CENTER FOR DRUG EVALUATION AND RESEARCH

APPROVAL PACKAGE FOR:

APPLICATION NUMBER

21-184/002

Statistical Review(s)

STATISTICAL REVIEW AND EVALUATION

Medical Division: **Division of Dermatologic & Dental Drug Products (DDDDP)**
HFD-540
Biometrics Division: **Division of Biometrics III**
HFD-725

NDA NUMBER:	21-184/SE1
SERIAL NUMBER:	002
DATE RECEIVED BY CENTER:	06/28/01
DRUG NAME:	Tazarotene Cream 0.1%
INDICATION:	
SPONSOR:	Allergan
DOCUMENTS REVIEWED:	Electronic Submission (\\CDSESUB1\N21184\S_002\2001-06-28)
STATISTICAL REVIEWER:	Kathleen Fritsch, Ph.D. (HFD-725)
STATISTICAL TEAM LEADER:	Mohamed Alosh, Ph.D. (HFD-725)
BIOMETRICS DIVISION DIRECTOR:	M. F. Huque, Ph.D. (HFD-725)
CLINICAL REVIEWER:	Phyllis Huene, M.D. (HFD-540)
PROJECT MANAGER:	Kalyani Bhatt (HFD-540)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Executive Summary of Statistical Findings

1.1 Conclusions and Recommendations

The sponsor has conducted studies to demonstrate the efficacy and safety of tazarotene cream 0.1% [redacted] The following signs and symptoms demonstrated a statistically significant effect for tazarotene in the phase 3 program: fine wrinkling, mottled hyperpigmentation, lentigines, elastosis, pore size, and irregular depigmentation.

In the phase 3 studies, the primary endpoints were clinical improvement in fine wrinkling and mottled hyperpigmentation at Week 24. Each endpoint is graded on a 5-point scale from 'none' to 'severe', and clinical improvement is defined as at least one grade improvement from baseline to Week 24. The sponsor also included an evaluation of the proportion of subjects with scores of 'none' or 'minimal' at Week 24 (hereafter referred to as treatment success) at the request of the Agency. Fine wrinkling and mottled hyperpigmentation demonstrated statistically significant clinical improvement at Week 24 with p-values for tazarotene versus vehicle less than 0.001 for both endpoints in both phase 3 studies. These two endpoints also demonstrated statistically significant treatment success at Week 24 with p-values less than or equal to 0.005.

Several secondary and "additional" endpoints were also assessed in the phase 3 studies. These endpoints included 8 additional signs and symptoms of photodamage (lentigines, elastosis, tactile roughness, course wrinkling, pore size, irregular depigmentation, telangiectasia, and actinic keratoses) and three global evaluations (investigator's overall integrated assessment, global response to treatment, and patient's overall self-assessment). Of these secondary endpoints, four of the signs and symptoms (lentigines, elastosis, pore size, and irregular depigmentation) had statistically significant clinical improvement ($p \leq 0.002$) and treatment success ($p \leq 0.003$) in two studies. The statistical significance of these primary and secondary endpoints is maintained when adjusted for multiplicity. All three global evaluations were also statistically significant ($p \leq 0.001$).

1.2 Overview of Clinical Program and Studies Reviewed

The clinical program for tazarotene cream 0.1% for [redacted] included a phase 2 dose ranging study (025C), and two phase 3 efficacy and safety studies (033C and 034C). Study 025C enrolled 349 subjects, including 58 subjects on the tazarotene cream 0.1% arm. Study 025C evaluated tazarotene cream at concentrations 0.01%, 0.025%, 0.5%, and 0.1%, along with vehicle and tretinoin cream 0.05%. Study 033C enrolled 563 subjects, including 283 on the tazarotene 0.1% arm. Similarly, Study 034C enrolled 568 subjects, including 284 on tazarotene. All of the investigative sites for these three studies were located in the United States. Most of the investigators from the phase 2 dose ranging study also participated in one of the phase 3 pivotal studies.

1.3 Principal Findings

Studies 033C and 034C demonstrate the statistical significance of the effect of tazarotene cream 0.1% on the clinical improvement (one grade decrease from baseline) at Week 24 of fine wrinkling and mottled hyperpigmentation. In addition to the two primary endpoints, statistical significance was also obtained for the following secondary and "additional" endpoints: lentigines, elastosis, pore size, irregular depigmentation, investigator's overall integrated assessment, global response to treatment, and patient's overall self-assessment.

Following a request by the Agency, the sponsor also analyzed the response to treatment at Week 24 defined in three additional ways: (i) achieving no evidence of sign/symptom (grade 0), (ii) achieving no or minimal evidence (grade 0 or 1), or (iii) achieving at least two grades improvement from baseline. Very few subjects on either tazarotene or vehicle were able to achieve grade 0, and definition (i) did not distinguish between the treatment arms for many of the signs and symptoms. The results for the second two definitions of treatment response were generally consistent with the sponsor's planned analysis of one grade improvement. For definition (ii) no or minimal evidence, all of the endpoints which were significant in the sponsor's primary analysis were also significant under this definition, plus coarse wrinkling and tactile roughness also achieved significance.

The sponsor's protocols included a multiplicity adjustment for the two primary endpoints, and statistical significance is attained under this adjustment (p-values were less than 0.001 in both studies for clinical improvement and less than 0.005 for achieving grade 0 or 1). The sponsor conducted all secondary analyses without adjustment at $\alpha = 0.05$. However, for clinical improvement, all endpoints that were significant when no adjustment was applied, also remain significant when adjusted for multiplicity (p-values were less than or equal to 0.002). For treatment success, lentigines, elastosis, pore size, and irregular depigmentation remain significant when adjusted for multiplicity ($p \leq 0.003$), while course wrinkling and tactile roughness do not ($p \geq 0.040$). Table 1.1 displays the signs and symptoms of photoaging which are statistically significant, and the weeks for which significant results are obtained in both phase 3 studies ($p \leq 0.05$).

Table 1.1 – Significant Efficacy Endpoints

Endpoint	Clinical Improvement[a]	Treatment Success[b]
Fine Wrinkling	Weeks 8-24	Weeks 12-24
Mottled Hyperpigmentation	Weeks 2-24	Weeks 8-24
Lentigines	Weeks 4-24	Weeks 8-24
Elastosis	Weeks 12-24	Weeks 12-24
Pore Size	Weeks 12-24	Weeks 16-24
Irregular Depigmentation	Weeks 16-24	Weeks 12-24
Course Wrinkling	-	Weeks 16-24[c]
Tactile Roughness	-	Week 24[c]

[a] At least a one grade decrease in severity from baseline
[b] Achieving grade 0 or 1
[c] Endpoint not significant if secondary endpoints are adjusted for multiplicity

Thus, the sponsor's clinical program has statistically demonstrated that tazarotene cream 0.1% is effective for the following ——— ——— fine wrinkling, and mottled hyperpigmentation (primary endpoints), and lentigines, elastosis, pore size, and irregular depigmentation (secondary endpoints).

APPEARS THIS WAY
ON ORIGINAL

2 Statistical Review and Evaluation of Evidence

2.1 Introduction and Background

Tazarotene creams 0.1% and 0.05% were approved in September 2000 for the treatment of stable plaque psoriasis (original NDA 21-184), and tazarotene cream 0.1% was approved in October 2001 for the treatment of acne vulgaris (supplement 001). In this supplement, the sponsor has submitted safety and efficacy studies to support the use of tazarotene cream 0.1% in ———.

The following signs and symptoms of photodamage were assessed: fine wrinkling, mottled hyperpigmentation, lentigines, elastosis, pore size, irregular depigmentation, tactile roughness, coarse wrinkling, telangiectasia, and actinic keratoses. Also assessed were the investigator's overall integrated assessment, global response to treatment, and patient's overall self-assessment.

The sponsor conducted two pivotal phase 3 efficacy and safety studies, 190168-033C and 190168-034C. In addition, the sponsor conducted a phase 2 dose ranging study of various concentrations of tazarotene cream, (190168-025C), and a study to assess the inter- and intra-rater reliability of investigators assessing the signs of photodamage, using the sponsor's photonumeric guidelines (190168-037C).

2.2 Data Analyzed and Sources

Table 2.1 lists the studies included in the sponsor's clinical program, along with the level of statistical evaluation included in this review.

Table 2.1 - Clinical Program for Tazarotene Cream 0.1% for Photodamage

Study	*Type of Study*	*Number of Subjects*	*Level of Review*
025C	Dose Ranging	Taz. 0.01% (59) Taz. 0.025% (58) Taz. 0.05% (58) Taz. 0.1% (58) Tretinoin 0.05% (58) Vehicle (58)	Brief Evaluation
033C	Phase 3 Efficacy/Safety	Taz. 0.1% (283) Vehicle (280)	Full Evaluation
034C	Phase 3 Efficacy/Safety	Taz. 0.1% (284) Vehicle (284)	Full Evaluation
036C	Histological Safety Profile	Taz. 0.1% (25) Vehicle (24)	None
037C	Inter- and Intra-Rater Reliability	No treatment administered (40 subj./10 inv.)	Brief Evaluation

The sponsor provided the electronic data sets for the phase 3 efficacy and safety studies (033C and 034C) used in this review. The efficacy data set for each study is PHOTO.XPT (clinical improvement of at least one grade). Alternate definitions of treatment success can be found for each study in the data sets PHOTO2.XPT (clinical improvement of at least 2 grades), PHOTO3.XPT (treatment success defined as grade 0), and PHOTO4.XPT (treatment success defined as grade 0 or 1). Data sets are archived in the Electronic Document Room at \\CDSESUB1\N21184\S_002\2001-06-28\crt\datasets \033cderived and \034cderived.

2.3 Statistical Evaluation of Evidence on Efficacy and Safety

2.3.1 Studies 033C and 034C

2.3.1.1 Study Design

Studies 033C and 034C are phase 3, randomized, multi-center, double-blind, parallel group, vehicle-controlled safety and efficacy studies. Both studies involved once daily application of tazarotene cream or vehicle cream for 24 weeks. Study 033C also followed subjects for a 28 week open-label follow-up period. Data from the 28 week follow-up period were submitted as part of the 120-day safety update on 10/30/2001. For the double-blind portion of the study, the primary efficacy timepoint was Week 24. Thirteen efficacy endpoints were defined and classified in the protocol as either primary endpoints, secondary endpoints, or other endpoints. Fine wrinkling and mottled hyperpigmentation at Week 24 were designated as the primary efficacy endpoints. The secondary endpoints were lentigines and elastosis at Week 24. The remaining nine endpoints (pore size, irregular depigmentation, tactile roughness, coarse wrinkling, telangiectasia, actinic keratoses, investigator's overall integrated assessment, global response to treatment, and patient's overall self-assessment) were classified as "other" endpoints. Subjects were evaluated at baseline, and Weeks 2, 4, 8, 12, 16, 20, and 24. In addition to the Week 24 assessment, the sponsor also considers the efficacy of each endpoint at each of the intermediate time points.

Fine wrinkling, mottled hyperpigmentation, lentigines, elastosis, irregular depigmentation, tactile roughness, coarse wrinkling, and telangiectasia were all assessed on the scale 0 = none, 1 = minimal, 2 = mild, 3 = moderate, and 4 = severe. The investigator's overall integrated assessment used the same categories plus also included the category 5 = very severe. Pore size was assessed on the scale 0 = barely visible, 1 = very small, 2 = small, 3 = medium, and 4 = large. Actinic keratoses were counted. Global response to treatment was evaluated on the scale 0 = complete response, 1 = almost complete response (90% improvement), 2 = marked response (75% improvement), 3 = moderate response (50% improvement), 4 = slight response (25% improvement), 5 = no response, and 6 = condition worsened. The patient's overall self-assessment was evaluated on the scale 1 = much improved, 2 = somewhat improved, 3 = no change, 4 = somewhat worse, and 5 = much worse.

The sponsor defined treatment success for the primary and secondary endpoints as an improvement of at least one grade from baseline to Week 24 ("clinical improvement"). At the pre-NDA meeting held with the sponsor on February 21, 2001, the Agency requested that the following definitions of treatment success also be considered:

(1) Achieving a severity score of 0 at study endpoint
(2) Achieving a severity score of 0 or 1 at study endpoint
(3) Improvement of at least 2 grades from baseline to Week 24

At the End of Phase 2 Meeting held on August 20, 1999, the sponsor was advised that specific signs and symptoms "could be dichotomized, by taking the proportion of subjects who appear at the end of the study to have minimal involvement in that response, or, by taking the proportion with minimal or mild involvement. Alternatively, the Sponsor could consider the proportion who achieve at least a one or two step improvement from baseline." (pg. 4 of End of Phase 2 Meeting Minutes) Therefore, the sponsor's definition of clinical improvement is consistent with the advice provided at the End of Phase 2 Meeting. Treatment success, defined as achieving grade 0 or 1, is also considered in this review.

2.3.1.2 Statistical Methods

The primary analysis population is the intent-to-treat (ITT) population. The ITT population consisted of all randomized patients. Last observation carried forward (LOCF) was used to impute missing data. A per protocol population was not defined in the protocol and no per protocol analysis was carried out, however, the sponsor did conduct an analysis on the observed cases only, with no imputation for missing data.

