CENTER FOR DRUG EVALUATION AND RESEARCH

Application Number NDA 21-108

STATISTICAL REVIEW(S)

Statistical Review and Evaluation

AUG 11 2000

NDA/ Drug Class: 21-108 / 3S

Name of Drug: Renova® (Tretinoin Emollient Cream) 0.02%

Applicant: Johnson and Johnson
199 Grandview Road
Skillman, NJ 08558-9418

Indications: Mitigation (Palliation) of fine ~~wrinkling, mottled hyperpigmentation, tactile roughness, and skin laxity.~~

Documents Reviewed: Volumes 1.1,1.2, 1.88-1.154 and diskettes containing SAS data sets from the sponsor

Medical Officer: Markham Luke, M.D., Ph.D. (HFD-540)

Contents:

I. Introduction . . . . . page 2
II. Experimental Designs . . . . . page 2-3
II. A. TEC-II 0.02% Formulations . . . . . page 3-4
II. B. Endpoints . . . . . page 4-5
II. C. Statistical Methods . . . . . pages 5-6
III. Primary Efficacy Results
III. A. Protocol J89-024 . . . . . pages 7-9
III. B. Protocol J89-025 . . . . . pages 10-12
III. C. Protocol J89-045 . . . . . pages 13-14
III. D. Protocol L91-026 . . . . . pages 15-16
IV. Supporting Study: Protocol K90-011 . . . . . pages 17-18
V. Overall Efficacy . . . . . pages 19-21
VI. Adverse Events . . . . . pages 22-24
References . . . . . page 25
Conclusions . . . . . pages 26-27
Signature Page . . . . . page 28
Appendix
Supporting Tables for J89-024 (A.1 & A.2) . . . . pages 29-30
Supporting Tables for J89-025 (A.3 & A.4) . . . . pages 31-32
Supporting Tables for J89-045 (A.5 & A.6) . . . . pages 33-34
Supporting Tables for L91-026 (A.7 & A.8) . . . . pages 35-36
Supporting Tables for K90-011 (A.9 & A.10) . . . . pages 37-38
Skin Irritation Table A.11. . . . . pages 39-42
Figures 1.-5. . . . . page 43-45

I. Introduction

When this NDA 21-108, RENOVA 0.02% (tretinoin emollient cream), formulation TEC-II 0.02%, was submitted, the sponsor originally claimed the indication of reducing the general signs and symptoms of photoaging. It was proposed that this general indication be measured at the end of the study (24 weeks) by the primary endpoints: an investigator's global evaluation, the change from baseline in an investigator's evaluation of overall severity of photodamage, and the overall subject self-assessment of photodamage. Most of the sponsor's original submission addressed these primary endpoints. Concurrently, as secondary measures, six general signs and symptoms of such damage were assessed: tactile roughness, fine wrinkling, coarse wrinkling, mottled hyperpigmentation, yellow-brown discoloration (labelled as "yellowing" by the sponsor), and skin laxity. Each of these latter six endpoints were scored by each investigator on a 10 point scale (0-9, with small numbers being more favorable). Photographs were provided to normalize the scale.

However, the consensus of the medical officers in the Division of Dermatological and Dental Drug Products was that there was no medical condition that corresponded to "photoaging", and hence that the three global evaluations of photodamages cited above were not easily interpretable. However, the six general signs and symptoms of photoaging, originally defined as secondary variables, were felt to be interpretable, and were thus appropriate primary endpoints. It was felt that these general signs and symptoms were manifested through a variety of possibly separate and possibly obscure biological processes. While some of these responses were inherently correlated (e.g. fine and coarse wrinkling), others were induced by processes so independent that the process that induced each condition may be treated as "orthogonal" to the process generating the others. Thus, pooling these measures to give a global measure of photoaging was not considered to be clinically appropriate.

Results from five studies provided the primary support for results. Except as otherwise noted, all results are based on this reviewer's analysis, applied to the data sets provided by the sponsor. This differed from the sponsor's analysis in the latter's analyses either used endpoints not considered appropriate or failed to follow the original protocol.

II. Experimental Designs

Five studies, two originally labelled by the sponsor as "primary" and three originally labelled as "secondary", using the six (in one case five) endpoints cited above were analyzed in this review to investigate statistically the effects of TEC-II 0.02% emollient cream on the endpoints noted above. In fact, with the concurrence of the Medical Officer, we are using the four multicenter studies as primary, with the one single center study as supporting. Apparently, these (and most of the other supporting studies not discussed here) were conducted from 1989 to 1993.

Table 1. The Studies

Protocol Number	Description
Sponsor labelled primary studies:	
J89-024 and J89-025	A double-blind, parallel, U.S. multicenter trial comparing the efficacy and safety of Renova (Tretinoin Emollient Cream) 0.02% with its vehicle qid 24 weeks in the treatment of tactile roughness, fine wrinkling, coarse wrinkling, mottled hyperpigmentation, yellow-brown discoloration, and skin laxity. Centers for J89-024 were in Ann Arbor, Michigan, Tucson, Arizona, and Snellville, Georgia. Centers for J89-025 were in Cleveland, Ohio, New Haven, Connecticut, and Atlanta, Georgia.
Sponsor labelled supportive studies:	
J89-045	A double-blind, parallel, multicenter trial conducted in Germany and Sweden comparing the efficacy and safety of Renova (Tretinoin Emollient Cream) 0.02% with its vehicle qid 24 weeks in the treatment of tactile roughness, fine wrinkling, coarse wrinkling, mottled hyperpigmentation, yellow-brown discoloration, and skin laxity. There was a 12 week off-therapy follow-up phase.
L91-026	A double-blind, parallel, U.S. multicenter trial comparing the efficacy and safety of Renova (Tretinoin Emollient Cream) 0.02% with its vehicle qid 24 weeks in the treatment of tactile roughness, fine wrinkling, coarse wrinkling, local hyperpigmentation, general hyperpigmentation, and skin laxity among non-Caucasian, non-Asian subjects. The study included a 28 week follow-up where all patients were treated with RENOVA.
K90-011	A double-blind, parallel, U.S. single center trial comparing the efficacy and safety of Renova (Tretinoin Emollient Cream) 0.02% with its vehicle qid 24 weeks in the treatment of tactile roughness, fine wrinkling, coarse wrinkling, mottled hyperpigmentation, yellow-brown discoloration, and skin laxity. There was a 12 week off-therapy follow-up phase.

Again, despite the sponsor's labelling, we are treating the first four studies as primary, with the last single center study as supporting. SAS data sets were provided for each of the studies above, and unless otherwise noted, were used to generate each of the following tables or analyses.

In each study above, treatment was to be applied once nightly for 24 weeks, with a general dosing guideline of 0.25 g per application. Return visits were scheduled two and four weeks after starting in the study, and every four weeks thereafter until the subject completed the study.

II.A. TEC-II 0.02% Formulations

There was a slight problem in that all studies except the L91-026 study were conducted on a formulation of RENOVA that differs from the to-be-marketed formulation. That is, all of the five studies cited above, except L91-026, plus most of the other phase I/II studies not discussed here, were conducted using TEC-II 0.02% without fragrance However, the to be marketed formulation includes a fragrance.

The sponsor did include the results of a small tolerance study, K90-016 to justify the claim of equivalence. This was a single-center, double-blind, within subject study in 25 healthy

Caucasian subjects. Five study drugs were applied to semi-occlusive patches on locations randomly allocated on each subject's back. Each site was evaluated 24 hours (72 on weekends) after each application. The reported results are summarized in the following table (provided by the sponsor – page 410, volume 8.):

Table 2. Study K90-016: Total Cumulative Irritation

Study Drug	2 week total score/ max score	3 week total score/ max score
TEC-II Vehicle with Fragrance	6.5/1000	23/1500
TEC-II Vehicle without Fragrance	5.0/1000	17/1500
TEC-II 0.05% with Fragrance	58.5/1000	284.5/1500
TEC-II 0.05% without Fragrance	55.5/1000	257.5/1500
TEC-II 0.02% with Fragrance	31.5/1000	122.5/1500

No estimate of variation was provided, so it is difficult to compare these total scores. However, it is apparent that use of the fragrance seems to be associated with increased irritation in both the vehicle and TEC-II 0,05%. No evaluation of a TEC-II 0.02% treatment group without fragrance was reported. The endpoint seems to be a safety endpoint, not efficacy. Hence using this study to justify equivalence of the tested formulation to the to be marketed formulation seems problematical.

Whether or not this discrepancy in formulations is of importance is a decision requiring the expertise of the Medical Officer.

II.B. Endpoints

Clinical assessments were performed by the investigators at baseline and at four week intervals during the study. The six primary signs and symptoms of photodamage, originally defined as secondary variables, were:

tactile roughness	mottled hyperpigmentation,
fine wrinkling	yellow/brown discoloration
coarse wrinkling	skin laxity.

were each evaluated on a 10 point scale from 0-9, defined as 0=none (absent), 1 to 3=mild, 4 to 6=moderate, and 7 to 9=severe. The investigator evaluation of overall severity was also evaluated on the same scale. Patients were required to have a score of at least moderate (4 or higher) on this latter variable for entry to the study. "A set of reference photographs depicting various grades of photodamage was given to each study center prior to the study to standardize grading criteria over time and across investigators." In the L91-026 study, in patients with skin types III-V, assessments were made of local hyperpigmentation and general hyperpigmentation instead of simple hyperpigmentation and yellowing as in the other studies. However, these were measured on the same scale.

In each study, the double-blinded phase of treatment continued to week 24 (or beyond). Endpoints were assessed at week 24, with subjects who reached that time point in the study, and also using the last observation carried forward (LOCF) to week 24 to impute missing observations at the end of treatment. In addition, for each endpoint two patient groups were analyzed: the intent-to-treat (ITT) population, i.e., all patients randomized and dispensed

medication, and a modified intent-to-treat (MITT) population, defined as those subjects with a baseline score of at least two or greater on that particular endpoint. It was the opinion of the Medical Officer that such a population better reflected the patients who would receive this treatment.

These studies used difference scores from baseline to adjust for baseline differences. Unless the change from baseline is a much more clinically relevant endpoint than the original measure, this reviewer would usually prefer the original endpoint. Note that for the change from baseline to be interpretable, we need to treat the ten point scale of the original measure as interval level data, a assumption which might be debatable. Even assuming interval level data, all other considerations being equal, this reviewer would usually recommend that the original scores be used, with randomization used to balance the baseline scores, or that the outcomes be analyzed by a method allowing using baseline as a measured covariate. However, the original protocols provided by sponsor called for the use of change from baseline response measures. And even more important, it was the opinion of the Medical Officer that the change from baseline was a more clinically relevant endpoint than the original measure. Hence, the primary endpoints used here are for the change from baseline.

II.C. Statistical Methodology

The protocols for each of the studies originally proposed that these responses would be analyzed by analysis of variance on the difference scores from baseline. Both the original measures and the changes from baseline are often quite skewed. Because of this skewness, with no other considerations, this reviewer would have preferred a permutation test, where the statistical significance of the observed treatment differences is based on the randomization distribution of the observed data, stratified on investigator. Note that for each study the actual randomization was apparently performed in blocks of four subjects within each center, but the effect of such restrictions on the actual randomization distribution of the ANOVA test statistic is usually ignored, and was ignored here.

A permutation test is a test of hypotheses based solely on the original randomization of the data. Such tests are often also called randomization tests, sometimes "exact" tests, or more generally, design-based tests. No *a priori* model or distributional assumptions are needed. The analysis is based on the randomization. Clearly such tests have attractive robustness features. Restriction to a subgroup of subjects, as is done when using the MITT population, means that we no longer have a wholly design-based justification for using the permutation/ randomization distribution of the test statistic. It is true that since MITT is defined at baseline prior to allocation to treatment, we would expect that with repeated runs of the experiment, it would be independent of treatment allocation. But that is only from a model based point of view. From a wholly design based point of view, the restriction to the MITT group does invalidate the permutation/randomization distribution. Thus, statistically, it makes more sense to base the analysis on the ITT population. However, again, this was over-ridden by the need for clinical relevance expressed by the Medical Officer.

Further, unless the analysis proposed by the original protocol is clearly inappropriate, it is this reviewer's opinion that the protocol should generally be followed. Since the protocol specified ANOVA, it was used as the main analysis method. Note that the corresponding

results for the randomization analysis are also given. Results are given both for the MITT and the ITT patient groups as defined above.

It is no coincidence that results for ANOVA and the permutation tests are similar. As noted by Fisher (1935) the t-test, or its equivalent ANOVA test, can be looked at as an approximation to the results from the permutation/randomization distribution. These results apply best to the Type I sums of squares, where one compares simple treatment means. Most ANOVA analyses in the United States seem to use Type III sums of squares, where one analyzes a pooled within center comparison of means. And ANOVA tests reported here also used the Type III sums of squares. However, for balanced data, as here, especially for the ITT population, these sums of squares are identical.

Because there are six possible primary endpoints, correction for multiple endpoints is needed. Bonferroni corrections where the observed significance level is compared $\alpha/6$, for a familywise level α, could be used, but are extremely conservative. Holm (1979) provided a modified step-up Bonferroni method, described below:

Holm's method performs testing in decreasing order of significance, i.e. starting at the smallest p-value. Testing is continued until a null hypothesis is accepted, i.e. an observed p-value is larger than the corresponding Holm's p-value, or until the hypotheses corresponding to all comparisons are rejected. With 6 endpoints, for a specified family-wise Type I error rate α, ordering the tests from k=1 to 6, from largest to smallest, the corresponding Holm's p-value is α/k. Thus the first test, k=1, with the smallest p-value, is compared to $\alpha/6$. If it is not statistically significant then stop. If it is significant, the next p-value is compared to $\alpha/5$. The procedure continues by comparing the observed increasing p-values to increasingly large α/k's until a non-significant comparison is reached. Once we reach a non-significant test no further comparisons are made. Comparisons whose significance levels are less than the corresponding Holm's α/k are declared to be statistically significant. Again, this procedure also maintains family-wise Type I error for partial or complete null hypotheses at or below α. Note that these corrections are performed separately within each of the four primary studies.

