CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

207071Orig1s000

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

CLINICAL STUDIES

NDA/Serial Number:	207071 / 000
Drug Name:	Finacea (azelaic acid) foam, 15%
Indication(s):	Rosacea
Applicant:	Bayer
Dates:	Submitted: 9/30/2014 PDUFA: 7/30/2015
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:	Gary Chiang, M.D. / David Kettl, M.D.
Project Manager:	Omolara Laiyemo

Keywords: Rosacea, erythema

Table of Contents

1 Executive Summary

Azelaic acid foam, 15% was superior to vehicle foam on the primary efficacy endpoints in two studies conducted on subjects with rosacea. One of the two studies was designed as a Phase 2 study (Study 120) and the other study was designed as a Phase 3 study (Study 846). Even though it was designed as a Phase 2 study, the primary efficacy endpoints and corresponding analysis methods were adequately prespecified in Protocol 120. However, the analysis methods for secondary endpoints, including multiplicity adjustments, were not adequately prespecified in Protocol 120. The protocol for the Phase 3 study (Study 846) adequately prespecified statistical methods for both the primary and secondary endpoints. The studies enrolled subjects age 18 and older with a diagnosis of papulopustular rosacea with an IGA score of moderate to severe, 12-50 inflammatory lesions, and persistent erythema with or without telangiectasia. Subjects applied treatment twice daily for 12 weeks.

The co-primary efficacy endpoints of IGA treatment success (clear or minimal) at Week 12 and absolute change in inflammatory lesions at Week 12 were statistically significant ($p<0.017$) in both studies. See Table 1. Treatment effects for the co-primary endpoints were generally consistent across subgroups and centers, and the treatment effect trends were generally consistent across various assumptions regarding missing data.

Table 1 – Efficacy Results at Week 12

	Study 120		Study 846	
	Azelaic Acid N=198	Vehicle N=203	Azelaic Acid N=483	Vehicle N=478
Primary Endpoints				
IGA clear or minimal	86 (43.4%)	66 (32.5%)	155 (32.1%)	112 (23.4%)
	p=0.017		p=0.001	
Change in inflammatory lesions	-13.0 (0.6)	-9.7 (0.6)	-13.0 (0.4)	-10.2 (0.4)
	p < 0.001		p < 0.001	
Secondary Endpoint				
Grouped erythema rating				
Improved	123 (62.1%)	108 (53.2%)	297 (61.5%)	245 (51.3%)
No change	68 (34.3%)	91 (44.8%)	178 (36.9%)	221 (46.2%)
Worsened	7 (3.5%)	4 (2.0%)	8 (1.7%)	12 (2.5%)
	p=0.138		p=0.001	

The protocol for Phase 3 Study 846 was submitted as a Special Protocol Assessment. The Agency provided agreement regarding the overall design and primary endpoints, but the Agency did not provide agreements regarding the secondary endpoints. Study 846 defined three secondary endpoints: (1) percent change in inflammatory lesions from baseline to Week 12, (2) response rate (clear, minimal, or mild on the IGA) at Week 12, and (3) grouped change in erythema rating (improved, no change, or worsened) at Week 12. The Phase 2 study (Study 120) defined percent change in inflammatory lesions and

clear, minimal, or mild on the IGA as key secondary endpoints and grouped change in erythema rating as an 'other' secondary endpoint, one of a number of 'other' secondary endpoints. The secondary endpoints of IGA clear, minimal, and mild and percent reduction in inflammatory lesions were closely related to the primary efficacy endpoints, and the results, which are similar to the primary endpoint results, can be found in the body of this review.

(b) (4)

2 Introduction

2.1 Overview

2.1.1 Clinical Studies

The applicant is developing azelaic acid foam 15% for the treatment of rosacea. The applicant currently markets Finacea (azelaic acid) gel 15% for the topical treatment of inflammatory papules and pustules of mild to moderate rosacea. This submission is a 505(b)(1) application. Azelaic acid foam was evaluated in an exploratory Phase 2 study (Study 140), a larger Phase 2 study (Study 120), and a Phase 3 study (Study 846). The applicant has submitted the larger Phase 2 Study 120 and the Phase 3 Study 846 as confirmatory studies. All studies were conducted in the United States. The basic design details are summarized in Table 2. This review will focus on Studies 120 and 846.

Table 2 – Clinical Studies Overview

Study Numbers	140	120	846
Study Design	Phase 2, Randomized, double-blind, vehicle-controlled	Phase 2, Randomized, double-blind, vehicle-controlled	Phase 3, Randomized, double-blind, vehicle-controlled
Treatment arms and Sample Size	Azelaic acid – 41 Vehicle - 42	Azelaic acid – 198 Vehicle - 203	Azelaic acid – 483 Vehicle - 478
Inclusion criteria	≥18 years 10-50 inflam. lesions	≥18 years 12-50 inflam. lesions IGA mod. or severe	≥18 years 12-50 inflam. lesions IGA mod. or severe
Treatment regimen	Twice daily for 12 weeks	Twice daily for 12 weeks	Twice daily for 12 weeks
Co-primary endpoints	-IGA clear or minimal -Abs. reduction in inflam. lesions -Grouped change in erythema	-IGA clear or minimal -Abs. reduction in inflam. lesions	-IGA clear or minimal -Abs. reduction in inflam. lesions
Study location	US	US	US
Study dates	Jan. 2008-June 2008	Dec. 2009 – Aug. 2010	Sep. 2012 – Jan. 2014

2.1.2 Regulatory History

The IND for azelaic acid foam was opened in 2007 with the protocol for Phase 2 Study 140. The following meetings were held with the sponsor:

- Pre-IND meeting (7/17/2007)
- Guidance meeting (6/10/2009)
- End of Phase 2 meeting (11/9/2011)
- Pre-NDA meeting (7/9/2014)

After the initial Phase 2 study (Study 140) was completed, the overall clinical development program was discussed at a Guidance meeting (6/10/2009). Because the results for the Study 140 were equivocal (IGA success rate (clear or minimal) of 46% for azelaic acid versus 48% for vehicle), the Agency recommended additional Phase 2 explorations before conducting Phase 3 studies. Thus Study 120 was designed as a Phase 2 study. Following completion of Study 120 with statistically significant results for the primary endpoints, the applicant stated at the End of Phase 2 meeting (11/9/2011) that in their view Study 120 fulfilled the 'requirements of a pivotal study.' The Agency stated that in order for the study to be considered an adequate and well-controlled study, the study needs to have been designed and executed with all of the characteristics of a Phase 3 study (e.g., adequate blinding, appropriate control arm, appropriate endpoints and scales, pre-defined statistical analysis plan, scientifically sound method for handling of missing data along with alternative approaches as sensitivity analyses, appropriate controls for multiplicity, sensitivity analyses, etc.), and that the adequacy of the study would be a review issue. Following the End-of-Phase 2 meeting, the applicant submitted a new Phase 3 protocol (Protocol 846) as a Special Protocol Assessment (SPA)

(2/5/2012). An SPA agreement letter was issued on 3/30/2012. The Agency provided agreement with the general design and the co-primary efficacy endpoints of IGA success (clear or minimal) and absolute change in inflammatory lesions at the end of treatment for Study 846. (b) (4)

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\nda207071\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. Most reviewer analyses were conducted using the datasets associated with the Integrated Summary of Effectiveness (ISE) with crosschecking for key analyses using the datasets associated with the individual study reports. The primary reason for using the ISE datasets was that one subject in Study 846 had their treatment assignment misclassified in the datasets associated with the clinical study report. On the case report form (CRF) for subject 20-020, the investigator mistakenly recorded that treatment kit 516 was dispensed, when kit 615 was the kit randomized and actually dispensed. In the clinical study report, Subject 20-020 is associated with azelaic acid treatment (the recorded kit number 516 was associated with azelaic acid treatment) rather than vehicle treatment (the intended and actually dispensed kit 615 was vehicle). All analyses in the clinical study report have Subject 20-020 misclassified as an azelaic acid subject. However, the applicant re-coded the treatment for Subject 20-020 for the ISE datasets. As the ISE datasets reflect the treatment as randomized and actually dispensed, the ISE datasets were used to correctly classify Subject 20-020 in all data presentations in this review. The analyses in this review are consistent with the ISE, but differ slightly from those presented in the clinical study report.

