CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

202428Orig1s000

STATISTICAL REVIEW(S)

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

STATISTICAL REVIEW AND EVALUATION

ADDENDUM

NDA/Serial Number:	202428 / 000
Drug Name:	TRADENAME (tazarotene) foam 0.1%
Indication(s):	Acne vulgaris
Applicant:	Stiefel
Dates:	Submitted: 7/29/2011 PDUFA: 5/29/2012
Review Priority:	Standard review
Biometrics Division:	Division of Biometrics III
Statistics Reviewer:	Kathleen Fritsch, Ph.D.
Concurring Reviewer:	Mohamed Alosh, Ph.D.
Medical Division:	Division of Dermatology and Dental Products
Clinical Team:	Denise Cook, M.D. / Gordana Diglisic, M.D.
Project Manager:	Cristina Attinello

Summary

This addendum includes additional details from the analysis of the secondary endpoint of percent reduction in lesions from the studies for tazarotene foam 0.1% in the treatment of acne vulgaris. Studies 301 and 302 were evaluated by this reviewer in the statistical review dated 3/14/2012. As pre-specified in the protocol, all lesion count endpoints (absolute and percent reduction) were analyzed with an ANCOVA model with terms for baseline value, treatment, center, and treatment-by-center interaction. The original statistical review presented results for both arithmetic and least squares means for the primary lesion count analyses (absolute reduction), but only least squares means for the secondary lesion count analyses (percent reduction). Least squares means are more closely associated with the Type III sum of squares (by adjusting for all other covariates in the model) used to calculate the p-values for the hypotheses. However, when the studies are reasonably balanced across centers and baseline counts, the results for the two mean types will be similar. Arithmetic means reflect the data as it was observed in the study rather than being impacted by the choice of the model used in the analysis. A comparison of the means and least squares means for the two studies is presented in Table 1. The two estimate types yield similar results in Studies 301 and 302. For ease in interpretation, presenting the study results in labeling using arithmetic means is recommended.

Table 1 – Means and Least Squares Means for Lesion Count Endpoints

Endpoint	Study 301		Study 302	
	Tazarotene N=371	Vehicle N=372	Tazarotene N=373	Vehicle N=369
Absolute change:				
Means				
Inflammatory	-18.0	-14.1	-17.8	-14.7
Non-inflammatory	-27.9	-16.7	-25.6	-18.2
Total	-45.8	-30.8	-43.3	-32.9
Least Squares Means				
Inflammatory	-17.6	-13.3	-17.6	-14.3
Non-inflammatory	-28.1	-16.2	-25.9	-17.0
Total	-45.8	-29.5	-43.5	-31.3
Percent change:				
Means				
Inflammatory	-57.5%	-45.2%	-54.5%	-45.3%
Non-inflammatory	-55.1%	-33.2%	-56.7%	-41.2%
Total	-56.3%	-39.0%	-56.0%	-42.6%
Least Squares Means				
Inflammatory	-56.1%	-43.6%	-53.7%	-44.3%
Non-inflammatory	-56.2%	-33.2%	-56.1%	-40.7%
Total	-56.4%	-37.6%	-55.2%	-42.0%

Signatures/Distribution List

Primary Statistical Reviewer: Kathleen Fritsch, Ph.D.
Date: 3/27/2012

Statistical Team Leader: Mohamed Alosh, Ph.D.

cc:
DDDP/Walker
DDDP/Diglisic
DDDP/Cook
DDDP/Attinello
OBIO/Patrician
DBIII/Wilson
DBIII/Alosh
DBIII/Fritsch

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

/s/

KATHLEEN S FRITSCH
03/27/2012

MOHAMED A ALOSH
03/27/2012

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	202428 / 000
Drug Name:	TRADENAME (tazarotene) foam 0.1%
Indication(s):	Acne vulgaris
Applicant:	Stiefel
Dates:	Submitted: 7/29/2011 PDUFA: 5/29/2012
Review Priority:	Standard review
Biometrics Division:	Division of Biometrics III
Statistics Reviewer:	Kathleen Fritsch, Ph.D.
Concurring Reviewer:	Mohamed Alosh, Ph.D.
Medical Division:	Division of Dermatology and Dental Products
Clinical Team:	Denise Cook, M.D. / Gordana Diglisic, M.D.
Project Manager:	Cristina Attinello

Keywords: Acne, multiplicity, missing data

Table of Contents

1 Executive Summary

Tazarotene foam 0.1% was superior to vehicle in the treatment of acne vulgaris in two studies. The studies enrolled subjects age 12 to 45 with an Investigator's Static Global Assessment (ISGA) of 3 (moderate) or greater, 25 to 50 facial inflammatory lesions, and 30 to 125 facial non-inflammatory lesions (excluding nasal lesions). The primary efficacy endpoints were the absolute change in lesions (inflammatory, non-inflammatory, and total), and ISGA success (two definition of success were defined: at least two grades reduction from baseline and achieving clear or almost clear). The studies were designed so that a successful outcome was defined as achieving statistical significance for two out of three lesion types and both definitions of ISGA success. The protocol stated that multiplicity with regard to the lesion count assessments would be handled with Holm's method. Although no additional details about how Holm's method would be applied were included in the protocol or statistical analysis plan, the p-values for the reduction in lesion count endpoints are statistically significant if the usual application of Holm's method is applied (smallest p-value compared to $\alpha/3$, next smallest compared to $\alpha/2$, and largest compared to α). The p-values for all 5 primary efficacy endpoints are less than 0.001 in both studies and therefore met the protocol-specified criteria for establishing efficacy. The efficacy results are summarized in Table 1.

Table 1 – Primary Efficacy Endpoints at Week 12 (ITT)

Endpoint	Study 301			Study 302		
	Tazarotene N=371	Vehicle N=372	p-value	Tazarotene N=373	Vehicle N=369	p-value
Absolute change[1]:						
Inflammatory	-17.6	-13.3	<0.001	-17.6	-14.3	<0.001
Non-inflammatory	-28.1	-16.2	<0.001	-25.9	-17.0	<0.001
Total	-45.8	-29.5	<0.001	-43.5	-31.3	<0.001
ISGA:						
2-Grade Improvement	132 (36%)	89 (24%)	<0.001	120 (32%)	67 (18%)	<0.001
Clear or Almost Clear	107 (29%)	60 (16%)	<0.001	103 (28%)	49 (13%)	<0.001

[1]Least squares means

Because the protocol does not include details about the specific hypotheses for the endpoints or exactly how Holm's method was intended to be used, it is not entirely clear what errors the method is attempting to control. Assuming that the statement that the absolute reduction in at least two out of three lesion types is statistically significant describes a clinically meaningful description regarding reductions in lesions, we can conclude that the Studies 301 and 302 have met their efficacy objectives.

2 Introduction

2.1 Overview

Tazarotene is in the retinoid drug class. The applicant (Stiefel) is seeking approval for tazarotene foam 0.1% for the topical treatment of acne vulgaris in patients 12 years of age and older. Two topical tazarotene formulations previously have been approved for the

treatment of acne. Tazaorac gel 0.1% was approved in 1997 and Tazorac cream 0.1% was approved in 2001. Both Tazorac formulations are owned by Allergan. This application for tazarotene foam 0.1% is a 505(b)(1) application with right of reference to Tazorac gel and cream. The applicant intends to rely on some of the findings of safety from pharmacology, toxicology, and pharmacokinetics studies conducted for the Tazorac gel NDA. To support the efficacy and safety of tazarotene foam 0.1%, the applicant has conducted two Phase 3 vehicle-controlled studies of tazarotene foam versus vehicle foam in the treatment of acne. The two studies enrolled 1485 subjects age 12 to 45 with acne vulgaris. An overview of the two studies is presented in Table 2. Study 301 was conducted at 21 centers, of which 18 were in the U.S. (667 subjects) and 3 were in Canada (76 subjects). Study 302 was conducted at 18 centers, of which 14 were in the U.S. (568 subjects) and 4 were in Canada (174 subjects).

Table 2 – Clinical Studies Overview

Study Numbers	W0260-301 and W0260-302		
Study Design	Randomized, double-blind vehicle-controlled		
Inclusion criteria	Age 12-45, 25-50 inflammatory lesions, 30-125 non-inflammatory lesions, ISGA ≥ 3		
Treatment regimen	Once daily for 12 weeks		
Primary endpoints	(1) absolute change in 2 out or 3 lesion counts (total, inflammatory, and non-inflammatory), (2) the proportion of subjects with at least a 2-grade improvement in ISGA score, and (3) the proportion of subjects with and ISGA score of 0 or 1		
Treatment arms and Sample Size		*301*	*302*
	Tazarotene	371	373
	Vehicle	372	369
Study location	United States and Canada		

The applicant opened the IND for tazarotene foam 0.1% on 8/28/2009 with the protocols for Studies 301 and 302, which were reviewed by the Agency. The protocols were not reviewed under a Special Protocol Assessment. The only meeting held with the sponsor was the Pre-NDA meeting held on 6/15/2011.

2.2 Data Sources

This reviewer evaluated the applicant's clinical study reports, datasets, clinical summaries, and proposed labeling. This submission was submitted in eCTD format and was entirely electronic. Both SDTM and analysis datasets were submitted. The analysis datasets used in this review are archived at \\cdsesub1\EVSPROD\NDA202428\0000\m5\datasets.

3 Statistical Evaluation

3.1 Data and Analysis Quality

The databases for the studies required minimal data management prior to performing analyses and no requests for additional datasets were made to the applicant.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Statistical Analysis

Studies 301 and 302 were randomized, double-blind, vehicle-controlled studies of the efficacy and safety of tazarotene foam in the treatment of acne. Subjects applied study product to their entire face once daily for 12 weeks. Study 301 enrolled 743 subjects (371 tazarotene, 372 vehicle) at 21 centers and Study 302 enrolled 742 subjects (373 tazarotene, 369 vehicle) at 18 centers. The studies enrolled subjects age 12 to 45 with an Investigator's Static Global Assessment (ISGA) of 3 (moderate) or greater, 25 to 50 facial inflammatory lesions, and 30 to 125 facial non-inflammatory lesions (excluding nasal lesions). The ISGA scale is presented in Table 3. Subjects were evaluated for efficacy at baseline and Weeks 2, 4, 8, and 12.

Table 3 – Investigator's Static Global Assessment (ISGA) Scale

Grade	**Description**	
0	Clear	Clear skin with no inflammatory or non-inflammatory lesions.
1	Almost clear	Rare non-inflammatory lesions with no more than rare papules.
2	Mild	Greater than Grade 1, some non-inflammatory lesions with no more than a few inflammatory lesions (papules/pustules only, no nodular lesions).
3	Moderate	Greater than Grade 2, up to many non-inflammatory lesions and may have some inflammatory lesions, but no more than one small nodular lesion.
4	Severe	Greater than Grade 3, up to many non-inflammatory and inflammatory lesions, but no more than a few nodular lesions.
5	Very severe	Many non-inflammatory and inflammatory lesions and more than a few nodular lesions. May have cystic lesions.