A final multiplicity issue is that we have 5 studies, 4 of which are being considered as primary. The usual interpretation of the CFR has been that we need to show statistical significance in "studies," i.e., at least two studies. Despite some question of its scientific merit this has been further interpreted to mean that if any two studies show statistically significant outcomes, we accept this as a statistically significant result. Clearly, generalizing this procedure to many studies with one to few endpoints could be quite anti-conservative. However, here we have four studies with six endpoints, and the decision procedure of requiring at least two out four studies to show statistical significance coupled with Holm's procedure within each study was felt to provide reasonable control of error. Work on these justifying (or disproving) this claim is proceeding in the Division of Biometrics 3. However, preliminary simulation results seem to suggest that under the circumstances here, family-wise Type I error is quite well controlled.

One further note on these endpoints, is that prior to the explanation of these conditions by the Medical Officer, this reviewer conducted a factor analysis of the six response measures within each of the two studies initially labelled a "primary" by the sponsor. This was done to see if they could be considered as being indicators of some one-dimensional construct, which one

might label as "photo-damage". However, even with just the six measures neither analysis was consistent with the notion of a single dimensional factor with independent uniquenesses.

III. Primary Efficacy Results:

III.A. Protocol J89-024

All patients were Caucasian, with demographic characteristics as summarized in Table A.1 of the appendix.

The following table, text Table 3., displays the mean change from baseline for each of the six endpoints: tactile roughness, fine wrinkling, coarse wrinkling, mottled hyperpigmentation, yellow-brown discoloration, and skin laxity. Again, the ITT population consists of all subjects dispensed treatment. Response measures are given at week 24 both for all subjects with data at week 24, i.e. essentially a slight superset of the fully evaluable group, and for the set with the imputed values for all subjects dispensed medication missing this week 24 measurement, using LOCF imputation. A "modified" intent to treat (MITT) group was formed for each endpoint by excluding those subjects in the ITT population with a 0 or 1 at baseline on the specified endpoint. The number of subjects involved, the mean change, and the standard deviation of the change are provided for both the ITT and the MITT populations at week 24 and using imputation with LOCF. In general, following apparent ICH guidelines, we emphasize the LOCF results.

Significance levels are given for the corresponding test of treatment differences using both an analysis of variance and a corresponding permutation test. Because the protocol specified the use of ANOVA, it was preferred for analysis. Note that while several measures had relatively large investigator effects, there were no statistically significant treatment by investigator interactions at any of these time points ($p>0.15$).

For both populations the values corresponding to those subjects who completed the 24 week course of treatment are given only as supporting results. Instead, for the statistical reasons given earlier, the statistician had a slight preference for using the ITT population at week 24 for the primary analysis. However, the Medical Officer preferred to use the corresponding MITT population, since these would better reflect the population of potential users. Thus, while results will be given for both groups, if only to show that the results are quite consistent between the population groups, the final emphasis will be on the MITT group chosen by the Medical Officer. Again, because there are a six possible endpoints, correction for multiple endpoints is needed before final conclusions can be drawn.

Prior to the adjustment for multiple endpoints, note that only for fine wrinkling was there a statistically significant difference at the .05 level. Others are close, but these will be inflated by the adjustment for multiplicity. Note that this statistically significant difference is associated with a difference of about 0.3 or 0.4 when these are measured on a 10 unit scale.

Table 3. Study J89-024: Differences From Baseline

	Population							
	ITT				MITT			
	Week 24		LOCF		Week 24		LOCF	
	Treat-ment	Veh-icle	Treat-ment	Veh-icle	Treat-ment	Veh-icle	Treat-ment	Veh-icle
Tactile Roughness								
Mean	-0.9	-0.9	-0.8	-0.8	-1.5	-1.5	-1.3	-1.5
Std Dev	1.1	1.1	1.2	1.1	1.2	1.1	1.2	1.1
n	77	83	90	90	43	46	49	48
p-value(ANOVA)	0.8051		0.6716		0.5652		0.4916	
p-value(Exact)	0.8183		0.7240		0.7794		0.5267	
Fine Wrinkling								
Mean	-0.9	-0.5	-0.8	-0.5	-0.9	-0.5	-0.8	-0.5
Std Dev	0.8	0.7	0.8	0.7	0.8	0.7	0.8	0.7
n	77	83	90	90	76	83	89	90
p-value(ANOVA)	0.0004		0.0021		0.0003		0.0017	
p-value(Exact)	0.0004		0.0030		0.0004		0.0021	
Coarse Wrinkling								
Mean	-0.5	-0.3	-0.5	-0.3	-0.5	-0.3	-0.5	-0.3
Std Dev	0.7	0.6	0.7	0.6	0.7	0.6	0.7	0.6
n	77	83	90	90	77	83	90	90
p-value(ANOVA)	0.0333		0.0547		0.0333		0.0547	
p-value(Exact)	0.0371		0.0693		0.0371		0.0693	
Mottled Hyperpigmentation								
Mean	-1.2	-1.0	-1.1	-0.9	-1.3	-1.0	-1.1	-0.9
Std Dev	1.1	0.9	1.1	0.9	1.1	0.9	1.1	0.9
n	77	83	90	90	71	78	84	85
p-value(ANOVA)	0.0831		0.2041		0.0622		0.1741	
p-value(Exact)	0.0809		0.2329		0.0725		0.2236	
Yellow-brown discoloration								
Mean	-1.0	-0.7	-0.9	-0.7	-1.5	-1.0	-1.3	-1.0
Std Dev	1.2	0.9	1.1	0.9	1.1	0.9	1.2	0.9
n	77	83	90	90	48	55	57	59
p-value(ANOVA)	0.0178		0.0952		0.0044		0.0663	
p-value(Exact)	0.0176		0.1122		0.0029		0.1038	
Laxity								
Mean	-0.5	-0.4	-0.5	-0.4	-0.6	-0.4	-0.5	-0.4
Std Dev	0.8	0.6	0.8	0.6	0.8	0.7	0.8	0.6
n	77	83	90	90	68	78	81	85
p-value(ANOVA)	0.2806		0.3821		0.1142		0.1834	
p-value(Exact)	0.3228		0.4555		0.1935		0.3106	

The following table, text Table 4., provides multiplicity corrected results for the significance levels associated with the LOCF subjects at week 24 from the table above.

Table 4. Study J89-024: P-values for Holm's test

Holms p-values	J89-024	MITT original p-values	ITT original p-values	MITT adjusted p-values	ITT adjusted p-values
0.0084	Fine Wrinkling	.0017 *	.0021 *	.0099 *	.0129 *
0.010	Coarse Wrinkling	.0547	.0547	.2734	.2734
0.0125	Yellow-brown Discoloration	.0663	.0952	.2734	,3807
0.0167	Mott Hyperpig.	.1741	.2041	.5227	.6122
0.0250	Laxity	.1834	.3821	.5227	.7643
0.0500	Tactile Roughness	.4916	.6716	.5227	.7643

* - denotes a statistically significant (at 0.05 level) comparison

Thus, adjusting for the multiple endpoints, only the comparison for fine wrinkling is statistically significant at the .05 level ($p \leq 0.0099$).

Note that all subjects were Caucasian, and almost all were female. Ages ranged between 45 to 69. So it was felt that the usual subgroup analyses would be superfluous.

Appendix table A.2 provides some descriptive frequencies of the measures at nominal week 24. In particular, this table displays the proportion of patients whose difference from baseline was less than or equal to –3 (corresponding to at least 3 unit improvement), less than or was less than or equal to –2 (corresponding to at least 2 unit improvement), less than or equal to –1, equal to 0, or greater than or equal to 1 (corresponding to at least 1 unit deterioration). Note that the groups labeled ≤ -1, $= 0$, or ≥ 1 partition the set of subjects at nominal week 24 (so the percentages add to 100%). Turning to that table one can see that, for example, for tactile roughness 41 (46%) LOCF subjects in the TEC-II treatment group showed no change over baseline in the ITT population versus 8 (16%) in the MITT population. The corresponding frequencies for the vehicle are 42 (47%) LOCF subjects in the ITT population versus 6 (13%) in the MITT population.

Again, only treatment differences in fine wrinkling were statistically significant. From Appendix table A.2 one can note that the difference in favor of treatment (in the MITT LOCF population) is due about 22% of the subjects in the tretinoin cream 0.02% group having a difference of –2 from baseline versus only 8% in the vehicle group. Similarly, 38% of the subjects in the tretinoin cream 0.02% group show a difference of –1 from baseline, versus 29% in the vehicle group.

Figure 1 in the appendix provides a plot of baseline and endpoint mean values for each response measure. Both the baseline value and the LOCF value are given for each endpoint. Note that our analysis is based on comparing the change from baseline between the TEC-II 0.02% group and it vehicle, i.e., comparing the differences between the two adjacent bars for each variable. For this study, these tend to be fairly small on the 10 point scales, but do generally favor the TEC-II treatment.

III.B. Protocol J89-025

All patients were Caucasian, with demographics summarized in Table A.3 of the appendix.

There was a highly statistically significant qualitative interaction between treatment and investigator for fine wrinkling in the J89-025 study. Using the F-ratio as a rough measure of effect size, the effect size of this interaction was about the same as treatment effect size. Estimated population marginal means, also called "least squares means," of the difference from baseline for fine wrinkling are displayed in the following layout:

Treatment \ Investigator	ID 747	ID 1690	ID 1980
Tretinoin Cream 0.02%	-.47	-.5	-1,6
Vehicle	-.7	-.2	- .9
Significance level*	.3345	.2151	.0027

*Of within investigator treatment difference

For investigator 747 the vehicle mean difference was less than (i.e. better than) the treatment mean difference from baseline. For the other two investigators the vehicle mean difference greater than (i.e., worse than) the treatment mean difference. However, when analyzing these as simple effects within each investigator, the differences between treatment and vehicle for investigator 747 were not statistically significantly different ($p \leq 0.3345$). The other investigators had mean vehicle differences greater than the corresponding treatment mean difference (and had statistically significant differences between treatment and vehicle). Thus, while descriptively the interaction appears to be qualitative, we would not reject the hypothesis that it was quantitative. It seems that a reasonable case can be made for treating this as an artifact of the experiment.

As before, the following Table 5. displays the mean change from baseline for each of the six endpoints: tactile roughness, fine wrinkling, coarse wrinkling, mottled hyperpigmentation, yellow-brown discoloration, and skin laxity. Again, the ITT population consists of all subjects dispensed treatment, while the "modified" intent to treat (MITT) group was formed for each endpoint by excluding those subjects in the ITT population with a 0 or 1 at baseline on the specified endpoint. Response measures are given at week 24 for all subjects with data at week 24, and for all ITT or MITT subjects using LOCF.

Table 5. Study J89-025: Differences From Baseline

	Population							
	ITT				MITT			
	Week 24		LOCF		Week 24		LOCF	
	Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle
Tactile Roughness								
Mean	-1.7	-1.3	-1.6	-1.3	-1.8	-1.4	-1.6	-1.4
Std Dev	1.3	1.5	1.3	1.5	1.3	1.4	1.3	1.4
n	82	86	90	90	79	83	87	87
p-value(ANOVA)	0.0553		0.1855		0.0426		0.1789	
p-value(Exact)	0.0618		0.2047		0.0436		0.1960	

Table 5. (cont.) Study J89-025: Differences From Baseline

	Population							
	ITT				MITT			
	Week 24		LOCF		Week 24		LOCF	
	Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle
Fine Wrinkling								
Mean	-0.9	-0.6	-0.9	-0.6	-0.9	-0.6	-0.9	-0.6
Std Dev	1.1	1.0	1.1	1.0	1.1	1.0	1.1	1.0
n	82	86	90	90	82	86	90	90
p-value(ANOVA)	0.0204		0.0571		0.0204		0.0571	
p-value(Exact)	0.0294		0.0727		0.0294		0.0727	
Coarse Wrinkling								
Mean	-0.5	-0.3	-0.5	-0.2	-0.5	-0.3	-0.5	-0.2
Std Dev	0.7	0.6	0.7	0.6	0.7	0.6	0.7	0.6
n	82	86	90	90	82	86	90	90
p-value(ANOVA)	0.0158		0.0201		0.0158		0.0201	
p-value(Exact)	0.0162		0.0266		0.0162		0.0266	
Mottled Hyperpigmentation								
Mean	-1.1	-0.4	-1.0	-0.4	-1.2	-0.4	-1.0	-0.4
Std Dev	0.9	0.7	1.0	0.7	0.9	0.7	1.0	0.7
n	82	86	90	90	79	84	87	88
p-value(ANOVA)	0.0001		0.0001		0.0001		0.0001	
p-value(Exact)	0.0001		0.0001		0.0001		0.0001	
Yellow-brown discoloration								
Mean	-0.9	-0.5	-0.8	-0.5	-0.9	-0.5	-0.8	-0.5
Std Dev	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
n	82	86	90	90	82	84	90	88
p-value(ANOVA)	0.0023		0.0073		0.0041		0.0127	
p-value(Exact)	0.0025		0.0091		0.0041		0.0140	
Laxity								
Mean	-0.5	-0.3	-0.5	-0.3	-0.5	-0.3	-0.5	-0.3
Std Dev	0.8	0.7	0.8	0.7	0.8	0.7	0.8	0.7
n	82	86	90	90	82	86	90	90
p-value(ANOVA)	0.0552		0.0802		0.0552		0.0802	
p-value(Exact)	0.0568		0.1005		0.0568		0.1005	

Note that prior to the adjustment for multiple endpoints, we would conclude that there were statistically significant differences between TEC-II and its vehicle in mottled hyperpigmentation ($p \leq 0.0001$), yellow brown discoloration ($p \leq 0.0127$), and coarse wrinkling ($p \leq 0.0201$). Using Holm's (1979) modified Bonferroni corrections we get the following table, text Table 6., giving multiplicity corrected significance levels for both the ITT and the MITT populations.

