not negatively impacted by results of the interim analysis (i.e. no favorable bias towards efficacy is seen by increasing sample size).

3.2 Evaluation of Safety

3.2.1 Local Skin Reactions

In an evaluation of the local safety of TN® Gel, the local safety profile of TN® Gel is very similar to that of Tretinoin. Five skin irritation factors were measured at baseline and at weeks 2, 4, 8, and 12. These factors are burning, drying, erythema, itching, and scaling. Each factor is based upon a 6 point scale (0 = normal to 5 = extensive). Figure 4 on the following page depicts the means of each of the five skin irritation factors for the four treatment arms for the combined data of Studies VLC.C.304 and VLC.C.305. As can be seen in the plot below and as discussed in more detail in Section 3.1.2.3, each of the four treatment arms have similar baseline values. Also note that the mean never goes above 1 for any treatment arm for any of the irritation factors. No matter the irritation factor, after treatment is administered Velac® Gel and Tretinoin have higher mean levels than Clindamycin and Vehicle. The general trend of Velac® Gel and Tretinoin is to increase the mean value of irritation from the mean baseline value up until week 2. Then, gradually the mean value of irritation starts to decrease and approach the baseline value. In terms of itching and erythema, all treatment arms showed an improvement in the mean value from baseline to week 12. In all of the irritation factors, the mean value of irritation for Clindamycin and vehicle tends to decrease as time progresses resulting in lower mean values than at baseline.

3.2.2 Adverse Events

Studies VLC.C.304 and VLC.C.305 were combined to look at the rate of incidence for adverse events. The tabled results of the adverse events occurring in at least 2% of the subjects are shown in Table 23 on the following page. From this table, TN® Gel has significantly more adverse events than either Clindamycin or its vehicle (p=.0002 and p=.0006, respectively). However, both TN® Gel and Tretinoin have very similar adverse event rates, and this seems to follow the trend seen in Figure 4.

Figure 4: Mean Value of Local Skin Reactions

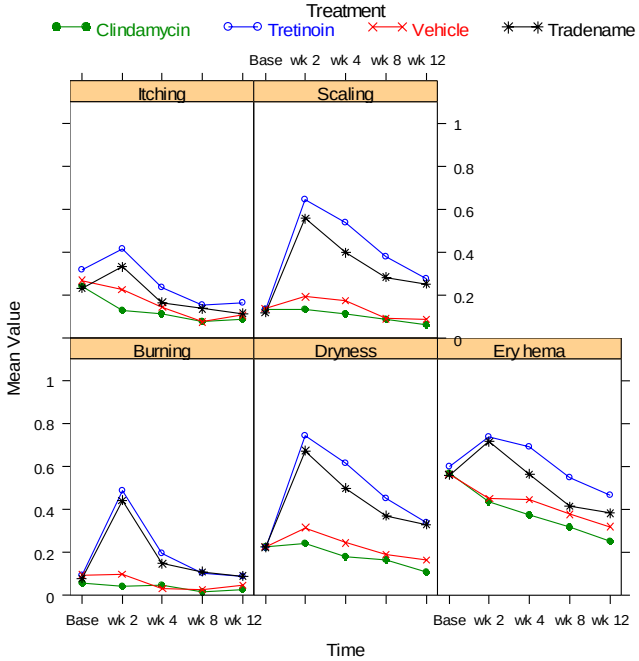


Table 23: Combined Adverse Events for Study VLC.C.305 and Study VLC.C.305

Term	TN® Gel	Clindamycin	Tretinoin	Vehicle	Total
N	634	635	635	315	2219
Subjects with AE	232 (37%)	172 (27%)	232 (37%)	91 (29%)	727 (33%)
p-value		p =.0002	p = .9829	p = .0006	
Application Site Dryness	58 (9%)	12 (2%)	50 (8%)	3 (1%)	123 (6%)
Application Site Desquamation	52 (8%)	2 (0%)	47 (7%)	2 (1%)	103 (5%)
Nasopharyngitis	28 (4%)	25 (4%)	26 (4%)	15 (5%)	94 (4%)
Application Site Burning	40 (6%)	3 (0%)	41 (6%)	5 (2%)	89 (4%)
Application Site Erythema	36 (6%)	6 (1%)	33 (5%)	3 (1%)	78 (4%)
Application Site Pruritus	25 (4%)	7 (1%)	21 (3%)	6 (2%)	59 (3%))
Upper Respiratory Tract Infection NOS	19 (3%)	11 (2%)	11 (2%)	11 (3%)	52 (2%)

Source: Module 2: Section 2.7.4 Page 87 Table 20
[1] p-value is based on a CMH test for the 2x2 contingency table.

A summary of subjects that experienced a serious or life-threatening adverse experience are provided in Table 24. Of the serious adverse events, none resulted in death of a subject and all were considered to be unrelated to treatment as reported by the study investigator.

Table 24: Combined Serious Adverse Events for Studies VLC.C.304 and VLC.C.305

Term	TN® Gel	Clindamycin	Tretinoin	Vehicle	Total
N	634	635	635	315	2219
Subjects with Serious AE	3 (0%)	1 (0%)	3 (1%)	0 (0%)	7 (0%)

Source: Module 5 Section 5.3.5.1.1 Page 294 Table 14.3.1.3 and Module 5 Section 5.3.5.1.2 Page 276 Table 14.3.1.1.

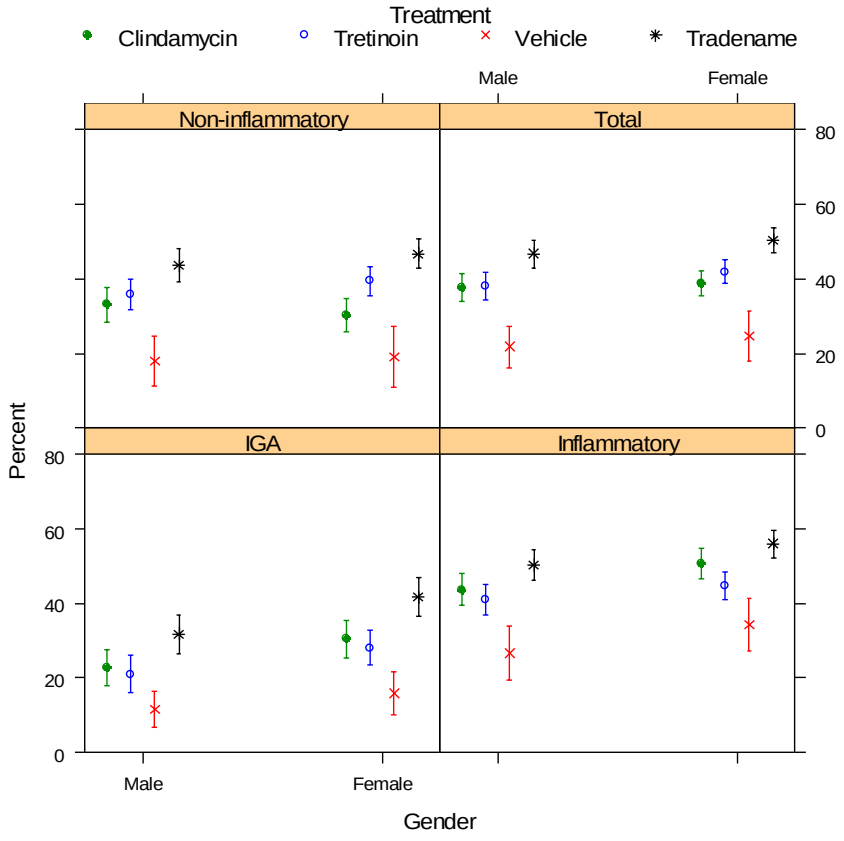
4 Findings in Special/Subgroup Populations

Tabled efficacy results across subgroups for each study are provided in the Appendix Section A.4. In the figures provided below, the data from Studies VLC.C.304 and VLC.C.305 are combined to provide a 'snapshot' of efficacy results across subgroups.

4.1 Gender, Race, and Age

Figure 5 depicts efficacy results across gender. This figure shows that males and females exhibited the same general trend across treatment groups for each endpoint. Overall, TN® Gel seems to be better than its components and vehicle for each of the endpoints across both genders.

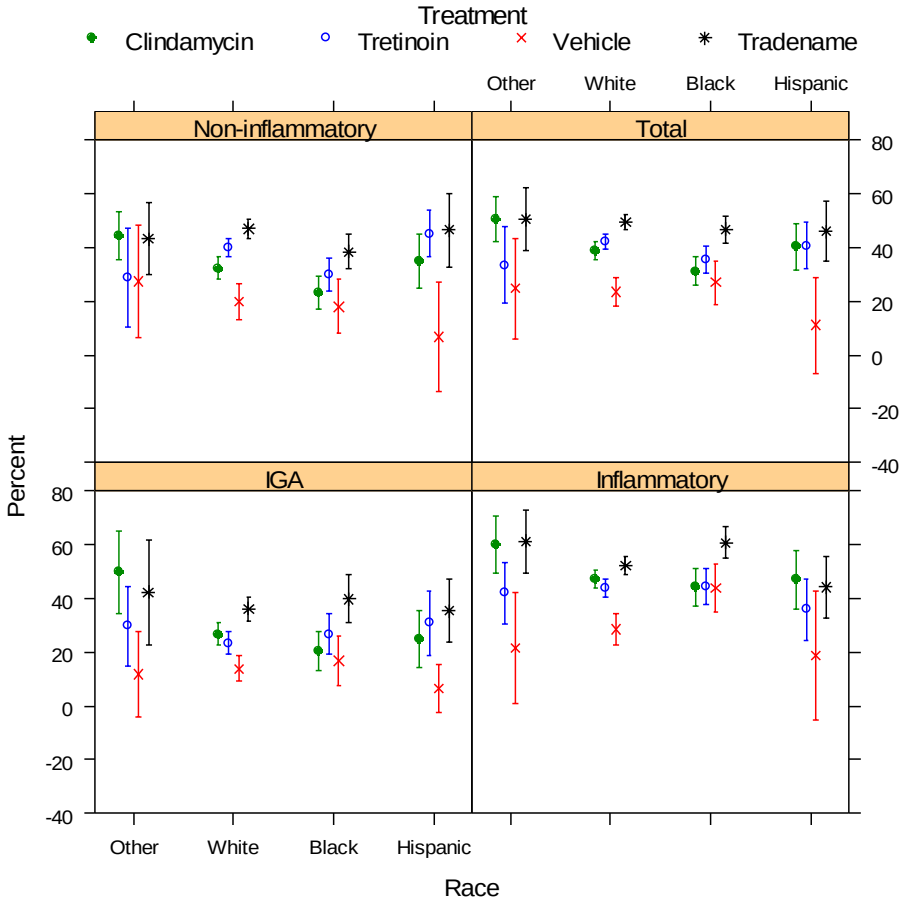
Figure 5: Efficacy Results for each Gender



Race is broken down into four subgroups: Caucasian, African American (denoted as Black in Figure 6), Hispanic and Other. Composition of the Other subgroup consists of Asians and the Missing and these two subgroups comprise approximately only 5% to 6% of subjects (refer to Table A.4.5 and Table A.4.6 in the Appendix Section A.4). Figure 6 depicts the combined study results across subgroups. In general, TN® Gel tends to have higher point estimates (as seen by the points) than its components and vehicle across all

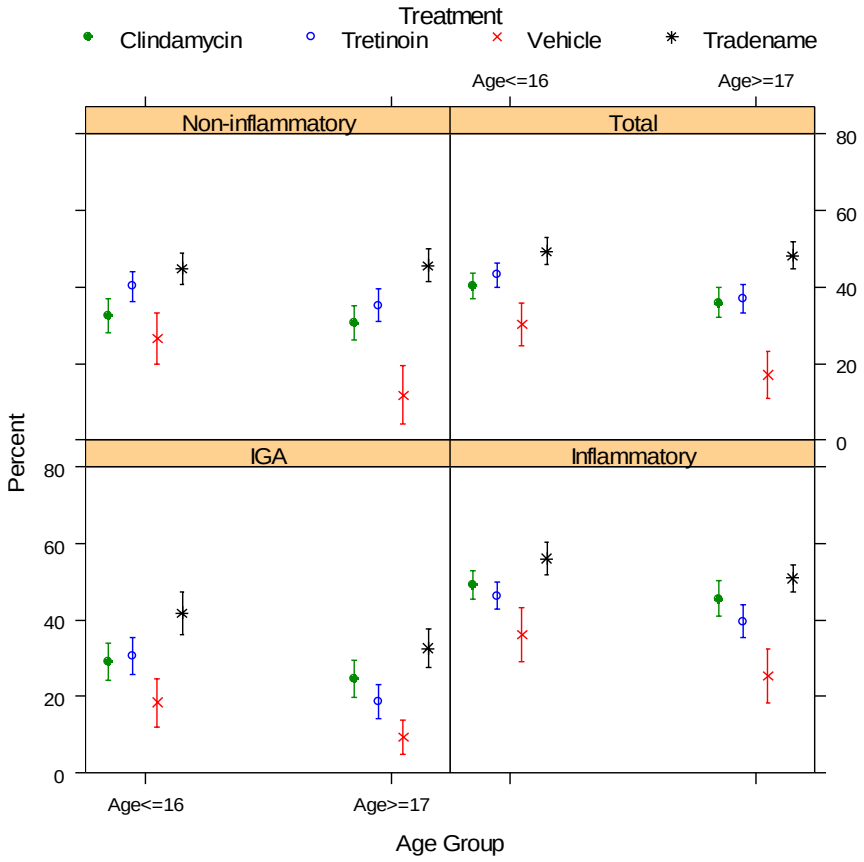
subgroups with the exception of the Other subgroup whose sample size is small. Also the efficacy of the four treatment arms tends to be relatively consistent across all subgroups.

Figure 6: Efficacy Results for each Race



Age is broken down into two categories: 12 ≤ Age ≤ 16 and Age ≥ 17. Dichotomizing Age into these two categories results in roughly 50% representation for each category. Figure 7 on the next page shows the efficacy across the two age subgroups. TN® Gel tends to have higher point estimates than its components and vehicle for all primary endpoints across age subgroups. Overall, regardless of treatment subjects 17 and older had smaller estimates of treatment effects than those subjects aged between 12 and 16 for each endpoint.

Figure 7: Efficacy Results for each Age Category



4.2 Other Special/Subgroup Populations

No other subgroups were analyzed.

5 Summary and Conclusions

5.1 Statistical Issues and Collective Evidence

The sponsor's submission contained two large pivotal Phase 3 trials that established the superiority of TN® Gel to each of its components and vehicle based upon the statistical analysis plan in the protocol which was agreed upon with the Agency. A question about the validity of the interim analysis and the effects of increasing the sample size were examined, but the examination did not detect verifiable reasons to doubt the efficacy conclusions. Overall, the sponsor's analysis plan was agreed upon by the Agency and based upon this analysis the objectives of the study were met.

Both Study VLC.C.304 and Study VLC.C.305 established superiority of TN® Gel to each of its components and its vehicle. The results of each study are reported in Table 25.

Referring to the table, Study VLC.C.304 firmly establishes the superiority of TN® Gel on both the lesion counts and IGA as seen by the small p-values. Study VLC.C.305 was not as convincing as Study VLC.C.304 in establishing the superiority of TN® Gel to its components but still met study objectives. In Study VLC.C.305, TN® Gel was clearly superior to its components and vehicle on the basis of percent reduction in 2 of 3 lesion counts. However the percent of subjects achieving a 'clear' or 'almost 'clear' determination for Clindamycin and Tretinoin were quite high in this study compared to Study VLC.C.304. Thus, for the proportion of successes in IGA, a lower strength of evidence is observed in Study VLC.C.305 than compared to Study VLC.C.304.

Table 25: Efficacy Results Summary: Percent Reduction from Baseline for Lesion Counts and Percent Success according to Investigator Global Assessment (IGA)

	TN® Gel	**Clindamycin**	**Tretinoin**	**Vehicle**
Study VLC.C.304[1]				
Sample Size	309	311	310	153
Total Lesions	46.2%	33.8% (p < .0001)	35.6% (p < .0001)	20.0% (p < .0001)
Non-Inflammatory Lesions	41.5%	25.0% (p < .0001)	32.9% (p = .0002)	14.3% (p < .0001)
Inflammatory Lesions	52.3%	46.4% (p = .0529)	39.3% (p < .0001)	27.2% (p < .0001)
IGA	35%	21% (p = .0002)	19% (p < .0001)	12% (p < .0001)
Study VLC.C.305[2]				
Sample Size	325	324	324	162
Total Lesions	51.1%	42.5% (p = .0001)	44.9% (p = .0034)	26.2% (p < .0001)
Non-Inflammatory Lesions	48.8%	37.8% (p < .0001)	42.7% (p = .0121)	22.5% (p < .0001)
Inflammatory Lesions	54.5%	48.6% (p = .0306)	47.2% (p = .0041)	33.3% (p < .0001)
IGA	39%	32% (p = .0441)	31% (p = .0108)	15% (p < .0001)

[1] Tested at a significance level of α=.04806 due to interim analysis.
[2] Tested at a significance level of α=.05.

