CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER: 21-145

STATISTICAL REVIEW(S)

Statistical Review and Evaluation

JUN 19 2000

NDA#	**21-145**
Name of Drug:	Eflornithine hydrochloride 15% topical cream
Applicant:	Westwood-Squibb Colton Holdings Partnership
Indication:	Excessive Facial Hair in Women
Documents Reviewed:	Vol. 1.1, 1.2, 1.51-1.85/submitted on Sept. 24, 1999, subsequent submissions on April 25 and May 8, 2000
Medical Reviewer:	Denise Cook, MD
Statistical Reviewer:	Qian Li, Sc.D.
Period of Review:	February-May 2000

I. Introduction:

In this NDA, the sponsor pursues marketing approval for the use of eflornithine hydrochloride 15% topical cream twice daily for controlling hair growth in women with excessive facial hair. For efficacy evaluation, two identically designed double-blinded, vehicle controlled, pivotal studies (DE140-001 and DE140-002) were conducted to determine the safety and efficacy of eflornithine 15% cream in the treatment of women with excessive facial hair.

This statistical review focuses on issues of the efficacy aspect in using eflornithine 15% topical cream for treating women with excessive facial hair. The two pivotal studies have presented convincing efficacy results in controlling hair growth. Detailed statistical issues were discussed in reviewer's comment section. There is no major statistical issue identified that would alter the study results and their interpretation.

II. Study Design and Statistical Methodology:

The two pivotal studies for efficacy evaluation were identically designed Phase III studies that were conducted within United States (only one study had one center from Europe). Adult women who removed facial hair at least twice per week and had an average hair density of at least five hairs per square centimeter on two facial evaluation areas (chin and upper lip) as determined by video image analysis were eligible for enrollment. Subjects were randomized to elfornithine treatment and vehicle in 2:1 ratio stratified by centers. The duration of treatment was 24 weeks followed by an eight-week no-treatment period. Subjects were scheduled to visit at pre-study visit (including day 2), Weeks 2, 4, 8, 12, 16, 20, 24 and 32.

1. Blinding:

Because reports concerning skin related adverse events (especially sting, burning, tingling, itching, etc, immediately after application of study medication) could provide

clues to the identity of study medication, an individual other than the physician responsible for completing global evaluations would query subjects about adverse events at visits. If non-serious skin related adverse events were reported, an individual other than the physician responsible for completing global evaluations collected information about the events and completed with appropriate case report forms.

2. Efficacy evaluation:

Physician's global assessment (PGA) and video image analysis were made 48 hours after facial hair removal by shaving at each visit except the visits at weeks 12 and 20. Subject's self assessment was evaluated at baseline, Weeks 8, 16, 24 and 32.

Primary efficacy endpoints:

Dichotomized physician's global assessment was designated as the primary endpoint that covered hair parameters such as length of hairs, density and darkening of skin. Four point scale was used in physician's global assessment: clear/almost clear, marked improvement, improvement, and no improvement. 'Success' was defined to include the subjects who were assessed as clear/almost clear and marked improvement, and 'failure' included improvement and no improvement. A difference of 20% (30% for eflornithine and 10% for vehicle) between the treatment groups was considered to be clinically significant.

Secondary efficacy variables:

Video image analyses for reduction in hair growth and spatial mass were defined as secondary response measures. Reduction in hair growth will be dichotomized into "success" (subjects with at least a 50% reduction in hair growth relative to baseline) and "failure" (less than 50% reduction). Spatial mass was measured as hair area per square centimeter of skin surface.

Subject's self-perception questionnaire that comprises six questions was another secondary endpoints, which evaluated the level of a subject's "bother" and "discomfort" with her excessive facial hair and its effect on quality of life. The six questions were

1) Bothered by facial hair?
2) Uncomfortable when meet new people?
3) Uncomfortable at work or class?
4) Uncomfortable at social gatherings?
5) Uncomfortable in exchanges of affection?
6) Bothered by time spent removing hair?

2. Analysis populations:

Two analysis populations (datasets) were defined in protocol. Intent-to-treat (ITT), or all subjects randomized (ASR), dataset consisted of all the subjects randomized into the

study. This data set should be used for the primary analysis. Another analysis population was evaluable dataset which consisted of all subjects who were without significant protocol violation.

3. Statistical Analyses:

The analysis of the primary endpoint was performed at Week 24, the end of treatment. The dichotomized physician's global assessment was analyzed using Cochran-Mantel-Haenszel test for general association, controlling for investigators.

The analysis of the secondary endpoint, reduction in hair growth was analyzed by Cochran-Mantel-Haenszel test, controlling for investigator. An ANOVA was used to analyze treatment difference in spatial mass with treatment, investigator and treatment-investigator interaction as effects in the model.

A multivariate analysis of variance was performed for subject's self-perception questionnaire with treatment, investigator and the interaction as covariates in the model.

Differences between treatments in the time to the first skin-related adverse event were evaluated by a time to event analysis using Wilcoxon test.