Table 6. Study J89-025: P-values adjusted for multiplicity

Holms p-values	J89-025	MITT original p-values	ITT original p-values	MITT adjusted p-values	ITT adjusted p-values
0.0084	Mott Hyperpig.	.0001 *	.0001 *	.0001*	.0001*
0.010	Yellow-brown Discoloration	.0127	.0073 *	.0634	.0366*
0.0125	Coarse Wrinkling.	.0201	.0201	.0805	.0805
0.0167	Fine Wrinkling.	.0571	.0571	.1712	.1712
0.0250	Laxity	.0802	.0802	.1712	.1712
0.0500	Tactile Roughness	.1789	.1855	.1789	.1855

* - denotes a statistically significant (at 0.05 level) comparison

Thus, adjusting for the multiple endpoints, in the MITT population only the comparison between treatment and vehicle for fine mottled hyperpigmentation is statistically significant at the .05 level. It is quite significant ($p \leq 0.0001$, in adjusted p-value). The corresponding comparison for yellow-brown discoloration is almost, but not quite, statistically significant ($p \leq 0.0634$, in adjusted p-value). But note that for this yellow-brown discoloration the differences between TEC-II 0.02% and its vehicle were statistically significant in the ITT population ($p \leq 0.0366$, in adjusted p-value). Whether or not that these results for yellow-brown discoloration are close enough to statistical significance in the MITT population to be of clinical significance is a decision for the Medical Officer.

Again, all subjects were Caucasian, and almost all were female, with restricted age ranges. So presumably the usual subgroup analyses would be superfluous.

Table A.4 in the appendix displays the numbers and percentages of subjects with a decrease from baseline of 3 or more, 2 or more, 1 or more, and those whose value at the specified endpoint equals the baseline (i.e., difference = 0), as well as those who show an increase over baseline (i.e., difference≥1).

From this table A.4, in the MITT population, for mottled hyperpigmenation, 29% of the tretinoin cream group showed a decrease of –2 or more. Only 11% in the vehicle group showed a decrease of –2. Similarly, 68% of the tretinoin cream group showed a decrease of –1 or more, versus only 30% in the vehicle group. For yellow-brown discoloration, 24% of the tretinoin cream group showed a decrease of –2 or more versus only 15% in the vehicle group. Similarly, 53% of the tretinoin cream group showed a decrease of –1 or more, versus 35% in the vehicle group.

Figure 2 in the appendix is a plot of baseline and endpoint mean values for each response measure. Both the baseline value and the LOCF value are given for each endpoint. When inspecting the bars it should be noted that our analysis is based on comparing the change from baseline between the TEC-II 0.02% group and it vehicle, i.e., comparing the differences between the two adjacent bars for each variable. These tend to be fairly small on the 10 point scales, but do generally favor the TEC-II treatment.

III. C. Protocol J89-045

All patients were Caucasian, with demographic characteristics described in table A.5 of the appendix.

The following text Table 7. shows the mean change from baseline for each of the six endpoints: tactile roughness, fine wrinkling, coarse wrinkling, mottled hyperpigmentation, yellow-brown discoloration, and skin laxity plus supporting statistics and tests of differences between treatment and vehicle. While several measures had relatively large investigator effects, there were no statistically significant treatment by investigator interactions at these time points.

Table 7. Study J89-045: Differences From Baseline

	Population							
	ITT				MITT			
	Week 24		LOCF		Week 24		LOCF	
	Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle
Tactile Roughness								
Mean	-1.1	-1.1	-1.0	-1.1	-1.3	-1.4	-1.3	-1.3
Std Dev	1.4	1.3	1.4	1.3	1.3	1.3	1.3	1.3
n	56	58	60	60	49	48	51	50
p-value(ANOVA)	0.8725		0.7246		0.7498		0.7153	
p-value(Exact)	0.8828		0.7768		0.9321		0.9340	
Fine Wrinkling								
Mean	-1.6	-0.6	-1.6	-0.6	-1.6	-0.6	-1.6	-0.6
Std Dev	1.2	1.0	1.2	1.0	1.2	1.0	1.2	1.0
n	56	58	60	60	56	58	60	60
p-value(ANOVA)	0.0001		0.0001		0.0001		0.0001	
p-value(Exact)	0.0001		0.0001		0.0001		0.0001	
Coarse Wrinkling								
Mean	-1.2	-0.7	-1.1	-0.7	-1.2	-0.7	-1.1	-0.7
Std Dev	1.2	1.0	1.2	1.0	1.2	1.0	1.2	1.0
n	56	58	60	60	56	58	60	60
p-value(ANOVA)	0.0262		0.0274		0.0262		0.0274	
p-value(Exact)	0.0260		0.0347		0.0260		0.0347	
Mottled Hyperpigmentation								
Mean	-1.9	-1.6	-1.8	-1.5	-2.0	-1.6	-1.9	-1.5
Std Dev	1.4	1.5	1.4	1.5	1.4	1.5	1.4	1.5
n	56	58	60	60	52	57	56	59
p-value(ANOVA)	0.1412		0.2394		0.0774		0.1494	
p-value(Exact)	0.1576		0.2659		0.0789		0.1535	
Yellow-brown discoloration								
Mean	-1.7	-0.8	-1.6	-0.8	-1.7	-0.8	-1.6	-0.8
Std Dev	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
n	56	58	60	60	56	58	60	60
p-value(ANOVA)	0.0004		0.0006		0.0004		0.0006	
p-value(Exact)	0.0004		0.0007		0.0004		0.0007	
Laxity								
Mean	-1.7	-1.0	-1.7	-1.0	-1.7	-1.0	-1.7	-1.0
Std Dev	1.6	1.3	1.6	1.3	1.6	1.3	1.6	1.3
n	56	58	60	60	56	58	60	60
p-value(ANOVA)	0.0129		0.0059		0.0129		0.0059	
p-value(Exact)	0.0144		0.0069		0.0144		0.0069	

While several of these are highly statistically significant, because there are a six possible endpoints, a correction for multiple endpoints is needed. Holm's (1979) modified Bonferroni corrections give the following Table 8. of corrected significance levels:

Table 8. Study J89-045: P-values adjusted for multiplicity

Holms p-values	J89-045	MITT original p-values	ITT original p-values	MITT adjusted p-values	ITT adjusted p-values
0.0084	Fine Wrinkling	.0001 *	.0001 *	.0001	.0001
0.010	Yellow-brown discoloration	.0006 *	.0006 *	.0029	.0029
0.0125	Laxity	.0059 *	.0059 *	.0235	.0235
0.0167	Coarse Wrinkling	.0274	.0274	.0821	.0821
0.0250	Mott Hyperpig.	.1494	.2394	.2987	.4789
0.0500	Tactile Roughness	.7153	.7246	.7153	.7246

* - denotes a statistically significant (at 0.05 level) comparison

Thus, adjusting for the multiple endpoints, the differences between TEC-II 0.02% emollient cream and its vehicle were statistically significant at the .05 level for fine wrinkling, yellow-brown discoloration, and skin laxity. ($p \leq 0.0001$, $p \leq 0.0029$, and $p \leq 0.0235$ respectively, in adjusted p-value).

Again, since most subjects were Caucasian and female, with restricted ages, no subgroup analysis was conducted.

For this study, appendix Table A.6 displays the numbers and percentages of subjects with a decrease from baseline of 3 or more, 2 or more, 1 or more, and those whose value at the specified endpoint equals the baseline (i.e., difference = 0), as well as those who show an increase over baseline (i.e., difference$\geq$1).

Note treatment differences for fine wrinkling, yellow-brown discoloration, and skin laxity were statistically significant. From the appendix table 3 we see that in the MITT population, for fine wrinkling, 27% of the TEC-II (tretinoin) 0.02% cream group showed a decrease of –3 or more. Only 3% in the vehicle group showed an equivalent decrease. Similarly, 43% and 83% of the tretinoin cream group showed a decrease of -2 or more, or –1 or more, versus only 20% and 45%, respectively, in the vehicle group. For yellow-brown discoloration, 28% of the TEC-II cream group showed a decrease of –3 or more versus 12% in the vehicle group. In the tretinoin cream 0.02% group, some 52% and 80% showed a decrease of -2 or more, or –1 or more, respectively. The corresponding proportions in the vehicle group were 25% and 50%, respectively. For skin laxity, 23% of the tretinoin cream group showed a decrease of –3 or more versus 15% in the vehicle group. In the tretinoin cream 0.02% group, some 48% and 80% showed a decrease of -2 or more, or –1 or more, respectively. The corresponding proportions in the vehicle group were 35% and 53%, respectively.

Figure 3. in the appendix provides a plot of baseline and endpoint mean values for each response measure.

III. D. Protocol L91-026

Unlike the other primary studies, this double-blind, randomized, multicenter, parallel group, vehicle controlled was conducted in patients with non-typical Caucasian type skin. The demographic characteristics of patients are given in Table A.7 of the appendix.

The six primary endpoints used in this study differed somewhat from those used in the other four studies. Tactile roughness, fine wrinkling, coarse wrinkling, and laxity were all measured on the 0-9 scale as before. However, instead of yellowing and simple hyperpigmentation, hyperpigmentation was assessed both locally (i.e., the presence of a ill defined patch in the zygoma area) and generally on the face or sun-exposed areas. The following table, Table 9., displays summary information on the mean change from baseline for each of these six primary endpoints. While several measures had relatively large investigator effects, there were no statistically significant treatment by investigator interactions at these time points.

Table 9. Study L91-026: Differences From Baseline

	Population							
	ITT				MITT			
	Week 24		LOCF		Week 24		LOCF	
	Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle
Tactile Roughness								
Mean	-0.7	-0.6	-0.6	-0.5	-1.0	-0.9	-1.0	-0.8
Std Dev	1.0	0.9	1.0	0.9	1.0	1.0	1.0	1.0
n	55	53	60	60	37	33	40	36
p-value(ANOVA)	0.7148		0.4557		0.7408		0.4768	
p-value(Exact)	0.7677		0.5111		0.8049		0.6351	
Fine Wrinkling								
Mean	-0.1	-0.5	-0.2	-0.4	-0.3	-0.6	-0.3	-0.5
Std Dev	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
n	55	53	60	60	40	45	45	52
p-value(ANOVA)	0.0481		0.0617		0.1509		0.1868	
p-value(Exact)	0.0591		0.0807		0.2210		0.2869	
Coarse Wrinkling								
Mean	-0.2	-0.2	-0.2	-0.2	-0.6	-0.5	-0.5	-0.5
Std Dev	1.0	0.9	0.9	0.8	1.2	1.1	1.1	1.1
n	55	53	60	60	26	26	28	28
p-value(ANOVA)	0.9537		0.9181		0.9911		0.8767	
p-value(Exact)	1.000		1.000		1.000		1.000	
Local Hyperpigmentation								
Mean	-0.1	-0.2	-0.2	-0.1	-0.5	-0.7	-0.6	-0.6
Std Dev	1.0	1.1	1.0	1.1	0.8	1.1	0.8	1.0
n	55	53	60	60	35	27	38	31
p-value(ANOVA)	0.9589		0.7895		0.7470		0.8539	
p-value(Exact)	1.000		0.8580		0.7820		0.8933	
General Hyperpigmentation								
Mean	-0.2	-0.3	-0.2	-0.2	-0.6	-0.7	-0.6	-0.6
Std Dev	0.7	0.9	0.7	0.8	0.8	1.1	0.7	1.0
n	55	53	60	60	27	26	29	30
p-value(ANOVA)	0.7891		0.9045		0.9556		0.7387	
p-value(Exact)	0.8093		1.000		1.000		0.9937	

Table 9. (cont.) Study L91-026: Differences From Baseline

	Population							
	ITT				MITT			
	Week 24		LOCF		Week 24		LOCF	
	Treat-ment	Veh-icle	Treat-ment	Veh-icle	Treat-ment	Veh-icle	Treat-ment	Veh-icle
Laxity								
Mean	-0.6	-0.3	-0.5	-0.3	-0.8	-0.5	-0.8	-0.4
Std Dev	1.0	0.9	1.0	0.9	1.2	1.0	1.2	0.9
n	55	53	60	60	39	37	41	42
p-value(ANOVA)	0.1192		0.1303		0.1054		0.0669	
p-value(Exact)	0.1234		0.1544		0.1756		0.1223	

Again, because there are a six possible endpoints, a correction for multiple endpoints is needed. Using Holm's (1979) modified Bonferroni corrections we get the following table.

Table 10. Study L91-026: P-values adjusted for multiplicity

Holms p-values	L91-026	MITT original p-values	MITT adjusted p-values	L91-026	ITT original p-values	ITT adjusted p-values
0.0084	Laxity	.0669	.6322	Fine Wrinkling	.0617	.3702
0.010	Fine Wrinkling	.1868	.7545	Laxity	.1303	.6513
0.0125	Tactile Roughness	.4768	1.00	Tactile Roughness	.4557	1.00
0.0167	General Mott. Hyperpig.	.7387	1.00	Local Mott. Hyperpig.	.7895	1.00
0.0250	Local Mott. Hyperpig.	.8539	1.00	General Mott. Hyperpig.	.9045	1.00
0.0500	Coarse Wrinkling	.8767	1.00	Coarse Wrinkling	.9181	1.00

Even prior to adjusting for multiplicity, there were no statistically significant differences associated with treatment in any of the endpoints.

Almost all subjects were Black, and mostly female, so again, the usual subgroup analyses were felt to be superfluous.

Appendix Table A.8 displays the numbers and percentages of subjects with a decrease from baseline of 3 or more, 2 or more, 1 or more, etc. If one inspects this table, it is apparent that for all endpoints the distributions seem to be roughly equivalent. This is quite consistent with the observation above that, even without adjusting for multiplicity, no treatment differences were statistically significant.

Figure 4 in the appendix shows a plot of baseline and endpoint mean values for each response measure. Both the baseline value and the LOCF value are given for each endpoint. Note that our analysis is based on comparing the change from baseline between the TEC-II 0.02% group and it vehicle, i.e., comparing the differences between the two adjacent bars for each variable. Unlike the charts in the previous figures, these do not seem to particularly favor the TEC-II treatment.

IV. Supporting Study: Protocol K90-011

This was a double-blind, randomized, single center, parallel group, vehicle controlled study of the safety and efficacy of Tretinoin Emollient Cream (TEC-II) 0.02% in the treatment of signs and symptoms associated with photodamaged skin. Treatment was to be applied once nightly for 24 weeks, with a general dosing guideline of 0.25 g per application. Return visits were scheduled two and four weeks after starting in the study, and every four weeks there after until the subject completed the study.

