Study Visit Schedule:

TABLE 10.1.2.1: SHOWING EVALUATION AND VISIT SCHEDULE FOR STUDY VOSG-PE-315

Examination	Phase	Screening Washout Period	Randomization / Baseline	Treatment Period				Final Visit
	Weeks	-1	0	1	2	4	6	8
	Days	-7 to -5	1	8	15	29	43	57
	Visit	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Written informed consent		X						
Background /Medical History		X						
Examination of the hands (X-ray if not already available)		X						
Prior/Concomitant medications and/or significant non-drug therapies		X	X	X	X	X	X	X
Hematology and blood chemistry		X						X
Urine pregnancy test			X					X
Inclusion/exclusion criteria		X	X					
Physical exam and vital signs			X					X
Diagram of affected joints (clinical assessment)		X	X		X	X	X	X
OA pain on VAS [1], global rating of disease activity, and AUSCAN [1,2]		X	X	X	X	X	X	X
FIHOA			X		X	X	X	
Global rating of benefit				X	X	X	X	X
Global rating of efficacy								X
Dispense study medication			X [3]		X	X	X	
Dispense rescue medication		X	X	X	X	X	X	
Dispense diary		X	X		X	X	X	
Review and collection of diary			X	X [4]	X	X	X	X
Accountability of study medications					X	X	X	X
Check consumption of rescue medication			X	X	X	X	X	X
AEs			X	X	X	X	X	X

[1] Separate assessments for the right hand and for the left hand.

[2] For the non-dominant hand, only pain and stiffness subscales were assessed.

[3] Investigator instructed subject in proper application of the study medication and supervised application.

[4] Only a review of the diary.

Statistical Analysis Plan and Definition of Analyzed Study Populations:

Handling of treatment failures

A treatment failure was designated as such:

If there was a series of 4 or more consecutive days, (starting after Day 7), in which a patient took either:

(a) At least 2 grams acetaminophen (paracetamol),

or

(b) At least half the maximum daily over the counter (OTC) dose of a NSAID

or

(c) One or more single prescription strength doses of a nonselective or COX-2 selective NSAID, specifically to treat hand OA pain.

No imputation of missing diary data was used for this purpose except for missing doses of rescue medication.

Sensitivity analyses:
Sensitivity analyses were to have been conducted in which imputation was done only for the visit immediately following the designation as *treatment failure*. For the sensitivity analysis of daily assessment of pain, only assessments within the period for which the definition of treatment failure was satisfied were replaced by imputation using LOCF.

Non-missing assessments after this period were not replaced.

An additional sensitivity analysis was conducted with a definition of treatment failure requiring:

(a) 4 grams paracetamol (acetaminophen)
or
(b) The full daily dose of an OTC NSAID
or
(c) A single dose of NSAIDs as above, daily for 4 consecutive days to treat the pain of hand OA.

Sensitivity analyses were to have been conducted on the primary outcomes in the final study model to assess the impact of a variety of issues. These included the following:
1) The impact of imputing by LOCF for early termination

a) At each Visit X, missing values in the diclofenac group due to early termination were imputed (replaced) by the mean of all non-missing values in the diclofenac group, and correspondingly for vehicle.
b) At each Visit X, missing values in the diclofenac group due to early termination were imputed by the mean of all non-missing values in the vehicle group, and vice versa.

2) The impact of treatment failures

a) Their post-failure efficacy data were imputed only at the immediately following visit rather than at all subsequent visits.
b) The definition of treatment failure was changed requiring (a) 4 grams paracetamol (acetaminophen) or (b) the full daily dose of an OTC NSAID, or (c) a single dose of NSAIDs, daily for 4 consecutive days to treat the pain of hand OA.

3) The impact of patients who did not stay in the study long enough to supply efficacy assessments at the Week 1 visit – the analyses were rerun with these patients excluded.

Protocol Amendments:

The study protocol VOSG-PE-315 (dated 18 February 2005) was amended once. This amendment was implemented prior to any site starting recruitment.

Amendment 1 (23 March 2005)
The visit schedule in the protocol was modified to reflect the lack of a FIHOA index assessment at the final visit. It was also to have been specified that subjects mark their response on the OA pain intensity, global rating of disease activity, and global rating of benefit VAS using an " X " rather than a vertical line (|). The text concerning the laboratory evaluations was to be modified to include erythrocyte sedimentation rate and C-reactive protein and to clarify the terminology concerned with existing parameters. In addition, the requirement for investigators to weigh returned tubes was to have been deleted.

Changes to the planned statistical analysis were to have been effected before the study was unblinded:

- The protocol specified that the statistical analysis was to be stratified by a 3-category classification of OA category: (1) only CMC-1 is painful, (2) mixed OA, (3) painful joints do not include CMC-1. When the clinical phase of the study was completed it was found that only 33 subjects qualified in category (1). To prevent the complications that arises in the statistical analysis when one category is much smaller than the others, categories (1) and (2) were to be pooled.
- Before the study was unblinded, the time to achieve OA pain intensity ≤ 20 mm in the target hand was to have been added as a secondary outcome, because it was noted that only about 20% of the ITT population achieved OA pain intensity of ≤ 10 mm by the end of the study.
- According to the protocol, ***missing baseline assessments*** for a subset of the efficacy outcomes were to have been imputed from a regression equation in which the efficacy assessments that the subject did provide at the baseline visit were the predictors. Following discussions with the FDA, instead of this regression approach ***the mean values of the missing assessments*** over *all subjects in the ITT efficacy population with non-missing values at the baseline visit* were to have been used to impute missing baseline assessments.
- The procedure for imputing individual questions that were not answered for the AUSCAN and FIHOA at the screening and baseline visits was to have been extended to cover all visits.
- Based on the experiences of two preceding studies on knee OA pain, a primary efficacy visit was to have been discounted if a subject stopped dosing with study medication two or more days before the primary efficacy visit or took certain disallowed concomitant medication on the day of the primary efficacy visit for any reason other than OA pain in the hands.
- Exclusion of data from certain evaluations when the baseline score indicated little or

no room for improvement upon treatment.

RESULTS

- **Disposition**

Of a total of 809 patients who were screened, 385 were randomized.
The most common reasons for screened subjects not being randomized were as follows:

- did not have the protocol-specified posterior-anterior X-ray of the dominant hand showing signs of OA in the same painful joints with Kellgren-Lawrence grade 1, 2, or 3 disease in the dominant hand (71 subjects).
- withdrawal of consent (46 subjects).
- pain in the non-dominant hand during the 24 hours before baseline assessment was not at least 20 mm lower (on a 100 mm VAS) than the corresponding rating in the target hand (41 subjects).

The proportions of subjects in the DSG and vehicle groups who completed the study were 87.4% and 86.1%, respectively. The proportion of subjects who prematurely discontinued the study in the DSG and vehicle groups were 12.6% (25 subjects) and 13.9% (26 subjects), respectively. The most common reasons for discontinuing the study were AEs (DSG: 5.1%; vehicle: 2.1%), unsatisfactory therapeutic effect (DSG: 4.0%; vehicle: 7.0%), and withdrawal of consent (DSG: 2.0%; vehicle: 3.2%).

The case report forms for the patients who discontinued due to withdrawal of consent were reviewed to ascertain whether these patients actually discontinued due to an adverse event or lack of efficacy; this review found that one vehicle randomized subject withdrew because of efficacy, and no study subjects in either arm withdrew because of adverse event.

Please refer to Table 10.1.2.2 noted below.

TABLE 10.1.2.2: SHOWING SUBJECT DISPOSITION BY TREATMENT GROUP; STUDY-315

	DSG	Vehicle
Total number of subjects		
Screened	809	
Randomized	198	187
Completed – n (%)	173 (87.4)	161 (86.1)
Discontinuations – n (%)		
Total	25 (12.6)	26 (13.9)
AEs	10 (5.1)	4 (2.1)
Unsatisfactory therapeutic effect	8 (4.0)	13 (7.0)
Protocol deviation	1 (0.5)	1 (0.5)
Subject withdrew consent	4 (2.0)	6 (3.2)
Lost to follow-up	2 (1.0)	1 (0.5)
Administrative problems	0	1 (0.5)

Note: % are relative to the total number of subjects randomized.
Source: Post-text Table 7.1; Appendix 7, Listing 7.3

- **Protocol Deviations**

Similar proportions of subjects in each treatment group had at least one protocol violation; 23.2% (46 subjects) in the DSG group and 23.5% (44 subjects) in the vehicle group. Of these, 11 (5.6%) and 12 (6.4%) were entry violations, and 6 (3%) and 6 (3.2%) were dosing violations, and 28 (14.1) and 28 (15.0) were concomitant medication violations occurring in the DSG and vehicle arms respectively. (See Table 10.2.1 noted below).
The most common protocol violation in each treatment group was the failure to discontinue rescue medication 36 hours before the baseline or the primary efficacy visits (27 subjects at Visit 2, 15 subjects at Visit 5, and 12 subjects at Visit 6).

TABLE 10.1.2.3: SHOWING SUMMARY OF PROTOCOL VIOLATIONS AND TYPES OF VIOLATIONS IN STUDY-315

	DSG, n (%)	Vehicle, n (%)
Total randomized	198	187
Total with protocol violations	46 (23.2)	44 (23.5)
Entry violations	11 (5.6)	12 (6.4)
Dosing violations	6 (3)	6 (3.2)
Visit violations	5 (2.5)	11(5.9)
Concomitant medications	28 (14.1)	28 (15.0)

- **Baseline**

Baseline hand OA assessments in the all randomized subject population are summarized in Table 10.1.2.4. Mean baseline values of the various assessments in the target hand were consistent with moderate pain and disability, e.g. mean OA pain values were 73.6 and mean total AUSCAN scores were between 66 and 68 in both treatment groups. Mean baseline values in the non-dominant hand were consistent with mild pain, mean OA pain values were approximately 30 in both treatment groups.
The difference between the treatment groups in the AUSCAN stiffness index in the non-dominant hand achieved significance ($p = 0.05$ by CMH Chi-squared test of treatment means). None of the other measures between treatment groups in the dominant or non-dominant hand were statistically significantly different.

TABLE 10.1.2.4: SHOWING BASELINE HAND OA ASSESSMENTS- ITT POPULATION – STUDY -315 – IN THE DOMINANT AND NON-DOMINANT HAND

	DSG (N = 198)	Vehicle (N = 187)
Global rating of disease [a]		
Mean ± SD	57.6 ± 19.0	56.5 ± 19.9
Range	5 to 97	9 to 97
Target (dominant) hand		
OA pain intensity [b]		
Mean ± SD	73.6 ± 15.6	73.6 ± 14.2
Range	40 to 100	41 to 100
Total AUSCAN index [c]		
Mean ± SD	67.2 ± 17.4	66.7 ± 16.8
Range	13 to 96	10 to 98
AUSCAN pain index [c]		
Mean ± SD	66.3 ± 17.9	66.8 ± 16.2
Range	12 to 98	11 to 99
AUSCAN stiffness index [c]		
Mean ± SD	66.0 ± 22.8	66.6 ± 23.9
Range	1 to 98	4 to 100
AUSCAN physical function index [c]		
Mean ± SD	67.9 ± 18.8	66.7 ± 18.4
Range	9 to 99	8 to 99
Non-dominant hand		
OA pain intensity [b]		
Mean ± SD	27.8 ± 17.9	30.2 ± 18.2
Range	1 to 77	0 to 78
AUSCAN pain index[c]		
Mean ± SD	31.3 ± 19.7	33.9 ± 20.3
Range	0 to 98	0 to 94
AUSCAN stiffness index[c]		
Mean ± SD	32.4 ± 22.8	37.1 ± 23.7
Range	0 to 98	1 to 98
FIHOA index[d]		
Mean ± SD	12.8 ± 4.4	12.5 ± 4.6
Range	1 to 22	1 to 25

[a] 100 mm visual analogue scale: 0 = very good, 100 = very poor.
[b] 100 mm visual analogue scale: 0 = no pain, 100 = unbearable pain.
[c] Average over multiple questions: 0 = no pain / stiffness / difficulty, 100 = extreme pain / stiffness / difficulty.
[d] FIHOA index: 0-30.
Source: Post-text table 7.10; Appendix 7, Listing 9.1, Listing 9.4 and Listing 9.6

PRIMARY EFFICACY RESULTS

Applicant's efficacy findings

DSG was superior to vehicle in all three co-primary efficacy outcomes (OA pain intensity, total AUSCAN index, and global rating of disease).

At Week 4, differences between the DSG and vehicle groups were statistically significant for OA pain intensity ($p = 0.018$) and total AUSCAN ($p = 0.011$), and separation on the global rating of disease activity was borderline significant ($p = 0.06$). Both Week 4 and Week 6 were designated as primary endpoints, but testing at Week 6 was conditional on positive primary outcomes at Week 4.

At Week 6, differences between the DSG and vehicle groups were statistically significant for OA pain intensity ($p = 0.023$), total AUSCAN ($p = 0.006$), and global rating of disease activity ($p = 0.023$). The treatment effect in particular with respect to the total AUSCAN index (difference DSG – vehicle) 6.3 mm and 7.1 mm at Week 4 and 6, respectively). Please refer to Table 10.2.2 shown below.