Subgroup analyses were performed to assess the effects of age, baseline hair growth, race, prior hair removal technique and change in dosage regimen on the primary endpoint.

III. Study Results

1. Study 1 (DE140-001):

Ten investigators at ten study centers in the United States enrolled 287 subjects between July 10, 1997 to July 30, 1998. One hundred ninety subjects were randomized to receive eflornithine 15% cream, ninety-seven subjects were randomized to vehicle cream. Based on the definition of ITT (or ASR), all the subjects randomized belonged to this analysis population. Subject accounting information was summarized in Table 1-1.

Demographic information showed reasonable balance between treatment groups except there was some imbalance in race. There were more white (62%) in eflornithine group than that in vehicle group (53%). However, the difference did not reach statistical significance (p-value 0.096). Another imbalance occurred was the skin type – there was more subjects having skin type II and III in eflornithine group (46%) than that in vehicle group (31%). The overall skin type difference was not statistically significant (p-value 0.158).

Table 1-1: Subject accounting information for Study 1.

	Eflornithine 15%		Vehicle	
	n	%	n	%
Randomized	190		97	
Received medication	188	100%	97	100%
Complete 24 wks	143	76.1%	73	75.3%
Complete 32 wks	139	73.9%	70	72.2%
Discontinued before Wk 24	47	25.0%	24	24.7%
Death	0	0.0%	0	0.0%
Due to AE	6	3.2%	5	5.2%
Due to lack of efficacy	0	0.0%	0	0.0%
Lost to follow-up	14	74.5%	8	8.3%
Others	25	13.3%	11	11.3%
Non-compliant	6		0	
Patient request	16		10	
Pregnancy	2		1	
Others	1		0	

Source: Based on sponsor's response requested by this reviewer received on May-8-2000.

Protocol violations in study conduct:

The protocol required that a study staff member other than the individual performing the PGA query the subjects concerning the status of non-serious skin related adverse events. This restriction was designed to avoid unintentionally unblinding the study medication through reports of certain skin related adverse events. However, this restriction was not followed at various investigation sites throughout the entire course of the study. Since the protocol violation may affect the study results, robust analysis was performed to see how sensitive the result is to the violation. The detail was documented in the reviewer's comment section.

Sponsor's efficacy results of primary endpoint:

The primary efficacy variable --- PGA assessed at week 24 showed that eflornithine treatment group was statistically significantly effective compared to vehicle group with p-value=0.001. The proportion of success defined as marked improvement and clear/almost clear was about 20% higher in eflornithine treatment group compared with vehicle group (24.4% in eflornithine group and 4.3% in vehicle). Results by visit were summarized in Table 2-1. As can be seen, the treatment difference increased as the increase of therapy duration. Also can be seen in this table, the analysis population was not ITT data set since it did not include all subjects randomized and received study medication as discussed later in reviewer's comment section.

Table 2-1: Physician's global assessment by visit for Study 1.

Week	Assessment	Elornithine15%	Vehicle	p-values
2	Clear/almost clear	0/176 (0.0%)	0/90 (0.0%)	0.273
	Marked improvement	7/176 (4.0%)	1/90 (1.1%)	
	Improved	49 /176 (27.8%)	15 /90 (16.7%)	
	No improvement	120/176 (68.2%)	74/90 (82.2%)	
4	Clear/almost clear	2/177 (1.1%)	0/92 (0.0%)	0.017
	Marked improvement	8/177 (4.5%)	0/92 (0.0%)	
	Improved	60/177 (33.9%)	23/92 (25.0%)	
	No improvement	107/177 (60.5%)	69/92 (75.0%)	
8	Clear/almost clear	4/178 (2.3%)	0/93 (0.0%)	0.007
	Marked improvement	20/178 (11.2%)	3/93 (3.2%)	
	Improved	87/178 (48.9%)	30/93 (32.3%)	
	No improvement	67/178 (37.6%)	60/90 (64.5%)	
16	Clear/almost clear	10/177 (5.7%)	0/92 (0.0%)	0.001
	Marked improvement	26/177 (14.7%)	0/92 (0.0%)	
	Improved	80/177 (45.2%)	31/92 (33.7 %)	
	No improvement	61/141 (34.5%)	61/92 (66.3%)	
24	Clear/almost clear	11/176 (6.3%)	0/92 (0.0%)	0.001
	Marked improvement	32/176 (18.2)	4/92 (4.3%)	
	Improved	75/176 (42.6%)	32/92 (34.8%)	
	No improvement	58/176 (33.0%)	56/92 (60.9%)	
32	Clear/almost clear	2/139 (1.4%)	0/70 (0.0%)	0.123
	Marked improvement	13/139 (9.4%)	3/70 (4.3%)	
	Improved	50/139 (36.0%)	24/67 (34.3%)	
	No improvement	74/139 (53.2%)	43/70 (61.4%)	

Source: Based on sponsor's tables 10.1.1, 10.1.2-1 – 10.1.2-4, and 10.1.3 on pages 85 – 88 in vol. 1.51.