Since this was only a single center study, results from this study were only considered to be potentially supportive of any outcomes from the other four studies, and not the basis of any claim on its own.

All patients were Caucasian, with demographic characteristics summarized in appendix Table A.9.

The following Table 11. displays the mean change from baseline for each of the six endpoints: tactile roughness, fine wrinkling, coarse wrinkling, mottled hyperpigmentation, yellow-brown discoloration, and skin laxity plus supporting statistics and tests of differences between treatment and vehicle. Note that in this study no subjects had scores of 0 or 1 at baseline on any of these endpoints, and thus the MITT and ITT populations were coincident.

Table 11. Study K90-011: Differences From Baseline

	ITT/MITT			
	Week 24		LOCF	
	Treat-ment	Veh-icle	Treat-ment	Veh-icle
Tactile Roughness				
Mean	-1.0	-0.9	-1.0	-0.9
Std Dev	1.0	0.9	1.0	0.9
n	36	35	40	40
p-value(ANOVA)	0.7089		0.7293	
p-value(Exact)	0.7111		0.8169	
Fine Wrinkling				
Mean	-0.4	-0.2	-0.4	-0.2
Std Dev	0.9	0.7	0.9	0.7
n	36	35	40	40
p-value(ANOVA)	0.2643		0.2515	
p-value(Exact)	0.3090		0.3165	
Coarse Wrinkling				
Mean	-0.2	0.1	-0.2	0.2
Std Dev	0.6	0.6	0.6	0.6
n	36	35	40	40
p-value(ANOVA)	0.0089		0.0049	
p-value(Exact)	0.0151		0.0085	

APPEARS THIS WAY ON ORIGINAL

Table 11. (cont.) Study K90-011: Differences From Baseline

	ITT/MITT Week 24		LOCF	
	Treat-ment	Veh-icle	Treat-ment	Veh-icle
Mottled Hyperpigmentation				
Mean	-0.8	-0.6	-0.7	-0.6
Std Dev	0.8	1.1	0.9	1.1
n	36	35	40	40
p-value(ANOVA)	0.4421		0.6529	
p-value(Exact)	0.4708		0.7370	
Yellow-brown discoloration				
Mean	-0.5	-0.4	-0.5	-0.4
Std Dev	0.8	0.8	0.8	0.8
n	36	35	40	40
p-value(ANOVA)	0.7182		0.5098	
p-value(Exact)	0.7758		0.5971	
Laxity				
Mean	-0.4	-0.2	-0.4	-0.2
Std Dev	0.8	0.9	0.8	0.9
n	36	35	40	40
p-value(ANOVA)	0.2369		0.1466	
p-value(Exact)	0.2760		0.1861	

APPEARS THIS WAY ON ORIGINAL

Again, when we adjust results by use Holm's (1979) corrections we get the following text Table 12.

Table 12. Study K90-011: P-values adjusted for multiplicity

Holms p-values	K90-011	Original p-values	Holm's adjusted p-value
0.0084	Coarse Wrinkling	.0049 *	.0295 *
0.010	Laxity	.1466	.7330
0.0125	Fine Wrinkling	.2515	1.0
0.0167	Yellow-brown discoloration	.5098	1.0
0.0250	Mott Hyperpig.	.6529	1.0
0.0500	Tactile Roughness	.7293	1.0

* - denotes a statistically significant (at 0.05 level) comparison

Thus, adjusting for the multiple endpoints, only the comparison for coarse wrinkling is statistically significant at the .05 level (adjusted p-value: $p \leq 0.0295$). Again, all subjects were Caucasian, and almost all were female. So it was felt that the usual subgroup analyses was not needed. Appendix table A.10 displays the numbers and percentages of subjects with a decrease from baseline of 3 or more, 2 or more, etc. From this table it seems that the significant difference between TEC-II and vehicle is associated with the fact that for coarse wrinkling, 28% of the LOCF subjects in the TEC-II treatment group showed a change of –1 over baseline versus 10% in the vehicle group.

Figure 5 in the appendix is a plot of baseline and endpoint mean values for each response measure.

V. Overall Efficacy

After redefining our study categories, we had four primary studies: two U.S. randomized, multi-center, double-blind studies, a multicenter study in northern Europe, and a multicenter center U.S. study limited to patients with (darker) skin type III or above (J89-024, J89-025, J89-045, and L91-026 respectively). Because study K90-011 was a single center study, it was felt that it could only be used to support indications in two of the other four multi-center studies. For each indication primary emphasis was to be placed on the MITT population, i.e. patients whose baseline score on the endpoint was greater than 1.

The original protocols for the primary studies specified that the endpoints for the signs and symptoms would be the change from baseline, analyzed with an analysis of variance with treatment (versus vehicle), investigator, and interaction as factors. So, despite some preference for a randomization test, the ANOVA tests were used. Within each study, it was proposed to adjust for the multiple comparison's using Holm's modified Bonferroni adjustment (starting with the smallest p-value). The following Table 13. gives the original p-values and the Holm's modified Bonferroni limits to which to compare the original p-values (to get a 0.05 family-wise significance level, within each study[1]).

Table 13. MITT populations: Holm's p-values to compare to original p-values.

Holms p-values	Study J89-024		Study J89-025		Study J89-045	
0.0084	Fine Wrinkling	.0017 *	Mott Hyperpig .	.0001 *	Fine Wrinkling	.0001 *
0.010	Coarse Wrinkling	.0547	Yellow-brown discoloration	.0127	Yellow-brown discoloration	.0006 *
0.0125	Yellow-brown discoloration	.0663	Coarse Wrinkling.	.0201	Laxity	.0059 *
0.0167	Mott Hyperpig.	.1741	Fine Wrinkling.	.0571	Coarse Wrinkling	.0274
0.0250	Laxity	.1834	Laxity	.0803	Mott Hyperpig.	.1494
0.0500	Tactile Roughness	.4916	Tactile Roughness	.1789	Tactile Roughness	.7153

Holms p-values	Study L91-026		Study K90-011	
0.0084	Fine Wrinkling	.0617	Coarse Wrinkling	.0049 *
0.010	Laxity	.1303	Laxity	.1466
0.0125	Tactile Roughness	.4557	Fine Wrinkling	.2515
0.0167	L. Mott. Hyperpig.	.8464	Yellow-brown discoloration	.5098
0.0250	G. Mott. Hyperpig.	.9045	Mott Hyperpig.	.6529
0.0500	Coarse Wrinkling	.9181	Tactile Roughness	.7293

* - denotes a statistically significant (at 0.05 level) comparison.

The usual interpretation of the requirements for efficacious studies is that we need at least two studies with significant results to justify a claim of efficacy. Adjusting for the multiplicity of outcomes only the difference in fine wrinkling between treatment and vehicle is statistically

[1] Note that adjustment across studies is not applied. As discussed in the section on statistical methods, this issue being investigated at the Division of Biometrics.

significant in two studies, namely J89-024 and J89-045. However, results for yellow-brown discoloration are nearly statistically significant at the 0.05 level in the J89-025 study (compare the observed p-value of 0.0127 to the Holm's p-value of 0.010), and are statistically significant in the J89-045 study. Whether this is close enough to clinical significance is a decision for the Medical Officer. Both differences are statistically significant in the ITT population (See text Table 15. below).

The significance levels in the table above are those from the original tests, unadjusted for multiplicity. The adjustment is applied when these are compared to the Holm's p-values to see if the observed significance level is small enough to be declared significant at a 0.05 level. For some purposes an adjusted significance level analogous to the observed significance in a single test would be useful. Suppose there are K endpoints. For the kth test, the observed significance level is best represented as the maximum of k times the observed significance level of the kth test, k+1 times the observed level of the (k+1)st test, up to K times the observed level of the Kth test. This allows a simple comparison to any potential family-wise error rate. The following Table 14. provides these adjusted p-values. Thus, for example, the observed statistical significance of the difference between treatment and vehicle in the J89-045 study can be assessed as 0.0634. However, such adjusted significance levels do seem to be more complicated to relate back to the original tests.

Table 14. MITT populations: p-values adjusted for multiplicity.

Study J89-024		Study J89-025		Study J89-045	
Fine Wrinkling	.0099 *	Mott Hyperpig .	.0001 *	Fine Wrinkling	.0001 *
Coarse Wrinkling	.2734	Yellow-brown discoloration	.0634	Yellow-brown discoloration	.0029 *
Yellow-brown discoloration	.2734	Coarse Wrinkling.	.0805	Laxity	.0235 *
Mott Hyperpig.	.5223	Fine Wrinkling.	.1712	Coarse Wrinkling	.0821
Laxity	.5223	Laxity	.1712	Mott Hyperpig.	.2987
Tactile Roughness	.5223	Tactile Roughness	.1789	Tactile Roughness	.7153

Study L91-026		Study K90-011	
Laxity	.4015	Coarse Wrinkling	.0295 *
Fine Wrinkling	.9341	Laxity	.7330
Tactile Roughness	1.0	Fine Wrinkling	1.0
G. Mott. Hyperpig.	1.0	Yellow-brown discoloration	1.0
L. Mott. Hyperpig.	1.0	Mott Hyperpig.	1.0
Coarse Wrinkling	1.0	Tactile Roughness	1.0

* - denotes a statistically significant (at 0.05 level) comparison

Of course, at a 0.05 familywise error, any conclusions are identical to those associated with the preceding Table 13. For comparison with these two tables, the overall results using the ITT population are given in the text Table 15. following:

Table 15. ITT populations: Holm's p-values to compare to original p-values.

Holms p-values	Study J89-024		Study J89-025		Study J89-045	
0.0084	Fine Wrinkling	.0021 *	Mott Hyperpig .	.0001 *	Fine Wrinkling	.0001 *
0.010	Coarse Wrinkling	.0547	Yellow-brown discoloration	.0073 *	Yellow-brown discoloration	.0006 *
0.0125	Yellow-brown discoloration	.0952	Fine Wrinkling.	.0194	Laxity	.0059 *
0.0167	Mott Hyperpig.	.2041	Coarse Wrinkling.	.0201	Coarse Wrinkling	.0274
0.0250	Laxity	.3821	Laxity	.1661	Mott Hyperpig.	.2394
0.0500	Tactile Roughness	.6716	Tactile Roughness	.1855	Tactile Roughness	.5484

Holms p-values	Study L91-026		Study K90-011	
0.0084	Fine Wrinkling	.0617	Coarse Wrinkling	.0049 *
0.010	Laxity	.1303	Laxity	.1466
0.0125	Tactile Roughness	.4557	Fine Wrinkling	.2515
0.0167	L. Mott. Hyperpig.	.8464	Yellow-brown discoloration	.5098
0.0250	G. Mott. Hyperpig.	.9045	Mott Hyperpig.	.6529
0.0500	Coarse Wrinkling	.9181	Tactile Roughness	.7293

* - denotes a statistically significant (at 0.05 level) comparison

Adjusting for the multiplicity of outcomes the difference in fine wrinkling between treatment and vehicle is statistically significant at a 0.05 level in two studies, J89-024 and J89-045 (i.e. both .0021 and .0001 are less than 0.0084), as in the MITT population. However in this ITT population results for yellow-brown discoloration are statistically significant in both the J89-025 and J89-045 studies (since in both studies, the observed p-value is less than 0.01).

APPEARS THIS WAY
ON ORIGINAL

VI. Adverse Events:

In each of the five studies emphasized here, measurements were also made on five signs and symptoms of skin irritation, in particular: erythema, peeling, itching, dryness, and burning/stinging. These were measured on the same 0 to 9 scale used with the primary endpoints:

0 = none 4-6 = moderate
1-3 = mild 7-9 = severe

Table 16. below summarizes baseline original values and the maximum change from baseline over all measured time points for each of these signs and symptoms of skin irritation, pooled over the five studies. When reading this table, note that the two left columns give the baseline distribution of severity for each of the five adverse conditions. Thus, for example, at baseline 268 subjects in the pooled treatment groups scored 0 on erythema and 264 in the pooled vehicle groups scored 0 on erythema. The second set of scores indicate the distribution of the maximum change from the baseline score. If this maximum change is negative (< 0) we can say that at all measured time points after baseline, the observed score was less than the baseline score, i.e., shows improvement over baseline on that sign or symptom. Thus 8 subjects in the treatment group and 16 in the vehicle group showed improvement over baseline in erythema at all observed time points. The other score groups (i.e. 1+, 2+, etc.) give the frequencies of those whose score was at least one over baseline, at least two over baseline, at least three over baseline, at least four over baseline, and finally at least five over baseline. Thus in the TEC-II 0.02% treatment group 207 patients had an erythema score that was at least one unit greater than baseline, 151 had scores at least two units greater than baseline, 97 had scores at least three units greater than baseline, etc. A Fisher Exact test with significance adjusted for the 30 comparisons by a Bonferroni correction is used to test the hypothesis that the distributions of those with a score at least k units greater than baseline is the same for the treatment and vehicle groups (for k=1,2,3,4,5).

Note that the Fisher Exact test was used since some of these tables are extremely sparse, and the stratified Cochran-Mantel-Haenszel tests would not be appropriate. Of course, other approaches could have been used.