3.2 Evaluation of Efficacy

3.2.1 Studies 120 and 846

3.2.1.1 Study Design and Statistical Analysis

Studies 120 and 846 were randomized double-blind, vehicle-controlled studies evaluating the safety and efficacy of azelaic acid foam 15% in the treatment of rosacea. Study 120 was designed as a Phase 2 study and Study 846 was designed as a Phase 3 study. The studies enrolled subjects age 18 and older with a diagnosis of papulopustular rosacea with an IGA score of moderate to severe, 12-50 inflammatory lesions, and persistent erythema with or without telangiectasia. Subjects were randomized in a 1:1 ratio to either azelaic

acid foam or vehicle foam. Subjects applied treatment twice daily for 12 weeks. Subjects were evaluated at screening, baseline, and Days 28, 56, 84 (end of treatment), and 112 (post-treatment follow-up). The primary efficacy timepoint was Week 12 (Day 84/end of treatment).

Although Study 120 was designed as a Phase 2 study, statistical methods for the primary efficacy endpoints were prespecified. However, the analysis methods for the secondary endpoints were not clearly prespecified and no multiplicity adjustments were prespecified for the secondary endpoints in the protocol. Study 846 was designed as a Phase 3 study and had adequate prespecification of analysis methods for the primary and secondary endpoints.

Both studies had co-primary endpoints of success on the IGA (clear or minimal) at Week 12 and absolute change in inflammatory lesions at Week 12. Both endpoints needed to demonstrate statistical significance. The studies both used an IGA scale with categories clear, minimal, mild, moderate, and severe; however the morphological descriptors used in each of the studies differed slightly as follows:

Table 3 – Investigator's Global Assessment (IGA)

	Study 120	Study 846
Clear	Virtually no rosacea, ie. no papules and/or pustules; no erythema	No papules and/or pustules; no erythema
Minimal	Rare papules and/or pustules; residual to mild erythema	Rare papules and/or pustules; faint, up to but no including mild erythema
Mild	Few papules and/or pustules; mild erythema	Few papules and/or pustules; mild erythema
Moderate	Pronounced number of papules and/or pustules; moderate erythema	Pronounced number of papules and/or pustules; moderate erythema
Severe	Numerous papules and/or pustules, occasionally with confluent areas of inflamed lesions; moderate to severe erythema	Numerous papules and/or pustules, occasionally with confluent area of inflamed lesions; moderate to severe erythema

Treatment success on the IGA was analyzed using a Cochran-Mantel-Haenszel (CMH) test stratified on pooled treatment center. Centers with fewer than 10 subjects were combined into one analysis center. Homogeneity of the odds ratios across center was tested using the Breslow-Day test.

The change in inflammatory lesion count was analyzed using an ANCOVA model with fixed effect terms for treatment and pooled center and baseline lesion count as a covariate. A supportive analysis including treatment-by-center interaction was conducted to assess the interaction.

The studies differed in their designation of secondary endpoints. Study 120 defined two secondary endpoints. The first secondary endpoint was defined as 'response rate' where a response was defined as achieving clear, minimal, or mild on the IGA at Week 12. The other secondary endpoint was percent change in inflammatory lesions from baseline to Week 12. Response rate was analyzed using a CMH test stratified on pooled center. Percent change in lesions was analyzed with ANCOVA with fixed effect terms for treatment and pooled center and baseline lesion count as a covariate. Study 120 was designed as Phase 2 study and did not include methods for adjusting multiplicity across secondary endpoints. All secondary endpoints were to be analyzed at the nominal 0.05 two-sided significance level.

Protocol 120 also defined a number of other secondary efficacy endpoints. The other secondary endpoints included assessments of erythema and telangiectasia. Erythema was assessed on a 4-point scale (clear or almost clear, mild, moderate, severe). Telangiectasia was also assessed on a 4-point scale (no, mild, moderate, severe). Mean erythema and telangiectasia scores were to be analyzed at each visit. The scores were also to be categorized into improved, unchanged, or worsened relative to baseline for analysis at each visit. The protocol also defined additional analyses on the IGA scores (mean nominal scores and change from baseline), and additional analyses based on the subject's global assessment of treatment response, subject's opinion on local tolerability, rating of facial skin color, and quality of life.

Study 846 defined three secondary endpoints. The three secondary endpoints were to be analyzed in sequential order (1) percent change in inflammatory lesions from baseline to Week 12, (2) response rate (clear, minimal, or mild) at Week 12, and (3) grouped change in erythema rating (improved, no change, or worsened) at Week 12. Percent change in lesions was analyzed with ANCOVA with fixed effect terms for treatment and center and baseline lesion count as a covariate. Response rate was analyzed using a CMH test stratified on pooled center. Grouped change in erythema was analyzed with a CMH (van Elteren) test controlling for centers.

Both studies defined the full analysis set population (FAS) as the primary efficacy analysis population. The FAS is defined as all subjects randomized and dispensed study medication. The per protocol population was defined as subjects who did not discontinue prematurely or had any major protocol deviations.

The primary method of handling missing data in the efficacy analyses was last observation carried forward (LOCF). Sensitivity analyses for IGA success include treating subjects with missing IGA observations as failures, and treating subjects with missing IGA observations as successes. Sensitivity analyses for change in lesion counts include imputing the median value within each treatment group among subjects with complete data, and conducting a repeated measures analysis on the lesion counts with factors for treatment, study week, and treatment-by-study week interaction using the unstructured covariance model.

Phase 3 Protocol 846 was reviewed under a Special Protocol Assessment (SPA). The Agency provided agreement with the general design, the co-primary efficacy endpoints of IGA success (clear or minimal) and absolute change in inflammatory lesions at the end of treatment and the proposed analysis methods for the primary endpoints. However, the Agency did not provide agreement regarding the proposed secondary endpoints of [redacted] (b) (4).

3.2.1.2 Subject Disposition

Study 120 randomized 198 subjects to azelaic acid and 203 to vehicle. Study 846 randomized 483 subjects to azelaic acid and 478 to vehicle. The clinical study report for Study 846 states that 484 subjects were randomized to azelaic acid and 477 subjects were randomized to vehicle, because Subject 20-020 was misclassified in the original database due to an investigator reporting error. (The investigator reported that Kit 516 (azelaic acid) was dispensed when Kit 615 (vehicle) was actually the randomized and dispensed kit number.) The applicant corrected the database used in the ISS and ISE reports and proposed labeling. Results presented in this review use the corrected treatment assignment for Subject 20-020.

Similar proportions of azelaic acid and vehicle subjects discontinued treatment prior to the end of the treatment period in Study 120 (approximately 10-11%); however a larger percentage of vehicle subjects than azelaic acid subjects discontinued treatment in Study 846 (17% vs. 13%). Subjects were to be followed for 4 weeks after the end of the treatment period. Thus, subjects could complete the treatment period, but discontinue before the end of the study. In Study 120, all subjects who discontinued treatment also discontinued the study; however, some subjects in Study 846 discontinued treatment but completed follow-up. The most common reasons for discontinuing treatment or follow-up were lost to follow-up and withdrawal by subject. See Table 4.