Clinical improvement for the primary and secondary efficacy endpoints (fine wrinkling, mottled hyperpigmentation, elastosis, and lentigines) was analyzed with Cochran-Mantel-Haenszel (CMH) tests stratified by center. Treatment by center interactions were tested using the Breslow-Day test at $\alpha = 0.10$. If a significant interaction was found, a sensitivity analysis was performed by deleting two centers—the one most favoring tazarotene, and the one most favoring vehicle—and re-running the analysis.

Clinical improvement (at least one grade improvement) is also defined as the endpoint for pore size, irregular depigmentation, tactile roughness, coarse wrinkling, telangiectasia, and the overall integrated assessment. For the global response to treatment, moderate response or better (≥50% improvement) is considered a success. These variables, along with the definitions of success suggested by the Agency (at least 2 grades improvement, no evidence of sign/symptom, no or minimal evidence of sign/symptom) were all analyzed similarly with CMH tests stratified on center.

The number of actinic keratoses, the distribution of scores on the global response to treatment, and the distribution of scores on the patient's overall assessment were analyzed with CMH tests stratified on center using modified ridit scores to test for row mean score differences.

Since the sponsor proposed two primary endpoints, Hochberg's (1988) procedure was used to adjust for multiplicity. This procedure first considers the endpoint with the largest p-value. If the p-value < 0.05, then both hypotheses are rejected and statistical significance is attained for both endpoints. If the larger p-value > 0.05, but the smaller p-value < 0.025, then statistical significance is attained for the endpoint with the smaller p-value. Otherwise, no statistical significance is attained for either endpoint. No adjustments were made for multiple secondary endpoints and timepoints.

2.3.1.3 Patient Disposition and Demographics

Study 033C. Study 033C enrolled 563 patients, 283 in the tazarotene arm and 280 in the vehicle arm. The study was completed by 87.6 % of tazarotene and 93.9% of vehicle patients. Twenty patients (7.1%) in the tazarotene group and 1 patient (0.4%) in the vehicle group discontinued due to adverse events. Table 2.2 lists the disposition of enrolled subjects.

Study participants were overwhelmingly white and female. Eighty-nine percent of subjects were female and 96% of subjects were Caucasian. The average patient age was 56 years. No significant differences were found between the treatment groups in terms of age, race, or sex. Table A.1 in the appendix summarizes the age, race, and sex of study participants.

Study 033C involved 15 centers. Each center enrolled between 24 and 52 subjects.

Table 2.2 – Patient Disposition in Studies 033C and 034C

	Study 033C		Study 034C	
	Tazarotene	Vehicle	Tazarotene	Vehicle
Enrolled	283	280	284	284
Completed	248 (87.6%)	263 (93.9%)	255 (89.8%)	250 (88.0%)
Discontinued	35 (12.4%)	17 (6.1%)	29 (10.2%)	34 (12.0%)
Lack of Efficacy	0	0	0	4 (1.4%)
Adverse Event	20 (7.1%)	1 (0.4%)	10 (3.5%)	4 (1.4%)
Lost to Follow-up	4 (1.4%)	2 (0.7%)	2 (0.7%)	5 (1.8%)
Relocated/Pers. Reasons	9 (3.2%)	3 (1.1%)	14 (4.9%)	20 (7.0%)
Improper Entry	0	2 (0.7%)	0	0
Non-Compliance	1 (0.4%)	1 (0.4%)	0	1 (0.4%)
Concomitant Therapy	0	0	1 (0.4%)	0
Other	1 (0.4%)	1 (0.4%)	2 (0.7%)	0

Source: Table 1, file 033c.pdf, pg. 77, and file 034c.pdf, pg. 73

Study 034C. Study 034C enrolled 568 patients, 284 in the tazarotene arm and 284 in the vehicle arm. The study was completed by 89.8 % of tazarotene and 88.0% of vehicle patients. Ten patients (3.5%) in the tazarotene group and 4 patients (1.4%) in the vehicle group discontinued due to adverse events. Table 2.2 lists the disposition of enrolled subjects.

Study participants were overwhelmingly white and female. Eighty-six percent of subjects were female and 98% of subjects were Caucasian. The average patient age was 54 years. No significant differences were found between the treatment groups in terms of age, race, or sex. Table A.1 in the appendix summarizes the age, race, and sex of study participants.

Study 034C involved 15 centers. One center enrolled only 7 patients. The remaining centers each enrolled between 30 and 48 subjects.

2.3.1.4 Sponsor's Primary Efficacy Results

The primary endpoints proposed by the sponsor were clinical improvement in fine wrinkling and hyperpigmentation, defined as at least one grade improvement from baseline to Week 24. At baseline, subjects were required to have a score of at least 2 (mild) for these two variables, with one of the scores at least 3 (moderate). Tables A.2 and A.3 in the appendix list the baseline scores for fine wrinkling and hyperpigmentation.

The primary analysis population was the ITT population. The last observation was carried forward to the study endpoint for patients who discontinued the study. Clinical improvement (at least one grade reduction) was analyzed with a Cochran-Mantel-Haenszel test, stratified on center. Treatment by center interaction was tested with a Breslow-Day test at level 0.10. The primary efficacy results are presented in Table 2.3. Results for the two primary efficacy endpoints at all time points are presented in Tables A.4 – A.7 in the appendix. In the sponsor's analyses, no imputation of missing data was performed for the intermediate time points, observed cases only were analyzed.

Table 2.3 – Percentage of Patients with Clinical Improvement at Study Endpoint (Week 24) for Fine Wrinkling and Mottled Hyperpigmentation (ITT)

Endpoint	Study	Tazarotene	Vehicle	Trt. Effect p-value[a]	Interaction p-value[b]
Fine Wrinkling	033C	40.3%	16.1%	<0.001	0.013
	034C	58.1%	22.5%	<0.001	<0.001
Mottled Hyperpigmentation	033C	59.0%	17.9%	<0.001	0.581
	034C	81.7%	39.4%	<0.001	<0.001

[a] p-values based on CMH test stratified on center
[b] p-values based on Breslow-Day test
Source: Table 12, file 033c.pdf, pg. 149, and file 034c.pdf, pg. 138.

Since the p-values for fine wrinkling and mottled hyperpigmentation are <0.001 in each study, statistical significance is attained for both endpoints at the primary timepoint (Week 24) in each study. Under Hochberg's procedure, if the p-values for both endpoints are less than 0.05, then statistical significance for each endpoint is demonstrated. At the intermediate timepoints (Weeks 2, 4, 8, 16, and 20), p-values were less than 0.05 from Week 8 through Week 24 in Study 033C and Week 2 through Week 24 in Study 034C for fine wrinkling, and from Week 2 through Week 24 in both studies for mottled hyperpigmentation. (See Tables A.4 – A.7 in the appendix.)

Significant treatment by center interactions were detected in both studies for fine wrinkling, and in Study 034C for mottled hyperpigmentation. Figures B.1 – B.4 in the appendix display the clinical improvement rates for fine wrinkling and mottled hyperpigmentation by center. In Study 033C, Center 3260 had a higher clinical improvement rate with vehicle than with tazarotene for fine wrinkling. At Center 3260 7/20 (35%) of tazarotene subjects demonstrated clinical improvement for fine wrinkling, while 12/20 (60%) of vehicle subjects demonstrated clinical improvement. At all other centers in Studies 033C and 034C, tazarotene either equaled or exceeded the clinical improvement rates for vehicle in terms of fine wrinkling. For mottled hyperpigmentation, tazarotene clinical improvement rates equaled or exceeded the rates for vehicle at all centers in both studies.

Center 3259 (Study 033C) demonstrated the largest treatment effect for tazarotene for both fine wrinkling and mottled hyperpigmentation. For fine wrinkling, 7/12 (58%) of tazarotene subjects demonstrated clinical improvement, while 0/12 (0%) of vehicle subjects did. Similarly for mottled hyperpigmentation, 9/12 (75%) of tazarotene subjects demonstrated clinical improvement, while 0/12 (0%) of vehicle subjects did. The sponsor's study report reports the following protocol deviation for Center 3259 "All patients participating in the Week 24 therapeutic drug monitoring applied the open-label medication for the Week 24 therapeutic drug monitoring visit" (Appendix 16.2.2, file 033c.pdf, pg. 1071). According to the protocol, subjects were to enter the open-label phase of the study *after* the Week 24 visit. It is unclear what effect this protocol violation might have had on the Week 24 efficacy results.

The sponsor conducted sensitivity analyses for those endpoints with significant treatment by center interactions. In Study 033C, the sponsor re-analyzed clinical improvement for fine wrinkling after deleting the center favoring vehicle (3260), and the center that most favored tazarotene (3259). The results of the CMH test were still significant with the two centers removed ($p < 0.001$), and the Breslow-Day test was no longer significant ($p = 0.818$). [Source: Table 1.2, file 033c.pdf, pg. 515.] In Study 034C, the sponsor had to remove 6 centers (out of 15) from both the fine wrinkling and mottled hyperpigmentation analyses before non-significant results ($p > 0.10$) for the Breslow-Day test were achieved. [Source: Tables 1.4 and 2.4, file 034c.pdf, pg. 498 and 502.] This suggests that the interaction effect was not just due to results from only a few centers in Study 034C, but due to a more persistent hetereogeneity in effect size across centers.

2.3.1.5 Sponsor's Secondary Efficacy Results

The secondary endpoints proposed by the sponsor were clinical improvement in lentigines and elastosis, defined as at least one grade improvement from baseline to Week 24. These endpoints were analyzed in the same way as fine wrinkling and mottled hyperpigmentation. The efficacy results for clinical improvement in lentigines and elastosis are presented in Table 2.4. Week-by-week results are presented in Tables A.8 – A.11. P-values for both lentigines and elastosis at study endpoint are < 0.001 in both studies. P-values < 0.05 were observed for lentigines from Week 4 through 24 in Study

033C, and Week 2 through 24 in Study 034C. P-values < 0.05 were observed for elastosis from Week 12 through 24 in Study 033C, and Week 8 through 24 in Study 034C.

Table 2.4 – Percentage of Patients with Clinical Improvement at Study Endpoint (Week 24) for Lentigines and Elastosis (ITT)

Endpoint	Study	Tazarotene	Vehicle	Trt. Effect p-value[a]	Interaction p-value[b]
Lentigines	033C	50.2%	15.7%	<0.001	0.032
	034C	54.6%	23.9%	<0.001	0.010
Elastosis	033C	20.5%	4.6%	<0.001	0.121
	034C	28.5%	10.9%	<0.001	0.158

[a] p-values based on CMH test stratified on center
[b] p-values based on Breslow-Day test
Source: Tables 15 & 16, file 033c.pdf, pg. 152 & 153, and file 034c.pdf, pg. 141 & 142.

Significant treatment by center interactions ($p < 0.10$) were observed for lentigines in both studies. One center in Study 034C (3160) had a higher rate of clinical improvement in lentigines in the vehicle arm than in the tazarotene arm. At Center 3160 4/20 (20%) of tazarotene subjects demonstrated clinical improvement for lentigines, while 5/20 (25%) of vehicle subjects demonstrated clinical improvement. All other centers in Studies 033C and 034C had a higher success rate on the tazarotene arm than the vehicle arm for lentigines. To achieve Breslow-Day p-values > 0.10, two centers in Study 033C and 4 centers in Study 034C had to be removed from the analysis

2.3.1.6 Sponsor's Additional Efficacy Results

In addition to the four primary and secondary efficacy endpoints, the sponsor also evaluated 9 additional endpoints: pore size, irregular depigmentation, tactile roughness, coarse wrinkling, telangiectasia, actinic keratoses, investigator's overall integrated assessment, global response to treatment, and patient's overall self-assessment. Tables 2.5 to 2.7 display the efficacy results for the additional endpoints. Pore size, irregular depigmentation, Overall Investigator Assessment, Global Response to Treatment, and Patient's Overall Self-Assessment all have p-values < 0.05 for both studies.

Table 2.5 – Percentage of Patients with Clinical Improvement at Study Endpoint (Week 24) for Additional Efficacy Endpoints (ITT)

Endpoint	Study	Tazarotene	Vehicle	p-value[a]
Tactile Roughness	033C	44.2%	34.6%	0.005
	034C	44.4%	37.3%	0.055
Course Wrinkling	033C	13.1%	5.7%	0.002
	034C	14.4%	10.2%	0.075
Telangiectasia	033C	14.8%	11.8%	0.283
	034C	15.5%	13.0%	0.333

[a] p-values based on CMH test stratified on center
Source: Tables 17 - 23, file 033c.pdf, pg. 154 - 161, and file 034c.pdf, pg. 143 - 150.

Table 2.6– Percentage of Patients with Clinical Improvement at Study Endpoint (Week 24) for Additional Efficacy Endpoints (ITT)

Endpoint	Study	Tazarotene	Vehicle	p-value[a]
Pore Size	033C	27.2%	9.6%	<0.001
	034C	39.8%	18.0%	<0.001
Irregular Depigmentation	033C	19.8%	9.3%	<0.001
	034C	22.5%	13.4%	0.002
Overall Investigator Assessment	033C	32.6%	8.2%	<0.001
	034C	53.5%	16.5%	<0.001

[a] p-values based on CMH test stratified on center
Source: Tables 17 - 23, file 033c.pdf, pg. 154 - 161, and file 034c.pdf, pg. 143 - 150.