Table 16. Signs and Symptoms of Skin Irritation: Baseline Scores and Maximum Change from Baseline

Observed Value	Number having value at baseline: Treatment N	%	Vehicle N	%	Observed Maximum Difference	Number with maximum difference from baseline by specified value: Treatment N	%	Vehicle N	%	p-value for Fisher Exact Test
Erythema										
=0	268	0.788	264	0.776	<0	8	0.024	16	0.048	0.144
=1	34	0.100	32	0.094	>0 (i.e., 1+)	207	0.620	123	0.367	0.0001 ***
=2	8	0.024	18	0.053	>1 (i.e., 2+)	151	0.452	55	0.164	0.0001 ***
=3	18	0.053	18	0.053	>2 (i.e., 3+)	97	0.290	15	0.045	0.0001 ***
=4	3	0.009	4	0.012	>3 (i.e., 4+)	53	0.159	2	0.006	0.0001 ***
≥5	9	0.026	4	0.012	>4 (i.e., 5+)	23	0.069	0	0.000	0.0001 ***
ALL	340		340			334		335		

Table 16. (cont.) Signs and Symptoms of Skin Irritation: Baseline Scores and Maximum Change from Baseline

Peeling											
=0	321	0.944	325	0.956	0<	3	0.009	5	0.015	0.725	
=1	12	0.035	7	0.021	>0 (i.e., 1+)	196	0.587	69	0.206	0.0001	***
=2	3	0.009	7	0.021	>1 (i.e., 2+)	148	0.443	28	0.084	0.0001	***
=3	4	0.012	1	0.003	>2 (i.e., 3+)	101	0.302	11	0.033	0.0001	***
=4	0	0.000	0	0.000	>3 (i.e., 4+)	61	0.183	5	0.015	0.0001	***
≥5	0	0.000	0	0.000	>4 (i.e., 5+)	29	0.087	1	0.003	0.0001	***
ALL	340		340			334		335			
Itching											
=0	328	0.965	333	0.979	0<	5	0.015	4	0.012	0.752	
=1	7	0.021	5	0.015	>0 (i.e., 1+)	147	0.440	55	0.164	0.021	*
=2	2	0.006	1	0.003	>1 (i.e., 2+)	108	0.323	37	0.110	0.0001	***
=3	2	0.006	1	0.003	>2 (i.e., 3+)	58	0.174	13	0.039	0.0001	***
=4	0	0.000	0	0.000	>3 (i.e., 4+)	35	0.105	6	0.018	0.0001	***
≥5	1	0.003	0	0.000	>4 (i.e., 5+)	18	0.054	3	0.009	0.0001	***
ALL	340		340			334		335			
Dry Skin											
=0	292	0.859	297	0.874	0<	9	0.027	13	0.039	0.516	
=1	22	0.065	21	0.062	>0 (i.e., 1+)	206	0.617	108	0.322	0.0001	***
=2	14	0.041	16	0.047	>1 (i.e., 2+)	159	0.476	58	0.173	0.0001	***
=3	10	0.029	4	0.012	>2 (i.e., 3+)	97	0.290	27	0.081	0.0001	***
=4	0	0.000	0	0.000	>3 (i.e., 4+)	61	0.183	10	0.030	0.0001	***
≥5	2	0.006	2	0.006	>4 (i.e., 5+)	28	0.084	3	0.009	0.0001	***
ALL	340		340			334		335			
Burning/Stinging											
=0	332	0.976	335	0.985	0<	1	0.003	3	0.009	0.624	
=1	4	0.012	1	0.003	>0 (i.e., 1+)	222	0.665	81	0.242	0.0001	***
=2	1	0.003	1	0.003	>1 (i.e., 2+)	164	0.491	44	0.131	0.0001	***
=3	1	0.003	3	0.009	>2 (i.e., 3+)	99	0.296	16	0.048	0.0001	***
=4	0	0.000	0	0.000	>3 (i.e., 4+)	68	0.204	7	0.021	0.0001	***
≥5	2	0.006	0	0.000	>4 (i.e., 5+)	45	0.135	3	0.009	0.0001	***
ALL	340		340			334		335			

* - Denotes statistically significant at the .05 level after Bonferroni adjustment (for 30 comparisons).
*** - Denotes statistically significant at the .001 level after Bonferroni adjustment (for 30 comparisons).

Thus we would estimate that during the study roughly 45% of the subjects would experience a 2-unit increase in erythema in the Tec-II 0.02% treatment group versus some 16% using the vehicle alone. Some 16% of the Tec-II treatment group would experience a 4-unit increase versus 1% using vehicle. Similarly, we would estimate that during the study roughly 44% of the subjects would experience a 2-unit increase over baseline in peeling in the Tec-II treatment group versus some 8% using the vehicle alone. Some 18% of the Tec-II treatment group would experience a 4-unit increase versus 2% using vehicle. Roughly 32% of the subjects experienced a 2-unit increase over baseline in itching in the Tec-II treatment group versus some 11% using the vehicle alone. About 10% of the Tec-II treatment group

experienced a 4-unit increase versus 2% using vehicle. Approximately 48% of the subjects in the Tec-II group experienced a 2-unit increase over baseline in dryness versus some 17% using the vehicle alone. About 18% of this Tec-II treatment group experienced a 4-unit increase versus 3% using vehicle. For burning/stinging roughly 49% of the subjects reported a 2-unit increase over baseline in the Tec-II treatment group versus some 13% using the vehicle alone. Roughly 20% of the Tec-II treatment group experienced a 4-unit increase versus 2% using vehicle. Even adjusting for the (30) multiple comparisons these differences were all highly statistically significant ($p \leq 0.0001$ for all comparisons discussed here).

Appendix Table A.11 gives more detailed summaries of the distributions of the original signs and symptoms of skin irritation cited above. The tables above again indicate that each of erythema, peeling, itching, dryness, and burning/stinging do initially get worse with both TEC-II 0.02% cream and with vehicle, though clearly worse with the former than the latter. But these skin conditions do tend to improve after 4-8 weeks of treatment.

The sponsor provided tables of other adverse events during the study. The Medical Officer felt that no detailed analysis of these events was necessary. However, it was felt that a multiplicity adjusted test of differences between TEC-II 0.02% Cream and its vehicle in the various adverse event might be useful. These analyses are based on the pooled adverse event data from the five efficacy studies cited above.

To test the statistical significance of any differences in reported adverse events between TEC-II 0.02% Cream and its vehicle, the adverse events were first screened for those with five or more subjects experiencing the event. The number five was arbitrary, but reduces the number of adjustments required, and hence should increase power in the tests adjusted for multiplicity. Thirty-three adverse events met this criterion in the pooled data set. Note that only the following comparisons were close to statistically significant (prior to adjusting for multiplicity of tests):

AE Code	Description	Incidence TEC-II	Incidence Vehicle	Unadjusted p-value	Adjusted p-value
1201101212	Facial Dryness	18/340	4/340	0.0038	0.0276
1201101411	Peeling	11/340	3/340	0.0549	0.4996.
1201200012	Erythema	15/340	6/340	0.0738	0.8156
1201200012	Facial Irritation	50/340	12/340	0.0001	0.0001

The unadjusted p-value is the p-value from a Fisher Exact test of differences between TEC-II 0.02% and its vehicle. All other unadjusted p-values were greater than .15. Adjusting the tests for this multiplicity of comparisons using the techniques of Westfall and Young (1993) gives the "Adjusted p-value" cited above. In this particular case the adjustments were done using by sampling 5000 replicates from the permutation distribution of each table. These are used to approximate the distribution of the minimum p-value of all the tests. Unlike most other methods for correcting for multiplicity, features of the distribution and inter-test correlations are incorporated into the analysis.

Thus, we would conclude that as reported adverse events, facial irritation and dryness are statistically significantly worse in the TEC-II 0.02% group than in the vehicle group.

REFERENCES:

Fisher, R.A. (1935) *The Design of Experiments,* London: Oliver & Boyd.

Holm, S. (1979) A simple sequentially rejective multiple test procedure, *Scandinavian Journal of Statistics,* **6**, 655-660.

Westfall, P.H. and Young, S.S. (1993) *Resampling-Based Multiple Testing: Examples and Methods for p-value Adjustment,* New York: John Wiley & Sons, Inc.

APPEARS THIS WAY
ON ORIGINAL

Conclusions (Which may be conveyed to the Sponsor):

1. In the original submission of NDA 21-108, RENOVA 0.02% (tretinoin emollient cream), formulation TEC-II 0.02%, the sponsor originally claimed the indication of reducing the general signs and symptoms of photoaging. Concurrently, six general signs and symptoms of such damage were assessed: tactile roughness, fine wrinkling, coarse wrinkling, mottled hyperpigmentation, yellow-brown discoloration (labelled as yellowing by the sponsor), and skin laxity. Each of these latter six endpoints were scored by each investigator on a 10-point scale (0-9 with small numbers being more favorable). Photographs were provided to normalize the scale.

2. It was the opinion of the Medical Officer that the various signs and symptoms of "photoaging" were manifested through a variety of possibly separate and possibly obscure biological processes that though linked, did not constitute a single process. Thus, instead of the single process of photoaging, each of the six signs and symptoms noted above was chosen as a separate endpoint. It was decided to address the issue of multiplicity of outcomes using Holm's stepwise modification of the Bonferroni corrections (see the statistical methods section for a brief discussion of these).

3. Results from five studies provided the primary support for results, two multicenter studies among U.S. Caucasian patients, one Northern European multicenter study among Caucasian patients, one single center study among U.S. Caucasian patients, and a multicenter study among U.S. non-Caucaian patients. The sponsor proposes to market TEC-II 0.02% with a fragrance. However only the last study used this formulation. The other studies used the same formulation, but without the fragrance. Whether this is of import is a decision for the Medical Officer.

4. The usual interpretation of the requirements for efficacious studies is that we need at least two studies with significant results to justify a claim of efficacy. In the MITT population adjusting for the multiplicity of outcomes the difference in fine wrinkling between treatment and vehicle is statistically significant in two studies, namely J89-024 and J89-045 ($p \leq .0099$ and $p \leq .0001$, respectively, using Holm's adjusted p-values). Whether this is close enough to clinical significance is a decision for the Medical Officer. Note that both differences are statistically significant in the ITT population ($p \leq .0366$ and $p \leq .0029$). Again, all reported p-values are adjusted for multiplicity using Holm's procedure.

5. Information on adverse events was also collected. For most of these there were no statistically significant differences between the TEC-II 0.02% group and its vehicle. However, even correcting for the multiplicity of performed tests, facial irritation and dryness are statistically significantly worse in the TEC-II 0.02% group than in the vehicle group.

6. In addition, in each of the five studies emphasized here, measurements were also made on five signs and symptoms of skin irritation, in particular: erythema, peeling, itching, dryness, and burning/stinging. Defining a failure as having an increase of at least one unit over baseline (at any time during the study), at least two over baseline, at least three over baseline, at least

four over baseline, and finally at least five over baseline, we can compare the proportions of these failures in the TEC-II 0.02% treatment group versus its vehicle for each of these safety endpoints. Even correcting for the 30 comparisons using simple Bonferroni corrections all differences are extremely highly statistically significant. Thus there is strong evidence that TEC-II 0.02% use in associated with more erythema, peeling, itching, dryness, and burning/stinging than its corresponding vehicle.

7. Provided the difference in formulations can be ignored, using the rule that two statistically significant studies are needed, this would seem to be sufficient to conclude that there is a statistically significant difference between TEC-II 0.02% and its vehicle in terms of fine wrinkling. [redacted] differences are statistically significant in one study, and close to statistically significant in another study. Whether this is sufficient is a decision for the Medical Officer. Again, there is strong evidence that TEC-II 0.02% use in associated with more erythema, peeling, itching, dryness, and burning/stinging than its corresponding vehicle.

APPEARS THIS WAY
ON ORIGINAL

[/S/] 08/07/00

Steve Thomson
Mathematical Statistician, Biometrics III

[/S/] 08/11/00

concur: **Mohamed Alosh, Ph.D.**
Team Leader, Biometrics III

APPEARS THIS WAY
ON ORIGINAL

cc:
Archival NDA: 21-108
HFD-540/Division File
HFD-540/Dr. Wilkin
HFD-540/Dr. Okun
HFD-540/Dr. Luke
HFD-540/Ms. Cintron
HFD-725/Dr. Huque
HFD-725/Dr. Alosh
HFD-725/Mr. Thomson
HFD-340/Dr. Lepay
This review has 33 pages, plus a 14 page appendix, and this signature page.
Chron.

\Thomson\MyDocuments\x 7-2044\July 31, 2000\c:\mydocument\n21108renova.doc

Appendix Table A.1: Study J89-024: Demographics

	TEC –II 0.02%	Vehicle
No. Enrolled	90	90
No. Completed	77	83
No. Discontinued:	13	7
Adverse Event	4	0
Personal	7	3
Loss to Follow-Up	2	4
Mean Age (Range)	58.5 (45-69)	58.5 (45-69)
No. Male/No. Female	12 / 78	9 / 81

Appendix Table A.2: Study J89-024: Differences From Baseline

The following table displays the numbers and percentages of subjects with a decrease from baseline of 3 or more, 2 or more, 1 or more, and those whose value at the specified endpoint equals the baseline (i.e., difference = 0), as well as those who show an increase over baseline (i.e., difference≥1).