In the initial submission of Protocols 301 and 302 (submitted 8/28/2009), the sponsor proposed defining the co-primary endpoints in the following way:

1. The absolute change in lesion counts (total, inflammatory, and non-inflammatory) from baseline to Week 12 (end of treatment).
2. The proportion of subjects who have a minimum 2-grade improvement in ISGA score from baseline to Week 12 (end of treatment).

The clinical reviewer commented that "The second co-primary endpoint will need to include that the subjects must also attain a clear or almost clear on the ISGA severity scale. Thus, it will not be treated as a secondary endpoint." (Source: Clinical Review dated 10/22/2009) The proportion of subjects who have an ISGA score of 0 or 1 at Week 12 had been specified as a secondary endpoint. However, although the intention of the clinical comment was that the Agency's preferred ISGA analysis was the proportion of subjects who have an ISGA score of 0 or 1 plus a minimum 2-grade improvement from baseline at Week 12, the comment was conveyed to the sponsor (Advice/Information

Request dated 12/4/2009) in such a way that it implied that two separate ISGA analyses were recommended. The advice letter stated the following:

"Recommended co-primary endpoints are:

- The absolute change in lesion counts (total, inflammatory, and non-inflammatory) from baseline to Week 12 (end of treatment).
- The proportion of subjects who have a minimum 2-grade improvement in ISGA score from baseline to Week 12 (end of treatment).
- The proportion of subjects who have an ISGA score of 0 or 1 (clear or almost clear) at Week 12 (end of treatment)."

The sponsor modified the protocol to list three co-primary endpoints using the same wording as the Agency advice letter. Note, however, that because the inclusion criteria required subjects to have an ISGA of 3 or greater at baseline, all subjects who achieved a score of 0 or 1 at Week 12 would have had at least a 2-grade improvement from baseline. Thus the applicant's list of co-primary endpoints includes the endpoints that the Agency generally asks for in acne studies (absolute change in total, inflammatory, and non-inflammatory lesion counts and the proportion of subjects with as ISGA score of 0 or 1 (with at least a 2-grade improvement implied)) plus an additional endpoint defined as the proportion of subjects with a minimum 2-grade improvement in ISGA score. Thus, if the clinical studies demonstrate statistical significance on all three co-primary endpoints, the study will have met the statistical criteria that the Agency generally recommends for acne trials. However, if either of the studies fails to demonstrate statistical significance for the 2-grade improvement on the ISGA endpoint, there would be difficulties with interpreting the findings of the study, as the study would not have met its pre-specified success criteria.

The secondary endpoints were specified in the protocol as

1. The percent change in lesion counts (total, inflammatory, and non-inflammatory) from baseline to Weeks 2, 4, 8, and 12.
2. Time to 50% reduction in total lesion counts.
3. The absolute change in lesion counts (total, inflammatory, and non-inflammatory) from baseline to Weeks 2, 4, and 8.
4. The proportion of subjects who have a minimum 2-grade improvement in ISGA score from baseline to Weeks 2, 4, and 8.
5. The proportion of subjects who have an ISGA score of 0 or 1 at Weeks 2, 4, and 8.
6. The proportion of subjects who have an SGA score of 0 or 1 at Weeks 2, 4, 8, and 12.

The SGA or Subject's Global Assessment scale was defined as

Table 4 – Subject's Global Assessment (SGA) Scale

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

The ITT population was defined as all randomized subjects who were dispensed study product. The per protocol population excluded subjects who

- missed more than 16 product applications
- had more than 5 days between the date of last application and the date of the final efficacy evaluation
- did not have efficacy evaluations at baseline and week 12
- used prohibited medications expected to interfere with the efficacy assessment (including receiving more than10 consecutive days of any antibiotic)
- enrolled in error with respect to inclusion or exclusion criteria expected to impact the efficacy assessment (lesion counts, ISGA, prior medications)
- had other major violations that could have an impact on the efficacy analysis

The absolute change in total, inflammatory, and non-inflammatory lesion counts were to be analyzed using an ANCOVA model with terms for baseline value, treatment, center, and treatment-by-center interaction. If the treatment-by-center interaction was not significant at the 0.1 level, it was to be removed from the model. The ISGA response rate analyses were to be analyzed with a Cochran-Mantel-Haenszel test stratified by center. The Breslow-Day test will be performed to test for homogeneity across centers. Statistical tests for each endpoint will be conducted at the two-sided significance level of 0.05.

To draw the conclusion that the studies demonstrated efficacy, the following results need to be statistically significant

- 2 out of 3 lesion count endpoints, where Holm's method will be used to adjust for multiple comparisons
- 2-grade improvement in the ISGA score
- ISGA score of 0 or 1

The protocol did not provide any discussion or elaboration as to how Holm's method would be used to assess whether 2 out of 3 lesions counts were statistically significant. The general procedure for applying Holm's procedure to three hypotheses would be to order the p-values for the three lesion count endpoints and compare the smallest to 0.0167 (0.05/3), the next smallest to 0.025 (0.05/2) and the largest to 0.05. Because the

decision rule stated that 2 out of 3 lesion count endpoints need to demonstrate statistical significance, the smallest p-value would need to be less than 0.0167 and the second smallest p-value would need to be less than 0.025 in order to have a significant finding for the lesion count endpoints. The study reports for the two studies simply note that all lesion count endpoints had p-values <0.001 and therefore are statistically significant. The study reports make no attempt to explain whether Holm's method was actually used to assess whether the p-values were statistically significant or whether this reviewer's interpretation of how the method was to be applied is how the applicant used the method.

The secondary endpoints of percent change in lesion counts or absolute change in lesion counts were to be analyzed in the same way as the primary lesion count analyses. Similarly, the ISGA and SGA response secondary endpoints were to be analyzed the same way as the primary analyses. Time to 50% reduction in total lesion counts was to be analyzed using Kaplan-Meier estimates and the log-rank test. The protocol did not include any methods for controlling multiplicity among the secondary endpoints, although the sponsor was advised in the original comments on the protocols that for "all secondary endpoints intended for labeling claims the sponsor should include appropriate multiplicity adjustments." (Advice letter dated 12/4/2009)

For continuous endpoints, the primary method of handling missing data was last observation carried forward (LOCF). Ordinary least-squares multiple regression where the predicted mean value for each treatment was imputed was used as a sensitivity analysis. For the response endpoints, LOCF was the primary method of handling missing data and imputing missing values as failures was the sensitivity analysis.

3.2.2 Subject Disposition

Study 301 randomized 744 subjects to tazarotene (372) or vehicle (372), however, one subject (randomized to tazarotene) was not dispensed medication and therefore was not included in the ITT population. This subject had a positive pregnancy test at the baseline visit and was not dispensed medication. Study 302 randomized 742 subjects to tazarotene (373) or vehicle (369) and all were included in the ITT population. Discontinuation rates were higher on the tazarotene arm than the vehicle arm in each study with approximately 18% of tazarotene subjects discontinuing the study early compared to approximately 10% of vehicle subjects. Subjects were more likely to discontinue due to adverse events or withdrawal by subject on the tazarotene arm than the vehicle arm. The most common reason for discontinuation was subject request. See Table 5 and Table 6.

Table 5 – Disposition of Subjects (Study 301)

	Tazarotene	Vehicle
Subjects Randomized	372	372
Dispensed Medication (ITT)	371	372
Completed study	306 (82%)	333 (90%)
Discontinued study	65 (18%)	39 (10%)
Reasons for discontinuation		
Adverse Event	11 (3%)	1 (<1%)
Lost to follow-up	14 (4%)	14 (4%)
Noncompliance with study drug	1 (<1%)	1 (<1%)
Withdrawal by subject	32 (9%)	16 (4%)
Other	7 (2)%	7 (2%)

Source: pg. 86 of csr-w0260-301.pdf and reviewer analysis

Table 6 – Disposition of Subjects (Study 302)

	Tazarotene	Vehicle
Subjects Randomized	373	369
Dispensed Medication (ITT)	373	369
Completed study	307 (82%)	334 (91%)
Discontinued study	66 (18%)	35 (9%)
Reasons for discontinuation		
Adverse Event	9 (2%)	0 (0%)
Lost to follow-up	13 (3%)	11 (3%)
Lack of Efficacy	1 (<1%)	0 (0%)
Withdrawal by subject	39 (10%)	21 (6%)
Other	4 (1%)	3 (1%)

Source: pg. 84 of csr-w0260-302.pdf and reviewer analysis

3.2.3 Baseline Characteristics

Baseline demographics were generally balanced across the treatment groups in the two studies. The mean age of the subjects was approximately 18-19 years, with 55-60% of subjects aged 12 to 17 years. The studies had nearly equal proportions of male and female subjects. The majority of subjects were white (77%) and approximately 15% were black, with smaller proportions of Asians, American Indian/Alaskan natives, or other races. Approximately 18% of subjects reported their ethnicity as Hispanic/Latino. See Table 7.

Table 7 – Demographics

	Study 301		Study 302	
	Tazarotene N=371	Vehicle N=372	Tazarotene N=373	Vehicle N=369
Age (years)				
Mean	18.2	18.6	19.2	19.2
Range	12 - 43	12 - 44	12 - 45	12 - 45
12 to 17 years	223 (60%)	227 (61%)	205 (55%)	205 (56%)
18 to 25 years	104 (28%)	99 (27%)	117 (31%)	108 (29%)
26 to 45 years	44 (12%)	46 (12%)	51 (14%)	56 (15%)
Gender				
Male	189 (51%)	180 (48%)	176 (47%)	184 (50%)
Female	182 (49%)	192 (52%)	197 (53%)	185 (50%)
Race				
White	285 (77%)	307 (83%)	278 (75%)	280 (76%)
African-American	53 (14%)	50 (13%)	59 (16%)	57 (15%)
Asian	11 (3%)	4 (1%)	27 (7%)	21 (6%)
Amer. Indian/AK native	12 (3%)	3 (1%)	2 (1%)	2 (1%)
Other	10 (3%)	8 (2%)	7 (2%)	9 (2%)
Ethnicity				
Hispanic or Latino	62 (17%)	55 (15%)	71 (19%)	72 (20%)
Not Hispanic or Latino	309 (83%)	317 (85%)	302 (81%)	297 (80%)

Source: pg. 42 of csr-w0260-301.pdf and pg 42 of csr-w0260-302.pdf

Baseline lesion counts and ISGA scores were balanced across the treatment arms in both studies. Subjects in Study 301 had a mean of about 82 total lesions at baseline and 76% of subjects had a baseline ISGA score of moderate. Subjects in Study 302 had a mean of about 78 total lesions at baseline and 85% of subjects had a baseline ISGA score of moderate. See Table 8.

