CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER: NDA 20-600

STATISTICAL REVIEW(S)

Statistical Review and Evaluation
(Addendum 2)

APR 15 1997

NDA/ Drug Class: 20-600 / 1S

Name of Drug: Tazarotene Gel, 0.05%, 0.1%,

Applicant: Allergan, Inc.
P.O. Box 19534
2525 Dupont Drive
Irvine, CA 92713-9534

Type of Report: Clinical / Statistical Addendum

Indication: Plaque Type Psoriasis and Acne Vulgaris

Documents Used: Volumes 1 & 2, dated 17 January 1997, response to the approvable action letter dated December 30, 1996.

Medical Officer: Dr. Hon S. Ko (HFD-540)

Introduction:

A study was provided by the sponsor to "evaluate the safety, efficacy, and acceptability of treatment with tazarotene (AGN 190168) 0.1% gel together with a placebo cream or a low-, mid-, or high-potency corticosteroid cream in the treatment of plaque psoriasis." This was a multicenter, investigator-masked, randomized parallel-group, four armed study, extending for 16 weeks (12 weeks treatment period, 4 weeks post-treatment). Treatments were either placebo, Synalar® 0.05%, Elocon® 0.1%, or Lidex® 0.05% cream applied once daily in the morning, in each case followed by tazarotene 0.1% gel applied once daily in the evening. Visits were to be made at week 0 (baseline), 2,4,8,12 (treatment period) and week 16 (post-treatment period).

The Medical Officer expressed some concerns relative to the analysis of several of the safety tables contained in this response. This report is an attempt to answer those concerns. In particular, tables 9.1.18.a-d were presented in the sponsor's report without associated test statistics.

We are interested in the general association of treatments with severity of response. In particular, for each subject, adverse events were categorized by severity (severe, moderate, or mild) or no adverse events. To retain consistency with the tables as presented by the sponsor, the last category, no adverse events, is not directly displayed in the tables below. However, for each treatment group the count in the no adverse event cell is N- the total count.

One way to approach appropriate tests for these tables would be a "generic"

approach, first testing for association of drug combination with severity of response. This corresponds to the "usual" chi-square test of no association, equivalent to the test of no interaction in a two-way loglinear model. This is analyzed using maximum likelihood technology. The second is a test of equal mean response scores, i.e.,

mean response = 0*Pr(no adverse event) + 1*Pr(mild) + 2*Pr(moderate) + 3*Pr(severe),

does not differ across treatment groups. This test uses the so-called Grizzle-Starmer-Koch weighted least squares approach to test differences across the treatment groups.

A more focused analysis is to directly address the so-called clinical hypotheses, provided by the sponsor, concerning adverse events, as listed in the synopsis of the sponsor's report:

i) H_0^1: Tazarotene 0.1% gel in combination with Synalar 0.01% cream, Elocon 0.1% cream, or Lidex 0.05% cream is associated with an incidence of treatment related adverse events which is 10 percentage points less than that of patients receiving tazarotene 0.1% gel in combination with a placebo cream.

ii) H_0^2: There is a trend towards lower incidence and severity of treatment related adverse events as the potency of the corticosteroid used in conjunction with tazarotene 0.1% gel increases.

Note that strictly speaking, these are inconsistent hypotheses. The first implies a constant adverse event rate over corticosteroids, the second a trend.

For the first set of hypotheses above the proportions can be tested directly. That is, if π_v and π_i are the proportion of subjects with appropriate adverse events in the vehicle group and comparator treatment, we can test:

$$\pi_v - \pi_i = 0.1,$$

by replacing the parameters by the observed proportions, squaring, and dividing by the estimated variance. Under the null hypothesis this term will be approximately chi-square with one degree of freedom.

For the observed proportion of adverse events the second hypotheses can be tested by specifying the observed proportion as the response function and using the weighted least squares technology. Further, with the mean response as defined above, we can test both sets of hypotheses using weighted least squares. In particular, a natural way to formulate H_0^1 in terms of mean response would seem to be as:

$$\mu_v - 1.1\,\mu_i = 0.0.$$

This tests that the vehicle mean is 10% more than the corticosteroid mean. For either response:

i) proportion of subjects experiencing any adverse event, labeled as the "proportion response" below,

or ii) the mean response (as defined above),

the test of trend across levels of putatively increasing potency can be implemented as testing:

response = $\alpha + \beta i$, for i = 0 (vehicle) to 3 (lidex).

Note that this will allow a 2 degree of freedom test for lack of fit.

The first set of two tables below (tables 9.1.18a & b) displays the subject's most severe adverse event during the treatment period, weeks 1-12, comparing the different treatment groups. The second set of two tables (tables 9.1.18c & d) displays the corresponding tables during the post treatment group. In each of these sets of two tables, the first table (tables 9.1.18a & c) displays those adverse events that are assessed as possibly treatment related, the second displays all adverse events.

Table 9.1.18a. Incidence by Severity - Treatment Related
Treatment Period
(Maximum Severity)

	N	Mild	Moderate	Severe	Total
taz/veh	75	11	20	7	38
		14.7%	26.7%	9.3%	50.7%
taz/synalar	78	9	27	3	39
		11.5%	34.6%	3.9%	50.0%
taz/elocon	74	16	11	2	29
		21.6%	14.9%	2.7%	39.2%
taz/lidex	73	11	15	5	31
		15.1%	20.6%	6.9%	42.5%

Generic Hypotheses:	No association (ML loglinear test)		$p \leq 0.116$ (9 df)
	Means equal (WLS test)		$p \leq 0.061$ (3 df)
Specific Hypotheses:	Proportions of Adverse Events		
	$\pi_v - \pi_i = 0.1$	veh vs synalar	$p \leq 0.248$
		veh vs elocon	$p \leq 0.855$
		veh vs lidex	$p \leq 0.826$
	Trend		$p \leq 0.167$
	Lack of Fit		$p \leq 0.598$
	Means of Adverse Events		
	$\mu_v - 1.1\,\mu_i = 0$	veh vs synalar	$p \leq 0.751$
		veh vs elocon	$p \leq 0.060$
		veh vs lidex	$p \leq 0.514$
	Trend		$p \leq 0.065$
	Lack of Fit		$p \leq 0.138$

For the generic hypotheses above in the table 9.1.18a above there is no strong, completely unequivocal evidence of an association between adverse events and treatment group ($p \leq 0.116$). However, the test of mean differences in severity is close to statistically significant ($p \leq 0.061$). Though not shown here, in fact most of this semi-apparent difference is due to the difference in profiles of tazarotene/elocon and tazarotene/lidex.

To address the specific hypotheses provided by the sponsor, if we consider these to be in terms of the proportion of adverse events, note that the first three tests provide no reason to reject the null hypothesis the adverse event rate for tazarotene with synalar, elocon, or lidex is

10% less than for tazarotene with vehicle ($p \leq 0.248$, $p \leq 0.855$, and $p \leq 0.826$ respectively). There is no strong evidence of a trend, that is, the test that the trend parameter is zero is accepted ($p \leq 0.167$).

Using the mean response formulations, results differ somewhat. In terms of this mean response, the first and last tests provide no reason to reject the null hypothesis that the adverse event rate for tazarotene with synalar or lidex is 10% less than for tazarotene with vehicle ($p \leq 0.751$ and $p \leq 0.514$ respectively). The comparison with elocon is close to statistical significance ($p \leq 0.060$), since the estimated adverse event mean of tazarotene with elocon is more than 10% less than the corresponding adverse mean of tazarotene with vehicle. Note that there is marginal evidence of a trend, that is, the test that the trend parameter is zero is nearly statistically significant ($p \leq 0.065$). This is due to a decreasing mean from vehicle through synalar to elocon, increasing to lidex. Given trend, note that there is no strong evidence for lack of fit ($p \leq 0.138$)

To summarize, with either response, in this experiment there is no statistically significant evidence to reject the claim that the adverse event rate with tazarotene and the corticosteroids is at least 10% less than with tazarotene and vehicle. For the proportion of adverse events there is no statistically significant evidence of a trend, however for the mean response there is some weak evidence of trend.

Table 9.1.18b, below, displays the same cross-tabulation, this time over all adverse events, not merely those putatively associated with treatment:

Table 9.1.18b. Incidence by Severity - All Adverse Events
Treatment Period
(Maximum Severity)

	N	Mild	Moderate	Severe	Total
taz/veh	75	13	29	8	50
		17.3%	38.7%	10.7%	66.7%
taz/synalar	78	12	30	5	47
		15.4%	38.5%	6.4%	60.3%
taz/elocon	74	19	23	3	45
		25.7%	31.1%	4.1%	60.8%
taz/lidex	73	12	22	9	43
		16.4%	30.1%	12.3%	58.9%

Generic Hypotheses:	No association (ML loglinear test)		$p \leq 0.476$ (9 df)
	Means differ (WLS test)		$p \leq 0.432$ (3 df)
Specific Hypotheses:	Proportions of Adverse Events		
	$\Pi_v - \Pi_i = 0.1$	veh vs synalar	$p \leq 0.644$
		veh vs elocon	$p \leq 0.598$
		veh vs lidex	$p \leq 0.778$
	Trend		$p \leq 0.357$
	Lack of Fit		$p \leq 0.856$
	Means of Adverse Events:		
	$\mu_v - 1.1\mu_i = 0$	veh vs synalar	$p \leq 0.819$
		veh vs elocon	$p \leq 0.323$
		veh vs lidex	$p \leq 0.931$
	Trend		$p \leq 0.307$
	Lack of Fit		$p \leq 0.426$

Again, considering the generic hypotheses first, in the table 9.1.18b above there is no particular evidence of any particular association between the occurrence and severity of adverse events and treatment group ($p \leq 0.476$) or in mean severity across treatment groups ($p \leq 0.432$).