		ITT				MITT			
		Week 24		LOCF		Week 24		LOCF	
		Treat-ment	Veh-icle	Treat-ment	Veh-icle	Treat-ment	Veh-icle	Treat-ment	Veh-icle
Tactile Roughness									
difference ≤ -3	Count	7	8	7	8	7	8	7	8
	%	9	10	8	9	16	17	14	17
difference ≤ -2	Count	25	23	25	23	25	23	25	23
	%	32	28	28	26	58	50	51	48
difference ≤ -1	Count	42	43	45	45	36	39	38	40
	%	55	52	50	50	84	85	78	83
difference = 0	Count	32	37	41	42	4	5	8	6
	%	42	45	46	47	9	11	16	13
difference ≥ 1	Count	3	3	4	3	3	2	3	2
	%	4	4	4	3	7	4	6	4
Fine Wrinkling									
difference ≤ -3	Count	0	2	0	2	0	2	0	2
	%	0	2	0	2	0	2	0	2
difference ≤ -2	Count	19	7	20	7	19	7	20	7
	%	25	8	22	8	25	8	22	8
difference ≤ -1	Count	51	31	53	33	51	31	53	33
	%	66	37	59	37	67	37	60	37
difference = 0	Count	26	52	37	57	25	52	36	57
	%	34	63	41	63	33	63	40	63
difference ≥ 1	Count	0	0	0	0	0	0	0	0
	%	0	0	0	0	0	0	0	0

Appendix Table A.2: (cont). Study J89-024: Differences From Baseline

		ITT				MITT			
		Week 24		LOCF		Week 24		LOCF	
		Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle
Coarse Wrinkling									
difference ≤ -3	Count	1	1	2	1	1	1	2	1
	%	1	1	2	1	1	1	2	1
difference ≤ -2	Count	6	4	7	4	6	4	7	4
	%	8	5	8	4	8	5	8	4
difference ≤ -1	Count	33	21	34	22	33	21	34	22
	%	43	25	38	24	43	25	38	24
difference = 0	Count	44	62	56	68	44	62	56	68
	%	57	75	62	76	57	75	62	76
difference ≥ 1	Count	0	0	0	0	0	0	0	0
	%	0	0	0	0	0	0	0	0
Mottled Hyperpigmentation									
difference ≤ -3	Count	8	7	8	7	8	7	8	7
	%	10	8	9	8	11	9	10	8
difference ≤ -2	Count	32	21	32	21	32	21	32	21
	%	42	25	36	23	45	27	38	25
difference ≤ -1	Count	56	51	58	51	55	51	57	51
	%	73	61	64	57	77	65	68	60
difference = 0	Count	20	32	31	39	15	27	26	34
	%	26	39	34	43	21	35	31	40
difference ≥ 1	Count	1	0	1	0	1	0	1	0
	%	1	0	1	0	1	0	1	0
Yellow-brown Discoloration									
difference ≤ -3	Count	9	2	9	2	9	2	9	2
	%	12	2	10	2	19	4	16	3
difference ≤ -2	Count	26	15	26	15	26	15	26	15
	%	34	18	29	17	54	27	46	25
difference ≤ -1	Count	39	40	41	41	37	38	38	39
	%	51	48	46	46	77	69	67	66
difference = 0	Count	38	43	49	49	11	17	19	20
	%	49	52	54	54	23	31	33	34
difference ≥ 1	Count	0	0	0	0	0	0	0	0
	%	0	0	0	0	0	0	0	0
Laxity									
difference ≤ -3	Count	2	0	3	0	2	0	3	0
	%	3	0	3	0	3	0	4	0
difference ≤ -2	Count	9	6	10	6	9	6	10	6
	%	12	7	11	7	13	8	12	7
difference ≤ -1	Count	27	30	30	31	27	30	30	31
	%	35	36	33	34	40	38	37	36
difference = 0	Count	50	52	60	58	41	47	51	53
	%	65	63	67	64	60	60	63	62
difference ≥ 1	Count	0	1	0	1	0	1	0	1
	%	0	1	0	1	0	1	0	1

Appendix Table A.3: Study J89-025: Patient Demographics

	TEC –II 0.02%	Vehicle
No. Enrolled	90	90
No. Completed	82	86
No. Discontinued:	8	4
Adverse Event	2	1
Personal	1	1
Loss to Follow-Up	5	2
Mean Age (Range)	58.6 (45-70)	58.5 (43-70)
No. Male/No. Female	10 / 80	10 / 80

Appendix Table A.4: Study J89-025: Differences From Baseline

		ITT				MITT			
		Week 24		LOCF		Week 24		LOCF	
		Treat-ment	Veh-icle	Treat-ment	Veh-icle	Treat-ment	Veh-icle	Treat-ment	Veh-icle
Tactile Roughness									
difference ≤ -3	Count	14	15	14	15	14	15	14	15
	%	17	17	16	17	18	18	16	17
difference ≤ -2	Count	43	31	44	31	43	31	44	31
	%	52	36	49	34	54	37	51	36
difference ≤ -1	Count	70	61	72	63	70	61	72	63
	%	85	71	80	70	89	73	83	72
difference = 0	Count	11	22	16	24	9	20	14	22
	%	13	26	18	27	11	24	16	25
difference ≥ 1	Count	1	3	2	3	0	2	1	2
	%	1	3	2	3	0	2	1	2
Fine Wrinkling									
difference ≤ -3	Count	10	4	10	4	10	4	10	4
	%	12	5	11	4	12	5	11	4
difference ≤ -2	Count	16	13	17	13	16	13	17	13
	%	20	15	19	14	20	15	19	14
difference ≤ -1	Count	47	34	48	35	47	34	48	35
	%	57	40	53	39	57	40	53	39
difference = 0	Count	35	52	42	54	35	52	42	54
	%	43	60	47	60	43	60	47	60
difference ≥ 1	Count	0	0	0	1	0	0	0	1
	%	0	0	0	1	0	0	0	1
Coarse Wrinkling									
difference ≤ -3	Count	0	0	0	0	0	0	0	0
	%	0	0	0	0	0	0	0	0
difference ≤ -2	Count	10	5	10	5	10	5	10	5
	%	12	6	11	6	12	6	11	6
difference ≤ -1	Count	31	19	31	19	31	19	31	19
	%	38	22	34	21	38	22	34	21
difference = 0	Count	51	66	59	69	51	66	59	69
	%	62	77	66	77	62	77	66	77
difference ≥ 1	Count	0	1	0	2	0	1	0	2
	%	0	1	0	2	0	1	0	2

Appendix Table A.4: (cont.) Study J89-025: Differences From Baseline

		ITT				MITT			
		Week 24		LOCF		Week 24		LOCF	
		Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle
Mottled Hyperpigmentation									
difference ≤ -3	Count	5	0	5	0	5	0	5	0
	%	6	0	6	0	6	0	6	0
difference ≤ -2	Count	25	10	25	10	25	10	25	10
	%	30	12	28	11	32	12	29	11
difference ≤ -1	Count	59	26	59	26	59	26	59	26
	%	72	30	66	29	75	31	68	30
difference = 0	Count	23	58	31	62	20	57	28	61
	%	28	67	34	69	25	68	32	69
difference ≥ 1	Count	0	2	0	2	0	1	0	1
	%	0	2	0	2	0	1	0	1
Yellow-Brown Discoloration									
difference ≤ -3	Count	4	4	4	4	4	4	4	4
	%	5	5	4	4	5	5	4	5
difference ≤ -2	Count	22	13	22	13	22	13	22	13
	%	27	15	24	14	27	15	24	15
difference ≤ -1	Count	48	30	48	31	48	30	48	31
	%	59	35	53	34	59	36	53	35
difference = 0	Count	34	52	42	55	34	51	42	54
	%	41	60	47	61	41	61	47	61
difference ≥ 1	Count	0	4	0	4	0	3	0	3
	%	0	5	0	4	0	4	0	3
Laxity									
difference ≤ -3	Count	2	1	2	1	2	1	2	1
	%	2	1	2	1	2	1	2	1
difference ≤ -2	Count	8	7	9	7	8	7	9	7
	%	10	8	10	8	10	8	10	8
difference ≤ -1	Count	31	19	32	19	31	19	32	19
	%	38	22	36	21	38	22	36	21
difference = 0	Count	49	64	56	68	49	64	56	68
	%	60	74	62	76	60	74	62	76
difference ≥ 1	Count	2	3	2	3	2	3	2	3
	%	2	3	2	3	2	3	2	3

APPEARS THIS WAY ON ORIGINAL

Appendix Table A.5: Study J89-045: Demographics

	TEC –II 0.02%	Vehicle
No. Enrolled	60	60
No. Completed	56	58
No. Discontinued:	4	2
Adverse Event	3	0
Personal	1	2
Mean Age (Range)	56.7 (45-68)	56.5 (44-74)
No. Male/No. Female	6 / 54	10 / 50

Appendix Table A.6: Study J89-045: Differences From Baseline

		ITT				MITT			
		Week 24		LOCF		Week 24		LOCF	
		Treat-ment	Veh-icle	Treat-ment	Veh-icle	Treat-ment	Veh-icle	Treat-ment	Veh-icle
Tactile Roughness									
difference ≤ -3	Count	10	9	10	9	10	9	10	9
	%	18	16	17	15	20	19	20	18
difference ≤ -2	Count	21	21	21	21	21	21	21	21
	%	38	36	35	35	43	44	41	42
difference ≤ -1	Count	37	40	39	40	37	37	38	37
	%	66	69	65	67	76	77	75	74
difference = 0	Count	10	12	11	14	6	8	7	10
	%	18	21	18	23	12	17	14	20
difference ≥ 1	Count	9	6	10	6	6	3	6	3
	%	16	10	17	10	12	6	12	6
Fine Wrinkling									
difference ≤ -3	Count	15	2	16	2	15	2	16	2
	%	27	3	27	3	27	3	27	3
difference ≤ -2	Count	25	12	26	12	25	12	26	12
	%	45	21	43	20	45	21	43	20
difference ≤ -1	Count	47	27	50	27	47	27	50	27
	%	84	47	83	45	84	47	83	45
difference = 0	Count	9	26	10	28	9	26	10	28
	%	16	45	17	47	16	45	17	47
difference ≥ 1	Count	0	5	0	5	0	5	0	5
	%	0	9	0	8	0	9	0	8
Coarse Wrinkling									
difference ≤ -3	Count	10	3	10	3	10	3	10	3
	%	18	5	17	5	18	5	17	5
difference ≤ -2	Count	19	13	19	13	19	13	19	13
	%	34	22	32	22	34	22	32	22
difference ≤ -1	Count	35	27	36	27	35	27	36	27
	%	63	47	60	45	63	47	60	45
difference = 0	Count	20	29	23	31	20	29	23	31
	%	36	50	38	52	36	50	38	52
difference ≥ 1	Count	1	2	1	2	1	2	1	2
	%	2	3	2	3	2	3	2	3

Appendix Table A.6: (cont.) Study J89-045: Differences From Baseline

		ITT				MITT			
		Week 24		LOCF		Week 24		LOCF	
		Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle
Mottled Hyperpigmentation									
difference ≤ -3	Count	17	17	17	17	17	17	17	17
	%	30	29	28	28	33	30	30	29
difference ≤ -2	Count	32	32	33	32	32	32	33	32
	%	57	55	55	53	62	56	59	54
difference ≤ -1	Count	45	42	47	42	44	41	46	41
	%	80	72	78	70	85	72	82	69
difference = 0	Count	11	11	13	13	8	11	10	13
	%	20	19	22	22	15	19	18	22
difference ≥ 1	Count	0	5	0	5	0	5	0	5
	%	0	9	0	8	0	9	0	8
Yellow-brown Discoloration									
difference ≤ -3	Count	17	7	17	7	17	7	17	7
	%	30	12	28	12	30	12	28	12
difference ≤ -2	Count	30	15	31	15	30	15	31	15
	%	54	26	52	25	54	26	52	25
difference ≤ -1	Count	46	30	48	30	46	30	48	30
	%	82	52	80	50	82	52	80	50
difference = 0	Count	8	23	10	25	8	23	10	25
	%	14	40	17	42	14	40	17	42
difference ≥ 1	Count	2	5	2	5	2	5	2	5
	%	4	9	3	8	4	9	3	8
Laxity									
difference ≤ -3	Count	13	9	14	9	13	9	14	9
	%	23	16	23	15	23	16	23	15
difference ≤ -2	Count	27	21	29	21	27	21	29	21
	%	48	36	48	35	48	36	48	35
difference ≤ -1	Count	45	32	48	32	45	32	48	32
	%	80	55	80	53	80	55	80	53
difference = 0	Count	10	23	11	25	10	23	11	25
	%	18	40	18	42	18	40	18	42
difference ≥ 1	Count	1	3	1	3	1	3	1	3
	%	2	5	2	5	2	5	2	5

APPEARS THIS WAY
ON ORIGINAL

Appendix Table A.7: Study L91-026: Demographics

	TEC –II 0.02%	Vehicle
No. Enrolled	60	60
No. Completed	59	56
No. Discontinued:	11	14
Adverse Event	3	6
Personal	1	6
Loss to Follow-Up	7	2
Mean Age (Range)	55.8 (40-74)	55.2 (40-74)
No. Male/No. Female	12 / 48	12 / 48
Black	52 (43%)	57 (48%)
Hispanic	4 (3%)	3 (3%)
American Indian	2 (2%)	0
Other	2 (2%)	0

Appendix Table A.8: Study L91-026: Differences From Baseline

		ITT				MITT			
		Week 24		LOCF		Week 24		LOCF	
		Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle
Tactile Roughness									
difference ≤ -3	Count	2	1	2	1	2	1	2	1
	%	4	2	3	2	5	3	5	3
difference ≤ -2	Count	14	11	14	11	14	11	14	11
	%	25	21	23	18	38	33	35	31
difference ≤ -1	Count	23	21	24	21	21	20	22	20
	%	42	40	40	35	57	61	55	56
difference = 0	Count	29	31	33	37	16	12	18	15
	%	53	58	55	62	43	36	45	42
difference ≥ 1	Count	3	1	3	2	0	1	0	1
	%	5	2	5	3	0	3	0	3
Fine Wrinkling									
difference ≤ -3	Count	0	0	0	0	0	0	0	0
	%	0	0	0	0	0	0	0	0
difference ≤ -2	Count	3	6	3	6	3	6	3	6
	%	5	11	5	10	8	13	7	12
difference ≤ -1	Count	12	22	13	23	12	21	13	22
	%	22	42	22	38	30	47	29	42
difference = 0	Count	39	27	43	33	27	22	31	28
	%	71	51	72	55	68	49	69	54
difference ≥ 1	Count	4	4	4	4	1	2	1	2
	%	7	8	7	7	3	4	2	4
Coarse Wrinkling									
difference ≤ -3	Count	2	1	2	1	2	1	2	1
	%	4	2	3	2	8	4	7	4
difference ≤ -2	Count	5	5	5	5	5	5	5	5
	%	9	9	8	8	19	19	18	18
difference ≤ -1	Count	14	12	15	14	12	11	12	13
	%	25	23	25	23	46	42	43	46
difference = 0	Count	33	35	37	40	11	12	13	12
	%	60	66	62	67	42	46	46	43
difference ≥ 1	Count	8	6	8	6	3	3	3	3
	%	15	11	13	10	12	12	11	11