Table 8 – Baseline Disease Characteristics

	Study 301		Study 302	
	Tazarotene N=371	Vehicle N=372	Tazarotene N=373	Vehicle N=369
Mean Lesion Counts				
Total Lesions	81.5	81.7	77.3	78.7
Non-Inflammatory Lesions	50.1	49.8	45.2	46.2
Inflammatory Lesions	31.4	31.9	32.1	32.4
ISGA				
Moderate	282 (76%)	282 (76%)	317 (85%)	311 (84%)
Severe	89 (24%)	90 (24%)	56 (15%)	58 (16%)

Source: pg. 43-44 of csr-w0260-301.pdf and pg 43-44 of csr-w0260-302.pdf

3.2.4 Primary Efficacy Endpoints

The set of co-primary efficacy endpoints was defined in the protocol as

- The absolute change in lesion counts (total, inflammatory, and non-inflammatory) from baseline to Week 12 (end of treatment), where 2 out of 3 need to demonstrate statistical significance.
- The proportion of subjects who have a minimum 2-grade improvement in ISGA score from baseline to Week 12 (end of treatment).
- The proportion of subjects who have an ISGA score of 0 or 1 (clear or almost clear) at Week 12 (end of treatment).

Tazarotene foam was superior to vehicle foam in all 5 of the specified endpoints ($p < 0.001$). If we take into account the protocol specification that the absolute change in lesion count endpoints would be assessed using Holm's method, we find that all three lesion types had p-values < 0.0167 ($\alpha/3$) and therefore were statistically significant using Holm's method. The lesion reduction associated with the use of tazarotene foam versus vehicle was approximately a reduction of an additional 4 inflammatory lesions and an additional 8 non-inflammatory lesions. The ITT analysis results are presented in Table 9. Recall that because the inclusion criteria required a baseline ISGA score of 3 or greater, achieving a Week 12 score of clear or almost clear (0 or 1) would necessarily require at least 2 grades reduction. The lesion reduction endpoints were analyzed with an ANCOVA with terms for baseline value, treatment, center, and treatment-by-center interaction. The treatment-by-center interaction term was removed from the model if it was not significant at the 0.1 level. Among the lesion reduction analyses conducted for the two studies, the treatment-by-center interaction was retained in all analyses except for the change in inflammatory lesions endpoint in Study 302. The ISGA endpoints were analyzed with the Cochran-Mantel-Haenszel test stratified on center. The ITT analyses were conducted using LOCF imputation for missing data. The results in the per protocol population were similar to those in the ITT population, though the point estimates in each arm, as well as the magnitude of the treatment effect, were generally larger in the per protocol population, than the ITT population (see Table 10).

Table 9 – Primary Efficacy Endpoints at Week 12 (ITT)

Endpoint	Study 301			Study 302		
	Tazarotene N=371	Vehicle N=372	p-value	Tazarotene N=373	Vehicle N=369	p-value
Baseline Count						
Inflammatory	31.4	31.9		32.1	32.4	
Non-inflammatory	50.1	49.8		45.2	46.2	
Total	81.5	81.7		77.3	78.7	
Absolute change: Means						
Inflammatory	-18.0	-14.1		-17.8	-14.7	
Non-inflammatory	-27.9	-16.7		-25.6	-18.2	
Total	-45.8	-30.8		-43.3	-32.9	
Least Squares Means						
Inflammatory	-17.6	-13.3	<0.001	-17.6	-14.3	<0.001
Non-inflammatory	-28.1	-16.2	<0.001	-25.9	-17.0	<0.001
Total	-45.8	-29.5	<0.001	-43.5	-31.3	<0.001
ISGA:						
2-Grade Improvement	132 (36%)	89 (24%)	<0.001	120 (32%)	67 (18%)	<0.001
Clear or Almost Clear	107 (29%)	60 (16%)	<0.001	103 (28%)	49 (13%)	<0.001

Source: pg 47 of csr-w0260-301.pdf and pg 47 of csr-w0260-302.pdf and reviewer analysis

Table 10 – Primary Efficacy Endpoints at Week 12 (PP)

Endpoint	Study 301		Study 302	
	Tazarotene N=270	Vehicle N=318	Tazarotene N=276	Vehicle N=311
Absolute change: Least Squares Means				
Inflammatory	-20.4	-14.6	-20.4	-15.5
Non-inflammatory	-31.3	-17.0	-29.0	-18.5
Total	-51.8	-31.2	-49.6	-33.7
ISGA:				
2-Grade Improvement	112 (41%)	83 (26%)	106 (38%)	62 (20%)
Clear or Almost Clear	91 (34%)	56 (18%)	93 (34%)	48 (15%)

Source: pg 99, 115, 131, 147, 156 of csr-w0260-301.pdf and pg 98, 114, 130, 146, 155 of csr-w0260-302.pdf and reviewer analysis

Whether or not multiplicity is adequately controlled among the lesion count endpoints depends upon how the hypotheses are constructed and how the hypotheses will be used to draw conclusions. The protocol does not spell out the specific hypotheses to be tested, only that the decision rule for the absolute change in lesion count endpoints was to demonstrate statistical significance for two out of three lesion types. The protocol stated that Holm's method would be used to control multiplicity, however the protocol did not specify details on how the method would be applied or how it would relate to the decision rule or any other conclusions. Note that total lesions are defined as the sum of the

inflammatory and non-inflammatory lesions and thus there are really only two lesion counts (inflammatory and non-inflammatory) plus a function of the two types (total).

Due to the structural relationship among the lesion types only a limited number of possible lesion count configurations are possible under the paradigm 'the lesion reduction treatment effect for at least two out of three types is greater than 0'. If the reduction in total lesions treatment effect is greater than 0 then at least one of the two lesion subtypes must also have a reduction treatment effect greater than 0 (the sum of two negative numbers cannot be a positive number). If one subtype experiences a reduction in lesions while the other subtype experiences an increase in lesion, then the total will experience a reduction as long as the reduction in the one subtype exceeds the increase in the other type (so that the sum is positive). Thus, when a study decision rule states that two out of three lesion types must have reductions that are statistically significant, that means that the study was adequately powered to detect the treatment effect for the reduction in total lesions and the treatment effect for the reduction in least one of the lesion subtypes. Though note that if the correct decision was made on the reduction in total lesions (i.e. the effect for the reduction in total lesions is real), then necessarily at least one of the subtypes must also have a real positive treatment effect, even if the study was not adequately powered to detect it. Note that the structure of the 'two out of three' testing paradigm only guarantees that any potential increase in lesion counts for one lesion subtype is slightly smaller than the decrease in lesion counts for the other type. The likelihood of an extreme scenario where one type of lesions greatly improves and the other type greatly worsens (relative to the vehicle) is not controlled by the decision rule for the lesion count endpoints itself, but rather by the natural limitations of the disease and the likelihood (presumably low based on experience with the disease) that the two lesion types would have changes that are so strongly negatively correlated. Note also that because the full decision rule for acne studies also requires success on the ISGA to be demonstrated, that this additional endpoint (subjects must be clear or almost clear to be classified as a success) mitigates the chance that a drug that worsens one type of lesions could be considered a success.

In addition to the decision rule for the lesion count endpoints requiring statistical significance for the reductions for two out of three lesion types, the protocol also stated that Holm's procedure would be applied to the lesion count endpoints, though the protocol did not include any details about how the multiplicity procedure would be applied. Because Holm's procedure is generally applicable in settings where the goal is to demonstrate that at least one of a set of endpoints is statistically significant, if all three endpoints meet the Holm's criteria then we can state that all three endpoints are statistically significant. However, this procedure would be overly conservative in the acne setting where one of the endpoints is the sum of the other two endpoints, as it does not take into account the structural relationships between the endpoints and the corresponding logical restrictions on the conclusions. Studies should be designed with careful thought about what hypotheses should be tested so that the requirements are clinically meaningful. Multiplicity procedures should be planned with a clear idea of the how they will be applied to control the Type I error. Vague hypotheses and unclear roles of multiplicity procedures leave questions about how to interpret the conclusions. If the

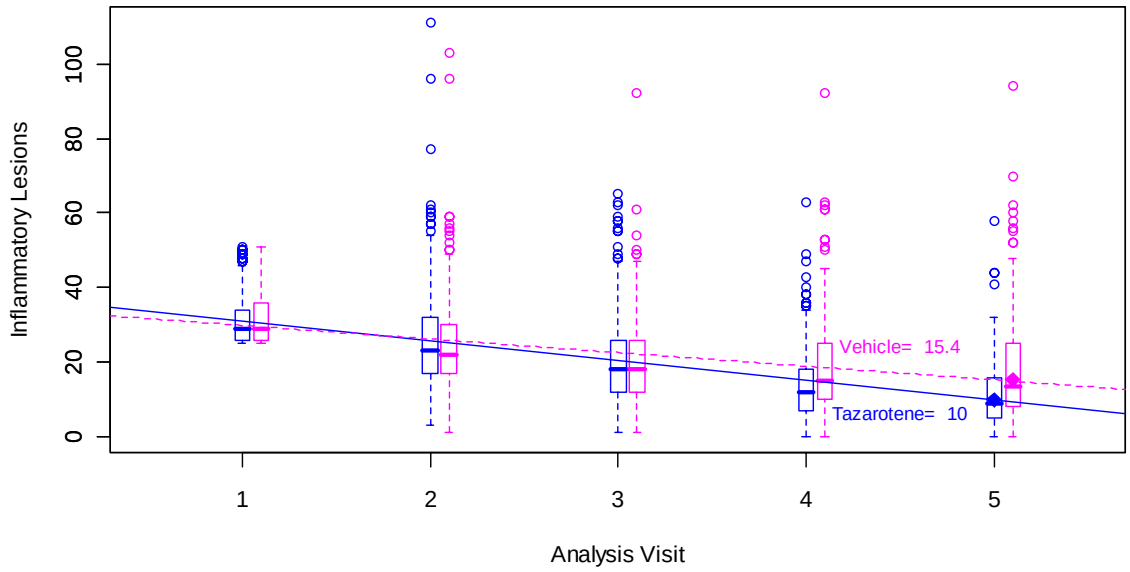
hypotheses do not match the clinical questions of interest, the study would not be designed to adequately assess the treatment effect. Inappropriate use of multiplicity procedures can lead to either underpowered tests (if the procedure is too conservative) or lead to inadequate Type I error control (if the procedure is not appropriate for the hypotheses).