Table 2.7 – Additional Efficacy Results at Study Endpoint (Week 24) (ITT)

Endpoint	Study	Tazarotene	Vehicle	p-value[a]
Mean Change in Actinic Keratoses from Baseline	033C	-0.1	-0.2	0.294
	034C	-0.2	-0.3	0.690
Global Response to Treatment (>50% Improvement)	033C	36.8%	3.2%	<0.001
	034C	65.5%	18.9%	<0.001
Patient's Overall Self-Assessment (Much Improved)	033C	36.7%	5.8%	<0.001
	034C	30.5%	5.7%	<0.001

[a] p-values are from CMH test for row mean score differences with modified ridits, stratified by center.
Source: Tables 22, 24, & 26, file 033c.pdf, pg. 159 - 166, and file 034c.pdf, pg. 148- 155.

2.3.1.7 Sponsor's Efficacy Conclusions

Based on the results of Studies 033C and 034C, the sponsor claims they have demonstrated efficacy in terms of clinical improvement (at least one grade improvement from baseline to Week 24) for the endpoints listed in Table 2.8. The timepoints represent the week from which p-values < 0.05 were observed for the treatment effect from that week onward through Week 24 in both studies. The sponsor also claims to have demonstrated significance for the global response to treatment, and the patient's overall self-assessment.

Table 2.8 – Significant Efficacy Endpoints (Sponsor's Conclusion)

Endpoint	Week[a]	Endpoint	Week[a]
Fine Wrinkling	8	Pore Size	12
Mottled Hyperpigmentation	2	Irregular Depigmentation	16
Lentigines	4	Overall Integrated Assessment	8
Elastosis	12		

[a] Week from which the p-values for the endpoint were < 0.05 from that week on through Week 24

The sponsor's analysis at the intermediate timepoints was conducted on observed cases only, rather than on LOCF. This reviewer re-analyzed the intermediate timepoint data using LOCF for missing values. P-values from the LOCF analysis were comparable to

the observed case analysis. The weeks with p-values < 0.05 in both studies for each of the above signs and symptoms are identical under both the observed cases and LOCF analyses.

2.3.1.8 Alternate Efficacy Criteria

At the Agency's request, the sponsor conducted post-hoc analyses of alternate definitions of success for the signs and symptoms of photodamage. The alternate definitions of success were defined as

- Improvement of at least 2 grades from baseline to Week 24 (for those patients with at least grade 2 at baseline.
- Grade 0 (none) at Week 24 (for those patients with at least grade 2 at baseline).
- Grade 0 or 1 (none or minimal) at Week 24 (for those patients with at least grade 2 at baseline).

Tables 2.9 and 2.10 display the success rates for fine wrinkling and mottled hyperpigmentation for tazarotene and vehicle for the alternate definitions of success.

Table 2.9 – Alternate Definitions of Success for Fine Wrinkling (Week 24) (ITT)

Fine Wrinkling	Study	Tazarotene	Vehicle	Trt. Effect p-value[a]	Interaction p-value[b]
At least 2 grades improvement	033C	5.3%	1.4%	0.011	0.193
	034C	13.4%	4.9%	<0.001	0.005
Grade 0 (none)	033C	0.4%	0.0%	0.332	NA
	034C	1.4%	1.4%	0.619	0.230
Grade 0 or 1 (none or minimal)	033C	7.1%	2.1%	0.005	0.161
	034C	19.7%	7.4%	<0.001	<0.001

[a] p-values based on CMH test stratified on center
[b] p-values based on Breslow-Day test
Source: Table 13, file 033c.pdf, pg. 486 - 488, and file 034c.pdf, pg. 467 – 469.

Table 2.10 – Alternate Definitions of Success for Mottled Hyperpigmentation (Week 24) (ITT)

Mottled Hyperpigmentation	Study	Tazarotene	Vehicle	Trt. Effect p-value[a]	Interaction p-value[b]
At least 2 grades improvement	033C	17.3%	0.7%	<0.001	0.155
	034C	28.2%	9.5%	<0.001	0.030
Grade 0 (none)	033C	2.8%	0.0%	0.005	NA
	034C	4.6%	1.8%	0.030	0.506
Grade 0 or 1 (none or minimal)	033C	27.9%	6.8%	<0.001	0.113
	034C	42.6%	17.6%	<0.001	0.008

[a] p-values based on CMH test stratified on center
[b] p-values based on Breslow-Day test
Source: Table 14, file 033c.pdf, pg. 489 - 491, and file 034c.pdf, pg. 470- 472.

For fine wrinkling, when treatment success is defined as "at least 2 grades improvement" or "grade 0 or 1" the treatment effect is still statistically significant ($p \leq 0.011$). However, since very few subjects achieved grade 0 for fine wrinkling at Week 24 (1 tazarotene subject in Study 033C and 4 tazarotene and 4 vehicle subjects in Study 034C), no significant treatment effect for this definition of success is demonstrated. For mottled hyperpigmentation, all three alternate definitions of treatment success demonstrate statistical significance ($p \leq 0.030$), though again, few subjects achieved grade 0 for mottled hyperpigmentation.

Tables A.12 and A.13 in the Appendix list the week-by-week percentages of subjects achieving grade 0 or 1 for fine wrinkling and mottled hyperpigmentation. For these tables, LOCF was used to impute missing data. Significant results ($p < 0.05$) were observed in both studies for Weeks 12 – 24 for fine wrinkling, and Weeks 8 – 24 for mottled hyperpigmentation. Study 034C had significant treatment by center interactions ($p \leq 0.008$) for both fine wrinkling and mottled hyperpigmentation, when treatment success is defined as grade 0 or 1 at study endpoint. In Study 033C, no significant interaction was detected for fine wrinkling or mottled hyperpigmentation for this definition of success ($p \geq 0.113$). Figures B.5 and B.6 in the appendix display the percentage of subjects achieving grade 0 or 1 at study endpoint by center for fine wrinkling and mottled hyperpigmentation for Study 034C. Center 1421 in study 034C demonstrated the largest treatment effect among the centers for both fine wrinkling and mottled hyperpigmentation. Center 1421 also had the highest treatment effect among centers for clinical improvement of fine wrinkling.

Tables 2.11 and 2.12 present the efficacy results for the alternate definitions of success for lentigines and elastosis. Defining success as grades 0 or 1 continues to demonstrate efficacy for both endpoints in both studies ($p < 0.001$). When success is defined as at least two grades improvement, tazarotene still demonstrates a treatment effect for lentigines. However, defining success as at least two grades improvement for elastosis, failed to generate enough successes to demonstrate an effect for tazarotene. Similarly, very few subjects in either arm achieved grade 0 for both lentigines and elastosis. Interaction p-values of approximately 0.06 were observed for lentigines in Study 034C for two of the definitions of success, while no significant interactions were observed for elastosis. Tables A.14 and A.15 in the Appendix list the percentage of subjects achieving grade 0 or 1 by week for lentigines and elastosis. Significant results ($p < 0.05$) were observed in both studies for Weeks 8 – 24 for lentigines, and Weeks 12 – 24 for elastosis.

Table 2.11 – Alternate Definitions of Success for Lentigines (Week 24) (ITT)

Lentigines	Study	Tazarotene	Vehicle	Trt. Effect p-value[a]	Interaction p-value[b]
At least 2 grades improvement	033C	19.6%	2.0%	<0.001	0.705
	034C	19.8%	5.2%	<0.001	0.065
Grade 0 (none)	033C	2.4%	0.0%	0.034	NA
	034C	4.4%	0.5%	0.015	0.108
Grade 0 or 1 (none or minimal)	033C	37.8%	8.1%	<0.001	0.175
	034C	37.9%	14.1%	<0.001	0.064

[a] p-values based on CMH test stratified on center
[b] p-values based on Breslow-Day test
Source: Table 15, file 033c.pdf, pg. 492 - 494, and file 034c.pdf, pg. 473- 475.

Table 2.12 – Alternate Definitions of Success for Elastosis (Week 24) (ITT)

Elastosis	Study	Tazarotene	Vehicle	Trt. Effect p-value[a]	Interaction p-value[b]
At least 2 grades improvement	033C	0.6%	0.0%	0.359	NA
	034C	7.8%	2.7%	0.016	0.369
Grade 0 (none)	033C	0.0%	0.0%	NA	NA
	034C	4.6%	0.7%	0.024	0.621
Grade 0 or 1 (none or minimal)	033C	14.3%	2.6%	<0.001	0.226
	034C	22.2%	6.1%	<0.001	0.120

[a] p-values based on CMH test stratified on center
[b] p-values based on Breslow-Day test
Source: Table 16, file 033c.pdf, pg. 495 - 497, and file 034c.pdf, pg. 476- 478.

Table 2.13 displays the success rates for additional efficacy endpoints, where success is defined as achieving grade 0 or 1 at study endpoint (Week 24). Tactile roughness, course wrinkling, pore size, and irregular depigmentation all have p-values < 0.05 in both studies. Tactile roughness and course wrinkling are significant (using the sponsor's definition of $p < 0.05$ for secondary and other endpoints) when success is defined as grade 0 or 1 at study endpoint, while they just missed the cutoff ($0.05 < p < 0.10$) under the sponsor's planned endpoint of clinical improvement of 1 grade. However, tactile roughness and course wrinkling would not be considered significant under these definitions of success if any adjustment for multiple endpoints were taken into account. Tables A.16 through A.19 in the Appendix list the percentage of subjects achieving grade 0 or 1 by week for pore size, irregular depigmentation, course wrinkling, and tactile roughness. Significant results ($p < 0.05$) were observed in both studies for Weeks 16 – 24 for pore size, and Weeks 12 – 24 for irregular depigmentation. Tactile roughness is significant at Week 24, and course wrinkling is significant for Weeks 16 – 24.

Table 2.13 – Treatment Success (Grade 0 or 1) at Study Endpoint (Week 24) for Additional Efficacy Endpoints (ITT)

Endpoint	Study	Tazarotene	Vehicle	p-value[a]
Tactile Roughness	033C	69.6%	57.7%	0.011
	034C	54.1%	44.8%	0.043
Course Wrinkling	033C	8.3%	3.1%	0.040
	034C	10.2%	4.3%	0.020
Telangiectasia	033C	16.3%	8.7%	0.024
	034C	17.5%	16.9%	0.906
Pore Size	033C	15.6%	4.8%	<0.001
	034C	26.5%	13.3%	<0.001
Irregular Depigmentation	033C	29.9%	10.0%	<0.001
	034C	41.0%	23.1%	0.003

Percentages include only subjects with Grade 2 or higher at baseline for endpoint
[a] p-values based on CMH test stratified on center
Source: Tables 17 - 21, file 033c.pdf, pg. 498 - 510, and file 034c.pdf, pg. 479 – 493.

2.3.1.9 Multiplicity Issues

In the protocols for Studies 033C and 034C, the sponsor proposed Hochberg's (1988) method to adjust for multiplicity for the two primary endpoints. However, no adjustment was proposed for the two secondary and nine additional endpoints analyzed in the study. These 11 endpoints were all individually tested at 0.05. Of the 11 secondary and other endpoints, both secondary endpoints (lentigines and elastosis) and five other endpoints (pore size, irregular depigmentation, overall integrated assessment, global response to treatment, and patient's overall self-assessment) met the criteria of having p-values <. 0.05 in both studies (using the sponsor's protocol definitions of success). The 13 efficacy endpoints analyzed consist of 10 signs and symptoms (fine wrinkling, mottled hyperpigmentation, lentigines, elastosis, tactile roughness, course wrinkling, telangiectasia, pore size, irregular depigmentation, and actinic keratoses) and 3 global evaluations (overall integrated assessment, global response to treatment, and patient's overall self-assessment).

With 10 different signs and symptoms, the risk of making a type I error on at least one of the signs and symptoms is high, unless the error rate is adjusted for multiple tests. The sponsor's proposed label claims efficacy , by claiming that tazarotene is '

" (pg. 10 of proposed.pdf). If the sponsor intends to claim efficacy beyond the two primary efficacy endpoints, then it is worth exploring whether claims regarding the secondary endpoints hold up under multiplicity adjustments.

To assess the robustness of the sponsor's claim for efficacy in 4 non-primary signs and symptoms, this reviewer applied two post-hoc methods for adjusting for multiple endpoints. Since the sponsor's protocol used Hochberg's (1988) method to adjust for

multiplicity in the two designated primary endpoints, one reasonable extension would be to apply Hochberg's method to the secondary endpoints. Since the secondary endpoints would be evaluated only if significance is attained for at least one primary endpoint, the multiplicity method need only adjust for the 8 secondary signs and symptoms.

To extend Hochberg's method, the p-values for n endpoints are sorted from largest to smallest ($p_{(n)}, \ldots, p_{(1)}$). Starting with the largest p-value, each p-value $p_{(i)}$ is compared in sequence to $\alpha/(n-i+1)$, until the first time $p_{(i^*)} < \alpha/(n-i^*+1)$. At this point testing stops and the tests corresponding to $p_{(1)}, \ldots, p_{(i^*)}$ are rejected, and $p_{(i^*+1)}, \ldots, p_{(n)}$ are accepted. Table 2.14 presents the p-values and results of the Hochberg procedure for Studies 033C and 034C. Results are provided for both clinical improvement (one grade decrease) and treatment success (grade 0 or 1). The shaded cells represent the endpoints which are significant under Hochberg's procedure.