Table 4 – Disposition of Subjects in Studies 120 and 846

	Study 120		Study 846	
	Azelaic Acid	Vehicle	Azelaic Acid	Vehicle
Subjects randomized and dispensed medication (ITT)	198	203	483	478
Completed treatment	177 (89%)	183 (90%)	419 (87%)	399 (83%)
Discontinued treatment	21 (11%)	20 (10%)	64 (13%)	79 (17%)
Completed treatment and follow-up	175 (88%)	178 (88%)	402 (83%)	381 (80%)
Completed treatment but discontinued follow-up	2 (1%)	5 (2%)	17 (4%)	18 (4%)
Death	--	--	--	1 (<1%)
Lost to follow-up	1 (<1%)	3 (1%)	7 (1%)	7 (1%)
Other	--	--	2 (<1%)	3 (<1%)
Protocol violation	1 (<1%)	--	2 (<1%)	1 (<1%)
Withdrawal by subject	--	1 (<1%)	5 (1%)	3 (<1%)
Lack of efficacy	--	1 (<1%)	--	--
Discontinued treatment but completed follow-up	--	--	7 (1%)	13 (3%)
Adverse event	--	--	2 (<1%)	5 (1%)
Other	--	--	1 (<1%)	--
Protocol Violation	--	--	3 (<1%)	3 (<1%)
Withdrawal by subject	--	--	1 (<1%)	5 (1%)
Discontinued treatment and follow-up[a]	21 (11%)	20 (10%)	57 (12%)	66 (14%)
Adverse event	4 (2%)	1 (<1%)	4 (<1%)	7 (1%)
Lost to follow-up	8 (4%)	10 (5%)	28 (6%)	23 (5%)
Other	2 (1%)	1 (<1%)	1 (<1%)	3 (<1%)
Protocol violation	2 (1%)	2 (1%)	1 (<1%)	2 (<1%)
Withdrawal by subject	5 (3%)	6 (3%)	23 (5%)	31 (6%)

[a] Reason for discontinuing treatment.
Source: pg. 229 and 291 of ise-iss-ise.pdf and reviewer analysis

3.2.1.3 Baseline Characteristics

Baseline demographics were generally balanced across the treatment groups in the two studies. The mean age was approximately 48 years with approximately 8% age 65 and older in Study 120 and 51 years with approximately 17% age 65 and older in Study 846. More than 70% of subjects were female. At least 95% of subjects were white. Approximately 28% of subjects in Study 120 and 20% of subjects in Study 846 were Hispanic/Latino. See Table 5.

Table 5 – Demographics in Studies 120 and 846

	Study 120		Study 846	
	Azelaic acid N=198	Vehicle N=203	Azelaic acid N=483	Vehicle N=478
Age (years)				
Mean	48.1	48.9	51.2	51.9
Range	19-78	20-83	19-92	19-83
< 65	180 (91%)	187 (92%)	407 (84%)	389 (81%)
≥ 65	18 (9%)	16 (8%)	76 (16%)	89 (19%)
Gender				
Male	43 (22%)	60 (30%)	129 (27%)	130 (27%)
Female	155 (78%)	143 (70%)	354 (73%)	348 (73%)
Race				
White	190 (96%)	196 (97%)	463 (96%)	455 (95%)
Black or Afric.-Amer.	--	2 (1%)	4 (<1%)	5 (1%)
Amer. Ind./AK Native	1 (<1%)	--	1 (<1%)	1 (<1%)
Asian	7 (4%)	4 (2%)	5 (1%)	3 (<1%)
Native HI/Pac. Islander	--	--	--	1 (<1%)
Multiple	--	1 (<1%)	2 (<1%)	4 (<1%)
Not reported	--	--	8 (2%)	9 (2%)
Ethnicity				
Hispanic or Latino	58 (29%)	53 (26%)	98 (20%)	98 (21%)
Not Hispanic or Latino	140 (71%)	150 (74%)	373 (77%)	373 (78%)
Not reported	--	--	12 (2%)	7 (1%)

Source: pg. 231-232 and 292-293 of ise-iss-ise.pdf

To be enrolled in Studies 120 and 846 subjects were to have IGA scores of moderate to severe and 12-50 inflammatory lesion counts. Approximately 87% of subjects had IGA scores of moderate at baseline. The average number of inflammatory lesions at baseline was 21. Most subjects had moderate erythema scores. See Table 6.

Table 6 – Baseline Disease Characteristics in Studies 120 and 846

	Study 120		Study 846	
	Azelaic acid N=198	Vehicle N=203	Azelaic acid N=483	Vehicle N=478
IGA				
Moderate	172 (87%)	189 (93%)	418 (87%)	416 (87%)
Severe	26 (13%)	14 (7%)	65 (13%)	62 (13%)
Lesion Count				
Mean	21.6	20.4	21.7	21.2
Range	12-50	12-48	12-50	12-50
Erythema Rating				
Mild	19 (10%)	12 (6%)	43 (9%)	39 (8%)
Moderate	160 (81%)	171 (84%)	364 (75%)	370 (77%)
Severe	19 (10%)	20 (10%)	76 (16%)	69 (14%)

Source: pg. 233-234 and 295-296 of ise-iss-ise.pdf

3.2.1.4 Primary Efficacy Endpoints

Azelaic acid foam was superior to vehicle foam on both IGA success rate and change in inflammatory lesions at Week 12 in both Studies 120 and 846 ($p<0.017$, two-sided). The applicant presented two-sided p-values in the study report for Study 120 and one-sided p-values in the study report for Study 846 and the ISE. This review will use two-sided p-values unless otherwise indicated. The primary method of handling missing data was LOCF. The treatment effects were consistent across the two studies for the two primary endpoints, with approximately 10% more subjects achieving IGA success on the azelaic arm versus the vehicle arm and a reduction of approximately 3 additional inflammatory lesions on the azelaic acid arm versus the vehicle arm in both studies. The vehicle IGA success rate was higher in Study 120 than in Study 846 (33% vs. 23%). Primary efficacy results are presented in Table 7, Table 8, and Table 9.

Table 7 – IGA Success Rate (clear or minimal) at Week 12 (Studies 120 and 846) [FAS]

Study 120		Study 846	
Azelaic Acid N=198	Vehicle N=203	Azelaic Acid N=483	Vehicle N=478
86 (43.4%)	66 (32.5%)	155 (32.1%)	112 (23.4%)
p=0.017		p=0.001	

Source: pg. 407 and 469 of ise-iss-ise.pdf and reviewer analysis

Table 8 – Change in Inflammatory Lesions from Baseline to Week 12 (Study 120) [FAS]

	Nominal value		Change from baseline	
	Azelaic Acid N=198	Vehicle N=203	Azelaic Acid N=198	Vehicle N=203
Baseline [mean (SD)]	21.6 (9.9)	20.4 (8.8)		
End of treatment				
[mean (SD)]	8.2 (8.9)	10.8 (10.3)	-13.3 (10.4)	-9.5 (9.7)
[adjusted mean (SE)]	8.0 (0.6)	11.3 (0.6)	-13.0 (0.6)	-9.7 (0.6)
			$p < 0.001$	

Source: pg. 422 of ise-iss-ise.pdf and reviewer analysis

Table 9 - Change in Inflammatory Lesions from Baseline to Week 12 (Study 846) [FAS]

	Nominal value		Change from baseline	
	Azelaic Acid N=483	Vehicle N=478	Azelaic Acid N=483	Vehicle N=478
Baseline [mean (SD)]	21.7 (9.1)	21.2 (8.7)		
End of treatment				
[mean (SD)]	8.5 (8.9)	10.8 (11.3)	-13.2 (9.5)	-10.3 (9.8)
[adjusted mean (SE)]	8.5 (0.4)	11.2 (0.4)	-13.0 (0.4)	-10.2 (0.4)
			<0.001	

Source: pg. 488 of ise-iss-ise.pdf and reviewer analysis

3.2.1.5 Missing Data Handling

The primary method of handling missing data in the efficacy analyses was last observation carried forward (LOCF). When Protocol 846 was submitted for Agency review under an SPA, it included only a primary method of handling missing data—LOCF. The Agency recommended providing a rationale for using LOCF and to propose two or three sensitivity analyses for handling missing data that rely on different assumptions. The applicant then added sensitivity analyses to the protocol which were not further reviewed by the Agency.