TABLE 10.1.2.4: SHOWING PRIMARY EFFICACY OUTCOMES AT WEEK 4 AND WEEK 6 IN STUDY VOSG-PE-315

	DSG (N = 198)	Vehicle (N = 187)	p-value
Week 4			
OA pain intensity (100 mm VAS) [a]			
Mean	42.6	49.7	
Mean change from baseline	31.1	23.9	0.018
Total AUSCAN (100 mm VAS) [b]			
Mean	43.7	50.2	
Mean change from baseline	23.5	16.8	0.011
Global rating of disease (100 mm VAS) [c]			
Mean	37.5	41.9	
Mean change from baseline	20.8	14.8	0.06
Week 6			
OA pain intensity (100 mm VAS) [a]			
Mean	39.9	46.9	
Mean change from baseline	33.7	26.7	0.023
AUSCAN (100 mm VAS) [b]			
Mean	41.4	48.5	
Mean change from baseline	25.9	18.6	0.006
Global rating of disease (100 mm VAS) [c]			
Mean	35.2	40.4	
Mean change from baseline	23.1	16.3	0.023

[a] 0 = no pain, 100 = unbearable pain.
[b] Calculated as the unweighted mean over all 15 questions (0 = no pain / stiffness / difficulty, 100 = extreme pain / stiffness / difficulty).
[c] 0 = very good, 100 = very poor.
Change from baseline was calculated as baseline minus post-baseline score. Analysis was analysis of covariance (ANCOVA) with main effects of treatment, center, hand OA category and baseline covariate.
Source: Post-text Table 9.2, Post-text Table 9.3, and Post-text Table 9.4

SECONDARY EFFICACY RESULTS – STUDY -315

Secondary outcomes were analyzed only in the ITT efficacy population. Comparisons of every measurement at all post-baseline visits favored DSG treatment, with statistically significant superiority demonstrated at most weeks for numerous outcomes.

Secondary measures of efficacy showed greater improvement in OA symptoms for the DSG group, and the differences in comparison with the vehicle group were frequently statistically significant. The proportion of subjects in the DSG group who rated treatment as very good or excellent was 47.7%, compared with 36.5% in the vehicle group (p = 8). (See Table 10.1.2.5 in three parts noted below).

In general, the data demonstrate an improvement in DSG-treated subjects compared to vehicle treated subjects improvements increased over time but did not reach statistical significance except in 2 instances.

TABLE 10.1.2.5 SHOWING SECONDARY EFFICACY OUTCOME MEASURES IN THE TARGET HAND –STUDY-315

Secondary efficacy outcomes (ITT population)	DSG N = 198	Vehicle N = 187	
	Mean ± SD	Mean ± SD	p-value
OA Pain intensity (100 mm VAS)			
Baseline	73.6 ± 15.6	73.6 ± 14.2	>0.99
Week 1	50.1 ± 26.3	55.0 ± 25.3	0.031
Week 2	46.6 ± 27.3	52.7 ± 26.3	0.013
Week 4	42.6 ± 30.5	49.7 ± 28.8	0.018
Week 6	39.9 ± 31.6	46.9 ± 29.9	0.023
Week 8	38.1 ± 32.7	44.0 ± 30.9	0.06
OA Pain intensity daily diary (100 mm VAS)			
Week 1 average	53.7 ± 21.4	55.5 ± 20.5	0.47
Week 2 average	47.4 ± 25.2	52.4 ± 23.3	0.018
Total AUSCAN index (Scale 0-100)			
Baseline	67.2 ± 17.4	66.7 ± 16.8	0.77
Week 1	50.6 ± 23.9	55.6 ± 23.0	0.007
Week 2	47.8 ± 24.9	52.3 ± 24.5	0.027
Week 4	43.7 ± 28.2	50.2 ± 27.3	0.011
Week 6	41.4 ± 28.8	48.5 ± 28.1	0.006
Week 8	40.5 ± 29.9	46.5 ± 28.7	0.028
AUSCAN pain index (Scale 0-100)			
Baseline	66.3 ± 17.9	66.8 ± 16.2	0.81
Week 1	48.7 ± 25.0	53.4 ± 23.5	0.022
Week 2	45.7 ± 26.0	51.1 ± 24.9	0.020
Week 4	42.2 ± 28.7	48.3 ± 27.4	0.027
Week 6	40.2 ± 29.1	46.7 ± 28.7	0.021
Week 8	39.2 ± 30.1	44.2 ± 29.5	0.09
AUSCAN stiffness index (Scale 0-100)			
Baseline	66.0 ± 22.8	66.6 ± 23.9	0.80
Week 1	48.6 ± 27.2	55.7 ± 27.9	0.003
Week 2	45.8 ± 27.5	53.1 ± 28.2	0.004
Week 4	42.6 ± 30.1	50.4 ± 30.1	0.009
Week 6	40.9 ± 31.1	49.5 ± 30.8	0.005
Week 8	39.4 ± 32.1	45.5 ± 31.4	0.048
AUSCAN physical function index (Scale 0-100)			
Baseline	67.9 ± 18.8	66.7 ± 18.4	0.54
Week 1	51.9 ± 24.5	56.3 ± 24.1	0.009
Week 2	49.2 ± 25.4	52.5 ± 25.4	0.06
Week 4	44.7 ± 28.6	50.8 ± 28.3	0.010
Week 6	42.0 ± 29.3	48.9 ± 28.7	0.005
Week 8	41.4 ± 30.4	47.5 ± 29.3	0.017

P-values at baseline are based on the CMH Chi-squared test of treatment means.
P-values after baseline are based on ANCOVA with main effects of treatment and center and baseline covariate.
Source: VOSG-PE-315 (CTD 5.3.5.1.4) Post-text Table 7.10, Post-text Table 9.2, Post-text Table 9.3, Post-text Table 9.5, Post-text Table 9.9.

TABLE 10.1.2.5 (contd): SHOWING SECONDARY EFFICACY MEASURES IN TARGET HAND (CONTD –STUDY 315

Selected efficacy outcomes (ITT population)	DSG N = 198	Vehicle N = 187	
	n (%)	n (%)	p-value
OARSI response			
Week 1	110 (55.6)	78 (41.7)	0.008
Week 2	117 (59.1)	94 (50.3)	0.06
Week 4	124 (62.6)	94 (50.3)	0.013
Week 6	127 (64.1)	103 (55.1)	0.054
Week 8	130 (65.7)	106 (56.7)	0.06
Pain/rescue response			
Week 1	55 (27.8)	41 (21.9)	0.16
Week 2	65 (32.8)	52 (27.8)	0.27
Week 4	97 (49.0)	59 (31.6)	<0.001
Week 6	92 (46.5)	72 (38.5)	0.11
Week 8	90 (45.5)	73 (39.0)	0.21
Use of rescue medication			
Week 1	133 (67.2)	130 (69.5)	0.66
Week 2	120 (60.6)	117 (62.6)	0.77
Week 3	113 (57.1)	122 (65.2)	0.12
Week 4	117 (59.1)	119 (63.6)	0.39
Week 5	112 (56.6)	113 (60.4)	0.53
Week 6	109 (55.1)	114 (61.0)	0.28
Week 7	109 (55.1)	103 (55.1)	0.98
Week 8	106 (53.5)	106 (56.7)	0.56
Entire study	167 (84.3)	156 (83.4)	0.84

P-values are based on a logistic regression model with main effects of treatment and hand OA category. Main effect of center was dropped to permit convergence.
Source: VOSG-PE-315 (CTD 5.3.5.1.4) Post-text Table 9.7, Post-text Table 9.8 and Post-text Table 9.12.

Overall conclusions from sensitivity analyses:

The analyses incorporating the "same mean" imputation scheme are somewhat supportive of the conclusions from the protocol-specified primary efficacy analyses. As expected, the least squares mean for each endpoint and each treatment group is numerically higher in the primary efficacy analysis; however, the numerical differences between treatment groups favor the Voltaren group.

Statistically significant by-treatment group differences are not achieved in the analyses incorporating the "alternate mean" imputation scheme. However, as highlighted by the sponsor, this analysis is conservative and biased in favor of the vehicle group in this case since more favorable results are imputed for the missing data in the vehicle group than in the Voltaren group. The lack of statistically significant findings in this analysis is not unexpected.

The results of the analyses when imputing according to the BOCF approach yield similar conclusions to the primary efficacy analyses and therefore, provide support for the primary efficacy analyses and minimize the concern regarding the possible impact of the missing data.

In summary, the qualitative conclusions from the missing data sensitivity analyses are largely supportive of the protocol-specified primary efficacy analysis. Several instances with p-values greater than 0.05 do occur but the numerical by-treatment group differences continue to favor Voltaren.

At the request of the division, the sponsor provided cumulative distribution plots (i.e., a continuous responder analyses) for the primary efficacy endpoints for weeks 4 and 6. The descriptive conclusions from these plots are supportive of the efficacy of Voltaren over vehicle for the primary efficacy endpoints and are provided in Figure 4.

OVERALL CONCLUSIONS STUDY VOSG-PE-315

DSG was significantly superior to vehicle in all three co-primary efficacy outcomes (OA pain intensity, total AUSCAN index, and global rating of disease).

At Week 4, differences between the DSG and vehicle groups were statistically significant for OA pain intensity ($p = 0.018$) and total AUSCAN ($p = 0.011$), and borderline significant for global rating of disease activity ($p = 0.06$).

At Week 6, differences between the DSG and vehicle groups were statistically significant for OA pain intensity ($p = 0.023$), total AUSCAN ($p = 0.006$), and global rating of disease activity ($p = 0.023$).

The sensitivity analyses did not modify the above conclusion.

0.1.3 INDIVIDUAL STUDY REPORT- VOSG-PN-304

Title of study: A 12-week, randomized, double-blind, multi-center, vehicle-controlled, parallel group study to assess the efficacy and safety of the DSG 1% for the relief of signs and symptoms in subjects with osteoarthritis of the knee

Primary objective: The primary objective was to have been to compare the efficacy of daily topical applications of diclofenac sodium gel 1% (DSG, 1%) with vehicle when 4 g were applied four times a day for 12 weeks by subjects with mild to moderate knee osteoarthritis (OA). The primary objective was evaluated at Week 12 with regard to reducing pain (Western Ontario McMaster Osteoarthritis Index (WOMAC) pain index), improving functional capacity (WOMAC function index), and improving global disease rating (Visual Analog Scale - VAS).
Secondary Objective: The secondary objective was to have been to evaluate and compare the safety of DSG, 1% with vehicle.

Study Design: This was to have been a 12-week, prospective, randomized, double-blind, multi-center, parallel group study that compared DSG, 1% with vehicle in subjects with OA of the knee.

The study population consisted of male and female ambulatory subjects ≥35 years of age with OA of one or both knees, but with a history of clinically symptomatic OA in one knee only, diagnosed at least 6 months previously and verified by X-ray (Kellgren-Lawrence Grade 1-3). Subjects had OA-related knee pain for 15 days in the month preceding screening and the pain in the target knee had (at least once) required the use of non steroidal anti-inflammatory drugs (NSAIDs) or acetaminophen. After washing out any prior analgesics, subjects had a baseline score of ≥ 50 mm on a 100 mm VAS when rating Pain on Movement (POM) and a baseline WOMAC pain score ≥ 9 (out of 20) immediately prior to randomization. Subjects with a POM score of > 20 mm in the contralateral knee at the baseline visit were excluded.

Safety was to have been assessed by monitoring adverse events (AEs), clinical laboratory evaluations, vital sign measurements, and physical examinations.

Restricted use of rescue medication (acetaminophen) was allowed up to a maximum of 4 g per day.

Inclusion criteria and exclusion criteria:

Eligibility criteria were the same as Study -310

Treatment:
DSG 1% or vehicle 4 gm was to have been applied to the knee four times per day for 12 weeks.

Permitted Concomitant medication:
Generally patients were to have been allowed to continue taking stable (non-analgesic) medications that would not interfere with the metabolism of DSG.

The protocol was amended by the sponsor in October 2004 in order to allow study subjects to continue to receive anxiolytics, stipulating that only stable, low doses which were present at entry and maintained throughout the study were permitted.

The amount of aspirin a subject was to have been allowed to receive (under the specified circumstances) was changed from 160 mg/day to 162 mg/day, in the October 2004 amendment.

Rescue medication:
Restricted use of rescue medication (acetaminophen) was to have been allowed up to a maximum of 4 g per day.

Outcome measures and Endpoints:
Primary efficacy Outcome
There were to have been three primary efficacy outcomes. Statistical significance was required on all 3 measures. The outcomes that were to have been evaluated at the final visit were:
1. WOMAC pain scale at Week 12.
2. WOMAC physical function scale at Week 12.
3. Patient's global disease rating at Week 12.

Comparison between treatment group on each primary outcome was to have been performed with an ANOVA model including main effects of treatment and center, treatment-by-center interaction and including a baseline covariate.

Primary Efficacy Measures:
The primary efficacy measures were:
1) WOMAC pain
2) WOMAC physical function of disease
3) Global Rating of disease

Secondary Efficacy Measures:
The secondary efficacy measures were designed to have been collected at all post-baseline visits, excluding those outcomes assessed at Visit 6 (Week 12) that were to have been designated as primary.
The secondary efficacy endpoints were to have been designated as follows:
1) WOMAC pain score
2) WOMAC stiffness score
3) WOMAC Physical function score
4) Global rating of disease

5) Global rating of benefit
6) Pain on movement (POM), spontaneous pain
7) Global evaluation of treatment (at the final visit)
8) Difference between target knee and contralateral knee on subset of four WOMAC questions
9) Osteoarthritis Research Society International (OARSI) response
10) Use of pain/rescue response
11) Pain on movement [POM] (by diary) – daily and averaged by week
12) Use of rescue medication.