The sponsor conducted an interim analysis to determine if sample sizes would be sufficient to achieve study objectives or to stop the trial if all primary endpoints were significant at α=.005 (O'Brien Fleming adjustment). The interim analysis was carried out by an in-house manager and issues about retaining blinding of study findings resulted as the original sample size was decided upon by an executive committee who chose to increase the sample size by approximately 196 subjects. This decision was made despite the reported conclusion that the in-house manager advised the executive committee that no sample size increases were needed to meet study objectives. The Agency requested further information of the sponsor about the conduct and data from the interim. From a review of these materials and by examining data as it accumulated in the trials, no apparent bias in efficacy claims could be found. Refer to Section 3.1.7 for further details.

The safety examination showed the local safety of TN® Gel is quite similar to that of Tretinoin. Both of these drug products seemed to elicit an initial local skin reaction to the

drug product up until about 2 weeks after which the reactions tended to decrease. Both Clindamycin and Vehicle showed little change in local skin reactions and actually moved toward some improvement over baseline measurements. No serious adverse events that could be attributed to treatment were reported by study investigators for any treatment arm.

5.2 Conclusions and Recommendations

TN® Gel is statistically superior to each of its components and its vehicle in the treatment of moderate to severe acne vulgaris. Significance is established on two Phase 3 Studies, VLC.C.304 and VLC.C.305. Study VLC.C.304 showed a clear significance in terms of the percent reduction in 2 of 3 lesion counts and percent achieving success on IGA. Study VLC.C.305 also showed strong significance in the percent reduction in 2 of 3 lesion counts, but with less strong evidence in the percent achieving success on IGA. However, all primary endpoints met the pre-defined, agreed upon α-levels.

The sponsor's proposed clinical labeling section list the following statements about studies VLC.C.304 and VLC.C.305.

(b) (4)

The data do not support that TN® Gel is significantly more effective in reducing all lesion counts, but rather support that TN® Gel is significantly more effective in reducing *at least 2 of 3* lesion counts. In Study VLC.C.304, TN® Gel failed to show a significant reduction in inflammatory lesion counts (p=.0503 when testing at α=.04806).

The combined analysis is not appropriate (b) (4) as this (b) (4) relies on the strong results from Study VLC.C.304 to an extent that it makes otherwise insignificant study findings from Study VLC.C.305 significant. In both the reviewer's and the sponsor's analysis of time to 50% reduction in total lesion counts, the comparison of TN® Gel to Tretinoin failed to reach statistical significance (p=.1009 and p=.0539, respectively). Hence, (b) (4) the combining of the data creates significance.

Overall, from a statistical perspective, the studies established the superiority of TN® Gel to each of its components and vehicle in the treatment of moderate to severe acne vulgaris.

Appendix

A.1 Baseline Demographic Characteristics

Table A.1.1: Demographic Characteristics for Study VLC.C.304

	TN® Gel	Clindamycin	Tretinoin	Vehicle	p-value[1]
Number of Subjects	309	311	310	153	
Age					
Mean (SD)	19.6 (7.8)	19.6 (7.8)	20.1 (7.6)	20.3 (7.6)	.3316
Age Category					
12- 16	157 (50.8%)	148 (47.7%)	138 (44.8%)	66 (43.1%)	.3382
17-65	152 (49.2%)	162 (52.3%)	170 (55.2%)	87 (56.9%)	
>65	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Missing	0 (0%)	1 (0%)	2 (0%)	0 (0%)	
Gender					
Male	147 (47.6%)	140 (45.0%)	122 (39.4%)	72 (47.1%)	.1764
Female	162 (52.4%)	171 (55.0%)	188 (60.6%)	81 (52.9%)	
Race					
African-American	62 (20.1%)	67 (21.5%)	72 (23.2%)	33 (21.6%)	.9987
Asian	8 (2.6%)	9 (2.9%)	4 (2.6%)	8 (2.6%)	
Caucasian	196 (63.4%)	196 (63.0%)	188 (60.7%)	94 (61.4%)	
Hispanic	35 (11.3%)	31 (10.0%)	30 (9.7%)	18 (11.8%)	
Other	8 (2.6%)	9 (2.9%)	11 (3.6%)	4 (2.6%)	

[1] p-values are based upon a Kruskal-Wallis test for continuous variables and a Chi-square test for categorical variables.
Source: Module 5 Section 5.3.5.1.1 Page 61 Table 9.

Table A.1.2: Demographic Characteristics for Study VLC.C.305

	TN® Gel	Clindamycin	Tretinoin	Vehicle	p-value[1]
Number of Subjects	325	324	325	162	
Age					
Mean (SD)	19.0 (6.8)	19.5 (7.1)	20.3 (8.7)	17.8 (6.7)	.0002
Age Category					
12- 16	171 (52.6%)	152 (46.9%)	144 (44.3%)	103 (63.6%)	.0022
17-65	154 (47.4%)	172 (53.1%)	180 (55.4%)	59 (36.4%)	
>65	0 (0%)	0 (0%)	1 (.3%)	0 (0%)	
Missing	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Gender					
Male	149 (45.9%)	157 (48.5%)	144 (44.3%)	92 (56.8%)	.0607
Female	176 (54.1%)	167 (51.5%)	181 (55.7%)	70 (43.2%)	
Race					
African-American	56 (17.2%)	56 (17.3%)	59 (18.2%)	32 (20.0%)	.6295
Asian	3 (.9%)	10 (3.1%)	6 (1.9%)	5 (3.1%)	
Caucasian	229 (70.5%)	210 (64.8%)	221 (68.0%)	108 (66.7%)	
Hispanic	30 (9.2%)	33 (10.2%)	28 (8.62%)	13 (8.0%)	
Other	7 (2.2%)	15 (4.6%)	11 (3.4%)	4 (2.5%)	

[1] p-values are based upon a Kruskal-Wallis test for continuous variables and a Chi-square test for categorical variables.
Source: Module 5 Section 5.3.5.1.1 Page 57 Table 9.

A.2 Full ANOVA Results for Lesion Counts (ITT Population)

Table A.2.1: Full Listing of ANOVA Results for Study VLC.C.304

	TN® Gel	Clindamycin	Tretinoin	Vehicle
Total Lesions				
Mean	46.2%	33.8%	35.6%	20.0%
SD	35.3%	33.2%	30.8%	38.7%
p-value interaction[1]		.5725	.5960	.1923
p-value treatment[2]		<.0001	<.0001	<.0001
p-value normality[3]		<.0001	<.0001	<.0001
p-value transform[4]		<.0001	<.0001	<.0001
p-value transform normality[5]		<.0001	<.0001	<.0001
Non-Inflammatory Lesions				
Mean	41.5%	25.0%	32.9%	14.3%
SD	43.5%	41.3%	36.0%	49.3%
p-value interaction[1]		.6319	.8423	.0383
p-value treatment[2]		<.0001	.0065	<.0001
p-value normality[3]		<.0001	<.0001	<.0001
p-value transform[4]		<.0001	.0002	<.0001
p-value transform normality[5]		<.0001	<.0001	<.0001
Inflammatory Lesions				
Mean	52.3%	46.4%	39.3%	27.2%
SD	36.2%	38.7%	37.4%	44.6%
p-value interaction[1]		.2665	.4721	.6773
p-value treatment[2]		.0503	<.0001	<.0001
p-value normality[3]		<.0001	<.0001	<.0001
p-value transform[4]		.0529	<.0001	<.0001
p-value transform normality[5]		<.0001	<.0001	<.0001

[1] p-value for the interaction term is based upon ANOVA results tested at $\alpha = .10$.
[2] p-value is based upon ANOVA results excluding the interaction term tested at $\alpha = .04806$.
[3] p-value is based upon Shapiro-Wilk test for normality of residuals tested at $\alpha = .05$.
[4] p-value is based upon ANOVA using the rank transform of endpoint tested at $\alpha = .04806$.
[5] p-value is based upon Shapiro-Wilk test for normality of the residuals w/ rank transform tested at $\alpha = .05$.
Source: Module 5 Section 5.3.5.1.1 Pages 71-72 Table 16.

Table A.2.2: Full Efficacy Results for Study VLC.C.305

	Velac	Clindamycin	Tretinoin	Vehicle
Total Lesions				
Mean	51.1%	42.5%	44.9%	26.2%
SD	28.4%	31.7%	31.0%	38.5%
p-value interaction[1]		.6464	.2410	.2574
p-value treatment[2]		.0001	.0034	<.0001
p-value normality[3]		<.0001	<.0001	<.0001
p-value transform[4]		.0002	.0029	<.0001
p-value transform normality[5]		<.0001	<.0001	.0439
Non-Inflammatory Lesions				
Mean	48.8%	37.8%	42.7%	22.5%
SD	31.3%	39.7%	37.1%	45.3%
p-value interaction[1]		.3965	.2028	.4122
p-value treatment[2]		<.0001	.0121	<.0001
p-value normality[3]		<.0001	<.0001	<.0001
p-value transform[4]		<.0001	.0271	<.0001
p-value transform normality[5]		<.0001	<.0001	.0154
Inflammatory Lesions				
Mean	54.5%	48.6%	47.2%	33.3%
SD	35.5%	37.2%	33.9%	46.9%
p-value interaction[1]		.8384	.1999	.3563
p-value treatment[2]		.0306	.0041	<.0001
p-value normality[3]		<.0001	<.0001	<.0001
p-value transform[4]		.0209	.0003	<.0001
p-value transform normality[5]		<.0001	<.0001	.0004

[1] p-value for the interaction term is based upon ANOVA results tested at $\alpha = .10$.
[2] p-value is based upon ANOVA results excluding the interaction term tested at $\alpha = .05$.
[3] p-value is based upon Shapiro-Wilk test for normality of residuals tested at $\alpha = .05$.
[4] p-value is based upon ANOVA using the rank transform of endpoint tested at $\alpha = .05$.
[5] p-value is based upon Shapiro-Wilk test for normality of the residuals w/ rank transform tested at $\alpha = .05$.
Source: Module 5 Section 5.3.5.1.2 Pages 67-68 Table 16.

A.3 Efficacy Figures

Figure A.3.1: Percent Reduction in Total Lesion Counts

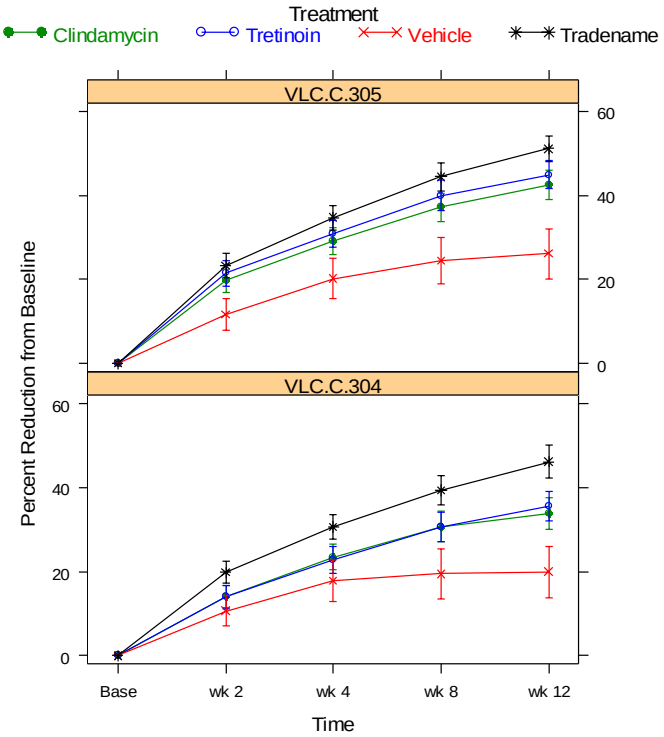


Figure A.3.2: Percent Reduction in Non-Inflammatory Lesion Counts

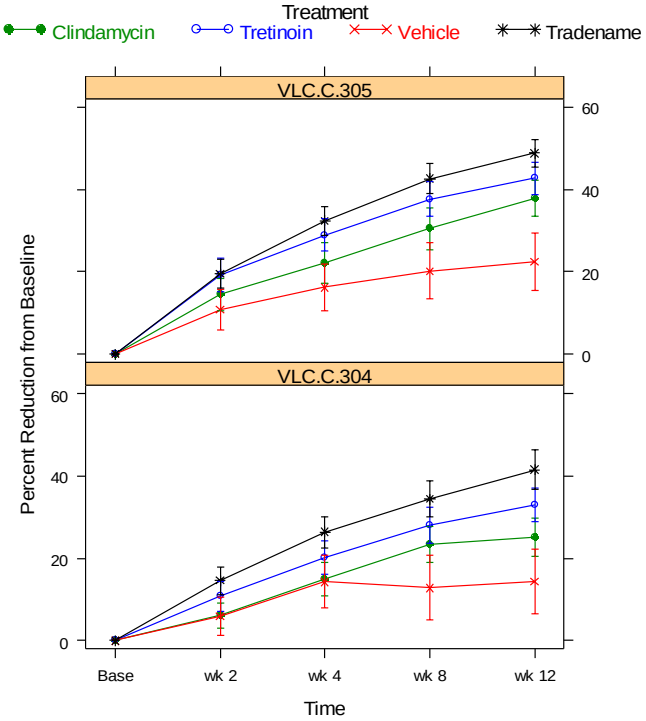


Figure A.3.3: Percent Reduction in Inflammatory Lesion Counts

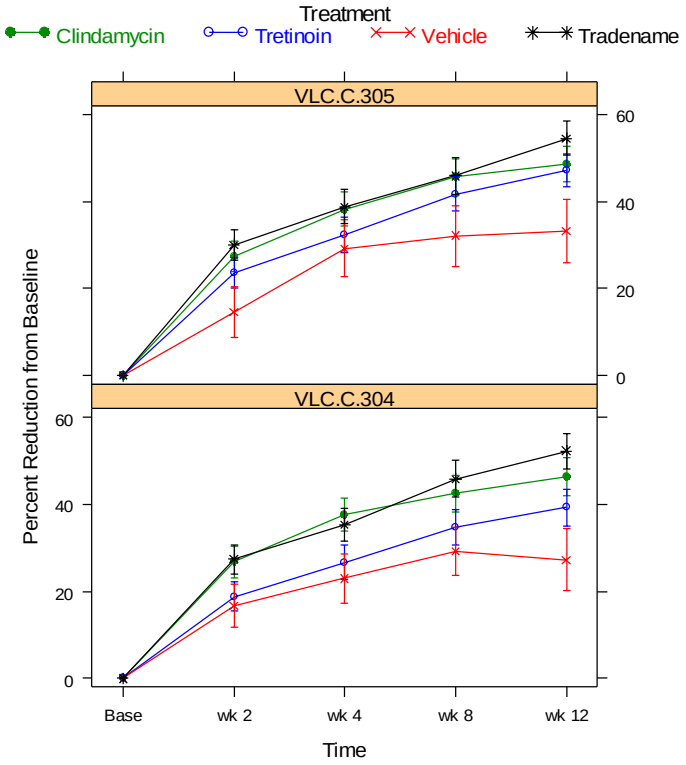
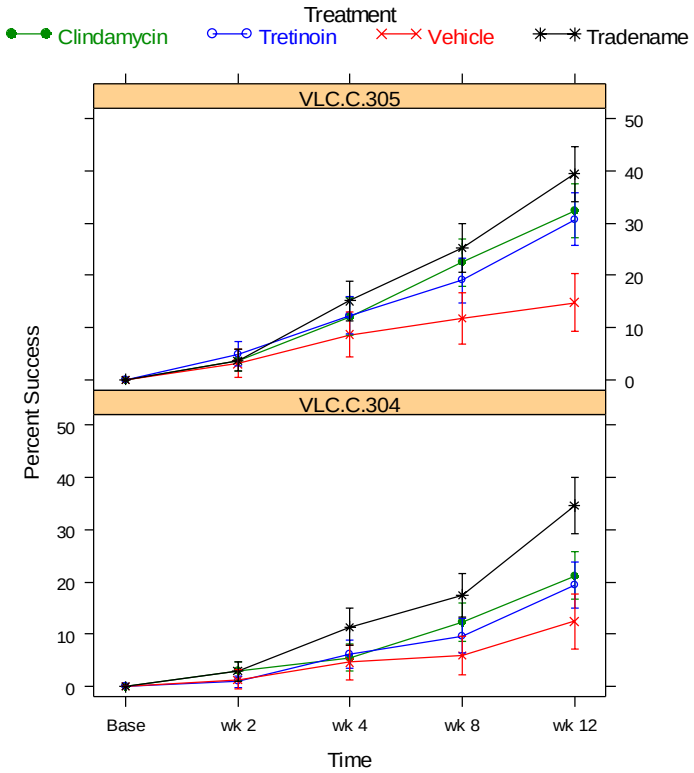


Figure A.3.4: Percent Success for Investigator Global Assessment



A.4 Efficacy Results by Subgroup

Table A.4.1: Efficacy Results by Gender for Study VLC.C.304

	TN® Gel	Clindamycin	Tretinoin	Vehicle
Total Lesions				
Male	43.4% (37.9%)	30.2% (36.1%)	34.0% (31.0%)	20.2% (34.2%)
Female	48.7% (32.7%)	36.8% (30.4%)	36.6% (30.7%)	19.9% (42.5%)
Non-Inflammatory Lesions				
Male	39.6% (46.5%)	22.2% (44.1%)	32.3% (35.1%)	15.8% (43.8%)
Female	43.2% (40.6%)	27.4% (38.8%)	33.3% (36.6%)	13.1% (54.0%)
Inflammatory Lesions				
Male	47.9% (39.7%)	41.8% (40.4%)	37.0% (38.5%)	24.1% (46.4%)
Female	56.2% (32.4%)	50.1% (37.0%)	40.7% (36.8%)	30.0% (43.0%)
Investigator's Global Assessment				
Male	46 (31.3%)	26 (18.6%)	19 (15.6%)	8 (11.1%)
Female	61 (37.7%)	40 (23.4%)	41 (21.8%)	11 (13.6%)

Source: Module 5 section 5.3.5.1.1 Tables 14.2.38-14.2.43 and 14.2.62-14.14.2.63.
Values in parentheses for lesion counts correspond to standard deviations.
Values in parentheses for IGA correspond to percent success.