To address the specific hypotheses provided by the sponsor, In terms of the adverse event rate, the first three tests provide no reason to reject the null hypothesis that the adverse event rate for tazarotene with synalar, elocon, or lidex is 10% less than for tazarotene with vehicle ($p \leq 0.644$, $p \leq 0.598$, and $p \leq 0.778$ respectively). There is no strong evidence of a trend, that is, the test that the trend parameter is zero is accepted ($p \leq 0.357$).

For these corresponding hypotheses in terms of response means, the first three tests provide no reason to reject the null hypothesis that the adverse event rate for tazarotene with synalar, elocon, or lidex is 10% less than for tazarotene with vehicle ($p \leq 0.819$, $p \leq 0.323$, and $p \leq 0.931$ respectively). Again, there is no strong evidence of a trend, that is, the test that the trend parameter is zero is accepted ($p \leq 0.307$).

To summarize, with either response, in this experiment there is no statistically significant evidence to reject the claim that the adverse event rate with tazarotene and the corticosteroids is at least 10% less than with tazarotene and vehicle. For either response there is no particular evidence of trend related to the strength of the corticosteroid.

Relatively few adverse events were recorded in the post treatment period. Tables of these appear below, tables 9.1.18c & d. For testing purposes, but not for display, in both tables the severe adverse event category was pooled with the moderate category. Table 9.1.18c displays putatively treatment related incidence during the post treatment period (weeks 13-16 inclusive).

Table 9.1.18c. Incidence by Severity - Treatment Related
Post Treatment Period
(Maximum Severity)

	N	Mild	Moderate	Severe	Total
taz/veh	52	4	3		7
		7.7%	5.8%		13.5%
taz/synalar	55	6	6		12
		10.9%	10.9%		21.8%
taz/elocon	65	5	0 (1E-20)		5
		7.7%			7.7%
taz/lidex	60	6	2		8
		10.0%	3.3%		13.3%

Generic Hypotheses:	No association (ML loglinear test)	$p \leq 0.170$ (6 df)
	Means differ (WLS test)	$p \leq 0.035$ (3 df)
Specific Hypotheses:	Proportions of Adverse Events	
	$\Pi_v - \Pi_i = 0.1$ veh vs synalar	$p \leq 0.012$
	veh vs elocon	$p \leq 0.464$
	veh vs lidex	$p \leq 0.126$
	Trend	$p \leq 0.463$
	Lack of Fit	$p \leq 0.109$
	Means of Adverse Events	
	$\mu_v - 1.1\ \mu_i = 0$ veh vs synalar	$p \leq 0.169$
	veh vs elocon	$p \leq 0.183$
	veh vs lidex	$p \leq 0.926$
	Trend	$p \leq 0.262$
	Lack of Fit	$p \leq 0.026$

So in the table 9.1.18c above there is no strong evidence of an association between adverse events and treatment group ($p \leq 0.170$). Empty cells, i.e., cells with zero frequency, as in the moderate/severe adverse count in the tazarotene/elocon treatment group present problems to the weighted least squares technology. One solution is to represent the zero cell count by a very small number, which is computationally almost zero. For passing to the weighted least squares program the cell frequency of the moderate/severe by tazarotene/elocon treatment is entered as 10^{-20} (i.e., a decimal with 1 in the 20th position preceded by 19 zeros) instead of zero. With this adjustment, however, the more specific test of mean differences is statistically significant, though barely. Also again, though not shown here, in fact most of this difference seems to be due to the difference in profiles of tazarotene /elocon and tazarotene/lidex.

To address the specific hypotheses provided by the sponsor, in terms of the proportion of adverse events, note that the second and third tests provide no strong reason to reject the null hypothesis the adverse event rate for tazarotene with elocon, or lidex is 10% less than for tazarotene with vehicle ($p \leq 0.464$ and $p \leq 0.126$ respectively). The adverse event rate for synalar with tazarotene is higher than the rate for vehicle and tazarotene, so for this experiment we would reject the hypothesis that the adverse event rate of synalar with tazarotene is 10% less than the rate with vehicle ($p \leq 0.012$). Again there is no strong evidence of a trend ($p \leq 0.463$).

Using the mean response formulations, there is no strong evidence to reject the null

hypothesis that the adverse event rate for tazarotene with synalar, elocon, or lidex is 10% less than for tazarotene with vehicle ($p \leq 0.169$, $p \leq 0.183$, and $p \leq 0.926$ respectively). There is no strong evidence of a trend ($p \leq 0.262$). However, the statistically significant lack of fit ($p \leq 0.026$) does suggest some pattern of differences between treatments.

Table 9.1.18d displays the frequencies over all adverse events during the post-treatment period.

Table 9.1.18d. Incidence by Severity - All Adverse Events
Post Treatment Period
(Maximum Severity)

	N	Mild	Moderate	Severe	Total
taz/veh	52	7	5	1	13
		13.5%	9.6%	1.9%	25.0%
taz/synalar	55	6	8	1	15
		10.9%	14.6%	1.8%	27.3%
taz/elocon	65	10	3		13
		15.4%	4.6%		20.0%
taz/lidex	60	8	6	1	15
		13.3%	8.3%	1.7%	23.3%

Generic Hypotheses:	No association (ML loglinear test)	$p \leq 0.578$ (6 df)
	Means differ (WLS test)	$p \leq 0.419$ (3 df)
Specific Hypotheses:	Proportions of Adverse Events	
	$\Pi_v - \Pi_i = 0.1$ veh vs synalar	$p \leq 0.148$
	veh vs elocon	$p \leq 0.521$
	veh vs lidex	$p \leq 0.305$
	Trend	$p \leq 0.621$
	Lack of Fit	$p \leq 0.701$
	Means of Adverse Events	
	$\mu_v - 1.1\ \mu_i = 0$ veh vs synalar	$p \leq 0.434$
	veh vs elocon	$p \leq 0.425$
	veh vs lidex	$p \leq 0.992$
	Trend	$p \leq 0.445$
	Lack of Fit	$p \leq 0.325$

So in the table above there is no particular evidence of any association between adverse events and treatment group ($p \leq 0.578$), or in mean differences in severity and occurrence across treatment groups ($p \leq 0.419$).

For the specific hypotheses above, in terms of either response function, the first three tests provide no reason to reject the null hypothesis that the adverse event rate for tazarotene with synalar, elocon, or lidex is 10% less than for tazarotene with vehicle ($p \leq 0.148$, $p \leq 0.521$, and $p \leq 0.305$ or $p \leq 0.434$, $p \leq 0.425$, and $p \leq 0.992$, respectively). Neither response shows a statistically significant evidence of a trend ($p \leq 0.621$ and $p \leq 0.445$, respectively).

In fact these were multicenter studies, a fact ignored in the analyses above. The original data was not supplied to this reviewer, so an analysis stratified on investigator was not feasible. Assuming that adverse event profiles will tend to be more similar within centers than between

centers, the impact of ignoring centers will generally be anti-conservative. That is, true p-values of test statistics will tend to be larger than their nominal levels computed ignoring the centers. However, there are at least 10 investigators, so assuming the patient counts are fairly uniform across investigators, the impact of this intracenter correlation may not be all that large. This at least suggests that the analysis above, ignoring center effects, should be adequate.

Conclusions

1. The sponsor provided one study to "evaluate the safety, efficacy, and acceptability of treatment with tazarotene 0.1% gel together with a placebo cream or a low-, mid-, or high-potency corticosteroid cream in the treatment of plaque psoriasis." This was a multicenter, investigator-masked, randomized parallel-group, four armed study, extending for 16 weeks (12 weeks treatment period, 4 weeks post-treatment). Treatments were either placebo, Synalar® 0.05%, Elocon® 0.1%, or Lidex® 0.05% cream applied once daily in the morning, in each case followed by tazarotene 0.1% gel applied once daily in the evening. Visits were to be made at week 0 (baseline), 2,4,8,12 (treatment period) and week 16 (post-treatment period). The Medical Officer expressed some concerns relative to the analysis of tables 9.1.18.a-d in the sponsor's report, in particular, they were given without associated test statistics.

2. Each subject was categorized by the severity of their most extreme adverse event, categorized as none, mild, moderate, or severe. Such ordinal data can be easily tested by tests of association between severity of adverse event and treatment, or mean severity across treatment. One of the specific hypotheses provided by the sponsor was that the adverse event rate for tazarotene with Synalar, Elocon, or Lidex is 10% less than for tazarotene with vehicle. Another hypothesis was that there was a trend of decreasing adverse events across levels of increasing strength of the corticosteroid.

3. For supposedly treatment related adverse events noted during the treatment period there was no strong evidence of an association between treatment and severity. However, the more specific test of differences across treatment group was close to statistical significance ($p \leq 0.061$). The data provided no reason to reject the null hypothesis that the adverse event rate for tazarotene with Synalar, Elocon, or Lidex is 10% less than for tazarotene with vehicle ($p \leq 0.248$, $p \leq 0.855$, and $p \leq 0.826$ respectively). However, the evidence for a trend was weak.

4. For all treatment related adverse events noted during the treatment period there was no strong evidence of an association between treatment and severity or of a difference between means. Again the data provided no reason to reject the null hypothesis that the adverse event rate for tazarotene with Synalar, Elocon, or Lidex is 10% less than for tazarotene with vehicle ($p \leq 0.644$, $p \leq 0.598$, and $p \leq 0.778$ respectively), with similar results for mean response. For either response there was no statistically significant evidence of a trend.

5. To summarize results during the post-treatment period, restricting to adverse events classified as treatment related, there was no strong evidence of an association between treatment and severity. However, the more specific test of differences across treatment group was statistically significant ($p \leq 0.035$). The data provided no reason to reject the null hypothesis that the adverse event rate for tazarotene with Elocon or Lidex is 10% less than for tazarotene with vehicle ($p \leq 0.464$ or $p \leq 0.126$). The rate for Synalar was sufficiently higher than the rate for tazarotene to reject this hypothesis ($p \leq 0.012$). Again

there was no strong evidence of a trend.