Study visit schedule

TABLE # 10.1.3.1 : SHOWING EVALUATION AND VISIT SCHEDULE FOR STUDY VOSG-PN-304

Procedure	Screening / washout	Random-ization	Treatment phase				Termination
Visit	1	2	3	Phone contact	4	5	6
Week			1	mid break[5]	4	8	12
Day	-7 (-5 to -14)	1	8 (+3)		29 (± 7)	57 (± 7)	85 (± 7)
Written informed consent	X						
Background information	X	X					
Assessment of osteoarthritis	X						
X-ray evaluated/ordered	X	X					
Safety laboratory	X				X[3]		X
Physical examination		X					X
Vital signs		X					X
Urine pregnancy test		X					X
Inclusion/exclusion criteria	X	X					
Randomization		X					
Efficacy assessments[1]	X	X	X		X	X	X
Dispense drug		X[4]	X		X	X	
Collect drug			X		X	X	X
Dispense diary[2]	X	X	X		X	X	
Collect and check diary		X	X		X	X	X
Dispense rescue medication	X	X	X		X	X	
Collect rescue medication		X	X		X	X	X
Concomitant medication	X	X	X		X	X	X
Adverse event reporting		X	X		X	X	X
Global Evaluation of Treatment							X

[1] **Efficacy assessments**: Spontaneous pain, POM, global rating of disease, global rating of benefit (Visit 3-Visit 6 only), and WOMAC at the study site.
[2] **Diary efficacy assessment**: POM assessed daily from Visit 1 through Day 14.
[3] **Laboratory**: only measure liver function tests (LFTs).
[4] **Treatment**: To standardize application, the investigator or designee applied the first dose.
[5] **Telephone contact**: Principal Investigator or designee contacted subject approximately midway between Visits 3-4, 4-5, and 5-6.

Statistical analysis plan and definition of Analyzed Study Populations

The primary efficacy endpoints were the WOMAC pain score and physical function score and the global rating of disease measured at Visit 6 (Week 12) at the study site. The following were secondary efficacy variables at all post-baseline visits, excluding those outcomes assessed at Visit 6 (Week 12) that were designated as primary: WOMAC pain

score, WOMAC stiffness score, WOMAC physical function score, global rating of disease, global rating of benefit, POM, spontaneous pain, global evaluation of treatment (at the final visit), difference between target knee and contra lateral knee on subset of four WOMAC questions, Osteoarthritis Research Society International (OARSI) response, pain / rescue response, POM (diary) – daily and averaged by week, and use of rescue medication.

The efficacy analysis population was initially designated to be ITT. Efficacy analysis of this population did not reveal convincing evidence of efficacy. Post hoc analysis of this population exposed the subjects the subject most likely to respond to topical medication. As a result of this analysis, the protocol of the identical Study -310 was amended prior to unblinding of Study -310, and the modified efficacy subpopulation (MES) was integrated as the analysis population of Study -310.

AMENDMENTS:
Amendment 1 (October 20th 2004)

Amendment 1 changed the study entrance criteria as follows:

- Inclusion criteria were to have been revised to state that:
  - Subjects had to have a clinical diagnosis of OA of the knee per ACR criteria for at least the previous 6 months with symptoms. The stipulation that pain in the target knee for >25 days preceding screening that required the use of NSAID's or acetaminophen was removed.
  - The minimum age for enrollment in the study was changed from 45 years to 35 years.

- Exclusion criteria were revised to exclude:
  - Subjects who had a history of pain in the contralateral knee were further defined to exclude subjects who had a history of pain in the contralateral knee within the last year.
  - "History of rheumatoid arthritis or laboratory values indicative of rheumatoid arthritis with subsequent diagnosis by a physician." Originally stated "History of rheumatoid arthritis or positive laboratory values for Rheumatoid Factor (RF), C-Reactive Protein (CRP) and Sedimentation Rate (BSR) at screening",
  - Subjects who had a history of chronic inflammatory disease or fibromyalgia.
  - Subjects who had received anticoagulants such as warfarin or heparin in the preceding week or antiaggregants other than aspirin (such as Plavix®) in the preceding month.

Amendment 2 (12-Nov-2004)
Amendment 2 changed a study entrance criterion as follows: • Exclusion criterion #2: The criterion that excluded subjects who had a baseline contralateral POM VAS score of more than 10 mm was changed to exclude subjects who had a score of more than 20 mm.
Amendment 3 (03-May-2005)

Amendment 3 (03 May, 2004)
Amendment 3 was generated to accommodate a request by a EU health authority for the addition of a fourth primary outcome. The added primary outcome was already listed among the secondary outcomes. The analysis of the new primary outcome was unchanged.

RESULTS:

Most subjects in each treatment group were female (DSG, 63.3%; vehicle, 61.2%) and Caucasian (DSG, 83.4%; vehicle, 82.7%). The mean ages of the DSG and vehicle groups were 62.2 and 62.8 years, respectively. The treatment groups were well balanced with respect to baseline knee examination parameters.

- DISPOSITION

A summary of subject disposition is provided in Table 10.3.1.

The total number of randomized subjects was 514, with 259 subjects in the DSG group and 255 in the vehicle group. A total of 487 of the 1001 subjects screened for this study were not randomized. The most common reasons screened subjects were not randomized were failure to meet entry criteria.

The proportions of subjects in the DSG and vehicle groups who completed the study were 81.5% and 79.2%, respectively. The proportions of subjects who prematurely discontinued the study in the DSG and vehicle groups were 18.5% (48 subjects) and 20.8% (53 subjects), respectively. The most common reasons for discontinuing the study were subject withdrawal of consent (DSG: 6.6%; vehicle: 7.5%), unsatisfactory therapeutic effect (DSG: 5.0%; vehicle: 5.9%), adverse events (DSG: 4.6%; vehicle: 2.4%), and lost to follow-up (DSG: 2.3%; vehicle: 2.7%). Subjects who discontinued due to AEs are discussed in Section 10.2.2.

The case report forms for the patients who discontinued due to withdrawal of consent were reviewed to ascertain whether these patients actually discontinued due to an adverse event; this review found that two (2) subjects in the vehicle arm withdrew because of lack of therapeutic effect, while 1 subject randomized to DSG withdrew because of an AE (skin rash).

TABLE 10.1.3.2: SHOWING DISPOSITION OF SUBJECTS IN STUDY -304 BY TREATMENT GROUP

	DSG	Vehicle
Total no. of subjects		
Screened	1001	
Randomized	259	255
Completed- n (%)	211 (81.5)	202 (79.2)
Discontinuations – n (%)		
Total	48 (18.5)	53 (20.8)
Adverse events	12 (4.6)	6 (2.4)
Unsatisfactory therapeutic effect	13 (5.0)	15 (5.9)
Protocol deviation	0	4 (1.6)
Subject withdrew consent	17 (6.6)	19 (7.5)
Lost to follow-up	6 (2.3)	7 (2.7)
Administrative problems	0	2 (0.8)

Source: Post-text Table 7.1; Appendix 7, Listing 7.3

- DEMOGRAPHICS

In the all randomized population, the majority of subjects in each treatment group were female (DSG, 63.3%; vehicle, 61.2%) and Caucasian (DSG, 83.4%; vehicle, 82.7%). The mean ages of the DSG and vehicle groups were 62.2 (range: 25 to 82 years) and 62.8 years (range: 36 to 84 years), respectively. The largest proportion of the DSG group was > 60 to 70 years of age (40.2%). The largest proportions of the vehicle group were > 50 to 60 or > 60 to 70 years of age (32.5% each category). No demographic characteristics were statistically significantly different between the treatment groups.

For the MES, demographic characteristics were very similar to those in the all randomized population. The mean body mass index (BMI) was statistically significantly different (CMH chi-squared test of treatment means) between the DSG and vehicle groups (32.0 kg/m2 versus 30.2 kg/m2; p = 0.026). No other demographic characteristics were significantly different between the treatment groups.

TABLE 10.1.3.3: SHOWING A DEMOGRAPHIC SUMMARY BY TREATMENT GROUP AND ALL RANDOMIZED AND MES POPULATION

	All randomized population		MES*	
	DSG N = 259	Vehicle N = 255	DSG N = 156	Vehicle N = 155
Sex – n (%)				
Male	95 (36.7)	99 (38.8)	58 (37.2)	66 (42.6)
Female	164 (63.3)	156 (61.2)	98 (62.8)	89 (57.4)
Race – n (%)				
Caucasian	216 (83.4)	211 (82.7)	128 (82.1)	132 (85.2)
Black	32 (12.4)	27 (10.6)	21 (13.5)	12 (7.7)
Asian	1 (0.4)	3 (1.2)	0 (0.0)	1 (0.6)
Other	10 (3.9)	14 (5.5)	7 (4.5)	10 (6.5)
Age (yr)				
≤40	5 (1.9)	5 (2.0)	4 (2.6)	3 (1.9)
>40-50	24 (9.3)	21 (8.2)	16 (10.3)	15 (9.7)
>50-60	74 (28.6)	83 (32.5)	42 (26.9)	44 (28.4)
>60-70	104 (40.2)	83 (32.5)	66 (42.3)	54 (34.8)
>70-80	51 (19.7)	56 (22.0)	28 (17.9)	35 (22.6)
>80	1 (0.4)	7 (2.7)	0 (0.0)	4 (2.6)
N	259	255	156	155
Mean ± SD	62.2 ± 9.6	62.8 ± 10.0	61.5 ± 9.7	63.2 ± 10.3
Range	25 – 82	36 – 84	25 - 80	36 - 84
Height (cm)				
N	258	254	155	154
Mean ± SD	167.7 ± 10.2	168.4 ± 10.3	168.3 ± 9.9	169.1 ± 10.6
Range	137 – 193	142 – 198	137 - 193	144 - 193
Weight (kg)				
N	257	254	155	154
Mean ± SD	89.4 ± 21.8	89.7 ± 19.5	90.8 ± 23.5	86.7 ± 19.4
Range	49 – 191	55 – 191	49 - 191	55 - 191
BMI (kg/m^2)				
N	257	254	155	154
Mean ± SD	31.7 ± 7.0	31.7 ± 6.4	32.0 ± 7.5	30.2 ± 5.8
Range	19.3 – 62.0	19.6 – 55.4	19.3 – 62.0	20.7 – 55.4

*See Section 9.3.2 for a discussion of efficacy outcomes for the MES
Source: Post-text Table 7.6 and Post-text Table 7.6b; Appendix 7; Listing 7.7

- BASELINE CHARACTERISTICS

The baseline OA assessments of both treatment groups and both the all randomized population and the MES populations were similar at baseline on all measures of OA

visit assessment. The global rating of disease, spontaneous pain, pain on movement of target and contralateral knee, the WOMAC pain , stiffness and physical function index were similar between the two treatment groups.

TABLE 10.1.3.4: SHOWING A SUMMARY OF BASELINE VISIT OA ASSESSMENTS- ALL RANDOMIZED POPULATION AND MODIFIED EFFICACY SUBPOPULATION (MES)

	All randomized subjects		MES*	
	DSG N = 259	Vehicle N = 255	DSG N = 156	Vehicle N = 155
Global rating of disease[1]				
Mean ± SD	63.1 ± 19.4	63.3 ± 18.2	65.0 ± 17.7	64.0 ± 17.2
Range	2 – 100	2 – 100	2 – 100	2 – 100
Spontaneous pain[2]				
Mean ± SD	59.7 ± 20.5	57.8 ± 24.1	61.0 ±19.6	56.8 ± 25.3
Range	2 – 99	1 – 100	3 - 99	1 - 100
Pain on movement[2]				
Target knee				
Mean ± SD	72.7 ± 11.8	71.8 ± 12.5	74.7 ± 12.1	73.5 ± 12.2
Range	50 – 100	20 – 100	51 - 100	50 - 100
Contralateral knee				
Mean ± SD	4.8 ± 5.3	4.6 ± 5.2**	3.7 ± 4.4	3.6 ± 4.3**
Range	0 – 21	0 – 24	0 - 18	0 - 18
WOMAC pain index[3]				
Mean ± SD	11.7 ± 2.1	11.7 ± 2.5	11.9 ± 2.0	11.6 ± 2.4
Range	2 – 19	4 – 20	9 - 19	6 - 19
WOMAC stiffness index[3]				
Mean ± SD	4.81 ± 1.29	4.84 ± 1.47	4.87 ± 1.31	4.72 ± 1.48
Range	0 – 8	0 – 8	0 - 8	0 - 8
WOMAC physical function index[3]				
Mean ± SD	38.6 ± 8.8	38.7 ± 10.4	38.9 ± 8.7	38.1 ± 10.3
Range	13 – 63	9 – 68	16 - 63	9 - 63

*See Section 9.3.2 for a discussion of efficacy outcomes for the MES
**N = 254 in All randomized, N = 154 in MES
[1] 0 = very good, 100 = very poor
[2] 0 = no pain, 100 = unbearable pain
[3] Pain: 0-20; stiffness: 0–8; physical function: 0-68
Source: Post-text Table 7.10 and Post-text Table 7.10a; Appendix 7, Listing 9.1 and Listing 9.2

- EFFICACY
  - Primary

For the ITT population, there was greater improvement in the DSG group as compared with the vehicle group throughout the study for each assessment, but the differences were not statistically significant. The mean WOMAC pain index scores (scale: 0 to 20) at Week 12 in the DSG and vehicle groups were 6.9 and 7.3, respectively. These scores reflect mean decreases from baseline of 4.8 in the DSG group and 4.4 in the vehicle group.