Table A.4.2: Efficacy Results by Gender for Study VLC.C.305

	TN® Gel	Clindamycin	Tretinoin	Vehicle
Total Lesions				
Male	50.1% (26.1%)	44.2% (28.2%)	41.6% (28.1%)	23.1% (38.2%)
Female	52.0% (30.2%)	41.0% (34.8%)	47.5% (32.9%)	30.2% (38.9%)
Non-Inflammatory Lesions				
Male	47.6% (30.2%)	42.8% (33.2%)	38.8% (35.1%)	19.8% (44.4%)
Female	49.8% (32.3%)	33.2% (44.7%)	45.8% (38.5%)	26.0% (46.5%)
Inflammatory Lesions				
Male	52.7% (33.0%)	45.4% (35.4%)	44.7% (29.8%)	28.6% (49.0%)
Female	56.0% (37.4%)	51.6% (38.7%)	49.1% (36.8%)	39.4% (43.4%)
Investigator's Global Assessment				
Male	48 (37.7%)	42 (26.8%)	37 (25.7%)	11 (12.0%)
Female	80 (45.5%)	63 (37.7%)	63 (34.8%)	13 (18.6%)

Source: Module 5 section 5.3.5.1.2 Tables 14.2.38-14.2.43 and 14.2.62-14.14.2.63.
Values in parentheses for lesion counts correspond to standard deviations.
Values in parentheses for IGA correspond to percent success.

Table A.4.3: Efficacy Results by Age Category for Study VLC.C.304

	TN® Gel	Clindamycin	Tretinoin	Vehicle
Total Lesions				
Age ≤ 16	48.3% (32.6%)	37.2% (30.9%)	39.4% (29.6%)	30.8% (29.6%)
Age ≥ 17	44.2% (37.8%)	30.1% (35.2%)	31.1% (31.7%)	5.8% (44.5%)
Non-Inflammatory Lesions				
Age ≤ 16	42.4% (38.5%)	27.9% (39.7%)	37.6% (35.7%)	27.1% (36.6%)
Age ≥ 17	40.7% (48.0%)	22.0% (42.9%)	27.4% (35.7%)	-2.4% (58.4%)
Inflammatory Lesions				
Age ≤ 16	57.1% (33.6%)	48.0% (35.2%)	40.8% (35.3%)	35.4% (43.8%)
Age ≥ 17	47.6% (38.1%)	44.6% (42.3%)	37.7% (40.0%)	16.4% (43.5%)
Investigator's Global Assessment				
Age ≤ 16	62 (40.8%)	38 (23.5%)	40 (23.5%)	12 (13.8%)
Age ≥ 17	45 (28.7%)	28 (18.8%)	20 (14.4%)	7 (10.6%)

Source: Module 5 section 5.3.5.1.1 Tables 14.2.44-14.2.49 and 14.2.64-14.14.2.65.
Values in parentheses for lesion counts correspond to standard deviations.
Values in parentheses for IGA correspond to percent success.

Table A.4.4: Efficacy Results by Age Category for Study VLC.C.305

	TN® Gel	Clindamycin	Tretinoin	Vehicle
Total Lesions				
Age ≤ 16	51.9% (26.0%)	41.9% (30.6%)	42.7% (31.0%)	24.2% (26.4%)
Age ≥ 17	50.3% (30.9%)	43.1% (32.8%)	46.6% (30.9%)	29.6% (42.1%)
Non-Inflammatory Lesions				
Age ≤ 16	50.0% (28.8%)	38.9% (33.4%)	42.8% (36.7%)	20.6% (43.0%)
Age ≥ 17	47.4% (34.0%)	36.9% (44.7%)	42.6% (37.6%)	25.7% (49.3%)
Inflammatory Lesions				
Age ≤ 16	54.0% (30.1%)	46.7% (39.2%)	41.5% (32.7%)	31.0% (49.2%)
Age ≥ 17	55.1% (40.7%)	50.3% (35.5%)	51.7% (34.2%)	37.2% (42.6%)
Investigator's Global Assessment				
Age ≤ 16	62 (36.3%)	46 (30.3%)	33 (22.9%)	9 (8.7%)
Age ≥ 17	66 (42.9%)	59 (34.3%)	67 (37.0%)	15 (25.4%)

Source: Module 5 section 5.3.5.1.2 Tables 14.2.44-14.2.49 and 14.2.64-14.14.2.65.
Values in parentheses for lesion counts correspond to standard deviations.
Values in parentheses for IGA correspond to percent success.

Table A.4.5: Efficacy Results by Race for Study VLC.C.304

	TN® Gel	Clindamycin	Tretinoin	Vehicle
Total Lesions				
Caucasian	46.5% (33.7%)	33.9% (34.8%)	39.0% (31.6%)	22.5% (36.4%)
African American	42.9% (29.5%)	28.7% (26.2%)	28.5% (25.7%)	21.4% (29.1%)
Hispanic	48.3% (52.4%)	34.6% (36.3%)	26.5% (33.2%)	-4.3% (57.2%)
Other[1]	51.1% (32.5%)	51.2% (28.7%)	41.7% (31.5%)	38.7% (32.5%)
Non-Inflammatory Lesions				
Caucasian	43.1% (39.3%)	25.3% (43.8%)	36.3% (36.6%)	18.0% (48.3%)
African American	32.9% (38.5%)	17.1% (36.0%)	21.4% (34.1%)	11.0% (39.6%)
Hispanic	46.6% (70.1%)	28.9% (37.8%)	33.0% (33.0%)	-11.6% (66.4%)
Other[1]	43.7% (34.5%)	44.0% (31.2%)	41.6% (34.0%)	43.4% (31.8%)
Inflammatory Lesions				
Caucasian	50.5% (36.1%)	46.3% (38.5%)	41.9% (35.9%)	28.2% (37.1%)
African American	56.8% (33.2%)	46.2% (33.3%)	39.8% (34.8%)	36.3% (37.2%)
Hispanic	49.1% (43.1%)	39.0% (51.6%)	18.3% (46.7%)	4.8% (77.6%)
Other[1]	63.2% (31.1%)	61.2% (33.0%)	43.9% (38.5%)	28.7% (47.4%)
Investigator's Global Assessment				
Caucasian	64 (32.3%)	45 (23.0%)	34 (18.1%)	13 (13.8%)
African American	20 (32.2%)	9 (13.4%)	14 (19.4%)	3 (9.1%)
Hispanic	16 (45.7%)	6 (19.4%)	6 (20.0%)	1 (5.5%)
Other[1]	7 (43.8%)	6 (25.3%)	6 (30.0%)	2 (25.0%)

Source: Module 5 section 5.3.5.1.1 Tables 14.2.50-14.2.61 and 14.2.66-14.14.2.69.
Values in parentheses for lesion counts correspond to standard deviations.
Values in parentheses for IGA correspond to percent success.
[1] Other subgroup consists of Asians and Missing

Table A.4.6: Efficacy Results by Race for Study VLC.C.305

	TN® Gel	Clindamycin	Tretinoin	Vehicle
Total Lesions				
Caucasian	52.2% (27.1%)	43.3% (31.4%)	45.2% (28.1%)	24.5% (39.6%)
African American	51.0% (28.2%)	34.3% (32.7%)	44.4% (31.1%)	33.1% (36.6%)
Hispanic	43.4% (37.1%)	46.1% (34.1%)	56.0% (26.7%)	32.6% (31.2%)
Other[1]	49.9% (29.1%)	50.2% (26.7%)	24.0% (56.4%)	12.5% (42.3%)
Non-Inflammatory Lesions				
Caucasian	50.3% (31.2%)	38.7% (41.8%)	43.5% (33.7%)	21.3% (47.4%)
African American	44.8% (29.9%)	29.4% (32.8%)	40.5% (33.2%)	25.2% (42.2%)
Hispanic	46.6% (34.0%)	41.0% (43.3%)	58.4% (28.4%)	32.1% (31.7%)
Other[1]	42.8% (36.1%)	44.8% (29.1%)	13.7% (74.4%)	13.1% (50.8%)
Inflammatory Lesions				
Caucasian	53.9% (33.5%)	48.1% (34.6%)	46.0% (31.7%)	28.8% (49.1%)
African American	65.2% (31.6%)	42.1% (46.2%)	50.2% (42.2%)	51.4% (34.2%)
Hispanic	38.5% (50.2%)	54.6% (35.4%)	54.8% (32.4%)	38.1% (49.4%)
Other[1]	58.2% (30.9%)	59.2% (38.0%)	39.9% (31.7%)	15.2% (41.6%)
Investigator's Global Assessment				
Caucasian	90 (39.3%)	64 (30.5%)	62 (28.1%)	15 (13.9%)
African American	27 (48.2%)	16 (28.6%)	21 (35.6%)	8 (25.0%)
Hispanic	7 (23.3%)	10 (30.3%)	12 (42.9%)	1 (7.7%)
Other[1]	4 (40.0%)	15 (60.0%)	5 (29.4%)	0 (0.0%)

Source: Module 5 section 5.3.5.1.2 Tables 14.2.50-14.2.61 and 14.2.66-14.14.2.69.
Values in parentheses for lesion counts correspond to standard deviations.
Values in parentheses for IGA correspond to percent success.
[1] Other subgroup consists of Asians and Missing

A.5 References

Cui, L., Hung, H.M., and Wang, S., (1999), "Modification of Sample Size in Group Sequential Clinical Trials," *Biometrics*, **55**, 853-857.

Mehta, C.R. and Tsiatis, A.A., (2001), "Flexible Sample Size Considerations Using Information-Based Interim Monitoring," *Drug Information Journal*, **35**, 1095-1112.

Chen, Y.H., DeMets, D.L., and Lan, K.K., (2004), "Increasing the Sample Size when Unblinded Interim Result is Promising," *Statistics in Medicine*, **23**, 1023-1038.

Signatures/Distribution List

Primary Statistical Reviewer: Mat Soukup, Ph.D.
Date: 04/12/2005

Statistical Team Leader: Mohamed Alosh, Ph.D.

cc:
Archival NDA
HFD-540/Wilkin
HFD-540/Luke
HFD-540/Nikhar
HFD-540/Owens
HFD-700/Anello
HFD-725/Dubey
HFD-725/Huque
HFD-725/Lin
HFD-725/Alosh
HFD-725/Soukup
4/25/05

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

/s/

Matt Soukup
4/25/05 08:51:11 AM
BIOMETRICS

Mohamed Alosh
4/26/05 09:10:14 AM
BIOMETRICS
Concur with review

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

STATISTICAL REVIEW AND EVALUATION

CARCINOGENICITY STUDY

NDA Number:	50,803 / Type 3S
Drug Name:	Velac® Gel (Clindamycin 1% / tretinoin 0.025%) 5% cream
Indication(s):	Topical Treatment of Acne Vulgaris
Applicant:	Connectics Corporation
Date(s):	Received 10/29/03, user fee (10 months) 4/29/04
Review Priority:	Standard

Biometrics Division:	Division 3, HFD-725
Statistical Reviewer:	Steve Thomson, HFD-725
Concurring Reviewers:	Team Leader: Mohamed Alosh, Ph. D., HFD-725
Medical Division:	Dermatological and Dental Drug Products, HFD-540
Toxicologist:	Reviewer: Jill Merrill, Ph.D., HFD-540 Team Leader: Paul Brown, Ph.D., HFD-540
Project Manager:	Margo Owens, HFD-540

Keywords: Bayesian analysis, Carcinogenicity, Cox regression, Kaplan-Meier product limit, Mixed models, Survival analysis, Trend test

-

Table of Contents

1. EXECUTIVE SUMMARY 3

1.1. CONCLUSIONS AND RECOMMENDATIONS 3
1.2. BRIEF OVERVIEW OF STUDY 4
1.3. STATISTICAL ISSUES AND FINDINGS 5
1.3.1. Statistical Issues 5
1.3.2. Statistical Findings 7

2. INTRODUCTION 9

2.1. OVERVIEW 9
2.2. DATA SOURCES 9

3. STATISTICAL EVALUATION 10

3.1. EVALUATION OF EFFICACY 10
3.2. EVALUATION OF SAFETY 10
3.2.1. Sponsor's Results and Conclusions 11
3.2.2. FDA Reviewer's Results 13

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS 18

5 SUMMARY AND CONCLUSIONS 19

5.1. STATISTICAL ISSUES AND COLLECTIVE EVIDENCE 19
5.2. CONCLUSIONS AND RECOMMENDATIONS 22

APPENDICES: 23

APPENDIX 1. OBSERVED INDIVIDUAL SKIN LESIONS AT END OF STUDY 23
APPENDIX 2. SURVIVAL ANALYSIS 25
APPENDIX 3. WEEKLY SOA TUMOR INCIDENCE 26
APPENDIX 4. TIME TO APPEARANCE OF FIRST TUMOR 29
APPENDIX 5. TIME TO APPEARANCE OF MAXIMAL NUMBER OF TUMORS 31
APPENDIX 6. DUNSON MODEL 33
APPENDIX 7. PRELIMINARY BAYESIAN ESTIMATES OF THE DUNSON MODEL 34
APPENDIX 8. REFERENCES 35

SIGNATURES/DISTRIBUTION LIST 36

1. EXECUTIVE SUMMARY

This submission was intended to assess whether daily dermal administration of Velac Gel in four dose groups (vehicle only, and 40, 120, and 200 mg/kg/day) for 26 weeks increased the incidence of dermal tumors (papillomas) at the site of application in hemizygous Tg.AC transgenic mice.

1.1. Conclusions and Recommendations

The study involved three Velac Gel treatment groups (40, 120, and 200 mg/kg/day nominal dose applied daily) and three control groups (Velac vehicle gel, untreated, and active/positive). As is typical of these studies, the active/positive control group was treated with a known carcinogen to demonstrate the propensity of the mice to develop tumors. The focus of the analyses was on the four Velac Gel dose groups (vehicle, i.e., dose = zero, and 40, 120, and 200 mg/kg/day). For both genders, in each of the four dose groups, tumor incidence was statistically significantly higher than in the untreated control group. In addition, for both genders, there was an apparent increasing response to dose in the four dose groups in terms of time to tumor development, numbers of animals with tumors, and number of tumors per animal However, actual tumor incidence and burden in the four dose groups was far less than observed in the active/positive control group.