Table 2.14 – Significant Secondary Endpoints using Hochberg's (1988) Procedure

	Clinical Improvement			
	033C		034C	
$\alpha/(n-i+1)$	*Endpoint*	*p-value*[a]	*Endpoint*	*p-value*[a]
0.0063	Lentigines	<0.001	Lentigines	<0.001
0.0071	Elastosis	<0.001	Elastosis	<0.001
0.0083	Pore Size	<0.001	Pore Size	<0.001
0.0100	Irr. Depig.	<0.001	Irr. Depig.	0.002
0.0125	Course Wrin.	0.002	Tact. Rough.	0.055
0.0167	Tact. Rough.	0.005	Course Wrin.	0.075
0.0250	Telang.	0.283	Telang.	0.333
0.0500	Act. Ker.	0.294	Act. Ker.	0.690

	Treatment Success			
	033C		034C	
$\alpha/(n-i+1)$	*Endpoint*	*p-value*[a]	*Endpoint*	*p-value*[a]
0.0063	Lentigines	<0.001	Lentigines	<0.001
0.0071	Elastosis	<0.001	Elastosis	<0.001
0.0083	Pore Size	<0.001	Pore Size	<0.001
0.0100	Irr. Depig.	<0.001	Irr. Depig.	0.003
0.0125	Tact. Rough.	0.011	Course Wrin.	0.020
0.0167	Telang.	0.024	Tact. Rough.	0.043
0.0250	Course Wrin.	0.040	Act.Ker.	0.690
0.0500	Act. Ker.	0.294	Telang.	0.906

[a] p-values based on CMH test stratified on center
Source: Reviewer Analysis

For both clinical improvement and treatment success, the endpoints which are significant in both studies under this procedure are lentigines, elastosis, pore size, and irregular depigmentation. For clinical improvement, these are the same endpoints which the sponsor found to be significant without adjusting for multiplicity. Applying the even

more conservative Bonferroni adjustment yields the same results. The same 4 endpoints all have p-values less than $\alpha/8 = 0.00625$ in both studies. Thus, even though these methods for adjusting for multiplicity have been applied post-hoc, the fact that the significance conclusions hold up under the most conservative adjustment for multiplicity (Bonferroni) provides assurance that the significant treatment effects are unlikely to be due to chance.

2.3.1.10 Per Protocol Analyses

A per protocol population was neither specified in the protocol nor analyzed by the sponsor. However, the sponsor did conduct an analysis of the observed cases (no imputation for missing data) at Week 24. Week 24 observed cases results for the primary and secondary efficacy endpoints (clinical improvement) can be found in Tables A.4 – A.11 in the appendix. For these four endpoints, all p-values were less than 0.001 at Week 24 in both studies. Results are similar with success defined as Grade 0 or 1 at Week 24, with all p-values ≤ 0.007 for the four endpoints in both studies (Source: Tables 13-21, file 033c.pdf, pg. 486 - 512, and file 034c.pdf, pg. 467 – 493). Thus the observed case analysis does not differ substantially from the ITT analysis, and all conclusions remain the same.

2.3.1.11 Subgroup Analyses

Results of subgroup analyses based on age (< 40, 40 – 65, > 65), gender (male, female), and race (white, non-white) are presented in Tables A.20 and A.21 in the appendix for clinical improvement in fine wrinkling and mottled hyperpigmentation. The majority of subjects were white, female, and in the 40 to 65 age group. Treatment effects were significant in the larger subgroups (40 – 65, female, white). Treatment effects in the smaller subgroups were either significant, or trended in the direction favoring tazarotene. Some of the subgroups were not statistically significant due to small sample sizes. It should be noted, however, that this is a post-hoc analysis, and the studies were not powered to test for efficacy in subgroups.

Subgroup analysis results for treatment success (grade 0 or 1) for fine wrinkling and mottled hyperpigmentation are presented in Tables A.22 and A.23 in the appendix. Treatment effects were significant in favor of tazarotene in the larger subgroups (40 – 65, female, white). Treatment effects in the smaller subgroups were either significant or trended in the direction favoring tazarotene, except for non-white subjects, where the trend favored vehicle in at least one study for both endpoints. However, the number of non-white subjects in each study was small (25 subjects in Study 033C, and 13 subjects in Study 034C).

2.3.1.12 Safety Assessment

Treatment duration for subjects in Studies 033C and 034C ranged from 4 to 220 days. In Study 033C, the average exposure to tazarotene was 162 days, and the average exposure to vehicle was 166 days. In Study 034C, the average exposure to tazarotene was 162

days, and the average exposure to vehicle was 161 days. In Study 033C, 80.2% of tazarotene subjects were exposed for at least 168 days (24 weeks), as were 79.2% of tazarotene subjects in Study 034C. In the vehicle arm, 82.1% of subjects in Study 033C, and 73.6% of subjects in Study 034C were exposed for at least 168 days.

Adverse Events for Studies 033C and 034C are summarized in Tables 2.15 and 2.16. The information in Table 2.16 has been pooled across both studies. Each study had significantly higher percentages of patients with all adverse events and treatment related adverse events on the tazarotene arm than the vehicle arm. A similar number of patients on each arm experienced treatment unrelated adverse events. The majority of treatment related events were in the skin and appendages system. In the skin and appendages system, rates for desquamation, erythema, burning sensation, dry skin, pruritis, skin irritation, irritant contact dermatitis, rash, and stinging sensation were significantly higher for tazarotene than for vehicle.

Table 2.15 – Adverse Events

	Study 033C			Study 034C		
	Tazar.	Vehicle	p-value[a]	Tazar.	Vehicle	p-value[a]
All Adverse Events	83.4%	52.5%	<0.001	81.3%	35.2%	<0.001
Treatment Related	70.0%	13.2%	<0.001	72.5%	8.8%	<0.001
Treatment Unrelated	44.2%	44.3%	0.978	34.2%	29.2%	0.207

[a] p-values based on chi-square test
Source: Table 30, file 033c.pdf, pg. 171 and file 034c.pdf, pg. 160

Table 2.16 – Adverse Events Reported by at least 3% of Patients in Either Treatment Arm, Pooled Across Studies 033C and 034C

Adverse Event	Tazarotene (N=567)	Vehicle (N=564)	p-value[a]
BODY AS A WHOLE			
infection	45 (7.9%)	45 (8.0%)	0.979
RESPRIRATORY SYSTEM			
sinus infection	20 (3.5%)	16 (2.8%)	0.508
rhinitis	13 (2.3%)	17 (3.0%)	0.450
SKIN AND APPENDAGES			
desquamation	224 (39.5%)	11 (2.0%)	<0.001
erythema	190 (33.5%)	14 (2.5%)	<0.001
burning sensation on skin	143 (25.2%)	1 (0.2%)	<0.001
dry skin	90 (15.9%)	15 (2.7%)	<0.001
pruritis	54 (9.5%)	7 (1.2%)	0.005
skin irritation	54 (9.5%)	3 (0.5%)	<0.001
irritant contact dermatitis	47 (8.3%)	6 (1.1%)	<0.001
rash	19 (3.4%)	7 (1.2%)	0.018
stinging sensation on skin	19 (3.4%)	1 (0.2%)	<0.001
acne	16 (2.8%)	18 (3.2%)	0.716

[a] p-values based on either chi-square or Fisher's exact test
Source: Table 8.8.5.2, file iss.pdf, pg. 17

2.3.1.13 Randomization

The randomization lists and actual treatment allocation lists with date of enrollment have been provided in the sponsor's study reports. Several centers assigned one or more patient numbers out of sequence. In Study 033C, 7 centers assigned at least one subject number out of sequence. Center 3263 assigned patient numbers at the screening visit, rather than at the baseline visit. In Study 034C, 5 centers assigned at least one subject number out of sequence. Investigator 2762 inadvertently assigned two subjects to the same patient number. The second subject was discontinued from the study due to mis-randomization. Due to the magnitude of the treatment effect observed in these studies and the fact that the number of out of sequence randomizations was small, any impact of the out of sequence randomizations should be negligible.

2.3.1.14 Reviewer Conclusions

The sponsor has demonstrated statistical significance for the two protocol-specified primary efficacy endpoints, clinical improvement in fine wrinkling and mottled hyperpigmentation at Week 24 ($p < 0.001$). Significant results ($p < 0.05$) were observed for fine wrinkling from Week 8 to Week 24 and for mottled hyperpigmentation from Week 2 to Week 24 in both studies. For the sponsor's designated secondary endpoints, clinical improvement in lentigines and elastosis at Week 24, statistical significance was also demonstrated ($p < 0.001$). Lentigines were significant ($p < 0.05$) from Week 4 to Week 24, and elastosis was significant from Week 12 to Week 24 in both studies.

Of the remaining six signs and symptoms of photoaging evaluated in Studies 033C and 034C, two demonstrated statistical significance. Clinical improvement in pore size had p-values less than 0.001 and irregular depigmentation had p-values less than or equal to 0.002 at Week 24. For the earlier evaluations, pore size was significant from Week 12 to Week 24, and irregular depigmentation was significant from Week 16 to Week 24 in both studies. The studies also demonstrated significance in the Overall Investigator Assessment, the Global Response to Treatment, and the Patient's Overall Self-Assessment ($p < 0.001$).

Since a large number of secondary/other endpoints were evaluated (8 signs and symptoms and 3 global evaluations) this reviewer also analyzed the secondary/other signs and symptoms endpoints with a multiplicity adjustment. The signs and symptoms that are significant in the unadjusted analysis are also significant in the analysis which takes into account multiplicity adjustments (Bonferroni or Hochberg): lentigines, elastosis, pore size, and irregular depigmentation. The fact that the adjusted analysis yields the same conclusion as the unadjusted analysis gives support to the sponsor's conclusion of efficacy in 6 signs and symptoms of photoaging.

At the Agency's request, the sponsor also conducted efficacy analyses where success is defined as achieving grade 0 or 1 at endpoint. The primary endpoints, fine wrinkling and mottled hyperpigmentation are statistically significant at Week 24 ($p \leq 0.005$). For the

secondary and additional endpoints, lentigines, elastosis, pore size, and irregular depigmentation were also significant ($p \leq 0.003$). In addition, two endpoints were significant at $\alpha = 0.05$ under this definition of success, that were not significant using clinical improvement, tactile roughness ($p \leq 0.043$) and course wrinkling ($p \leq 0.040$). However, these two endpoints are not significant if an adjustment for multiple secondary endpoints is applied. Treatment success is significant ($p < 0.05$) for fine wrinkling from Weeks 12 – 24, for mottled hyperpigmentation from Weeks 8 – 24, for lentigines from Weeks 8 – 24, for elastosis from Weeks 12 – 24, for pore size from Weeks 16 – 24, and for irregular depigmentation from Weeks 12 – 24.

2.3.2 Study 025C

2.3.2.1 Study Design

Study 025C was a phase 2, randomized, investigator blind, multi-center, vehicle and active control safety and efficacy dose ranging study for tazarotene cream. The study had 6 arms: tazarotene cream in concentrations 0.1%, 0.05%, 0.025%, and 0.01%, vehicle cream, and tretinoin cream 0.05%. The treatment design was similar to that used in Studies 033C and 034C. The treatment duration was 24 weeks with evaluation visits at Weeks 0, 2, 4, 8, 12, 16, 20, and 24, plus a follow-up visit at Week 26. The study enrolled 349 patients, with 58 or 59 subjects per arm. The study was conducted at 6 investigative centers.

The signs and symptoms endpoints (fine wrinkling, mottled hyperpigmentation, lentigines, elastosis, irregular depigmentation, course wrinkling, telangiectasia, pore size, and actinic keratoses) were the same as in Studies 033C and 034C. The only difference was that the signs and symptoms were measured on 6-point scales (none, minimal, mild, moderate, severe, very severe) rather than 5-point scales (Studies 033C and 034C did not use the 'very severe' category). Overall integrated assessment, global response to treatment, and patient's self assessment were also recorded. Clinical improvement was defined as at least one grade decrease in scale from baseline to Week 24.

2.3.2.2 Efficacy results

Clinical improvement was analyzed with a Cochran-Mantel-Haenszel test, stratified on center. As this was a phase 2 dose ranging study, no adjustments were made for multiple doses or multiple endpoints. All tests were conducted at $\alpha = 0.05$. Table 2.17 lists the clinical improvement rates for the signs and symptoms at Week 24. The results for fine wrinkling, mottled hyperpigmentation, lentigines, and elastosis are consistent with the results found in Studies 033C and 034C. This table also presents some evidence of a dose dependent response for some of the endpoints.