3.2.5 Missing Data Handling

Approximately 18% of tazarotene subjects and 10% of vehicle subjects discontinued the study prior to Week 12. The protocol specified LOCF as the primary method of imputation for missing data for both the lesion count and ISGA analyses. The protocol also specified one missing data sensitivity analysis for each endpoint. For the ISGA endpoints, treating missing subjects as failures was specified as the sensitivity analysis. For the lesion count endpoints, ordinary least-squares multiple regression was specified as the sensitivity analysis. The original versions of Protocols 301 and 302 did not include any sensitivity analyses for handling missing data. The sponsor was advised (Advice letter dated 12/4/2009) to propose sensitivity analyses for handling missing data. The statistical analysis plan included a brief description of the proposed regression method ('an ordinary least-squares multiple regression model will be used to impute the predicted mean value for each treatment') but did include specific details about the implementation. The details regarding the regression model imputation were included only in SAS programs submitted with the NDA that were used by the applicant to construct the analysis data sets for Studies 301 and 302.

The regression model analysis used the observed lesion count values from each visit and fit a linear model for the lesion counts with analysis visit (coded as 1 through 5) as the independent variable. A separate model was fit for each treatment arm. Missing values were imputed using the predicted value for each treatment arm for the Week 12 visit (Visit 5). Note that this regression imputation assigns all subjects on a particular treatment arm with missing data at Week 12 with the same value. To illustrate this method, Figure 1 displays the observed inflammatory lesion counts by visit, the regression lines fit to the observed inflammatory lesion counts (with analysis visit as the independent variable), and the predicted Week 12 (Visit 5) values for the two treatment arms in Study 301. Imputed counts were rounded to the nearest integer. So in this example, all tazarotene subjects with missing Week 12 inflammatory lesion counts had an imputed Week 12 count of 10 and all vehicle subjects with missing Week 12 inflammatory lesion counts had an imputed Week 12 count of 15. Missing values for non-inflammatory lesion counts were imputed similarly.

Figure 1 – Least Squares Regression Imputation Model for Inflammatory Lesion Counts for Study 301



One definition of a useful sensitivity analysis would be a methodology that, relative to an observed cases analysis, generally narrowed the observed treatment effect, increased standard errors, and yet had imputed values that were still within the range of plausibility. Comparisons of the point estimates for the efficacy endpoints from observed case analysis, LOCF, regression imputation (for lesion count endpoints), and missing as failure (for ISGA endpoints) are presented in Table 11 and Table 12. All of the p-values for the imputation sensitivity analyses, like those for the primary LOCF imputation, are <0.001. Note that the regression imputation method for the absolute change endpoints has similar treatment effect estimates to the observed case analysis while also having smaller standard errors. These characteristics could be anticipated by the regression methodology which by design

- assumes that subjects who drop out all have the same outcome as the 'typical' subject who completed the study (leading to point estimates similar to the observed case analysis)
- assumes that there is no variability among the outcomes for subjects who did not complete the trial (and thus reducing standard errors by returning the sample size to its full value without increasing the standard deviation of the observations)

This methodology assumes that the set of subjects who drop out is a random sample from the total population, implying that the data is missing completely at random. However, in the studies the rate of dropout was higher on the tazarotene arm than the vehicle arm (18% vs. 10%), and subjects on the tazarotene arm were more likely to discontinue due to adverse events or subject request. Thus an assumption of data missing completely at random would not be reasonable and this method is likely to introduce bias that would favor a conclusion of efficacy for tazarotene. As such, the regression imputation method is not useful for assessing the impact of missing data on the conclusions of the study as it has the opposite effect on treatment effects and standard errors (no decrease in treatment

effects and reducing standard errors) than is desirable for a sensitivity analysis method for handling missing data. The LOCF method for these studies led to smaller treatment effect estimates than the observed case analysis, as most subjects had smaller reductions at the time they dropped out than the reductions observed in subjects who completed the study. However the LOCF method did not have much effect on the standard errors. For the ISGA endpoints, the missing as failure imputation leads to estimates which are similar to the LOCF analysis, with the missing as failure estimates generally within about 1% of the LOCF estimates. Both of these methods have lower point estimates and treatment effect estimates than the observed cases analysis which ignores subjects who dropped out.

Table 11 – Primary Efficacy Endpoints at Week 12 Using Different Imputation Methods (Observed Cases, LOCF, Regression Imputation or Missing as Failure) – Study 301

Endpoint	Observed		LOCF		Regression/MVF	
	Tazar N=304	Vehicle N=330	Tazar N=371	Vehicle N=372	Tazar N=371	Vehicle N=372
Abs. change						
Inflamm.	-19.9	-14.0	-17.6	-13.3	-20.2	-14.3
Diff. (SE)	-5.9 (0.9)		-4.3 (0.9)		-5.9 (0.7)	
Non-inflamm.	-30.9	-16.8	-28.1	-16.2	-31.6	-17.7
Diff. (SE)	-14.1 (1.5)		-11.9 (1.5)		-13.9 (1.4)	
Total	-51.0	-30.8	-45.8	-29.5	-52.0	-31.8
Diff. (SE)	-20.2 (2.0)		-16.2 (2.0)		-20.2 (1.8)	
ISGA:						
2-Grade	129	84	132	89	129	84
Improvement	(42%)	(25%)	(36%)	(24%)	(35%)	(23%)
Clear or	106	57	107	60	106	57
Almost Clear	(35%)	(17%)	(29%)	(16%)	(29%)	(15%)

MVF = Missing values as failure; SE = Standard error
Source: pg 98, 114, 130, 146, 155 of csr-w0260-301.pdf and reviewer analysis

Table 12 – Primary Efficacy Endpoints at Week 12 Using Different Imputation Methods (Observed Cases, LOCF, Regression Imputation or Missing as Failure) – Study 302

Endpoint	Observed		LOCF		Regression/MVF	
	Tazar N=307	Vehicle N=331	Tazar N=373	Vehicle N=369	Tazar N=373	Vehicle N=369
Abs. change						
Inflamm.	-20.2	-15.3	-17.6	-14.3	-20.6	-15.4
Diff. (SE)	-4.9 (0.7)		-3.3 (0.7)		-5.2 (0.6)	
Non-inflamm.	-29.1	-18.2	-25.9	-17.0	-29.2	-19.1
Diff. (SE)	-10.9 (1.1)		-8.9 (1.1)		-10.1 (1.0)	
Total	-49.5	-33.3	-43.5	-31.3	-49.3	-35.1
Diff. (SE)	-16.2 (1.6)		-12.2 (1.6)		-14.3 (1.3)	
ISGA:						
2-Grade Improvement	116 (38%)	67 (20%)	120 (32%)	67 (18%)	116 (31%)	76 (18%)
Clear or Almost Clear	100 (33%)	49 (15%)	103 (28%)	49 (13%)	100 (27%)	49 (13%)

MVF = Missing values as failure; SE = Standard error
Source: pg 97, 113, 129, 145, 154 of csr-w0260-302.pdf and reviewer analysis

3.2.6 Secondary Efficacy Endpoints

The protocols included a number of secondary endpoints at various timepoints. These were

1. The percent change in lesion counts (total, inflammatory, and non-inflammatory) from baseline to Weeks 2, 4, 8, and 12.
2. Time to 50% reduction in total lesion counts.
3. The absolute change in lesion counts (total, inflammatory, and non-inflammatory) from baseline to Weeks 2, 4, and 8.
4. The proportion of subjects who have a minimum 2-grade improvement in ISGA score from baseline to Weeks 2, 4, and 8.
5. The proportion of subjects who have an ISGA score of 0 or 1 at Weeks 2, 4, and 8.
6. The proportion of subjects who have an SGA score of 0 or 1 at Weeks 2, 4, 8, and 12.

Although the applicant was advised that Protocols 301 and 301 should include multiplicity adjustments for the set of secondary endpoints (see Advice Letter for IND 105564 dated 12/9/2009), the applicant did not include any multiplicity adjustments for the secondary endpoints in the protocols or analyses. Typically in labeling, percent reduction in lesions counts has been included as a supportive analysis for the absolute reduction in lesions. Without a protocol that included adjustments for secondary endpoints, it would not be appropriate to include other secondary endpoints in labeling, because the Type I error rate was not controlled. Efficacy over time as a supportive analysis is assessed in the following section.

Results for the percent change in lesion counts at Week 12 and the time to 50% reduction in total lesion counts are presented in Table 13. The mean percent reduction in inflammatory lesions was approximately 55% for tazarotene versus 44% for vehicle in both studies, while the mean percent reduction in non-inflammatory lesions was approximately 56% versus 34% in Study 301 and 56% versus 41% in Study 302. The median time to 50% reduction in total lesions was 8 weeks for tazarotene and 12 weeks for vehicle.

Table 13 – Secondary Efficacy Endpoints (Percent Change in Lesion Counts at Week 12 and Time to 50% Reduction in Total Lesion Counts)

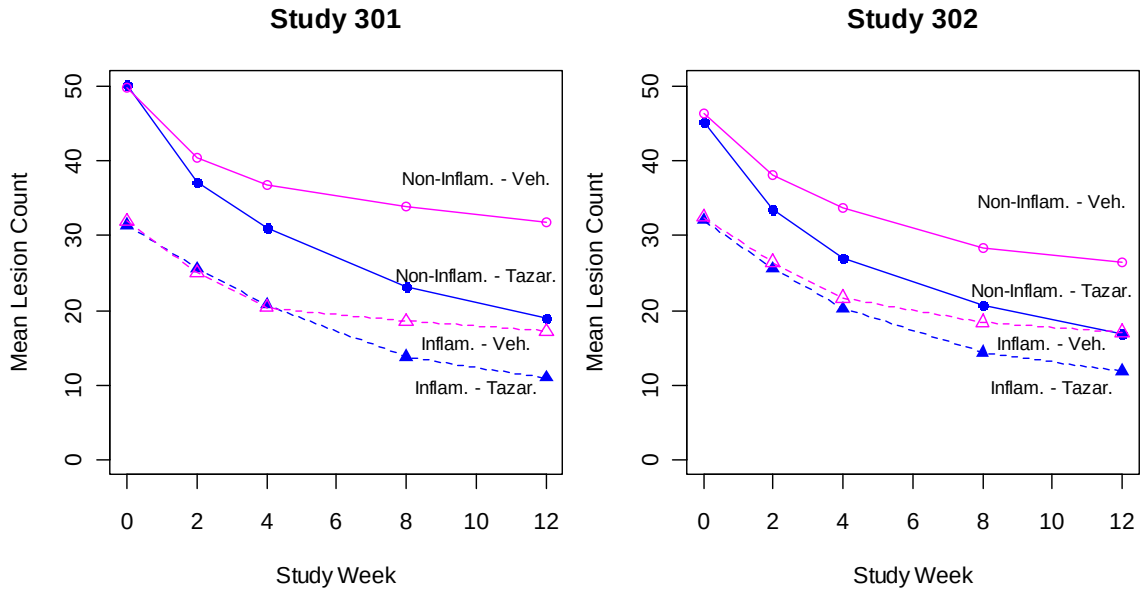
Endpoint	Study 301			Study 302		
	Tazarotene N=371	Vehicle N=372	p-value	Tazarotene N=373	Vehicle N=369	p-value
Percent change: Least Squares Means						
Inflammatory	-56.1%	-43.6%	<0.001	-53.7%	-44.3%	<0.001
Non-inflammatory	-56.2%	-33.2%	<0.001	-56.1%	-40.7%	<0.001
Total	-56.4%	-37.6%	<0.001	-55.2%	-42.0%	<0.001
Median days to 50% reduction	57	85	<0.001	57	85	<0.001