Secondary analysis:

Video image data was incomplete due to technical problems and the implementation of the method. As a result, complete image data for the baseline and final (WK 24 or early discharge) visits were only 71% of all the subjects. This included 128 of 190 subjects (67%) in eflornithine group and 77 of 97 subjects (79%) in the vehicle group. For reduction in hair growth, analysis at week 24 showed no statistically significant treatment difference (p-value=0.158). Only 6.3% subjects were categorized as success in eflornithine group compared to 1.3% for vehicle.

An analysis of variance was used to analyze treatment differences in spatial mass with treatment and investigator in the model. Results of this analysis for spatial mass at week 24 showed a statistically significant treatment difference (p-value=0.0001). The mean spatial mass for the subjects treated with eflornithine was 0.037 mm^2, while 0.046 mm^2 for vehicle.

Subject's self-assessment questionnaire at baseline showed statistically significant difference between the two treatment groups (p-value=0.015), with less bothered/uncomfortable score in eflornithine treatment group. The individual questions that showed statistically significantly better at baseline in eflornithine group were uncomfortable at work or class and uncomfortable at social gathering. Analysis of covariate adjusting for the baseline measurement performed after 24 weeks of treatment showed statistically significant reduction of score in eflornithine group in all six questions. The six individual p-values for each question at Week 24 were 0.0046, 0.0005, 0.0011, 0.0022, 0.0045, and 0.0182.

Subgroup analyses:

The effects of age, race and prior hair removal technique on the primary efficacy variable were analyzed. The age was divided to <65 years old and 65 years and older. Since very few subjects were 65 years and older, this was a meaningless subgroup analysis. The success rates observed in eflornithine groups between white subjects and non-white subjects had quite large difference, 30.6% for white and 13.8% for non-white.

Skin related adverse events:

Time to onset of skin-related adverse events was analyzed by Wilcoxon test. No statistically significant difference was observed between treatment groups. Analyses on skin-related adverse effect by race suggested that white subjects were more vulnerable to skin-related adverse events than the non-white subjects. The skin-related AE in eflornithine treatment group was 70% for white vs. 55% for non-white.

2. Study 2 (DE140-002):

Nine study centers in the United States (8) and Europe (1) enrolled 309 subjects between July 1997 tc July 1998. Two hundred five (205) subjects were randomized to receive eflornithine treatment, 104 subjects were randomized to vehicle treatment. All the subjects randomized in this study belonged to ITT analysis population. Patient accounting information was summarized in table 1-2.

Table 1-2: Subject accounting information for Study 2.

	Eflornithine 15%		Vehicle	
	n	%	n	%
Randomized	205		104	
Receive medication	205	100%	104	100%
Complete 24 wks	161	78.5%	80	76.9%
Complete 32 wks	156	76.1%	78	75.0%
Discontinued before WK24	44	21.5%	24	23.1%
Death	0	0.0%	0	0.0%
Due to AE	4	2.0%	1	1.0%
Due to lack of efficacy	0	0.0%	0	0.0%
Lost to follow-up	15	7.3%	10	9.6%
Others	25	12.19%	13	12.5%
Patient request	16		10	
Pregnancy	1		1	
Physician's decision	1		0	
Others	7		2	

Source: Based on sponsor's response requested by this reviewer received on 5-8-2000.

Demographic information showed reasonable balance between treatment groups. White subjects constituted 67% of the overall subjects.

Problems in study conduct:

The protocol required that a different study staff member other than the individual performing the PGA query the subjects concerning the status of non-serious skin related adverse events. This restriction was designed to avoid unintentionally unblinding the study medication through reports of certain skin related adverse events. However, this restriction was not followed at various investigation sites throughout the entire course of the study. Similar to Study 1, robust analysis was performed to test the sensitivity of the results to the violation.

Sponsor's efficacy results of the primary endpoint:

The primary efficacy variable --- PGA assessed at week 24 showed that eflornithine treatment group was statistically significantly effective compared to vehicle group with p-value=0.001. The success rate was about 30% higher in eflornithine treatment group compared with vehicle group (43.9% in eflornithine group and 12.9% in vehicle). Results by visit were summarized in Table 2-2. Again it can be seen, the treatment difference increased as the increase of therapy duration. The analysis data set was not ITT population since it did not include all subjects randomized as discussed later in reviewer's comment section.

Table 2-2: Physician's global assessment by visit for Study 2.