Table 2.17 – Clinical Improvement at Week 24 from Study 025C

Endpoint	Taz. 0.1% N=58	Taz. 0.05% N=58	Taz. 0.025% N=58	Taz. 0.01% N=59	Tret. 0.05% N=58	Veh. N=58	p-value[a]
Fine Wrinkling	53.4%	48.3%	34.5%	45.8%	53.4%	19.0%	<0.001
Mottled Hyperpig.	86.2%	81.0%	69.0%	72.9%	84.5%	67.2%	0.050
Lentigines	73.2%	77.2%	68.4%	71.2%	85.5%	49.1%	<0.001
Elastosis	54.9%	43.1%	44.0%	28.8%	43.1%	30.6%	0.011
Tactile Roughness	61.1%	64.3%	54.5%	65.5%	61.1%	63.6%	0.715
Course Wrinkling	14.0%	10.5%	10.9%	10.2%	10.5%	3.6%	0.587
Telangiectasia	25.5%	31.5%	16.0%	26.9%	24.5%	20.0%	0.471
Pore Size	41.4%	46.6%	42.1%	42.4%	35.7%	30.9%	0.515
Irr. Depig.	68.6%	72.2%	57.9%	51.4%	65.6%	54.3%	0.307

[a] Among-group p-values are from Cochran-Mantel-Haenszel test.
Source: Tables 10 – 18, file 025c.pdf, pg 90 – 123.

2.3.2.3 Safety Results

Table 2.18 presents the incidence of treatment related adverse events in the skin and appendages system. Local irritation rates for tazarotene cream 0.1% are consistent with those observed in Studies 033C and 034C. Local irritation rates appear to be dose dependent for a number of symptoms, however the rates for tazarotene cream 0.1% and 0.05% are similar.

Table 2.18 – Treatment Related Adverse Events in the Skin and Appendages System in Study 025C

	Taz. 0.1% N=58	Taz. 0.05% N=58	Taz. 0.025% N=58	Taz. 0.01% N=59	Tret. 0.05% N=58	Veh. N=58
Desquamation	22 (37.9%)	27 (46.6%)	13 (22.4%)	8 (13.6%)	13 (22.4%)	5 (8.6%)
Burning skin	17 (29.3%)	22 (37.9%)	9 (15.5%)	3 (5.1%)	11 (19.0%)	3 (5.2%)
Erythema	16 (27.6%)	19 (32.8%)	13 (22.4%)	7 (11.9%)	7 (12.1%)	5 (8.6%)
Pruritus	12 (20.7%)	6 (10.3%)	5 (8.6%)	2 (3.4%)	2 (3.4%)	0 (0.0%)
Dry skin	11 (19.0%)	14 (24.1%)	14 (24.1%)	10 (16.9%)	11 (19.0%)	5 (8.6%)
Irritation skin	10 (17.2%)	7 (12.1%)	7 (12.1%)	2 (3.4%)	4 (6.9%)	1 (1.7%)
Irritant CD	7 (12.1%)	5 (8.6%)	7 (12.1%)	2 (3.4%)	5 (8.6%)	0 (0.0%)
Stinging skin	6 (10.3%)	3 (5.2%)	1 (1.7%)	2 (3.4%)	5 (8.6%)	0 (0.0%)
Papules	4 (6.9%)	3 (5.2%)	3 (5.2%)	1 (1.7%)	3 (5.2%)	3 (5.2%)
Rash	3 (5.2%)	3 (5.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Acne	3 (5.2%)	7 (12.1%)	2 (3.4%)	4 (6.8%)	4 (6.9%)	6 (10.3%)
Seborrhea	2 (3.4%)	1 (1.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Fissure skin	2 (3.4%)	0 (0.0%)	0 (0.0%)	1 (1.7%)	3 (5.2%)	0 (0.0%)
Excoriation	1 (1.7%)	0 (0.0%)	2 (3.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Skin tightness	1 (1.7%)	2 (3.4%)	0 (0.0%)	1 (1.7%)	0 (0.0%)	1 (1.7%)
Reaction skin	1 (1.7%)	1 (1.7%)	0 (0.0%)	2 (3.4%)	2 (3.4%)	1 (1.7%)

Source: Table 12.2.3.1, file 025c.pdf, pg 58.

2.3.2.4 Sponsor's Conclusions

Based on the results of Study 025C, the sponsor selected 0.1% concentration of tazarotene cream to pursue in phase 3 studies. The sponsor also used this study to select fine wrinkling and mottled hyperpigmentation as primary endpoints, and lentigines and elastosis as secondary endpoints for the phase 3 studies.

2.3.3 Study 037C

2.3.3.1 Study Design

Study 037C was designed to evaluate the inter- and intra-rater reliability among investigators evaluating signs of photodamage, particularly fine wrinkling and mottled hyperpigmentation. Twenty subjects with varying degrees of fine wrinkling, and 20 subjects with varying degrees of mottled hyperpigmentation were recruited to be evaluated by 10 investigators. Each subject was evaluated by each investigator for either fine wrinkling or mottled hyperpigmentation on the scale 0 = none, 1 = minimal, 2 = mild, 3 = moderate, and 4 = severe. Each investigator also rated all 40 subjects' lentigines, elastosis, irregular depigmentation, course wrinkling, telangiectasia, and pore size and counted actinic keratoses. These signs and symptoms were all evaluated in the same way as in Studies 033C and 034C. For mottled hyperpigmentation and fine wrinkling, investigators were provided with the sponsor's photonumeric guidelines, as they were in the phase 3 studies. These guidelines provide 3 example photographs for each grade for the target symptom. Subjects were selected for enrollment by the medical monitor so that in the monitor's opinion, at least two subjects in each group (fine wrinkling and mottled hyperpigmentation) represented each grade level (1- 4) for the subject's target symptom. Investigators were blinded to the distribution of subjects. Each investigator evaluated each subject twice. The ordering of the subjects was randomized for each evaluation session. All evaluations were conducted on the same day at the same center. In the study, the minimum time between evaluations on the same patient by the same investigator was 2.6 hours and the maximum time was 4.2 hours (mean 3.4 hours).

2.3.3.2 Statistical Methods

Inter-rater reliability of each symptom was summarized with Kendall's coefficient of concordance (Kendall's W), using the average of the two evaluations. Two-sided 95% confidence intervals based on the chi-square approximation were also computed. The inter-rater reliability estimates and confidence intervals are presented in Table 2.19. The primary endpoints of fine wrinkling and mottled hyperpigmentation have estimates of 0.926 and 0.879, respectively. Thus, there was fairly consistent inter-rater agreement on these two endpoints. For the secondary endpoints, lentigines, elastosis, telangiectasia, and course wrinkling had estimates in the range 0.70 – 0.87, while the remaining endpoints had estimates less than 0.54.

Table 2.19 – Kendall's W Statistics and Confidence Intervals of Inter-Rater Agreement

Sign/Symptom	Stat./Conf. Int.	Sign/Symptom	Stat./Conf. Int.
Fine Wrinkling	0.926 (0.535, 1.000)	Course Wrinkling	0.872 (0.586, 1.000)
Mottled Hyperpig.	0.879 (0.508, 1.000)	Tactile Roughness	0.531 (0.357, 0.875)
Lentigines	0.704 (0.473, 1.000)	Irregular Depig.	0.483 (0.325, 0.796)
Elastosis	0.804 (0.541, 1.000)	Pore Size	0.517 (0.348, 0.853)
Telangiectasia	0.778 (0.523, 1.000)	Actinic Keratoses	0.540 (0.363, 0.891)

95% Confidence intervals based on chi-square approximation.
Fine Wrinkling and Mottled Hyperpigmentation were assessed on independent samples of 20 patients each. The remaining symptoms were assessed on the combined sample of 40 patients. Each patient was evaluated twice by 10 raters.
Source: Tables 15 – 16, file 037c.pdf, pg 66-67.

To estimate intra-rater reliability, a kappa statistic was computed for each rater. A weighted kappa statistic was then calculated to estimate the intra-rater agreement across all 10 raters. The weighted kappa statistics are presented in Table 2.20, along with the 95% confidence intervals from the normal approximation. As with inter-rater agreement, the fine wrinkling and mottled hyperpigmentation evaluations demonstrated fairly consistent results, with kappa statistics of 0.929 and 0.911 respectively. These estimates were followed by slightly lower estimates for lentigines, elastosis, telangiectasia and course wrinkling, and lower still estimates for tactile roughness, irregular depigmentation, pore size, and actinic keratoses.

Table 2.20 – Kappa Statistics and Confidence Intervals of Intra-Rater Agreement

Sign/Symptom	Stat./Conf. Int.	Sign/Symptom	Stat./Conf. Int.
Fine Wrinkling	0.929 (0.820, 1.000)	Course Wrinkling	0.841 (0.706, 0.977)
Mottled Hyperpig.	0.911 (0.811, 1.000)	Tactile Roughness	0.353 (0.183, 0.523)
Lentigines	0.734 (0.560, 0.908)	Irregular Depig.	0.557 (0.317, 0.797)
Elastosis	0.834 (0.690, 0.977)	Pore Size	0.615 (0.425, 0.806)
Telangiectasia	0.750 (0.581, 0.920)	Actinic Keratoses	0.646 (0.433, 0.858)

95% Confidence intervals based on the normal approximation.
Fine Wrinkling and Mottled Hyperpigmentation were assessed on independent samples of 20 patients each. The remaining symptoms were assessed on the combined sample of 40 patients. Each patient was evaluated twice by 10 raters.
Source: Tables 17 – 18, file 037c.pdf, pg 68-69.

2.3.3.3 Sponsor's Conclusions

The sponsor claims that the results of this study demonstrate that fine wrinkling and mottled hyperpigmentation, the primary efficacy variables from the pivotal clinical trials, have good inter- and intra-rater agreement when used with the sponsor's photonumeric guidelines. In addition, good inter- and intra-rater agreement was observed for lentigines, elastosis, telangiectasia, and course wrinkling, with less inter- and intra-rater agreement for tactile roughness, irregular depigmentation, pore size, and actinic keratoses. Thus, the sponsor claims that their photonumeric guidelines for fine wrinkling and mottled hyperpigmentation, which also were used in the phase 3 trials, are highly effective for helping investigators achieve consistently reliable ratings. In addition, evaluators were able to achieve reliable scores for the signs and symptoms for which the photo-numeric guidelines were not provided.

2.4 Statistical Evaluation of Collective Evidence

The sponsor has conducted two phase 3 studies (033C and 034C) to support the efficacy and safety of tazarotene cream 0.1% [redacted]

[redacted] These two studies support the claim that tazarotene cream 0.1% is statistically superior to vehicle in terms of clinical improvement of the primary endpoints fine wrinkling and mottled hyperpigmentation. Statistical significance was also obtained for the secondary endpoints of clinical improvement for lentigines, elastosis, pore size, and irregular depigmentation. Clinical improvement is defined as at least one grade improvement from baseline to Week 24 where each sign is evaluated on the scale 0 = none, 1 = minimal, 2 = mild, 3 = moderate, and 4 = severe.

The sponsor designated fine wrinkling and mottled hyperpigmentation as the primary endpoints. Hochberg's (1988) method was applied to the two primary endpoints to adjust for multiplicity. Of the remaining endpoints, lentigines and elastosis were designated as secondary endpoints by the sponsor, with the remaining endpoints (pore size, irregular depigmentation, course wrinkling, telangiectasia, tactile roughness, actinic keratoses, overall investigator assessment, global response to treatment, and patient's overall self-assessment) classified as "other" endpoints. Of the secondary and other endpoints, clinical improvement in lentigines, elastosis, pore size, irregular depigmentation are significant along with the overall investigator assessment, global response to treatment, and the patient's overall self-assessment. The same secondary signs and symptoms endpoints are significant whether or not an adjustment for multiplicity is applied. At Week 24, p-values for lentigines, elastosis, pore size, and irregular depigmentation are all less than or equal to 0.002 in both studies, while the p-values for tactile roughness, course wrinkling, telangiectasia, and actinic keratoses are greater than or equal to 0.055 in at least one study. The overall investigator's assessment is significant at $p < 0.001$ at Week 24. If treatment success is defined as achieving grade 0 or 1 at study endpoint, the same signs and symptoms are statistically significant as with clinical improvement: fine wrinkling, mottled hyperpigmentation, lentigines, elastosis, pore size, and irregular depigmentation.

The results from the phase 2 dose ranging study (025C) support the results of the phase 3 studies. In study 025C, a significant treatment effects for clinical improvement of fine wrinkling, mottled hyperpigmentation, lentigines, and elastosis were observed. No adjustments for multiplicity were made in this exploratory phase 2 study designed to select the endpoints and tazarotene concentration for the phase 3 studies.

Study 037C evaluated the inter- and intra-rater variability of the signs and symptoms of photoaging. Fine wrinkling and mottled hyperpigmentation demonstrated a good degree of inter- and intra-rater agreement. In addition, good inter- and intra-rater agreement was observed for lentigines, elastosis, telangiectasia, and course wrinkling, with less inter- and intra-rater agreement for tactile roughness, irregular depigmentation, pore size, and actinic keratoses. This study supports the results of the phase 3 studies by providing evidence that the primary and secondary efficacy evaluations can be reproduced with reasonable reliability by both the same and different investigators for a given subject.

