CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

21-644

STATISTICAL REVIEW(S)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Pharmacoepidemiology and Statistical Science
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	21-644 / S-000
Drug Name:	Clobetasol Propionate Shampoo, 0.05%
Indication(s):	Scalp Psoriasis
Applicant:	Galderma Laboratories, L.P.
Date(s):	Received 5/06/03, user fee (10 months) 3/06/04
Review Priority:	Standard
Biometrics Division:	Division 3, HFD-725
Statistical Reviewer:	Steve Thomson, HFD-725
Concurring Reviewers:	Team Leader: Mohamed Alosh, Ph. D., HFD-725 Division Director: Mohammed Huque, Ph. D., HFD-725
Medical Division:	Dermatological and Dental Drug Products, HFD-540
Clinical Team:	Reviewer: Jill Lindstrom, M.D. Team Leader: M. Luke, M.D., Ph.D.
Project Manager:	J. Smith, HFD-540

Keywords: Clinical studies, Cochran-Mantel-Haenszel, Analysis of covariance, Intent-to-treat, Bayesian analysis, Generalized estimating equations

Table of Contents

1. EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

Two pivotal trials (RD.06.SPR.18075 and RD.06.SPR.18076, denoted as Study 18075 and Study 18076 respectively) were submitted to investigate the efficacy and safety of Clobetasol Propionate Shampoo, 0.05%, versus its vehicle in subjects aged 12 years and older in the treatment of moderate to severe scalp psoriasis. Results from Study 18076 strongly support the primary and secondary efficacy claims. Results from Study 18075 are somewhat weaker, but show statistically significant superiority of clobetasol shampoo over its vehicle.

Reports from three active controlled European supporting studies were also provided. Two of these showed statistically significant superiority of clobetasol over its comparator. The other study showed statistically significant superiority of clobetasol over its vehicle, but tended to show inferiority relative to its active comparator.

1.2 Brief Overview of Clinical Studies

Two pivotal Phase 3 studies following very similar protocols were provided. As shown in the following tables, both studies were randomized, double-blind, parallel-group, placebo-controlled, multi-center evaluations of the safety and efficacy of Clobetasol Propionate Shampoo in the treatment of moderate to severe scalp psoriasis. Treatment was to be applied once daily for four weeks (or shorter if psoriasis cleared). Patients were to return two weeks after completing treatment for a follow-up assessment. Thus the end of study (EOS) is at 6 weeks.

Study report	# Centers	Study design	# of subjects/ group	Treatment regimen
Protocol 18075	12 U.S.	Vehicle Controlled (2:1)	99/Clobetasol 49/Vehicle	Once daily for 4 weeks
Protocol 18076	13 U.S./ Canada	Vehicle Controlled (2:1)	95/Clobetasol 47/Vehicle	Once daily for 4 weeks

A six point (0=None to 5=Very Severe) Global Severity Score (GSS) was the basis for the primary efficacy variable in both studies. With the concurrence of the Division of Dermatological and Dental Drug Products, this GSS was dichotomized so that a 0 or 1 score (i.e., "None" or "Minimal") was chosen as the primary endpoint, "Success". The end of treatment 4 week evaluation of this dichotomized endpoint was specified as the primary endpoint in the protocols of both pivotal studies.

Signs and symptoms of scalp psoriasis were investigated as supporting secondary variables. Each of erythema, scaling, plaque thickening, and pruritis were measured on a 4-point scale (0:none to 3:severe).

In both pivotal studies, about 92-93% of the subjects in each treatment group under each protocol completed the study. The European supporting studies are described in Appendix 6.

1.3 Statistical Issues and Findings

Statistical Issues

For pivotal trials the protocols specified that primary assessment of week 4 efficacy, i.e. treatment "success", was based on the intent-to-treat (ITT) population, with missing values imputed by last observation carried forward (LOCF). For both the primary and the secondary endpoints listed above, the protocols specified row mean comparisons using Cochran-Mantel-Haenszel tests stratified on center.

There seemed to be no major problems with the statistical aspects of this submission. However, the following may be worth noting:

1. In the primary endpoint, success on the GSS, there was evidence of a difference in effect size in the clobetasol group. In particular, in Study 18075 only one of 10 centers had success rate of 50% or greater in the clobetasol group, whereas five of 12 centers in Study 18076 achieved such a success rate. This issue is addressed in Appendix 4.

2. The demographics of the patients in this study might not represent the U.S. population as there were too few Asian and Black patients in either study to make any reliable conclusions concerning these subgroups.

3. The Sponsor listed a large number of secondary endpoints. Despite comments given in the End of Phase 2 review neither the Sponsor's protocols nor the final reports included a correction for the multiplicity of secondary endpoints. After discussion with the Medical team, only the individual scores for erythema, plaque thickening, scaling, and pruritus were classified as secondary endpoints in this statistical review, particularly the Week 4 values. To maintain overall type I error, Holm's method of adjusting for multiplicity of the secondary endpoints was chosen.

4. The remaining endpoints analyzed by the Sponsor as secondary were considered as tertiary by the Medical team. Descriptive results for these endpoints are given in Appendix 3. However, to limit the number of tests, and thus control type I error, no formal statistical tests are provided for these endpoints.

Statistical Findings

In the ITT-LOCF population in Study 18075, 28% of the Clobetasol patients achieved success at week 4, versus 10% in the vehicle group. In Study 18076 these success proportions were 42% of the Clobetasol patients versus 2% in the vehicle group. The dichotomized endpoints were analyzed using Mantel-Haenszel tests stratified on center ($p \leq 0.0118$ and $p < 0.0001$, respectively).

Both studies also included a comparison of success after a further two-week follow-up period (thus at 6 weeks), but Study 18076 showed a statistically significant difference at the end of the two week follow-up period, while the other seemed to show no particular difference.

In each study the four secondary endpoints were analyzed at the end of four weeks. Using Holm's method to correct for the multiplicity of comparisons within each study, there were statistically significant differences in favor of Clobetasol over its vehicle in the original scales. The Medical Officer also requested an analysis using dichotomized endpoints, which gave similar results. Descriptively, the remaining endpoints labelled as secondary by the Sponsor tended to support these conclusions.

Although too few Asian and Black patients were included for any claim of generalizability, results in the labelled demographic subgroups were consistent.

Appendix 7 includes a preliminary Bayesian analyses of a logistic model for the Week 4 success rate. With vague priors the posterior probability that clobetasol was greater than its vehicle was approximately 0.85 in Study 18075 and 0.998 in Study 18076. Further, as shown in Appendix 4, for this endpoint the p-values for homogeneity of odds ratios were quite large ($p=0.5811$ and $p=0.9543$, respectively), consistent with the notion of no large treatment by center interactions. Appendix 4 also assesses differences across studies.

2. INTRODUCTION

2.1 Overview

According to the Sponsor: "CLOBEX™ Shampoo 0.05% is a new dosage form of clobetasol propionate, 0.05%. There are other approved prescription forms of clobetasol propionate 0.05%, creams, ointments, gels, and scalp applications currently on the market. All current forms are indicated for twice daily application limited to 2 consecutive weeks for the relief of the inflammatory and pruritic manifestations of corticosteroid-responsive dermatoses and 4 consecutive weeks in the treatment of severe plaque-type psoriasis." This dosage form is to be applied once daily, to remain for 15 minutes on the affected scalp area, and then lathered and rinsed out.

The Sponsor noted that the "U.S. Clinical trials with CLOBEX™ Shampoo were initiated in 2001 under IND 60,934."

Summaries of the Phase 3 trials are presented below:

Table 1: Detailed Overview of the Phase 3 Studies

Study Number	Study Design	Treatment	# of Subjects	Treatment Duration
Pivotal Studies				
18075 USA	Multi-center, randomized, vehicle-controlled, double-blind, parallel group comparison	Clobetasol Shampoo, 15 min on dry scalp before rinsing qd Vehicle Shampoo 15 min on dry scalp before rinsing qd	99 49	4-week treatment followed by 2-week post-treatment follow-up
18076 USA/Canada	Multi-center, randomized, vehicle-controlled, double-blind, parallel group comparison	Clobetasol Shampoo, 15 min on dry scalp before rinsing qd Vehicle Shampoo 15 min on dry scalp before rinsing qd	95 47	4-week treatment followed by 2-week post-treatment follow-up
Non-Pivotal Studies (Europe)				
2638	Multi-center, randomized, active-controlled, investigator-blind, parallel group comparison	Clobetasol Shampoo, 15 min on dry scalp before rinsing qd Daivonex® Solution on dry scalp without rinsing bid	76 75	4-week treatment
2648	Multi-center, randomized, active-controlled, investigator-blind, parallel group comparison	Clobetasol Shampoo, 15 min on dry scalp before rinsing qd Polytar® Liquid Shampoo on wet scalp bid	121 41	4-week treatment
2665	Multi-center, randomized, active- and vehicle-controlled, investigator-blind, parallel group comparison	Clobetasol Shampoo, 15 min on dry scalp before rinsing qd Vehicle Shampoo 15 min on dry scalp before rinsing qd Dermoval® Gel on dry scalp without rinsing qd	63 20 61	4-week treatment

Statistical reviews of the two pivotal efficacy studies are given in Section 2.3 below. Summary reviews of the three supporting studies are given in Appendix 6.

2.2 Data Sources

Data for the two pivotal studies were downloaded from the FDA Electronic Data Room as SAS transport files. Two collections of data sets were provided for each study. The first set had five derived data sets including codes for population (Intent-to-treat or Per Protocol) and other computed variables. The second set had 17 data sets for Study 18075 and 14 for Study 18076, mainly derived from case report forms. Note the only known discrepancy between the Sponsor's response counts and those derived from the data sets was in the tertiary response variable Global Assessment of Improvement (Per Investigator). In particular in Study 18076 this reviewer counted 3 subjects whose assessment was a score of -1 (worse) while the Sponsor reported 4 such subjects. Other totals agreed. In particular, tabulations of the primary and secondary variables used in this analysis seem to be consistent.

Reports for Studies 18075 and 18076 were in volumes 33 and 39 of the Sponsor's submission, with protocols concluding in the successive volumes 34 and 40 respectively.

Analyses of the three non-pivotal studies, 2638, 2648, and 2665, were summarized from the Sponsor's reports of these studies and are reported in Appendix 6. Original data sets for these

data sets were not provided. These reports for the supporting studies were in volumes 43, 46, and 49 of the Sponsor's submission, respectively.

All analyses were conducted using SAS 6.12 and SAS 8.2.