Week	Assessment	Elornithine15%	Vehicle	p-values
2	Clear/almost clear	0/193 (0.0%)	0/98 (0.0%)	1.000
	Marked improvement	1/193 (0.5%)	0/98 (0.0%)	
	Improved	72/193 (37.3%)	22/98 (22.5%)	
	No improvement	120/193 (62.2%)	76/98 (77.6%)	
4	Clear/almost clear	0/193 (0.0%)	0/101 (0.0%)	0.258
	Marked improvement	19/193 (9.8%)	6/101 (5.9%)	
	Improved	95/193 (49.2%)	30/101 (29.7%)	
	No improvement	79/193 (40.9%)	65/101 (64.4%)	
8	Clear/almost clear	6/194 (3.1%)	0/98 (0.0%)	0.001
	Marked improvement	34 /194 (17.5%)	5/98 (5.1%)	
	Improved	93/194 (47.9%)	33/98 (33.7%)	
	No improvement	61/194 (31.4%)	60/98 (61.2%)	
16	Clear/almost clear	6/196 (3.1%)	0/97 (0.0%)	0.001
	Marked improvement	61/198 (31.1%)	5/97 (5.2%)	
	Improved	66/196 (33.7%)	33/97 (34.0%)	
	No improvement	63/196 (32.1%)	59/97 (60.8%)	
24	Clear/almost clear	10/198 (5.1%)	0/101 (0.0%)	0.001
	Marked improvement	77/198 (38.9%)	13/101 (12.9%)	
	Improved	57/198 (28.8%)	31/101 (30.7%)	
	No improvement	54/198 (27.3%)	57/101 (56.4%)	
32	Clear/almost clear	1/155 (0.7%)	1/75 (1.3%)	0.151
	Marked improvement	19/155 (12.3%)	4/75 (5.3%)	
	Improved	57/155 (36.8%)	31/75 (41.3%)	
	No improvement	78/155 (50.3%)	39/75 (93.3%)	

Source: Based on sponsor's tables 10.1.1, 10.1.2-1 – 10.1.2-4, and 10.1.3 on pages 87 – 90 in vol. 1.68.

Secondary analysis:

Due to problems of video image data, complete image data for the baseline and final (WK 24 or early discharge) visits were only 74% of the total subjects. This included 151 of 205 subjects (64%) in eflornithine group and 77 of 104 subjects (74%) in the vehicle group.

For reduction in hair growth, analysis at week 24 showed no statistically significant treatment difference (p-value=0.085). Only 8.6% subjects were categorized as success in eflornithine group compared to 2.6% for vehicle.

An analysis of variance was used to analyze treatment differences in spatial mass with treatment and investigator in the model. Results of the analysis for spatial mass at week 24 showed a statistically significant treatment difference (p-value=0.0004) favoring eflornithine group. The mean spatial mass for the subjects treated with eflornithine 15% cream was 0.036 mm^2, while 0.043 mm^2 for vehicle.

Multivariate analysis for subject's self-assessment questionnaire at week 24 showed statistically significant improvement in eflornithine treatment group compared with vehicle (p-value=0.0027). Analysis of covariate performed after 24 weeks of treatment showed statistically significant reduction of score in eflornithine group in all six questions. The six individual p-values at Week 24 were 0.0001, 0.0002, 0.0001, 0.0003, 0.0002, and 0.0001.

Subgroup analyses:

The effects of age, race and prior hair removal technique on the primary efficacy variable were analyzed. The age was divided to <65 years old and 65 years old and older. This was not a meaningful analysis since very few subjects were 65 years and older. The treatment difference observed between white subjects and non-white subjects was not as large as it was observed in Study 1. To aid the comparison between the two studies, the results of subgroup analyses by race were listed in Table 3 for the two studies side by side. The difference in success rates between white and non-white was 16.8% in Study 1 and only 6.8% in Study 2.

Table 3: Subgroup Analysis of Success Rate by Race for Study 1 and Study 2.

Race	Study 1				Study 2			
	Eflornithine		Vehicle		Eflornithine		Vehicle	
	n/N	%	n/N	%	n/N	%	n/N	%
White	34/111	30.6	2/48	4.2	61/132	46.2	10/66	15.2
Non-white	9/65	13.8	2/44	4.5	26/66	39.4	3/35	8.6
Total	176		92		198		101	

Source: Table 10.3-2 on page 100 in Vol.1.51 and Table 10.3-2 on page 104 in Vol.1.68.

Skin related adverse events:

Time to onset of skin-related adverse events was analyzed by Wilcoxon test. Statistically significant difference at level 0.05 was observed between treatment groups (p-value=0.044) with more adverse events occurred earlier in the eflornithine group than that in the vehicle group. In contrast to Study 1, analyses on skin-related adverse event by race showed almost equal vulnerability to skin-related adverse events between white and non-white in eflornithine treatment group (67% in white vs. 62% in non-white).

IV. Reviewer's Comments:

1. Study population:

In the sponsor's primary analysis on the primary efficacy variable (PGA), the analysis data set was not ITT population. The ITT population defined by the Division is all subjects randomized and received study medications. By this definition, 285 subjects in Study 1 and 309 in Study 2 should be included in ITT populations. For subjects who did not have physician's global assessment on Week 24, we classified their response to treatment as "No improvement". Therefore, no information on PGA assessment was carried forward to Week 24. The results of this reviewer's analysis based on ITT population showed that p-values were consistent with sponsor's analyses in both studies from Week 2 through Week 24. This was because the treatment difference between the treatment groups was quite large and insensitive to small change in data. The success rates at Week 24 for Study 1 were 22.9% for eflornithine and 4.1% for vehicle compared to 24.5% for eflornithine and 4.3% for vehicle, according to the sponsor's analysis. The success rates for Study 2 were 40.5% for eflornithine and 12.5 % for vehicle, as compared to sponsor's results 44% for eflornithine and 12.9% for vehicle. Tables 4-1 and 4-2 present the detailed results of ITT analyses for Study 1 and Study 2 respectively. Note in Study 2 that the number of success at Week 24 is 83, which is different from the results (87 success) obtained in sponsor's LOCF analysis.