Evidence from these four clinical studies statistically support the claim that tazarotene cream 0.1% [redacted]

/S/

Kathleen Fritsch, Ph.D.
Mathematical Statistician, Biometrics III

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

cc:
Archival NDA 21-184/SE1 - 002
HFD-540/Dr. Wilken
HFD-540/Dr. Walker
HFD-540/Dr. Huene
HFD-540/Ms. Bhatt
HFD-700/Dr. Anello
HFD-725/Dr. Huque
HFD-725/Dr. Alosh
HFD-725/Dr. Fritsch

NDA 21-184/SE1-002/Stat Rev N 21184 002.doc 2/11/02

2.5 Appendix of Additional Tables

Table A.1 – Demographic Data for Studies 033C and 034C

	Study 033C		Study 034C	
	Tazarotene (N = 283)	Vehicle (N = 280)	Tazarotene (N = 284)	Vehicle (N=284)
Age (Mean ± SD)	56.2 ± 11.4	56.2 ± 10.9	53.7 ± 11.7	53.9 ± 11.6
p-value[a]	$p = 0.931$		$p = 0.571$	
Race				
Caucasian	268 (94.7%)	270 (96.4%)	279 (98.2%)	276 (97.2%)
Asian	6 (2.1%)	3 (1.1%)	0	2 (0.7%)
Hispanic	7 (2.5%)	5 (1.1%)	5 (1.8%)	5 (1.8%)
Other	2 (0.7%)	2 (0.7%)	0	1 (0.4%)
p-value[b]	$p = 0.723$		$p = 0.607$	
Sex				
Female	253 (89.4%)	249 (88.9%)	249 (87.7%)	239 (84.2%)
Male	30 (10.6%)	31 (11.1%)	35 (12.3%)	45 (15.8%)
p-value[b]	$p = 0.857$		$p = 0.228$	

[a] p-values based on the Wilcoxon rank sum test
[b] p-values based on Pearson's chi-square test
Source: Table 4, file 033c.pdf, pg. 83, and file 034c.pdf, pg. 79.

Table A.2 – Baseline Scores for Fine Wrinkling in Studies 033C and 034C

	Study 033C		Study 034C	
	Tazarotene (N = 283)	Vehicle (N = 280)	Tazarotene (N = 284)	Vehicle (N = 284)
None	0	0	0	0
Minimal	0	0	0	0
Mild	54 (19.1%)	41 (14.6%)	61 (21.5%)	58 (20.4%)
Moderate	169 (59.7%)	171 (61.1%)	148 (52.1%)	145 (51.1%)
Severe	60 (21.2%)	68 (24.3%)	75 (26.4%)	81 (28.5%)
p-value[a]	$p = 0.154$		$p = 0.522$	

[a] p-values based on CMH test for row mean score differences with modified ridits, stratified by center
Source: Table 10, file 033c.pdf, pg. 144, and file 034c.pdf, pg. 133

Table A.3 – Baseline Scores for Mottled Hyperpigmentation in Studies 033C and 034C

	Study 033C		Study 034C	
	Tazarotene (N = 283)	Vehicle (N = 280)	Tazarotene (N = 284)	Vehicle (N = 284)
None	0	0	0	0
Minimal	0	1 (0.4%)	0	0
Mild	93 (32.9%)	106 (37.9%)	84 (29.6%)	110 (38.7%)
Moderate	158 (55.8%)	150 (53.6%)	167 (58.8%)	136 (48.9%)
Severe	32 (11.3%)	23 (8.2%)	33 (11.6%)	38 (13.4%)
p-value[a]	$p = 0.150$		$p = 0.130$	

[a] p-values based on CMH test for row mean score differences with modified ridits, stratified by center
Source: Table 10, file 033c.pdf, pg. 144, and file 034c.pdf, pg. 133.

Table A.4 – Percentage of Patients with Clinical Improvement at Each Visit for Fine Wrinkling in Study 033C (Observed Cases, Study Endpoint is LOCF)

Visit[a]	Tazarotene (N=283)	Vehicle (N=280)	Trt. Effect p-value[b]	Interaction p-value[c]
Week 2	5/267 (1.9%)	2/264 (0.8%)	0.276	0.174
Week 4	22/267 (8.2%)	14/267 (5.2%)	0.141	0.036
Week 8	43/262 (16.4%)	27/271 (10.0%)	0.018	<0.001
Week 12	60/254 (23.6%)	30/256 (11.7%)	<0.001	0.075
Week 16	82/245 (33.5%)	35/248 (14.1%)	<0.001	0.083
Week 20	92/240 (38.3%)	45/254 (17.7%)	<0.001	0.037
Week 24	108/244 (44.3%)	43/249 (17.3%)	<0.001	0.010
Study Endpoint	114/283 (40.3%)	45/280 (16.1%)	<0.001	0.013

[a] Week 2-24 data is observed cases, Study Endpoint is last observation on each patient (LOCF, ITT)
[b] p-values based on CMH test stratified on center [c] p-values based on Breslow-Day test
Source: Table 13, file 033c.pdf, pg. 150.

Table A.5 – Percentage of Patients with Clinical Improvement at Each Visit for Fine Wrinkling in Study 034C (Observed Cases, Study Endpoint is LOCF)

Visit[a]	Tazarotene (N=284)	Vehicle (N=284)	Trt. Effect p-value[b]	Interaction p-value[c]
Week 2	11/259 (4.2%)	3/259 (1.2%)	0.021	0.018
Week 4	35/267 (13.1%)	12/262 (4.6%)	<0.001	0.035
Week 8	71/259 (27.4%)	31/250 (12.4%)	<0.001	0.015
Week 12	99/249 (39.8%)	36/243 (14.8%)	<0.001	0.001
Week 16	119/242 (49.2%)	42/240 (17.5%)	<0.001	0.001
Week 20	141/238 (59.2%)	54/235 (23.0%)	<0.001	<0.001
Week 24	155/246 (63.0%)	58/240 (24.2%)	<0.001	<0.001
Study Endpoint	165/284 (58.1%)	64/284 (22.5%)	<0.001	<0.001

[a] Week 2-24 data is observed cases, Study Endpoint is last observation on each patient (LOCF, ITT)
[b] p-values based on CMH test stratified on center [c] p-values based on Breslow-Day test
Source: Table 13, file 034c.pdf, pg. 139

Table A.6 – Percentage of Patients with Clinical Improvement at Each Visit for Mottled Hyperpigmentation in Study 033C (Observed Cases, Study Endpoint is LOCF)

Visit[a]	Tazarotene (N=283)	Vehicle (N=280)	Trt. Effect p-value[b]	Interaction p-value[c]
Week 2	18/267 (6.7%)	2/264 (0.8%)	<0.001	0.939
Week 4	59/267 (22.1%)	16/267 (6.0%)	<0.001	0.228
Week 8	102/262 (38.9%)	33/271 (12.2%)	<0.001	0.698
Week 12	123/254 (48.4%)	43/256 (16.8%)	<0.001	0.316
Week 16	138/245 (56.3%)	46/248 (18.5%)	<0.001	0.343
Week 20	142/240 (59.2%)	53/254 (20.9%)	<0.001	0.525
Week 24	155/244 (63.5%)	48/249 (19.3%)	<0.001	0.728
Study Endpoint	167/283 (59.0%)	50/280 (17.9%)	<0.001	0.581

[a] Week 2-24 data is observed cases, Study Endpoint is last observation on each patient (LOCF, ITT)
[b] p-values based on CMH test stratified on center [c] p-values based on Breslow-Day test
Source: Table 14, file 033c.pdf, pg. 151

Table A.7 – Percentage of Patients with Clinical Improvement at Each Visit for Mottled Hyperpigmentation in Study 034C (Observed Cases, Study Endpoint is LOCF)

Visit[a]	Tazarotene (N=284)	Vehicle (N=284)	Trt. Effect p-value[b]	Interaction p-value[c]
Week 2	30/259 (11.6%)	9/259 (3.5%)	<0.001	0.002
Week 4	84/267 (31.5%)	39/262 (14.9%)	<0.001	0.104
Week 8	147/259 (56.8%)	61/250 (24.4%)	<0.001	<0.001
Week 12	173/249 (69.5%)	79/243 (32.5%)	<0.001	<0.001
Week 16	189/242 (78.1%)	92/240 (38.3%)	<0.001	<0.001
Week 20	202/238 (84.9%)	100/235 (42.6%)	<0.001	<0.001
Week 24	213/246 (86.6%)	103/240 (42.9%)	<0.001	0.014
Study Endpoint	232/284 (81.7%)	112/284 (39.4%)	<0.001	<0.001

[a] p-values based on CMH test stratified on center [b] p-values based on Breslow-Day test
[c] Week 2-24 data is observed cases, Study Endpoint is last observation on each patient (LOCF)
Source: Table 14, file 034c.pdf, pg. 140

Table A.8 – Percentage of Patients with Clinical Improvement at Each Visit for Lentigines in Study 033C (Observed Cases, Study Endpoint is LOCF)

Visit[a]	Tazarotene (N=283)	Vehicle (N=280)	Trt. Effect p-value[b]	Interaction p-value[c]
Week 2	13/267 (4.9%)	6/264 (2.3%)	0.122	0.567
Week 4	50/267 (18.7%)	19/267 (7.1%)	<0.001	0.318
Week 8	85/262 (32.4%)	24/271 (8.9%)	<0.001	0.680
Week 12	117/254 (46.1%)	32/256 (12.5%)	<0.001	0.046
Week 16	125/245 (51.0%)	31/248 (12.5%)	<0.001	0.056
Week 20	133/240 (55.4%)	40/254 (15.7%)	<0.001	0.028
Week 24	135/244 (55.3%)	40/249 (16.1%)	<0.001	0.048
Study Endpoint	142/283 (50.2%)	44/280 (15.7%)	<0.001	0.032

[a] Week 2-24 data is observed cases, Study Endpoint is last observation on each patient (LOCF, ITT)
[b] p-values based on CMH test stratified on center [c] p-values based on Breslow-Day test
Source: Table 15, file 033c.pdf, pg. 152

Table A.9 – Percentage of Patients with Clinical Improvement at Each Visit for Lentigines in Study 034C (Observed Cases, Study Endpoint is LOCF)

Visit[a]	Tazarotene (N=284)	Vehicle (N=284)	Trt. Effect p-value[b]	Interaction p-value[c]
Week 2	19/259 (7.3%)	3/259 (1.2%)	<0.001	0.310
Week 4	50/267 (18.7%)	14/262 (5.3%)	<0.001	0.429
Week 8	80/259 (30.9%)	23/250 (9.2%)	<0.001	0.153
Week 12	106/249 (42.6%)	35/243 (14.4%)	<0.001	0.003
Week 16	124/242 (51.2%)	50/240 (20.8%)	<0.001	0.023
Week 20	137/238 (57.6%)	53/235 (22.6%)	<0.001	0.024
Week 24	145/246 (58.9%)	66/240 (27.5%)	<0.001	0.044
Study Endpoint	155/284 (54.6%)	68/284 (23.9%)	<0.001	0.010

[a] Week 2-24 data is observed cases, Study Endpoint is last observation on each patient (LOCF, ITT)
[b] p-values based on CMH test stratified on center [c] p-values based on Breslow-Day test
Source: Table 15, file 034c.pdf, pg. 141

Table A.10 – Percentage of Patients with Clinical Improvement at Each Visit for Elastosis in Study 033C (Observed Cases, Study Endpoint is LOCF)

Visit[a]	Tazarotene (N=283)	Vehicle (N=280)	Trt. Effect p-value[b]	Interaction p-value[c]
Week 2	1/267 (0.4%)	3/264 (1.1%)	0.327	0.496
Week 4	11/267 (4.1%)	5/267 (1.9%)	0.116	0.149
Week 8	19/262 (7.3%)	13/271 (4.8%)	0.221	0.067
Week 12	32/254 (12.6%)	12/256 (4.7%)	0.001	0.090
Week 16	37/245 (15.1%)	10/248 (4.0%)	<0.001	0.084
Week 20	42/240 (17.5%)	10/254 (3.9%)	<0.001	0.173
Week 24	57/243 (23.5%)	13/249 (5.2%)	<0.001	0.095
Study Endpoint	58/283 (20.5%)	13/280 (4.6%)	<0.001	0.121

[a] Week 2-24 data is observed cases, Study Endpoint is last observation on each patient (LOCF, ITT)
[b] p-values based on CMH test stratified on center [c] p-values based on Breslow-Day test
Source: Table 16, file 033c.pdf, pg. 153

Table A.11 – Percentage of Patients with Clinical Improvement at Each Visit for Elastosis in Study 034C (Observed Cases, Study Endpoint is LOCF)

Visit[a]	Tazarotene (N=284)	Vehicle (N=284)	Trt. Effect p-value[b]	Interaction p-value[c]
Week 2	8/259 (3.1%)	6/259 (2.3%)	0.509	0.498
Week 4	21/267 (7.9%)	13/262 (5.0%)	0.161	0.101
Week 8	32/259 (12.4%)	20/250 (8.0%)	0.046	0.132
Week 12	52/249 (20.9%)	20/243 (8.2%)	<0.001	0.287
Week 16	61/242 (25.2%)	22/240 (9.2%)	<0.001	0.481
Week 20	73/238 (30.7%)	24/235 (10.2%)	<0.001	0.561
Week 24	73/246 (29.7%)	26/240 (10.8%)	<0.001	0.307
Study Endpoint	81/284 (28.5%)	31/284 (10.9%)	<0.001	0.158