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Study 18075

A Randomized, Double-Blind, Parallel Group Evaluation of Clobetasol Propionate Shampoo, 0.05% Versus Its Vehicle – An Efficacy and Safety Study in Subjects with Scalp Psoriasis

The primary objective of the study was to evaluate the efficacy and safety of Clobetasol Propionate Shampoo, 0.05%, versus its corresponding vehicle in the treatment of moderate to severe scalp psoriasis.

Design

This study was conducted as a multi-center, randomized, vehicle-controlled, double-blinded, parallel-group comparison in subjects aged 12 years and older. Patients were randomized 2:1 ratio to either Clobetasol Propionate Shampoo, 0.05%, or to Clobetasol Propionate Shampoo vehicle. Subjects were to apply the study drug once daily to the affected areas of the scalp then wait 15 minutes before lathering and rinsing. Treatment was to continue for a period of four weeks (or until clearance), with a two week treatment-free follow-up period. Subjects were evaluated at baseline and nominal weeks 2, 4, and 6. The Week 4 endpoint is the time point for the primary analysis.

Originally 12 centers were scheduled for patient recruitment. However, due to low recruitment, as was specified in the protocol, three small centers were pooled for analysis.

Efficacy Endpoints

The basis for the primary endpoint was the Global Severity Scale (GSS), on a 0-5 scale equivalent to that defined in Table 2 below. With the concurrence of the Division of Dermatological and Dental Drug Products, the protocol specified a dichotomized version of this endpoint evaluated at week 4 as the primary efficacy endpoint.

Table 2: Global Severity Scale

Score	Category	Category Description
0	Clear	Plaque thickening = none (no elevation or thickening over normal skin) Scaling = none (no evidence of scaling) Erythema = possible hyperpigmentation or residual red coloration
1	Minimal	Plaque thickening = possible but difficult to ascertain whether there is a slight elevation above normal skin level Scaling = possible residual surface dryness and scaling Erythema = up to mild (up to light red or pink coloration)
2	Mild	Plaque thickening = slight (slight but definite elevation) Scaling = fine (fine scales partially or mostly covering lesions) Erythema = up to moderate (up to definite red coloration)
3	Moderate	Plaque thickening = moderate (moderate elevation with rounded or sloped edges) Scaling = coarser (most lesions at least partially covered) Erythema = moderate (definite red coloration)
4	Severe	Plaque thickening = marked (marked elevation typically with hard or sharp edges) Scaling = coarse (non-tenacious scale predominates, covering most or all of the lesions) Erythema = very severe (very bright red coloration)
5	Very Severe	Plaque thickening = very marked (very marked elevation typically with hard or sharp edges) Scaling = very coarse (thick tenacious scale covers most or all of the lesions) Erythema = very severe (extreme red coloration; deep red coloration)

A score of 0 or 1, i.e., "Clear" or "Minimal", on this Global Severity Scale defined treatment "success". Note that a baseline GSS score of 3, or Moderate, was necessary for enrollment. Thus a subject needed to improve by at least two units to achieve success (grade 0 or 1).

For secondary endpoints the Sponsor's protocol specified the actual global severity scale, not dichotomized; individual scores for erythema, plaque thickening, scaling, and pruritus; total severity score (TSS), defined as is the sum of the erythema, plaque thickening, and scaling scores; percent scalp surface area of involvement; global assessment of improvement by the investigator; and global assessment of improvement by the subject. Despite a comment by the Division given in the End of Phase 2 (EOP2) review, neither the Sponsor's protocol nor the final report included a correction for the multiplicity of endpoints. After discussion with the Medical team, only the individual scores for erythema, plaque thickening, scaling, and pruritus were classified as secondary endpoints in this statistical review. The other endpoints listed above were considered tertiary.

These individual secondary endpoints were defined as below:

Table 3. Secondary Endpoints

Erythema (abnormal redness of the skin)

0	None	No erythema
1	Mild	Slight pinkness present
2	Moderate	Definite redness; easily recognized
3	Severe	Intense redness

Scaling (scales attached to the scalp)

0	None	No scale visible on the scalp
1	Mild	Some scales, which may often be fine, on the scalp
2	Moderate	Numerous flakes of scaling present on the scalp
3	Severe	Presence of very numerous flakes of scaling, usually large, on the scalp

Plaque Thickening (a thickening or elevation of a circumscribed lesion or plaque)

0	None	No plaque thickening
1	Mild	Slight thickening
2	Moderate	Definite but not solid thickening
3	Severe	Marked, solid thickening

Pruritus (an itching sensation)

0	None	No itching
1	Mild	Slight itching, not really bothersome
2	Moderate	Definite itching, somewhat bothersome, without loss of sleep
3	Severe	Intense itching that has caused pronounced discomfort; night rest interrupted. Excoriation of the skin from scratching may be present.

The Medical team noted that the Division has not historically recognized pruritus as a secondary endpoint in psoriasis trials. However, the team indicated that pruritus is often a symptom of scalp psoriasis and may be appropriate as a secondary endpoint and hence was retained in this analysis.

Tertiary Endpoints

The Sponsor defined the total severity score (TSS) as the sum of the erythema, plaque thickening, and scaling scores above. However, the Medical team considered the TSS to be of limited clinical use. Further, from a statistical point of view use of the TSS might not have added value when the individual scores are already used. Thus, despite the Sponsor's specification of this particular sum as a primary variable, it is considered tertiary in this analysis. Another endpoint specified as secondary by the Sponsor was the percent scalp surface area of involvement, however no details of the computation of this endpoint are provided. The Medical reviewer noted, in comments provided to the Sponsor regarding IND 60,934, serial number 004, that it was "... unclear from the protocol how surface area of involvement can be accurately estimated on the scalp." The current submission does not seem to include further details on this assessment. Hence, except for descriptive statistics, it is ignored in this analysis.

The other endpoints specified by the Sponsor as secondary included the GSS in the original scale and the global assessment of improvement by the investigator and by the subject, both measured on the following scale (the investigator's assessment includes descriptors not included here).

Global Assessment of Improvement

Score	Category
5	Clear
4	Almost clear
3	Marked improvement
2	Moderate improvement
1	Minimal improvement
0	No change
-1	Worse

This endpoint is defined in reference to the subject's baseline assessment. Because of concerns about recall bias the Division does not generally recommend such measures.

These endpoints were scheduled to be evaluated as indicated below:

Table 4. Times of Efficacy Evaluations

Variables	Baseline	Week 2	Week 4	Week 6
Global severity	X	X	X	X
Erythema	X	X	X	X
Scaling	X	X	X	X
Plaque thickening	X	X	X	X
Pruritus	X	X	X	X
Scalp surface area of involvement	X	X	X	X
Global Assessment of Improvement by Investigator		X	X	
Global Assessment of Improvement by Subject		X	X	

Recall that week 4 is the end of treatment and the time point for the primary endpoint.

Patient Disposition, Demographic, and Baseline Characteristics

Three populations are provided for analysis, starting with the intent-to-treat (ITT) group, defined as all subjects randomized and dispensed medication. The subset of those patients with observed data at each time point can be called the group of completers at that time. The ITT group where dropouts or missing are imputed using last-observation-carried-forward (LOCF), is the ITT-LOCF group. The Week 4 set of these subjects is the primary analysis group. The per protocol group is the subset of completers with no major protocol violations.

For both studies, baseline demographic characteristics and patient disposition are summarized in Appendix 5. Baseline scores for the various endpoints are included in appendices 1-4.

Statistical Methodologies

As specified in the protocol the superiority of the Clobetasol treatment group versus its vehicle is to be tested using Cochran-Mantel-Haenszel tests, stratified on centers. The protocol specified RIDIT scores and those are used here. The Week 4 comparison of success on the Global Severity Score is "the" Primary Endpoint in each study.

The Sponsor provided tests for all the secondary and tertiary endpoints cited above, but provided no correction for multiplicity. After discussion with the Medical Officer, it was decided to restrict attention to the four secondary endpoints: erythema, plaque thickening, scaling, and pruritus. To control family-wise type I error Holm's Step-down method was used. For this test, p-values are sorted by increasing size. For k comparisons at level α, the smallest observed p-value is compared to α/k. If it is significant, compare the next smallest to $\alpha/(k-1)$, then the next p-value to $\alpha/(k-2)$, etc., until the last is compared to α. Stop at the first non-significant comparison and declare all remaining comparisons statistically non-significant. This is a post hoc choice. The Sponsor did not adjust for multiplicity.

Each of the pivotal study centers with less than 3 subjects in each treatment group were to be pooled, as stated in the Sponsor's protocol.

Detailed evaluations of the primary and secondary endpoints are presented in Appendices 1-4 and 7. Appendix 1 has a detailed analysis of the Global Severity Score at each assessed timepoint. Appendices 2 and 3 include results on secondary and tertiary endpoints. Appendix 4 includes assessments of effects of studies, centers, and treatment by center interaction on the GSS success rate. Appendix 7 is a preliminary Bayesian analysis of a logistic model for Week 4 success rate. Appendix 5 provides baseline demographics and patient disposition by the end of the study. Appendix 6 has summaries of the supporting studies.

Efficacy Results

The results of the primary efficacy variable, success in Global Severity Score at the week 4 endpoint, are presented in Table 5.

Table 5. Study 18075 Efficacy Evaluations

Population	Clobetasol Propionate Shampoo, 0.05% n/N(%)	Clobetasol Propionate ShampooVehicle n/N(%)	CMH p-value
ITT -LOCF	28/99 (28.3 %)	5/49 (10.2 %)	0.012
Completers (ITT)	27/89 (30.3 %)	5/45 (11.1 %)	0.013
Per Protocol	27/88 (30.7 %)	4/42 (9.5 %)	0.008

where N=total number of evaluable subjects at week 4
n=number of subjects with success
ITT=intent to treat
LOCF=last observation carried forward

In the pivotal trial Study 18075, clobetasol shampoo was superior to its vehicle for the primary efficacy endpoint success in global severity for the ITT (LOCF), completers, and the per protocol population (all p≤0.0013). As shown in Appendix 2, using the mean ridit scores or using the dichotomized endpoints, and correcting for multiplicity using Holm's method, each of erythema, plaque thickening, scaling, and pruritis showed statistically significant differences between Clobetasol Propionate Shampoo, 0.05%, and its vehicle.

3.1.2. Study 18076

A Randomized, Double-Blind, Parallel Group Evaluation of Clobetasol Proprionate Shampoo, 0.05% Versus Its Vehicle – An Efficacy and Safety Study in Subjects with Scalp Psoriasis

This study was conducted as a multi-center, randomized, vehicle-controlled, double-blinded, parallel-group comparison involving subjects aged 12 years and older with moderate to severe scalp psoriasis, under virtually the same protocol as Study 18075. In particular endpoints for Study 18076 are as given in the discussion of Study 18075. Baseline demographic measures are tabulated in Appendix 5. Other comments about Study 18075 given above, apply here as well.