Table 4-1: ITT analysis for the primary efficacy variable PGA at week 24 for Study 1.

		Eflornithine	Vehicle	p-value
Success	Clear/Almost clear	11 (5.9%)	0 (0.0%)	0.001
	Marked Improvement	32 (17.0%)	4 (4.1%)	
	Subtotal	43 (22.9%)	4 (4.1%)	
Failure	Improved	56 (29.8%)	24 (24.7%)	
	No Improvement/worse	89 (47.3%)	69 (71.1%)	
	Subtotal	147 (77.1%)	93 (95.9%)	
Total		188	97	

Table 4-2: ITT analysis for the primary efficacy variable PGA at week 24 for Study 2.

		Eflornithine	Vehicle	p-value
Success	Clear/Almost clear	10 (4.9%)	0 (0.0%)	0.001
	Marked Improvement	73 (35.6%)	13 (12.5%)	
	Subtotal	83 (40.5%)	13 (12.5%)	
Failure	Improved	45 (22.0%)	28 (26.9%)	
	No Improvement/worse	77 (37.6%)	63 (60.6%)	
	Subtotal	122 (59.5%)	91 (87.5%)	
Total		205	104	

2. By center analyses:

Since the sponsor did not provide detailed treatment by center analysis, in Table 5-1 and Table 5-2, by center analyses were presented for Studies 1 and 2 respectively for the completeness of the information. As can be seen from the two tables, there were quite large variations among treatment centers in eflornithine group for both studies. However, in vehicle treatment group, success rates were consistent among centers except one center in Study 2 had quite large success rate in Study 2. This tells us that vehicle treatment group showed no treatment effect consistently in two studies among the centers, however, the treatment effect in eflornithine treatment group varies from 0% to 64%.

Table 5-1: By center analysis on PGA for Study 1.

Center	Eflornithine		Vehicle		Total
	Success		Success		
	N	%	N	%	
00005	1/20	5.0	0/10	0.0	30
00006	3/29	10.3	1/14	7.1	43
00007	6/20	30.0	0/10	0.0	30
00008	8/30	26.7	2/16	12.5	47
00009	6/11	54.5	0/5	0.0	16
00010	6/19	31.6	0/10	0.0	29
00011	1/20	5.0	0/10	0.0	30
00012	0/9	0.0	0/6	0.0	16
00013	6/13	46.2	0/7	0.0	20
00014	6/17	35.3	1/9	11.1	26

Table 5-2: By center analysis on PGA for Study 2.

Center	Eflornithine		Vehicle		Total
	Success		Success		
	n/N	%	n/N	%	
00002	5/11	45.5	0/5	0.0	16
00003	6/18	33.3	1/9	11.1	27
00004	12/48	25.0	2/24	8.3	72
00005	8/17	47.1	2/9	22.2	26
00006	4/10	40.0	0/5	0.0	15
00007	7/21	33.3	5/11	45.5	32
00008	23/36	63.9	3/18	16.7	54
00012	4/7	57.1	0/4	0.0	11
00014	14/37	37.8	0/19	0.0	56

3. Robust Analyses:

Protocol violations in the conduct of both studies were reported, mainly because the same individual performing the PGA assessment also queried skin related adverse reaction. Such violation may cause unintentionally unblinding the study treatment since some skin related adverse reactions such as stinging, burning, tingling, and itching were treatment related. To assess the impact of such violation, robust analyses were performed by setting the PGA score as failure for those subjects who experienced stinging, tingling and rash in eflornithine treatment group. The robust analyses showed only small changes in both studies in the eflornithine treatment group. The success rate was 20.2% in robustness analysis vs. 22.9% in regular analysis in Study 1, while in Study 2 was 33.7% vs. 40.5%.

4. Difference between studies:

As can be seen from the primary analyses, the success rates between the two studies in eflornithine treatment groups were large, 22.6% in Study 1 vs. 40.5% in Study 2. Demographic information was comparable between the two studies. No baseline and background difference was identified that could explain the difference in the success rates between the two studies in the eflornithine treatment group. Since there were some imbalance in race and skin type in Study 1, to explore the effect of such factors, subgroup analyses were conducted for race and skin type.

The results for race subgroup analyses were listed in Table 6 for both studies side by side. The race was regrouped into three groups, White, Black and Other. As can be seen from the tables, the success rates in Study 2 were consistently high for all three race subgroups. The results for skin type were listed in Table 7. Skin type was regrouped into three subgroups. Again, the success rates in Study 2 were consistently higher in Study 2 than that in Study 1.