[a] Week 2-24 data is observed cases, Study Endpoint is last observation on each patient (LOCF, ITT)
[b] p-values based on CMH test stratified on center [c] p-values based on Breslow-Day test
Source: Table 16, file 033c.pdf, pg. 142

Table A.12– Percent Success (Grade 0 or 1) at Each Visit for Fine Wrinkling (ITT, LOCF)

Visit	Study 033C Tazarotene (N=283)[a]	Vehicle (N=280)[a]	p-value[b]	Study 034C Tazarotene (N=284)[a]	Vehicle (N=284)[a]	p-value[b]
Week 2	2 (0.7%)	0 (0.0%)	0.173	2 (0.7%)	0 (0.0%)	0.168
Week 4	2 (0.7%)	1 (0.4%)	0.593	4 (1.4%)	1 (0.4%)	0.187
Week 8	8 (2.8%)	3 (1.1%)	0.140	8 (2.8%)	4 (1.4%)	0.251
Week 12	11 (3.9%)	3 (1.1%)	0.031	23 (8.1%)	7 (2.5%)	0.002
Week 16	13 (4.6%)	3 (1.1%)	0.012	35 (12.3%)	10 (3.5%)	<0.001
Week 20	15 (5.3%)	6 (2.1%)	0.048	49 (17.3%)	18 (6.3%)	<0.001
Week 24	20 (7.1%)	6 (2.1%)	0.005	56 (19.7%)	21 (7.4%)	<0.001

[a] Number of subjects with baseline grade ≥ 2 [b] p-values based on CMH test stratified on center
Source: Reviewer analysis

Table A.13– Percent Success (Grade 0 or 1) at Each Visit for Mottled Hyperpigmenation (ITT, LOCF)

Visit	Study 033C Tazarotene (N=283)[a]	Vehicle (N=279)[a]	p-value[b]	Study 034C Tazarotene (N=284)[a]	Vehicle (N=284)[a]	p-value[b]
Week 2	5 (1.8%)	2 (0.7%)	0.285	5 (1.8%)	2 (0.7%)	0.234
Week 4	18 (6.3%)	2 (0.7%)	<0.001	20 (7.0%)	10 (3.5%)	0.053
Week 8	35 (12.4%)	10 (3.6%)	<0.001	37 (13.0%)	18 (6.3%)	0.005
Week 12	48 (17.0%)	15 (5.4%)	<0.001	67 (23.6%)	28 (9.9%)	<0.001
Week 16	65 (23.0%)	17 (6.1%)	<0.001	94 (33.1%)	44 (15.5%)	<0.001
Week 20	74 (26.1%)	20 (7.1%)	<0.001	113 (39.8%)	48 (16.9%)	<0.001
Week 24	79 (27.9%)	19 (6.8%)	<0.001	121 (42.6%)	50 (17.6%)	<0.001

[a] Number of subjects with baseline grade ≥ 2 [b] p-values based on CMH test stratified on center
Source: Reviewer analysis

Table A.14– Percent Success (Grade 0 or 1) at Each Visit for Lentigines (ITT, LOCF)

Visit	Study 033C Tazarotene (N=209)[a]	Vehicle (N=198)[a]	p-value[b]	Study 034C Tazarotene (N=227)[a]	Vehicle (N=213)[a]	p-value[b]
Week 2	2 (1.0%)	2 (1.0%)	0.951	4 (1.8%)	0 (0.0%)	0.054
Week 4	18 (8.6%)	5 (2.5%)	0.006	11 (4.8%)	4 (1.9%)	0.081
Week 8	38 (18.2%)	10 (5.1%)	<0.001	25 (11.0%)	6 (2.8%)	<0.001
Week 12	54 (25.8%)	10 (5.1%)	<0.001	52 (22.9%)	9 (4.2%)	<0.001
Week 16	67 (32.1%)	14 (7.1%)	<0.001	68 (30.0%)	18 (8.5%)	<0.001
Week 20	73 (34.9%)	15 (7.6%)	<0.001	80 (35.2%)	25 (11.7%)	<0.001
Week 24	79 (37.8%)	16 (8.1%)	<0.001	86 (37.9%)	30 (14.1%)	<0.001

[a] Number of subjects with baseline grade ≥ 2
[b] p-values based on CMH test stratified on center
Source: Reviewer analysis

Table A.15– Percent Success (Grade 0 or 1) at Each Visit for Elastosis (ITT, LOCF)

Visit	Study 033C Tazarotene (N=161)[a]	Vehicle (N=154)[a]	p-value[b]	Study 034C Tazarotene (N=153)[a]	Vehicle (N=147)[a]	p-value[b]
Week 2	1 (0.6%)	2 (1.3%)	0.548	4 (2.6%)	1 (0.7%)	0.233
Week 4	5 (3.1%)	1 (0.6%)	0.104	10 (6.5%)	5 (3.4%)	0.316
Week 8	9 (5.6%)	2 (1.3%)	0.032	12 (7.8%)	5 (3.4%)	0.173
Week 12	15 (9.3%)	2 (1.3%)	0.002	18 (11.8%)	7 (4.8%)	0.048
Week 16	16 (9.9%)	2 (1.3%)	0.001	25 (16.3%)	7 (4.8%)	0.002
Week 20	18 (11.2%)	3 (1.9%)	0.001	32 (20.9%)	8 (5.4%)	<0.001
Week 24	23 (14.3%)	4 (2.6%)	<0.001	34 (22.2%)	9 (6.1%)	<0.001

[a] Number of subjects with baseline grade ≥ 2
[b] p-values based on CMH test stratified on center
Source: Reviewer analysis

Table A.16– Percent Success (Grade 0 or 1) at Each Visit for Pore Size (ITT, LOCF)

Visit	Study 033C Tazarotene (N=205)[a]	Vehicle (N=208)[a]	p-value[b]	Study 034C Tazarotene (N=196)[a]	Vehicle (N=196)[a]	p-value[b]
Week 2	5 (2.4%)	2 (1.0%)	0.300	6 (3.1%)	5 (2.6%)	0.674
Week 4	12 (5.8%)	7 (3.4%)	0.286	13 (6.6%)	11 (5.6%)	0.532
Week 8	14 (6.8%)	9 (4.3%)	0.343	26 (13.3%)	15 (7.7%)	0.025
Week 12	16 (7.8%)	8 (3.8%)	0.108	34 (17.3%)	15 (7.7%)	0.001
Week 16	23 (11.2%)	9 (4.3%)	0.007	45 (22.9%)	22 (11.2%)	<0.001
Week 20	29 (14.1%)	9 (4.3%)	<0.001	48 (24.5%)	23 (11.7%)	<0.001
Week 24	32 (15.6%)	10 (4.8%)	<0.001	52 (26.5%)	26 (13.3%)	<0.001

[a] Number of subjects with baseline grade ≥ 2
[b] p-values based on CMH test stratified on center
Source: Reviewer analysis

Table A.17– Percent Success (Grade 0 or 1) at Each Visit for Irregular Depigmentation (ITT, LOCF)

Visit	Study 033C Tazarotene (N=87)[a]	Vehicle (N=90)[a]	p-value[b]	Study 034C Tazarotene (N=100)[a]	Vehicle (N=104)[a]	p-value[b]
Week 2	1 (1.1%)	3 (3.3%)	0.125	4 (4.0%)	5 (4.8%)	0.661
Week 4	5 (5.7%)	6 (6.7%)	0.665	12 (12.0%)	9 (8.7%)	0.200
Week 8	14 (16.1%)	5 (5.6%)	0.064	22 (22.0%)	13 (12.5%)	0.028
Week 12	16 (18.4%)	6 (6.7%)	0.030	31 (31.0%)	16 (15.4%)	0.001
Week 16	18 (20.7%)	8 (8.9%)	0.039	35 (35.0%)	19 (18.3%)	0.001
Week 20	21 (24.1%)	8 (8.9%)	0.003	39 (39.0%)	24 (23.1%)	0.005
Week 24	26 (29.9%)	9 (10.0%)	<0.001	41 (41.0%)	24 (23.1%)	0.003

[a] Number of subjects with baseline grade ≥ 2 [b] p-values based on CMH test stratified on center
Source: Reviewer analysis

Table A.18– Percent Success (Grade 0 or 1) at Each Visit for Course Wrinkling (ITT, LOCF)

Visit	Study 033C Tazarotene (N=205)[a]	Vehicle (N=194)[a]	p-value[b]	Study 034C Tazarotene (N=196)[a]	Vehicle (N=187)[a]	p-value[b]
Week 2	1 (0.5%)	0 (0.0%)	0.371	1 (0.5%)	1 (0.5%)	0.911
Week 4	4 (2.0%)	1 (0.5%)	0.245	5 (2.6%)	1 (0.5%)	0.146
Week 8	5 (2.4%)	1 (0.5%)	0.177	8 (4.1%)	1 (0.5%)	0.027
Week 12	9 (4.4%)	1 (0.5%)	0.020	12 (6.1%)	1 (0.5%)	0.002
Week 16	12 (5.9%)	1 (0.5%)	0.006	18 (9.2%)	3 (1.6%)	<0.001
Week 20	18 (8.8%)	4 (2.1%)	0.006	21 (10.7%)	5 (2.7%)	<0.001
Week 24	17 (8.3%)	6 (3.1%)	0.040	19 (9.7%)	8 (4.3%)	0.032

[a] Number of subjects with baseline grade ≥ 2 [b] p-values based on CMH test stratified on center
Source: Reviewer analysis

Table A.19– Percent Success (Grade 0 or 1) at Each Visit for Tactile Roughness (ITT, LOCF)

Visit	Study 033C Tazarotene (N=125)[a]	Vehicle (N=123)[a]	p-value[b]	Study 034C Tazarotene (N=135)[a]	Vehicle (N=134)[a]	p-value[b]
Week 2	18 (14.4%)	15 (12.2%)	0.418	16 (11.9%)	24 (17.9%)	0.028
Week 4	32 (25.6%)	38 (30.9%)	0.425	35 (25.9%)	44 (32.8%)	0.085
Week 8	56 (44.8%)	65 (52.8%)	0.264	45 (33.3%)	45 (33.6%)	0.659
Week 12	75 (60.0%)	63 (51.2%)	0.059	54 (40.0%)	48 (35.8%)	0.355
Week 16	80 (64.0%)	69 (56.1%)	0.083	65 (48.1%)	50 (37.3%)	0.052
Week 20	84 (67.2%)	78 (63.4%)	0.299	67 (49.6%)	55 (41.0%)	0.057
Week 24	87 (69.6%)	71 (57.7%)	0.011	74 (54.8%)	60 (44.8%)	0.032

[a] Number of subjects with baseline grade ≥ 2 [b] p-values based on CMH test stratified on center
Source: Reviewer analysis

Table A.20 – Subgroup Analyses for Fine Wrinkling (Clinical Improvement) in Studies 033C and 034C

	033C Tazarotene % (N)	033C Vehicle % (N)	033C p-value	034C Tazarotene % (N)	034C Vehicle % (N)	034C p-value
Age						
< 40	28% (18)	11% (19)	0.232	67% (24)	19% (26)	0.001
40 – 65	44% (202)	16% (200)	<0.001	59% (209)	24% (209)	<0.001
> 65	32% (63)	18% (61)	0.098	51% (51)	18% (49)	<0.001
Gender						
Female	42% (253)	17% (249)	<0.001	57% (249)	22% (239)	<0.001
Male	30% (30)	7% (31)	0.022	66% (35)	24% (45)	<0.001
Race						
White	39% (268)	16% (270)	<0.001	59% (279)	23% (276)	<0.001
Non-White	60% (15)	20% (10)	0.099	20% (5)	13% (8)	>0.999

[a] p-values based on Fisher's exact test
Source: Table 13, file 033c.pdf, pg. 331-337, and file 034c.pdf, pg. 313-319.

Table A.21 – Subgroup Analyses for Mottled Hyperpigmentation (Clinical Improvement) in Studies 033C and 034C

	033C			034C		
	Tazarotene % (N)	Vehicle % (N)	p-value[a]	Tazarotene % (N)	Vehicle % (N)	p-value[a]
Age						
< 40	39% (18)	21% (19)	0.295	88% (24)	58% (26)	0.028
40 – 65	60% (202)	18% (200)	<0.001	81% (209)	37% (209)	<0.001
> 65	62% (63)	18% (61)	<0.001	80% (51)	41% (49)	<0.001
Gender						
Female	59% (253)	19% (259)	<0.001	83% (249)	39% (239)	<0.001
Male	60% (30)	10% (31)	<0.001	74% (35)	40% (45)	0.003
Race						
White	59% (268)	18% (270)	<0.001	82% (279)	40% (276)	<0.001
Non-White	67% (15)	10% (10)	0.012	80% (5)	38% (8)	0.266

[a] p-values based on Fisher's exact test
Source: Table 14, file 033c.pdf, pg. 370-376, and file 034c.pdf, pg. 352-358.