Efficacy Endpoint Outcomes

The results of the primary efficacy variable, success in Global Severity at the week 4 endpoint, are presented in Table 6 below.

Table 6. Study 18076 Efficacy Evaluations

Population	Clobetasol Propionate Shampoo, 0.05% n/N(%)	Clobetasol Propionate ShampooVehicle n/N(%)	CMH p-value
ITT (LOCF)	40/95 (42.1 %)	1/47 (2.1 %)	<0.001
Completers (ITT)	40/91 (44.0 %)	1/45 (2.2 %)	<0.001
PP	39/84 (46.4 %)	1/42 (2.4 %)	<0.001

Thus, for all three populations, in the pivotal trial Study 18076, clobetasol shampoo was statistically significantly better than its vehicle for the primary efficacy endpoint success in global severity (all p <0.001). As shown in Appendix 2, the results from secondary endpoints were supportive. In particular, using either the mean ridit scores or using the dichotomized endpoints, correcting for multiplicity using Holm's method, each of erythema, plaque thickening, scaling, and pruritis showed statistically significant differences between Clobetasol Propionate Shampoo, 0.05%, and its vehicle.

3.2 Evaluation of Safety

The primary safety issue with this submission is HP axis suppression, addressed in the Medical Officer's review.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

Note that results seem to be fairly consistent across different gender, race, and age groups, as defined in the study.

Table 7. Week 4 Subgroup Success Rates in ITT-LOCF in Study 18075

Subgroup:	Clobetasol Propionate Shampoo, 0.05% n/N(%)	Clobetasol Propionate ShampooVehicle n/N(%)
Gender		
Female	13/53 (24.5%)	4/29 (13.8%)
Male	15/46 (32.6%)	1/20 (5.0%)
Age		
12 - 17	3/3 (100%)	0/3 (0)
18 - 64	22/80 (27.5%)	4/37 (10.8%)
≥ 65	3/16 (18.8%)	1/9 (11.1%)
Race		
Caucasian	23/85 (27.1%)	4/45 (8.9%)
Other	5/14 (35.7%)	1/4 (25.0%)

Table 8. Week 4 Subgroup Success Rates in ITT-LOCF in Study 18076

Subgroup:	Clobetasol Propionate Shampoo, 0.05% n/N(%)	Clobetasol Propionate ShampooVehicle n/N(%)
Gender		
Female	23/57 (40.4%)	1/25 (4.0%)
Male	17/38 (44.7%)	0/22 (0%)
Age		
12 - 17	1/2 (50.0%)	0/3 (0%)
18 - 64	32/80 (40.0%)	1/39 (2.6%)
≥ 65	7/13 (53.8%)	0/6 (0%)
Race		
Caucasian	36/88 (40.9%)	1/43 (2.3%)
Other	4/7 (57.1%)	0/6 (0%)

Although results in the labelled demographic subgroups were consistent, it should be noted that too few Asian and Black patients were included to make any reliable conclusions concerning these subgroups. Consequently, patient populations in the studies might not be representative of the U.S. population.

4.2 Other Special/Subgroup Populations

NA

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

The primary endpoint "success" on the Global Severity Scale was analyzed using a CMH test using ridit scores, stratified on center. The primary analysis group was the ITT population at week 4.

Table 9. Efficacy Evaluation of Week 4 Success on Global Severity Score.

Study	Clobetasol Propionate Shampoo, 0.05% n/N(%)	Clobetasol Propionate ShampooVehicle n/N(%)	CMH p-value
18075	28/99 (28.3 %)	5/49 (10.2 %)	0.012
18076	40/95 (42.1 %)	1/47 (2.1 %)	<0.001

However, for both studies, the results were similar and consistent among the Per protocol and Completer populations at week 4.

The Sponsor listed a large number of secondary endpoints. Despite a comment given in the EOP2 review neither the Sponsor's protocol nor the final report included a correction for the multiplicity of secondary endpoints. After discussion with the Medical team, only the individual scores for erythema, plaque thickening, scaling, and pruritus were classified as secondary endpoints in this statistical review, particularly the Week 4 values.

Table 10. Efficacy Evaluation of Week 4 Success on Secondary Endpoints.

Endpoint	Study	Clobetasol Propionate Shampoo, 0.05% n/N(%)	Clobetasol Propionate ShampooVehicle n/N(%)	Nominal CMH p-value
Erythema	18075	62/99 (62.6 %)	20/49 (40.8 %)	0.007
	18076	65/95 (68.4 %)	16/47 (34.1 %)	<0.001
Scaling	18075	15/99 (15.2 %)	2/49 (4.1 %)	0.032
	18076	21/95 (22.1 %)	0/47 (0 %)	<0.001
Plaque Thickening	18075	34/99 (34.3 %)	5/49 (10.2 %)	<0.001
	18076	35/95 (38.5 %)	5/47 (10.6 %)	0.002
Pruritis	18075	41/99 (41.4 %)	8/49 (16.3 %)	0.002
	18076	43/95 (45.3 %)	6/47 (12.8 %)	<0.001

Using Holm's method of adjusting for multiplicity for the four endpoints within each study, all differences are statistically significant. Further details on these endpoints are given in Appendix 2.

There was some question of the consistency of results across centers. At least for the primary endpoint, this issue is addressed in Appendix 4.

Finally randomization was informally assessed. First randomization was done in blocks of 3, and seemed to be generally appropriate. The table below gives the randomization pattern of the first block of 3 patients, and over all consecutive blocks of 3 patients in the studies.

Table 11. Counts of Runs in Treatment Allocation

	Study 18075		Study 18076	
Run	1st run	All runs	1st run	All runs
VCC	4	18	3	14
CVC	5	19	2	9
CCV	3	15	8	24

Thus in Study 18075, 4 centers assigned vehicle to the first subject entered in the study, and 8 (3+5) assigned clobetasol. More generally, in Study 18075, among all randomization blocks of 3 patients, 18 assigned vehicle to the first subject in the block and 34 (19+15) assigned clobetasol.

Thus Study 18075 seems almost preternaturally balanced. Study 18076 is much more unbalanced in terms of treatment allocation within blocks, but is within the range one would expect with valid randomization.

5.2 Conclusions and Recommendations

This New Drug Application was submitted to investigate the efficacy and safety of Clobetasol Propionate Shampoo, 0.05%, versus its vehicle in subjects aged 12 years and older, when used once daily for four weeks in patients with moderate to severe scalp psoriasis. To study the efficacy of clobetasol propionate shampoo the sponsor provided results from two virtually identical randomized, double-blind, placebo-controlled, multi-center studies (Studies 18075 and 18076, respectively). A six point (0=None to 5=Very Severe) Global Severity Score was the basis for the primary efficacy variable in both studies. With the concurrence of the Division of Dermatological and Dental Drug Products, this was dichotomized so that a 0 or 1 score (i.e., "None" or "Minimal") was chosen as the primary endpoint, "Success". In the ITT - LOCF population in Study 18075 28% of the clobetasol patients achieved such success at week 4, versus 10% in the vehicle group. In Study 18076 these success proportions were 42% of the clobetasol patients versus 2% in the vehicle group. The dichotomized endpoints were analyzed using Mantel-Haenszel tests stratified on center ($p \leq 0.0118$ and $p \leq 0.0001$, respectively). Thus from a statistical point of view, in terms of the primary endpoint, we would conclude that there were statistically significant differences between clobetasol and its vehicle, particularly in Study 18076. There was also a comparison after a further two-week follow-up period, but one study showed a statistically significant difference at the end of this two week follow-up period, while the other seemed to show no particular difference. These results were supported by the results of three studies conducted in Europe (see Appendix 6), and a preliminary Bayesian analysis (see Appendix 7). Results from secondary endpoints are supporting (see Appendix 2).

APPENDICES

Appendix 1. Success on Dichotomized Global Severity

First note that comparisons are based on the dichotomized endpoint, 0-1 defining success, otherwise failure. The population used is the intent-to-treat, which without last observation carried forward (LOCF), is the population of completers by each time point. The protocol specified ridit scores for the CMH tests, and these are used here. The week 4 ITT-LOCF values are the primary endpoints.

Table A.1.1 Study 18075

Global Severity		Visit: Baseline		Week 2		Week 4 Completers		Week 4 LOCF		Week 6 EOS	
		Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
0 Clear	n	.	.	.	1	7	1	7	1	7	2
	%	.	.	.	2.1	7.9	2.2	7.1	2.0	7.6	4.4
1 Minimal	n	.	.	8	3	20	4	21	4	11	4
	%	.	.	8.1	6.3	22.5	8.9	21.2	8.2	12.0	8.9
2 Mild	n	.	.	39	7	28	11	31	11	13	7
	%	.	.	39.4	14.6	31.5	24.4	31.3	22.4	14.1	15.6
3 Moderate	n	76	35	48	31	28	25	33	28	49	25
	%	76.8	71.4	48.5	64.6	31.5	55.6	33.3	57.1	53.3	55.6
4 Minimal	n	20	13	4	6	6	3	7	4	12	6
	%	20.2	26.5	4.0	12.5	6.7	6.7	7.1	8.2	13.0	13.3
5 Very severe	n	3	1	.	.	.	1	.	1	.	1
	%	3.0	2.0	.	.	.	2.2	.	2.0	.	2.2
Overall	n	99	49	99	48	89	45	99	49	92	45
p-value[1]		.		0.9869		0.0133		0.0118		0.3815	
Homogeneity p-value[2]		.		0.6936		0.5544		0.5811		0.2275	

[1]CMH test of differences in success (clear or minimal) proportions using RIDIT scores stratified on center.
[2]Breslow Day test of homogeneity of odds ratios

Thus only differences among the week 4 completers and the week 4 ITT-LOCF are considered to be statistically significant.

Appendix 1. (cont.) Success on Dichotomized Global Severity

In Table A.1.1 above, note that the pooled center 2001, 2028, 2128 had the largest differential success rate at week 4 (6/13 successes in the clobetasol group, 0/7 in the vehicle group). The majority of these subjects were from center 2128. The contribution of each of the three pooled centers is given below. The Medical team requested (meeting 7 September 2003) a sensitivity analysis to assess the effect of deleting the three pooled centers. From a statistical point of view the usefulness of this analysis is debatable. If treatment differences are significant after removing about 1/7 of the most successful patients, results could be interpreted as showing the robustness of the effect. But failure to show differences is not particularly interpretable.