No apparent reason could explain the difference in success rate between the two studies.

Table 6: Subgroup analysis by Race for Study 1 and Study 2.

Success Rate by Race	Study 1				Study 2			
	Eflornithine		Vehicle		Eflornithine		Vehicle	
	n/N	%	n/N	%	n/N	%	n/N	%
White	34/117	29.1	2/51	3.9	61/137	44.5	10/69	14.5
Black	6/50	12.0	2/33	6.1	17/56	30.4	1/30	3.3
Other	3/21	14.3	0/13	0.0	5/12	41.7	2/5	40.0
Total	190		97		205		104	

Table 7: Subgroup analysis by skin type for Study 1 and Study 2:

Success Rates by Skin Types	Study 1				Study 2			
	Eflornithine		Vehicle		Eflornithine		Vehicle	
	n/N	%	n/N	%	n/N	%	n/N	%
I & II	11/40	27.5	1/20	5.0	26/57	45.6	3/30	10.0
III & IV	26/92	28.3	1/40	2.5	42/95	44.2	9/45	20.0
V & VI	6/56	10.7	2/37	5.4	15/52	28.9	1/29	3.4
Total	190		97		205		104	

5. Multiplicity in self-assessment questionnaire:

Since the results of self-assessment questionnaire will appear on the label, which consisted of six individual questions, multiplicity adjustment procedure should be applied to ensure the appropriate significance level of each individual p-values for the six questions. Although there was no prespecified multiplicity adjustment approach, to require all six questions to be statistically significant at 0.05 is a stringent criteria. Since all the six individual p-values were highly statistically significant (<0.05) in both studies, which satisfied the stringent criteria. Therefore, it was appropriate to conclude that all the individual questions showed conceivable treatment difference.

6. Analysis of video image results – growth length at Week 24:

Requested by the medical officer Dr. Cook, p-values for mean hair growth length at Week 24 were calculated for both studies. A simple t-test was performed with only treatment included in the model. Without missing data imputation, the analysis included 160 subjects in eflornithine treatment group and 87 in vehicle group for Study 1 and 178 subjects in eflornithine treatment group and 92 in vehicle group for Study 2. The analyses yielded p-values 0.001 for both studies. Detail was listed in Table 8 and Table 9 for Study 1 and Study 2 respectively.

Table 8: Analysis for hair length at Week 24 for Study 1.

Hair length	Eflornithine	Vehicle	Overall	p-value
N	160	87	247	0.001
Mean	0.404	0.484	0.432	
S.E.	0.009	0.015	0.008	
Range				

Table 9: Analysis for hair length at Week 24 for Study 2.

Hair length	Eflornithine	Vehicle	Overall	p-value
N	178	92	270	0.001
Mean	0.404	0.469	0.426	
S.E.	0.009	0.013	0.008	
Range				

V. Conclusion:

Both studies presented statistically significant treatment differences between eflornithine and vehicle treatment groups. The treatment differences were 18.8% in Study 1 and 28.0% in Study 2. Subject self-assessment questionnaire also showed statistically significant treatment difference in all six individual questions in both studies. For all the secondary efficacy variables, only hair growth by video image analysis failed to show statistically significance at 0.05 (p-values were 0.158 for Study 1 and 0.085 for Study 2). However, the results generated by video image analysis were not reliable due to technical problems in devices and large volume of missing data.

Although the results of both studies were statistically significantly in favor of eflornithine treatment group in treating women with excessive facial hair, however, the data could not explain the large treatment difference in eflornithine group observed in the two pivotal studies.

/S/ 6-19-00

Qian Li, Sc.D
Mathematical Statistician

Concur:

/S/ 6/19/00

Mohamed Al-Osh, Ph.D
Acting Team Leader

CC:
HFD-540/Division File
HFD-540/Dr. Wilkin
HFD-540/Dr. Walker
HFD-540/Dr. Cook
HFD-540/Ms. Wright

HFD-725/Division File
HFD-725/Dr. Huque
HFD-725/Dr. Al-Osh
HFD-725/Dr. Li

This review consists of 14 pages of text and tables.

Pre-clinical Statistical Consult

NDA: 21-145

APR 4 2000

Drug Class: Hair Removal Product

Name of Product: Vaniqa, BMS-203522 (eflornithine) lotion

Applicant: Bristol-Myers Squibb
Pharmaceutical Research Institute
Department of Toxicology
Buffalo, New York

Indication: Treatment of excessive facial hair in women.

Documents Reviewed: Volumes 24 and 25 of NDA 21-145 dated 28 September 1999.

I. Background:

At the request of Dr. Barbara Hill, HFD-540, Division of Dermatologic and Dental Drug Products, the reviewing toxicologist and pharmacologist for this submission, a dermal carcinogenicity study in mice was selected for statistical review of the sponsor's analysis.