The following discussion provides an overview of the analyses used to assess the tumorigenicity of Velac Gel. This analysis included:
1. Comparison of the four dose groups at the end of the study to the untreated control using Fisher exact tests
2. Tests of trend in tumor incidence and mean number of tumors at the end of the study over the four dose groups
3. Comparisons over the four dose groups of the time to first tumor and time to maximum number of tumors
4. A model to simultaneously assess both incidence and number of tumors over time in the four dose groups

There were site of application (SOA) tumors in the four Velac Gel dose groups and the active/positive control group. However, there were no SOA tumors in the untreated control group. Fisher exact tests of tumor incidence comparing each of the four dose groups gave statistically significant differences with the untreated control groups (all $p \leq 0.0236$, and most much less). For both genders, there was an apparent dose effect in that as dose of the comparator increased, the statistical significance of the exact test increased dramatically (i.e., p-values quickly decreased).

Tests of trend in dose generally confirmed these dose effects. For the final SOA tumor count, the Cochran-Armitage test of trend in increasing dose for the number of all SOA papillomas was statistically significant for both genders (males $p \leq 0.0173$, females $p < 0.0001$).

A possible model for the tumor count per animal at risk is a Poisson regression model, with a log link. With such a model, the effect of increasing dose from Velac vehicle to 200mg/kg/day in the four Velac dose groups was statistically significant in both genders (males $p \leq 0.0001$, females $p \leq 0.0011$). Tests of nonlinearity in trend in SOA tumors were highly non-significant (p-value at least 0.58 for both genders). This suggests that group differences in tumor count were mainly associated with linear differences in dose. Somewhat different results were found when restricting the analysis to animals with tumors. For female mice the likelihood chi-square test of the linear effect of dose was not statistically significant ($p \leq 0.5217$). For male mice similar tests were statistically significant ($p \leq 0.0065$). These discrepancies may be explained by the much smaller sample sizes when restricting the analysis to the mice with tumors.

Tests of linear trend for time to first SOA tumor using the Cox proportional hazards model (and thus equivalent to a partition of the logrank test) were barely statistically significant in males ($p \leq 0.0431$) and quite significant in females ($p \leq 0.0002$). For both time to event variables and both genders the corresponding tests of nonlinearity in trend were statistically quite non-significant (all p-values at least 0.62). So for both tests and both genders, group differences were essentially due to differences in linear dose.

Dunson, et al. (2000), proposed a Markov generalized linear mixed model combining time to first SOA tumor and the subsequent number of SOA tumors. Using this model, for both genders there were statistically significant dose related trends in papilloma initiation ($p < 0.0001$ for both genders). In males there was evidence of a dose related trend in the development of further tumors after tumor initiation ($p \leq 0.0497$). However, there was no evidence a dose related effect after initiation on the development of further tumors in females ($p < 0.8990$). Specifically, there seemed to be a statistically significant dose related effect on tumor initiation, but only males showed a statistically significant dose related difference after the initiation of the first tumors.

1.2. Brief Overview of Study

Animals were randomized by body weight into five groups of 20 animals per gender and one group of 15 animals per gender. Three of these groups were treated with the test article, clindamycin phosphate in Velac gel vehicle once daily. Nominal doses were 40, 120, and 200 mg/kg/day. The vehicle control group was treated once daily with Velac Gel vehicle. These define the four Velac treatment groups. To verify the sensitivity of the animals, the positive control was treated with a known carcinogenic, 50 µg 12-O-tetradeconyl-phorbol 13-acetate (TPA), in Velac Gel vehicle, three times a week. An untreated control group was shaved, but not treated. Animals in all groups except the active control were treated for 26-27 weeks and then sacrificed. Dosing was initiated on December 3, 2003, and the study was completed on June 4, 2004.

1.3. Statistical Issues and Findings

1.3.1. Statistical Issues

According to Dunson, et al. (2000), the current standard for statistical analysis of skin papillomas separately tests for differences between the experimental group and the control group with respect to the following endpoints (in a slightly different order):
(1) percentage (or count) of animals with tumors, presumably at the end of the study
(2) average number of tumors per animal at risk
(3) average number of tumors per tumor-bearing animal
(4) average latency time to appearance of the first skin tumor
(5) average latency time to the appearance of the maximal number of tumors

In discussion with the FDA toxicologist it was noted that the Executive Carcinogenicity Assessment Committee (Exec CAC) tends to prefer endpoint (1), with support from endpoints (2) or (3). However, other endpoints seem to be recommended by the National Center for Toxicological Research. Results for all five endpoints are reported below. The Sponsor proposed analyses for endpoint (1), were based on Fisher exact tests of incidence comparing each level of treatment to one of the controls, either vehicle or untreated, and for endpoints (2) and (3), based on ANOVAs of the number of tumors, followed by Dunnett's tests. For both the number of tumors per animal at risk and the number of tumors per animal with tumors, tumor counts were skewed to the right, resembling an exponential or gamma distribution more than a normal distribution. Under such circumstances, for robust results ANOVA may require more than the 80 observations present in the four Velac dose groups. For this reason this reviewer had doubts about the appropriateness of the Sponsor's analysis.

Dunson (2000) proposed an alternative analysis using a mixed effects Poisson model for the increase over time in the maximum number of papillomas. In addition to this Dunson model for both incidence and tumor count, the endpoints cited above were analyzed as follows:

1. Analysis of the count of animals with tumors at the end of the study:

Fisher exact tests were used to compare the number of animals with tumors in each dose group to the untreated controls, primarily to determine whether or not there was any evidence of site of application (SOA) tumorigenicity. Given such evidence, interest would be focused on the effect of dose. In such circumstances a Cochran-Armitage test of trend over the vehicle group (dose=0), and the 40, 120, and 200 mg/kg/day groups would seem to be more appropriate and more powerful than the pair-wise Fisher exact tests comparing dose to vehicle, as proposed by the sponsor. Note that the relatively small sample size should limit possible overpowering of the tests.

2. Analysis of the number of tumors at the end of the study per animal at risk:

The distribution of the mean number of SOA tumors at the end of the study was skewed to the right, with many animals having no SOA papillomas and thus having tumor counts scored as zeros. This distribution suggests a Poisson regression model, possibly a zero-inflated model. Alternatively, a model that explicitly limits the dependent variable may be appropriate, as with the Tobit models used here. Recall that this is a normal linear regression model with dose as a factor, but response censored at zero. For both models, the effect of linear dose and nonlinearity in dose was tested. Overall, both the Poisson regression model and the Tobit regression model gave consistent results.

3. Analysis of the number of tumors at the end of the study per animal with tumors:

As in the analysis of animals at risk, the distribution of the number of SOA tumors was skewed, and a Poisson regression model seemed to fit, though less well than for the number of tumors per animal at risk.

4. & 5. Analysis of time to appearance of the first SOA skin tumor and time to appearance of the maximum number of tumors.

Time to event curves for each of these endpoints were estimated using Kaplan-Meier product limit techniques in each of the four Velac Gel dose groups. Overall homogeneity of these time to event curves was tested using logrank and Wilcoxon tests. The logrank test can be interpreted as a Cox proportional hazards regression test of linear trend pooled with a two degrees of freedom test of nonlinearity in trend in the four doses. The latter test was used to test for linear trend in dose for both of these time to event endpoints.

This reviewer also considered other issues, discussed below:

1. Almost all animals survived to sacrifice at the end of the study. However, for completeness, survival proportions were compared using Kaplan-Meier product limit techniques.

2. The untreated control group had no SOA tumors. Therefore, with the possible exception of incidence, tests comparing the untreated control to the four dose groups would be superfluous, and were not done.

3. A number of different analyses are presented in this review. Clearly, the results of a number of these will be correlated. If we were primarily interested in controlling overall type I error, i.e., the error involved in rejecting the true null hypothesis of no treatment differences, a correction for multiplicity would be needed. This approach seems to be suggested by Dunson et al. (2000). However, this reviewer considers that type II error, the error involved in not rejecting a false null hypothesis when there are differences, is at least as important. However, such type II

error is not generally controlled in standard frequentist formulations of hypothesis testing. Note that Haseman's rules for multiplicity, used for standard rodent carcinogenicity analyses, are not applicable in this situation. An appropriate approach to the multiplicity problems involved in these analyses is not clear and should be considered as a matter for future development. When considering the analyses presented, it should be noted that, due to the multiple comparisons, family-wise type I error is not strictly controlled.

4. Another issue is the exact definition of tumor burden. The Sponsor defined tumor burden as the sum of both active and disappeared tumors. At the end of the study, tumor was redefined to include latent tumors. This definition was followed in most analyses here. However, the Dunson definition was based on active tumors only.

1.3.2. Statistical Findings

1. Analysis of the count of animals with tumors at the end of the study:

There were no site of application (SOA) tumors in the untreated group. Pairwise Fisher exact tests of tumor incidence comparing each of the four Velac dose groups, the vehicle (dose=0), and the 40, 120, and 200 mg/kg/day treatment groups to the untreated control group showed statistically significant differences (for both genders, all $p \leq 0.02$, and usually much less). The Cochran-Armitage test of trend in the number of animals with SOA papillomas (active, disappearing, and latent) in the four Velac gel dose groups was statistically significant for both genders (two-sided test, males: $p \leq 0.0173$, females $p: < 0.0001$).

2. Analysis of the number of tumors at the end of the study per animal at risk:

When modeling mean tumor count per animal at risk as a function of dose in the four Velac dose groups using a Poisson model, the linear effect of dose was statistically significant ($p < 0.0001$ both genders). Conditional on this linearity, tests of non-linearity in dose were not statistically significant (males: $p \leq 0.5845$, females $p \leq 0.7699$), suggesting treatment differences were largely due to increasing dose. Results using a Tobit regression model were similar.

3. Analysis of the number of tumors at the end of the study per animal with tumors:

There is also interest in modeling tumor count per animal with tumors. Using a Poisson regression model for this restricted subset of mice, the linear effect of dose, Velac Gel vehicle, and the 40, 120, and 200 mg/kg/day treatment groups, was statistically significant in males ($p \leq 0.0071$), and not significant for females ($p \leq 0.5312$). Further tests of non-linearity in dose were not statistically significant (males: $p \leq 0.9719$, females: $p \leq 0.8236$). However, in the four Velac treatment groups, only 39 female mice and 52 male mice displayed tumors. With such small sample sizes, the maximum likelihood techniques used may not be appropriate, and tests based on these estimates may not have much power to detect differences.

4. Analysis of time to appearance of the first SOA skin tumor:

The overall tests of differences among the four treatment groups in the Kaplan-Meier product limit curves for the time to first tumor were statistically non-significant in males (logrank $p \leq 0.2088$ and Wilcoxon $p \leq 0.2844$) and statistically significant in females (logrank $p \leq 0.0010$ and Wilcoxon $p \leq 0.0019$). More focused tests of linear trend over the four Velac doses (including vehicle) using the Cox proportional hazards model were barely statistically significant in males ($p \leq 0.0431$) and quite statistically significant in females ($p \leq 0.0002$). The two degree of freedom tests of nonlinearity in trend were statistically quite non-significant ($p \leq 0.9561$ and $p \leq 0.6154$ for each gender, respectively).

5. Analysis of time to appearance of the maximum number of tumors:

The overall test of differences in the product limit curves of time to maximum tumor count was statistically non-significant in males (logrank $p \leq 0.1140$ and Wilcoxon $p \leq 0.1148$) and statistically significant in females (logrank $p \leq 0.0070$ and Wilcoxon $p \leq 0.0013$). The more focused tests of linear trend using the Cox proportional hazards model were statistically significant in females ($p \leq 0.0193$) and non-significant in males ($p \leq 0.1680$). For males this is largely due to overlapping time to event curves in the non-vehicle groups. For females the two degree of freedom test of nonlinearity in trend over dose was statistically non-significant ($p \leq 0.6902$). Thus, there was no evidence of a strong evidence of a dose related trend in males, but some evidence in females.

6. Dunson model:

Dunson, et al. (2000), proposed a Markov generalized linear mixed model combining time to first tumor and the subsequent number of tumors. For both genders there were statistically significant dose related trends in papilloma initiation ($p < 0.0001$ for both genders). In male mice there was evidence of a dose related trend in the development of further tumors after tumor initiation ($p \leq 0.0497$). However, there was no strong evidence for the development of further tumors in females ($p < 0.8990$). Specifically, there seems to be a statistically significant dose related effect on tumor initiation, but only males show a statistically significant dose related difference in tumor development after the initiation of the first tumors.

7. Testing treatment differences in survival:

Over the four Velac dose groups there were no strong dose related trends in survival. Tests of differences in survival were statistically non-significant (Males: logrank $p \leq 0.6494$, Wilcoxon $p \leq 0.6611$, Females: logrank $p \leq 0.1131$, Wilcoxon $p \leq 0.1008$).

2. INTRODUCTION

2.1. Overview

The purpose of this study was to assess whether daily dermal administration of Velac Gel in four dose groups (vehicle only, and 40, 120, and 200 mg/kg/day) for 26 weeks increased the incidence of dermal tumors at the site of application in hemizygous Tg.AC transgenic mice. Dosing was initiated on December 3, 2003, and the study was completed on June 4, 2004.

For each gender, three groups of 20 Tg.AC transgenic mice were treated daily with one of three formulations of the test article, clindamycin phosphate foam in Velac Gel vehicle. Nominal clindamycin doses were 40, 120, and 200 mg/kg/day (Groups 4-6). Another similar group of 20 animals/gender were treated daily with Velac vehicle, and served as the vehicle control group (Group 1). Most analyses were restricted to these four groups of animals. Additionally there was a similar group of 20 animals/gender that were shaved but untreated (Group 2), and a smaller group of 15 animals/gender (Group 3) that were administered a known carcinogen, 50 µg 12-O-tetradeconyl-phorbol 13-acetate (TPA), in Velac Gel vehicle, three times weekly to verify that the animals did respond to a tumor promoter. The four Velac dose groups (vehicle, 40, 120, and 200 mg/kg/day) were treated for 26 consecutive weeks with daily application on a defined body surface area of approximately 6 cm^2.

Thus, some 115 male mice and 115 female mice were allocated to treatment groups as summarized in Table 1 below:

Table 1. Description of Design for 28-week Dermal Carcinogenicity Study in Tg.AC Mice`

Group No.	Treatment	# Males	# Females
1	Vehicle Control	20	20
2	Untreated Control	20	20
3	Positive Control	15	15
4	1% (40 mg/kg/day)	20	20
5	3% (120 mg/kg/day)	20	20
6	5% (200 mg/kg/day)	20	20
Total		115	115

2.2. Data Sources

Initially, the Sponsor sent data sets sufficient for the standard analyses (1)-(5) listed in 1.3.1 above. Because the data sets originally provided were not sufficient to analyze the Dunson model for pooled incidence and tumor count, the Sponsor was requested to send a data set providing weekly tumor counts and an indication of the weekly animal status. However, the data set subsequently provided was defined so that animals who had no papillomas during the study were deleted from the data set, rather than being present with zero tumor counts. With

some programming, this and several other smaller deficiencies in the data set were corrected by this reviewer and used in the analyses presented here.

The data sets provided were limited to the four Velac Gel dose groups (vehicle, 40, 120, and 200 mg/kg/day). These were the groups required for most analyses. The other treatment groups were not used in most of the statistical analyses presented here. Note that when the papilloma count reached 20 in the active control group, animals were sacrificed; consequently, final tumor counts in that group are actually depressed. The active control group is used only to demonstrate sensitivity of the animal model, and was ignored in the remaining analysis, as was the no treatment group.

Data for the carcinogenicity study were provided in the FDA electronic data room with 8 SAS transport data sets in the directory:
\\Cdsesub1\n50803\N_000\2004-08-23\pharmtox\datasets\aa81ew.7d8t.btl

Data for the weekly tumor counts was provided in the FDA electronic data room as a SAS transport file in the directory:
\\Cdsesub1\n50803\N_000\2004-10-15\pharmtox\datasets\aa81ew.7d8t.btl

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

NA

3.2. Evaluation of Safety

Study AA81EW.7D8T.BTL: 26-Week Dermal Carcinogenicity Study in Tg.AC Mice

According to the Sponsor: "Three groups of 20 Tg.AC transgenic mice/sex (Groups 4, 5, and 6) were treated daily with 1 of 3 formulations of the test article, clindamycin phosphate in Velac Gel vehicle (1%, 3%, and 5% concentrations, respectively). Nominal clindamycin doses were 40, 120, and 200 mg/kg/day, respectively. A group consisting of the same number and strain of animals (Group 1) was treated daily with Velac Gel vehicle and served as the vehicle control group. Additionally, a group consisting of the same number and strain of animals (Group 2) served as a shaved but untreated control group. The positive control group (Group 3, 15 mice/sex) was administered 50 µg tetradeconylphorbol 13-acetate (TPA) in Velac Gel vehicle three times per week (Monday, Wednesday, and Friday). The positive control served as a qualitative and quantitative indicator of the test system's response to a known tumor promoter. The vehicle control and clindamycin treated animals were dosed via daily dermal application for 26 consecutive weeks at a fixed volume of 0.1mL over a defined body surface area of approximately 6 cm^2." (page 8 of report) Animals were approximately 7-8 weeks old at the start

of dosing. The study was initiated on November 19, 2003, with dosing initiated on December 3, 2003. The study was completed on June 4, 2004.