6. Generalizing to all adverse events during the post-treatment period, there was no statistically significant evidence of an association between treatment and severity, or of mean differences across treatment. Again, the data provided no reason to reject the null hypothesis that the adverse event rate for tazarotene with Synalar, Elocon, or Lidex is 10% less than for tazarotene with vehicle ($p \leq 0.148$, $p \leq 0.521$, or $p \leq 0.305$). Again, there was no strong evidence of a trend.

Steve Thomson 04/15/97

Steve Thomson
Mathematical Statistician, Biometrics IV

04/15/'97

concur: R. Srinivasan, Ph.D.
Team Leader, Biometrics IV

cc:
Archival NDA: 20-600
HFD-540/Division File
HFD-540/Dr. Wilkin
HFD-540/Dr. Ko
HFD-540/Mr. Cross
HFD-725/Dr. Harkins
HFD-725/Dr. Srinivasan
HFD-725/Mr. Thomson
HFD-340/Dr. LePay
This review has 10 pages.
Chron.

\Thomson\WP Text\(301)827-2078\April 11, 1997\c:\wpfiles\nd20600b.wp

NOV 1 5 1996

Statistical Review and Evaluation (Addendum)

NDA/ Drug Class: 20-600 / 1S

Name of Drug: Tazarotene Gel, 0.05%, 0.1%,

Applicant: Allergan, Inc.
P.O. Box 19534
2525 Dupont Drive
Irvine, CA 92713-9534

Type of Report: Clinical / Statistical Addendum

Indication: Plaque type Psoriasis and Acne Vulgaris

Documents Used: Volume 1.1 dated 27 June 1996, and volumes 9.1-9.3 dated 30 July 1996, and data sets provided on diskettes.

Medical Officer: Dr. Hon S. Ko (HFD-540)

Introduction:

The Medical Officer has expressed some concerns relative to the analysis of this NDA. This report is an attempt to answer those concerns. The primary efficacy studies involved in this NDA are summarized in the following tables:

Table 1a. Phase III Clinical Studies
Acne Vulgaris

Study no	design	objective	duration of study	No. enrolled*
R168-220-7997	multicenter, double blind, randomized, parallel-group (acne)	safety/efficacy vs vehicle od	12-week treatment	T.1% T.05% V 150 148 148
R168-221-8606		safety/efficacy vs vehicle od	12-week treatment	T.1% T.05% V 149 149 149

*T=Tazarotene, V=vehicle.

Table 1b. Phase III Clinical Studies
Stable Plaque Psoriasis

Study no	design	objective	duration of study	No. enrolled*
R168-T20-8606	multicenter, double blind, randomized, parallel-group (psoriasis)	safety/efficacy & duration of therapeutic effect vs vehicle od	12-week treatment, post-tr: 12 weeks	T.1% T.05% V 108 108 108
R168-121-8606		safety/efficacy vs vehicle od	12-week treatment	T.1% T.05% V 112 111 113
R168-125-8606	multicenter, investigator-masked randomized, parallel-group (psoriasis)	safety/efficacy & duration of therapeutic effect od vs Lidex cream .05% bid	12-week treatment, post-tr: 12 weeks	T.1% T.05% L 116 117 115
R168-126-8606		safety/efficacy & duration of therapeutic effect od vs Lidex cream .05% bid	12-week treatment, post-tr: 12 weeks	T.1% T.05% L 110 111 110
R168-145-8606		safety/efficacy & duration of therapeutic effect od vs Dovonex ointment .005% bid	12-week treatment, post-tr: 12 weeks	T.1% T.05% D 122 124 123

*T = Tazarotene, V = vehicle, L = Lidex cream and D = Dovonex ointment.

Concerns expressed by the Medical Officer

1. Concerns addressed to the Sponsor:

i. Present an appropriate analysis of efficacy data of women in the acne studies (Studies R168-220-7997 and R168-221-8606) who were using estrogen versus those that were not using estrogen.

These were presented in the sponsor's tables 1.1-1.8 (pages 2.355-2.363, i.e., volume 2, pages 355-363, of the 30 July submission), with separate tables for female estrogen users and nonusers.

The following table is taken from those tables, and presents the percent change from baseline at the 12th week of treatment for inflammatory, non-inflammatory, and total lesions, as well as the sponsor defined percent treatment success.

Table 1. Female Estrogen Users versus Nonusers: Comparison of Endpoints

	Estrogen Users			Not Estrogen Users		
	Tazarotene 0.1%	Tazarotene 0.05%	Vehicle	Tazarotene 0.1%	Tazarotene 0.05%	Vehicle
p-values:	Taz 0.1% vs .05%	Taz .05% vs veh		Taz 0.1% vs .05%	Taz .05% vs veh	
	--------- Taz .1% vs veh ---------			--------- Taz .1% vs veh ---------		
n	14	15	12	97	96	95
Total Inflammatory Lesions	-36.0	-50.1	-50.9	-50.0	-41.8	-34.6
P-values(a)	0.969	0.065		0.028	0.442	
	--------0.048----------			--------0.152----------		
Total Non-inflammatory Lesions:	-36.0	-44.4	-38.9	-52.4	-45.2	-37.5
P-values(a)	0.875	0.247		0.001	0.021	
	--------0.170----------			--------0.109----------		
Total Lesions:	-37.1	-47.1	-42.3	-52.1	-44.4	-32.7
P-values(a)	0.756	0.117		0.001	0.020	
	--------0.053----------			--------0.082----------		
% Treatment Success	50.0%	66.7%	58.3%	65.0%	51.0%	39.0%
P-values(b)	0.462	0.706		0.059	0.110	
	--------0.713----------			--------0.001----------		

(a) From an ANOVA with study and treatment as factors, percent change from baseline as specified response. P-values are from contrasts of specified effect.

(b) From a Fisher Exact test of specified effects

Note that the subgroup that did not use estrogen showed the expected trend for all four response variables. That is, for all four response variables, within the non-estrogen subgroup, numerically the tazarotene 0.1% treatment group is better than the tazarotene 0.05% group, which in turn is better than the vehicle group. In fact, six to eight of these differences are statistically significant (see table 1 above). By comparison for each of the response measures in the estrogen user subgroup the tazarotene 0.1% treatment group is numerically worse than the vehicle group. On the other hand, except for the inflammatory lesions, the tazarotene 0.05% group is numerically better than the vehicle. These differences were not statistically significant in the sponsor's analysis. However, as explained below, this could be easily modified to make most of these differences statistically significant.

Statistical note:

Again, from the sponsor's tables 1.1-1.8, (pages 2.355-2.363), within group variances of estrogen user and non user subgroups are roughly homogeneous (say always with a variance ratio of less than two or so). So assuming the variances are homogeneous, pooling the within estrogen user group variance with the within nonuser group variance would have resulted in a much more precise estimate of variance for the estrogen user subgroup, with many more degrees of freedom. Just from the magnitude of the differences in treatment means with the estrogen user subgroup, it is clear that with

such an estimate of variance many of the differences between treatment groups in the estrogen subgroup would in fact be found to be statistically significant.

However, no randomization was done on estrogen use, so any results from comparing estrogen users to nonusers are not really the results of designed trial, but rather the result of an observational study. This makes any interpretation with such a small number of estrogen users much more problematical. Besides, though again not displayed here, estrogen use is highly unbalanced across investigators. So to some extent, the effect of estrogen use is confounded with investigator differences. So all these observations do tend to cast doubt on any of the potentially statistically significant differences within the estrogen user subgroup. So this reviewer would suggest that any apparent difference in treatment response between estrogen users and nonusers is at best mildly suggestive, and certainly not conclusive.

ii. **Present "separate meta-analyses of subsets for both safety and efficacy, by combining the data of the vehicle-controlled pivotal trials for each indication so that there may be adequate power to detect significant differences between some of the demographic subsets. Please present data on global 'treatment success' for demographic subsets in acne."**

Strictly speaking, the usual definition of a "meta-analysis" is an analysis of the pooled outcomes of the separate experiments essentially treating each experiment as an outcome. To be of use such studies require a number of near replicates. Here, even among the psoriasis studies there are a maximum of only two or five replicates, depending upon whether one uses a study or treatment subgroup as a replicate. This is probably too small a number of replicates to be of use for true meta-analytic techniques.

Instead of giving a more usual meta-analysis the sponsor has interpreted this as a request for various analyses with data pooled across studies to increase sample size. Personally this reviewer would have preferred to pool the data and incorporate the demographic effects as explicit factors or covariates, and then study the interaction or association of these demographic factors or covariates with treatment. However, the sponsor's approach may be adequate.

These pooled data analyses in the vehicle controlled pivotal trials are presented for demographic variables in sponsor's tables 2.1 through 2.81 (plaque elevation, scaling, erythema, and treatment success rate), pages 2.364-2.44 of the 30 July submission, for the psoriasis studies, and in sponsor's tables 5.1 through 5.28 (total inflammatory, total non-inflammatory, total lesion counts, and treatment success rates), pages 3.005-3.042, of the 30 July submission, for the acne studies.

In interpreting those tables (see sponsor's tables 2.1 through 2.54, pages 2.364-2.417) note that the sponsor has included within group p-values at each time point. These provide tests of the statistical significance of change from baseline. However, subjects are selected on the basis of severity of symptoms. Since only subjects who show a certain level of severity are chosen at baseline, just due to the normal process of change we would

expect that over time more patients would improve than would worsen. These so called "regression effects," plus secular trends in the population, rendered it problematical to interpret any within group change from baseline as a valid measure of treatment effect. Hence this reviewer would recommend that they be ignored.