Source: pg 51 and 54 of csr-w0260-301.pdf and pg 51 and 54 of csr-w0260-302.pdf and reviewer analysis

3.2.7 Efficacy over Time

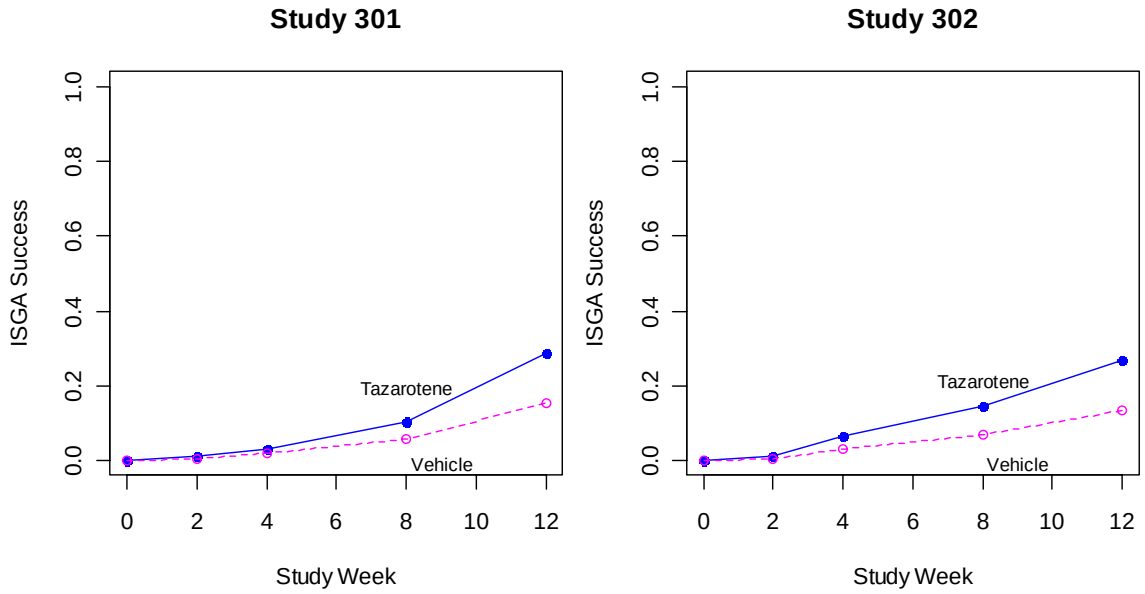
The mean inflammatory and non-inflammatory lesion counts decreased over time and the ISGA success rate (clear or almost clear) increased over time during the study. Although the mean lesion counts decreased on both treatment arms, the decrease was greater on the tazarotene arm than the vehicle arm. The separation between the treatment arms began earlier for non-inflammatory lesions than for inflammatory lesions. The mean lesion counts are presented in Figure 2 and the ISGA response rates are presented in Figure 3.

Figure 2 – Mean Lesion Counts over Time (Observed Cases)



Source: reviewer analysis

Figure 3 – ISGA Success (Clear or Almost Clear) over Time (Observed Cases)



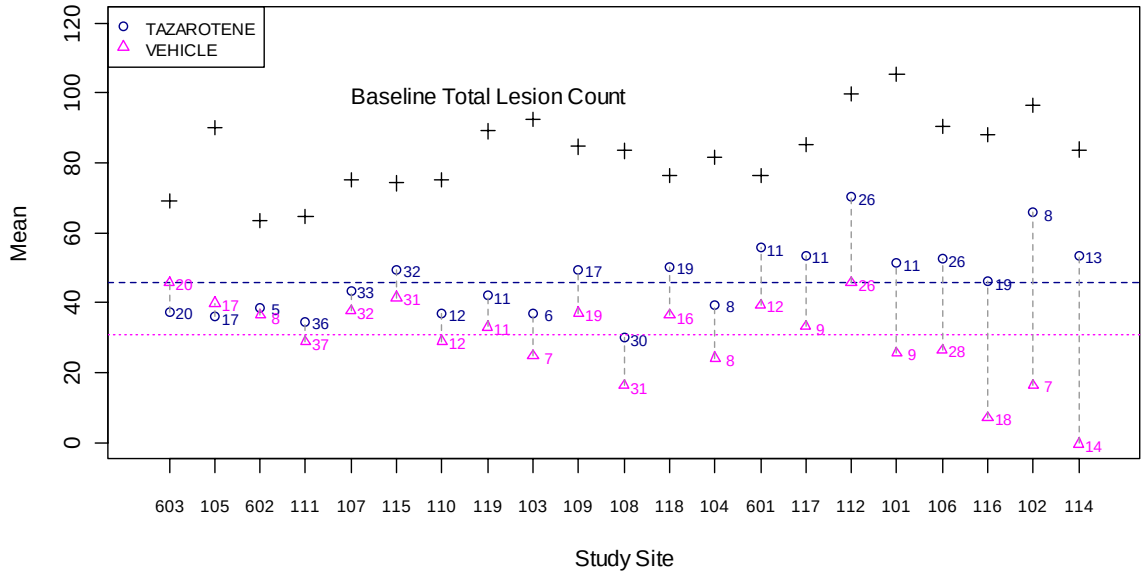
Source: reviewer analysis

3.2.8 Efficacy by Center

Treatment effects varied somewhat across centers, but no one center dominated the efficacy results. Sample sizes per center ranged from 10 to 73 subjects per center. The mean absolute change in total lesion count by center is presented in Figure 4 and Figure 5. Note that the baseline mean total lesion count varied somewhat across centers and that

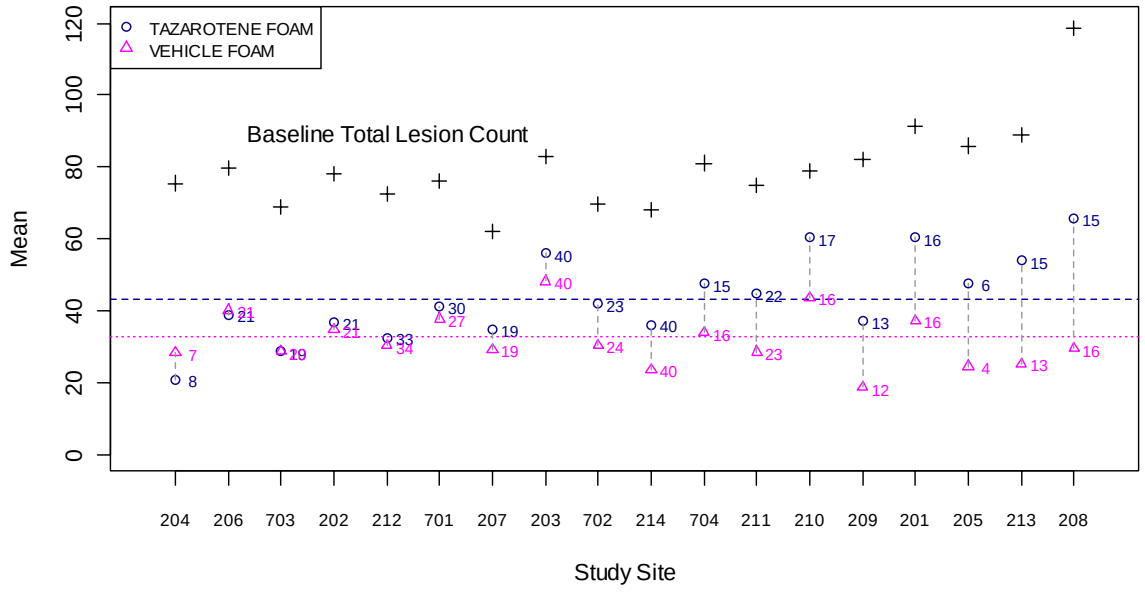
the centers with the larger observed treatment effects tended to be the centers with the larger total lesion counts at baseline. The proportions of subjects classified as ISGA success (clear or almost clear) by center are presented in Figure 6 and Figure 7.

Figure 4 – Absolute Change in Total Lesions by Center (Study 301)



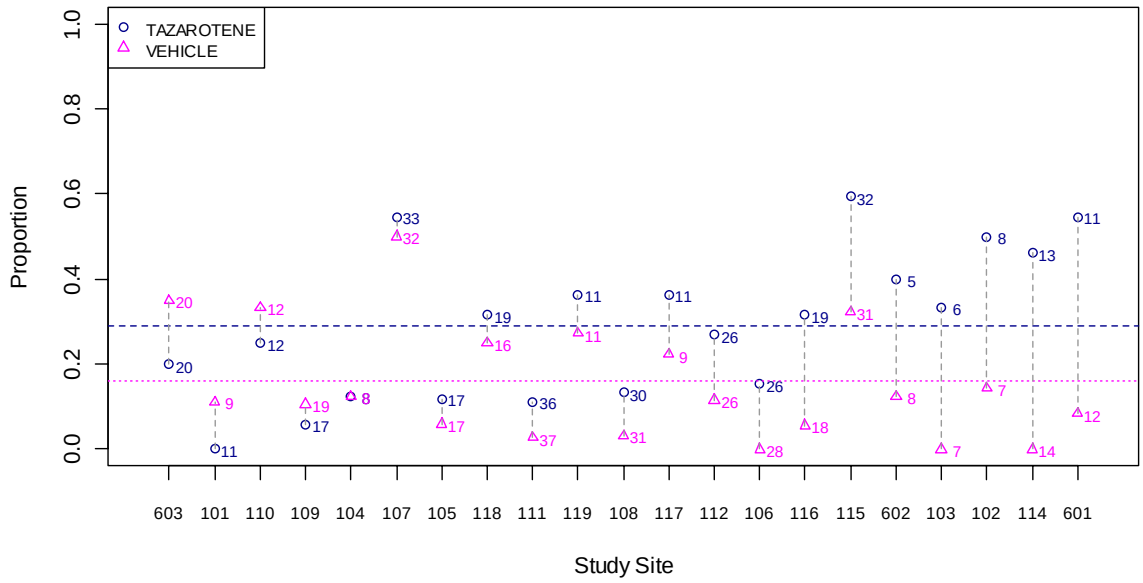
Source: Reviewer analysis

Figure 5 – Absolute Change in Total Lesions by Center (Study 302)