The mean WOMAC physical function scores (scale: 0 to 68) at Week 12 in the DSG and vehicle groups were 24.2 and 25.9, respectively. These scores reflect mean decreases from baseline of 14.4 in the DSG group and 12.8 in the vehicle group.

The mean global ratings of disease VAS scores (scale: 0 = very good to 100 = very poor) at Week 12 in the DSG and vehicle groups were 37.9 and 40.9, respectively. These scores reflect mean decreases (indicating improvement in rating) of 25.1 in the DSG group and 22.4 from baseline in the vehicle group.

TABLE 10.1.3.4: SHOWING PRIMARY EFFICACY OUTCOMES AT WEEK 12 IN THE ITT POPULATION, AND CONTRASTED WITH THE MES POPULATION

	ITT population			MES population		
	DSG N=259	Vehicle N=255	P value	DSG N=156	Vehicle N=155	P value
WOMAC pain score (Scale =0-20)						
Mean	6.91	7.25		6.19	7.03	
Mean change from baseline	4.83	4.41	0.31	5.67	4.58	0.023
WOMAC physical function score (Scale=0-68)						
Mean	24.2	25.9		22.0	25.3	
Mean change from baseline	14.4	12.8	0.17	16.9	12.8	0.011
Global rating of disease (100 mm VAS)						
Mean	37.9	40.9		34.5	41.1	
Mean change from baseline	25.1	22.4	0.23	30.4	23.0	0.013

In the MES population, there was a statistically significant difference between the treatment groups favoring DSG treatment for each primary outcome measure.

- Secondary

Table 10.1.3.5 provides a summary of the Week 1 and Week 12 secondary efficacy

outcomes for the MES. Most of the outcomes assessed in the MES subjects at Week 12 showed a statistically significant difference between treatments favoring DSG. In categories where secondary efficacy assessments did not achieve statistical significance, there was greater improvement in the DSG group as compared with the vehicle group throughout the study for each assessment.

The sponsor found that subjects with a decline of POM score in the target knee between the screening visit and the baseline visit and subjects with a score exceeding 1 on the WOMAC abridged pain index for the contralateral knee at baseline were not responsive to topical DSG, 1% compared to vehicle. All such subjects were excluded and the remaining subjects were defined as the *modified efficacy subpopulation* **(MES)** and were analyzed in a manner identical to the original ITT efficacy population.

VOSG-PN-304 was the first of the controlled Phase 3 studies to be completed by the applicant, and based on results of analysis of the ITT population study, the applicant was able to identify the flaws in the clinical model utilized, and potentially also in VOSG-PN-310, a study of identical design, that remained blinded at the time of completion of Study -304. As a result of this analysis, the protocol of Study-310 was amended.

TABLE 10.1.3.5: SHOWING SECONDARY EFFICACY OUTCOME ANALYSIS – MES SUBJECTS (STUDY -304)

	DSG N = 156	Vehicle N = 155	
WOMAC Functional Disability Index	**Mean ± SD**	**Mean ± SD**	**p-value**
Pain (Scale 0-20)			
Baseline	11.85 ± 2.03	11.61 ± 2.38	
Week 1	7.81 ± 3.62	8.32 ± 3.68	0.08
Stiffness (Scale 0-8)			
Baseline	4.87 ± 1.31	4.72 ± 1.48	
Week 1	3.43 ± 1.64	3.54 ± 1.72	0.30
Week 12	2.79 ± 1.80	3.13 ± 1.92	0.027
Physical function (Scale 0-68)			
Baseline	38.9 ± 8.7	38.1 ± 10.3	
Week 1	27.4 ± 12.7	29.7 ± 12.6	0.042
Pain – time-weighted index	6.66 ± 3.14	7.34 ± 3.48	0.019
Spontaneous pain			
Baseline	61.0 ± 19.6	56.8 ± 25.3	
Week 1	42.2 ± 25.7	41.5 ± 25.8	0.61
Week 12	31.0 ± 26.9	35.4 ± 27.5	0.036
Global ratings			
Disease (Scale 0-100)			
Baseline	65.0 ± 17.7	64.0 ± 17.2	
Week 1	44.7 ± 23.9	46.8 ± 23.8	0.28
Benefit (Scale 0-100)			
Week 1	36.6 ± 24.7	41.3 ± 25.7	0.16
Week 12	27.7 ± 25.1	38.0 ± 29.1	0.001
Global evaluation of treatment (Scale 0-4)	2.40 ± 1.29	1.95 ± 1.38	0.007
Pain on movement (Scale 0-100)			
Assessed at site			
Baseline	74.7 ± 12.1	73.5 ± 12.2	
Week 1	47.0 ± 25.4	50.1 ± 25.0	0.13
Week 12	36.6 ± 27.0	44.0 ± 28.0	0.006
Assessed by subject diary			
Baseline	58.3 ± 17.4	55.6 ± 17.9	
Day 7	43.7 ± 22.7	46.3 ± 23.3	0.10
Day 14	41.8 ± 23.9	42.6 ± 23.9	0.27

P-values are based on ANCOVA with main effects of treatment and center and (except for Global Rating of Benefit) baseline covariate. For global evaluation of treatment, p-value is based on the CMH Chi-squared test of treatment mean ridits, stratified by center; DSG (N = 149), Vehicle (N = 143).
Source: Post-text Table 7.9a, Post-text Table 7,10a, Post-text Table 9.2b, Post-text Table 9.3a, Post-text Table 9.4b, Post-text Table 9.5a, Post-text Table 9.6a, Post-text Table 9.7a, Post-text Table 9.8a and Post-text Table 9.12a; Appendix 7, Listing 9.1, Listing 9.2 and Listing 9.5

SOURCE: VOSG-PN-304- main report, page 57

SAFETY

The proportion of subjects who experienced at least one treatment-emergent AE was 53.7% in the DSG group, compared with 47.1% in the vehicle group. The most common treatment-emergent AE was headache, which occurred in 16.6% and 16.5% of subjects in the DSG and vehicle groups, respectively. Other relatively frequent treatment-emergent AEs that occurred in similar proportions of subjects in each group were arthralgia (DSG: 6.9%; vehicle: 5.9%) and back pain (DSG: 6.9%; vehicle: 7.5%). Nasopharyngitis occurred in more subjects in the DSG group than in the vehicle group, 6.2%

TABLE 10.1.3.5: SHOWING THE NUMBER (%) OF SUBJECTS WITH THE MOST FREQUENT TREATMENT-EMERGENT AE's (> 3%) IN EITHER TREATMENT GROUP (ALL TREATED SUBJECTS)

All treatment-emergent adverse events	DSG	Vehicle
Total treated subjects	N = 259	N = 255
Any adverse event n (%)	139 (53.7)	120 (47.1)
Headache	43 (16.6)	42 (16.5)
Arthralgia	18 (6.9)	15 (5.9)
Back pain	18 (6.9)	19 (7.5)
Nasopharyngitis	16 (6.2)	6 (2.4)
Application site dermatitis	14 (5.4)	0
Pain in extremity	13 (5.0)	4 (1.6)
Upper respiratory tract infection	11 (4.2)	10 (3.9)
Sinusitis	6 (2.3)	11 (4.3)
Influenza	4 (1.5)	8 (3.1)

Source: Post-text Table 10.1; Appendix 7, Listing 10.2

CONCLUSIONS

In the ITT population, subjects who received DSG showed a numerically greater response on efficacy assessments of OA and greater treatment satisfaction as compared with subjects who received vehicle. However, the differences were not statistically significant.

The MES excluded subjects with confounding factors (spontaneous improvement in the target knee or significant pain in the contralateral knee) as is typical in clinical trial 'flare' designs of OA. In that population, all primary efficacy outcomes assessed at Week 12 yielded statistically significant outcomes favoring DSG treatment.

The overall AE profiles of the two treatment groups were similar. The incidence of application site dermatitis in the SOC Skin and subcutaneous tissue disorders was 5.4% in the DSG group and none was observed in the vehicle group. The incidence of gastrointestinal AEs was 3.1% in the DSG group and 3.9% in the vehicle group. Laboratory

findings, including liver function, were unremarkable.

In conclusion, topical treatment with DSG improves pain and functional impairment in patients with knee OA, but avoids the high systemic exposure and the consequent systemic AE profile of oral NSAIDs.

10.1.4. INDIVIDUAL STUDY REPORT - VOSG-PN-314

Title of Study: An 8-week, multi-center, randomized, double-blind, placebo-controlled, parallel group trial of diclofenac sodium gel 1% in patients with primary osteoarthritis of the hand.

Primary objective:
The primary objective of this 8 week study was to have been to compare the efficacy of diclofenac sodium gel, 1% (DSG), applied four times a day versus vehicle in osteoarthritis (OA) of the hand. The results were to have been based on assessment of three efficacy outcomes, each assessed on a 100 mm visual analogue scale (VAS), at Week 4 and Week 6.

Secondary objective:
The secondary objectives of this study were to have been to evaluate:

- Onset of efficacy, by assessment of above efficacy outcomes at the study site visits at Weeks 1 and 2, and by assessment of daily OA pain intensity in the diary over Days 1-14.
- Durability of efficacy, by assessment of above efficacy outcomes at the study site visit at Week 8.
- Effect of DSG on the non-dominant hand by assessment of efficacy outcomes for the non-dominant.

Study Design:
This was to have been an 8-week, prospective, randomized, double-blind, multi-center, placebo-controlled, parallel group study in subjects with OA of the hand.

A total of 398 subjects were randomized to treatment, 202 to DSG and 196 to placebo.

Subjects were to have had an initial 1 week washout of analgesics. Following screening and baseline visits, the subjects visited the study site 5 times for assessments of efficacy, safety, and compliance. Subjects also completed daily diaries throughout the washout and treatment periods in which efficacy and study medication compliance information were to have been recorded.

Safety assessments were to have consisted of monitoring and recording all AEs, serious adverse events (SAE's) with their severity and relationship to study drug and pregnancies, monitoring of hematology and blood chemistry performed at the central laboratory and assessments of vital signs and physical condition.

Use of rescue medication (acetaminophen) was allowed up to a maximum of 4 g per day. All randomized subjects were to have been included in the intent-to-treat (ITT) safety analysis.
The trial population comprised symptomatic subjects aged $\geq$ 40 years with a diagnosis of primary OA in their dominant hand as defined by The American College of Rheumatology

(ACR) criteria.

Inclusion and Exclusion Criteria:
Inclusion and exclusion criteria were the same as for Study VOSG-PE-315.

Treatment:
The DSG or placebo vehicle control was applied four times daily for 8 weeks (2 g to the dominant hand and 2 gm to the non-dominant hand).

Rescue medication:
The use of acetaminophen (up to doses of 4gm per day) as rescue medication was to have been allowed.

Outcome Measures:
The ***primary efficacy outcome measures*** were to have been:

- OA pain intensity in the dominant target hand over the previous 24 hours.
- Total Australian/Canadian Hand Index (AUSCAN) score for the dominant hand
- Global rating of disease activity

The ***secondary efficacy outcome measures*** were to have been measured after 1, 2 and 8 weeks of treatment.
These measures were to have been as follows:

- OA pain intensity in the target hand (in the previous 24 hours)
- Global rating of disease activity
- Total AUSCAN score in the target hand (unweighed sum of the scores on 15 questions)

Other measures that were to have been derived from the above assessment were:

- Time to resolution of pain (OA pain intensity $\leq$ 10 mm in the target hand
- Time to OA pain intensity $\leq$ 20 mm in the target hand
- Osteoarthritis Research Society International response at each visit defined as either of the following:
  - improvement in pain $\geq$ 50% and absolute change $\geq$ 20
    or *at least 2* of the following
  - improvement in pain $\geq$ 20% and absolute change $\geq$ 10
  - improvement in function $\geq$ 20% and absolute change $\geq$ 10
  - improvement in global rating of disease $\geq$ 20% and absolute change $\geq$ 10

Function was to have been measured by the AUSCAN total score standardized to a 0-100 scale. Pain was to have been measured as OA pain intensity in the target hand in the previous 24 hours on the 100 mm VAS

- Pain/ rescue response at each visit defined as not taking any rescue medication in the 3 days prior to the day of the study visit and showing a reduction from baseline

of at least 20 mm on the VAS assessment of pain intensity in the target hand in the previous 24 hours

- Weekly averages of the diary assessment of daily OA pain intensity

Pain intensity outcomes:
OA pain intensity in the target and in the non-dominant hand was to have been assessed at each visit in the CRF and daily in the diary over Days 1 to 14. The treatment groups were to have been compared on the assessments by day over Days 1-14 and averaged by week (week 1: study days 1-7, week 2: study days 8-14).

On consideration of the daily diary washout OA pain intensity values in the target hand over Days -7 to -1 and of the OA pain intensity assessment in the target hand completed at the baseline visit, it was to have been determined that:

1. The baseline for all analyses of the post-baseline OA diary pain intensity values was to have been the mean of the daily diary OA pain intensity values over Days -7 to -1. If the pain VAS was not assessed on any of these days, then the average over all patients with non-missing baseline covariate was to have been used.