Sensitivity analyses for IGA success include treating subjects with missing IGA observations as failures, and treating subjects with missing IGA observations as successes. Most of the observations imputed as LOCF were imputed as failures, though a slightly higher proportion of subjects with missing values on the azelaic acid arm were imputed as successes relative to subjects on the vehicle arm. Missing as failure imputation led to similar results as the LOCF imputation. In Study 120, the proportion of subjects with missing data at Week 12 was similar (9.5% on the azelaic acid arm and 10.8% on the vehicle arm), and thus the missing as success imputation lead to a similar treatment effect as was observed using LOCF and missing as failure. In Study 846, a greater proportion of subjects on the vehicle arm had missing data relative to the proportion of subjects on the azelaic acid arm (15.3% vs. 12.4%). Thus, imputing missing as success narrowed the treatment effect, however the trend still favors azelaic acid. See Table 10.

Table 10 – IGA Success Rates under Missing Data Sensitivity Analyses

	Study 120		Study 846	
	Azelaic Acid N=198	Vehicle N=203	Azelaic Acid N=483	Vehicle N=478
LOCF (primary)	86 (43.4%)	66 (32.5%) p=0.017	155 (32.1%)	112 (23.4%) p=0.001
Missing as failure (sensitivity)	84 (42.4%)	65 (32.0%) p=0.0238	149 (30.9%)	110 (23.0%) p=0.0035
Missing as success (sensitivity)	103 (52.0%)	87 (42.9%) p=0.0531	209 (43.3%)	183 (38.3%) p=0.0792

Source: reviewer analysis

As a sensitivity analysis for change in lesion counts the applicant imputed the median value within each treatment group among subjects with complete data. This analysis treats all subjects with missing data on a treatment arm as having the identical response as the 'typical' completer. Imputing the median of the completers for the subjects with missing data increased the reduction in lesions in both treatment arms relative to LOCF, as subjects who drop out early tend to have smaller reductions at the time of dropout than subjects who complete the study, though the estimated treatment effect was similar in both imputations. In addition, imputing the same value (the median) for all missing values on a treatment arm also reduces the variability relative to LOCF, as the same number is imputed for all subjects with missing data on a treatment arm, rather than

imputing the last observed value, which would at least have subject-to-subject variability. In addition the applicant conducted a repeated measures analysis on the nominal lesion counts with factors for treatment, study week, and treatment-by-study week interaction using the unstructured covariance model. The final lesion counts were lower using the repeated measures analysis relative to the LOCF imputation, but the treatment effects trended in the same direction. While the results of this sensitivity analysis were still statistically significant for Study 846, the results were not statistically significant for Study 120. See Table 11.

Table 11 – Lesion Count Results (Change from Baseline and Week 12 Counts) under Missing Data Sensitivity Analyses

	Study 120		Study 846	
	Azelaic Acid N=198	Vehicle N=203	Azelaic Acid N=483	Vehicle N=478
Change from baseline [adjusted mean (SE)]				
LOCF (primary)	-13.0 (0.6)	-9.7 (0.6)	-13.0 (0.4)	-10.2 (0.4)
	p < 0.001		p <0.001	
Median imputation (sensitivity)	-13.6 (0.5)	-10.7 (0.5)	-14.1 (0.4)	-11.6 (0.4)
	p <0.001		p <0.001	
Week 12 Counts [adjusted mean (SE)]				
LOCF	8.0 (0.6)	11.3 (0.6)	8.5 (0.4)	11.2 (0.4)
	p<0.001		p<0.001	
Repeated measures (sensitivity)	7.4 (0.7)	10.0 (0.7)	7.3 (0.5)	9.9 (0.5)
	p=0.191		p=0.004	

Source: pg. 462-463, 574-575 of ise-iss-ise.pdf

3.2.1.6 Secondary Efficacy Analyses

The secondary efficacy endpoints in Study 846 were (1) percent change in inflammatory lesions from baseline to Week 12, (2) response rate (IGA clear, minimal, or mild) at Week 12, and (3) grouped change in erythema rating (improved, no change, or worsened) at Week 12. In Study 846 the secondary endpoints were to be analyzed in sequential order to account for multiplicity. In the SPA agreement letter for Study 846, the Agency did not provide agreements regarding any of the proposed secondary endpoints.

Study 120 defined only two secondary endpoints (1) percent change in inflammatory lesions from baseline to Week 12, (2) response rate (IGA clear, minimal, or mild) at Week 12. No methods for accounting for multiplicity among the secondary endpoints were specified for Study 120. Study 120 also included grouped change in erythema rating as an 'other' secondary endpoint, among a large number of other secondary endpoints.

Percent reduction in lesions and response defined as IGA clear, minimal or mild were analyzed using comparable statistical methods to the primary endpoints (ANCOVA with fixed effect terms for treatment and pooled center and baseline lesion count as a covariate

for the percent change in lesions and CMH test stratified on pooled center for response rate). Grouped erythema rating was analyzed using a CMH test stratified on pooled center in Study 846. Although the protocol for Study 846 stated that the van Elteren CMH test would be used, which would imply using the 'modridit' scores option in SAS, the applicant's reported p-values correspond to the 'table' scores option of the CMH test instead. The protocol for Study 120 did not specify a specific analysis to be used for the erythema rating, noting simply that both absolute values and grouped ratings would be analyzed. In the study report, the applicant presented results for the grouped erythema rating using an analysis population that excluded subjects who had no post-baseline observations. The protocol did not clearly specify a primary method of analysis, and the sponsor presented p-values from two analyses, a Wilcoxon rank sum test and a CMH test stratified on pooled centers using table scores. Table 12 below presents the p-values from the CMH test using table scores for both studies.

In Study 846 all three secondary endpoints were statistically significant ($p \leq 0.001$) when analyzed sequentially as specified in the protocol. In Study 120, the percent reduction in lesions and response rate (clear, minimal, or mild on the IGA) were both nominally significant ($p \leq 0.012$), though no method of adjusting for multiplicity was specified in the protocol in order to adequately interpret the results. The p-value for grouped erythema rating was not nominally significant in Study 120. See Table 12.