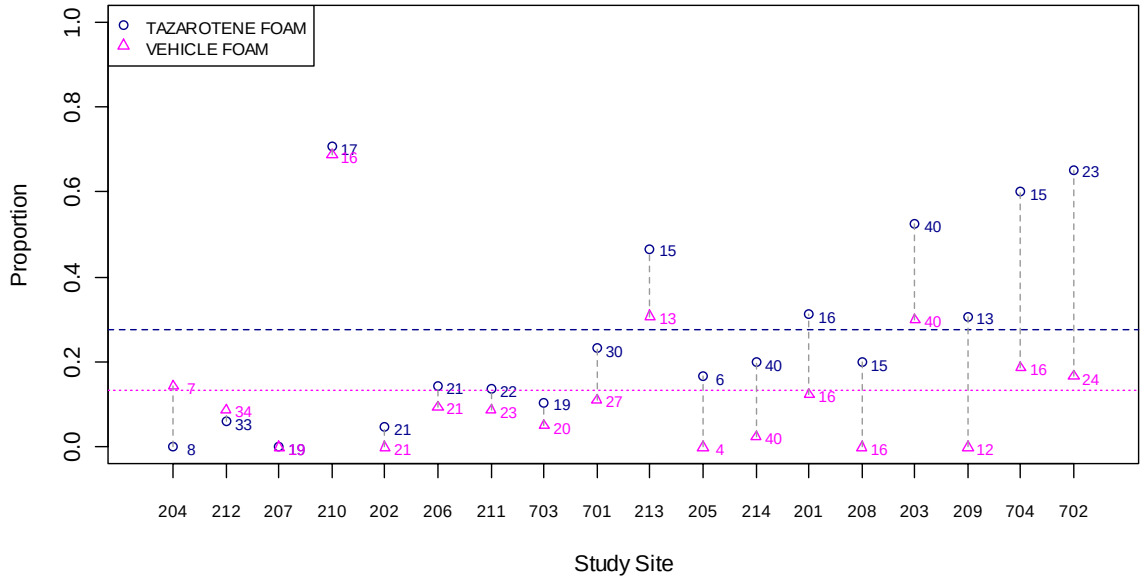
Source: Reviewer analysis

Figure 6 – ISGA Success Rate (Clear or Almost Clear) by Center (Study 301)



Source: Reviewer analysis

Figure 7 – ISGA Success Rate (Clear or Almost Clear) by Center (Study 302)



Source: Reviewer analysis

The observed variability across centers is reflected in the p-values from the ANCOVA models including the treatment-by-center interaction term for the absolute change in lesion endpoints and the from the Breslow-Day test for the ISGA success endpoints. All of the interaction p-values from Study 301 were <0.10 and in Study 302, the p-values for non-inflammatory lesions and total lesions were <0.10. See Table 14. However, only a small number of centers had observed treatment effects in the opposite direction and the differences among centers are more in terms of the magnitude of effect.

Table 14 – P-values for ANCOVA Treatment-by-Center Interaction and Breslow-Day Test for Homogeneity across Center

Endpoint	Study 301	Study 302
Absolute change:		
Inflammatory	0.0386	0.3596
Non-inflammatory	<0.0001	0.0012
Total	<0.0001	0.0113
ISGA:		
2-Grade Improvement	0.0312	0.2118
Clear or Almost Clear	0.0769	0.4316

Source: reviewer analysis

3.3 Evaluation of Safety

3.3.1 Extent of Exposure

Subjects on tazarotene had fewer total days of application of study product than subjects on vehicle. A greater number of subjects on tazarotene than vehicle discontinued the study. Also, a greater number of subjects on tazarotene had missed application days (mean difference of about 5 days between number of days of application and study product duration, which includes missed application days, for tazarotene versus about 3 missed days for vehicle). See Table 15.

Table 15 - Product Application Days (Excludes Intermittent Days of Missed Applications) and Product Duration (Does Not Exclude Intermittent Days of Missed Applications)

	Study 301		Study 302	
	Tazarotene N=371	Vehicle N=372	Tazarotene N=373	Vehicle N=369
Product Application Days				
≥ 4 weeks	319 (86%)	345 (93%)	319 (86%)	348 (94%)
≥ 8 weeks	296 (80%)	335 (90%)	301 (81%)	336 (91%)
≥ 11 weeks	224 (60%)	289 (78%)	212 (57%)	270 (73%)
Mean	68.5	76.9	67.4	76.2
Product Duration				
≥ 4 weeks	322 (87%)	349 (94%)	321 (86%)	348 (94%)
≥ 8 weeks	310 (84%)	336 (90%)	308 (83%)	339 (92%)
≥ 11 weeks	301 (81%)	327 (88%)	302 (81%)	332 (90%)
Mean	73.5	79.9	72.4	79.7

Source: pg 229 and 231 of csr-w0260-301.pdf and pg 230 and 232 of csr-w0260-302.pdf.

3.3.2 Adverse Events

A higher proportion of subjects on tazarotene than vehicle experienced adverse events (35-39% vs. 20%). The events that were observed at a higher rate in tazarotene subjects were application site reactions such as irritation, erythema, exfoliation, and dryness. The most common adverse events are presented in Table 16.

Table 16 – Adverse Events Occurring in at Least 3% of Subjects

	Study 301		Study 302	
	Tazarotene N=371	Vehicle N=372	Tazarotene N=373	Vehicle N=369
Any Adverse Event	145 (39%)	76 (20%)	132 (35%)	75 (20%)
Application site irritation	66 (18%)	5 (1%)	41 (11%)	4 (1%)
Application site erythema	32 (9%)	0	16 (4%)	2 (1%)
Application site exfoliation	28 (8%)	1 (<1%)	16 (4%)	2 (1%)
Application site dryness	22 (6%)	2 (1%)	28 (8%)	6 (2%)
Upper resp. tract infection	14 (4%)	14 (4%)	9 (2%)	13 (4%)
Nasopharyngitis	7 (2%)	8 (2%)	15 (4%)	12 (3%)

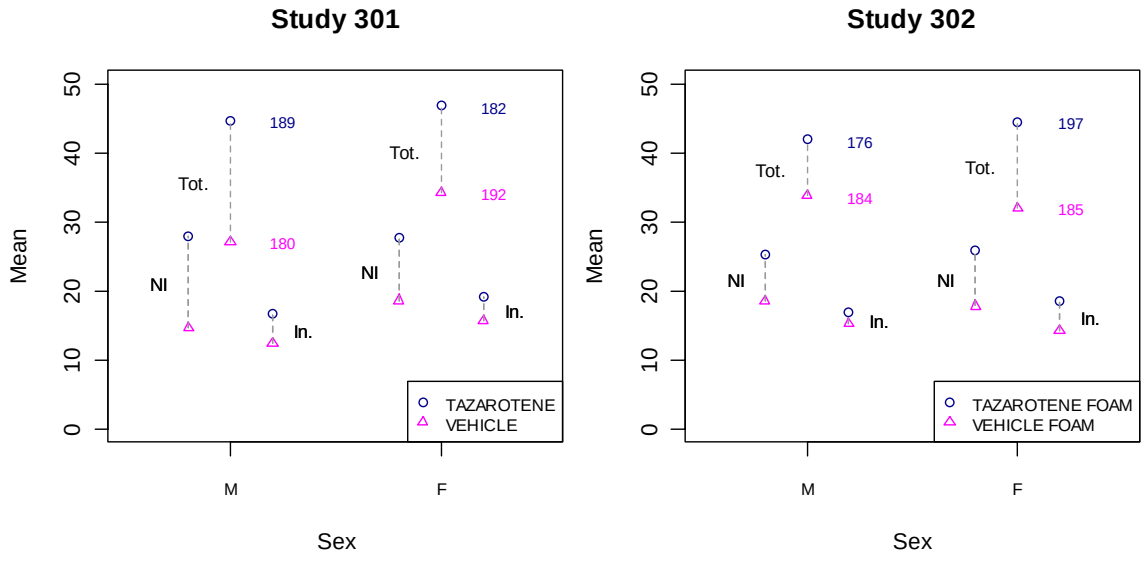
Source: pg 69 of csr-w0260-301.pdf and pg 69 of csr-w0260-302.pdf.

4 Findings in Special/Subgroup Populations

4.1 Gender, Race, Age, and Geographic Region

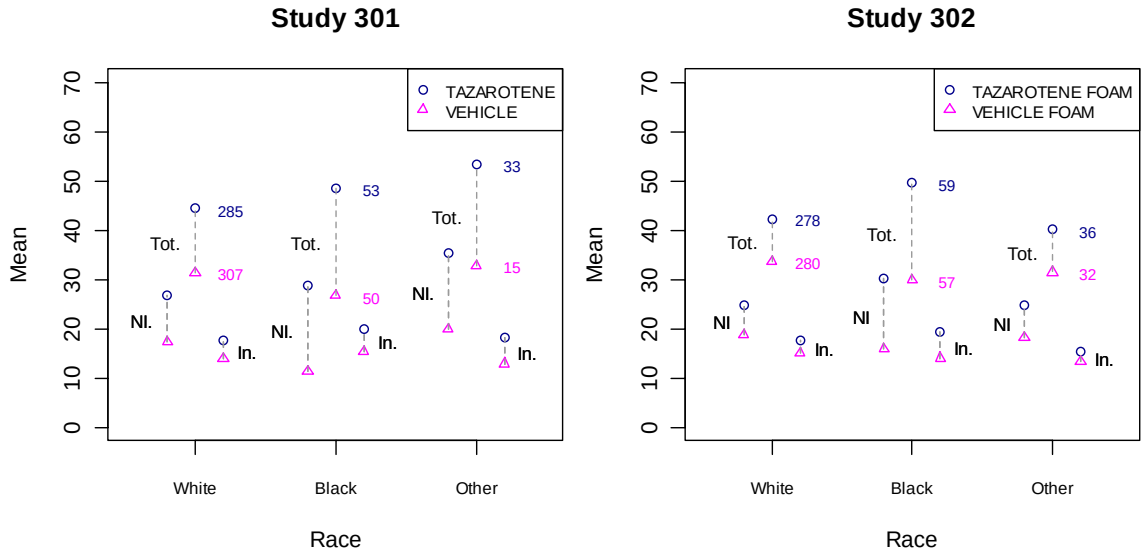
The treatment effects were generally consistent across gender, race, age and geographic region subgroups. The results for the absolute change in lesion counts are presented in Figure 8 through Figure 11. The results for the success on the ISGA were similar, and are presented in the Appendix.