2. For all analyses of the post-baseline assessments of OA pain intensity in the CRF two parameters would be used: (a) the assessment of OA pain intensity in the CRF at the baseline visit and (b) the mean of the daily diary OA pain intensity values over Days - 7 to -1. Both were found to have independent predictive capacity that did not disappear when both were in the model.
This consideration applied to both the target and the non-dominant hand.

Time to pain events:
Time to resolution of pain and time to OA pain intensity $\leq$20 mm in the target hand, were to have been both determined based on the VAS assessments from the CRF and from the diary. If for one day there was more than one assessment, the higher VAS value was to have been used. The treatments were to have been compared with the Cox proportional hazards model. A patient whose OA pain intensity in the target hand never achieved the target level of pain (whether 10 mm or 20 mm) contributed a censored observation, where the time to event was to have been taken as the number of days between baseline and the day of the last completed VAS assessment.

Dichotomous outcomes:
All dichotomous outcomes were to have been analyzed with regard to the difference between treatments with the logistic regression model.

Hand examinations:
The difference of treatment groups regarding the affected joints of the target and non-dominant hand was to have been analyzed with the CMH test for categorical data. No stratification for center or hand OA category was to have been made.

Primary Efficacy Endpoints:

- OA pain intensity
- Total AUSCAN Score
- Global rating of disease

Secondary Efficacy Measures:

- OA pain intensity @ Week 1, 2 and 8
- Total AUSCAN Score
- Global rating of disease @ Week 1,2 and 8
- AUSCAN sub-indices
- OA pain intensity (non-dominant hand)
- AUSCAN sub-indices (non-dominant hand)
- Global rating of benefit
- FIHOA (Weeks 2, 4 and 6)
- Daily OA pain intensity (target diary assessment)
- Daily OA pain intensity (non-dominant hand)
- Time to resolution of pain (target hand)
- Time to pain improvement (target hand)

Safety:

Safety assessments were to have consisted of monitoring and recording of all AE's, SAE's and their severity, monitoring of hematology and blood chemistry, and assessments of vital signs and physical condition.

Study Visit Schedule:

TABLE 10.1.4.1:SHOWING EVALUATION AND VISIT SCHEDULE FOR STUDY VOSG-PE-315

Examination	Phase	Screening Washout Period	Randomization / Baseline	Treatment Period				Final Visit
	Weeks	-1	0	1	2	4	6	8
	Days	-7 to -5	1	8	15	29	43	57
	Visit	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Written informed consent		X						
Background /Medical History		X						
Examination of the hands (X-ray if not already available)		X						
Prior/Concomitant medications and/or significant non-drug therapies		X	X	X	X	X	X	X
Hematology and blood chemistry		X						X
Urine pregnancy test			X					X
Inclusion/exclusion criteria		X	X					
Physical exam and vital signs			X					X
Diagram of affected joints (clinical assessment)		X	X		X	X	X	X
OA pain on VAS [1], global rating of disease activity, and AUSCAN [1, 2]		X	X	X	X	X	X	X
FIHOA			X		X	X	X	
Global rating of benefit				X	X	X	X	X
Global rating of efficacy								X
Dispense study medication			X [3]		X	X	X	
Dispense rescue medication		X	X	X	X	X	X	
Dispense diary		X	X		X	X	X	
Review and collection of diary			X	X [4]	X	X	X	X
Accountability of study medications					X	X	X	X
Check consumption of rescue medication			X	X	X	X	X	X
AEs			X	X	X	X	X	X

[1] Separate assessments for the right hand and for the left hand.

[2] For the non-dominant hand, only pain and stiffness subscales were assessed.

[3] Investigator instructed subject in proper application of the study medication and supervised application.

[4] Only a review of the diary.

Statistical Analysis Plan and Definition of Analyzed Study Populations:

Handling of treatment failures

A treatment failure was designated as such:

If there was a series of 4 or more consecutive days, (starting after Day 7), in which a patient took either:

(a) At least 2 grams acetaminophen (paracetamol),

or

(b) At least half the maximum daily over the counter (OTC) dose of a NSAID

or

(c) One or more single prescription strength doses of a nonselective or COX-2 selective NSAID, specifically to treat hand OA pain.

No imputation of missing diary data was used for this purpose except for missing doses of rescue medication.

Sensitivity analyses:
Sensitivity analyses were to have been conducted in which imputation was done only for the visit immediately following the designation as *treatment failure*. For the sensitivity analysis of daily assessment of pain, only assessments within the period for which the definition of treatment failure was satisfied were replaced by imputation using LOCF.

Non-missing assessments after this period were not replaced.

An additional sensitivity analysis was conducted with a definition of treatment failure requiring:
(a) 4 grams paracetamol (acetaminophen)
or
(b) The full daily dose of an OTC NSAID
or
(c) A single dose of NSAIDs as above, daily for 4 consecutive days to treat the pain of hand OA.

Sensitivity analyses were to have been conducted on the primary outcomes in the final study model to assess the impact of a variety of issues.

These included the following:

1) The impact of imputing by LOCF for early termination
   a) At each Visit X, missing values in the diclofenac group due to early termination were imputed (replaced) by the mean of all non-missing values in the diclofenac group, and correspondingly for vehicle.
   b) At each Visit X, missing values in the diclofenac group due to early termination were imputed by the mean of all non-missing values in the vehicle group, and vice versa.

2) The impact of treatment failures
   a) Their post-failure efficacy data were imputed only at the immediately following visit rather than at all subsequent visits.
   b) The definition of treatment failure was changed requiring (a) 4 grams paracetamol (acetaminophen) or (b) the full daily dose of an OTC NSAID, or (c) a single dose of NSAIDs, daily for 4 consecutive days to treat the pain of hand OA.

3) The impact of patients who did not stay in the study long enough to supply efficacy assessments at the Week 1 visit – the analyses were rerun with these patients excluded.

AMENDMENTS:

The study protocol (dated 22 December 2004) was amended twice. Both amendments were implemented prior to any site starting recruitment.

Amendment 1 (11 January 2005)

The following amendments were made to clarify the Inclusion/Exclusion criteria and the identity of the laboratory parameters to the investigators at the different study locations.

• Text was added to Exclusion criterion #1 Secondary post-traumatic OA, history and/or evidence of any other rheumatic disease involving the potential target hand or the arm: algodystrophy, septic arthritis, inflammatory joint disease (e.g. psoriatic arthritis), rapidly destructive osteoarthropathy, chondrocalcinosis, gout, recurrent episodes of pseudogout, Paget's disease of bone, articular fracture, ochronosis, acromegaly, hemochromatosis, primary osteochondromatosis, heritable disorders (e.g. hypermobility), collagen gene mutations, carpal tunnel syndrome, Dupuytren's disease and neurological diseases of the hand or arm.

• Additional text was provided to clarify forbidden concomitant therapies: "Chondroitin sulfate, glucosamine sulfate, avocado or soybean unsaponifiables.

• The text concerning the laboratory evaluations was modified to include the erythrocyte sedimentation rate and C-reactive protein and to clarify the terminology concerned with existing parameters.

Amendment 2 (23 March 2005)
It was decided that randomization would not be stratified to balance treatment allocation by presence or absence of pain in the CMC-1 joint. Rather, randomization was conducted at the site by assigning the lowest available randomization number at the site to the subject (the interactive voice response system was not to be used). The rationale for these changes was to minimize the possibility of logistical problems in the randomization process and of issues arising in the interpretation of the final study results.

The protocol was to have been amended such that medical conditions originating (or increasing in severity) between the time that the informed consent is signed and start of study treatment were to have been captured in the medical history form, rather than being recorded on the AE form and these events were not to be summarized as part of the statistical analysis of safety. This change was made since there is no requirement for these occurrences to be handled as AEs. Events occurring during the run-in period were not considered to be relevant to the safety of DSG.

The visit schedule in the protocol was modified to reflect the absence of a FIHOA index assessment at the final visit. It was also specified that subjects mark their response on the OA pain intensity and global rating of disease activity VAS using an " X " rather than a vertical line (|). In addition, the requirement for investigators to weigh returned tubes of study drug was deleted.

RESULTS:

- DISPOSITION

A total of 45 of the 443 subjects screened for this study were not randomized. The most common reasons for screened subjects not being randomized were:

- pain (on a 100 mm VAS) in the target hand during previous 24 hours to the baseline visit being < 40 mm (7 subjects)
- withdrawal of consent (7 subjects)
- pain not usually greater in the dominant hand (5 subjects)
- clinically significant laboratory abnormalities (5 subjects)
- pain (on a 100 mm VAS) over 24 hours prior to the baseline visit < 20 mm lower (on a 100 mm VAS) in the non-dominant hand than in the target hand (4 subjects).

The proportions of randomized subjects in the DSG and vehicle groups who completed the study were 89.6% and 94.9%, respectively.

TABLE 10.1.4.2: SHOWING THE SUBJECT DISPOSITION FOR EACH TREATMENT GROUP

	DSG	Vehicle
Total no. of subjects		
Screened	443	
Randomized	202	196
Completed – n (%)	181 (89.6)	186 (94.9)
Discontinuations – n (%)		
Total	21 (10.4)	10 (5.1)
AEs	9 (4.5)	5 (2.6)
Unsatisfactory therapeutic effect	3 (1.5)	4 (2.0)
Protocol deviation	1 (0.5)	0
Subject withdrew consent	4 (2.0)	1 (0.5)
Administrative problems	4 (2.0)	0

Note: percent basis is the total number of subjects randomized.
Source: Post-text Table 7.1; Appendix 7, Listing 7.3

The case report forms for the patients who discontinued due to withdrawal of consent were reviewed to ascertain whether these patients actually discontinued due to an

adverse event or due to lack of efficacy. The review found that one subject in the DSG arm discontinued due to an AE ("inconvenient sensation" in the hand) while another subject in the vehicle arm discontinued due to lack of efficacy.

PROTOCOL DEVIATIONS:

The proportion of subjects who had at least one protocol violation was 35.6% (72 subjects) in the DSG group, compared with 42.3% (83 subjects) in the vehicle group. The most common protocol violations in each treatment group related to subjects being randomized out of sequence (54 subjects overall), failure to discontinue rescue medication 36 hours before study visits (23 subjects at Visit 2, 16 subjects at Visit 5, and 17 subjects at Visit 6), pain in the non-dominant hand being < 20 mm lower than the target hand (33 subjects), overall compliance with study drug being < 75% (21 subjects), and a ≤ 15 mm increase in OA pain following washout of NSAID (18 subjects).

Subject 222-1451 (vehicle group) is listed among the major protocol violators. A discussion with the investigator revealed that a correction on the date of Visit 2 had not been entered into the database before the study was unblinded. It was decided not to correct the database. Thus this subject was in fact not a major protocol violator (due to the wash out period then being long enough).

- DEMOGRAPHICS

The mean age in the DSG and vehicle groups was 64.0 and 63.5 years, respectively. Most subjects in each treatment group were female (DSG, 81.7%; vehicle, 85.2%) and Caucasian (DSG, 98.5%; vehicle, 97.4%).
The majority of subjects in the DSG and vehicle groups were 50 to 70 years of age (66.3% and 65.8%, respectively). No statistically significant differences were detected between treatment groups in any of the demographic characteristics.

TABLE 10.1.4.3: SHOWING THE DEMOGRAPHIC SUMMARY BY TREATMENT GROUP—Study 314

	DSG (N = 202)	Vehicle (N = 196)
Sex – n (%)		
Male	37 (18.3)	29 (14.8)
Female	165 (81.7)	167 (85.2)
Race – n (%)		
Caucasian	199 (98.5)	191 (97.4)
Black	1 (0.5)	3 (1.5)
Asian	2 (1.0)	2 (1.0)
Age (yr) – n (%)		
≤ 40	1 (0.5)	1 (0.5)
> 40-50	17 (8.4)	17 (8.7)
> 50-60	61 (30.2)	59 (30.1)
> 60-70	73 (36.1)	70 (35.7)
> 70-80	39 (19.3)	39 (19.9)
> 80	11 (5.4)	10 (5.1)
Mean ± SD	64.0 ± 9.7	63.5 ± 9.8
Range	40 to 88	40 to 85
Height (cm)		
Mean ± SD	164.3 ± 9.1	164.3 ± 7.6
Range	142 to 193	142 to 187
Weight (kg)		
Mean ± SD	72.6 ± 12.9	71.4 ± 13.5
Range	45 to 111	31 to 120
BMI (kg/m^2)		
Mean ± SD	26.9 ± 4.3	26.4 ± 4.5
Range	17.4 to 37.9	14.3 to 41.9

Source: Post-text table 7.6; Appendix 7; Listing 7.7

o BASELINE CHARACTERISTICS

The treatment groups were well balanced with respect to baseline hand examination parameters. In the DSG and vehicle groups, the baseline mean values of the various assessments in the target hand were consistent with moderate pain and disability, e.g., mean OA pain values were 68.6 in both treatment groups while mean baseline total AUSCAN scores were between 61 and 63. The baseline mean values of the various assessments in the non-dominant hand were consistent with mild pain, e.g., mean OA pain values were between 30 and 35 in both treatment groups.