Group 3 animals were sacrificed when they displayed 20 or more tumors. Male mice in Group 1 were sacrificed at Week 26. The other mice were sacrificed at Week 27.

The site of each observed or latent tumor was recorded at each weekly evaluation. The Sponsor reports that skin papillomas were not counted as actual papillomas until they had been observed for three consecutive weeks. Tumor incidence was defined as the number of animals with tumor. Latent tumors and tumors that later disappeared were both accumulated separately. Prior to the end of the study, the Sponsor's definition of tumor burden was the total of these actual site of application (SOA) tumors and the site of application tumors that later disappeared (DSOA). The definition of the tumor burden at the end of the study also includes the latent papillomas (LSOA). Non site of application (NSOA) tumors were accumulated similarly, but were not included in the definition of tumor burden. Note that the definition of tumor burden given in Dunson et al. (2000) does not include the disappeared tumors and apparently not the latent tumors.

3.2.1. Sponsor's Results and Conclusions

Tables 2 and 3 below summarize the Sponsor's statistical analyses of the final tumor incidence and burden in male and female Tg.AC mice, respectively. Both final site of application (SOA), non-site of application (NSOA), and total papilloma incidence (both SOA and NSOA) were tabulated by the Sponsor. The SOA mean number of tumors per papilloma bearing animal and per animal at risk are also provided. For each treatment group, the Sponsor tested tumor incidence against both the Velac Gel vehicle and the untreated control using Fisher's exact test on the resulting 2x2 subtable. The SOA comparisons of tumor incidence (i.e., count of the animals with tumors) that showed statistically significant differences from a control group are denoted by the superscripts 1 and 2 below. When the comparison of incidence between a treatment group and the Velac Gel vehicle is statistically significant, the treatment group cell is denoted by a superscript 1. When the comparison of incidence between a treatment group and the no treatment group is statistically significant, that treatment group cell is denoted by a superscript 2.

Table 2. Sponsor's Summary of Male Tumor Incidence and Burden

Group No.	Treatment	Incidence			SOA Burden	
		SOA	NSOA	Both	Per papil-loma animal	Per animal at risk
1	Vehicle Control	9/20[2] 45%	5/20 25%	12/20 60%	1.8	0.8
2	Untreated Control	0/18 0%	4/18 22%	4/18 22%	0.0	0.0
3	Positive Control	NA	NA	NA	NA	NA
4	1% (40 mg/kg/day)	12/20[2] 60%	5/20 25%	15/20 75%	2.2	1.3
5	3% (120 mg/kg/day)	15/20[2] 75%	6/20 30%	16/20 80%	2.7	2.0[4]
6	5% (200 mg/kg/day)	16/20[1,2] 80%	4/20 20%	16/20 80%	3.4	2.8[3,4]

Significant Results from:
[1] - Fisher's Exact test compared to vehicle control.
[2] - Fisher's Exact test compared to untreated control.
[3] - ANOVA and Dunnett's t-test compared to vehicle control group.
[4] - ANOVA and Dunnett's t-test compared to untreated control group.

Thus, in terms of incidence, for male mice (Table 2) there were statistically significant differences between the untreated control and each of the four Velac Gel treatment groups. However, only the high dose group (200 mg/kg/day) was statistically significantly worse than the vehicle.

Furthermore, the protocol states that if there were statistically significant differences in incidence, then analysis of variance (ANOVA) tests were to be performed on group mean counts of tumors, both per tumor-bearing animal and the number of effective animals at risk. The protocol states that if the ANOVA results in a statistically significant F statistic ($p \leq 0.05$) for treatment differences, then Dunnett's test was to be used to test for pair-wise comparisons versus the Velac Gel vehicle and the untreated control. Similar to the results for tumor incidence, in Tables 2 (above) and 3 (below), statistically significant results found by the Sponsor are denoted by the superscript 3, for a comparison to vehicle, and superscript 4, for a comparison to the untreated control. For males, only differences between the high dose group (Group 6 - 200mg/kg/day) and the vehicle control among all animals at risk were statistically significant, though both groups 5 & 6 (120 & 200 mg/kg/day) were significantly worse than the untreated control. As discussed in section 1.3.1 and 5.1, this reviewer does not agree that ANOVA is appropriate for these analyses.

Table 3. Sponsor's Summary of Female Tumor Incidence and Burden

Group No.	Treatment	Incidence			SOA Burden	
		SOA	NSOA	Both	Per papil-loma animal	Per animal at risk
1	Vehicle Control	5/20[2] 25%	3/20 15%	6/20 30%	2.0	0.5
2	Untreated Control	0/16 0%	6/16 38%	6/16 38%	0.0	0.0
3	Positive Control	NA	NA	NA	NA	NA
4	1% (40 mg/kg/day)	6/20[2] 30%	2/20 10%	7/20 35%	1.7	0.5
5	3% (120 mg/kg/day)	12/20[1,2] 60%	2/20 10%	12/20 60%	1.8	1.1[4]
6	5% (200 mg/kg/day)	16/20[1,2] 80%	5/20 25%	16/20 80%	2.2	1.8[3,4]

Significant Results from:
[1] - Fisher's Exact test compared to vehicle control.
[2] - Fisher's Exact test compared to untreated control.
[3] - ANOVA and Dunnett's t-test compared to vehicle control group.
[4] - ANOVA and Dunnett's t-test compared to untreated control group.

Results for females (Table 3) were similar to the results for males. Thus, in terms of incidence, there were statistically significant differences between the untreated control and each treatment group. For tumor burden, differences between the high dose group (200 mg/kg/day) and the vehicle control among all animals at risk were statistically significant, though both groups 5 and 6 (120 and 200 mg/kg/day) were statistically significantly worse than the untreated control. Results for females virtually duplicated those for males, except in female mice that the Fisher exact test comparing Groups 1 (vehicle) and 5 (120 mg/kg/day) was also statistically significant.

The Sponsor reported that the analysis for first appearance of papillomas showed no statistically significant differences between doses, but provided no details of the analysis. This conclusion is not consistent with the FDA analysis, as described below.

3.2.2. FDA Reviewer's Results

Results for survival of the animals are presented first, followed by analyses of the following endpoints assessing papilloma incidence and burden:
(1) percentage (or count) of animals with tumors, presumably at the end of the study
(2) average number of tumors per animal at risk
(3) average number of tumors per tumor-bearing animal
(4) average latency time to appearance of the first skin tumor
(5) average latency time to the appearance of the maximal number of tumors

Finally, results from the Dunson, et al. (2000) model for joint analysis of tumor incidence and tumor burden over time are summarized.

3.2.2.1 Survival analysis:

Tables 4 (Males) and 5 (Females) below summarize overall mortality for the different dosages (Groups 1-6).

Table 4. Overall Survival: Males

Group / Disposition	1 Vehicle Control	2 Untreated Control	3 Positive Control	4 40 mg/ kg/day	5 120 mg/ kg/day	6 200 mg/ kg/day
Moribund Sacrifice	1/20	1/20	0/15	0/20	3/20	1/20
Scheduled Sacrifice	0/20	0/20	12/15	0/20	0/20	0/20
Found Dead	0/20	1/20	1/15	1/20	0/20	1/20
Terminal Sacrifice	19/20	18/20	2/15	19/20	15/20	18/20

Note that there was no evidence of a dose related trend in survival in Table 4. Kaplan-Meier plots of the survival distributions among the four Velac treatment groups (vehicle, 40 mg/kg/day, 120 mg/kg/day, and 200 mg/kg/day) are given in Appendix 2, and also show no strong dose related trends in survival. Tests of treatment differences in survival among the four Velac Groups were statistically non-significant (logrank $p \leq 0.6494$, Wilcoxon $p \leq 0.6611$).

Table 5. Overall Survival: Females

Group / Disposition	1 Vehicle Control	2 Untreated Control	3 Positive Control	4 40 mg/ kg/day	5 120 mg/ kg/day	6 200 mg/ kg/day
Moribund Sacrifice	0/20	1/20	0/15	1/20	1/20	1/20
Scheduled Sacrifice	0/20	0/20	12/15	0/20	0/20	0/20
Found Dead	1/20	3/20	0/15	0/20	4/20	1/20
Terminal Sacrifice	19/20	16/20	3/15	19/20	15/20	18/20

Again, there was no evidence of a dose related trend in survival either in the table above or the Kaplan-Meier plots in Appendix 2. Tests of homogeneity of survival among the four Velac treatment groups did not show statistically significant treatment differences (logrank $p \leq 0.1131$, Wilcoxon $p \leq 0.1008$).

So, for both genders there is no particularly strong evidence of a difference among the vehicle, 40 mg/kg/day, 120 mg/kg/day, and 200 mg/kg/day Velac treatment groups.

3.2.2.2 Papilloma incidence and burden:

The following tables (Tables 6 (Males) and 7 (Females)) display tumor incidence (number of animals with tumor) and several different ways of computing the mean number of tumors per animal per treatment group at the end of the study. The row labeled "SOA" lists the number of mice at Week 27 (or end of the study) with active papillomas at the site of application. The row labeled "SOA + DSOA" lists the number of mice at Week 27 with active or disappeared SOA papillomas. The row labeled "SOA + DSOA + LSOA" lists the number of mice with any SOA papillomas at Week 27, including active, disappeared, or latent. The latter entry fits the Sponsor's definition of tumor burden at the end of the study. Note that the corresponding definition of tumor burden given in Dunson et al. (2000) does not include disappeared tumors and apparently not latent tumors. The row labeled "NSOA" gives the number of mice with any active, disappeared, or latent non-SOA papillomas at Week 27. Finally the row labeled "All" indicates the number of mice with any papillomas.

For each treatment group, the end of study SOA tumor burden is first defined relative to all animals with papillomas using active and disappeared tumors. This mean is computed among animals with at least one papilloma. The next row is also computed among animals with at least one papilloma using all SOA tumors (active, disappeared, or latent). The final mean is the mean of all tumors defined over all animals at risk (thus including animals with no tumors).

Table 6. Summary of Papilloma Incidence at end of Study: Male Mice

Group / Summary	1 Vehicle Control	2 Untreated Control	3 Positive Control	4 40 mg/ kg/day	5 120 mg/ kg/day	6 200 mg/ kg/day
Incidence (Animals with tumor / Number animals)						
SOA	7 / 20	0 / 20	14 / 15	9 / 20	11 / 20	12 / 20
SOA + DSOA	8	0	14	11	13	14
SOA + DSOA + LSOA	9	0	14	12	15	16
NSOA	5	4	3	5	6	6
All	12	4	14	15	16	16
SOA Tumor Burden (Mean number of tumors)						
Per animal w/pap. (SOA+DSOA)	1.8	0	NA	2.1	2.1	3.2
Per animal w/pap.(SOA+DSOA+LSOA)	1.8			2.2	2.7	3.4
Per animal at risk	0.8			1.3	2.0	2.8

Fisher exact tests of tumor incidence comparing each of the four Velac dose groups above (vehicle, 40 mg/kg/day, 120 mg/kg/day, and 200 mg/kg/day) to the untreated groups gave statistically significant results (all $p \leq 0.0006$). The Cochran-Armitage test of trend in the number of all SOA papillomas (active, disappearing, and latent) among males in the four Velac Gel dose groups, was statistically significant (two-sided test, $p \leq 0.0173$).

Similar results for females are displayed in Table 7 below.

Table 7. Summary of Papilloma Incidence at end of Study: Female Mice

Summary \ Group	1 Vehicle Control	2 Untreated Control	3 Positive Control	4 40 mg/ kg/day	5 120 mg/ kg/day	6 200 mg/ kg/day
Incidence (Animals with tumor / Number animals)						
SOA	3 / 20	0 / 20	14 / 15	5 / 20	8 / 20	12 / 20
SOA + DSOA	5	0	15	5	10	14
SOA + DSOA + LSOA	5	0	15	6	12	16
NSOA	3	6	2	2	2	5
All	6	6	15	7	12	16
SOA Tumor Burden (Mean number of tumors)						
Per animal w/pap. (SOA+DSOA)	1.8	0	NA	1.4	1.7	1.9
Per animal w/pap.(SOA+DSOA+LSOA)	2.0			1.7	1.8	2.2
Per animal at risk	0.5	0		0.5	1.1	1.8

For the incidence of tumors at Week 27 in female mice, as displayed in Table 7 above, Fisher exact tests of tumor incidence comparing each of the four Velac dose groups to the untreated group also gave statistically significant results (all $p \leq 0.0236$). Over the four Velac dose groups the Cochran-Armitage test of trend in incidence in SOA papillomas was statistically significant ($p < 0.0001$).

Modeling tumor count per animal at risk with a Poisson regression model, with response as a function of dose, the linear effect of dose over the vehicle, 40 mg/kg/day, 120 mg/kg/day, and 200 mg/kg/day Velac dose groups was statistically highly statistically significant ($p < 0.0001$ for both genders). The two degree of freedom tests of nonlinearity in dose were highly non-significant ($p \leq 0.5845$ for males, $p \leq 0.7699$ for females). For the censored normal theory Tobit regressions, the effect of increasing dose was statistically significant in both genders ($p \leq 0.0023$ for males, $p \leq 0.0001$ for females). The two degree of freedom tests of nonlinearity in dose in the Tobit models were highly statistically non-significant. Frequentist hypothesis testing is structured so that only type I error is actually controlled, so failure to reject does not necessarily mean the null hypothesis is true. However, the very large significance levels associated with the tests of non-linearity in dose, are clearly suggestive that all group differences are mainly associated with linear trend in dose.

When modeling tumor count per animal with tumors using a Possion regression model, the linear effect of dose was statistically significant in males ($p \leq 0.0071$), and not significant for females ($p \leq 0.5312$). Further tests of non-linearity in dose were not statistically significant (males: $p \leq 0.9719$, females: $p \leq 0.8236$). Very similar results were obtained with a continuous response gamma model. However, in the four Velac dose groups, only 39 female mice and 52 male mice displayed tumors. With such small sample sizes the maximum likelihood techniques used may not be appropriate, and tests based on these estimates may not have much power to detect differences.

Plots of Kaplan-Meier product limit estimates of the time to first tumor for the four Velac dose groups are presented in Appendix 4. Table 8, below, summarizes the means and quartiles of these distributions:

Table 8. Summaries of the distribution of time (in weeks) to first tumor

Group	1 Vehicle/Control	4 40 mg/kg/day	5 120 mg/kg/day	6 200 mg/kg/day
Male				
1st quartile	20.0	16.0	16.0	15.0
Median	- -	25.0	25.0	18.0
Mean	22.5	22.3	21.1	19.4
3rd quartile	- -	- -	27.0	27.0
Female				
1st quartile	25.0	24.0	16.0	13.0
Median	- -	- -	25.0	25.0
Mean	23.8	24.7	27.0	20.1
3rd quartile	- -	- -	- -	27.0

Recall that, because of the censoring, the mean estimates of time to the event tend to be biased downwards. Furthermore, depending upon the proportion of mice that are censored some quantiles in the distribution of time to event may not be estimable. The overall test of differences in Kaplan-Meier curves of the time to first tumor was statistically non-significant in males (logrank $p \leq 0.2088$ and Wilcoxon $p \leq 0.2844$) and statistically significant in females (logrank $p \leq 0.0010$ and Wilcoxon $p \leq 0.0019$). More focused tests of linear trend using the Cox proportional hazards model were barely statistically significant in males ($p \leq 0.0431$) and quite significant in females ($p \leq 0.0002$). The two degree of freedom tests of nonlinearity in trend were statistically quite non-significant ($p \leq 0.9561$ and $p \leq 0.6154$, respectively), so treatment differences are primarily due to linear decreases in dose.