Appendix Table A.8: (cont.) Study L91-026: Differences From Baseline

		ITT Week 24		ITT LOCF		MITT Week 24		MITT LOCF	
		Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle
Local Mottled Hyperpigmentation									
difference ≤ -3	Count	0	1	0	1	0	1	0	1
	%	0	2	0	2	0	4	0	3
difference ≤ -2	Count	4	6	5	6	4	6	5	6
	%	7	11	8	10	11	22	13	19
difference ≤ -1	Count	17	15	19	15	17	14	19	14
	%	31	28	32	25	49	52	50	45
difference = 0	Count	29	30	32	37	17	11	18	15
	%	53	57	53	62	49	41	47	48
difference ≥ 1	Count	9	8	9	8	1	2	1	2
	%	16	15	15	13	3	7	3	6
General Mottled Hyperpigmentation									
difference ≤ -3	Count	0	2	0	2	0	2	0	2
	%	0	4	0	3	0	8	0	7
difference ≤ -2	Count	3	5	3	5	3	5	3	5
	%	5	9	5	8	11	19	10	17
difference ≤ -1	Count	14	12	15	12	13	12	14	12
	%	25	23	25	20	48	46	48	40
difference = 0	Count	37	37	41	44	13	13	14	17
	%	67	70	68	73	48	50	48	57
difference ≥ 1	Count	4	4	4	4	1	1	1	1
	%	7	8	7	7	4	4	3	3
Laxity									
difference ≤ -3	Count	4	2	4	2	4	2	4	2
	%	7	4	7	3	10	5	10	5
difference ≤ -2	Count	10	4	10	4	10	4	10	4
	%	18	8	17	7	26	11	24	10
difference ≤ -1	Count	19	16	19	16	19	15	19	15
	%	35	30	32	27	49	41	46	36
difference = 0	Count	34	31	38	38	18	19	20	24
	%	62	58	63	63	46	51	49	57
difference ≥ 1	Count	2	6	3	6	2	3	2	3
	%	4	11	5	10	5	8	5	7

APPEARS THIS WAY
ON ORIGINAL

Appendix Table A.9: Study K90-011: Demographics

	TEC –II 0.02%	Vehicle
No. Enrolled	40	40
No. Completed	40	40
No. Discontinued:	36	35
Adverse Event	1	1
Personal	1	2
Protocol Violation	1	0
Loss to Follow-Up	1	2
Mean Age (Range)	60.0 (46-71)	60.1 (49-70)
No. Male/No. Female	6 / 34	3 / 37

Appendix Table A.10: Study K90-011: Differences From Baseline

		ITT Week 24		ITT LOCF		MITT Week 24		MITT LOCF	
		Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle
Tactile Roughness									
difference ≤ -3	Count	2	1	2	1	2	1	2	1
	%	6	3	5	3	6	3	5	3
difference ≤ -2	Count	13	9	14	11	13	9	14	11
	%	36	26	35	28	36	26	35	28
difference ≤ -1	Count	24	24	27	26	24	24	27	26
	%	67	69	68	65	67	69	68	65
difference = 0	Count	10	10	10	13	10	10	10	13
	%	28	29	25	33	28	29	25	33
difference ≥ 1	Count	2	1	3	1	2	1	3	1
	%	6	3	8	3	6	3	8	3
Fine Wrinkling									
difference ≤ -3	Count	1	0	1	0	1	0	1	0
	%	3	0	3	0	3	0	3	0
difference ≤ -2	Count	5	0	5	0	5	0	5	0
	%	14	0	13	0	14	0	13	0
difference ≤ -1	Count	13	12	14	13	13	12	14	13
	%	36	34	35	33	36	34	35	33
difference = 0	Count	20	20	23	24	20	20	23	24
	%	56	57	58	60	56	57	58	60
difference ≥ 1	Count	3	3	3	3	3	3	3	3
	%	8	9	8	8	8	9	8	8
Coarse Wrinkling									
difference ≤ -3	Count	0	0	0	0	0	0	0	0
	%	0	0	0	0	0	0	0	0
difference ≤ -2	Count	0	0	0	0	0	0	0	0
	%	0	0	0	0	0	0	0	0
difference ≤ -1	Count	11	3	11	4	11	3	11	4
	%	31	9	28	10	31	9	28	10
difference = 0	Count	22	24	26	25	22	24	26	25
	%	61	69	65	63	61	69	65	63
difference ≥ 1	Count	3	8	3	11	3	8	3	11
	%	8	23	8	28	8	23	8	28

Appendix Table A.10: (cont.) Study K90-011: Differences From Baseline

		ITT Week 24		ITT LOCF		MITT Week 24		MITT LOCF	
		Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle	Treatment	Vehicle
Mottled Hyperpigmentation									
difference ≤-3	Count	1	1	1	1	1	1	1	1
	%	3	3	3	3	3	3	3	3
difference ≤-2	Count	3	5	4	6	3	5	4	6
	%	8	14	10	15	8	14	10	15
difference ≤-1	Count	26	21	27	23	26	21	27	23
	%	72	60	68	58	72	60	68	58
difference = 0	Count	7	9	8	12	7	9	8	12
	%	19	26	20	30	19	26	20	30
difference ≥ 1	Count	3	5	5	5	3	5	5	5
	%	8	14	13	13	8	14	13	13
Yellow-Brown Discoloration									
difference ≤-3	Count	0	0	0	0	0	0	0	0
	%	0	0	0	0	0	0	0	0
difference ≤-2	Count	3	2	4	3	3	2	4	3
	%	8	6	10	8	8	6	10	8
difference ≤-1	Count	20	17	22	18	20	17	22	18
	%	56	49	55	45	56	49	55	45
difference = 0	Count	11	15	13	18	11	15	13	18
	%	31	43	33	45	31	43	33	45
difference ≥ 1	Count	5	3	5	4	5	3	5	4
	%	14	9	13	10	14	9	13	10
Laxity									
difference ≤ -3	Count	0	0	0	0	0	0	0	0
	%	0	0	0	0	0	0	0	0
difference ≤ -2	Count	3	2	3	2	3	2	3	2
	%	8	6	8	5	8	6	8	5
difference ≤ -1	Count	16	13	18	14	16	13	18	14
	%	44	37	45	35	44	37	45	35
difference = 0	Count	16	14	18	17	16	14	18	17
	%	44	40	45	43	44	40	45	43
difference ≥ 1	Count	4	8	4	9	4	8	4	9
	%	11	23	10	23	11	23	10	23

APPEARS THIS WAY
ON ORIGINAL

Appendix Table A.11: Distributions of Signs and Symptoms of Skin Irritation

Period:	Baseline				Week 2				Week 4				Week 8			
	Treat		Veh		Treat		Veh		Treat		Veh		Treat		Veh	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Erythema																
0	268	78.8	264	77.6	139	44.6	223	70.1	175	54.0	230	70.8	199	63.0	245	76.8
1	34	10.0	32	9.4	44	14.1	36	11.3	34	10.5	35	10.8	35	11.1	33	10.3
2	8	2.4	18	5.3	44	14.1	29	9.1	40	12.3	36	11.1	36	11.4	27	8.5
3	18	5.3	18	5.3	38	12.2	19	6.0	40	12.3	17	5.2	33	10.4	9	2.8
4	3	0.9	4	1.2	20	6.4	9	2.8	24	7.4	5	1.5	8	2.5	3	0.9
5	2	0.6	3	0.9	14	4.5	2	0.6	6	1.9	1	0.3	5	1.6	2	0.6
6	5	1.5	.	.	10	3.2	.	.	4	1.2	1	0.3	.	.	.	.
7	2	0.6	.	.	2	0.6	.	.	1	0.3	.	.	.	.	.	.
8	.	.	1	0.3	1	0.3	.	.	.	.	.	.	.	.	.	.
ALL	340	100.0	340	100.0	312	100.0	318	100.0	324	100.0	325	100.0	316	100.0	319	100.0

Period:	Week 12				Week 16				Week 20				Week 24			
	Treat		Veh		Treat		Veh		Treat		Veh		Treat		Veh	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Erythema																
0	206	66.0	250	78.1	218	71.2	252	81.6	223	74.3	246	80.9	240	78.4	268	85.1
1	31	9.9	27	8.4	34	11.1	26	8.4	31	10.3	32	10.5	28	9.2	17	5.4
2	33	10.6	26	8.1	28	9.2	21	6.8	23	7.7	18	5.9	20	6.5	19	6.0
3	28	9.0	10	3.1	19	6.2	7	2.3	14	4.7	5	1.6	13	4.2	9	2.9
4	6	1.9	1	0.3	5	1.6	2	0.6	7	2.3	3	1.0	4	1.3	1	0.3
5	6	1.9	6	1.9	.	.	1	0.3	1	0.3	.	.	.	.	.	.
6	1	0.3	.	.	1	0.3	.	.	.	.	.	.	1	0.3	1	0.3
7	1	0.3	.	.	1	0.3	.	.	.	.	.	.	.	.	.	.
8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
9	.	.	.	.	.	.	.	.	1	0.3	.	.	.	.	.	.
ALL	312	100.0	320	100.0	306	100.0	309	100.0	300	100.0	304	100.0	306	100.0	315	100.0

Period:	Baseline				Week 2				Week 4				Week 8			
	Treat		Veh		Treat		Veh		Treat		Veh		Treat		Veh	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Peeling																
0	321	94.4	325	95.6	181	58.0	276	86.8	206	63.6	291	89.5	238	75.3	293	91.8
1	12	3.5	7	2.1	30	9.6	23	7.2	30	9.3	19	5.8	33	10.4	14	4.4
2	3	0.9	7	2.1	39	12.5	14	4.4	41	12.7	10	3.1	20	6.3	10	3.1
3	4	1.2	1	0.3	26	8.3	4	1.3	26	8.0	2	0.6	16	5.1	2	0.6
4	.	.	.	.	21	6.7	1	0.3	10	3.1	2	0.6	6	1.9	.	.
5	.	.	.	.	10	3.2	.	.	8	2.5	1	0.3	2	0.6	.	.
6	.	.	.	.	4	1.3	.	.	2	0.6	.	.	1	0.3	.	.
7	.	.	.	.	.	.	.	.	1	0.3	.	.	.	.	.	.
8	.	.	.	.	1	0.3	.	.	.	.	.	.	.	.	.	.
ALL	340	100.0	340	100.0	312	100.0	318	100.0	324	100.0	325	100.0	316	100.0	319	100.0

Appendix Table A.11: (cont.) Distributions of Signs and Symptoms of Skin Irritation

Period:	Week 12				Week 16				Week 20				Week 24			
	Treat		Veh		Treat		Veh		Treat		Veh		Treat		Veh	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Peeling																
0	238	76.3	302	94.4	243	79.4	293	94.8	246	82.0	291	95.7	261	85.3	303	96.2
1	27	8.7	9	2.8	26	8.5	10	3.2	26	8.7	5	1.6	21	6.9	6	1.9
2	28	9.0	6	1.9	26	8.5	6	1.9	17	5.7	7	2.3	18	5.9	5	1.6
3	12	3.8	1	0.3	7	2.3	.	.	9	3.0	1	0.3	5	1.6	1	0.3
4	5	1.6	1	0.3	3	1.0	.	.	2	0.7	.	.	.	.	.	.
5	2	0.6	.	.	.	.	.	.	.	.	.	.	1	0.3	.	.
6	.	.	.	.	1	0.3	.	.	.	.	.	.	.	.	.	.
7	.	.	1	0.3	.	.	.	.	.	.	.	.	.	.	.	.
ALL	312	100.0	320	100.0	306	100.0	309	100.0	300	100.0	304	100.0	306	100.0	315	100.0

Period:	Baseline				Week 2				Week 4				Week 8			
	Treat		Veh		Treat		Veh		Treat		Veh		Treat		Veh	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Dryness																
0	292	85.9	297	87.4	167	53.5	250	78.6	182	56.2	256	78.8	227	71.8	271	85.0
1	22	6.5	21	6.2	34	10.9	24	7.5	39	12.0	32	9.8	32	10.1	24	7.5
2	14	4.1	16	4.7	38	12.2	27	8.5	40	12.3	18	5.5	33	10.4	20	6.3
3	10	2.9	4	1.2	34	10.9	11	3.5	35	10.8	11	3.4	14	4.4	4	1.3
4	.	.	.	.	19	6.1	4	1.3	15	4.6	3	0.9	7	2.2	.	.
5	2	0.6	1	0.3	12	3.8	1	0.3	7	2.2	2	0.6	3	0.9	.	.
6	.	.	1	0.3	7	2.2	1	0.3	3	0.9	2	0.6	.	.	.	.
7	.	.	.	.	.	.	.	.	1	0.3	1	0.3	.	.	.	.
8	.	.	.	.	1	0.3	.	.	2	0.6	.	.	.	.	.	.
ALL	340	100.0	340	100.0	312	100.0	318	100.0	324	100.0	325	100.0	316	100.0	319	100.0

Period:	Week 12				Week 16				Week 20				Week 24			
	Treat		Veh		Treat		Veh		Treat		Veh		Treat		Veh	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Dryness																
0	243	77.9	283	88.4	238	77.8	286	92.6	236	78.7	287	94.4	255	83.3	295	93.7
1	20	6.4	19	5.9	33	10.8	19	6.1	33	11.0	10	3.3	28	9.2	13	4.1
2	32	10.3	15	4.7	21	6.9	4	1.3	21	7.0	6	2.0	17	5.6	6	1.9
3	11	3.5	.	.	9	2.9	.	.	6	2.0	1	0.3	5	1.6	1	0.3
4	5	1.6	2	0.6	3	1.0	.	.	3	1.0	.	.	1	0.3	.	.
5	1	0.3	.	.	.	.	.	.	.	.	.	.	.	.	.	.
6	.	.	.	.	1	0.3	.	.	.	.	.	.	.	.	.	.
7	.	.	1	0.3	1	0.3	.	.	1	0.3	.	.	.	.	.	.
ALL	312	100.0	320	100.0	306	100.0	309	100.0	300	100.0	304	100.0	306	100.0	315	100.0

Appendix Table A.11. (cont.) Distributions of Signs and Symptoms of Skin Irritation

Period:	Baseline				Week 2				Week 4				Week 8			
	Treat		Veh		Treat		Veh		Treat		Veh		Treat		Veh	
Itching	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
0	328	96.5	333	97.9	222	71.2	296	93.1	265	81.8	299	92.0	276	87.3	306	95.9
1	7	2.1	5	1.5	29	9.3	6	1.9	20	6.2	18	5.5	18	5.7	5	1.6
2	2	0.6	1	0.3	30	9.6	12	3.8	23	7.1	4	1.2	14	4.4	7	2.2
3	2	0.6	1	0.3	14	4.5	2	0.6	7	2.2	2	0.6	4	1.3	.	.
4	.	.	.	.	10	3.2	1	0.3	3	0.9	1	0.3	1	0.3	.	.
5	.	.	.	.	4	1.3	1	0.3	4	1.2	.	.	2	0.6	.	.
6	1	0.3	.	.	1	0.3	.	.	.	.	1	0.3	.	.	1	0.3
7	.	.	.	.	1	0.3	.	.	.	.	.	.	.	.	.	.
8	.	.	.	.	1	0.3	.	.	2	0.6	.	.	1	0.3	.	.
ALL	340	100.0	340	100.0	312	100.0	318	100.0	324	100.0	325	100.0	316	100.0	319	100.0