Of more use are the tests between treatment effect also displayed on the tables above. Note that in general, provided the subgroup has a sufficient number of cases, the two tazarotene subgroups are statistically significantly better than vehicle in all response measures. Again, in subgroups with a sufficient number of cases, the score in the high dose group (tazarotene 0.1%) is usually better than the score in the low dose group (tazarotene 0.05%), though the difference is seldom statistically significant. Such results are consistent with the pooled analyses.

iii. Give the "statistical significance" of the "adverse event data in each study, including among group comparisons of each adverse event with an incidence of >1% and termination of the study due to adverse events."

The tables of adverse event data appear as sponsor's tables 1a to 5b in the 27 June submission (pages 1.045-1.082, i.e. vol 1, page 045-082). The sponsor analyzed these using a Fisher Exact test. From the sponsor's table 3b (pages 1.055-1.064) it appears that in the psoriasis studies that pruritus, burning skin, erythema, and skin irritation were all statistically significantly worse for both tazarotene treatment levels than for tazarotene vehicle, Lidex cream 0.05%, or for Dovonex 0.005% (For all comparisons $p<0.001$). Similarly skin pain, desquamation, rash, contact dermatitis, and "psoriasis worsened" were usually worse for both levels of tazarotene than for the tazarotene vehicle or the alternative active treatments. However, p-values were not always as small, though usually statistically significant (i.e., $p\leq0.05$). Also, for each of these except "psoriasis worsened," the adverse event occurred statistically significantly more often in the tazarotene 0.1% group than in the tazarotene 0.05% group.

From the sponsor's table 4b (pages 1.068-1.070) it appears that in the acne studies that desquamation, burning skin, dry skin, and erythema were all statistically significantly worse for both tazarotene treatment levels than for tazarotene vehicle, Lidex cream 0.05%, or for Dovonex 0.005% (For all comparisons $p<0.001$). Similarly stinging skin was statistically significantly worse for both levels of tazarotene than for tazarotene vehicle ($p\leq0.02$). Pruritus, skin irritation, and skin pain were all worse for the high dose tazarotene group than for the vehicle group ($p\leq0.05$, $p\leq0.01$, $p\leq0.06$ respectively). By comparison, differences between the tazarotene 0.05% group and the vehicle group were not statistically significant for these three variables.

To summarize, there is statistically significant evidence of dose related trends in various measures of skin irritation. For both acne and psoriasis studies, other adverse events showed no general pattern of statistically significant differences across treatment groups. From sponsor's table 16a (page 5.087), based on the Fisher Exact test, statistically significantly more subjects dropped out from the pooled tazarotene 0.1% group

and the pooled tazarotene 0.05% group than from the pooled vehicle group ($p \leq 0.02$ and $p \leq 0.05$ respectively). Results are more extreme for the active controlled studies.

Statistical Note:

Note that the sponsor based analysis on Fisher Exact tests, comparing response across treatment groups. One problem with the Fisher Exact test is that it theoretically it assumes all subjects are equally likely to have appeared at any of the centers. When centers are geographically diverse, this is not likely to be true, and centers form a restriction on the randomization. Similar tests that adjust for centers are the Mantel-Haenszel (MH), or more generally the Cochrane-Mantel-Haenszel (CMH) tests. Although generally this reviewer would express a mild preference for the MH or CMH test to the usual Fisher Exact test, there are at least two reasons to use the Fisher Exact test:

1) When a frequency marginal of a response is zero within a center that center does not contribute to the MH (or CMH) statistic. That is, for rare events, as would be typical of most adverse events, any center that did not have the event in at least one of the treatment groups would be dropped from the computation of the test statistic. So the data for that center are dropped from this analysis. Contrariwise, all cases are retained with the Fisher Exact test.

2)Further, we would expect that within a treatment center, subjects would tend to be more alike than they are across the other centers. Under such a structure the "usual estimates" of variation underestimate population variation, and test statistics reject the null hypothesis somewhat too often. However, from a regulatory viewpoint, for adverse events this is acceptable, and, in fact, if not too extreme, is probably preferable to a conservative test.

iv. Present among-group "comparisons for efficacy data in R168-145-8606, giving significance levels between <u>two</u> treatment groups at a time."

Results are presented in the sponsor's tables 7.1 through 7.15 (pages 3.404 through 3.438). This study was a comparison of the two levels of tazarotene with Calcipotriol. By week 12 of the treatment period calcipotriol was better than either level of tazarotene on all response measures: plaque elevation, scaling, erythema, sum of scores (all averaged over target lesions), as well as the overall psoriasis evaluation, and the global response to treatment. For these responses, by treatment week 12, the difference between tazarotene 0.05% and calcipotriol is statistically significant for all measures except erythema. By week 12, only scaling, sum of scores, and the overall evaluation of psoriasis have statistically significant differences between tazarotene 0.01% and calcipotriol (but again calcipotriol is uniformly numerically superior to tazarotene 0.1% in this study). Results during the post-treatment period are numerically roughly similar, except that differences between tazarotene treatment groups and calcipotriol are seldom statistically significantly different and at week 12 the change in the overall evaluation of psoriasis is numerically better in the tazarotene 0.1% group than in the calcipotriol groups.

v. Show the comparison "between treatment groups for achievement of the global evaluation scores of [1] >75% improvement and [2] 100% improvement in each study."

These tables appear as sponsor tables 24a through 31b (pages 5.098-5.134) of the 21 June submission. Note that from the sponsor's table 24c, at the 12th week of treatment, in the R168-120-8606 study 38% of the patients had >75% inprovement in tazarotene 0.1% group, 28% in the tazarotene 0.05% versus 12% in the vehicle group. These differences between levels of tazarotene and vehicle were statistically significant ($p<0.001$ and $p\leq0.006$ respectively). in the R168-121-8606 study 25% of the patients had >75% inprovement in tazarotene 0.1% group, 18% in the tazarotene 0.05% versus 10% in the vehicle group. These less impressive differences between tazarotene and vehicle were still statistically significant or close to it ($p<0.012$ and $p\leq0.065$ respectively). For both studies, due to the few cases with 100% improvement, no differences were apparent when one tabulates only those cases. In the R168-125-8606 study, from sponsor's table 26c, at the 12th week the Mantel-Haenzsel tests of >75% improvement shows statistically significant differences between the 31% at that level in tazarotene 0.1% group, the 28% at that level in the 0.05% group, and the 42% in the Lidex group ($p<0.07$ and $p\leq0.04$ respectively). However the differences favor Lidex over either level of tazarotene. In the R168-126-8606 study, from sponsor's table 27c, at the 12th week the Mantel-Haenzsel tests of >75% improvement shows some differences between the 26% at that level in tazarotene 0.1% group, the 14% at that level in the 0.05% group, and the 29% in the Lidex group ($p<0.59$ and $p\leq0.02$ respectively). Again the differences favor Lidex over either level of tazarotene. Results are basically similar for both studies with Lidex when using the cases with 100% improvement as the response. In the R168-145-8606 study, from sponsor's table 28c, at the 12th week the Mantel-Haenzsel test of >75% improvement shows statistically significant differences among the 26% at that level in tazarotene 0.1% group, the 26% at that level in the 0.05% group, and the 47% in the Calcipotriol group ($p\leq0.02$ for both comparisons). Again, because the much smaller number of positive responses, results are less clear when one tabulates the 100% responses (sponsor's table 28d).

In the R168-220-8606 acne vulgaris study, from sponsor's table 30a, at the 12th week the Mantel-Haenzsel test of >75% improvement shows some differences among the 38% at that level in tazarotene 0.1% group, the 26% at that level in the 0.05% group, and the 20% in the vehicle group ($p<0.001$ and $p\leq0.0232$ respectively). Contrariwise, in the R168-221-8606 study, from the sponsor's table 31a, at the 12th week the Mantel-Haenzsel tests of >75% improvement show no statistically significant differences among the 18% at that level in tazarotene 0.1% group, the 11% at that level in the 0.05% group, and the 10% in the vehicle group ($p<0.090$ and $p\leq0.83$ respectively). In both studies only one patient had 100% clearing, so that 100% clearing is too stringent a criteria to be a valid measure of efficacy in evaluating the use of tazarotene in the acne studies.

Overall, relatively few patients showed 100% improvement for either the psoriasis or acne studies. Thus for this data this measure is not sensitive to differences in efficacy. On the other hand, >75% improvement does seem to have some sensitivity in the psoriasis studies, but less in the acne studies. Still, overall, in the psoriasis studies at the 12th week both levels of tazarotene seem to be superior to vehicle (For the R168-120 study: $p<0.001$ for tazarotene 0.1% and $p\leq0.006$ for tazarotene 0.05% versus vehicle. For the R168-121 study: $p<0.012$ for tazarotene 0.1% and $p\leq0.065$ for tazarotene 0.05% versus vehicle.)

vi. Give the "statistical significance of the dropouts among treatment groups in each study and significance of the differences in drug exposure among these groups."

A comparison among treatment groups of the rate of dropouts due to any cause is given in sponsor's provided tables 8.1 through 8.6 for the psoriasis studies and sponsor's tables 9.1 and 9.2 for the acne studies (pages 3.439-3.450 of the 30 July submission). Fishers exact test was used for the comparison of dropouts. As noted earlier this may be anticonservative, but from a regulatory point of view is probably adequate. During the treatment period, in the R168-120-8606 study roughly 25% of the patients dropped out of each treatment group (all comparisons have $p>0.999$). Dropouts were higher in the post treatment period, with a maximum of 72% in the vehicle group, versus 57% and 54% in the tazarotene 0.1% group and 0.,05% group, respectively. In the R168-121-8606 study 38% dropped out of the tazarotene 0.1% group during treatment, versus 22% in the other two groups ($p\leq0.01$ for comparing the 0.1% group to vehicle or the 0.05% group). In the treatment period of the R168-125-8606 study 28% dropped out of the tazarotene 0.1% group, 36% out of the 0.05% group, and 45% out of the Lidex group. The difference in dropouts between Lidex and the 0.1% group was statistically significant ($p\leq0.022$). In contradiction, in the treatment period of the R168-126-8606 study 33% dropped out of the tazarotene 0.1% group, 33% out of the 0.05% group, and 14% out of the Lidex group. The difference in dropouts between Lidex and either dose level of tazarotene was statistically significant ($p\leq0.001$). In the treatment period of the R168-145-8606 study 43% dropped out of the tazarotene 0.1% group, 48% out of the 0.05% group, and 25% out of the Calcipotriol group. The difference between Calcipotriol and the levels of tazarotene is statistically significant ($p\leq0.004$).