Table A.22 – Subgroup Analyses for Fine Wrinkling (Grade 0 or 1) in Studies 033C and 034C

	033C			034C		
	Tazarotene % (N)	Vehicle % (N)	p-value[a]	Tazarotene % (N)	Vehicle % (N)	p-value[a]
Age						
< 40	22.2% (18)	5.3% (19)	0.180	50.0% (24)	11.5% (26)	0.005
40 – 65	7.4% (202)	2.5% (200)	0.037	19.6% (209)	8.1% (209)	0.001
> 65	1.6% (63)	0% (61)	>0.999	5.9% (51)	2.0% (49)	0.618
Gender						
Female	7.5% (253)	2.4% (249)	0.012	20.5% (249)	7.5% (239)	<0.001
Male	3.3% (30)	0% (31)	0.492	14.3% (35)	6.7% (45)	0.288
Race						
White	6.0% (268)	1.9% (270)	0.015	20.1% (279)	7.3%(276)	<0.001
Non-White	26.7% (15)	10.0% (10)	0.615	0% (5)	12.5% (8)	>0.999

[a] p-values based on Fisher's exact test
Source: Reviewer Analysis

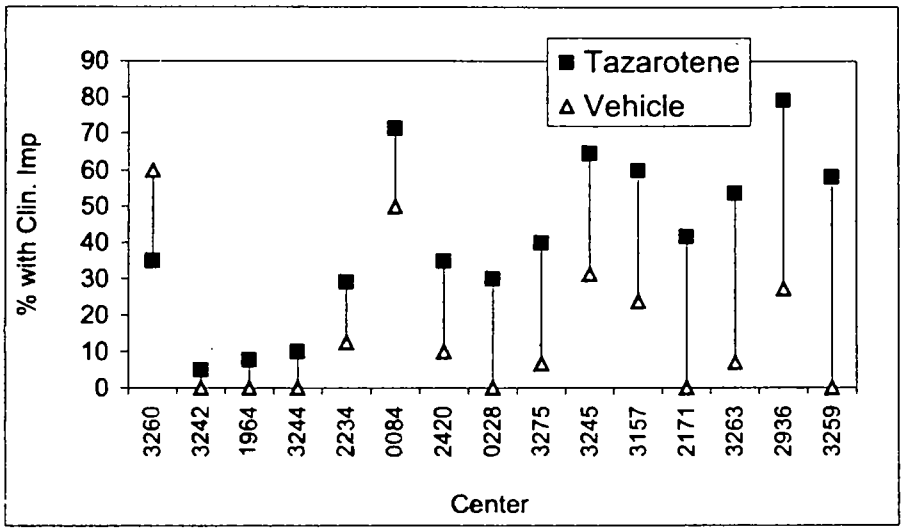
Table A.23 – Subgroup Analyses for Mottled Hyperpigmentation (Grade 0 or 1) in Studies 033C and 034C

	033C			034C		
	Tazarotene % (N)	Vehicle % (N)	p-value[a]	Tazarotene % (N)	Vehicle % (N)	p-value[a]
Age						
< 40	16.7% (18)	0% (19)	0.105	50.0% (24)	23.1% (26)	0.077
40 – 65	31.7% (202)	7.5% (200)	<0.001	43.1% (209)	17.2% (209)	<0.001
> 65	19.1% (63)	6.7% (60)	0.594	37.3% (51)	16.3% (49)	0.024
Gender						
Female	29.3% (253)	7.3% (248)	<0.001	43.0% (249)	17.6% (239)	<0.001
Male	16.7% (30)	3.2% (31)	0.080	40.0% (35)	17.8% (45)	0.043
Race						
White	29.1% (268)	6.7% (269)	<0.001	43.0% (279)	17.4% (276)	<0.001
Non-White	6.7% (15)	10.0% (10)	>0.999	20.0% (5)	25.0% (8)	>0.999

[a] p-values based on Fisher's exact test
Source: Reviewer Analysis

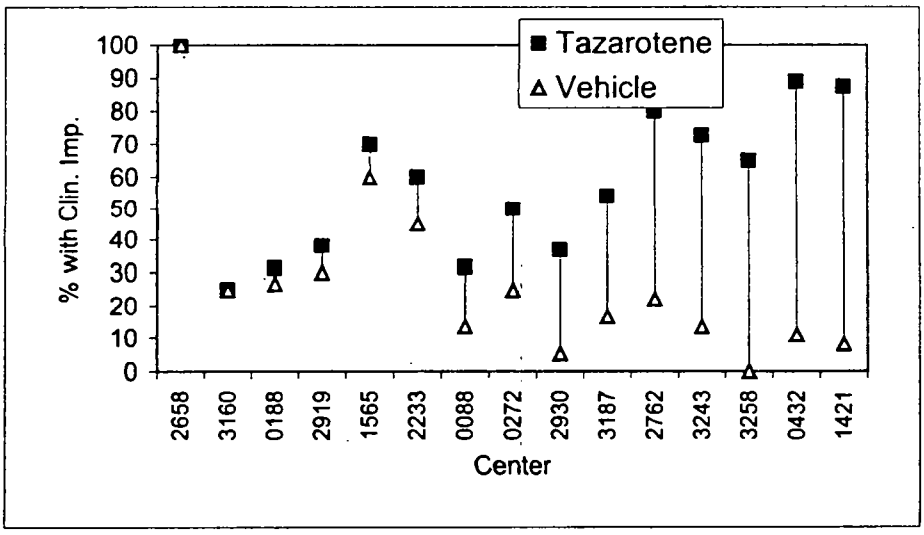
2.6 Appendix of Figures

Figure B.1 – Clinical Improvement for Fine Wrinkling by Center in Study 033C



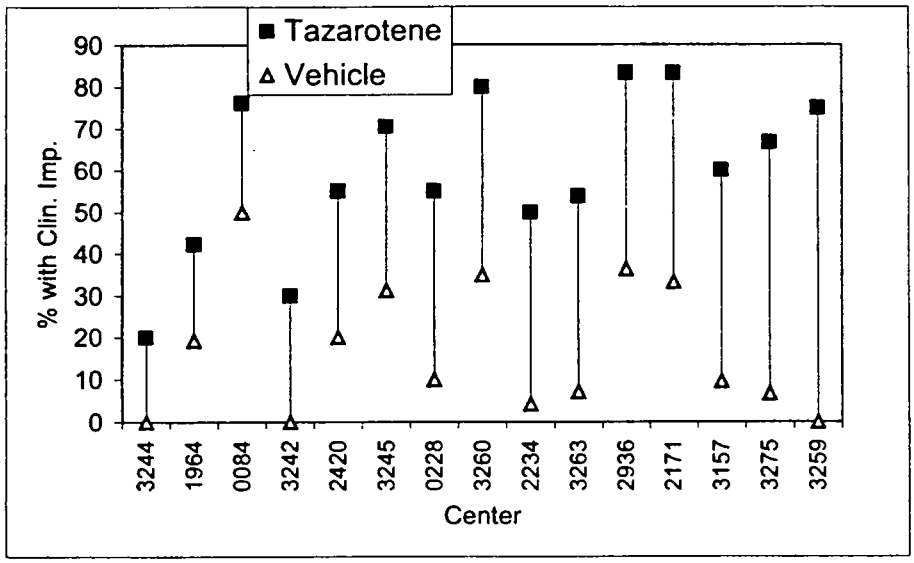
Source for Data: Table 1.1, file 033.pdf, pg 514 (Reviewer graphic)
Centers Sorted by treatment difference
Breslow-Day p-value = 0.013

Figure B.2 – Clinical Improvement for Fine Wrinkling by Center in Study 034C



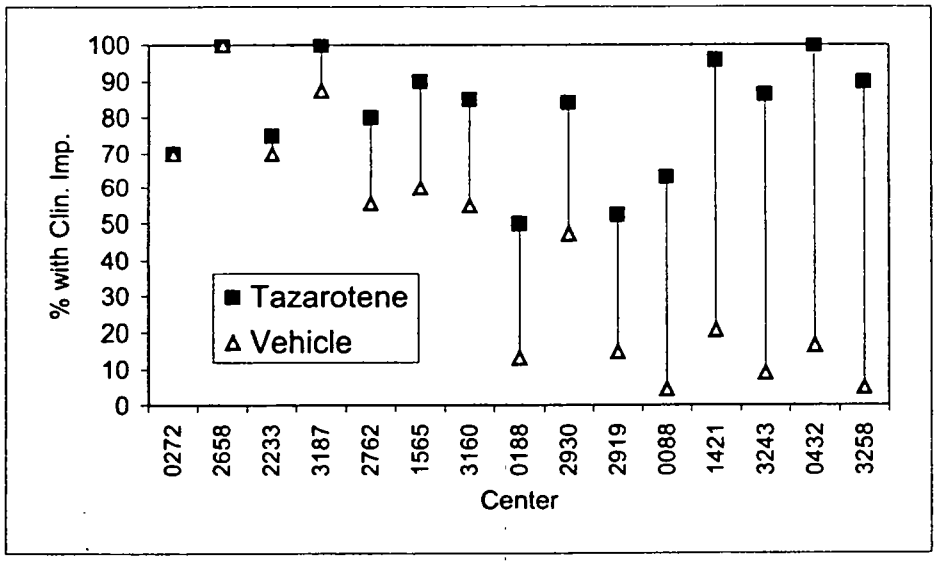
Source for Data: Table 1.1, file 034.pdf, pg 495 (Reviewer graphic)
Centers Sorted by treatment difference
Breslow-Day p-value < 0.001

Figure B.3 – Clinical Improvement for Mottled Hyperpigmentation by Center in Study 033C



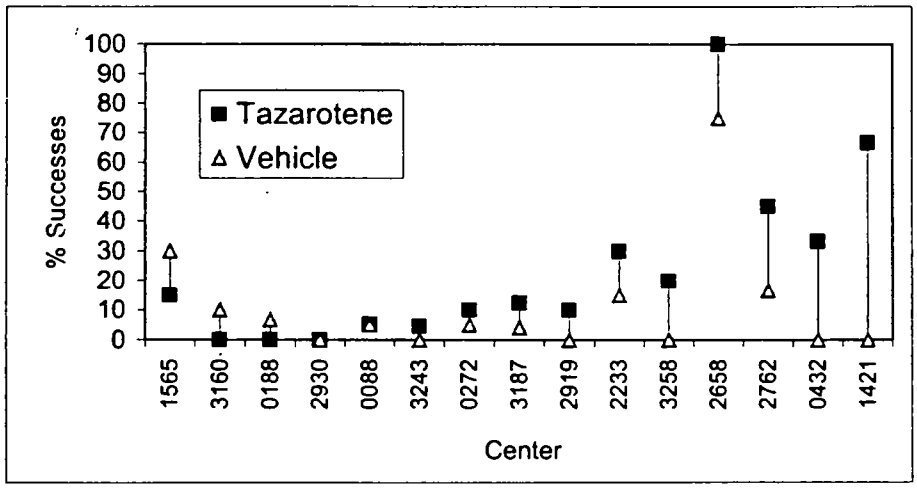
Source: Reviewer Analysis
Centers Sorted by treatment difference
Breslow-Day p-value = 0.581

Figure B.4 – Clinical Improvement for Mottled Hyperpigmentation by Center in Study 034C



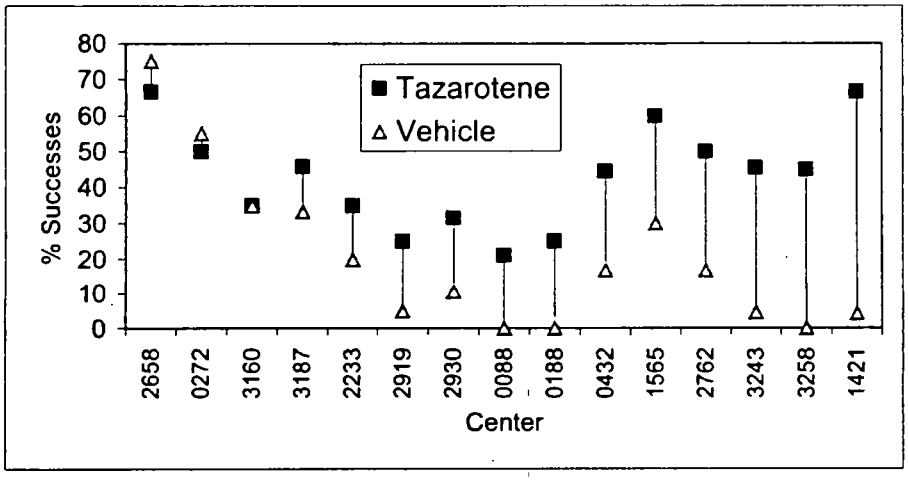
Source for Data: Table 2.1, file 034.pdf, pg 499 (Reviewer graphic)
Centers Sorted by treatment difference
Breslow-Day p-value < 0.001

Figure B.5 – Percent Success for Fine Wrinkling (Grade 0 or 1) by Center in Study 034C



Source: Reviewer Analysis
Centers Sorted by treatment difference
Breslow-Day p-value = <0.001

Figure B.6 – Percent Success for Mottled Hyperpigmentation (Grade 0 or 1) by Center in Study 034C



Source: Reviewer Analysis
Centers Sorted by treatment difference
Breslow-Day p-value = 0.008

2.7 List of References

Hochberg, Y. (1988) "A Sharper Bonferroni Procedure for Multiple Tests of Significance," *Biometrika*, 75:800-802.

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

/s/

Kathleen Fritsch
2/11/02 08:44:33 AM
BIOMETRICS

Mohamed Alosh
2/11/02 05:22:34 PM
BIOMETRICS
Concur with review