II. A Two Year Dermal Carcinogenicity Study in the Albino Mouse:

II. a. Summary:

According to the sponsor "BMS-203522, as 15% lotion, was administered from a precalibrated ——— pipette once daily 7 days a week for 2 years to three oncogenicity groups of 50 male and 50 female mice at doses of 150, 300, or 600 mg/kg/day (25, 50, or 100 µL/mouse/day). Two groups of 50 mice/ sex served as controls. One group was an untreated control. The other group was a vehicle control and received the control vehicle (vehicle lotion for BMS-203522). Criteria for evaluation included survival, clinical observations, dermal irritation assessments, body weights, food consumption, and gross and microscopic pathology following unscheduled and scheduled necropsies." (page 10, volume 1.24)

Table 1., below, summarizes the mortality for the mice classified by sex, dose, and duration of exposure.

Table 1. Mortality

Males:

Week	0 mg/kg/day Untreated Control	0 mg/kg/day Vehicle Control	150 mg/kg/ day	300 mg/kg/ day	600 mg/kg/ day
1-52	3	2	4	5	7
52-78	10	13	7	8	7
79-104	14	16	19	19	19
Final Sacrifice	20	17	19	18	17
Total	47	48	49	50	50

Females:

Week	0 mg/kg/day Untreated Control	0 mg/kg/day Vehicle Control	150 mg/kg/ day	300 mg/kg/ day	600 mg/kg/ day
1-52	3	11	5	5	4
52-78	13	11	5	10	9
79-104	19	18	23	18	20
Final Sacrifice	14	10	17	17	18
Total	49	50	50	50	50

Originally 50 animals were assigned per gender/dose group. It is this reviewer's opinion that the sponsor's explanations for the losses reflected above were adequate and obviously not related to the treatment under test.

II. b. Sponsor Analyses:

Mortality data were analyzed using logrank tests to compare the within treatment group product-limit/Kaplan-Meier estimated survival curves separately for males and females. (see curves pages 37 and 38 of volume 1.24). The sponsor reports that there was no statistically significant difference among treatment groups ($p>0.05$), as one might expect either from inspection of the survival curves or from the mortality tables above.

For body weight and food consumption the sponsor proposed a relatively complicated comparisons. First homogeneity of variance was tested using Levine's test. If Levine's test was statistically significant ($p<.01$), pair-wise tests using Welch's t-test were performed. Otherwise a Dunnett's test, presumably comparing treatment group to control was used. It is this reviewer's opinion that with the level of homogeneity found in these experiments, simple analysis of variance would be sufficient. But that is largely a matter of taste, and it is this reviewer's opinion that the sponsor's procedure is quite defensible.

According to the sponsor: "During Weeks 16-30, body weights for the 3 groups of treated males were statistically significantly decreased compared with the untreated controls. These differences between the treated groups and the untreated control were not evident in the 150 and 300 mg/kg/day groups after Week 30. The body weights in the 600 mg/kg/day males remained significantly decreased relative to the untreated controls during Weeks 31 to 61. "

Continuing: "During Weeks 15-30 and 46-61, all 3 groups of treated females had body weights that were statistically significantly decreased compared to the untreated control group. These differences were not evident during Weeks 31-45 and after Week 61." Further, the sponsor notes that there were sporadically statistically significant differences in food consumption between the control groups and the treated groups during the study. However, these seem to show no particular pattern, and like the results on comparing body weights above, it is this reviewer's opinion that the sporadically statistically significant differences are typical of the artifacts that appear in any study. Note that plots of mean body and mean food consumption are given on pages 39-42 of the report.

For carcinogenicity, the protocol specified that: "Differences in tumor rates between groups will be analyzed using Peto analysis for individual tumor types when at least two animals with comparable tumors are found in the high-dose group or at least four animals with comparable tumors are found in the combined intermediate- and high-dose groups. The control group will be compared to the vehicle group when the number of tumors in the vehicle group is three more than the number of tumors in the untreated control. The combined fatal and incidental Peto one-sided test for trend of increasing tumor rates over dose levels will be considered statistically significant at $p<0.005$ for common tumors and $p<0.025$ for rare tumors." (page 120, volume 25)

Note the limits on significance levels cited above are Haseman's rules for tests of trend. That is, based on his extensive experience with such analyses, Haseman (1983) proposed a p-value adjustment rule that is applicable to these comparisons. That is, for a roughly 0.10 (10%) overall false positive error rate in tests of trend, rare tumors (with a historical control incidence 1% or below) and common tumors should be tested as above. The corresponding rule for pairwise tests is that is, for a roughly 0.10 (10%) overall false positive error rate, rare tumors should be tested at a 0.05 level, and common tumors (with a historical control incidence greater than 1%) at a 0.01 level.

To this reviewer it is somewhat surprising that for a dermal study there is no explicit mention of the treatment of mortality independent tumors (i.e., observable, particularly skin tumors). Note these would normally be treated identically to fatal tumors in the tumor analysis (i.e., using a logrank test), but apparently there were few such tumors so presumably no explicit treatment was necessary.