TABLE 10.1.4.4: SHOWING A SUMMARY OF THE BASELINE HAND OA ASSESSMENTS –STUDY -314

	DSG (N = 202)	Vehicle (N = 196)
Global rating of disease (100 mm VAS)[a]		
Mean ± SD	56.3 ± 17.2	56.9 ± 17.6
Range	5 to 95	18 to 98
Target (dominant) hand		
OA pain intensity (100 mm VAS)[b]		
Mean ± SD	68.6 ± 14.5	68.6 ± 15.1
Range	26 to 98	4 to 99
Total AUSCAN index[c]		
Mean ± SD	62.5 ± 16.3	61.2 ± 18.1
Range	12 to 97	10 to 94
AUSCAN pain index[c]		
Mean ± SD	61.5 ± 16.9	60.3 ± 18.4
Range	15 to 98	10 to 97
AUSCAN stiffness index[c]		
Mean ± SD	54.8 ± 25.3	53.1 ± 25.7
Range	1 to 98	0 to 100
AUSCAN physical function index[c]		
Mean ± SD	64.0 ± 17.2	62.6 ± 19.0
Range	10 to 97	9 to 95
Contralateral (non-dominant) hand		
OA pain intensity[b]		
Mean ± SD	29.6 ± 18.6	29.4 ± 18.4
Range	0 to 82	0 to 85
AUSCAN pain index[c]		
Mean ± SD	34.4 ± 19.6	30.6 ± 19.3
Range	0 to 96	0 to 86
AUSCAN stiffness index[c]		
Mean ± SD	32.4 ± 24.1	28.7 ± 22.8
Range	0 to 97	0 to 94
FIHOA index[d]		
Mean ± SD	12.9 ± 4.1	13.3 ± 4.3
Range	3 to 24	2 to 24

[a] 100 mm visual analogue scale: 0 = very good, 100 = very poor.
[b] 100 mm visual analogue scale: 0 = no pain, 100 = unbearable pain.
[c] Average over multiple questions: 0 = no pain/stiffness/difficulty, 100 = extreme pain/stiffness/difficulty.
[d] FIHOA index: 0-30.
Source: Post-text table 7.10; Appendix 7, Listing 9.1, Listing 9.4 and Listing 9.6

- EFFICACY

- PRIMARY

The following table (10.1.4.4) presents results for the primary efficacy outcomes at Weeks 4 and 6. All primary outcomes favored the DSG group vs. the vehicle group. However, the differences were not statistically significant at the Week 4 and Week 6 time points, as were defined by protocol. Statistical superiority of efficacy end points were reached early at Week 2, and not at 4-6 weeks as was defined by protocol.

TABLE 10.1.4.4: SHOWING THE PRIMARY EFFICACY OUTCOMES AT WEEKS 4 and 6 – STUDY -314 (ITT POPULATION)

		Mean Decrease from baseline (100 mm VAS)			
Outcome variable	**Week**	**DSG (N = 202)**	**Vehicle (N = 196)**	**Difference V-D [d]**	**p-value**
OA pain intensity [a]	4	46.4	49.4	2.0	0.33
	6	41.9	46.2	3.4	0.14
Total AUSCAN [b]	4	46.2	48.4	2.6	0.16
	6	43.0	46.2	3.6	0.09
Global rating of disease [c]	4	42.7	43.3	0.3	0.14
	6	38.8	42.1	3.0	0.16

[a] 0 = no pain, 100 = unbearable pain
[b] Calculated as the unweighted mean over all 15 questions (0 = no pain / stiffness / difficulty, 100 = extreme pain / stiffness / difficulty)
[c] 0 = very good, 100 = very poor
[d] Least squares mean difference between vehicle and DSG
Change from baseline was calculated as baseline minus post-baseline score. Analysis was analysis of covariance (ANCOVA) with main effects of treatment, center, hand OA category and baseline covariate.

SECONDARY

All secondary efficacy outcomes favored treatment with DSG, although the differences between active and study drug were not statistically significant. All secondary efficacy outcomes favored treatment with DSG. Statistically significant differences ($p < 0.05$) between treatment groups were seen at Week 1 and/or Week 2 for OA pain intensity, total AUSCAN index score (as well as sub-indices of pain and physical function), daily OA pain intensity, and the proportion of OARSI responders.

- SAFETY

The safety profiles of the two treatment groups were comparable. The proportion of subjects who experienced at least one treatment-emergent AE was 29.7% in the DSG group compared with 29.1% in the vehicle group. The most common treatment-emergent AE was headache, which occurred
The incidence of treatment-emergent AEs was highest in the nervous system disorders body system: 7.9% in the DSG group compared with 11.7% in the vehicle group. The incidence of treatment-emergent AEs was second highest in the musculoskeletal and connective tissue disorders body system: 7.9% in the DSG group compared with 6.6% in the vehicle group. Infections and infestations were also frequent with incidence rates of 5.4% in the DSG group compared with 8.7% in the vehicle group. General disorders and administration site conditions were reported for 5.9% in the DSG group compared with 3.1% in the vehicle group. The overall incidence of gastrointestinal AEs was comparable in both treatment groups (DSG: 4%, vehicle: 5.1%). No hepatobiliary AEs were reported.

TABLE 10.1.4.5: SHOWING: NUMBER (%) OF SUBJECTS WITH MOST FREQUENT TREATMENT-EMERGENT AE's (> 2 % IN EITHER TREATMENT GROUP(ALL TREATED SUBJECTS)

All treatment-emergent adverse events	DSG (N = 202)	Vehicle (N = 196)
Any adverse event - n (%)	60 (29.7)	57 (29.1)
Headache	14 (6.9)	19 (9.7)
Nasopharyngitis	6 (3.0)	10 (5.1)
Back pain	5 (2.5)	5 (2.6)
Arthralgia	5 (2.5)	3 (1.5)
Toothache	1 (0.5)	6 (3.1)
Application site dermatitis	5 (2.5)	0
Source:[Post-text Table 10.1];[Appendix 7] [Listing 10.2]		

No subject died during the study. A total of 5 subjects, 2 in the DSG group and 3 in the vehicle group, experienced a total of 5 SAEs. In the DSG group, the SAEs were sepsis caused by Escherichia coli (severe) and depression. In the vehicle group, the SAEs were coronary artery stent insertion, wrist fracture and facial palsy. None of the SAEs were suspected to be related to the study drug. A total of 14 subjects, 9 (4.5%) in the DSG group and 5 (2.6%) in the vehicle group experienced AEs that led to discontinuation of the study drug. The only AE that led to discontinuation in more than 1% of subjects in either group was application site dermatitis (1.5% in the DSG group, none in the vehicle group). Five subjects (2.5%) in the DSG group discontinued due to application site conditions (including dermatitis, irritation, pruritus, and an unspecified application site reaction) compared to one subject (0.5%) in the vehicle group (application site burning). Gastrointestinal disorders leading to study

discontinuation were absent in the DSG group (none) and infrequent in the vehicle group (0.5%).

Laboratory, physical examination, and vital signs findings were unremarkable. Liver function test elevations were infrequent, with similar incidence in both treatment groups; ALT or AST elevations $\geq$ 3 x ULN at the latest post-baseline assessment were observed in less than 1% of subjects in either treatment group.

OVERALL CONCLUSION- Study -314 (hand OA):
Assessments of efficacy favored treatment with DSG vs. the vehicle group. However, differences vs. vehicle were not statistically significant in the primary endpoints at Week 4 and Week 6, and did not meet the primary efficacy endpoints defined in the protocol.

Statistically significant effects were found in the primary efficacy variables at Week 1 (total AUSCAN score) and Week 2 (OA pain intensity and total AUSCAN score). Statistically significant efficacy results were only obtained at the earlier timepoints.

Study VOSG-PE-314 produced a marginal efficacy result as compared with Study -315. The reasons for this are unclear, and may be related to:
1) Baseline characteristics of the enrolled subjects were different, with those subjects in Study -314 having a higher rate of sclerosis in X-ray, and approximately one-half of the NSAID or acetaminophen use prior to the study.

2) This was the only Phase 3 study that was performed outside of the United States (in France and Germany).
3) The different translations of the questionnaires may have influenced the efficacy outcomes.

10.1.5 INDIVIDUAL STUDY REPORT VOSG-PN-309

TITLE OF STUDY: An uncontrolled long-term safety trial of DSG 1% in patients with OA of the knee

Primary Objective:
The primary objective of the study was to have been to determine the long-term safety of 1% DSG, when 4 g per knee was to have been applied to one or both knees four times a day for up to 12 months as measured by rates of clinical adverse events (AE's) and monitoring of laboratory values.

Secondary Objective:
The secondary objective was to have been to evaluate the continuing effectiveness of 1% DSG with the use of the Western Ontario McMaster Osteoarthritis (WOMAC) Index.

Study Design:
The study was to have been designed as a multi-center, open-label, long-term safety study that included subjects who had completed either the VOSG-PN-304 or the VOSG-PN-310 double-blind study by 31-Mar-2005 and naïve subjects (who had not participated in the double-blind studies) with osteoarthritis (OA) of the knee.

There were to have been eight (8) scheduled visits for subjects from the double-blind trials; a baseline visit (this was also the final visit of the double-blind efficacy/safety trial), monthly visits for 6 months, and a visit at Month 9.
Naïve subjects had 2 additional visits. One was a preliminary screening visit to conduct laboratory tests. Results of these tests were then reviewed at the baseline visit prior to enrollment into the study. The other additional visit occurred at Month 12. At each visit the investigator or designee reviewed the diary to determine treatment compliance and rescue medication usage and the subject reported AE's and concomitant medication usage.

Subjects were to have had blood drawn for laboratory tests at Months 1 and 3 and then every 3 months and completed the WOMAC questionnaire every 3 months.

Duration: Subjects were to have applied 4 g of DSG per knee to one or both knees four times a day for up to 12 months. Subjects continuing from the double-blind studies were scheduled to be treated for 9 months and naïve subjects were scheduled to treat for 12 months. If there was a clinical diagnosis of OA of the contralateral knee, the subject was dispensed additional medication and was instructed to apply an additional 4 g of DSG 4 times a day to the contralateral knee.

Sample size: The study was to continue until at least 450 subjects had been exposed to DSG for 3 months, 225 subjects for 6 months and 75 subjects for 12 months (including prior exposure to DSG among those continuing from the double-blind trials).
Inclusion and exclusion criteria: Subjects were eligible to continue from VOSG-PN-304 or VOSG-PN-310 if they completed the double-blind studies without major protocol

violations. Naïve subjects comprised of male and female subjects, ≥ 35 years of age with OA of the knee that had been diagnosed at least 6 months previously and verified by X-ray (Kellgren-Lawrence Grade I-III) and who had OA-related knee pain for 15 days of the month preceding screening.

Treatment:
At Visit 1, all subjects were dispensed open label active DSG. Subjects from the double-blind studies retained their number from the double blind trial. Naïve subjects received a unique number.

Permitted Concomitant Medication:
Generally, patients were to have been allowed to continue taking concomitant non-analgesic medications that were on a stable dose, and would not be expected to compromise the safe use of topical diclofenac. NSAIDs or any other analgesics except the study medication and rescue medication provided were not permitted from the time that the Informed Consent Form was signed until study completion or early termination. Stable low doses of aspirin up to 162 mg/d were allowed.

Rescue Medication:
Acetaminophen tablets 500 mg were to be supplied as rescue medication. Individual doses of 1 to 2 tablets could be taken, up to a maximum of 8 tablets per day (4 g per day).

Diaries were dispensed at Visit 1 to all subjects. The number of tablets taken, the time, and the reason were recorded by the subject on the diary. Subjects were instructed to take only the rescue medication provided for any other aches they might experience during the trial, such as headache.

Outcome Measures:
Efficacy was to have been a secondary outcome in this study. WOMAC 3.1 Likert Scale scores (pain, stiffness, physical function) were assessed every three months.

Primary Endpoint:
The primary objective of the study was to have been long term safety. All AE's were to have been monitored and recorded, including their severity and relationship to study drug. All pregnancies were to have been recorded. Hematology, blood chemistry, and urine assessments were obtained at screening (naïve subjects only) or baseline (continuing subjects only) and at the conclusion of the trial. LFT's were monitored regularly throughout the trial. Vital signs and physical examinations were done at baseline and at the conclusion of the trial.

Secondary (Efficacy) Endpoint:
Summary statistics were to be given for the scores at each visit and for the change from baseline. No statistical testing was to be performed for the efficacy parameters.

The subject were to have assessed pain/stiffness/physical function for the target knee joint

using the WOMAC categorical scale at the initial visit and at Month 3, 6, and 9 (and 12, for naïve subjects only) study visits. The study nurse was to have assisted the subjects in case of questions and checked the completion of the questionnaire before the subject left the study site.

All safety analyses were to have been done in the '*All treated subjects*' population comprised all subjects who received at least one dose of DSG in studies VOSG-PN-309, VOSG-PN-304, or VOSG-PN-310.

Subjects excluded from the analysis population were to have included the following:
(i) All subjects who those who received placebo in VOSG-PN-304 or VOSG-PN-310 and did not continue into VOSG-PN-309
(ii) All treated subjects population who had previously received placebo, had continued into VOSG-PN-309, but attended only the baseline visit, and who had no follow-up information in VOSG-PN-309.

Two subset populations were defined for supportive analyses of safety:
(i) All subjects treated for at least 6 months – this included all treated subjects except those who applied no study medication on > 18 days over Months 1-6 (any 18 days, consecutive or not).
(ii) All subjects treated for 12 months – this included all treated subjects except those who applied no study medication on > 36 days over Months 1-12 (any 36 days, consecutive or not).