The conclusions for the grouped change in erythema rating are the same under the CMH analyses using both table and modridit scores. In Study 846; the p-value is 0.001 under both analyses, while in Study 120 the p-value is 0.138 using table scores and 0.083 using modridit scores. The p-value for the Wilcoxon rank sum test in Study 120 was also not significant (p=0.108)

Table 12 – Secondary Efficacy Endpoints at Week 12 in Studies 120 and 846

	Study 120		Study 846	
	Azelaic acid N=198	Vehicle N=203	Azelaic acid N=483	Vehicle N=478
Percent reduction in lesions				
[mean (SD)]	-62.4% (35.7%)	-47.7% (41.3%)	-61.6% (33.5%)	-50.8% (40.0%)
[adjusted mean (SE)]	-63.4% (2.7%)	-47.9% (2.7%)	-60.7% (1.7%)	-49.5% (1.7%)
	p<0.001		p<0.001	
IGA clear, minimal, mild	137 (69.2%)	117 (57.6%)	320 (66.3%)	260 (54.4%)
	p=0.012		p<0.001	
Grouped erythema rating				
Improved	123 (62.1%)	108 (53.2%)	297 (61.5%)	245 (51.3%)
No change	68 (34.3%)	91 (44.8%)	178 (36.9%)	221 (46.2%)
Worsened	7 (3.5%)	4 (2.0%)	8 (1.7%)	12 (2.5%)
	p=0.138[a]		p=0.001[a]	

[a] P-value from the CMH test stratified on pooled centers using table scores. Note that in the ISE, the applicant presented one-sided p-values (0.069 for Study 120 and <0.001 for Study 846) rather than two-sided p-values.
Source: pg. 422, 488 of ise-iss-ise.pdf, pg 194 of 1403120-report-body.pdf, and reviewer analysis

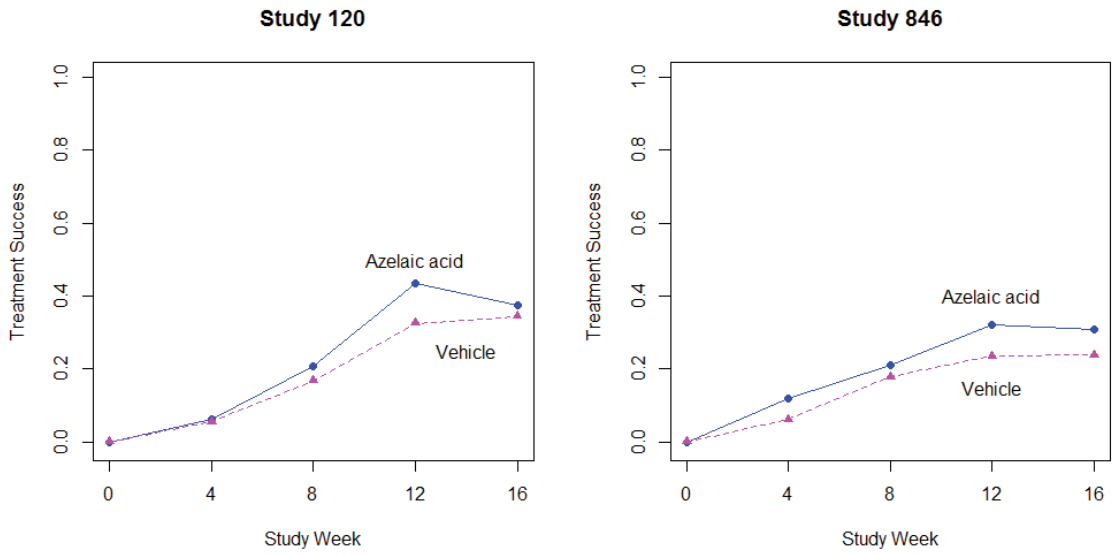
The treatment effect for the percent reduction in inflammatory lesions was about 15% in Study 120 and 11% in Study 846, and the results are consistent with the primary efficacy endpoint of absolute reduction in lesions. The secondary endpoint of response rate (IGA clear, minimal, or mild) differs from the primary endpoints of treatment success (IGA clear or minimal) due to the inclusion of subjects with scores of mild at the end of treatment. In Study 120, similar proportions of azelaic acid and vehicle subjects had an IGA score of mild at the end of treatment (25.8% for azelaic acid and 25.1% for vehicle) leading to very similar treatment effects when 'mild' is included in the response definition or not (about 11%). In Study 846, slightly more azelaic acid and vehicle subjects had an IGA score of mild at the end of treatment (34.2% vs 31.0%), leading to a slightly larger treatment effect when 'mild' is included in the response definition versus when it is not included (about 12% vs. 9%). Note that because the majority of subjects (≥ 87%) had an IGA score of moderate at baseline, improving 1 grade to a score of mild may not represent much of a clinically meaningful change.

Although the proportion of subjects who improved their erythema rating was greater on the azelaic acid arm than the vehicle arm in both studies, the results were not even nominally significant in Study 120 under either of the applicant's analysis methods (CMH or Wilcoxon). Note also that the proportion of subjects whose erythema worsened was greater on the azelaic acid arm than the vehicle arm in Study 120. As the grouped change in erythema endpoint allows for the category of improved for only a 1-point improvement on the erythema rating scale, such changes may not be clinically important.

3.2.1.7 Efficacy over Time

Treatment success rates increased over time through Week 12 (end of treatment) with treatment success rates higher on the azelaic acid arm than the vehicle arm, though the differences between the two arms were small. The difference between the two arms was greatest at the end of the treatment period (Week12). Treatment success on the azelaic acid arm decreased slightly after treatment ended between Weeks 12 and 16. Although the treatment success rates on both the azelaic acid arm and the vehicle arm were higher in Study 120 than 846, the treatment effects were similar in the two studies. See Figure 1.

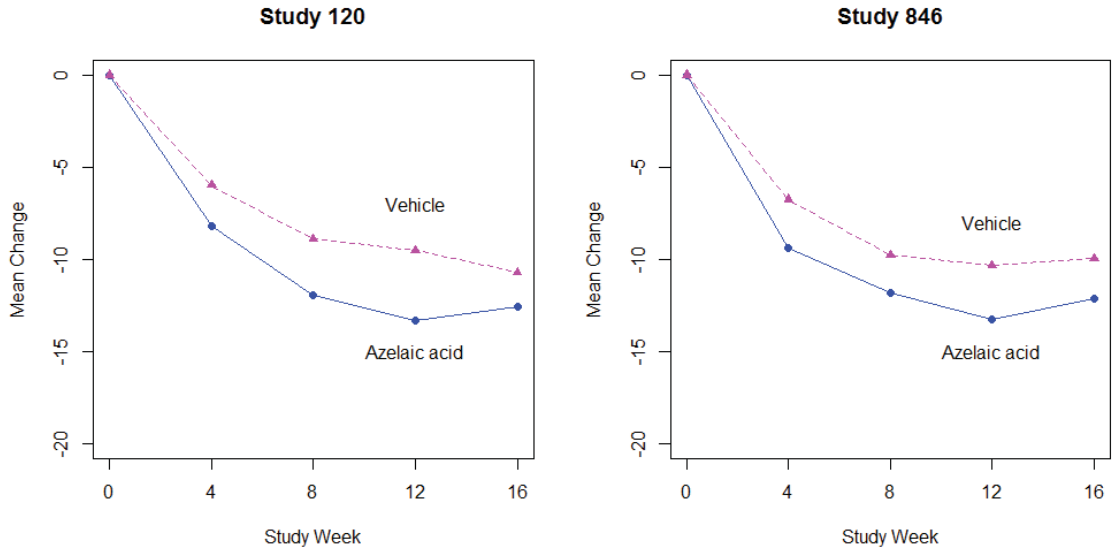
Figure 1 – Treatment Success Rates over Time



Source: reviewer analysis

The mean change in inflammatory lesions decreased over time through Week 12 (end of treatment) with greater decreases on the azelaic acid arm than the vehicle arm. The difference between the two arms was greatest at the end of the treatment period (Week12). The results were similar in the two studies. See Figure 2.