Table A.1.2 Sensitivity Analyses:
Scores of Removed Centers (2001, 2028, and 2128)

Global Severity (Dichotomized) Investigator		Baseline		Visit Week 2		Week 4 Completers		Week 4 LOCF		Week 6 EOS	
		Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
2001	0-1 Success	.	.	.	.	1	.	1	.	.	.
	2-5 Fail	2	1	2	1	1	1	2	1	2	1
2028	0-1 Success	.	.	.	.	2	.	2	.	1	.
	2-5 Fail	4	2	4	2	2	2	2	2	3	2
2128	0-1 Success	.	.	1	.	3	.	3	.	2	.
	2-5 Failure	7	4	6	3	2	2	3	4	5	2
Pooled	0-1 Success	.	.	1	.	6	.	6	.	3	.
	2-5 Failure	13	7	12	6	5	5	7	7	10	5
p-value[3]		.		0.8469		0.0713		0.0902		0.6072	
p-value[4]		.		0.8296		0.0566		0.0676		0.5832	

[3]corresponds to CMH test deleting pooled centers 2001, 2028, and 2128.
[4]corresponds to CMH test replacing observed proportion in pooled centers 2001, 2028, and 2128 with average of other centers.

Note that when deleting the pooled 2001 center, differences are no longer statistically significant, although the general trend in favor of clobetasol is still evident.

Appendix 1. (cont.) Success on Dichotomized Global Severity

Table A.1.3 Study 18076

Global Severity		Baseline		Visit Week 2		Week 4 Completers		Week 4 LOCF		Week 6 EOS	
		Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
0 Clear	n	.	.	.	.	10	.	10	.	9	.
	%	.	.	.	.	11.0	.	10.5	.	10.2	.
1 Minimal	n	.	.	17	3	30	1	30	1	12	2
	%	.	.	18.3	6.5	33.0	2.2	31.6	2.1	13.6	4.5
2 Mild	n	.	.	34	7	27	6	28	6	17	7
	%	.	.	36.6	15.2	29.7	13.3	29.5	12.8	19.3	15.9
3 Moderate	n	70	32	32	23	17	27	19	29	36	23
	%	73.7	68.1	34.4	50.0	18.7	60.0	20.0	61.7	40.9	52.3
4 Minimal	n	20	10	8	12	6	10	7	10	14	11
	%	21.1	21.3	8.6	26.1	6.6	22.2	7.4	21.3	15.9	25.0
5 Very severe	n	5	5	2	1	1	1	1	1	.	1
	%	5.3	10.6	2.2	2.2	1.1	2.2	1.1	2.1	.	2.3
Overall	n	95	47	93	46	91	45	95	47	88	44
p-value[1]		.		0.0805		<0.0001		<0.0001		0.0034	
Homogeneity p-value[2]		.		0.5582		0.9135		0.9543		0.5575	

[1]CMH test of differences in success (clear or minimal) proportions using RIDIT scores stratified on center.
[2]Breslow Day test of homogeneity of odds ratios

Thus in Study 18076, differences at Week 4 LOCF, Week 4, and Week 6 are considered to be statistically significant. At the request of the medical team, the sensitivity analyses were performed deleting center 439. This center had the largest differential success rate at Week 4 (6/10 successes in the clobetasol group, 0/5 in the vehicle group).

Table A.1.4 Sensitivity Analyses: Scores of Removed Center 439

		Baseline		Week 2		Week 4 Completers		Week 4 LOCF		Week 6 EOS	
		Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
0-1 Success	n	.	.	3	.	6	.	6	.	5	.
2-5 Failure	n	10	5	7	5	4	5	4	5	5	5
All	n	10	5	10	5	10	5	10	5	10	5
p-value[3]		.		0.1852		<0.0001		0.0001		0.0186	
p-value[4]		.		0.1578		<0.0001		<0.0001		0.0136	

[3]corresponds to CMH test deleting pooled centers 439.
[4]corresponds to CMH test replacing observed proportion in center 439 with average of other centers

Even after deleting this center, differences at Week 4 LOCF, Week 4, and Week 6 are considered to be statistically significant.

APPEARS THIS WAY
ON ORIGINAL

Appendix 2. Secondary Endpoints

The medical team specified the Week 4 LOCF values of erythema, scaling, plaque thickening and pruritis as the relevant secondary endpoints (i.e. of use in labelling).

1. To test treatment differences the Sponsor's protocol specifies CMH row mean tests using ridit scores based on the 0-3 scale. To control Type 1 error, Holm's stepdown method was used to control family-wise error over the four signs and symptoms. Statistical significance levels at weeks other than the fourth are provided for reference only. To control error they should not be used for decisions.

2. On September 11, 2003, the medical review team requested an MH test of differences using dichotomized scores such that an erythema score of 0 or 1 (none or mild) is considered a success, while scaling, plaque thickening, and pruritis are dichotomized so that a score of 0 (none) is considered a success. These are also included in the tables below.

Note that at the Week 4 LOCF success endpoints, four different comparisons are specified in each study. To control family-wise Type I error Holm's Step-down method is used. For this test, p-values are sorted by increasing size. For k comparisons at level α, the smallest observed p-value is compared to α/k. If it is significant compare the next smallest to $\alpha/(k-1)$, then the next p-value to $\alpha/(k-2)$, etc., until the last is compared to α. Stop at the first non-significant comparison and declare all remaining comparisons statistically non-significant.

For studies 18075 and 18076, at the Week 4 LOCF endpoint we get the table for testing differences in mean ridit scores on the 0-3 scale:

Comparison	Holmes Bound	P-values Study 18075	P-values Study 18076
Pruritis	0.0125	0.0013	<0.0001
Erythema	0.0167	0.0046	<0.0001
Plaque Thickening	0.025	0.0056	<0.0001
Scaling	0.05	0.0118	<0.0001

For both studies all differences are statistically significant. So at Week 4, we would conclude that for all four secondary endpoints, the differences between clobetasol and its vehicle are statistically significant.

Using the dichotomized endpoints requested by the Medical Officer, for studies 18075 at the Week 4 LOCF endpoint we get the table:

Appendix 2. (cont.) Secondary Endpoints

Comparison	Holmes Bound	P-values Study 18075
Plaque Thickening	0.0125	0.0005
Pruritis	0.0167	0.0021
Erythema	0.025	0.0070
Scaling	0.05	0.0317

So, for all four comparisons, differences are statistically significant for the dichotomized endpoint in Study 18075.

Using the dichotomized endpoints for Study 18076 at the Week 4 LOCF endpoint we get the table:

Comparison	Holmes Bound	P-values Study 18076
Erythema	0.0125	0.0001
Pruritis	0.0167	0.0002
Scaling	0.025	0.0006
Plaque Thickening	0.05	0.0015

Thus, as with Study 18075, in Study 18076 for all four comparisons, differences are statistically significant for the dichotomized endpoint. The complete tables used to generate these conclusions are as follows:

Table A.2.1 Study 18075 Erythema

		Visit									
		Baseline		Week 2		Week 4 Completers		Week 4 LOCF		Week 6 EOS	
		Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
0 None	N	.	.	3	1	12	1	12	1	11	2
	%	.	.	3.0	2.1	13.5	2.2	12.1	2.0	12.0	4.4
1 Mild	N	3	1	45	9	45	19	50	19	26	13
	%	3.0	2.0	45.5	18.8	50.6	42.2	50.5	38.8	28.3	28.9
2 Moderate	N	69	34	44	31	28	20	31	23	43	24
	%	69.7	69.4	44.4	64.6	31.5	44.4	31.3	46.9	46.7	53.3
3 Severe	N	27	14	7	7	4	5	6	6	12	6
	%	27.3	28.6	7.1	14.6	4.5	11.1	6.1	12.2	13.0	13.3
p-value[1]		0.7656		0.0020		0.0064		0.0046		0.3157	
p-value[2]		0.7132		0.0012		0.0187		0.0070		0.3982	

[1]CMH test using row mean ridits based on 0-3 scale.
[2]CMH test using ridits on dichotomized scale (0,1 vs 2,3).

Appendix 2. (cont.) Secondary Endpoints

Table A.2.2 Study 18075 Scaling

		Visit									
		Baseline		Week 2		Week 4 Completers		Week 4 LOCF		Week 6 EOS	
		Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
0 None	N	.	.	3	1	15	2	15	2	12	2
	%	.	.	3.0	2.1	16.9	4.4	15.2	4.1	13.0	4.4
1 Mild	N	1	.	41	11	39	15	43	15	19	12
	%	1.0	.	41.4	22.9	43.8	33.3	43.4	30.6	20.7	26.7
2 Moderate	N	59	32	46	29	29	23	33	26	42	22
	%	59.6	65.3	46.5	60.4	32.6	51.1	33.3	53.1	45.7	48.9
3 Severe	N	39	17	9	7	6	5	8	6	19	9
	%	39.4	34.7	9.1	14.6	6.7	11.1	8.1	12.2	20.7	20.0
p-value[1]		0.5160		0.0409		0.0163		0.0118		0.6851	
p-value[2]		.		0.7714		0.0314		0.0317		0.1166	

[1]CMH test using row mean ridits based on 0-3 scale.
[2]CMH test using ridits on dichotomized scale (0 vs 1,2,3).

Table A.2.3 Study 18075 Plaque Thickening

		Visit									
		Baseline		Week 2		Week 4 Completers		Week 4 LOCF		Week 6 EOS	
		Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
0 None	N	.	2	16	3	32	5	34	5	23	7
	%	.	4.1	16.2	6.3	36.0	11.1	34.3	10.2	25.0	15.6
1 Mild	N	22	6	43	13	33	21	35	21	29	15
	%	22.2	12.2	43.4	27.1	37.1	46.7	35.4	42.9	31.5	33.3
2 Moderate	N	65	35	37	29	22	17	27	21	35	21
	%	65.7	71.4	37.4	60.4	24.7	37.8	27.3	42.9	38.0	46.7
3 Severe	N	12	6	3	3	2	2	3	2	5	2
	%	12.1	12.2	3.0	6.3	2.2	4.4	3.0	4.1	5.4	4.4
p-value[1]		0.8715		0.0006		0.0057		0.0056		0.3229	
p-value[2]		0.0393		0.0369		0.0008		0.0005		0.2074	

[1]CMH test using row mean ridits based on 0-3 scale.
[2]CMH test using ridits on dichotomized scale (0 vs 1,2,3).