Study Visit Schedule:

TABLE 10.1.5.1 SHOWING EVALUATION AND VISIT SCHEDULE

Examination	Screening (naïve subjects only)	Entry to trial	Treatment Day (± 14 days)							End of study or early termination for continuing subjects only	End of study or early termination for naïve subjects only
Month		-	1	2	3	4	5	6	9[N]	9[C]	12[N]
Day	-7 (0 to -14)	1	30	60	90	120	150	180	270	270 (± 14 days)	365 (± 14 days)
Visit number	0	1[5]	2[5]	3[5]	4[5]	5[5]	6[5]	7[5]	8[5]	8[C]	9
Informed consent for open label	X	X[C]									
Collect Medical history	X										
Safety laboratory	X[3]	X[C3]	X[1]		X[1]			X[1]	X[1]	X	X
X-ray	X										
Physical examination		X								X	X
Vital signs		X								X	X
Urine pregnancy		X								X	X
Entrance criteria	X[4]	X[4]									
Dispensed medication[2] and diary	X[6]	X	X	X	X	X	X	X	X		
Collected/ checked medication and diary		X[7]	X	X	X	X	X	X	X	X	X
WOMAC		X			X			X	X	X	X
AEs		X	X	X	X	X	X	X	X	X	X
Concomitant medication		X	X	X	X	X	X	X	X	X	X

[C]Continuing subjects only
[N]Naïve subjects only
[1]Liver function tests (LFTs) only
[2]Medication included DSG and rescue medication

Statistical Analysis Plan and Definitions of Analyzed Study Populations:

Safety and Efficacy evaluation:
Safety was to have been assessed mainly as adverse events (AE's) and laboratory values.

Safety was assessed mainly as the frequency of treatment-emergent AEs and as the number of laboratory values at post-baseline visits that fell outside of pre-determined ranges. Statistical testing was not performed for any safety parameter.

HANDLING OF MISSING DATA:
If a single question in the WOMAC pain index was not answered, it was imputed as the median of the answers on the other 4 questions (rounded down if necessary). The same procedure was applied if at most three questions in the physical function index were not answered. Otherwise, the values of the WOMAC pain, stiffness and physical function indices were imputed using the procedures of

Missing WOMAC subscores at the screening visit were not imputed.

Missing WOMAC subscores at the baseline visit were *imputed as the mean values of the respective assessments over all patients with non-missing values* at the baseline visit.

If a *post-baseline visit or several consecutive visits were skipped* by a patient, each WOMAC outcome was imputed by *averaging the outcome values of the latest preceding (including baseline) and the earliest following non-missed visit*, rounded down to the nearest integer.

If there was no Visit X and there were two visits in the window for Visit X+1, then the earlier of these two visits was averaged with Visit X-1 to impute Visit X regardless of which of the two visits was used as Visit X+1.

Imputation for *treatment failure* or *treatment confounding* was not to have been done in this study; this included data collected during the double-blind phase in patients who were classified as treatment failures or treatment confounders during the double-blind phase).

Missing safety data was not replaced.

AMENDMENTS:
Amendment 1 (29-Sep-2004)
The Amendment 1 version of the protocol was the first submitted for IRB approval and should be considered the "original" protocol.

Amendment 2 (09-Dec-2004)
Changes implemented under Amendment 2 were to have included the following:
- The study duration of 12 months for subjects continuing from the double-blind

studies was to have been reduced to 9 months.
- The study inclusion minimum age of 45 years was to have been reduced to 35 years for consistency with the double-blind studies.
- The inclusion criterion regarding clinical diagnosis of OA of the knee was revised to specify "with onset of symptoms" at least 6 months previously, as in the double-blind studies.
- Protocol inconsistencies regarding the pregnancy testing schedule and issuance of rescue medication to the subjects were corrected.
- The AE section of the protocol was modified for clarity.
- Editorial revisions were made to support the items noted above.

Amendment 3 (01-Mar-2005)
Stopping enrollment of continuing and naïve subjects effective 31-Mar-2005 was planned, as it was determined that the study enrollment goals - at least 450 subjects exposed to DSG for 3 months, 225 subjects for 6 months, and 75 subjects for 12 months - would be met by that date.

PROTOCOL DEVIATIONS:
A summary of protocol violations is provided in [Table 10.1.5.1].
The proportion of subjects treated in VOSG-PN-309 who had at least one protocol violation was 18.0% (104 subjects); similar proportions were observed among subjects with one or both knees affected by OA.
The most common violation in the population of subjects treated in VOSG-PN-309 was taking a NSAID after enrollment (10.7%) and having a major protocol violation in one of the preceding double-blind studies (4.7%). Other individual protocol violations were noted for $\leq$ 2% of the population.

Uses of NSAIDs or aspirin after enrollment in all patients treated in VOSG-PN-309 were summarized in the results. In the total population, 13.5% of subjects used these disallowed medications at any time during the study. Use of disallowed medications was more frequent in subjects with two symptomatic knees (16.7%) than in subjects with one symptomatic knee (11.4%).

Study -309 is an open label, roll over, extension study, designed to look primarily at long term safety. The concomitant use of NSAIDs and blood thinning agents could only better define the safety profile of the topical product, since it is not unreasonable to believe that patients in the real life, non-clinical trial environment, could conceivably be taking these concomitant medications along with the topical NSAIDs, thereby mimicking the real life use of this topical product. The concomitant use of oral NSAIDs in this trial speaks more to the safety profile of the product.

TABLE 10.1.5.1: SHOWING A SUMMARY OF THE PROTOCOL VIOLATIONS IN ALL PATIENTS TREATED IN VOSG-PN-309

	Total	1-Knee	2-Knees
Total Treated Patients	578	350	228
Total With a Protocol Violation - N (%)	104(18.0)	66(18.9)	38(16.7)
Entry Violations			
Subtotal	35(6.1)	29(8.3)	6(2.6)
Baseline X-ray showed at least one exclus. criterion	5(0.9)	1(0.3)	4(1.8)
Concom. disease: GOUT	2(0.3)	1(0.3)	1(0.4)
Concom. disease: MICROSCOPIC COLITIS	1(0.2)	0	1(0.4)
Had a major protocol violation in the preceding core study	27(4.7)	27(7.7)	0
Rollover patient - did not complete the core double-blind trial	1(0.2)	1(0.3)	0
Con Med Violations			
Subtotal	72(12.5)	39(11.1)	33(14.5)
BLOOD THINNING AGENT taken after enrollment	10(1.7)	6(1.7)	4(1.8)
NSAID taken after enrollment	62(10.7)	33(9.4)	29(12.7)

Note: A patient can have multiple protocol violations.
All percents are relative to total treated patients.

RESULTS

DISPOSITION:

A summary of subject disposition is provided in Table 10.1.5.2.
A total of 583 subjects were enrolled in VOSG-PN-309 by 112 study sites in the United States. The study population comprised 169 subjects who continued from VOSG-PN-304, 122 subjects who continued from VOSG-PN-310, and 292 naïve (not continuing from the previous studies) subjects. All of the continuing subjects enrolled in the current study to treat one knee. Among the naïve subjects, 64 enrolled to treat one knee and 228 enrolled to treat both knees. Of the 396 naïve subjects screened for this study, 104 were not enrolled. The most common reasons screened subjects were not enrolled included not having the qualifying knee X-ray showing OA with Kellgren-Lawrence grade 1-3, having rheumatoid arthritis, and withdrawing consent.

About half of all subjects completed their pre-specified treatment duration, almost 60% of one-knee subjects and almost 40% of two-knee subjects. The most common reasons for prematurely discontinuing the study were withdrawal of consent in roughly 20% of subjects, AEs (roughly 10% of one-knee subjects and roughly 20%

of two-knee subjects, and unsatisfactory therapeutic effect in 7.0% of one-knee subjects and 14.0% of two-knee subjects.

TABLE 10.1.5.2: SHOWING DISPOSITION OF SUBJECTS IN STUDY -309

	Total n (%)	One Knee	Two Knees
TOTAL # SUBJECTS IN STUDY			
Enrolled	583	355 (18.0)	228
Treated	587 (99.1)	350 (98.6)	228 (100)
Completed	294 (50.4)	206 (58.0)	88 (38.6)
ORIGIN OF SUBJECTS			
De novo	292 (49.6)	64 (18.0)	228 (100)
Continuing from -304	169 (29.0)	169 (47.6)	0
Continuing from -310	122 (20.9)	122 (34.4)	0
DISCONTINUATION – n (%)			
Total	289 (49.6)	149 (42)	140 (61.4)
Subjects withdrew consent	105 (18)	57 (16.1)	48 (21.2)
AE's	88 (15.1)	42 (11.8)	46 (20.2)
Unsatisfactory effect	57 (9.8)	25 (7.0)	32 (14.0)
Lost to follow up	22 (3.8)	13 (3.7)	9 (3.9)
Protocol Deviation	9 (1.5)	6 (1.7)	3 (1.3)
Administrative	8 (1.4)	6 (1.7)	2 (0.9)

EFFICACY RESULTS:

Study VOSG-PN-309 an uncontrolled, long-term safety study in knee OA, included periodic assessments of a limited number of efficacy outcomes. The study was designed as a long-term safety study and limited efficacy measures were obtained for comparison with baseline values The measures were the WOMAC 3.1 LK Scale scores for pain, stiffness and physical function and assessments were made at baseline, 3, 6, 9 and 12 months shown below in Tables 10.1.5.x to 10.1.5.z. There was a consistent reduction in each of the measured WOMAC Indices at each post-baseline determination (Pain: 40% reduction, Stiffness 30% reduction and Physical Function 35% reduction).

TABLE 10.1.5.3 : SHOWING THE EFFICACY OUTCOMES: WOMAC PAIN INDEX- VOSG-PN-309

	WOMAC Pain Index (Scale = 0 to 20)	Total N = 578	One Knee N = 350	Two Knees N = 228
Baseline	N	578	350	228
	Mean ± SD	9.19 ± 3.85	9.12 ± 4.05	9.29 ± 3.52
Month 3	N	520	324	196
	Mean ± SD	5.90 ± 3.82	5.50 ± 3.73	6.55 ± 3.89
	Change from baseline Mean ± SD	3.39 ± 4.31	3.69 ± 4.50	2.90 ± 3.94
Month 6	N	437	285	152
	Mean ± SD	5.63 ± 3.89	5.18 ± 3.63	6.48 ± 4.21
	Change from baseline Mean ± SD	3.68 ± 4.41	3.96 ± 4.53	3.13 ± 4.14
Month 9	N	384	253	131
	Mean ± SD	5.63 ± 4.09	5.08 ± 3.73	6.71 ± 4.53
	Change from baseline Mean ± SD	3.62 ± 4.53	4.03± 4.51	2.82 ± 4.46
Month 12	N	268	146	122
	Mean ± SD	6.10 ± 4.17	5.82 ± 3.83	6.44 ± 4.54
	Change from baseline Mean ± SD	4.03 ± 4.61	4.84 ± 4.53	3.06 ± 4.54

Source: VOSG-PN-309 (CTD 5.3.5.2.1), Post-text Table 9.2

APPEARS THIS WAY
ON ORIGINAL

TABLE 10.1.5.v: SHOWING EFFICACY OUTCOMES: WOMAC STIFFNESS INDEX-VOSG-PN-309

	WOMAC Stiffness Index (Scale = 0 to 8)	Total N = 578	One Knee N = 350	Two Knees N = 228
Baseline	N	578	350	228
	Mean ± SD	4.14 ± 1.71	4.02 ± 1.80	4.32 ± 1.56
Month 3	N	520	324	196
	Mean ± SD	2.73 ± 1.73	2.53 ± 1.68	3.06 ± 1.75
	Change from baseline Mean ± SD	1.44 ± 1.85	1.56 ± 1.89	1.26 ± 1.77
Month 6	N	437	285	152
	Mean ± SD	2.77 ± 1.76	2.55 ± 1.73	3.18 ± 1.76
	Change from baseline Mean ± SD	1.42 ± 1.94	1.54 ± 2.02	1.18 ± 1.77
Month 9	N	384	253	131
	Mean ± SD	2.70 ± 1.92	2.50 ± 1.81	3.10 ± 2.06
	Change from baseline Mean ± SD	1.46 ± 2.11	1.58 ± 2.10	1.23 ± 2.11
Month 12	n	268	146	122
	Mean ± SD	2.99 ± 1.93	2.88 ± 1.90	3.12 ± 1.97
	Change from baseline Mean ± SD	1.50 ± 2.07	1.76 ± 2.14	1.20 ± 1.96

Source: VOSG-PN-309 (CTD 5.3.5.2.1), Post-text Table 9.2

TABLE SHOWING EFFICACY OUTCOMES: WOMAC PHYSICAL FUNCTION INDEX -VOSG-PN-309

	WOMAC Physical Function Index (Scale = 0 to 68)	Total N = 578	One Knee N = 350	Two Knees N = 228
Baseline	N	578	350	228
	Mean ± SD	32.21 ± 12.99	31.35 ± 13.48	33.53 ± 12.11
Month 3	N	520	324	196
	Mean ± SD	21.43 ± 13.33	19.95 ± 13.18	23.88 ± 13.25
	Change from baseline Mean ± SD	10.93 ± 13.53	11.69 ± 14.18	9.65 ± 12.30
Month 6	N	437	285	152
	Mean ± SD	20.70 ± 13.12	19.04 ± 12.65	23.83 ± 13.44
	Change from baseline Mean ± SD	11.54 ± 13.69	12.46 ± 14.21	9.81 ± 12.53
Month 9	N	384	253	131
	Mean ± SD	20.50 ± 14.07	18.78 ± 13.32	23.82 ± 14.91
	Change from baseline Mean ± SD	11.53 ±14.40	12.62 ± 14.49	9.43 ± 14.06
Month 12	N	268	146	122
	Mean ± SD	21.83 ± 14.45	20.91 ± 13.64	22.93 ± 15.34
	Change from baseline Mean ± SD	12.76 ± 14.79	14.89 ± 14.35	10.20 ± 14.98

Source: VOSG-PN-309 (CTD 5.3.5.2.1), Post-text Table 9.2

The potential implications for selection bias must be considered, as the subjects most likely to stay in the trial are those who are most satisfied with the product. On the other hand, only 57 subjects discontinued the study because of lack of therapeutic effect. Furthermore,

of the 289 subjects who discontinued prematurely, almost half did so at or before the Month 3 visit, so that the effect of selection bias would be attenuated considerably beyond that time point. The population for analysis of efficacy included all subjects treated in VOSG-PN-309. All three WOMAC indices showed improvement in OA symptoms over the treatment period, with mean scores typically lower by about 30-40% at end-of-study vs. baseline.