The sponsor reports the analyses of neoplastic lesions on pages 83-105 of volume 25. The sponsor's statistical analysis of differences among doses in various neoplastic lesions included the Cochran-Armitage Trend test, and pairwise comparisons with the untreated control

using a Fisher Exact test, as well as the Peto test for trend specified in the original protocol. The first two tests are NOT adjusted for differences in mortality, and hence would usually not be preferred by Division of Biometrics reviewers. However, since there seem to be no dose related trends in the data, for this study the Cochran-Armitage test and Fisher Exact test should be appropriate. This has some confirmation in the fact that for this data, the Cochran-Armitage test of trend (not adjusted for mortality) differs little from the Peto test of trend (adjusted for mortality).

Whether or not different reported neoplasms should have been combined is a matter for the scientific judgement of the toxicologist. Assuming the reported combinations of neoplasms are appropriate, even without adjustment for multiple comparisons using Haseman's rules, for those neoplasms analyzed no test of trend or pairwise comparison with the untreated control was even statistically significant at a 0.05 level. It is this reviewer's opinion that the few comparisons that had significance levels close to this level were typical of the types of artifacts that occur when numerous statistical tests are performed.

III. Validity of Design:

Lin and Ali (1994), quoting work by Haseman, have suggested that a 50% survival rate between weeks 80-90 of a two-year study may be considered a sufficient number of survivors as well as one measure of adequate exposure. From table 1 above, we see that this has been achieved.

In analyses performed in the United States, it is traditional that the highest dose should be close to the MTD to achieve the greatest likelihood of tumorigenicity. Chu, Ceuto, and Ward (1981) proposed three criteria to see if the high dose is close to the Maximum Tolerated Dose (MTD) and presents a reasonable tumor challenge to the animals. They recommended that a high dose be considered close to the MTD if:

i.) there was a detectable weight loss of up to 10% in the dosed group relative to controls.

ii.) there were exhibits of clinical signs or severe histopathologic toxic effects attributable to the chemical in the dosed animals, and/or

iii.) there was a slightly increased mortality in dosed animals compared to controls.

From the plots of mean body weight (pages 39 and 40 of volume 24), it does appear that i) is satisfied for females, but not males. However, there seems to be no strong evidence for the other criteria. Thus, using the Chu, Ceuto, and Ward criteria, it seems to this reviewer that the selected dose was not close to the MTD.

The above evaluation of the appropriateness of the design and whether or not the MTD was achieved is based only on bodyweight and survival data. Information regarding clinical signs and histopathological data, plus other possible considerations, are well beyond the expertise of this reviewer, but presumably would be used by the toxicologist in the final assessment of the adequacy of this experiment.

IV. Conclusion: –

According to the sponsor: "No drug related effects were observed in survival, clinical observations, body weights, food consumption, or gross microscopic observations. . . . There was no evidence of oncogenic or non-oncogenic effects in any organ following dermal application of BMS-203522 to male and female mice." Assuming the appropriate neoplasms were chosen, this reviewer sees no reason to dispute those conclusions.

APPEARS THIS WAY
ON ORIGINAL

References:

Chu, K.C., Ceuto, C., and Ward, J.M. (1981), Factors in the Evaluation of 200 National Cancer Institute Carcinogen Bioassays, *Journal of Toxicology and Environmental Health,* 8, 251-280

Haseman, J. K. (1983), A Reexamination of false-positive rates for carcinogenicity studies, *Fundamental and Applied Toxicology,* 3, 334-339.

Lin, K. K. and Ali, M.W. (1994), Statistical Review and Evaluation of Animal Tumorigenicity Studies, *Statistics in the Pharmaceutical Industry, Second Edition, Revised and Extended,* edited by C.R. Buncher and J.Y. Tsay, Marcel Dekker, Inc. New York.

Peto, R., Pike, M.C., Day, N.E., Gray, R.G., Lee, P.N., Parrish, S., Peto, J., Richards, S., and Wahrendorf, J. (1980) Guidelines for sample sensitive significance tests for carcinogenic effects in long-term animal experiments, *IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans, supplement 2: Long term and Short term Screening Assays for Carcinogens: A Critical Appraisal,* International Agency for Research Against Cancer, 311-426.

APPEARS THIS WAY
ON ORIGINAL

/S/ 4/4/00

for Steve Thomson
Mathematical Statistician, Biometrics III

/S/ 4/4/00

concur: Mohamed Al-Osh, Ph.D.
Acting Team Leader, Biometrics III

This review has 6 pages, including this signature page.
cc:
Archival: NDA 21-145
HFD-540/
HFD-540/Dr. Wilkin
HFD-540/Dr. Walker
HFD-540/Dr. Cook
HFD-540/Ms. Wright
HFD-540/Dr. Hill
HFD-540/Dr. Jacobs
HFD-725/Dr. Huque
HFD-725/Mr. Thomson
HFD-725/Dr. Al-Osh
Chron.

\Thomson\Word Document\(301)827-2044\March 30, 2000\c:\My Documents\Carc21145BMS.doc