In the acne study, R168-220-8606, 26% dropped out of the tazarotene 0.1% group, 30% out of the tazarotene 0.05% group, and 20% out of the vehicle group. In the R168-221-8606 study, 20% dropped out of the tazarotene 0.1% group, 21% out of the tazarotene 0.05% group, and 23% out of the vehicle group. None of these dropout rates showed any statistically significant differences across treatments.

A comparison among treatment groups of the drug exposure in patients who completed the study and those who dropped out appears in sponsor provided tables 10.1 through 10.10 for the psoriasis studies and 11.1 through 11.4 for the acne studies. These were analyzed using one-way ANOVA. However, this reviewer does not see much use in these comparisons that is not already reflected in the analysis of drop outs. Hence these tables will be considered only supportive of the dropout analysis and will not be summarized.

vii. Discuss the presence "of any bias, and the significance of such bias, if present, in the disproportionate sample sizes among treatment groups in the post-treatment periods of the psoriasis studies."

This reviewer agrees with sponsors' response (page 1.232 of the July 30 submission) which concludes: "In summary, both the study design and comparability analysis of demographic and medical history characteristics indicate that there was no [particular] bias favoring one treatment group over another during the post-treatment period of the psoriasis studies."

Concerns addressed to FDA Statisticians (from the Medical Officer):

i. In some of the studies in this NDA (e.g. R168-121, -126, and -145) centers were combined for the analysis. Address the validity of the combinations:

In general pooling center means will add to apparent intracenter variation, while pooling the variances (computed from the pooled mean) will decrease apparent intracenter variation. Almost always, the effect of the mean terms will dominate the effect of the variances, and the overall effect of pooling should be conservative. Thus, pooling sparse centers should have little effect on conclusions.

ii. Can we compare statistical significance across studies? For instance, in some studies tazarotene treatment was associated with better "treatment success" (e.g., R169-125-8606) than in others.

Apparently the question is whether we can claim that the level of treatment success, as in the noted study, is particularly "unusual" or not. That is, in general, when can one or more p-values be considered discrepant? For continuous data, under a simple null hypothesis, e.g., as in testing equality of means, the p-value of test statistic follows a uniform(0,1) distribution. This is also approximately true for discrete data. The problem here is that these tests need to be derived under the alternative hypothesis. Even for unbiased tests, if the null hypothesis is not true, the only fact we know about the alternative distribution is that the probability of the rejection region is greater than the alpha-level. The exact distribution of the p-values under the alternative depends upon the true and unknown alternative. Thus, the description of any feature of the alternative distribution requires large sample approximations. However, five observations or four (if one deletes the potentially discrepant test) are not enough observations to describe the alternative. In this case, the problems are compounded by the fact that the discrepant observation is chosen as the smallest. For most sets of random p-values, the smallest will seem to some extent to be discrepant. For a sufficient number of tests one can take these as multivariate binary and investigate if the p-value associated with one test is discrepant.

However, such tests require a fairly large sample size, i.e., a fairly large number of tests. In this case we have one test per study, so for each response variable, for the psoriasis study, there are only five observations in the study. Hence there is no power for detecting if one of them is discrepant. So it is this reviewer's opinion that there is no appropriate way to see if the specified test is discrepant.

iii. In the between treatment group comparisons in the post treatment period of the psoriasis studies (studies R168-120, -125, -126, and -145), the applicant used day-0 values as a baseline for computing clinical score reductions but used the week 12, end of treatment values, as the baseline for "treatment success". Further, what is a fair way to compare "duration of effect?"

Note that a test of "no treatment effect," as specified in the original protocol, on the change scores of say the week 24 response from the week 12 response will test a hypothesis of the sort:

$$\mu_{1,24}-\mu_{1,12}=\mu_{2,24}-\mu_{2,12}$$

where $\mu_{i,24}$ represents the 24th day post treatment mean of the i-th treatment group

and $\mu_{i,12}$ represents the 12th day end-of-treatment mean of the same i-th treatment group. With these effects, putative treatment differences as indicated by a significant test statistic could be due to differences at the week 24 endpoint, or at the week 12 endpoint, or both. Alternatively, differences at the 24-week endpoint could be canceled out by differences at the week 12 end of treatment, presumably leading to a nonsignificant test statistic. So there is no readily interpretable relation between the outcome of the test and the final treatment success. This does not hold for the similar effect on change from baseline, since by randomization theory we would expect that the baseline measures have the same expected mean across treatment groups. The only reason for including the baseline measurement is to take advantage of the possible increase in precision due adjusting each score by its corresponding baseline value. Overall, the change from baseline is interpreted as treatment effect. No such simple interpretation is available for the scores denoting change from week 12 to week 24.

Note that the sponsor modified the protocol in the psoriasis studies to provide analyses on the change from baseline, as well as the originally specified change from the end of treatment. The sponsor apparently recognized the weakness in the protocol specified analysis, as noted in the final report for study R168-120:

"The second analysis, presented in this final report, was based on a comparison with baseline (week zero of the treatment period). The rationale for the latter analysis, which seems more appropriate, was: 1) To show continued therapeutic effect or a return to pretreatment levels, it is necessary to show a relationship to the original therapeutic levels;

2) If the comparison is based on data collected at the end of treatment, the groups are not comparable because the active medications may significantly reduce the severity levels in comparison to vehicle."

One reasonable way to assess duration of treatment effect would be to compute the life-table displaying duration to various time points. For example, one could compute the time to from treatment success (if any) to the first failure. This is defined as a global evaluation of a good to excellent response. If a subject, who was a treatment success, completed the post-treatment period without failing, or dropped out before failing, he is a censored observation. Subjects who were not treatment success were dropped, so all analyses are conditional upon being in the subset of treatment successes. Besides the implicit conditional probability definition implied by restricting to successful subjects, there are some technical problems with this approach. For example, the exact times of treatment success or subsequent failure are not observed, only the time of the subsequent visit. Still, as a rough approximation this seems to this reviewer to be a reasonable approach. There were some other technical points in the actual implementation that are debatable, but again this is at least a reasonable first approximation.

Table 2. Global Successes and Duration of Global Success

	Not Treatment Success	Total Treatment Success	Failed	Censored	80th Percentile	60th Percentile	50th Percentile	40th Percentile	20th Percentile
	Counts				Days				
R168-125-8606									
Tazarotene 0.1%	69	46	26	20	26	56	74	97	NA
Tazarotene 0.05%	78	38	26	12	30	56	76	91	136
Lidex Cream 0.05%	58	56	39	17	41	63	87	106	155
p-value	0.027								
R168-126-8606									
Tazarotene 0.1%	80	28	20	8	21	29	41	41	80
Tazarotene 0.05%	87	21	11	10	28	29	57	71	NA
Lidex Cream 0.05%	64	46	35	11	18	43	59	78	111
p-value	0.001								
R168-145-8606									
Tazarotene 0.1%	77	46	33	13	28	50	54	61	94
Tazarotene 0.05%	88	34	20	14	36	53	53	86	114
Dovonex Ointment 0.005%	69	55	34	21	36	71	84	92	155
p-value	0.0035								

For each experiment above the table displays the in the second column the number

of treatment successes as definbed by the sponsor. The first column displays the number of subjects that were not successes. The p-values listed below provide CMH tests of the homogeneity of treatment mean scores over the investigators. Note that in all cases the homogeneity score is statistically significant, and most of that significance comes from the higher success rates of Lidex cream or Dovonex ointment. The failed column lists the count of those successes who eventually changed to failures, while the censored column lists those who never failed (or dropped out before failure). The last five columns reflect the percentiles, in days, of the survival distribution (i.e., ξ_p such that min $Pr(S(t)\geq\xi_p)\geq p$).

Table 3 below displays the p-values of the test of homogeneity of the product-limit survival estimates above:

Table 3. P-values of Tests of Homogeneity of Duration of Efficacy

	Log-Rank	Wilcoxon
R168-125-8606	0.6607	0.5529
R168-126-8606	0.7451	0.5860
R168-145-8606	0.0344	0.0235

From table 2 above and other more extensive plots not displayed here, descriptively, the product limit duration curve of the tazarotene 0.05% group slightly dominates the curve for the Lidex 0.05% in the R168-125-8606 study. In the R168-126-8606 these curves are closer to coincidence. In both studies the Lidex 0.05% curve and the tazarotene 0.05% curve dominate the tazarotene 0.1% curves. However, differences between curves are not statistically significant. So overall, using this definition of duration, there are no statistically significant differences among the two levels of tazarotene and Lidex cream. However in the R168-145-8606 study, the product limit curve estimate of duration of the Dovonex 0.005% ointment dominated the two levels of tazarotene. In this data, there were in fact statistically significant differences among the three product limit curves ($p\leq0.0344$ and $p\leq0.0235$). By inspection of the table 2 and the associated curves, it does appear that Dovonex is separated from the levels of tazarotene, and thus is responsible for most of this statistical significance.

The key observation relative to the sponsor's suggestion of greater duration in efficacy of tazarotene is that in none of the studies is there any evidence that the duration of the effect of tazarotene dominates the effects of the other medications. If anything, the data incline to suggest that both Lidex and Dovonex dominate Tazarotene.

iv. For the age studies, the applicant used age 45 as a cut off, and no patients were older than 45. So the subgroup analysis was of limited use.