Note that many animals did not develop SOA tumors. For these animals the time to the maximum number of tumors was one week. As the number of tumors in a dose group increases we would expect the time to the maximum dose to increase, so an increasing trend in time to the maximum dose would suggest tumorigenicity. Plots of Kaplan-Meier product limit estimates of the time to maximum tumor count for the four Velac Gel dose groups are presented in Appendix 4. Table 9 summarizes these distributions:

Table 9. Summaries of the distribution of times (in weeks) to maximum tumor count

Group	1 Vehicle/Control	4 40 mg/kg/day	5 120 mg/kg/day	6 200 mg/kg/day
Male				
1st quartile	1.0	1.0	6.5	12.0
Median	1.0	13.5	19.5	24.0
Mean	10.7	14.2	17.1	18.5
3rd quartile	25.0	27.0	27.0	27.0
Female				
1st quartile	1.0	1.0	1.0	10.5
Median	1.0	1.0	17.0	22.5
Mean	6.1	7.7	14.5	18.1
3rd quartile	7.0	18.5	27.0	26.5

The overall test of differences in the product limit curves of time to maximum tumor count was statistically non-significant in males (logrank $p \leq 0.1140$ and Wilcoxon $p \leq 0.1148$) and quite statistically significant in females (logrank $p \leq 0.0070$ and Wilcoxon $p \leq 0.0013$). The more focused tests of linear trend using the Cox proportional hazards model were statistically significant in female mice ($p \leq 0.0193$) and statistically non-significant in male mice ($p \leq 0.1680$). For males, this was largely due to overlapping time to event curves in the non-vehicle groups. For females, the two degree of freedom test of nonlinearity in trend over dose was statistically quite non-significant ($p \leq 0.6902$).

3.2.2.3 Dunson's model for incidence and tumor burden:

The Dunson model is a Markov generalized linear mixed model, with terms for initiation of the first tumor and development of further tumors (Please see Appendix 6 for details). The response variable was defined as the increase in the number of active tumors, not counting tumors that disappeared or were still categorized as latent. There were statistically significant terms related to papilloma initiation for both genders ($p < 0.0001$ for both males and females). There is evidence of a dose related trend in the development of further tumors after tumor initiation in males ($p \leq 0.0497$). However there is no strong evidence for the development of further tumors in females ($p < 0.8990$).

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

NA

5 SUMMARY AND CONCLUSIONS

5.1. Statistical Issues and Collective Evidence

According to Dunson et al. (2000), the current standard for statistical analysis of skin papillomas separately tests for differences between the experimental group and the control group with respect to (in a slightly different order):
(1) percentage (or count) of animals with tumors, presumably at the end of the study
(2) average number of tumors per animal at risk
(3) average number of tumors per tumor-bearing animal
(4) average latency time to appearance of the first skin tumor
(5) average latency time to the appearance of the maximal number of tumors

In discussion with the FDA toxicologist it was noted that the Executive Carcinogenicity Assessment Committee (Exec CAC) tends to prefer endpoint (1), with support from endpoints (2) or (3). However other endpoints seem to be recommended by the National Center for Toxicological Research. Results for all five endpoints are reported below. Note that the Sponsor proposed analyses of (1) above were based on Fisher exact tests of incidence comparing each level of treatment to control, and for (2) and (3) on ANOVAs of the number of tumors, followed by Dunnett's tests. For either the number of tumors per animal at risk or the number of tumors per animal with tumors, tumor counts were skewed to the right, apparently following an exponential or gamma distribution. Under such circumstances, ANOVA may require more than the 80 observations available at the end of the study for robust results. For this reason, this reviewer has doubts about the appropriateness of the Sponsor's analysis.

In the paper cited above, Dunson et al. (2000) proposed a mixed effects Poisson model for the increase in the maximum number of papillomas. Details of this model are provided in Appendix 6. In addition to the Dunson model for both incidence and number of tumors, the endpoints cited above were analyzed as follows:

1. Analysis of the count of animals with site of application (SOA) tumors at the end of the study:

Interest was focused on the effect of dose. In such circumstances a Cochran-Armitage test of trend over the vehicle, and 40, 120, and 200 mg/kg/day groups would seem to be more appropriate than the Fisher exact tests proposed by the sponsor. Furthermore, in the presence of trend, Cochran-Armitage tests of dose related trend would be more powerful than the Fisher exact tests used by the Sponsor. Note that the relatively small sample size should limit possible overpowering of the tests.

2. Analysis of the number of tumors per animal at risk:

The distribution of the number of tumors was skewed to the right, with many animals having no SOA papillomas and thus having tumor counts scored as zeros, which suggests a Poisson regression model. Alternatively, a model that explicitly limits the dependent variable may be appropriate, as with Tobit models. Recall that this is a normal linear regression model with dose as a factor, but response censored at zero. For both models, the effect of linear dose and nonlinearity in dose was tested. Overall, both the Poisson regression and the Tobit regression models gave consistent results.

3. Analysis of the number of tumors per animal with tumors:

Again the distribution of the number of tumors was skewed, and a Poisson regression model seemed to fit, though less well than for tumors per animal at risk. An alternative would be to consider the response as continuous and use a gamma response model. However, only about half the animals displayed tumors, mostly in the high dose groups. First modeling the dose response only for those animals with tumors affected the randomization to treatment. Furthermore, with the smaller sample size, the asymptotic analysis used to derive the likelihood based results is more questionable.

4. & 5. Analysis of time to appearance of the first skin tumor and time to appearance of the maximal number of tumors.

Time to event curves for each of the vehicle group, and the 40, 120, and 200 mg/kg/day groups were estimated using product limit techniques. Overall homogeneity of these curves was tested using logrank and Wilcoxon tests. The logrank test can be interpreted as a Cox proportional hazards regression test of linear trend pooled with a two degrees of freedom test of nonlinearity in trend in the four doses. This test was used to test for linear trend in dose.

This reviewer also considered other issues, discussed below:

1. Almost all animals survived to sacrifice at the end of the study. However, survival proportions were compared using Kaplan-Meier product limit techniques.

2. Note that the untreated control group had zero SOA tumors, so, with the possible exception of incidence, it was determined that tests comparing the untreated control to the vehicle and the three dose groups would be superfluous, and were not done.

3. A number of different analyses are presented in this review. Clearly the results of a number of these will be correlated. If we were primarily interested in controlling overall type I error, i.e., the error involved in rejecting the true null hypothesis of no treatment differences, a correction for multiplicity would be needed. This approach seems to be suggested by Dunson et al. (2000). However, this reviewer considers that type II error, the error involved in not rejecting a false null hypothesis when there are differences, is at least as important, but is not generally controlled in standard frequentist formulations of hypothesis testing. Note that Haseman's rules

for multiplicity, used for standard rodent carcinogenicity analyses are not applicable in this situation. An appropriate approach to the multiplicity problems involved in these analyses is not clear and should be considered as a matter for future development. When considering the analyses presented, it should be noted that, due to the multiple comparisons, family-wise type I error was not strictly controlled.

4. Another issue is the exact definition of tumor burden. The Sponsor defined tumor burden as the sum of both active and disappeared tumors. At the end of the study, tumor was redefined to include latent tumors. This definition was followed in most analyses here. However, the Dunson definition is based on active tumors only.

There were no SOA tumors in the untreated group. Fisher Exact tests of tumor incidence comparing each of the four dose groups above (including vehicle) gave statistically significant differences with the untreated group. Tests of trend in dose generally confirmed the dose effects noted above (for both genders, all $p \leq 0.0236$, and most much less).

For tumor incidence in the four nominal Velac gel dose groups, i.e., the vehicle, 40 mg/kg/day, 120 mg/kg/day, and 200 mg/kg/day, the Cochran-Armitage test of trend in the number of all SOA papillomas (active, disappearing, and latent) was statistically significant for both genders (two-sided test, males: $p \leq 0.0173$, females $p < 0.0001$). When modeling tumor count per animal at risk as a function of dose using a Poisson model, the linear effect of dose was statistically significant ($p < 0.0001$, for both genders). Further tests of non-linearity in dose were not statistically significant (males: $p \leq 0.5845$, females $p \leq 0.7699$).

There was also interest in modeling tumor counts per animal with tumors. Using a Poisson regression model for this restricted subset of mice, the linear effect of dose was statistically significant for males ($p \leq 0.0065$), and not significant for females ($p \leq 0.5312$). Further tests of non-linearity in dose were not statistically significant (males: $p \leq 0.9719$, females $p \leq 0.8236$). However, in the four Velac treatment groups, only 39 female mice and 52 male mice displayed tumors. With such small sample sizes the maximum likelihood techniques used may not be appropriate, and tests based on these estimates may not have much power to detect differences.

The overall test of differences in Kaplan-Meier curves of the time to first tumor was statistically non-significant in males (logrank $p \leq 0.2088$ and Wilcoxon $p \leq 0.2844$) and statistically significant in females (logrank $p \leq 0.0010$ and Wilcoxon $p \leq 0.0019$). More focused tests of linear trend using the Cox proportional hazards model were barely statistically significant in males ($p \leq 0.0431$) and quite significant in females ($p \leq 0.0002$). The two degree of freedom tests of nonlinearity in trend were statistically quite non-significant ($p \leq 0.9561$ and $p \leq 0.6154$ for each gender, respectively).

Again, the overall test of treatment differences in the curves of time to maximum tumor count was statistically non-significant in males (logrank $p \leq 0.1140$ and Wilcoxon $p \leq 0.1148$)

and statistically significant in females (logrank $p \leq 0.0070$ and Wilcoxon $p \leq 0.0013$). Similarly to the results for time to first tumor, the more focused tests of linear trend using the Cox proportional hazards model were not statistically significant in males ($p \leq 0.1680$) and quite significant in females ($p \leq 0.0193$). Again, the two degree of freedom tests of nonlinearity in trend were statistically quite non-significant ($p \leq 0.7507$ and $p \leq 0.6902$ for males and females, respectively).

Dunson, et al. (2000), proposed a Markov generalized linear mixed model combining time to first tumor and the subsequent number of tumors (Please see Appendix 6 for details). For both genders there were statistically significant dose related trends in papilloma initiation ($p < 0.0001$ for both genders). In males there was evidence of a dose related trend in the development of further tumors after tumor initiation ($p \leq 0.0497$). However, there was no strong evidence for the development of further tumors in females ($p < 0.8990$). Note that the tests based on tumor incidence and tumor counts described above were generally consistent with the results of the Dunson model. Specifically, there seems to be a statistically significant dose related effect on tumor initiation, but only males show a statistically significant dose related difference after the initiation of the first tumors.

Kaplan-Meier plots of the survival distribution are given in Appendix 2, and show no strong dose related trends in survival. Note that tests of homogeneity of survival were statistically non-significant (Males: logrank $p \leq 0.6494$, Wilcoxon $p \leq 0.6611$, Females: logrank $p \leq 0.1131$, Wilcoxon $p \leq 0.1008$).

5.2. Conclusions and Recommendations

This submission was intended to assess whether daily dermal administration of Velac Gel for 26 weeks increased the incidence of dermal tumors at the site of application in hemizygous Tg.AC transgenic mice. The analysis was based on results from four dose groups: a vehicle only group and nominal clindamycin doses in the vehicle of 40, 120, and 200 mg/kg/day. For both genders, there was statistically significant evidence of an increasing response to dose in terms of time to tumor development, numbers of animals with tumors, and number of tumors per animal. However, it should be noted that actual tumor incidence and burden for any of the four Velac Gel dose groups was far less than the active control group, which was treated with the known carcinogen.

APPENDICES:

Appendix 1. Observed Individual Skin Lesions At End of Study

The following tables display lesion counts for each animal with lesions. The Sponsor defined papilloma codes were as follows:
SOA = Site of application
NSOA = Non-site of application
L = Latent papilloma (Not yet observed for 3 consecutive weeks)
D = Observed for 3 consecutive weeks but subsequently disappeared
(Note DNSOA papillomas are indicated by parentheses in NSOA column, e.g., (1))

Table A.1.1 Male Tg.AC Papilloma Incidence at Week 27, by Animal

(b) (4)

(b) (4)

Table A.1.2 Female Tg.AC Papilloma Incidence at Week 27, by Animal

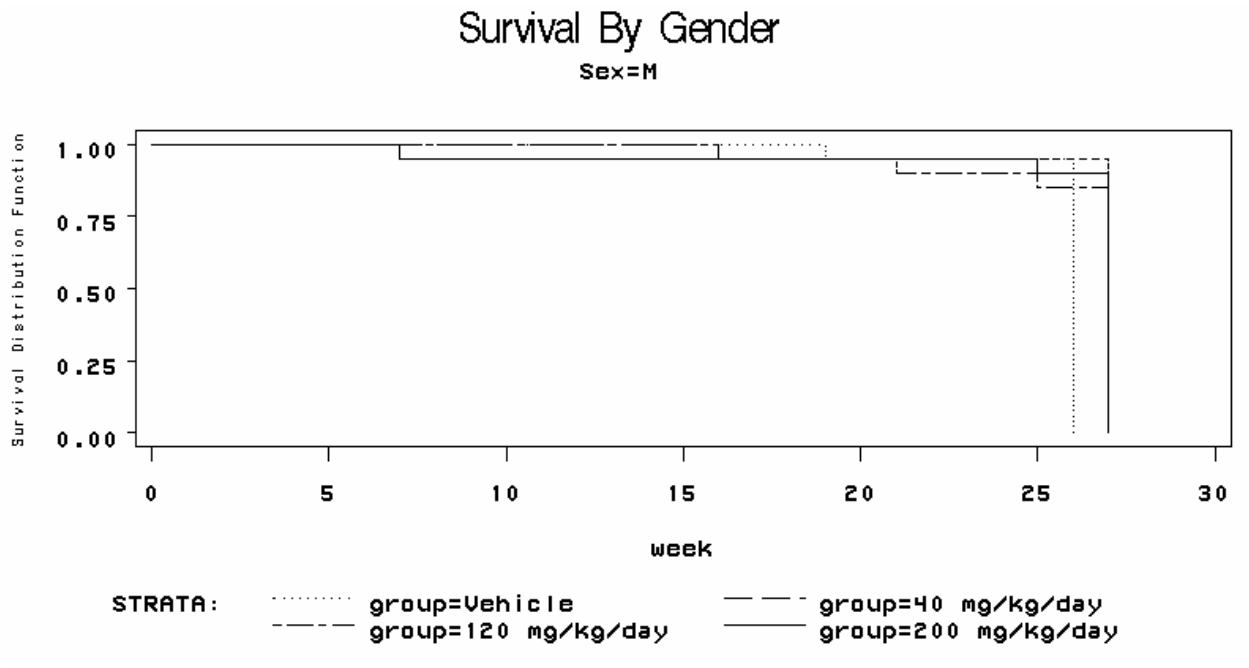
(b) (4)

Note that results from the active control (Group 3) demonstrate the sensitivity of the model, while results in Groups 1, 4, 5, and 6 show a general increase in tumor burden with dose.

Appendix 2. Survival Analysis

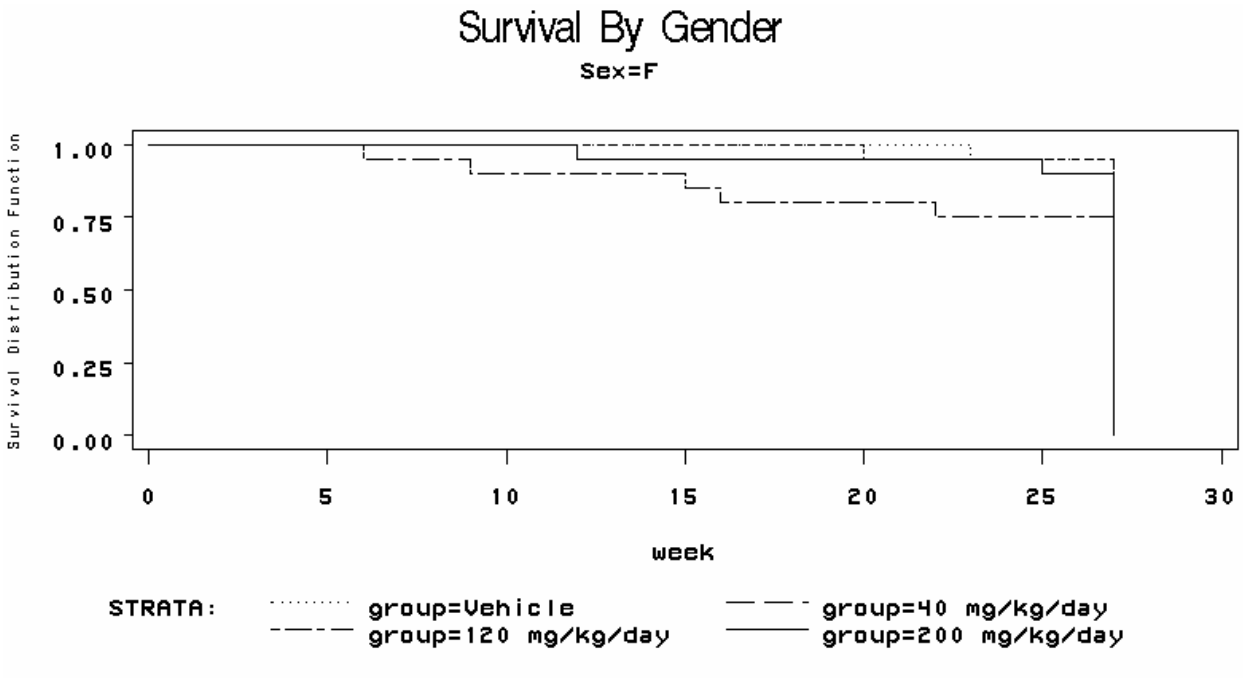
Group 1 is the vehicle/control group, Group 4 is the 40 mg/kg/day treatment group, Group 5 is the 120 mg/kg/day treatment group, and Group 6 is the 200 mg/kg/day group. The following are the survival curves comparing treatments for each gender. In the plots below the death of animals sacrificed at the end of the study are displayed as events. In the tests comparing the survival curves those mice sacrificed are considered as censored.