Period:	Week 12				Week 16				Week 20				Week 24			
	Treat		Veh		Treat		Veh		Treat		Veh		Treat		Veh	
Itching	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
0	272	87.2	304	95.0	273	89.2	304	98.4	275	91.7	297	97.7	289	94.4	310	98.4
1	16	5.1	9	2.8	17	5.6	5	1.6	10	3.3	5	1.6	12	3.9	2	0.6
2	11	3.5	4	1.3	12	3.9	.	.	9	3.0	2	0.7	4	1.3	2	0.6
3	5	1.6	2	0.6	.	.	.	.	3	1.0	.	.	1	0.3	1	0.3
4	3	1.0	1	0.3	.	.	.	.	2	0.7	.	.	.	.	.	.
5	3	1.0	.	.	2	0.7	.	.	1	0.3	.	.	.	.	.	.
6	.	.	.	.	1	0.3	.	.	.	.	.	.	.	.	.	.
7	1	0.3	.	.	1	0.3	.	.	.	.	.	.	.	.	.	.
8	1	0.3	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ALL	312	100.0	320	100.0	306	100.0	309	100.0	300	100.0	304	100.0	306	100.0	315	100.0

Period:	Baseline				Week 2				Week 4				Week 8			
	Treat		Veh		Treat		Veh		Treat		Veh		Treat		Veh	
Burning/Stinging																
0	332	97.6	335	98.5	148	47.4	275	86.5	207	63.9	284	87.4	243	76.9	304	95.3
1	4	1.2	1	0.3	52	16.7	21	6.6	47	14.5	20	6.2	26	8.2	8	2.5
2	1	0.3	1	0.3	45	14.4	13	4.1	41	12.7	17	5.2	30	9.5	6	1.9
3	1	0.3	3	0.9	21	6.7	5	1.6	18	5.6	3	0.9	8	2.5	.	.
4	.	.	.	.	19	6.1	3	0.9	4	1.2	.	.	3	0.9	.	.
5	.	.	.	.	12	3.8	1	0.3	3	0.9	.	.	3	0.9	.	.
6	2	0.6	.	.	6	1.9	.	.	1	0.3	.	.	1	0.3	1	0.3
7	.	.	.	.	8	2.6	.	.	1	0.3	.	.	.	.	.	.
8	.	.	.	.	1	0.3	.	.	2	0.6	.	.	2	0.6	.	.
9	.	.	.	.	.	.	.	.	.	.	1	0.3	.	.	.	.
ALL	340	100.0	340	100.0	312	100.0	318	100.0	324	100.0	325	100.0	316	100.0	319	100.0

Appendix Table A.11: (cont.) Distributions of Signs and Symptoms of Skin Irritation

Period:	Week 12				Week 16				Week 20				Week 24			
	Treat		Veh		Treat		Veh		Treat		Veh		Treat		Veh	
Burning/Stinging																
0	251	80.4	305	95.3	256	83.7	301	97.4	270	90.0	297	97.7	282	92.2	309	98.1
1	30	9.6	9	2.8	25	8.2	7	2.3	17	5.7	6	2.0	16	5.2	2	0.6
2	15	4.8	5	1.6	17	5.6	1	0.3	6	2.0	1	0.3	4	1.3	2	0.6
3	4	1.3	1	0.3	1	0.3	.	.	2	0.7	.	.	1	0.3	1	0.3
4	3	1.0	.	.	1	0.3	.	.	2	0.7	.	.	2	0.7	1	0.3
5	6	1.9	.	.	5	1.6	.	.	3	1.0	.	.	.	.	.	.
6	1	0.3	.	.	.	.	.	.	.	.	.	.	1	0.3	.	.
7	1	0.3	.	.	.	.	.	.	.	.	.	.	.	.	.	.
8	1	0.3	.	.	1	0.3	.	.	.	.	.	.	.	.	.	.
ALL	312	100.0	320	100.0	306	100.0	309	100.0	300	100.0	304	100.0	306	100.0	315	100.0

APPEARS THIS WAY
ON ORIGINAL

Figure 1: Study 89—024

MITT population: Compare Differences from Baseline

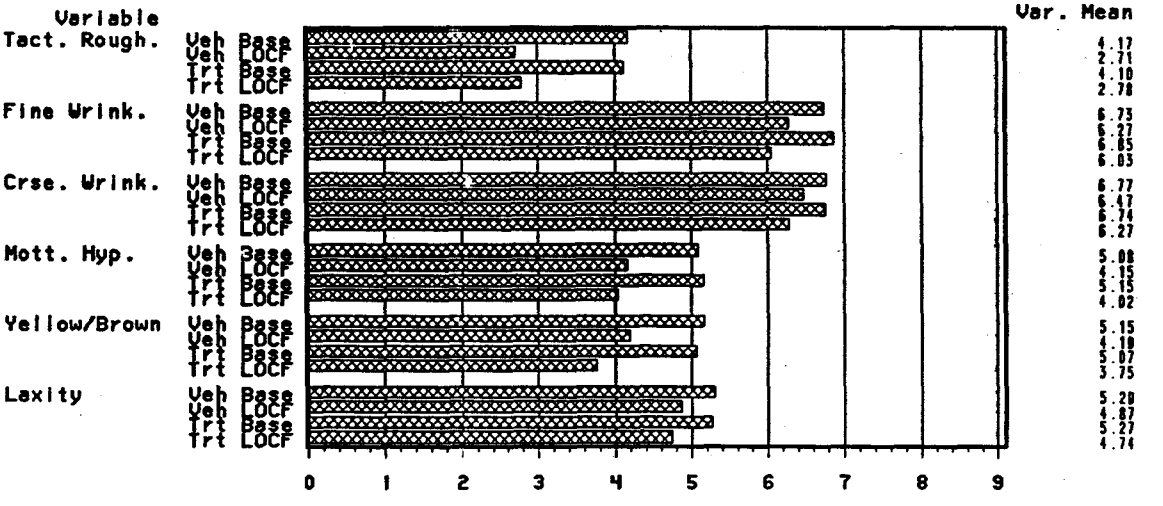


Figure 2: Study 89—025

MITT population: Differences from Baseline
Simple means

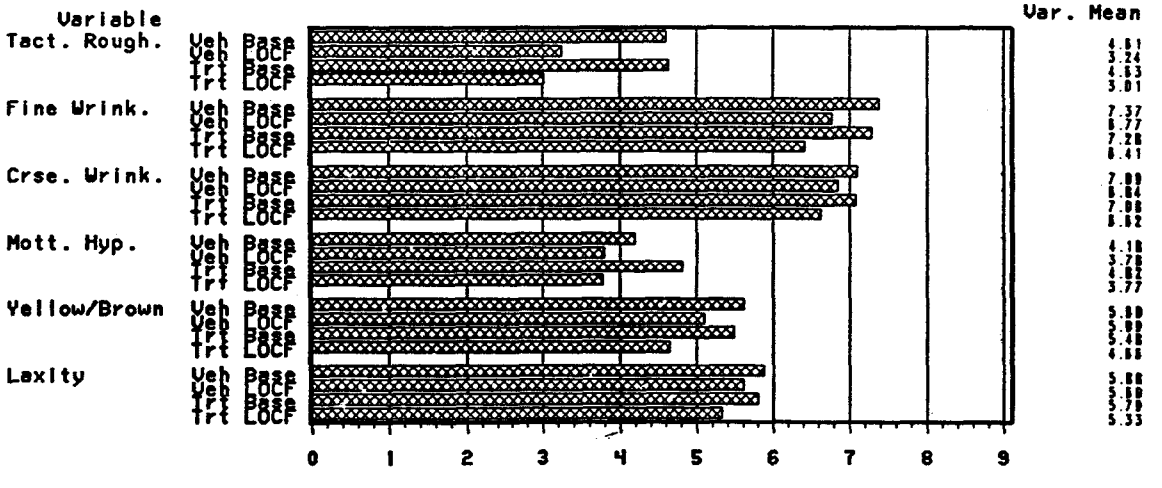


Figure 3: Study 89–045

MITT population: Differences from Baseline
Simple means

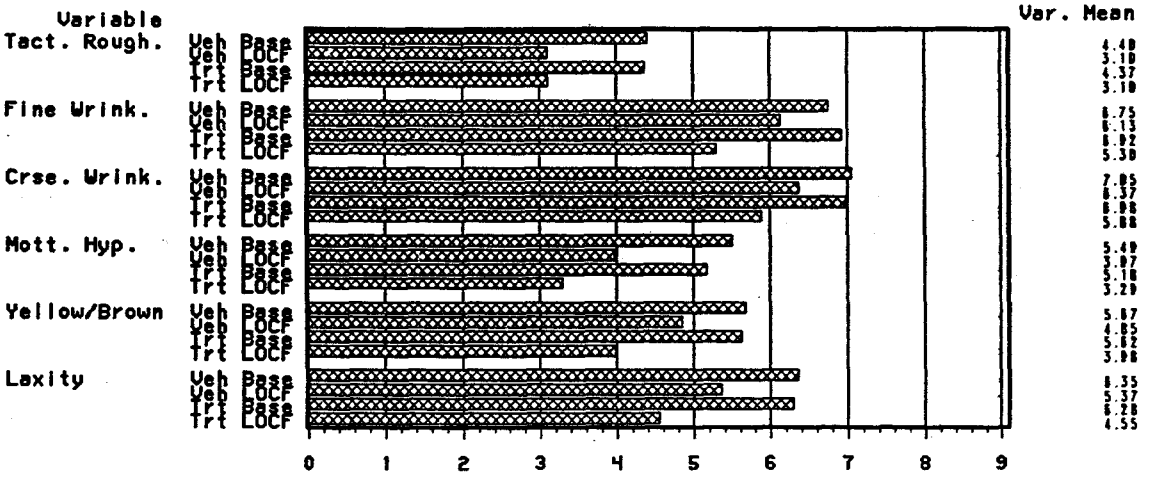


Figure 4: Study L91–026

MITT population: Differences from Baseline
Simple means

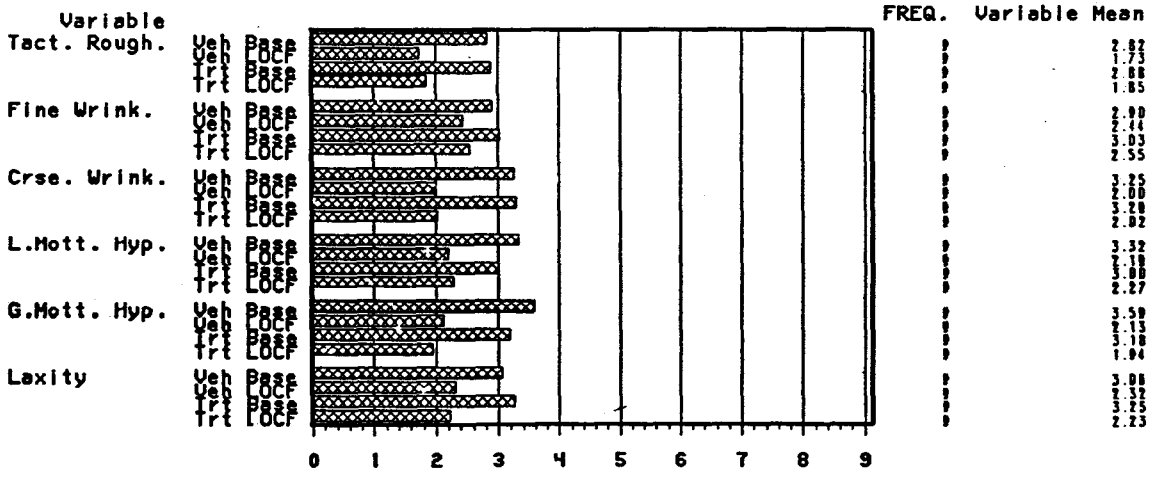
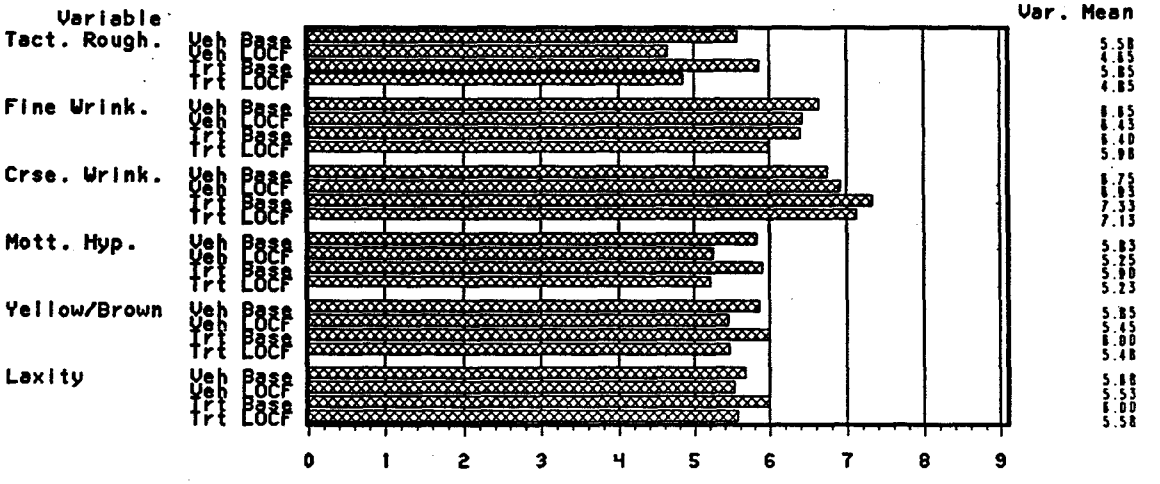


Figure 5: Study K90—011

MITT/ITT population: Differences from Baseline
Simple means



APPEARS THIS WAY
ON ORIGINAL