These tables are provided in tables 4 and 5 at the end of this report, for the age subgroups ages up to 18, 19-25, and aged 26 or over. Note that there is no particular evidence of a differential effect of age.

Conclusions

1. A total of seven randomized, multicenter studies were provided to support the claim of efficacy of Tazarotene Gel, at concentrations of 0.05% and 0.1% for the treatment of moderate stable plaque type psoriasis and general acne vulgaris. Of the five studies for stable plaque type psoriasis, two were double-blind vehicle-controlled and three were investigator masked active controlled studies. The two studies provided to support the claim of efficacy in the treatment of acne vulgaris were both double-blinded and vehicle-controlled. This supplement was to address a number of questions raised by the Medical Officer.

2. The first question concerned the differences in efficacy in the acne studies among female estrogen users versus nonusers. Although numerically the tazarotene 0.05% group was generally best, followed by the vehicle, and finally followed by the tazarotene 0.1%. However if one restricts attention to the estrogen users, these differences are not statistically significant. Although using a variance estimate from the non estrogen users would make some of these differences statistically significant, this would remain an essentially observational result based on a small number of cases. Hence, although there may be some evidence of a differential effect of treatment among estrogen users versus nonusers, results are not conclusive.

3. Generally speaking the results from combining the various subgroups across studies are consistent with the within study results.

4. Further, in general, in the psoriasis trials pruritus, burning skin, erythema, skin irritation, skin pain, desquamation, rash, contact dermatitis were significantly worse for tazarotene 0.1% and 0.05% treatment groups versus the vehicle. To summarize, there is statistically significant evidence of a dose related trend in various measures of skin irritation. For both acne and psoriasis studies, other adverse events showed no general pattern of statistically significant differences across treatment groups. Statistically significantly more subjects dropped out from the pooled tazarotene 0.1% group and the pooled tazarotene 0.05% group than from the pooled vehicle group ($p \leq 0.02$ and $p \leq 0.05$ respectively). Results are more extreme for the active controlled studies.

5. Results from the R168-145-8606 study are consistent with the results of the other psoriasis studies. Generally, calcipotriol 0.005% ointment (Dovonex) is superior to either dose of tazarotene for all endpoints.

6. Few subjects showed 100% success for either psoriasis or acne. However if one uses the much less restrictive response of "$>75\%$ improvement" there are statistically significant differences between the treatment groups and vehicle.

7. The dropout rate varied across studies, but does not seem to reflect any consistent pattern for either the acne or psoriasis studies.

8. "In summary, both the study design and comparability analysis of demographic and medical history characteristics indicate that there was no [particular] bias favoring one treatment group over another during the post-treatment period of the psoriasis studies."

9. In some cases the sponsor combined centers in the analysis. In general, combing such centers should be anti-conservative, but due to the few cases involved in these centers the impact should be very minor.

10. The medical officer inquired about the feasibility of comparing statistical significance across studies. In general this is difficult without a lot of assumed probability structure or else a large number of true replicates. These conditions do not hold, so this method of comparison is not really feasible.

11. In the original protocol tests of "no treatment effect" for several response variables in the post treatment period were specified as change scores from the end of treatment response. For example, the change scores of say the week 24 response from the week 12 response will test a hypothesis of the sort:

$$\mu_{1,24}-\mu_{1,12}=\mu_{2,24}-\mu_{2,12}$$

where $\mu_{i,24}$ represents the 24th day post treatment mean of the i-th treatment group

and $\mu_{i,12}$ represents the 12th day end-of-treatment mean of the same i-th treatment

group. With these effects, putative treatment differences as indicated by a significant test statistic could be due to differences at the week 24 endpoint, or at the week 12 endpoint, or both. Alternatively, differences at the 24-week endpoint could be canceled out by differences at the week 12 end of treatment, presumably leading to a nonsignificant test statistic. So there is no readily interpretable relation between the outcome of the test and the final treatment success. Note that the sponsor seems to have noted the problem with these scores and for several studies also provided the change scores from the start of treatment. To this reviewer these do seem to be interpretable.

12. Duration of efficacy was estimated using product limit estimates in the subsets of subjects who had a treatment success. The key observation relative to the sponsor's suggestion of greater duration in efficacy of tazarotene is that in none of the studies is there any evidence that the duration of the effect of tazarotene dominates the effects of the other medications. If anything, the data incline to suggest that Dovonex 0.005% ointment dominates both levels of tazarotene, with at best, Lidex roughly as effective as the lower dose of tazarotene, but still dominating both levels of tazarotene.

Steve Thomson 11/14/96

Steve Thomson
Mathematical Statistician, Biometrics IV

11/15/'96

concur: R. Srinivasan, Ph.D.
Acting Team Leader, Biometrics IV

cc:
Archival NDA: 20-600
HFD-540/Division File
HFD-540/Dr. Wilkin
HFD-540/Dr. Ko
HFD-540/Mr. Cross
HFD-725/Dr. Harkins
HFD-725/Dr. Srinivasan
HFD-725/Mr. Thomson
HFD-340/Dr. LePay
This review has 15 pages, and 2 tables (in multiple parts) in the appendix.
Chron.

\Thomson\WP Text\x 7-2078\November 14, 1996\c:\wpfiles\nd20600a.wp

Table 4a.: Acne Study R 168-220
Total Noninflammatory Lesions

				Age=<= 18 Years Week: 0	4	8	12
Drug	Tazarotene 0.1%	Diff from baseline (0)	n	67	65	62	65
			Mean	67.7	-6.0	-16.3	-16.4
			Std Dev	40.5	46.2	29.7	30.4
		% Diff	Mean	.	-5.5	-20.2	-21.8
			Std Dev	.	73.1	29.6	30.5
	Tazarotene 0.05%	Diff from baseline (0)	n	52	49	42	45
			Mean	64.3	-10.9	-19.7	-23.5
			Std Dev	37.1	21.9	26.6	27.9
		% Diff	Mean	.	-16.6	-30.5	-34.1
			Std Dev	.	37.6	28.8	33.3
	Vehicle	Diff from baseline (0)	n	67	61	60	61
			Mean	70.3	-20.5	-28.6	-36.9
			Std Dev	41.1	23.6	35.1	35.3
		% Diff	Mean	.	-27.7	-38.6	-46.7
			Std Dev	.	27.1	33.3	33.3
				Age=19 - 25 Years Week: 0	4	8	12
Drug	Tazarotene 0.1%	Diff from baseline (0)	n	39	39	34	31
			Mean	53.6	-19.2	-23.5	-30.3
			Std Dev	32.1	21.4	25.5	31.4
		% Diff	Mean	.	-32.3	-41.8	-50.1
			Std Dev	.	24.0	38.1	39.6
	Tazarotene 0.05%	Diff from baseline (0)	n	40	39	32	33
			Mean	56.5	-20.8	-25.2	-27.2
			Std Dev	33.8	33.5	30.9	32.1
		% Diff	Mean	.	-29.9	-42.2	-44.9
			Std Dev	.	30.0	33.9	37.4
	Vehicle	Diff from baseline (0)	n	33	31	25	27
			Mean	55.8	-24.6	-33.7	-38.4
			Std Dev	39.5	25.2	27.2	31.0
		% Diff	Mean	.	-37.2	-55.7	-65.1
			Std Dev	.	33.7	30.0	34.9

Table 4a (cont.).: Acne Study R 168-220
Total Noninflammatory Lesions

Age=> 25 Years							
			Week	0	4	8	12
Drug	Tazarotene 0.1%	Diff from baseline (0)	n	23	23	22	21
			Mean	47.1	-12.0	-22.5	-24.0
			Std Dev	36.1	23.2	26.0	29.5
		% Diff	Mean	.	-20.1	-44.8	-51.3
			Std Dev	.	38.7	25.5	35.8
	Tazarotene 0.05%	Diff from baseline (0)	n	32	30	23	22
			Mean	39.8	-12.0	-26.2	-28.9
			Std Dev	18.1	15.7	13.7	16.8
		% Diff	Mean	.	-33.5	-65.4	-68.6
			Std Dev	.	40.0	22.1	26.0
	Vehicle	Diff from baseline (0)	n	22	20	17	17
			Mean	45.4	-16.8	-29.5	-32.2
			Std Dev	29.3	18.0	21.9	22.8
		% Diff	Mean	.	-35.9	-59.2	-68.2
			Std Dev	.	23.9	20.6	26.4

Table 4b.: Acne Study R 168-220
Total Inflammatory Lesions

				Age=<= 18 Years Week: 0	4	8	12
Drug	Tazarotene 0.1%	Diff from baseline (0)	n	67	65	62	65
			Mean	25.2	-1.6	-5.7	-4.7
			Std Dev	14.4	9.7	11.3	12.9
		% Diff	Mean	.	-6.5	-24.4	-21.7
			Std Dev	.	37.1	44.9	45.0
	Tazarotene 0.05%	Diff from baseline (0)	n	52	49	42	45
			Mean	21.0	-2.5	-6.0	-6.0
			Std Dev	12.1	7.9	5.5	8.2
		% Diff	Mean	.	-12.9	-28.5	-26.7
			Std Dev	.	41.7	24.0	34.9
	Vehicle	Diff from baseline (0)	n	67	61	60	61
			Mean	21.4	-1.2	-5.4	-9.1
			Std Dev	12.4	8.6	11.7	11.6
		% Diff	Mean	.	-3.4	-18.0	-34.3
			Std Dev	.	48.9	45.3	41.9