Appendix 2. (cont.) Secondary Endpoints

Table A.2.4 Study 18075 Pruritis

		Visit									
		Baseline		Week 2		Week 4 Completers		Week 4 LOCF		Week 6 EOS	
		Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
0 None	N	2	.	20	7	39	8	41	8	20	7
	%	2.0	.	20.2	14.6	43.8	17.8	41.4	16.3	21.7	15.6
1 Mild	N	17	9	49	17	30	19	34	19	26	17
	%	17.2	18.4	49.5	35.4	33.7	42.2	34.3	38.8	28.3	37.8
2 Moderate	N	57	29	24	16	14	12	17	14	34	14
	%	57.6	59.2	24.2	33.3	15.7	26.7	17.2	28.6	37.0	31.1
3 Severe	N	23	11	6	8	6	6	7	8	12	7
	%	23.2	22.4	6.1	16.7	6.7	13.3	7.1	16.3	13.0	15.6
p-value[1]		0.9263		0.0288		0.0045		0.0013		0.8659	
p-value[2]		0.3312		0.3433		0.0027		0.0021		0.4393	

[1]CMH test using row mean ridits based on 0-3 scale.
[2]CMH test using ridits on dichotomized scale (0 vs 1,2,3).

Table A.2.5 Study 18076 Erythema

		Visit									
		Baseline		Week 2		Week 4 Completers		Week 4 LOCF		Week 6 EOS	
		Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
0 None	N	.	.	9	1	17	3	17	3	12	1
	%	.	.	9.7	2.2	18.7	6.7	17.9	6.4	13.6	2.3
1 Mild	N	5	3	41	10	47	13	48	13	27	10
	%	5.3	6.4	44.1	21.7	51.6	28.9	50.5	27.7	30.7	22.7
2 Moderate	N	67	34	38	28	24	23	26	25	41	27
	%	70.5	72.3	40.9	60.9	26.4	51.1	27.4	53.2	46.6	61.4
3 Severe	N	23	10	5	7	3	6	4	6	8	6
	%	24.2	21.3	5.4	15.2	3.3	13.3	4.2	12.8	9.1	13.6
p-value[1]		0.6863		0.0008		<0.0001		<0.0001		0.0173	
p-value[2]		0.0804		0.0013		0.0001		0.0001		0.0302	

[1]CMH test using row mean ridits based on 0-3 scale.
[2]CMH test using ridits on dichotomized scale (0,1 vs 2,3).

Appendix 2. (cont.) Secondary Endpoints

Table A.2.6 Study 18076 Scaling

		Visit									
		Baseline		Week 2		Week 4 Completers		Week 4 LOCF		Week 6 EOS	
		Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
0 None	N	.	.	4	1	21	.	21	.	9	.
	%	.	.	4.3	2.2	23.1	.	22.1	.	10.2	.
1 Mild	N	3	2	46	12	40	9	41	9	25	7
	%	3.2	4.3	49.5	26.1	44.0	20.0	43.2	19.1	28.4	15.9
2 Moderate	N	56	22	34	21	23	23	25	25	36	19
	%	58.9	46.8	36.6	45.7	25.3	51.1	26.3	53.2	40.9	43.2
3 Severe	N	36	23	9	12	7	13	8	13	18	18
	%	37.9	48.9	9.7	26.1	7.7	28.9	8.4	27.7	20.5	40.9
p-value[1]		0.3017		0.0019		<0.0001		<0.0001		0.0002	
p-value[2]		.		0.6097		0.0005		0.0006		0.0150	

[1]CMH test using row mean ridits based on 0-3 scale.
[2]CMH test using ridits on dichotomized scale (0 vs 1,2,3).

Table A.2.7 Study 18076 Plaque Thickening

		Visit									
		Baseline		Week 2		Week 4 Completers		Week 4 LOCF		Week 6 EOS	
		Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
0 None	N	.	.	19	3	35	5	35	5	19	4
	%	.	.	20.4	6.5	38.5	11.1	36.8	10.6	21.6	9.1
1 Mild	N	15	7	43	10	40	11	41	12	34	9
	%	15.8	14.9	46.2	21.7	44.0	24.4	43.2	25.5	38.6	20.5
2 Moderate	N	65	29	24	25	11	22	14	23	26	26
	%	68.4	61.7	25.8	54.3	12.1	48.9	14.7	48.9	29.5	59.1
3 Severe	N	15	11	7	8	5	7	5	7	9	5
	%	15.8	23.4	7.5	17.4	5.5	15.6	5.3	14.9	10.2	11.4
p-value[1]		0.2367		0.0001		<0.0001		<0.0001		0.0026	
p-value[2]		.		0.0414		0.0014		0.0015		0.0406	

[1]CMH test using row mean ridits based on 0-3 scale.
[2]CMH test using ridits on dichotomized scale (0 vs 1,2,3).

Appendix 2. (cont.) Secondary Endpoints

Table A.2.8 Study 18076 Pruritis

		Baseline		Week 2		Visit Week 4 Completers		Week 4 LOCF		Week 6 EOS	
		Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
0 None	N	3	.	28	4	43	6	43	6	21	6
	%	3.2	.	30.1	8.7	47.3	13.3	45.3	12.8	23.9	13.6
1 Mild	N	25	7	46	19	37	17	38	17	34	15
	%	26.3	14.9	49.5	41.3	40.7	37.8	40.0	36.2	38.6	34.1
2 Moderate	N	46	24	15	15	8	14	9	15	26	16
	%	48.4	51.1	16.1	32.6	8.8	31.1	9.5	31.9	29.5	36.4
3 Severe	N	21	16	4	8	3	8	5	9	7	7
	%	22.1	34.0	4.3	17.4	3.3	17.8	5.3	19.1	8.0	15.9
p-value[1]		0.0309		0.0002		<0.0001		<0.0001		0.1413	
p-value[2]		0.2014		0.0043		0.0002		0.0002		0.1632	

[1]CMH test using row mean ridits based on 0-3 scale.
[2]CMH test using ridits on dichotomized scale (0 vs 1,2,3).

Appendix 3. Tertiary Endpoints

The following tables are included for informational purposes only. On September 11, 2003, the Medical team indicated that these endpoints were not of particular clinical interest. Hence, no significance levels for tests of treatment differences were provided. For total severity score and % scalp area of involvement, means at baseline are provided. For later visits, change from baseline is provided. For both the investigator and the patient global assessment of improvement, the entire distribution over time is provided.

Table A.3.1 Study 18075 Total Severity Score (At Baseline) / Change in Severity Score (From Week 2 on)

	Visit									
	Baseline		Week 2		Week 4 Completers		Week 4 LOCF		Week 6 EOS	
	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
Mean	6.5	6.5	-2.1	-1.1	-2.9	-1.9	-2.9	-1.7	-1.9	-1.5
Std dev	1.1	1.2	1.7	1.3	2.0	1.5	2.0	1.6	2.1	1.6
n	99	49	99	48	89	45	99	49	92	45

Table A.3.2 Study 18075 Global Assessment of Improvement (As Per Investigator)

		Visit					
		Week 2		Week 4 Completers		Week 4 LOCF	
		Clob	Veh	Clob	Veh	Clob	Veh
5 Clear	N	.	1	7	1	7	1
	%	.	2.1	7.9	2.2	7.2	2.1
4 Almost Clear	N	11	2	17	3	19	3
	%	11.3	4.3	19.1	6.7	19.6	6.4
3 Mark improv	N	18	5	18	3	19	3
	%	18.6	10.6	20.2	6.7	19.6	6.4
2 Mod improv	N	29	4	15	12	17	12
	%	29.9	8.5	16.9	26.7	17.5	25.5
1 Min improv	N	18	18	17	19	18	19
	%	18.6	38.3	19.1	42.2	18.6	40.4
0 No change	N	19	14	10	5	12	6
	%	19.6	29.8	11.2	11.1	12.4	12.8
-1 Worse	N	2	3	5	2	5	3
	%	2.1	6.4	5.6	4.4	5.2	6.4

Appendix 3. (cont.) Tertiary Endpoints

Table A.3.3 Study 18075 Global Assessment of Improvement (As Per Subject)

		Visit					
		Week 2		Week 4 Completers		Week 4 LOCF	
		Clob	Veh	Clob	Veh	Clob	Veh
5 Clear	N	1	1	4	1	5	1
	%	1.0	2.1	4.5	2.2	5.2	2.1
4 Almost Clear	N	10	2	20	2	20	2
	%	10.3	4.3	22.5	4.4	20.6	4.3
3 Mark improv	N	19	1	18	2	20	2
	%	19.6	2.1	20.2	4.4	20.6	4.3
2 Mod improv	N	30	8	15	13	17	13
	%	30.9	17.0	16.9	28.9	17.5	27.7
1 Min improv	N	18	17	19	16	20	16
	%	18.6	36.2	21.3	35.6	20.6	34.0
0 No change	N	18	13	11	7	13	8
	%	18.6	27.7	12.4	15.6	13.4	17.0
-1 Worse	N	1	5	2	4	2	5
	%	1.0	10.6	2.2	8.9	2.1	10.6

Table A.3.4 Study 18075 % Scalp Surface Area of Involvement (At Baseline / Change from Baseline at Week 2 and later)

					Visit					
	Baseline		Week 2		Week 4 Completers		Week 4 LOCF		Week 6 EOS	
	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
mean	35.1	31.4	-8.6	-3.9	-14.8	-3.9	-13.8	-3.5	-10.2	-3.2
Std dev	23.4	23.4	15.3	12.1	20.8	14.6	20.1	14.0	18.7	16.9
n	99	49	99	48	89	45	99	49	92	45

Appendix 3. (cont.) Tertiary Endpoints

Study 18076

Table A.3.5 Study 18076 Total Severity Score (At Baseline) / Change in Severity Score (From Week 2 on)

	Baseline		Visit Week 2		Week 4 Completers		Week 4 LOCF		Week 6 EOS	
	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
mean	6.5	6.5	-2.4	-1.0	-3.4	-1.2	-3.3	-1.2	-2.0	-0.9
Std dev	1.1	1.2	1.6	1.6	1.9	1.7	2.0	1.7	2.0	1.7
n	95	47	93	46	91	45	95	47	88	44

Table A.3.6 Study 18076 Global Assessment of Improvement (As Per Investigator)

		Visit Week 2		Week 4 Completers		Week 4 LOCF	
		Clob	Veh	Clob	Veh	Clob	Veh
5 Clear	N	.	.	9	.	9	.
	%	.	.	9.9	.	9.7	.
4 Almost Clear	N	8	.	23	1	23	1
	%	8.6	.	25.3	2.3	24.7	2.2
3 Mark improv	N	22	6	24	4	24	4
	%	23.7	13.0	26.4	9.1	25.8	8.9
2 Mod improv	N	27	5	17	4	18	4
	%	29.0	10.9	18.7	9.1	19.4	8.9
1 Min improv	N	22	11	7	16	7	16
	%	23.7	23.9	7.7	36.4	7.5	35.6
0 No change	N	14	20	11	17	12	17
	%	15.1	43.5	12.1	38.6	12.9	37.8
-1 Worse	N	.	4	.	2	.	3
	%	.	8.7	.	4.5	.	6.7