Male Tg.AC Mice



There was no evidence of a dose related trend in survival in these male mice. Tests of differences in the the time to event curves were statistically non-significant (logrank $p \leq 0.6494$, Wilcoxon $p \leq 0.6611$).

Female Tg.AC Mice



For females there was no statistically significant evidence of a dose related trend in survival (logrank $p \leq 0.1131$, Wilcoxon $p \leq 0.1008$).

Appendix 3. Weekly SOA Tumor Incidence

The following tables summarize the number of animals that died during the study by week of death, before week 26 and the number of animals with active and disappeared tumors, and active, disappeared, and latent tumors prior to final sacrifice. The mean number of tumors is presented among mice with tumors. The minimum and maximum numbers of tumors are also displayed.

Group 1 = Velac Gel vehicle Group
Group 4 = Clindamycin phosphate 40 mg/kg/day
5 = Clindamycin phosphate 120 mg/kg/day
6 = Clindamycin phosphate 200 mg/kg/day

Table A.3.1 Male Site of Application Tumor Incidence

Treatment Group		week 1-5	6	7	8	9	10	11	12	13	14	15	16
1	# dead	0	0	0	0	0	0	0	0	0	0	0	0
	# animals w/ tumors	0	0	0	0	0	2	2	2	2	4	5	5
	mean tumors	.	.	.	.	.	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	min-max	.	.	.	.	.	1-1	1-1	1-1	1-1	1-1	1-1	1-1
4	# dead	0	0	0	0	0	0	0	0	0	0	0	1
	# animals w/ tumors	0	0	0	0	0	0	0	1	2	4	4	6
	mean tumors	.	.	.	.	.	.	.	1.0	1.5	1.3	1.3	1.2
	min-max	.	.	.	.	.	.	.	1-1	1-2	1-2	1-2	1-2
5	# dead	0	0	0	0	0	0	0	0	0	0	0	1
	# animals w/ tumors	0	0	0	0	1	2	2	3	3	4	4	6
	mean tumors	.	.	.	.	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	min-max	.	.	.	.	1-1	1-1	1-1	1-1	1-1	1-1	1-1	1-1
6	# dead	0	0	1	1	1	1	1	1	1	1	1	1
	# animals w/ tumors	0	0	0	1	2	2	2	3	3	4	5	7
	mean tumors	.	.	.	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.4	1.3
	min-max	.	.	.	1-	1-1	1-1	1-1	1-1	1-1	1-1	1-2	1-2

Treatment Group		week 17	18	19	20	21	22	23	24	25	26	27
1	# dead	0	0	1	1	1	1	1	1	1	20	20
	# animals w/ tumors	5	5	5	5	5	5	5	5	7	9	9
	mean tumors	1.0	1.0	1.0	1.2	1.2	1.2	1.4	1.6	1.6	1.7	1.8
	min-max	1-1	1-1	1-1	1-2	1-2	1-2	1-2	1-3	1-3	1-3	1-4
4	# dead	1	1	1	1	1	1	1	1	1	1	20
	# animals w/ tumors	6	6	7	7	7	7	7	8	10	10	12
	mean tumors	1.2	1.2	1.1	1.4	1.4	1.4	1.6	1.5	1.7	2.0	2.2
	min-max	1-2	1-2	1-2	1-2	1-2	1-2	1-2	1-2	1-3	1-3	1-4
5	# dead	1	1	1	1	2	2	2	2	3	3	20
	# animals w/ tumors	6	7	8	8	9	9	9	9	12	13	15
	mean tumors	1.3	1.3	1.4	1.6	1.7	1.9	1.9	2.4	2.1	2.0	2.7
	min	1-2	1-2	1-2	1-3	1-3	1-4	1-4	1-5	1-5	1-5	1-10
6	# dead	1	1	1	1	1	1	1	1	2	2	20
	# animals w/ tumors	8	10	11	12	12	12	12	12	13	14	16
	mean tumors	1.3	1.2	1.4	1.5	1.8	2.3	2.7	3.1	3.3	3.2	3.4
	min-max	1-2	1-2	1-3	1-3	1-4	1-5	1-5	1-10	1-10	1-11	1-15

Table A.3.2 Female Site of Application Tumor Incidence

Treatment Group		week 1-5	6	7	8	9	10	11	12	13	14	15	16
1	# dead	0	0	0	0	0	0	0	0	0	0	0	0
	# animals w/ tumor	0	0	0	0	0	0	0	0	1	1	1	1
	mean tumors	.	.	.	.	.	.	.	.	1.0	1.0	1.0	1.0
	min-max	.	.	.	.	.	.	.	.	1-1	1-1	1-1	1-4
4	# dead	0	0	0	0	0	0	0	0	0	0	0	0
	# animals w/ tumor	0	0	0	0	1	1	1	1	1	2	2	2
	mean tumors	.	.	.	.	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	min-max	.	.	.	.	1-1	1-1	1-1	1-1	1-1	1-1	1-1	1-1
5	# dead	0	1	1	1	2	2	2	2	2	2	3	4
	# animals w/ tumor	0	0	0	1	1	2	2	2	3	3	3	5
	mean tumors	.	.	.	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.2
	min-max	.	.	.	1-1	1-1	1-1	1-1	1-1	1-1	1-1	1-1	1-2
6	# dead	0	0	0	0	0	0	0	1	1	1	1	1
	# animals w/ tumors	0	0	1	2	2	4	4	4	5	5	7	7
	mean tumors	.	.	1.0	1.5	1.5	1.3	1.3	1.3	1.2	1.2	1.1	1.1
	min-max	.	.	1-1	1-2	1-2	1-2	1-2	1-2	1-2	1-2	1-2	1-2

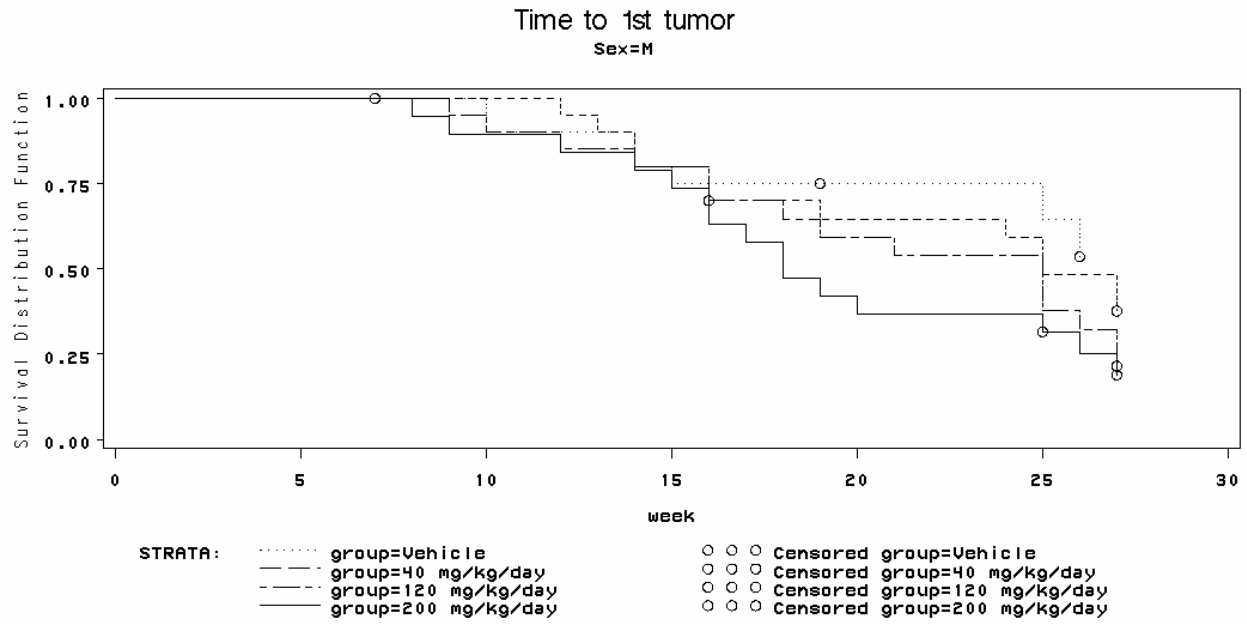
Treatment Group		week 17	18	19	20	21	22	23	24	25	26	27
1	# dead	0	0	0	0	0	0	1	1	1	1	20
	# animals w/ tumor	1	1	2	2	3	3	3	4	5	5	5
	mean tumors	1.0	1.0	1.0	1.0	1.3	1.3	1.7	1.5	1.4	1.8	2.0
	min-max	1-1	1-1	1-1	1-1	1-2	1-2	1-3	1-3	1-3	1-3	1-4
4	# dead	0	0	0	1	1	1	1	1	1	1	20
	# animals w/ tumors	2	2	2	3	3	3	4	5	5	5	6
	mean tumors	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.2	1.2	1.4	1.7
	min-max	1-1	1-1	1-1	1-1	1-1	1-1	1-1	1-2	1-2	1-2	1-4
5	# dead	4	4	4	4	4	5	5	5	5	5	20
	# animals w/ tumors	5	6	6	6	6	6	6	6	9	10	12
	mean tumor	1.2	1.3	1.5	1.5	1.5	1.7	1.7	1.7	1.7	1.6	1.8
	min-max	1-2	1-2	1-3	1-3	1-3	1-4	1-4	1-4	1-4	1-4	1-4
6	# dead	1	1	1	1	1	1	1	1	2	2	20
	# animals w/ tumors	7	7	8	9	9	9	9	9	12	14	16
	mean tumors	1.1	1.6	1.5	1.7	1.7	1.7	1.8	1.8	1.9	1.9	2.2
	min-max	1-2	1-2	1-2	1-3	1-3	1-3	1-3	1-3	1-5	1-5	1-6

Note that for both sexes there is clear trend over increasing dose, for an increasing number of animals with tumors and an increasing mean number of tumors as the study progressed.

Appendix 4. Time to Appearance of First Tumor

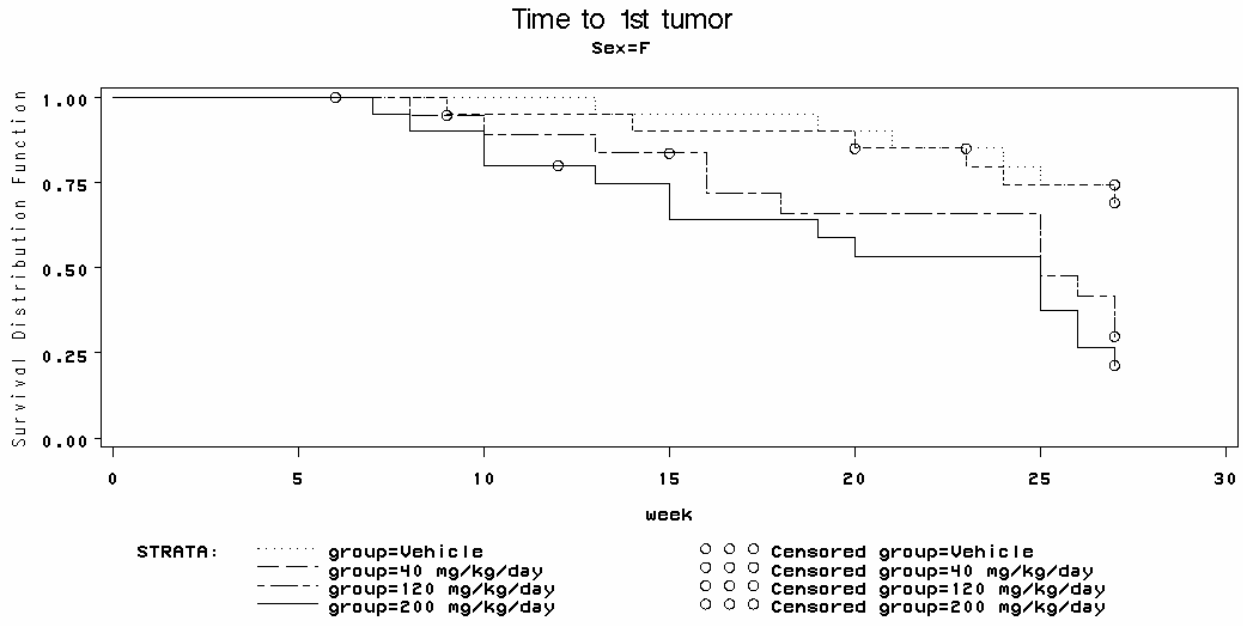
The overall test of differences in Kaplan-Meier curves among the four Velac dose groups (vehicle (dose=0), 40 mg/kg/day, 120 mg/kg/day, and 200 mg/kg/day) of the time to first tumor was statistically non-significant in males (logrank $p \leq 0.2088$ and Wilcoxon $p \leq 0.2844$) and statistically significant in females (logrank $p \leq 0.0010$ and Wilcoxon $p \leq 0.0013$). More focused tests of linear trend over the four Velac Gel dose levels using the Cox proportional hazards model were barely statistically significant in males ($p \leq 0.0431$) and quite significant in females ($p \leq 0.0002$). The two degree of freedom tests of nonlinearity in trend were statistically quite non-significant ($p \leq 0.9561$ and $p \leq 0.6154$, for males and females respectively). The Kaplan-Meier estimated curves of the time to first tumor are given below:

Male Tg.AC Mice



Best Available Copy

Female Tg.AC Mice

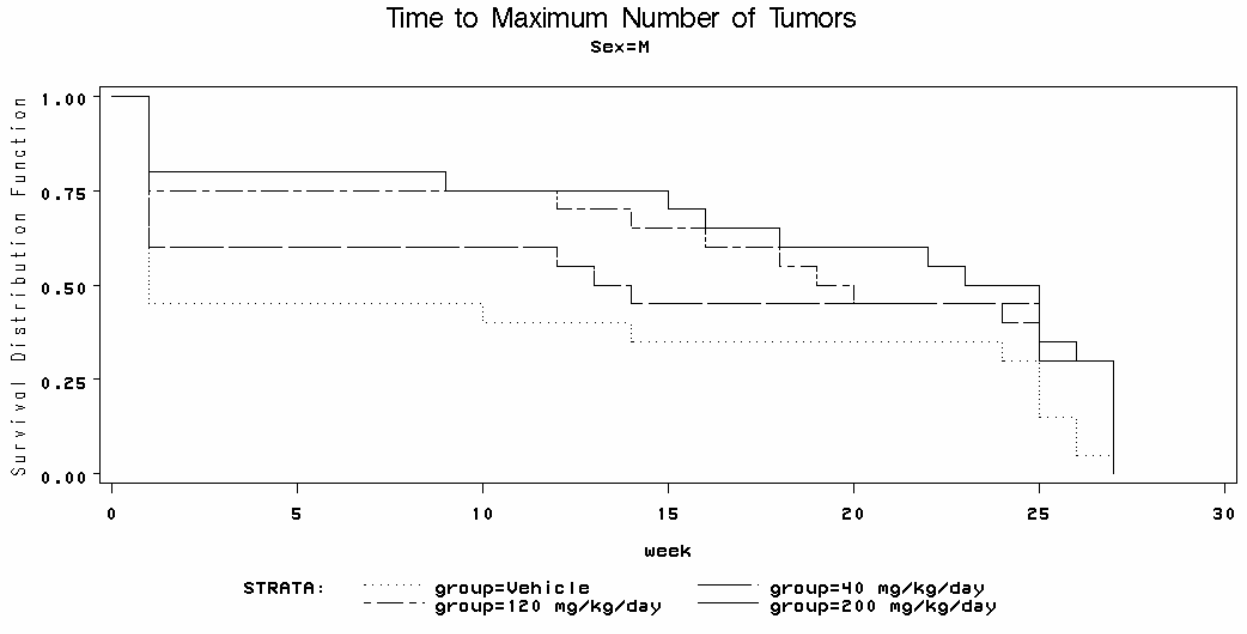


Best Available Copy

Appendix 5. Time to Appearance of the Maximum Number of Tumors

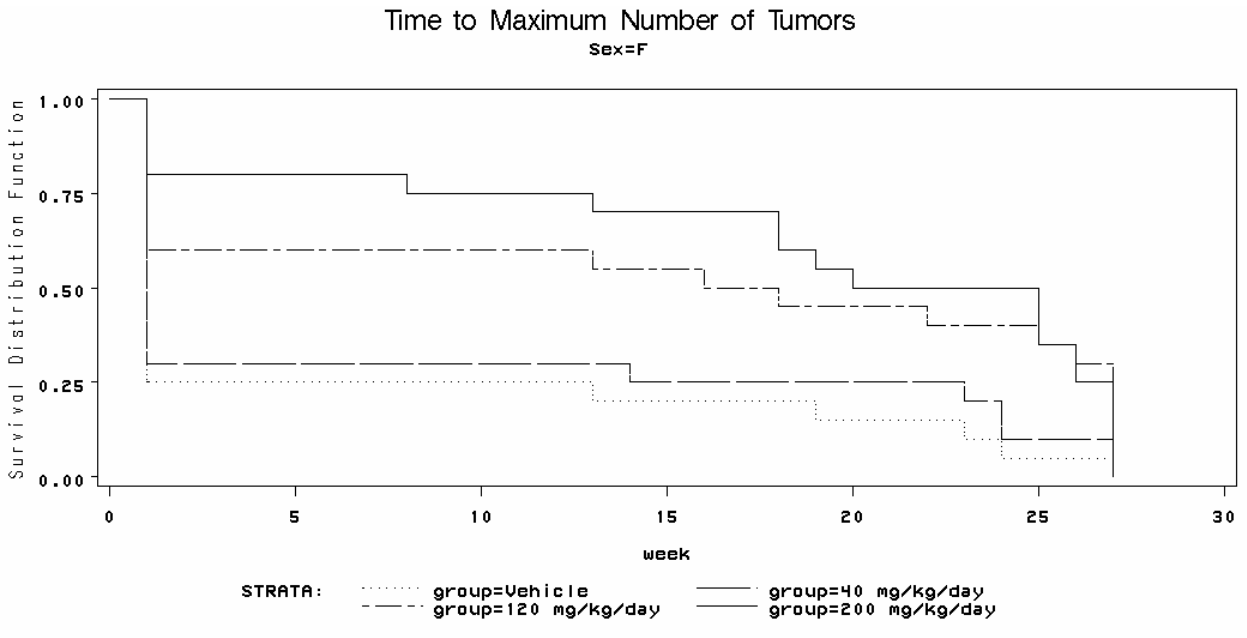
Again, the overall test of differences between the four Velac dose groups (vehicle (dose=0), 40 mg/kg/day, 120 mg/kg/day, and 200 mg/kg/day) in the product limit curves of time to maximum tumor count was statistically non-significant in male mice (logrank $p \leq 0.1140$ and Wilcoxon $p \leq 0.1148$) and was statistically significant in female mice (logrank $p \leq 0.0070$ and Wilcoxon $p \leq 0.0013$). The more focused tests of linear trend using the Cox proportional hazards model were not statistically significant in males ($p \leq 0.1680$) and were statistically significant in females ($p \leq 0.0193$). The two degree of freedom tests of nonlinearity in trend were statistically quite non-significant ($p \leq 0.7507$ and $p \leq 0.6902$, for male mice and female mice, respectively).

Male Tg.AC Mice



Best Available Copy

Female Tg.AC Mice



Best Available Copy

Appendix 6. Dunson Model

Let Z_{ij} be the number of detectable papillomas on the back of mouse i at week j. Let $M_{ij} = \max\{ Z_{1j}, \ldots, Z_{i,j-1}\}$ denote the maximum papilloma burden for mouse i prior to week j, and let $Y_{ij} = M_{i,j+1} - M_{i,j}$ be the increase in burden between week j-1 and j. Dunson models the Y_{ij} as a Poisson random variable with means:

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

where b_i denotes the mouse specific susceptibility ~ Normal(0, σ^2)

t_j = j/T where T is the number of weeks in the study

d_i = log (dose) (dose 0 is retained as a 0)

β_1, β_2 are intercept parameters associated with the rate of appearance of spontaneous papillomas.

γ_1, γ_2 are slope parameters associated the effect of dose.

Note that in the 0 dose control group the probability that a control animal develops its first tumor is 1 - exp (- exp (β_1)). Since this probability is likely to be small we would expect β_1 would be a relatively large negative number. The probability of detecting the first tumor should increase with time and dose, as reflected in the term $(b_i + \gamma_1)\, t_j\, d_i$.

This model is described in detail in Dunson et al. (2000). The SAS procedure NLMIXED can be used to estimate parameters as with the following program:

```
PROC NLMIXED DATA=dunson ; BY NOTSORTED sex;
     PARMS var 1 beta1 beta2 1 gam1 gam2 0.1;
       IF week=1 THEN eta = exp(beta1+(b+gam1)*t*dose);
            ELSE IF week>1 AND M0=0 THEN
                   eta = exp(beta1+(b+gam1)*t*dose);
            ELSE eta = exp(beta2+gam2*dose);
      MODEL y~Poisson(eta);
      RANDOM b~normal(0,var) subject=animal;
run;
```

Giving the following, somewhat edited, output:

```
--------------------- Sex=M -----------------
                   Standard
Parameter Estimate   Error    t Value  Pr > |t|  Lower      Upper   Gradient

  var       0.1664    0.1034    1.61    0.1116    -0.03945   0.3723 -0.01068
  beta1    -5.0581    0.3746  -13.50    <.0001    -5.8038   -4.3124  0.006305
  beta2    -3.0572    0.5017   -6.09    <.0001    -4.0559   -2.0586  0.005659
  gam1      0.7298    0.1477    4.94    <.0001     0.4358    1.0239  0.018456
  gam2      0.2142    0.1075    1.99    0.0497     0.000273  0.4281  0.024245


--------------- Sex=F --------------------
```

Parameter	Estimate	Standard Error	t Value	Pr > |t|	Lower	Upper	Gradient
var	0.02394	0.05931	0.40	0.6876	-0.09411	0.1420	0.000151
beta1	-5.0190	0.3448	-14.55	<.0001	-5.7054	-4.3326	0.000042
beta2	-2.6243	0.6156	-4.26	<.0001	-3.8496	-1.3991	4.231E-6
gam1	0.5317	0.1171	4.54	<.0001	0.2987	0.7648	3.886E-6
gam2	-0.01714	0.1346	-0.13	0.8990	-0.2851	0.2509	-0.00003

Note that the statistically significant γ_1 (gam1) terms for both genders imply dose related trends in papilloma initiation ($p < 0.0001$ for both males and females). The nearly statistically non-significant γ_2 (gam2) terms indicate that there is some evidence of a dose related trend in the development of further tumors after tumor initiation in males ($p \leq 0.0497$). However there is no strong evidence for the development of further tumors in females ($p < 0.8990$). Note the non-significance of the σ^2 (var) term suggests relative homogeneity among mice in tumor initiation, particularly among females.

Appendix 7. Preliminary Bayesian Estimates of the Dunson Model

Recall that we define M_{ij} as the maximum number of papillomas for mouse i prior to week j, and let $Y_{ij} = M_{i,j+1} - M_{i,j}$ denote the increase in burden between week j-1 and j. The responses are modeled with a typical Bayesian hierarchical model where the $Y_{ij} \sim$ Poisson (μ_{ij}), with:

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

with variables:

$t_j = j/T$ where T is the number of weeks in the study

$d_i =$ log (dose) (dose 0 is retained as a 0)

b_i denotes individual mouse specific susceptibility $\sim$ Normal$(0, 1/\tau_1^2)$

and parameters with priors:

$\beta_1, \beta_2, \gamma_1, \gamma_2 \sim N(0, 1/\tau_2^2)$

$\tau_1, \tau_2 \sim$ Gamma(0.01, 0.0001).

Note that these generally seem to be considered as relatively non-informative priors (see the WINBUGS manual). As an aside, this reviewer has some question that the Gamma priors above generally should be considered non-informative.

Posterior parameter estimates for tumor burden in Males:

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta1	-4.296	0.2896	0.009027	-4.893	-4.285	-3.76	5001	20000
beta2	-3.12	0.4092	0.0139	-3.975	-3.097	-2.389	5001	20000
gam1	0.6368	0.1347	0.004844	0.3958	0.6284	0.9232	5001	20000
gam2	0.2617	0.08759	0.002977	0.1044	0.2574	0.4431	5001	20000
tau1	62.63	220.5	13.16	3.111	11.69	567.1	5001	20000
tau2	0.1418	0.1028	8.99E-4	0.01646	0.1182	0.4025	5001	20000

Posterior parameter estimates for tumor burden in females:

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta1	-4.668	0.2979	0.005977	-5.281	-4.657	-4.122	5001	20000
beta2	-2.812	0.5741	0.01838	-4.089	-2.764	-1.849	5001	20000
gam1	0.5297	0.1081	0.002141	0.3227	0.5298	0.745	5001	20000
gam2	0.08941	0.1224	0.003906	-0.1212	0.08096	0.355	5001	20000
tau1	201.8	393.2	21.86	5.842	58.6	1257.0	5001	20000
tau2	0.1334	0.09688	9.172E-4	0.01535	0.1112	0.3753	5001	20000

Recall that the tau parameters measure precision, the inverse of the variance. For each gender, the beta1, beta2, gam1, gam2 parameter estimates are quite consistent with the corresponding maximum likelihood estimates.

These estimates were derived using WINBUGS 1.4 programs similar to the following:

```
Model{
     for (i in 1:N){
      eta[i]<-equals(M[i],0)*exp(beta1+(b[subj[i]]+gam1)*t[i]*dose[i] ) +
                        step(M[i]-0.5)*exp(beta2+gam2*dose[i]) + 0.0*week[i]
            y[i]~dpois(eta[i])
                    }
     for (j in 1:nsubj){
             b[j]~dnorm(0,tau1)
                        }
      tau1~dgamma(0.01, 0.001)
      tau2~dgamma(0.01, 0.001)
      beta1 ~ dnorm(0.0,tau2)
      beta2 ~ dnorm(0.0,tau2)
      gam1 ~ dnorm(0.0,tau2)
      gam2 ~ dnorm(0.0,tau2)
    }
inits
list(tau1=0.1,tau2=0.2,beta1=0.1,beta2=0.2,gam1=0.01,gam2=0.02)

data
 list(N=2081,nsubj=80)
subj[ ] week[ ] M[ ] dose[ ] t[ ] y[ ]
   1    1    0   0.000  0.037     0
   1    2    0   0.000  0.074     0
   1    3    0   0.000  0.111     0
   - data -
  80   26    1   5.298  0.963     0
  80   27    1   5.298  1.000     0
  END
```

Appendix 8. References

Dunson, DB, JK Haseman, APJM van Birgelen, S Stasiewicz, and RW Tennant, R.W., 2000, Statistical Analysis of Skin Tumor Data from Tg.AC Mouse Bioassays, Toxicological Sciences, 55:293-302.

JW Spalding, JE French, RR Tice, M Furedi-Machacek, JK Haseman, and RW Tennant, 1999, Development of a Transgenic Mouse Model for Carcinogenesis Bioassays: Evaluation of Chemically Induced Skin Tumors in Tg.AC Mice, Toxicological Sciences, 49:241-254.

SIGNATURES/DISTRIBUTION LIST

Primary Statistical Reviewer: Steve Thomson
Date: January 27, 2005

Concurring Reviewer(s):

Statistical Team Leader: Mohamed Alosh, Ph.D., HFD-725

cc:
HFD-540:/Jon Wilkin. M.D.
HFD-540/Mildred Wright
HFD-540/Jill Merrill, Ph.D.
HFD-540/Paul Brown, Ph.D.
HFD-725/Steve Thomson
HFD-725/ Mohamed Alosh, Ph.D.
HFD-725/ Mohammed Huque, Ph.D.,
HFD-725/ Charles Anello, Ph.D.
HFD-725/Robert O'Neill, Ph.D.
HFD-725/Satya Dubey, Ph.D.

c:\Data\Profiles\MyDocuments~\carcNDA50803VelacGelConnectics.doc

---

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
1/27/05 02:47:22 PM
BIOMETRICS

Mohamed Alosh
1/28/05 03:43:34 PM
BIOMETRICS
Concur with review

Statistical Review and Evaluation
Fileability Review

NDA:	(b) (4)
Name of Drug:	Velac Gel (clindamycin 1% -tretinoin .025%)
Applicant:	Connetics Corporation
Indication:	Acne Vulgaris
Filing Date:	10-25-2004
Fileability Meeting Date:	10-13-2004
User Fee Date:	06-25-2005
Statistical Reviewer:	Mat Soukup, Ph.D., HFD-725
Clinical Reviewer:	Bindi Nikhar, M.D., HFD-540
Project Manager:	Margo Owens, HFD-540

Clinical Studies: VLC.C.304 and VLC.C.305 (Two 4-arm Phase III trial: Velac gel, clindamycin, tretonin, and vehicle)

I. ORGANIZATION AND DATA PRESENTATION	**YES/NO/NA**
A. Is there a comprehensive table of contents with adequate indexing and pagination?	Y
B. Are the original protocols, protocol amendments, and proposed label provided?	Y
C. Are the following tables/listings provided in each study report?	
1. Patient profile listings by center, for all enrolled patients.	Y
2. Discontinued subject tables by center (includes reason and time of loss).	Y
3. Subgroup analysis summary tables (gender, age, race, etc.)	Y
4. Adverse event listings by center and time of occurrence.	Y
D. Have the data been submitted electronically?	Y
1. Has adequate documentation of the data sets been provided?	Y
2. Do the data appear to accurately represent the data described in the study reports?	Y
3. Can the data be easily merged across studies and indications?	Y

II. STATISTICAL METHODOLOGY	YES/NO/NA
A. Are all primary efficacy studies of appropriate design to meet basic approvability requirements within current Division policy, or to the extent agreed upon previously with the sponsor by the Division?	Y
B. For each study, is there a comprehensive statistical summary of the efficacy analyses which covers the intent-to-treat population and per protocol population?	Y
C. Based on the summary analyses of each study,	
1. Are the analyses appropriate for the type of data collected, the study design, and the study objectives (based on protocol objectives and proposed labeling claims?)	Y
2. Are the intent-to-treat and per protocol patient analyses properly performed?	Y
3. Has missing data been appropriately handled?	Y
4. Have multiplicity issues (regarding endpoints, timepoints, or dose groups) been adequately addressed?	Y
5. If interim analyses were performed, were they planned in the protocol and appropriate significance level adjustments made?	Y
D. Were sufficient and appropriate references included for novel statistical approaches?	NA
E. Are all of the pivotal studies complete?	Y
F. Has the safety data been comprehensively and adequately summarized?	Y

III. FILEABILITY CONCLUSIONS

From a statistical perspective this submission, or indications therein, **is** reviewable with only minor further input from the sponsor.

IV. 74-DAY LETTER COMMENTS

Filing Issues: Insufficient details of interim analysis results are submitted.

Request for Information: Please provide the following information regarding the interim analysis in Study VLC.C.304.

- Efficacy results at the interim along with the number of subjects.

- Whether any “Special Situations” as defined in the Interim Statistical Analysis Plan (Section 16.1.13) occurred and if any action was taken.
- Please provide any related information that ensured adequate blinding in the conduct of the interim analysis.

Mat Soukup, Ph.D.
Mathematical Statistician, Biometrics III

Concur: Mohamed Alosh, Ph.D.
Team Leader, Biometrics III

cc:
Archival NDA (b) (4) SUP-000
HFD-540/Wilkin
HFD-540/Luke
HFD-540/Nikhar
HFD-540/Owens
HFD-700/Anello
HFD-725/Huque
HFD-725/Alosh
HFD-725/Soukup

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

/s/

Matt Soukup
10/21/04 02:15:01 PM
BIOMETRICS

Mohamed Alosh
10/21/04 02:38:03 PM
BIOMETRICS