				Age=19 - 25 Years Week: 0	4	8	12
Drug	Tazarotene 0.1%	Diff from baseline (0)	n	39	39	34	31
			Mean	23.6	-7.6	-9.2	-11.3
			Std Dev	11.8	10.5	9.3	10.4
		% Diff	Mean	.	-23.5	-44.3	-44.6
			Std Dev	.	49.5	32.0	37.4
	Tazarotene 0.05%	Diff from baseline (0)	n	40	39	32	33
			Mean	20.7	-3.5	-8.3	-8.5
			Std Dev	9.0	9.5	7.8	8.6
		% Diff	Mean	.	-18.0	-36.6	-37.9
			Std Dev	.	45.2	36.1	40.5
	Vehicle	Diff from baseline (0)	n	33	31	25	27
			Mean	20.1	-6.2	-7.4	-9.1
			Std Dev	8.2	6.6	7.7	9.1
		% Diff	Mean	.	-30.5	-35.4	-45.1
			Std Dev	.	29.7	33.7	40.9

Table 4b (cont.).: Acne Study R 168-220
Total Inflammatory Lesions

				Age=> 25 Years Week: 0	4	8	12
Drug	Tazarotene 0.1%	Diff from baseline (0)	n	23	23	22	21
			Mean	17.0	-5.7	-7.8	-7.1
			Std Dev	10.0	6.2	13.6	12.2
		% Diff	Mean	.	-34.2	-37.3	-31.8
			Std Dev	.	32.1	68.3	58.8
	Tazarotene 0.05%	Diff from baseline (0)	n	32	30	23	22
			Mean	16.6	-6.6	-9.7	-11.8
			Std Dev	7.6	5.0	6.1	6.6
		% Diff	Mean	.	-40.3	-56.6	-67.8
			Std Dev	.	24.0	22.3	23.7
	Vehicle	Diff from baseline (0)	n	22	20	17	17
			Mean	18.3	-5.9	-11.1	-13.4
			Std Dev	8.9	10.4	10.0	7.6
		% Diff	Mean	.	-32.8	-55.7	-67.8
			Std Dev	.	44.3	30.5	21.5

Table 4c.: Acne Study R 168-220
Total Lesions

Age=<= 18 Years

				Week: 0	4	8	12
Drug	Tazarotene 0.1%	Diff from baseline (0)	n	67	65	62	65
			Mean	92.9	-7.7	-22.0	-21.1
			Std Dev	47.0	51.3	32.8	35.2
		% Diff	Mean	.	-9.3	-22.6	-22.4
			Std Dev	.	44.5	25.9	28.8
	Tazarotene 0.05%	Diff from baseline (0)	n	52	49	42	45
			Mean	85.3	-13.4	-25.7	-29.5
			Std Dev	41.5	26.9	29.0	30.7
		% Diff	Mean	.	-16.0	-29.4	-32.5
			Std Dev	.	33.3	25.6	30.7
	Vehicle	Diff from baseline (0)	n	67	61	60	61
			Mean	91.8	-21.7	-34.0	-45.9
			Std Dev	47.2	26.8	42.9	42.5
		% Diff	Mean	.	-22.4	-34.0	-44.7
			Std Dev	.	24.6	31.6	31.8

Age=19 - 25 Years

				Week: 0	4	8	12
Drug	Tazarotene 0.1%	Diff from baseline (0)	n	39	39	34	31
			Mean	77.2	-26.8	-32.7	-41.6
			Std Dev	38.1	25.3	28.5	37.3
		% Diff	Mean	.	-31.9	-41.9	-48.1
			Std Dev	.	21.8	30.9	34.2
	Tazarotene 0.05%	Diff from baseline (0)	n	40	39	32	33
			Mean	77.2	-24.4	-33.5	-35.7
			Std Dev	37.1	35.4	32.7	35.8
		% Diff	Mean	.	-27.9	-42.4	-43.3
			Std Dev	.	27.6	29.7	34.5
	Vehicle	Diff from baseline (0)	n	33	31	25	27
			Mean	75.8	-30.8	-41.1	-47.5
			Std Dev	40.4	27.1	28.7	33.6
		% Diff	Mean	.	-35.5	-50.4	-59.2
			Std Dev	.	26.6	26.9	33.3

Table 4c (cont.): Acne Study R 168-220
Total Lesions

Age=> 25 Years

				Week:	0	4	8	12
Drug	Tazarotene 0.1%	Diff from baseline (0)	n		23	23	22	21
			Mean		64.1	-17.7	-30.3	-31.0
			Std Dev		39.0	24.7	34.0	37.7
		% Diff	Mean		.	-26.3	-42.6	-45.9
			Std Dev		.	22.4	31.1	38.1
	Tazarotene 0.05%	Diff from baseline (0)	n		32	30	23	22
			Mean		56.4	-18.6	-35.9	-40.7
			Std Dev		21.7	17.2	17.6	21.1
		% Diff	Mean		.	-35.3	-62.2	-67.8
			Std Dev		.	28.4	15.5	18.5
	Vehicle	Diff from baseline (0)	n		22	20	17	17
			Mean		63.7	-22.7	-40.5	-45.6
			Std Dev		34.3	22.4	27.7	26.7
		% Diff	Mean		.	-36.1	-59.0	-69.1
			Std Dev		.	24.8	16.7	18.7

Table 4d: Acne Study R 168-220
Overall Evaluation of Acne

Age=<= 18 Years

		Week 0		4		8		12	
		n	%	n	%	n	%	n	%
Drug	Overall Evaluation								
Tazarotene 0.1%	Mild	27	40.3	33	50.8	39	62.9	40	61.5
	Moderate	39	58.2	30	46.2	20	32.3	21	32.3
	Severe	1	1.5	2	3.1	3	4.8	4	6.2
Tazarotene 0.05%	None	1	1.9	.	.	.	.	2	4.4
	Mild	27	51.9	28	57.1	25	59.5	32	71.1
	Moderate	23	44.2	21	42.9	17	40.5	11	24.4
	Severe	1	1.9	.	.	.	.	.	.
Vehicle	Mild	36	53.7	33	54.1	47	78.3	45	73.8
	Moderate	30	44.8	26	42.6	11	18.3	16	26.2
	Severe	1	1.5	2	3.3	2	3.3	.	.

Age=19 - 25 Years

		Week 0		4		8		12	
		n	%	n	%	n	%	n	%
Drug	Overall Evaluation								
Tazarotene 0.1%	Mild	14	35.9	25	64.1	25	73.5	22	71.0
	Moderate	19	48.7	12	30.8	8	23.5	9	29.0
	Severe	6	15.4	2	5.1	1	2.9	.	.
Tazarotene 0.05%	Mild	16	40.0	25	64.1	23	71.9	25	75.8
	Moderate	23	57.5	13	33.3	8	25.0	7	21.2
	Severe	1	2.5	1	2.6	1	3.1	1	3.0
Vehicle	None	.	.	.	.	.	.	3	11.1
	Mild	11	33.3	20	64.5	17	68.0	19	70.4
	Moderate	21	63.6	11	35.5	8	32.0	5	18.5
	Severe	1	3.0	.	.	.	.	.	.

Table 4d (cont.): Acne Study R 168-220
Overall Evaluation of Acne

Age=> 25 Years

		Week 0		4		8		12	
		n	%	n	%	n	%	n	%
Drug	Overall Evaluation								
Tazarotene 0.1%	None	.	.	.	.	.	.	1	4.8
	Mild	16	69.6	18	78.3	20	90.9	19	90.5
	Moderate	7	30.4	5	21.7	2	9.1	1	4.8
Tazarotene 0.05%	None	.	.	.	.	.	.	1	4.5
	Mild	23	71.9	24	80.0	22	95.7	21	95.5
	Moderate	8	25.0	6	20.0	1	4.3	.	.
	Severe	1	3.1	.	.	.	.	.	.
Vehicle	Mild	14	63.6	16	80.0	16	94.1	17	100.0
	Moderate	8	36.4	4	20.0	1	5.9	.	.

Table 4e.: Acne Study R 168-220
Global Evaluation of Efficacy

Age=<= 18 Years

		Week 4		Week 8		Week 12	
Drug	Global Evaluation	n	%	n	%	n	%
Tazarotene 0.1%	Excellent Response	5	7.7	1	1.6	5	7.7
	Good Response	4	6.2	10	16.1	12	18.5
	Fair Response	11	16.9	23	37.1	16	24.6
	Poor Response	22	33.8	12	19.4	14	21.5
	Condition Unchanged	23	35.4	16	25.8	18	27.7
Tazarotene 0.05%	Excellent Response	1	2.0	2	4.8	6	13.3
	Good Response	8	16.3	9	21.4	12	26.7
	Fair Response	11	22.4	12	28.6	10	22.2
	Poor Response	15	30.6	11	26.2	11	24.4
	Condition Unchanged	14	28.6	8	19.0	6	13.3
Vehicle	Excellent Response	2	3.3	7	11.7	15	24.6
	Good Response	10	16.4	17	28.3	20	32.8
	Fair Response	21	34.4	15	25.0	7	41.5
	Poor Response	17	27.9	12	20.0	11	18.0
	Condition Unchanged	11	18.0	9	15.0	8	13.1

Age=19 - 25 Years

		Week 4		Week 8		Week 12	
Drug	Global Evaluation	n	%	n	%	n	%
Tazarotene 0.1%	Excellent Response	3	7.7	6	17.6	8	25.8
	Good Response	7	17.9	7	20.6	9	29.0
	Fair Response	12	30.8	13	38.2	7	22.6
	Poor Response	14	35.9	5	14.7	5	16.1
	Condition Unchanged	3	7.7	3	8.8	2	6.5
Tazarotene 0.05%	Excellent Response	1	2.6	4	12.5	7	21.2
	Good Response	6	15.4	8	25.0	7	21.2
	Fair Response	14	35.9	11	34.4	9	27.3
	Poor Response	12	30.8	8	25.0	8	24.2
	Condition Unchanged	6	15.4	1	3.1	2	6.1
Vehicle	Completely cleared	.	.	.	.	1	3.7
	Excellent Response	5	16.1	9	36.0	12	44.4
	Good Response	6	19.4	3	12.0	6	22.2
	Fair Response	11	35.5	7	28.0	4	14.8
	Poor Response	7	22.6	4	16.0	1	3.7
	Condition Unchanged	2	6.5	2	8.0	3	11.1

Table 4e (cont.): Acne Study R 168-220
Global Evaluation of Efficacy

Age=> 25 Years

		Week 4		Week 8		Week 12	
Drug	Global Evaluation	n	%	n	%	n	%
Tazarotene 0.1%	Excellent Response	1	4.3	5	22.7	10	47.6
	Good Response	8	34.8	9	40.9	3	14.3
	Fair Response	4	17.4	3	13.6	2	9.5
	Poor Response	5	21.7	4	18.2	4	19.0
	Condition Unchanged	5	21.7	1	4.5	2	9.5
Tazarotene 0.05%	Excellent Response	3	10.0	8	34.8	13	59.1
	Good Response	10	33.3	11	47.8	6	27.3
	Fair Response	9	30.0	4	17.4	3	13.6
	Poor Response	5	16.7	.	.	.	.
	Condition Unchanged	3	10.0	.	.	.	.
Vehicle	Excellent Response	2	10.0	8	47.1	12	70.6
	Good Response	6	30.0	5	29.4	5	29.4
	Fair Response	7	35.0	3	17.6	.	.
	Poor Response	2	10.0	1	5.9	.	.
	Condition Unchanged	3	15.0	.	.	.	.