Appendix 3. (cont.) Tertiary Endpoints

Table A.3.7 Study 18076 Global Assessment of Improvement (As Per Subject)

		Visit					
		Week 2		Week 4 Completers		Week 4 LOCF	
		Clob	Veh	Clob	Veh	Clob	Veh
5 Clear	N	1	.	4	.	4	.
	%	1.1	.	4.4	.	4.3	.
4 Almost Clear	N	11	1	24	1	24	1
	%	11.8	2.2	26.4	2.2	25.8	2.2
3 Mark improv	N	18	3	25	8	25	8
	%	19.4	6.5	27.5	17.8	26.9	17.4
2 Mod improv	N	33	8	18	9	19	9
	%	35.5	17.4	19.8	20.0	20.4	19.6
1 Min improv	N	22	13	13	9	13	9
	%	23.7	28.3	14.3	20.0	14.0	19.6
0 No change	N	6	14	6	14	6	14
	%	6.5	30.4	6.6	31.1	6.5	30.4
-1 Worse	N	2	7	1	4	2	5
	%	2.2	15.2	1.1	8.9	2.2	10.9

Table A.3.8 Study 18076 % Scalp Surface Area of Involvement (At Baseline / Change from Baseline at Week 2 and later)

	Visit									
	Baseline		Week 2		Week 4 Completers		Week 4 LOCF		Week 6 EOS	
	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
Mean	39.1	42.8	-11.9	-1.5	-18.8	-2.9	-18.0	-2.7	-11.8	-0.6
Std dev	26.4	28.6	19.4	7.7	22.1	9.4	21.9	9.3	19.0	8.4
n	95	47	93	46	91	45	95	47	88	44

Appendix 4. Investigator/Center Effects on Dichotomized Global Severity

The following tables provide the relative proportions of success on the Global Severity Scale over the various Study visits. In Study 18075, at the Week 4 LOCF point, in centers 2020, 2065, and 2132 the relative proportions of success are equal across treatment groups, while at all other centers the relative proportions of success are higher in the clobetasol group. Statistically such observations are confirmed by the large significance levels associated with the Breslow-Day test of homogeneity of odds ratios across centers (p=0.5811 at Week 4 ITT-LOCF). Note that center 2001 is the label of pooled center 2001, 2028, and 2128,

Table A.4.1 Study 18075 Global Evaluation

Inv. Number		Visit: Baseline		Week 2		Week 4 Completers		Week 4 LOCF		Week 6 EOS	
		Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
2001	n/N	0/13	0/7	1/13	0/6	6/11	0/5	6/13	0/7	3/13	0/5
	%	.	.	7.7	0.0	54.5	0.0	46.2	0.0	23.1	0.0
2019	n/N	0/7	0/4	2/7	1/4	2/6	0/4	2/7	0/4	1/6	1/4
	%	0.0	0.0	28.6	25.0	33.3	0.0	28.6	0.0	16.7	25.0
2020	n/N	0/12	0/6	1/12	1/6	6/11	3/6	6/12	3/6	2/11	3/6
	%	0.0	0.0	8.3	16.7	54.5	50.0	50.0	50.0	18.2	50.0
2023	n/N	0/10	0/4	1/10	0/4	4/9	0/4	4/10	0/4	1/9	0/4
	%	0.0	0.0	10.0	0.0	44.4	0.0	40.0	0.0	11.1	0.0
2032	n	0/6	0/3	1/6	0/3	1/5	0/3	1/6	0/3	0/6	1/3
	%	0.0	0.0	16.7	0.0	20.0	0.0	16.7	0.0	0.0	33.3
2063	n	0/14	0/7	1/14	0/7	1/11	0/7	2/14	0/7	2/12	0/7
	%	0.0	0.0	7.1	0.0	9.1	0.0	14.3	0.0	16.7	0.0
2065	n	0/12	0/6	0/12	1/6	2/11	1/4	2/12	1/6	3/10	0/5
	%	0.0	0.0	0.0	16.7	18.2	25.0	16.7	16.7	30.0	0.0
2066	n	0/6	0/3	0/6	0/3	1/6	0/3	1/6	0/3	1/6	0/3
	%	0.0	0.0	0.0	0.0	16.7	0.0	16.7	0.0	16.7	0.0
2127	n	0/9	0/4	0/9	0/4	2/9	0/4	2/9	0/4	3/9	0/3
	%	0.0	0.0	0.0	0.0	22.2	0.0	22.2	0.0	33.3	0.0
2132	n	0/10	0/5	1/10	1/5	2/10	1/5	2/10	1/5	2/10	1/5
	%	0.0	0.0	10.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0
All	n	0/99	0/49	8/99	4/48	27/89	5/45	28/99	5/49	18/92	6/45
	%	0.0	0.0	8.1	8.3	30.3	11.1	28.3	10.2	19.6	13.3
Breslow-Day				0.6936		0.5544		0.5811		0.2275	

Appendix 4. (cont.) Investigator/Center Effects on Dichotomized Global Severity

In Study 18076, at the Week 4 LOCF endpoint, as well as the Week 4 endpoint, at all centers the relative proportions of success are higher in the Clobetasol group than in the vehicle group. Statistically such observations are confirmed by the large significance levels associated with the Breslow-Day test of homogeneity of odds ratios across centers (p=0.9543 at Week 4 ITT).

Table A.4.2 Study 18076 Global Evaluation

							Visit Number			
Inv.	Baseline		Week 2		Week 4		Week 4 LOCF		Week 6 EOS	
Number	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh	Clob	Veh
438 n/N	0/11	0/4	5/11	0/4	7/11	0/4	7/11	0/4	2/11	0/4
%	0.0	0.0	45.5	0.0	63.6	0.0	63.6	0.0	18.2	0.0
439 n/N	0/10	0/5	3/10	0/5	6/10	0/5	6/10	0/5	5/10	0/5
%	0.0	0.0	30.0	0.0	60.0	0.0	60.0	0.0	50.0	0.0
740 n/N	0/6	0/3	0/6	0/3	2/6	0/3	2/6	0/3	0/6	0/3
%	0.0	0.0	0.0	0.0	33.3	0.0	33.3	0.0	0.0	0.0
1086 n/N	0/7	0/3	1/7	0/3	2/6	0/3	2/7	0/3	2/6	0/3
%	0.0	0.0	14.3	0.0	33.3	0.0	28.6	0.0	33.3	0.0
1170 n/N	0/12	0/6	2/12	1/6	4/12	0/6	4/12	0/6	3/12	1/6
%	0.0	0.0	16.7	16.7	33.3	0.0	33.3	0.0	25.0	16.7
2094 n/N	0/9	0/4	0/9	0/4	2/8	0/4	2/9	0/4	1/8	0/4
%	0.0	0.0	0.0	0.0	25.0	0.0	22.2	0.0	12.5	0.0
2096 n	0/5	0/3	1/5	0/3	2/5	0/3	2/5	0/3	2/5	0/3
%	0.0	0.0	20.0	0.0	40.0	0.0	40.0	0.0	40.0	0.0
2102 n	0/9	0/5	3/9	2/4	5/9	1/4	5/9	1/5	2/9	1/4
%	0.0	0.0	33.3	50.0	55.6	25.0	55.6	20.0	22.2	25.0
2129 n	0/4	0/3	0/3	0/3	1/3	0/3	1/4	0/3	1/3	0/3
%	0.0	0.0	0.0	0.0	33.3	0.0	25.0	0.0	33.3	0.0
2150 n	0/6	0/3	1/5	0/3	3/5	0/3	3/6	0/3	0/4	0/3
%	0.0	0.0	20.0	0.0	60.0	0.0	50.0	0.0	0.0	0.06
2165 n	0/6	0/3	1/6	0/3	3/6	0/3	3/6	0/3	3/4	0/3
%	0.0	0.0	16.7	0.0	50.0	0.0	50.0	0.0	75.0	0.0
2166 n	0/10	0/5	0/10	0/5	3/10	0/4	3/10	0/5	0/10	0/3
%	0.0	0.0	0.0	0.0	30.0	0.0	30.0	0.0	0.0	0.0
All n/N	0/95	0/47	17/93	3/46	40/92	1/45	40/95	1/47	21/88	2/44
%	0.0	0.0	18.3	6.5	44.0	2.2	42.1	2.1	23.9	4.5
Breslow-Day			0.5582		0.9135		0.9543		0.5575	

Appendix 4. (cont.) Investigator/Center Effects on Dichotomized Global Severity

Observe that in Study 18075 only one of 10 centers had success rate of 50% or greater in the Clobetasol group, whereas five of 12 centers in Study 18076 achieved such a success rate. To investigate this issue further the Week 4 LOCF success rate data was analyzed using a GEE model, specifying a logit link, with factors for treatment and study, stratified on investigator with an exchangable provisional within investigator correlation. Note this consistent with the model

$$\text{Logit(success)} = \beta_0 + \beta_1 \text{ treatment} + \beta_2 \text{ study} + \beta_3 \text{ study by treat} + \text{investigator}$$

with investigator a random effect.

The following table of "Wald" statistics results:

Analysis of GEE Parameter Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	-3.0891	0.6205	-4.3051	-1.8730	-4.98	<.0001
Treatment	1.8948	0.6605	0.6003	3.1893	2.87	0.0041
study	1.1200	0.8369	-0.5203	2.7604	1.34	0.1808
Treat*study	-1.4319	0.9081	-3.2117	0.3480	-1.58	0.1148

A common adage that applies here is that absence of proof is not proof of absence. However, the statistically non-significant study effect and treatment by study effect is at least consistent with the hypothesis of no differential study effect. These estimates were derived from the pooled LOCF Week 4 data using the SAS GENMOD procedure as follows:

```
PROC GENMOD DESCENDING;
CLASS treat study invnum;
model success = treat study treat*study / d=bin;
repeated subject=invnum(study) / type=exch;
```

For an alternative Bayesian approach, the same model was fit in WINBUGS 1.4 using vague priors for the parameters and the variance of the investigator effect. For identification purposes the investigator effects were centered within each study. One way to assess effects is to compare models using the Deviance Information Criterion of Spiegelhalter et al (2002, *JRSS-B*, pg 583-640). The model with intercept, treatment effects, and centered investigator effect has DIC=266.302. The full model adding study effects above has DIC=266.454. So effectively ignoring study effects results in a slightly better model. This would be consistent with the notion that there is no large study effect.

Appendix 5. Patient Disposition and Demographics for the Pivotal Trials

Patient Disposition

A total of 148 subjects from 12 study centers were enrolled and randomized into the study to receive either Clobetasol Propionate Shampoo, 0.05% or Clobetasol Propionate ShampooVehicle.