Figure 2 – Mean Change in Inflammatory Lesion Counts over Time

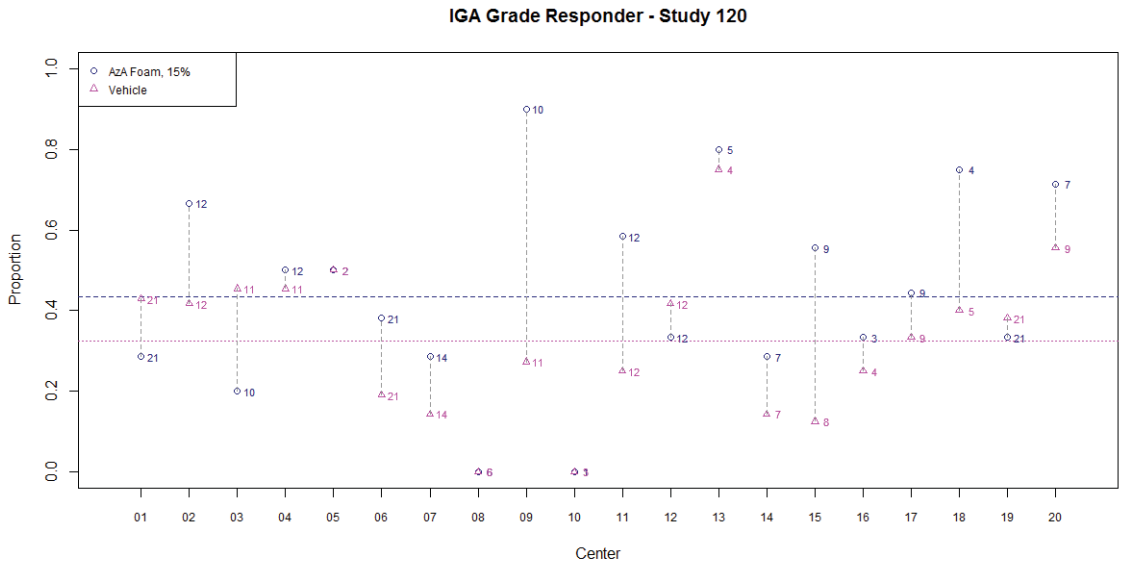


Source: reviewer analysis

3.2.1.8 Efficacy by Center

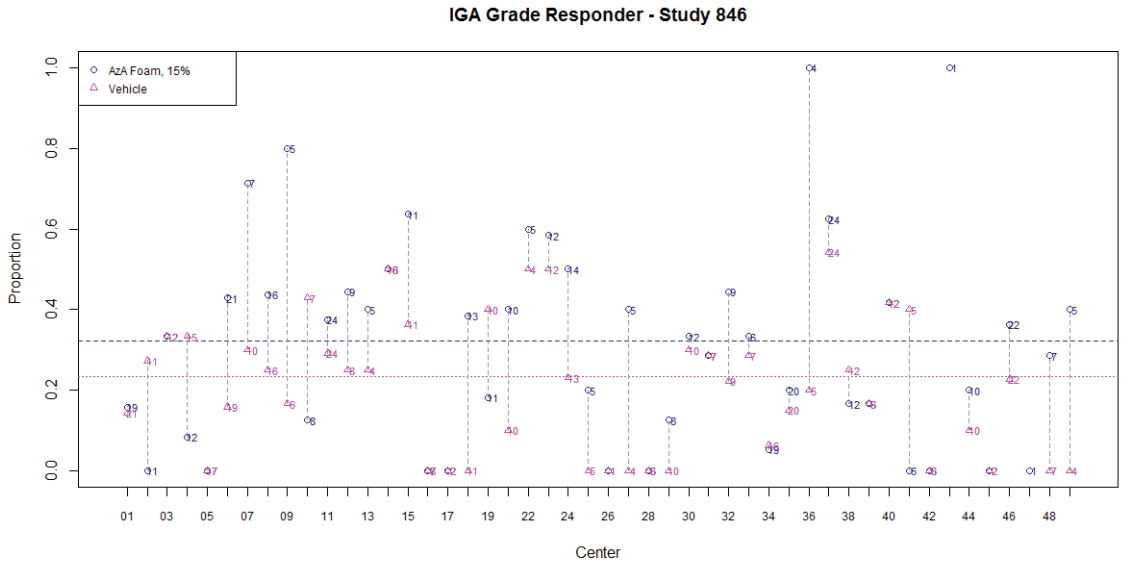
Study 120, with 401 subjects, was conducted at 20 centers in the US. The five smallest centers (fewer than 10 subjects) were pooled into an analysis center leading to 16 analysis centers. Study 846, with 961 subjects, was conducted at 48 centers in the US. The 10 smallest centers (fewer than 10 subjects) were pooled into an analysis center leading to 39 analysis centers. The response rates and mean changes in lesion counts across centers were variable, but because each center had only a small proportion of the overall sample size, no center is overly influential on the overall results. See Figure 3 through Figure 6. For the IGA treatment success endpoint, homogeneity across analysis centers was conducted using the Breslow-Day test. The p-values from the Breslow-Day test were 0.179 for Study 120 and 0.190 for Study 846. Thus the studies were not able to detect a significant lack of homogeneity for the treatment success endpoint. For the mean change in inflammatory lesions endpoints, treatment-by-center interaction was tested by including the interaction term in the ANCOVA model. The p-values for the treatment-by-center interaction term were 0.071 in Study 120 and 0.710 in Study 846. For Phase 2 Protocol 120, the applicant stated only that center effects would be assessed with descriptive statistics, and the protocol did not specify a p-value for assessing treatment-by-center interactions.

Figure 3 – IGA Response Rate by Center (Study 120)



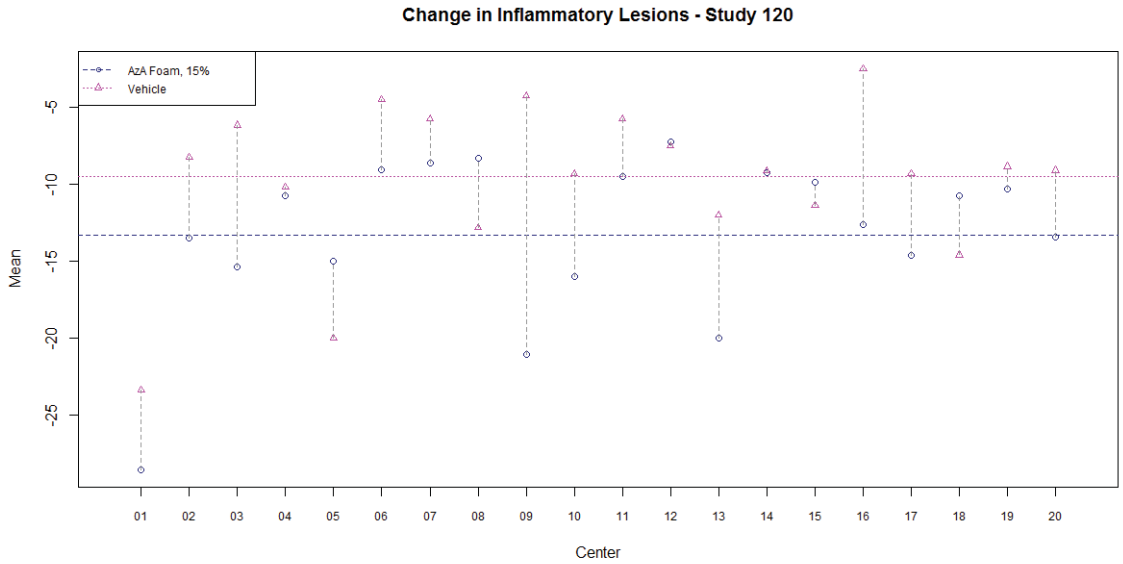
Source: reviewer analysis. Numbers represent the sample size on the treatment arm.