Table 5a.: Acne Study R 168-221
Total Noninflammatory Lesions

Age=<= 18 Years

Drug				Week	0	4	8	12
	Tazarotene 0.1%	Diff from baseline (0)	n		69	65	64	63
			Mean		55.5	-11.8	-21.6	-23.3
			Std Dev		33.2	25.6	26.2	26.0
		% Diff	Mean		.	-16.2	-33.9	-39.4
			Std Dev		.	37.8	31.4	34.9
	Tazarotene 0.05%	Diff from baseline (0)	n		55	54	51	51
			Mean		64.5	-5.6	-18.3	-22.9
			Std Dev		37.6	27.6	25.8	22.9
		% Diff	Mean		.	-8.7	-27.6	-35.5
			Std Dev		.	33.7	29.8	27.6
	Vehicle	Diff from baseline (0)	n		63	62	55	50
			Mean		57.5	-0.7	-7.3	-12.0
			Std Dev		30.8	25.8	26.0	21.3
		% Diff	Mean		.	-3.7	-12.0	-25.5
			Std Dev		.	37.9	57.1	36.2

Age=19 - 25 Years

Drug				Week:	0	4	8	12
	Tazarotene 0.1%	Diff from baseline (0)	n		39	39	29	32
			Mean		51.0	-11.2	-19.9	-24.0
			Std Dev		23.1	15.5	17.5	19.1
		% Diff	Mean		.	-17.6	-37.8	-46.4
			Std Dev		.	28.8	35.8	33.1
	Tazarotene 0.05%	Diff from baseline (0)	n		31	31	25	26
			Mean		49.9	-11.2	-17.6	-23.8
			Std Dev		23.4	15.0	17.7	20.9
		% Diff	Mean		.	-20.7	-32.6	-41.9
			Std Dev		.	33.0	34.4	31.7
	Vehicle	Diff from baseline (0)	n		37	36	33	30
			Mean		44.2	-3.9	-11.2	-12.8
			Std Dev		15.5	13.3	16.1	16.2
		% Diff	Mean		.	-7.1	-24.6	-26.1
			Std Dev		.	30.6	34.5	32.2

Table 5a (cont.): Acne Study R 168-221
Total Noninflammatory Lesions

Age=> 25 Years

Drug				Week: 0	4	8	12
	Tazarotene 0.1%	Diff from baseline (0)	n	25	24	23	22
			Mean	44.2	-13.6	-15.1	-18.0
			Std Dev	28.5	14.3	16.2	13.7
		% Diff	Mean	.	-29.1	-39.0	-47.8
			Std Dev	.	25.0	28.5	27.7
	Tazarotene 0.05%	Diff from baseline (0)	n	43	41	41	38
			Mean	38.7	-7.8	-12.2	-13.7
			Std Dev	11.9	12.0	15.1	14.5
		% Diff	Mean	.	-20.2	-32.9	-38.1
			Std Dev	.	33.1	34.5	34.3
	Vehicle	Diff from baseline (0)	n	36	34	33	30
			Mean	41.3	-7.5	-10.9	-13.8
			Std Dev	18.7	15.7	13.5	15.2
		% Diff	Mean	.	-15.5	-25.0	-30.7
			Std Dev	.	39.5	33.3	37.9

Table 5b.: Acne Study R 168-221
Total Inflammatory Lesions

Age=<= 18 Years

Drug				Week: 0	4	8	12
	Tazarotene 0.1%	Diff from baseline (0)	n	69	65	64	63
			Mean	25.7	-3.5	-10.0	-11.7
			Std Dev	12.9	10.9	9.9	9.4
		% Diff	Mean	.	-14.1	-38.6	-45.9
			Std Dev	.	38.1	27.2	22.9
	Tazarotene 0.05%	Diff from baseline (0)	n	55	54	51	51
			Mean	26.5	-1.2	-5.2	-9.5
			Std Dev	15.3	12.0	16.0	13.3
		% Diff	Mean	.	-3.0	-20.5	-31.4
			Std Dev	.	43.5	54.7	43.0
	Vehicle	Diff from baseline (0)	n	63	62	55	50
			Mean	23.2	-2.6	-6.1	-4.9
			Std Dev	11.4	10.3	12.9	12.4
		% Diff	Mean	.	-7.3	-22.8	-19.6
			Std Dev	.	43.5	55.4	57.6

Age=19 - 25 Years

Drug				Week: 0	4	8	12
	Tazarotene 0.1%	Diff from baseline (0)	n	39	39	29	32
			Mean	19.4	-3.3	-8.0	-8.3
			Std Dev	11.5	7.4	8.2	8.0
		% Diff	Mean	.	-15.6	-40.7	-43.6
			Std Dev	.	35.7	32.5	37.9
	Tazarotene 0.05%	Diff from baseline (0)	n	31	31	25	26
			Mean	19.0	-3.8	-7.4	-8.8
			Std Dev	7.8	7.0	6.3	7.8
		% Diff	Mean	.	-19.5	-38.6	-42.5
			Std Dev	.	32.3	38.4	31.7
	Vehicle	Diff from baseline (0)	n	37	36	33	30
			Mean	22.2	-2.3	-5.0	-7.8
			Std Dev	10.6	6.7	8.0	8.8
		% Diff	Mean	.	-7.7	-15.6	-32.1
			Std Dev	.	40.1	43.5	46.1

Table 5b (cont.): Acne Study R 168-221
Total Inflammatory Lesions

Age=> 25 Years

				Week:	0	4	8	12
Drug	Tazarotene 0.1%	Diff from baseline (0)	n		25	24	23	22
			Mean		18.1	-4.2	-8.0	-9.3
			Std Dev		7.8	9.4	6.2	5.0
		% Diff	Mean		.	-26.3	-44.7	-53.8
			Std Dev		.	49.3	28.3	23.5
	Tazarotene 0.05%	Diff from baseline (0)	n		43	41	41	38
			Mean		16.2	-4.3	-6.9	-7.2
			Std Dev		5.3	6.7	6.4	8.4
		% Diff	Mean		.	-23.8	-39.9	-39.7
			Std Dev		.	40.8	39.4	53.1
	Vehicle	Diff from baseline (0)	n		36	34	33	30
			Mean		16.5	-2.9	-5.0	-5.1
			Std Dev		7.9	4.8	4.3	6.0
		% Diff	Mean		.	-23.9	-37.5	-38.0
			Std Dev		.	30.7	31.7	41.8

Table 5c.: Acne Study R 168-221
Total Lesions

Age=<= 18 Years

				Week:	0	4	8	12
Drug	Tazarotene 0.1%	Diff from baseline (0)	n		69	65	64	63
			Mean		81.3	-15.3	-31.5	-35.0
			Std Dev		41.3	29.4	30.8	31.0
		% Diff	Mean		.	-16.0	-35.7	-42.0
			Std Dev		.	31.1	24.8	25.4
	Tazarotene 0.05%	Diff from baseline (0)	n		55	54	51	51
			Mean		91.0	-6.8	-23.5	-32.5
			Std Dev		47.4	26.8	29.0	28.0
		% Diff	Mean		.	-8.2	-26.3	-35.4
			Std Dev		.	25.4	25.9	23.6
	Vehicle	Diff from baseline (0)	n		63	62	55	50
			Mean		80.7	-3.3	-13.4	-17.0
			Std Dev		36.8	27.8	36.0	29.8
		% Diff	Mean		.	-5.4	-14.7	-22.8
			Std Dev		.	32.0	53.8	37.5

Age=19 - 25 Years

				Week:	0	4	8	12
Drug	Tazarotene 0.1%	Diff from baseline (0)	n		39	39	29	32
			Mean		70.4	-14.4	-28.0	-32.3
			Std Dev		28.9	19.5	22.8	23.1
		% Diff	Mean		.	-17.9	-38.9	-46.1
			Std Dev		.	22.1	26.5	29.3
	Tazarotene 0.05%	Diff from baseline (0)	n		31	31	25	26
			Mean		68.9	-15.0	-25.0	-32.6
			Std Dev		26.9	20.1	20.4	23.2
		% Diff	Mean		.	-20.3	-35.6	-43.8
			Std Dev		.	29.2	26.3	24.1
	Vehicle	Diff from baseline (0)	n		37	36	33	30
			Mean		66.4	-6.2	-16.2	-20.6
			Std Dev		20.4	16.4	18.9	19.0
		% Diff	Mean		.	-8.3	-23.7	-29.5
			Std Dev		.	27.4	30.2	29.1