Figure 8 – Absolute Change in Lesions by Sex



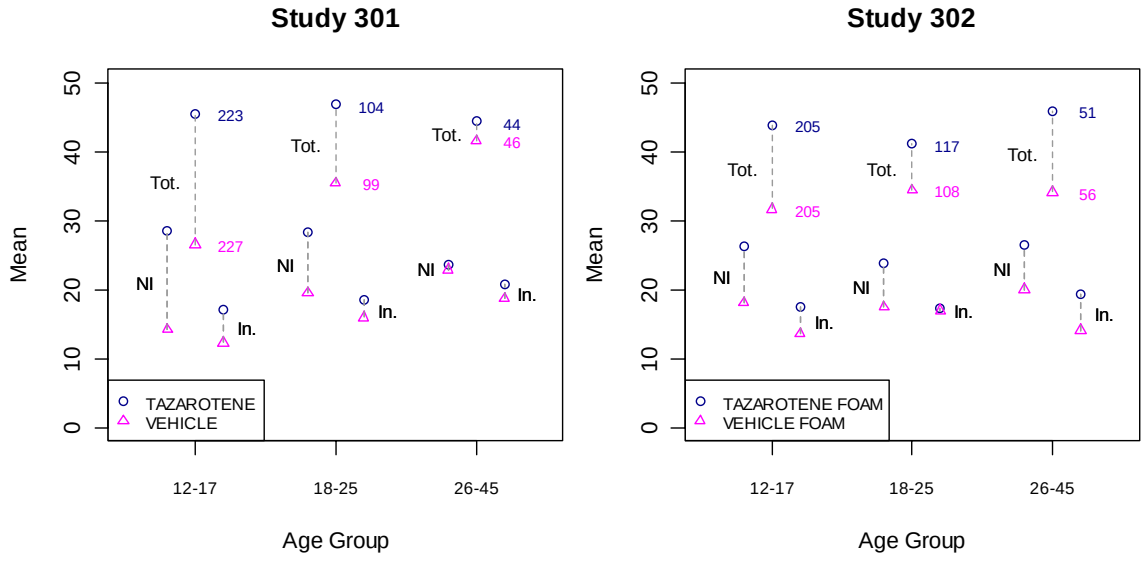
Source: Reviewer analysis

Figure 9 – Absolute Change in Lesions by Race



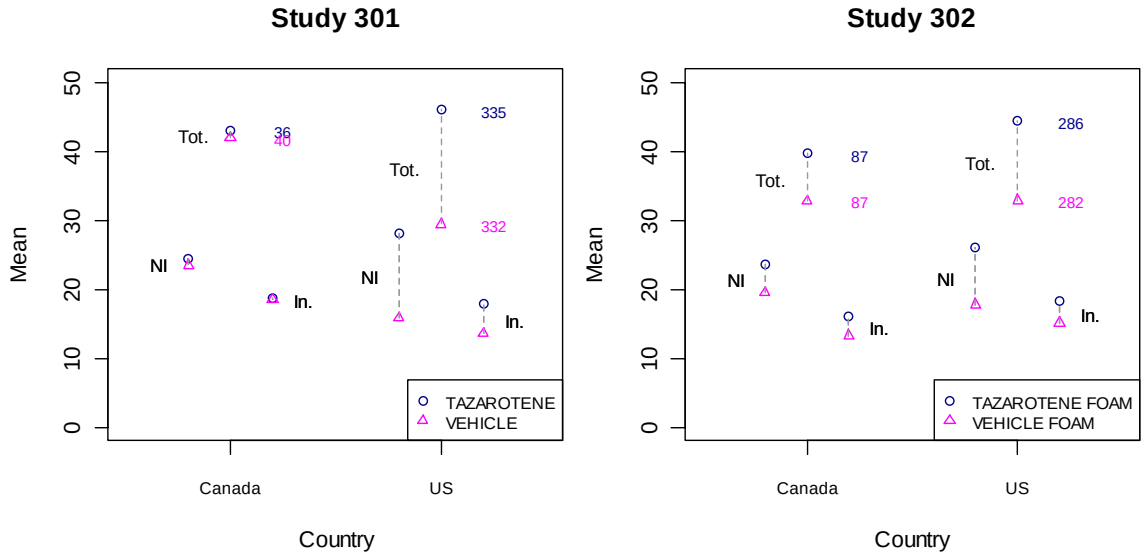
Source: Reviewer analysis

Figure 10 – Absolute Change in Lesions by Age Group



Source: Reviewer analysis

Figure 11 – Absolute Change in Lesions by Country



Source: Reviewer analysis

4.2 Other Special/Subgroup Populations

Not applicable.

5 Summary and Conclusions

5.1 Statistical Issues and Collective Evidence

The applicant has evaluated the efficacy of tazarotene foam 0.1% in two vehicle controlled studies. The studies met the criteria of the protocol-specified decision rule of achieving statistical significance for two out of three lesion types (inflammatory, non-inflammatory, and total) and two definitions of ISGA success ([1] clear or almost clear and [2] at least two grades reduction from baseline). All p-values were less than 0.001.

The protocol stated that multiplicity for the lesion count assessments would be handled with Holm's method. Although no additional details about how Holm's method would be applied were included in the protocol or statistical analysis plan, the p-values are statistically significant if the usual application of Holm's method is applied (smallest p-value compared to $\alpha/3$, next smallest compared to $\alpha/2$, and largest compared to α). Both studies met the requirements of the decision rule for the absolute reduction in lesions, namely, that two out of three reduction in lesion count endpoints were statistically significant. Even though the protocol specified a multiplicity control method for the three lesion count endpoints (Holm's), due to the lack of protocol details on hypotheses or the application of Holm's method, it is not entirely clear what specific claims would be associated with the tests for specific lesion types, and correspondingly what errors the method is attempting to control. Assuming that the statement that the absolute reduction in at least two out of three lesion types is statistically significant describes a clinically

meaningful description regarding reductions in lesions, we can conclude that the Studies 301 and 302 have met their efficacy objectives.

As a sensitivity analysis for the handling of missing data for the reduction in lesion endpoints, the applicant proposed a method based on a regression model analysis where a linear model for the lesion counts was fit with analysis visit as the independent variable. Missing values were imputed using the predicted value for each treatment arm for the Week 12 visit. This regression imputation assigns the same value to all subjects on a particular treatment arm with missing data at Week 12. This methodology tends to impute outcomes for subjects with missing data that could lead to favorable study outcomes because the method

- assumes that subjects who drop out all have the same outcome as the 'typical' subject who completed the study (leading to point estimates similar to the observed case analysis)
- assumes that there is no variability among the outcomes for subjects who did not complete the trial (and thus reducing standard errors by returning the sample size to its full value without increasing the standard deviation of the observations)

Because this imputation method has characteristics that are generally considered anti-conservative, its utility as a sensitivity analysis is limited and would not be recommended for use in other studies.

5.2 Conclusions and Recommendations

Tazarotene foam 0.1% was superior to vehicle in the treatment of acne vulgaris in two studies. The studies enrolled subjects age 12 to 45 with an Investigator's Static Global Assessment (ISGA) of 3 (moderate) or greater, 25 to 50 facial inflammatory lesions, and 30 to 125 facial non-inflammatory lesions (excluding nasal lesions). The primary efficacy endpoints were the absolute change in lesions (inflammatory, non-inflammatory, and total), and ISGA success (two definition of success were defined: at least two grades reduction from baseline and achieving clear or almost clear). The studies were designed so that a successful outcome was defined as achieving statistical significance for two out of three lesion types and both definitions of ISGA success. The protocol stated that multiplicity with regard to the lesion count assessments would be handled with Holm's method. The p-values for all 5 primary efficacy endpoints less than 0.001 in both studies and therefore met the protocol-specified criteria for establishing efficacy. The efficacy results are summarized in Table 17.

Table 17 – Primary Efficacy Endpoints at Week 12 (ITT)

Endpoint	Study 301			Study 302		
	Tazarotene N=371	Vehicle N=372	p-value	Tazarotene N=373	Vehicle N=369	p-value
Absolute change[1]*:*						
Inflammatory	-17.6	-13.3	<0.001	-17.6	-14.3	<0.001
Non-inflammatory	-28.1	-16.2	<0.001	-25.9	-17.0	<0.001
Total	-45.8	-29.5	<0.001	-43.5	-31.3	<0.001
ISGA:						
2-Grade Improvement	132 (36%)	89 (24%)	<0.001	120 (32%)	67 (18%)	<0.001
Clear or Almost Clear	107 (29%)	60 (16%)	<0.001	103 (28%)	49 (13%)	<0.001