Table A.5.1 Study 18075 Patient Disposition

Disposition	Clobetasol Propionate Shampoo, 0.05% N (%)	Clobetasol Proprionate Shampoo Vehicle N (%)	Total N (%)
Enrolled/Randomized	99 (100)	49 (100)	148 (100)
ITT population	99 (100)	49 (100)	148 (100)
PP population	91 (91.9)	43 (87.8)	134 (90.5)
Completed study	91 (91.9)	45 (91.8)	136 (91.6)
Discontinued	8 (8.1)	4 (8.2)	12 (8.1)
Lack of efficacy	0 (0)	1 (2.0)	1 (0.7)
Adverse event	2 (2.0)	1 (2.0)	3 (2.0)
Subject request	3 (3.0)	2 (4.1)	5 (3.4)
Protocol violation	1 (1.0)	0 (0)	1 (0.7)
Lost to follow-up	2 (2.0)	0 (0)	2 (1.4)

The Sponsor reported that among the 148 randomized subjects, 14 were found to have violated the protocol after receiving study medication and were not included in the PP (per protocol) population. Reasons included missing 2 or more visits (6 active and 1 vehicle subjects), receiving prohibited concomitant medication (2 active and 3 vehicle subjects), no post-baseline data (2 vehicle subjects).

Table A.5.2 below presents the patient demographic information in Study 18075.

Table A.5.2 Study 18075 Demographics

N (%)	Clobetasol Proprionate Shampoo 0.05% N=99 (67%)	Clobebetasol Propionate Shampoo Vehicle N=49(33%)	Total N=148(100%)
Gender Male	46 (46.5)	20 (40.8)	66 (44.6)
Female	53 (53.5)	29 (59.2)	82 (55.4)
Race Caucasian	85 (85.9)	45 (91.8)	130 (87.8)
Black	3 (3.0)	1 (2.0)	4 (2.7)
Asian	2 (2.0)	0 (0)	2 (1.4)
Hispanic	8 (8.1)	3 (6.1)	11 (7.4)
Other	1 (1.0)	0 (0)	1 (0.7)
Age groups			
12 to 17 years	3 (3.0)	3 (6.1)	6 (4.1)
18 to 64 years	80 (80.8)	37 (75.5)	117 (79.1)
≥65 years	16 (16.2)	9 (18.4)	25 (16.9)
Age Mean (SD)	47.1 (16.4)	46.4 (18.5)	46.9

Appendix 5. (cont.) Patient Disposition and Demographics for the Pivotal Trials

Table A.5.3 below presents the patient disposition in Study 18076.

Table A.5.3 Study 18076 Patient Disposition

Disposition	Clobetasol Propionate Shampoo, 0.05% N=99(67%)	Clobetasol Proprionate Shampoo Vehicle N=49(33%)	Total
Enrolled/Randomized	95 (100)	47 (100)	142 (100)
ITT population	95 (100)	47 (100)	142 (100)
PP population	84 (88.4)	42 (89.4)	126 (88.7)
Completed study	88 (92.6)	44 (93.6)	132 (93.0)
Discontinued	7 (7.4)	3 (6.4)	10 (7.0)
Lack of efficacy	0 (0)	0 (0)	0 (0)
Adverse event	1 (1.1)	0 (0)	1 (0.7)
Subject request	3 (3.2)	2 (4.3)	5 (3.5)
Protocol violation	2 (2.1)	1 (2.1)	3 (2.1)
Lost to follow-up	1 (1.1)	0 (0)	1 (0.7)

Among the 142 randomized subjects, 16 were found to have violated the protocol after receiving study medication and were not included in the PP (per protocol) population. Reasons included not meeting inclusion/exclusion criteria (2 active and 1 vehicle subjects), missing 2 or more visits (2 active subjects), missing doses for 5 or more consecutive days (1 active and 2 vehicle subjects), receiving prohibited concomitant medication (4 active and 1 vehicle subjects), no post-baseline data (1 active and 1 vehicle subject), and being discontinued from the study due to protocol violation per Investigator's judgment (1 active subject). Hence 126 subjects were included in the Per Protocol population.

Table A.5.4 below presents the patient demographic information in Study 18076.

Table A.5.4 Study 18076 Demographics

N (%)	Clobetasol Proprionate Shampoo 0.05% N=99 (67%)	Clobebetasol Propionate Shampoo Vehicle N=49(33%)	Total N=148(100%)
Gender Male	38 (40.0)	22 (46.8)	60 (42.3)
Female	57 (60.0)	25 (53.2)	82 (57.7)
Race Caucasian	88 (92.6)	43 (91.5)	131 (92.3)
Black	2 (2.1)	1 (2.1)	3 (2.1)
Asian	0 (0)	0 (0)	0 (0)
Hispanic	4 (4.2)	3 (6.4)	7 (4.9)
Other	1 (1.1)	0 (0)	1 (0.7)
Age 12 to 17 years	2 (2.1)	2 (4.3)	4 (2.8)
18 to 64 years	82 (86.3)	39 (83.0)	121 (85.2)
≥65 years	11 (11.6)	6 (12.8)	17 (12.0)
Age Mean (SD)	45.1 (15.3)	45.1 (15.7)	45.1

Thus, in both studies, the Sponsor's age, race and gender groups appear to be reasonable equitably distributed between the active and vehicle groups. However, it is apparent that too few Asian and Black patients were included to make any reliable conclusions concerning these subgroups.

Appendix 6. Non-pivotal Phase 3 Trials for Scalp Psoriasis

All three supportive trials were conducted in Europe, with generally similar protocols. Further, all three protocols are also similar to the protocols of the two pivotal studies, 18075 and 18076. In all studies, treatment was to last for up to four weeks. Subjects were evaluated at baseline and weeks 2 and 4.

As with the pivotal studies, secondary endpoints included erythema, scaling/ desquamation, plaque thickening, and pruritus, each scored 0(none) to 3(severe). The Total Severity Score (TSS) is the sum of the individual scores for erythema, scaling and plaque thickening. The Global Severity Score (GSS) utilizes a static 6-point scale, scored 0(none) to 5(very severe), similar to the GSS in the pivotal studies, but with slightly different descriptors. The primary efficacy endpoints specified in the protocols are the Week 4 Global Severity Score and Total Symptom Score. Other secondary endpoints were the subject's global assessment of improvement.

Note that the Medical team has questioned the utility of the TSS, and, unlike the pivotal studies the GSS is not dichotomized. However, these are the primary endpoints specified in the protocols. Both endpoints were analyzed using ANCOVA's with baseline as a covariate and center and treatment as factors. The Sponsor noted that in all cases treatment by center interactions were investigated, but they proved to be ignorable. Thus treatment comparisons are based on adjusted means, adjusted for center and baseline. For the both the TSS and the GSS, the smaller numbers are more favorable. Since the treatment difference is taken as Clobetasol - Comparator negative differences tend to be more favorable to the Sponsor's claims.

Study 2638

Parallel group comparison of 4-week treatment with Clobetasol 17-propionate 0.05% Shampoo versus Calcipotriol solution 0.005% (Dovonex/Daivonex™) – An efficacy and safety study in subjects with scalp psoriasis

The objectives of the study were to compare the efficacy and safety of Clobetasol Propionate Shampoo, 0.05% and calcipotriol 0.005% solution (Dovonex/Daivonex™) in subjects with moderate to severe scalp psoriasis, and to show the non-inferiority or superiority in efficacy of clobetasol versus the comparator product. The study was a randomized, multi-center, investigator-blinded, active-controlled comparison of two parallel groups. One hundred and fifty-one subjects aged 12 years or older at 14 study centers in Western Europe were randomized in a 1:1 ratio to receive either Clobetasol Propionate Shampoo, 0.05% once daily or Daivonex™ solution twice daily for 4 weeks.

Appendix 6. (cont.) Non-pivotal Phase 3 Trials for Scalp Psoriasis

The Sponsor specified a noninferiority bound of 1.5 for the TSS, stating that this would correspond to a roughly 20% change in expected baseline score. However the observed baseline score was 5, so a more appropriate 20% bound would be a 1 unit change. In addition, the Division has generally preferred a 10% bound, corresponding to a 0.5 unit change. For the GSS the baseline mean score is about 3.5, so a 20% margin would be 0.70, and a 10% margin would 0.35. Thus, if the 95% confidence interval is less than 0.35 (for 10% margin) or 0.70 (for 20% margin) we would accept the hypothesis of non-inferiority.

Table A.6.1 Study 2638 Global Severity Scores (Week 4)

Population	Treatment	N	Mean (SD)	Difference	Confidence Interval[1]	p-value[2]
ITT	Clobetasol	76	1.55 (1.20)	-0.43	-0.78, -0.08	0.016
	Calcipotriol	75	2.03 (1.31)			
PP	Clobetasol	67	1.42 (1.09)	-0.27	-0.59, 0.06	0.114
	Calcipotriol	61	1.74 (1.17)			

[1] 95% confidence interval of difference (Clobetasol-Calcipotriol)
[2] ANCOVA test of differences

Table A.6.2 Study 2638 Total Severity Scores (Week 4)

Population	Treatment	N	Mean (SD)	Difference	Confidence Interval[1]	p-value[2]
ITT	Clobetasol	76	1.76 (1.57)	-0.51	-0.97, -0.05	0.028
	Calcipotriol	75	2.36 (1.64)			
PP	Clobetasol	67	1.64 (1.49)	-0.24	-0.66, 0.18	0.267
	Calcipotriol	61	1.94 (1.35)			

[1] 95% confidence interval of difference (Clobetasol-Calcipotriol)
[2] ANCOVA test of differences

So whether we use the 10% margin generally recommended by the Division or the 20% margin specified by the Sponsor, we would accept noninferiority of clobetasol to Calcipotriol for both endpoints for both populations. For the ITT population clobetasol is statistically significantly better than Calcipotriol for both endpoints. While trends are somewhat similar for the PP population, differences are not statistically significant.

Study Number 2648

The Safety and Efficacy of Clobetasol Propionate Shampoo, 0.05% compared to Polytar Liquid® in the treatment of scalp psoriasis

The study objectives were to compare the efficacy of Clobetasol Propionate Shampoo, 0.05% versus the marketed topical product Polytar Liquid® in subjects with moderate to severe scalp psoriasis. This was another randomized, multi-center, investigator-blinded, parallel-group, and active-controlled study. One hundred and sixty-two subjects ages 18 years or older in 22

Appendix 6. (cont.) Non-pivotal Phase 3 Trials for Scalp Psoriasis

centers in Great Britain were randomized in a 3:1 ratio to receive either Clobetasol Propionate Shampoo, 0.05% once daily or Polytar Liquid® twice weekly for 4 weeks.