Figure 4 – IGA Response Rate by Center Study 846



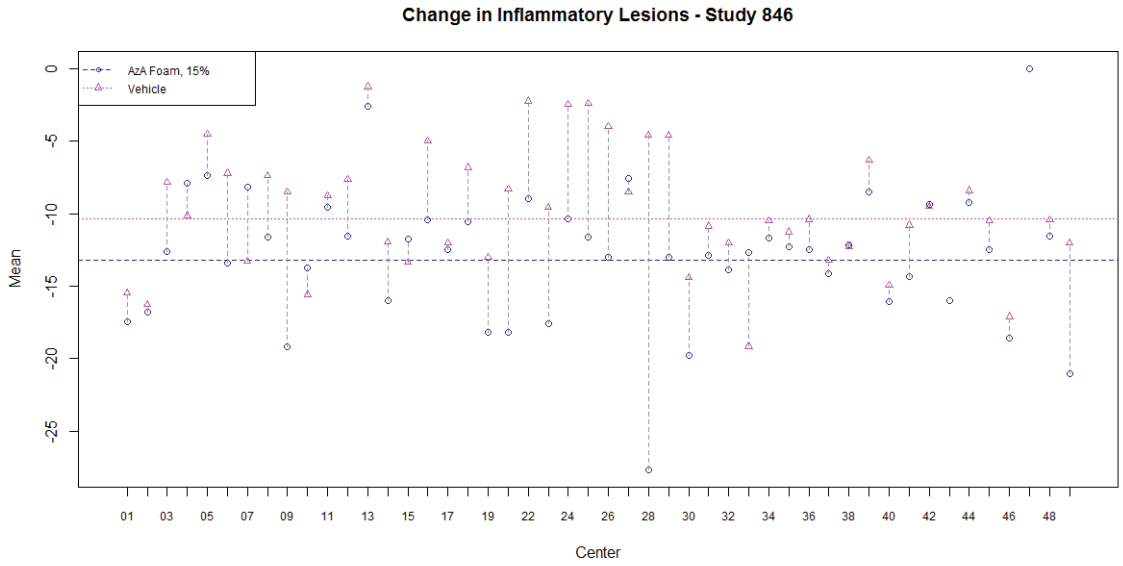
Source: reviewer analysis. Numbers represent the sample size on the treatment arm.

Figure 5 – Mean Change in Inflammatory Lesions by Center (Study 120)



Source: reviewer analysis

Figure 6 – Mean Change in Inflammatory Lesions by Center (Study 846)



Source: reviewer analysis

3.2.2 Supportive Study 140

The initial clinical study conducted for azelaic acid foam was a Phase 2 study (Study 140). The study was a randomized, double-blind, vehicle-controlled study in subjects age 18 and older with papulopustular rosacea with 10 to 50 inflammatory lesions. The study enrolled 83 subjects (41 on the azelaic acid arm and 42 on the vehicle arm) at 7 centers. Subjects applied treatment twice daily for 12 weeks. Subjects were evaluated using a 7-point IGA scale (clear, minimal, mild, mild-to-moderate, moderate, moderate-to-severe, and severe), inflammatory lesion counts, and erythema score. The protocol defined three primary efficacy endpoints: (1) change from baseline to end of study in inflammatory lesion count, (2) clear or minimal on the IGA at end of study, and (3) grouped change in erythema at end of study. Efficacy results are presented in Table 13. Little difference was observed between the two treatment arms at the end of the study in this small Phase 2 study. The azelaic acid arm had approximately a reduction of one additional inflammatory lesion relative to the vehicle arm (-11.7 v. -10.8), similar results for IGA treatment success, with a slight trend favoring vehicle (46% for azelaic acid and 48% for vehicle), and a slight trend for improvement on erythema favoring azelaic acid (61% vs. 48%).

Table 13 – Efficacy Results in Study 140

	Study 140	
	Azelaic acid N=41	Vehicle N=42
Absolute reduction in lesions [mean (SD)]		
Baseline	18.0 (10.6)	17.6 (8.4)
Change from baseline	-11.7 (8.5)	-10.8 (7.8)
	p=0.609	
IGA clear or minimal	19 (46.2%)	20 (47.6%)
	p=0.915	
Grouped erythema rating		
Improved	25 (61.0%)	20 (47.6%)
No change	15 (36.6%)	19 (45.2%)
Worsened	1 (2.4%)	3 (7.1%)
	p=0.117	

Source: pg. 53-57 of 1402140-report-body.pdf.

3.3 Evaluation of Safety

3.3.1 Extent of Exposure

Subjects on the azelaic acid and vehicle arms had similar treatment durations (last day of treatment – first day of treatment +1) of approximately 79 days in both studies. Subjects on the vehicle arm used more study product on average (approximately 19 g more) during the treatment period than subjects on the azelaic acid arm. The exposure calculations were conducted on subjects with available data. See Table 14.

Table 14 – Extent of Exposure in Studies 120 and 846

	Study 120		Study 846	
	Azelaic acid N=198	Vehicle N=203	Azelaic acid N=483	Vehicle N=478
Duration of treatment (Days)	*N=198*	*N=203*	*N=481*	*N=478*
Mean (SD)	79.3 (20.0)	79.1 (21.0)	80.4 (20.1)	79.3 (20.4)
Range	1-135	1-113	1-154	1-157
Amount used (g)	*N=192*	*N=191*	*N=433*	*N=436*
Mean (SD)	109.5 (52.6)	127.5 (56.3)	103.0 (57.1)	122.3 (64.7)
Range	0-306	17-302	0-310	0-389

Source: reviewer analysis

3.3.2 Adverse Events

Approximately 32% of azelaic acid subjects and 25% of vehicle subjects experienced at least one adverse event during the study. Approximately 1.5% of subjects discontinued treatment due to adverse events. The most common adverse reactions were the administration site conditions of pain, pruritus, and paresthesia. See Table 15 and Table 16.

Table 15 – Adverse Events in Studies 120 and 846

	Study 120		Study 846	
	Azelaic acid N=198	Vehicle N=203	Azelaic acid N=483	Vehicle N=478
Any adverse event	68 (34%)	48 (24%)	149 (31%)	119 (25%)
Serious adverse event	1 (<1%)	3 (1%)	3 (<1%)	4 (<1%)
Discontinued treatment due to AEs	4 (2%)	3 (1%)	6 (1%)	12 (3%)

Source: pg 1891-1892 and 2258-2259 of ise-iss-ise.pdf and reviewer analysis

Table 16 – Administration Site Conditions and Skin and Subcutaneous Tissue Disorders Observed in ≥1% of Azelaic Acid Subjects (Combined Studies)

	Study 120		Study 846		Combined	
	Azelaic acid N=198	Vehicle N=203	Azelaic acid N=483	Vehicle N=478	Azelaic acid N=681	Vehicle N=681
Disagreeable skin sensations[a]	18 (9%)	4 (2%)	24 (5%)	6 (1%)	42 (6%)	10 (1%)
Application site pain[a]	15 (8%)	3 (1%)	17 (4%)	6 (1%)	32 (5%)	9 (1%)
Application site paresthesia[a]	3 (1%)	1 (<1%)	3 (<1%)	1 (<1%)	6 (1%)	2 (<1%)
Skin burning sensation[a]	0 (0%)	0 (0%)	4 (<1%)	0 (0%)	4 (<1%)	0 (0%)
Pruritus[b]	7 (4%)	0 (0%)	10 (2%)	2 (<1%)	17 (2%)	2 (<1%)
Application site pruritus[b]	6 (3%)	0 (0%)	7 (1%)	2 (<1%)	13 (2%)	2 (<1%)
Pruritus[b]	1 (<1%)	0 (%)	3 (<1%)	0 (0%)	4 (<1%)	0 (0%)
Application site dryness	0 (0%)	2 (1%)	5 (1%)	3(<1%)	5 (<1%)	5 (<1%)
Application site erythema	1 (<1%)	2 (1%)	4 (<1%)	4 (<1%)	5 (<1%)	6 (1%)

[a] For labeling, application site pain, application site paresthesia, and skin burning sensation are combined into a single category.
[b] For labeling, application site pruritus and pruritus are combined into a single category.
Source: pg 608-612, 1903-1908 and 2272-2274 of ise-iss-ise.pdf and reviewer analysis

4 Findings in Special/Subgroup Populations

4.1 Gender, Race, Age, and Geographic Region

There were too few subjects who reported a race other than white for meaningful subgroup analysis by age group or race. Treatment effects across age group, gender, and ethnicity were generally consistent in favor of azelaic acid. All subjects were enrolled in the United States. See Table 17 and Table 18.