[1]Least squares means

Appendix

Figure 12 – ISGA Success (Clear or Almost Clear) by Sex

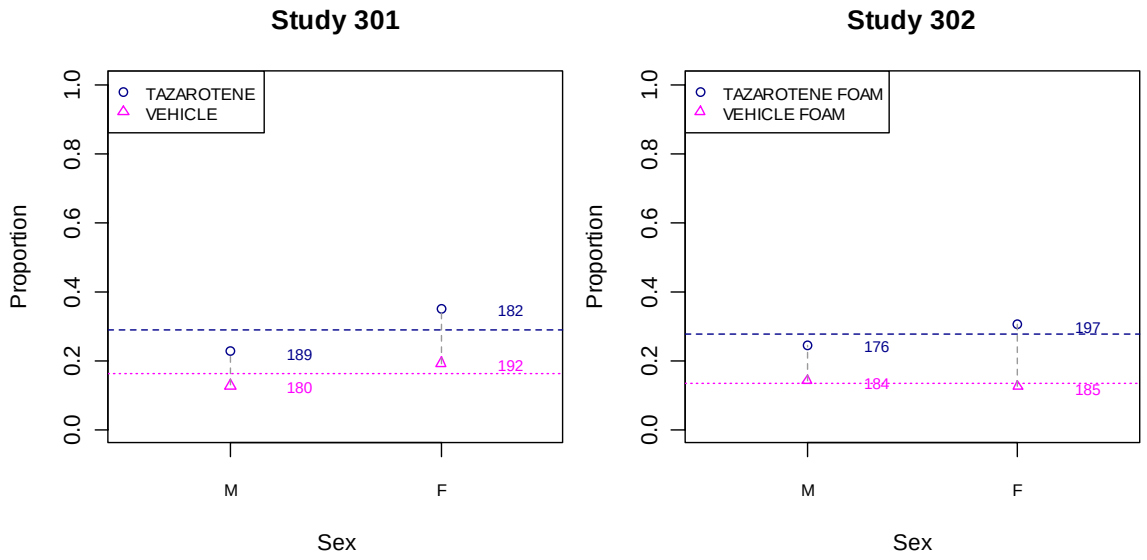


Figure 13 – ISGA Success (Clear or Almost Clear) by Race

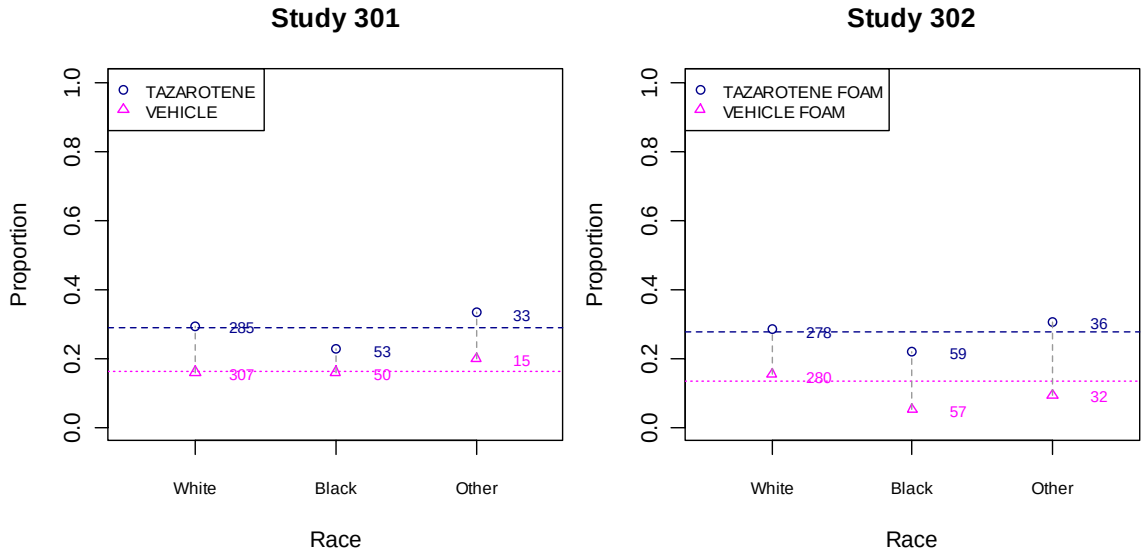


Figure 14 – ISGA Success (Clear or Almost Clear) by Age Group

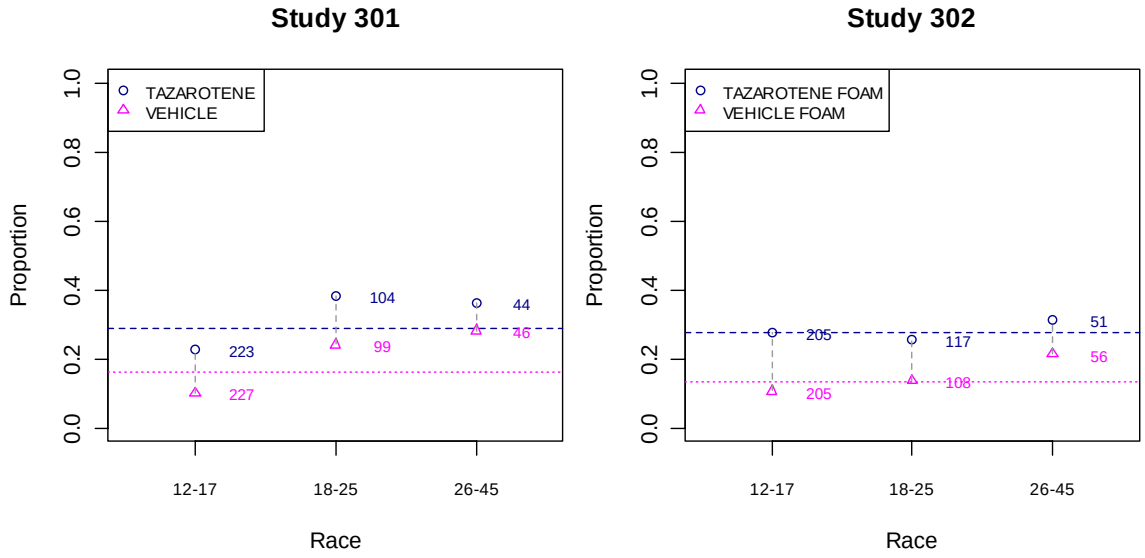
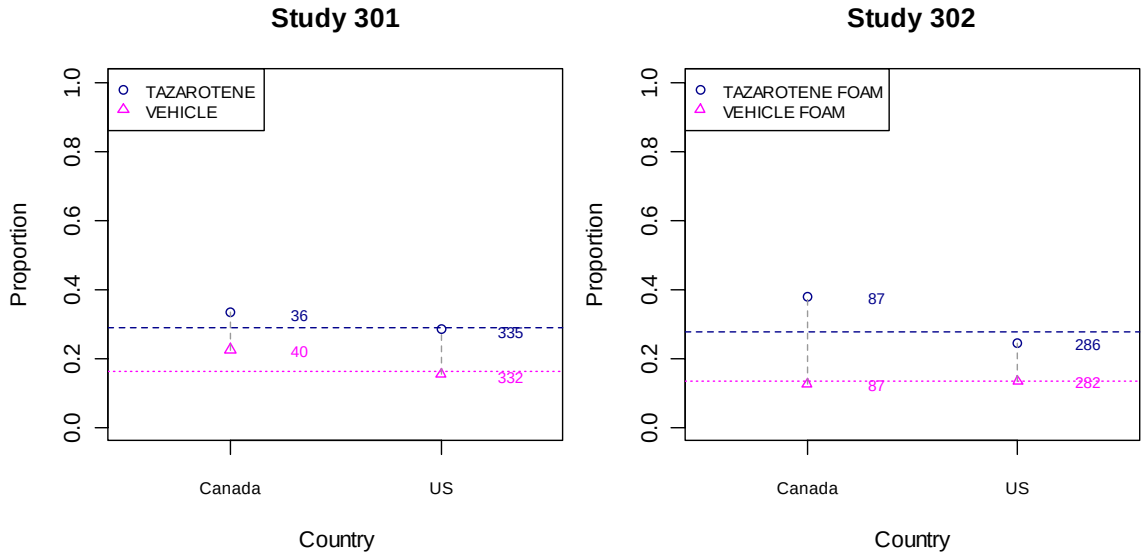


Figure 15– ISGA Success (Clear or Almost Clear) by Country



Signatures/Distribution List

Primary Statistical Reviewer: Kathleen Fritsch, Ph.D.
Date: 3/14/2012

Statistical Team Leader: Mohamed Alosh, Ph.D.

cc:
DDDP/Walker
DDDP/Diglisic
DDDP/Cook
DDDP/Attinello
OBIO/Patrician
DBIII/Wilson
DBIII/Alosh
DBIII/Fritsch

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

/s/

KATHLEEN S FRITSCH
03/14/2012

MOHAMED A ALOSH
03/14/2012
Concur with the review

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

NDA Number: 202428 **Applicant:** Stiefel **Stamp Date:** 7/29/11

Drug Name: Tazarotene Foam **NDA/BLA Type:** 505(b)(1) **Indication:** Acne

SUBMISSION SUMMARY

This submission contains two Phase 3 studies for tazarotene foam 0.1% vs. vehicle in the treatment of acne. Study 301 enrolled 744 subjects (372 tazaotene/372 vehicle) and Study 302 enrolled 742 subjects (373 tazarotene/369 vehicle). Both studies enrolled subjects age 12 and older with 25-50 inflammatory lesions, 30-125 non-inflammatory lesions and an ISGA of 3 or greater. Treatment was applied once daily to the face for 12 weeks. The co-primary efficacy endpoints were (1) absolute change in 2 out or 3 lesion counts (total, inflammatory, and non-inflammatory), (2) the proportion of subjects with at least a 2-trade improvement in ISGA score, and (3) the proportion of subjects with and ISGA score of 0 or 1. The applicant has secured a right of reference to NDAs 20600 and 21184 for Tazaorac gel and cream. The associated IND is 105564.

EFFICACY RESULTS SUMMARY

Endpoint	Study 301			Study 302		
	Tazarotene N=371	Vehicle N=372	p-value	Tazarotene N=373	Vehicle N=369	p-value
Absolute change:						
Inflammatory	-18.0	-14.1	<0.001	-17.8	-14.7	<0.001
Non-inflammatory	-27.9	-16.7	<0.001	-25.6	-18.2	<0.001
Total	-45.8	-30.8	<0.001	-43.3	-32.9	<0.001
ISGA:						
2-Grade Improvement	132 (36%)	89 (24%)	<0.001	120 (32%)	67 (18%)	<0.001
Clear or Almost Clear	107 (29%)	60 (16%)	<0.001	103 (28%)	49 (13%)	<0.001

I. On **initial** overview of the NDA/BLA application identify and list any potential Refuse to File issues:

	Content Parameter for RTF	Yes	No	NA	Comments
1	Indexing and reference links within the electronic submission are sufficient to permit navigation through the submission, including access to reports, tables, data, etc.	X			
2	ISS, ISE, and complete study reports are available (including original protocols, subsequent amendments, etc.)	X			
3	Safety and efficacy were investigated for gender, racial, and geriatric subgroups investigated.	X			
4	Data sets in EDR are accessible and conform to applicable guidances (e.g., existence of define.pdf file for data sets).	X			

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE?

Yes.

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

II. Identify and list any potential review issues to be forwarded to the Applicant for the 74-day letter.

Content Parameter (possible review concerns for 74-day letter)	Yes	No	NA	Comment
Designs utilized are appropriate for the indications requested.	X			
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.	X			
Interim analyses (if present) were pre-specified in the protocol and appropriate adjustments in significance level made. DSMB meeting minutes and data are available.			X	
Appropriate references for novel statistical methodology (if present) are included.			X	
Safety data organized to permit analyses across clinical trials in the NDA/BLA.	X			
Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.	X			

74-DAY LETTER REQUESTS TO THE APPLICANT

None.

Kathleen Fritsch, Ph.D.
Reviewing Statistician Date

Mohamed Alosh, Ph.D.
Supervisor/Team Leader Date

cc:
NDA 202428 / 0
DDDP/Walker
DDDP/Diglisic
DDDP/Cook
DDDP/Attinello
OBIO/Patrician
DBIII/Wilson
DBIII/Alosh
DBIII/Fritsch

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

/s/

KATHLEEN S FRITSCH
09/13/2011

MOHAMED A ALOSH
09/13/2011