Table A.6.3 Study 2648 Global Severity Scores (Week 4)

Population	Treatment	N	Mean (SD)	Difference	Confidence Interval[1]	p-value[2]
ITT	Clobetasol	121	1.9 (1.0)	-1.010	-1.357, -0.663	0.0001
	Polytar Liquid®	41	3.0 (1.0)			
PP	Clobetasol	105	1.9 (1.0)	-1.126	-1.494, -0.758	0.0001
	Polytar Liquid®	32	3.0 (1.0)			

[1] 95% confidence interval of difference (Clobetasol-Polytar)
[2] ANCOVA test of differences

Table A.6.4 Study 2648 Total Severity Scores (Week 4)

Population	Treatment	N	Mean (SD)	Difference	Confidence Interval[1]	p-value[2]
ITT	Clobetasol	121	3.2 (2.0)	-1.842	-2.475, -1.208	0.0001
	Polytar Liquid®	41	5.2 (1.9)			
PP	Clobetasol	105	3.1 (1.9)	-2.066	-2.727, -1.405	0.0001
	Polytar Liquid®	32	5.3 (1.9)			

[1] 95% confidence interval of difference (Clobetasol-Polytar)
[2] ANCOVA test of differences

Clearly for both populations, for both endpoints, both noninferiority and statistically significant superiority of Clobetasol Propionate Shampoo over Polytar Liquid is accepted.

Study 2665

Title: Efficacy and Safety of Clobetasol Propionate Shampoo, 0.05% as compared to it's Vehicle and Clobetasol Propionate 0.05% Gel (Dermoval™ Gel) in the Treatment of Subjects with Scalp Psoriasis

The study objectives were to demonstrate the non-inferiority in efficacy of Clobetasol Propionate Shampoo, 0.05% versus Dermoval™ Gel and superior efficacy of Clobetasol Propionate Shampoo, 0.05% versus Vehicle Shampoo, and to provide safety data to support the registration of the drug on a worldwide basis. This study was randomized, multi-center, investigator-blinded, parallel-group, and active- and vehicle-controlled. One hundred and forty subjects ages 18 years in several European countries were randomized in a 3:3:1 ratio to receive either Clobetasol Propionate Shampoo, 0.05%, Dermoval™ Gel, or vehicle shampoo once daily for 4 weeks.

Appendix 6. (cont.) Non-pivotal Phase 3 Trials for Scalp Psoriasis

Table A.6.5 Study 2665 Global Severity Scores (Week 4)

Population	Treatment	N	Mean (SD)	Difference	Confidence Interval[1]	p-value[2]
ITT	Clobetasol	63	1.7 (1.3)			
	Dermoval	61	1.1 (1.0)	0.55	0.21, 0.90	0.002
	Vehicle	20	2.4 (1.2)	-0.64	-1.14, -0.14	0.012
PP	Clobetasol	57	1.8 (1.7)			
	Dermoval	55	1.1 (1.3)	0.55	0.21, 0.90	0.002
	Vehicle	16	2.4 (1.6)	-0.48	-1.00, 0.05	0.074

[1] 95% confidence interval of difference (Clobetasol-Comparator) where comparator is the listed drug.
[2] ANCOVA test of differences

Table A.6.6 Study 2665 Total Severity Scores (Week 4)

Population	Treatment	N	Mean (SD)	Difference	Confidence Interval[1]	p-value[2]
ITT	Clobetasol	63	2.0 (1.9)			
	Dermoval	61	1.2 (1.6)	0.79	0.24, 1.34	0.005
	Vehicle	20	2.9 (2.0)	-0.97	-1.76, -0.18	0.016
PP	Clobetasol	57	1.8 (1.7)			
	Dermoval	55	1.1 (1.3)	0.77	0.25, 1.29	0.004
	Vehicle	16	2.4 (1.6)	-0.70	-1.48, 0.08	0.078

[1] 95% confidence interval of difference (Clobetasol-Comparator) where comparator is the listed drug.
[2] ANCOVA test of differences

Thus note that noninferiority of clobetasol to Dermoval is not established. In fact, Dermoval is statistically significantly better than clobetasol for both endpoints. However, Clobetasol Propionate Shampoo is statistically significantly better than vehicle.

APPEARS THIS WAY
ON ORIGINAL

Appendix 7. A Preliminary Bayesian Analysis

For the pivotal studies, a score of 0 or 1, i.e., clear or minimal, on the Global Severity Scale was defined as a treatment success. A simple Bayesian analysis of this endpoint in the ITT population was initiated. In each study the logit of success was modeled as with an intercept, a treatment effect, a random center effect, and a random interaction. The intercept and treatment effect was modeled with a normal prior, with a large variance. The variances of the random terms were modeled with gamma variances. As the names suggest the no interaction model deletes the interaction term while the no treatment or interaction model deletes both the random interaction and the treatment term.

Deviance Information Criteria (DIC)	Study	
	18075	18076
Full Model	141.13	125.74
No Interaction	142.75	124.67
No Treatment or Interaction	141.70	133.33

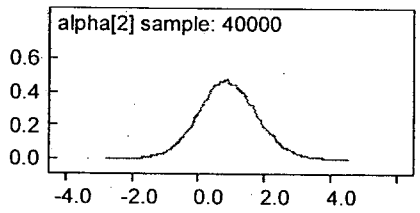
Since for a given data set smaller DICs are associated with better models, the DIC's for these studies suggest the full model is preferred in the 18075 study and the no interaction model in the 18076 study.

For Study 18075, fitting the full model in WINBUGS 1.4 using the program below gives the following summary statistics for the posterior distributions:

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
alpha[1]	-2.732	0.8336	0.01654	-4.43	-2.708	-1.147	4001	44000
alpha[2]	0.9177	0.8817	0.01671	-0.7664	0.8934	2.713	4001	44000
prob	0.8564	0.3507	0.004985	0.0	1.0	1.0	4001	44000
sigma1	1.652	0.5466	0.003832	0.7722	1.586	2.892	4001	44000
sigma2	1.751	0.5523	0.004687	0.8466	1.692	2.983	4001	44000

Note that alpha[2] is the differential effect of clobetasol on the probability of success, and that prob is the estimate of the posterior probability that the effect is positive.

A somewhat smoothed estimate of the posterior distribution of treatment effect is as follows:
Distribution of estimate of treatment effect.



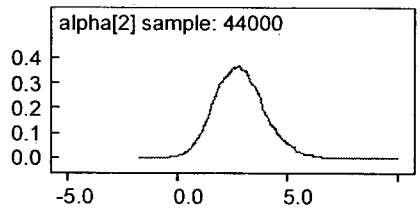
Appendix 7. (cont.) A Bayesian Analysis

So the posterior probability of a positive treatment effect in Study 18075 is approximately 0.856.

In Study 18076, fitting the full model in WINBUGS 1.4 gives the following summary statistics for the posterior distributions:

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
alpha[1]	-4.393	1.097	0.03031	-6.75	-4.325	-2.461	4001	44000
alpha[2]	2.838	1.118	0.03006	0.816	2.784	5.191	4001	44000
prob	0.9975	0.0499	5.475E-4	1.0	1.0	1.0	4001	44000
sigma1	1.741	0.5689	0.004245	0.817	1.673	3.051	4001	44000
sigma2	1.709	0.5488	0.005013	0.8086	1.65	2.946	4001	44000

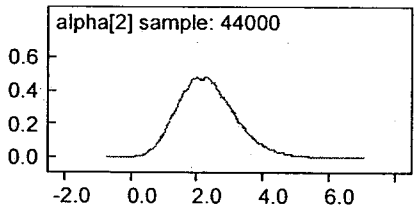
with posterior:



Fitting the sub-model with an intercept, a treatment effect, and a random center effect gives the following. Again, the intercept and treatment effects were modeled with a normal prior having a large variance.

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
alpha[1]	-3.788	0.9138	0.01999	-5.75	-3.716	-2.213	4001	44000
alpha[2]	2.346	0.8646	0.01837	0.8683	2.277	4.244	4001	44000
prob	0.9997	0.0178	1.35E-4	1.0	1.0	1.0	4001	44000
sigma	1.651	0.5385	0.003742	0.7728	1.592	2.866	4001	44000

The posterior distribution of treatment effect:



Appendix 7. (cont.) A Bayesian Analysis

So for either model, in Study 18076 the posterior probability of a positive treatment effect is well above 0.99. The probability of a positive treatment effect in Study 18075 is somewhat more problematical (i.e., estimated posterior probability of a positive effect = 0.856).

WINBUGS 1.4 Program:

```
model {
   for ( i in 1:N ) {
       treat[i]<- equals(nt[i],1)
       succ[i]~dbern(p[i])
       logit(p[i]) <- mu[i]
       mu[i]<- alpha[1] + alpha[2]*treat[i] + center[nc[i]] +
          intactn[nc[i],nt[i]]
                    }
   for (k in 1:c ) {
       center[k]~dnorm(0.0,tau1)
       intactn[k,1]~dnorm(0.0,tau2)
       intactn[k,2]~dnorm(0.0,tau2)
                  }
   for (m in 1:2 ) {
       alpha[m]~dnorm(0.0,0.001)
                  }
   sigma1~dgamma(10,5)
   sigma2~dgamma(10,5)
   tau1<-1/sigma1
   tau2<-1/sigma2
   prob<-step(alpha[2])
     }
inits
list(alpha=c(0,0),sigma1=1,sigma2=1)
list(alpha=c(1,-1),sigma1=0.5,sigma2=3.0)
list(alpha=c(-1,1),sigma1=3.0,sigma2=0.5)
data
 list(N=148,c=10)
 nc[ ] nt[ ] succ[ ]
  1     2     0
  1     1     0
     . . .
 10     1     0
END
```

Note that, due to time contraints, this was only a preliminary model. A more complete analysis would assess model specification and more complicated models. A traditional approach to compare nested models would be to assess them using Bayes factors. However, for these models direct estimation of Bayes factors using several techniques led to computational overflows and underflows.

SIGNATURES/DISTRIBUTION LIST PAGE

Steve Thomson
Mathematical Statistician, Biometrics III

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

cc:
Archival NDA 21644

HFD-540/Dr. Lindstrom
HFD-540/Dr. Wilkin
HFD-540/Dr. Luke
HFD-540/Ms. Smith
HFD-725/Dr. Alosh
HFD-725/Dr. Huque
HFD-725/Mr. Thomson
HFD-725/Dr. Anello
HFD-725/Dr. Dubey
HFD-725/Dr. Bhore
Chron.

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

/s/

Steven Thomson
1/12/04 04:01:43 PM
BIOMETRICS

Mohamed Alosh
1/12/04 04:28:57 PM
BIOMETRICS
Concur with review