Table 17 – IGA Treatment Success Rates by Subgroup

	Study 120		Study 846	
	Azelaic acid N=198	Vehicle N=203	Azelaic acid N=483	Vehicle N=478
Age (years)				
< 65	76/180 (42%)	60/187 (32%)	119/407 (29%)	90/389 (23%)
≥ 65	10/18 (56%)	6/16 (38%)	36/76 (47%)	22/89 (25%)
Gender				
Male	13/43 (30%)	17/60 (28%)	38/129 (29%)	29/130 (22%)
Female	73/155 (47%)	49/143 (34%)	117/354 (33%)	83/348 (24%)
Race				
White	83/190 (44%)	63/196 (32%)	150/463 (32%)	105/455 (23%)
Not white	3/8 (38%)	3/7 (43%)	5/20 (25%)	7/23 (30%)
Ethnicity				
Hispanic or Latino	22/58 (38%)	14/53 (26%)	35/98 (36%)	26/98 (27%)
Not Hispanic or Latino/Not reported	64/140 (46%)	52/150 (35%)	120/385 (31%)	86/380 (23%)

Source: reviewer analysis.

Table 18 – Mean Change in Inflammatory Lesions by Subgroup

	Study 120		Study 846	
Mean [N]	Azelaic acid N=198	Vehicle N=203	Azelaic acid N=483	Vehicle N=478
Age (years)				
< 65	-13.6 [180]	-9.5 [187]	-13.2 [407]	-10.1 [389]
≥ 65	-10.8 [18]	-9.4 [16]	-13.2 [76]	-11.2 [89]
Gender				
Male	-15.3 [43]	-9.2 [60]	-13.0 [129]	-9.8 [130]
Female	-12.8 [155]	-9.7 [143]	-13.3 [354]	-10.5 [348]
Race				
White	-13.3 [190]	-9.6 [196]	-13.4 [463]	-10.4 [455]
Not white	-13.3 [8]	-8.7 [7]	-9.8 [20]	-9.7 [23]
Ethnicity				
Hispanic or Latino	-12.4 [58]	-7.1 [53]	-11.5 [98]	-10.1 [98]
Not Hispanic or Latino/Not reported	-13.7 [140]	-10.4 [150]	-13.7 [385]	-10.4 [380]

Table presents Mean change and [subgroup sample size]
Source: reviewer analysis.

4.2 Other Special/Subgroup Populations

Not applicable.

5 Summary and Conclusions

5.1 Statistical Issues and Collective Evidence

The applicant evaluated the efficacy of azelaic acid foam 15% in two vehicle-controlled studies for the treatment of rosacea. One of the studies was designed as a Phase 2 study (Study 120) and the other study was designed as a Phase 3 study (Study 846) that underwent a Special Protocol Assessment. The Agency provided agreement regarding the overall design and primary endpoints for Study 846 under the SPA. The statistical methods for the primary efficacy endpoints were adequately prespecified in both Study 120 and Study 846. Both studies had statistically significant results for the co-primary efficacy endpoints of IGA treatment success (clear or minimal) at Week 12 and absolute reduction in inflammatory lesions at Week 12 ($p<0.017$, two-sided). Treatment effects for the co-primary endpoints were generally consistent across studies, subgroups, and centers, and the treatment effect trends were generally consistent across various assumptions regarding missing data.

Study 846 defined three secondary endpoints: (1) percent change in inflammatory lesions from baseline to Week 12, (2) response rate (clear, minimal, or mild on the IGA) at Week 12, and (3) grouped change in erythema rating (improved, no change, or worsened) at Week 12. The three endpoints were to be analyzed sequentially. Study 120 defined two secondary endpoints: (1) percent change in inflammatory lesions from baseline to Week 12, and (2) response rate (clear, minimal, or mild on the IGA) at Week 12. Protocol 120 did not specify any method for adjusting for multiplicity due to two endpoints. Grouped change in erythema rating was designated as an 'other' secondary endpoint, one of a number of 'other' secondary endpoints (also with no multiplicity adjustments). The Agency did not provide any agreements regarding the secondary endpoints in the SPA letter for Protocol 846.

The planned secondary endpoints of response rate (IGA clear, minimal, and mild) and percent reduction in inflammatory lesions were closely related to the primary efficacy endpoints. Although Protocol 120 did not prespecify multiplicity adjustments due to multiple secondary endpoints, the analysis of these two endpoints was generally supportive of the related primary endpoints in both studies (nominally significant in Study 120 and significant after adjusting for multiplicity in Study 846).

(b) (4)

5.2 Conclusions and Recommendations

Azelaic acid foam, 15% was superior to vehicle foam in the treatment of rosacea in two studies. Study 120 (a Phase 2 study) and Study 846 (a Phase 3 study) enrolled subjects age 18 and older with a diagnosis of papulopustular rosacea with an IGA score of moderate to severe, 12-50 inflammatory lesions, and persistent erythema with or without telangiectasia. Subjects applied treatment twice daily for 12 weeks. The co-primary efficacy endpoints of IGA treatment success (clear or minimal) at Week 12 and absolute change in inflammatory lesions at Week 12 were statistically significant ($p<0.017$). The secondary endpoint of grouped erythema rating was not adequately prespecified in Protocol 120, nor were the results statistically significant (nominal $p=0.138$). Although the results for grouped erythema rating were prespecified and statistically significant in Study 846, the Agency did not provide agreement regarding this endpoint. (b) (4)

Table 19 – Efficacy Results at Week 12

	Study 120		Study 846	
	Azelaic Acid N=198	Vehicle N=203	Azelaic Acid N=483	Vehicle N=478
Primary Endpoints				
IGA clear or minimal	86 (43.4%)	66 (32.5%)	155 (32.1%)	112 (23.4%)
	p=0.017		p=0.001	
Change in inflammatory lesions	-13.0 (0.6)	-9.7 (0.6)	-13.0 (0.4)	-10.2 (0.4)
	p < 0.001		p < 0.001	
Secondary Endpoint				
Grouped erythema rating				
Improved	123 (62.1%)	108 (53.2%)	297 (61.5%)	245 (51.3%)
No change	68 (34.3%)	91 (44.8%)	178 (36.9%)	221 (46.2%)
Worsened	7 (3.5%)	4 (2.0%)	8 (1.7%)	12 (2.5%)
	p=0.138		p=0.001	

Signatures/Distribution List

Primary Statistical Reviewer: Kathleen Fritsch, Ph.D.
Date: 5/28/2015

Statistical Team Leader: Mohamed Alosh, Ph.D.

cc:
DDDP/Marcus
DDDP/Kettl
DDDP/Chiang
DDDP/Laiyemo
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
05/28/2015

MOHAMED A ALOSH
05/28/2015