SAFETY RESULT

The population for analysis of safety included all subjects treated with DSG in the two controlled double-blind studies VOSG-PN-304 and VOSG-PN-310 and all subjects treated in VOSG-PN-309.

Serious AEs were reported in 3.1% of subjects. The most frequent SAE was pneumonia (including bronchopneumonia), which was experienced by 3 subjects. Diarrhea, acute pancreatitis, myocardial infarction, and asthma were reported as SAEs for 2 subjects each.

AEs were responsible for discontinuation of the study drug in 12.1% of patients. In comparison with the one-knee population, a larger proportion of subjects in the two-knee population experienced treatment-emergent AEs that led to discontinuation of the study drug (9.7%, versus 19.7%). The most common AE that led to discontinuation of the study drug was application site dermatitis, which was experienced by 6.2% of all treated subjects, 11.0% of the two-knee population, and 4.7% of the one-knee population. Three subjects discontinued due to increases in liver function tests.

An increase in ALT above 3x ULN in any post-baseline lab sample was reported in 1% of the subjects (0.7% and 1.8%, in patients who treated one or both knees, respectively), whereas the overall incidence of total bilirubin elevation above 2x ULN was 0.3% (0.3% and 0.4%, respectively). Only 1 subject had simultaneous elevations of ALT over 3 x ULN (417 U/L) and of total bilirubin over 2 x ULN (112.9 umol/L). In this subject, both parameters returned to normal ranges on continued treatment and remained normal until end of study.

Vital sign observations were unremarkable.

CONCLUSIONS

DSG was tolerated in long term use at doses of up to 32 g/day, and up to 12 months of dosing. The nature and distribution of AEs were influenced by dose (treatment of one vs. both knees) or duration of exposure.

Summary statistics for each efficacy parameter for all subjects treated at each of the times were calculated and differences with respect to baseline values were computed. Overall, there was a consistent reduction in each of the measured WOMAC Indices at each post-baseline determination (Pain: 40% reduction, Stiffness: 33% reduction and Physical Function: 35% reduction). The study demonstrated persistence of the treatment benefit of the topical drug up to 12 months.

SECTION 10.3:

STUDIES EVALUATING THE EFFICACY AND SYSTEMIC AVAILABILITY OF DEA DICLOFENAC PRODUCT.

TABLE : SHOWING PK/PD STUDIES IN HEALTHY VOLUNTEERS – STUDY VOSG-PN-107

STUDY # Goal(s) Country	STUDY DESIGN STUDY POPULATION ENDPOINTS	TOTAL # STUDIED AGE RANGE (MEAN) GENDER (M, F)	DOSAGE R & DURATION
VOSG-PN-107 1) Compare systemic exposure to diclofenac at steady state from DSG 1% and diclofenac DEA 1.16% 2) To determine the effect of applied heat and moderate exercise on bioavailability of diclofenac from DSG, 1%	Single-center, open-label, randomized, double 3-way crossover study	36 healthy volunteers representative of the target population with osteoarthritis. All Caucasians	4 g on one knee qid
	healthy volunteers	50 to 78 years (61.7 ± 7.8)	Three treatment periods of 7 days separated by 14-day washout periods.
Belgium	-Plasma diclofenac -urinary diclofenac 4'OH diclofenac	(12 m, 24 f)	

STUDY # Goal(s) Country	STUDY DESIGN STUDY POPULATION ENDPOINTS	TOTAL # STUDIED AGE RANGE (MEAN) GENDER (M/F)	DOSAGE Rx DURATION
VOSG-PE-113 1) To compare exposure to DSG 1% on one knee vs. on 2 knees and two hands (maximum exposure) vs. oral dosing of 50 mg Diclofenac sodium tablets. 2) Accumulation from D1 to D7 and pharmacodynamic effects of the treatments at D7	Single-center, open-label, randomized, 3-way crossover s Study	40	**topical**: **A**: 4 g on one knee (400 cm^2) qid **B**: 4 g to each of two knees (800 cm^2) and 2 g to each of two hands (400 cm^2) *i.e.*, 12 g on 1200 cm^2 qid **C**: **Oral** diclofenac sodium 50 mg enteric-coated tablets tid
	Healthy volunteers	59.9 ± 5.7 (range: 50 to 74) years old	7 days
Belgium	Plasma diclofenac and urinary diclofenac and 4'OHdiclofenac levels and derived PK parameters. Platelet aggregation and COX-1 and COX-2 inhibition on D7.	40 subjects (20m, 20 f)	

PLACEBO CONTROLLED STUDY USING DEA GEL- STUDY VOSG-PE-303

STUDY #	STUDY DESIGN	TOTAL # STUDIED	DOSAGE	RESULTS & RELEVANCE TO EFFICACY AND SAFETY
	STUDY POPULATION	AGE RANGE (MEAN)	Rx DURATION	
Country Goal	**ENDPOINTS**	**GENDER (M,F)**		
VOSG-PE-303 Germany To compare the efficacy of daily topical applications of diclofenac diethylamine (DEA) gel, 1.16% with vehicle	randomized, double blind, multi-center, placebo-controlled, parallel group study	237 patients (237 w)	4 g qid on one knee	statistically significant reduction in pain on movement averaged over days 1-14 vs. vehicle; significant improvement vs. vehicle for secondary efficacy criteria; comparable tolerability in both treatment groups
	patients with knee osteoarthritis	44-89 years (mean 66 in both groups)	3 weeks	
	primary efficacy endpoint: average over days 1-14 of pain on movement; safety	diclofenac DEA: 117 patients (38% m); vehicle: 120 patients (35% m)		
REFERENCE THERAPY CONTROLLED STUDY				
VE-OA-1 Germany To compare the efficacy of daily topical applications of diclofenac diethylamine (DEA) gel, 1.16% with oral ibuprofen	randomized, doubleblind, multi-center, active-controlled, double-dummy parallel group study	311 patients; safety population: 321 patients (310w, 1a)	diclofenac DEA gel: 3 g qid on both hands ibuprofen tablets: 400 mg tid orally	non-inferiority of topical diclofenac vs. oral ibuprofen; diclofenac DEA gel: 21.8% AEs, ibuprofen tablets: 26.9% AEs
	Patients with hand osteoarthritis	36-95 years (mean 60.7 in diclofenac DEA gel group, 63.2 in ibuprofen	3 weeks	

		tablets group)		
	primary efficacy endpoint: incidence of 40% improvement in global assessment of pain; safety	diclofenac DEA gel: 159 patients; ibuprofen tablets: 152 patients; safety population: diclofenac DEA gel: 165 patients (14% m); ibuprofen tablets: 156 patients (10% m)		

Other Pertinent Information

REFERENCES

EndNote Reference List Place Holder.

Reference Manager List Place Holder

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

/s/

Neville Gibbs
8/29/2007 06:17:29 PM
MEDICAL OFFICER

Mwango Kashoki
8/30/2007 08:56:09 AM
MEDICAL OFFICER

DEPARTMENT OF HEALTH & HUMAN SERVICES Public Health Service

Division of Dermatologic and Dental Drug Products
Office of Drug Evaluation V
Center for Drug Evaluation and Research
Food and Drug Administration
Silver Spring, MD 20993-002

Tel 301-796-2110
FAX 301-796-9894

MEMORANDUM

Date: June 4, 2007

From: Brenda E. Vaughan, M.D., Medical Officer
Through: Markham Luke, M.D., Ph.D. Dermatology Team Leader and
Susan Walker, M.D., Division Director, DDDP

To: Bob Rappaport, M.D., Division Director
Division of Anesthesia, Analgesia and Rheumatology Products

Cc: Julie Beitz, M.D., Office Director, ODE 3
Christy Cottrell, Supervisory PM, DDDP

Re: Consult # 968, received from Division of Anesthesia, Analgesia and Rheumatology Products on 04/25/2007, and assigned on 04/27/2007

Executive summary

1) The three dermal safety studies (Cumulative Irritation Potential Study VOSG-PN-108), Skin Sensitizing Potential Study VOSG-PN-111, and UVA and UVB Phototoxicity Potential Study VOSG-PE-112) are adequate in design and study duration. Each clinical trial provided data derived from sufficient numbers of evaluable subjects. Cumulative irritation potential was assessed in two studies (Study VOSG-PN- 108- cumulative irritation potential and during induction phase of Study VOSG-PN-111). Photocontact allergic potential data were not included in the NDA submission and are recommended potentially as post-marketing studies. Labeling should reflect potential for concern, e.g.

Applicant's interpretations of the findings are accurately reflected in the Discussion Sections for each study. Line listings for the dermal safety studies do not support the Applicant's Summary of Clinical Safety conclusion that none of the studies identified

"any evidence" that the drug had potential to produce irritation, sensitization or photo-toxicity.

The Applicant concludes that there is a weak sensitization potential signal for diclofenac sodium gel, 1%, and its vehicle gel based on conditions of Study VOSG-PN-111 (Section 11, Discussion and overall conclusions, pg. 38, Novartis Consumer Health, Inc., Clinical Study Report 9-November-2004 Study No. VOSG-PN-111). In this study, 233 subjects completed all phases. One subject exhibited a possible sensitization response; however, was negative at rechallenge. Potential for irritation was demonstrated during the induction phase (irritation potential assessment) where seven subjects exhibited strong erythema or erythema and papules (a score of 3) resulting in moving the patch applications to naïve adjacent sites. For 5 of the 7 subjects, the score of 3 was associated with active and vehicle only.

In Study VOSG-PN-108, 36 subjects completed all aspects of the cumulative irritation study. The cumulative irritation effect of diclofenac sodium gel, 1%, applied under occlusion on the skin over 21 days, as assessed by irritation rate, mean irritation score, cumulative irritation score, frequency indices and time to irritation was minimal when compared to that of SLS (positive irritant control) and slightly higher than that of the blank patch (negative irritant control). For Study VOSG-PE-112, the obtained results indicate an absence of a clinically relevant photo-toxicity potential of Diclofenac Sodium Topical Gel, 1%.

2. According to the Summary of Clinical Safety (CTD 2.7.4, Section 4.2.4, Severity of events, pg. 58) the severity of an AE was determined by the AE report in the subject's CRF. An event was to be reported as mild, moderate or severe in accordance with the investigator's interpretation of the AE. The severity categories used are acceptable for grading dermatologic adverse events, although for dermatologic AEs of interest, a grading scale with descriptors would have been useful in order to reduce inter-investigator variability.

3. This Reviewer recommends that labeling should include language to the effect that [redacted]. Photoallergenicity studies have not been conducted with diclofenac sodium topical gel, 1%. As previously mentioned, labeling should reflect that adequate precautions should be taken to minimize sunlight exposure.

Under clinical use conditions in larger numbers of subjects, an 11% incidence attributed to application site dermatitis (CTD 2.7.4 Summary of Clinical Safety) was observed. Overall, DSG-treated subjects had a reported rate of skin-related AEs of 11.4% compared to 4.3% for the vehicle-treated subjects. Application site dermatitis was the single AE with the highest overall reporting rate for both treatment groups (DSG 3.4% vs. Vehicle 0.7%) (Novartis Consumer Health Confidential, Page 32, CTD 2.5 Clinical Overview).

Background

Diclofenac sodium gel 1% (DSG) is a non-selective NSAID product being developed [redacted] joints amenable to [redacted] treatment, such as the hands and knees". The route of administration is topical.

The NDA contains reports of four Phase 3 safety and efficacy studies (2 studies of hand OA and 2 studies of knee OA). The safety population consists of 913 subjects exposed to DSG and 876 subjects exposed to vehicle (placebo).

The proportion of subjects reporting skin adverse events in the major safety population was 10.1% in subjects treated with active drug and 3.8% in subjects treated with vehicle. In both the active and the vehicle groups, the most frequently reported skin adverse events were application site dermatitis and rash. The difference in the incidence of application site dermatitis and application site erythema was statistically significant between the 2 treatment groups. All adverse events resolved when medication was discontinued. No skin-related serious adverse events (SAE's) occurred.

All the adverse events were coded from a common dictionary (MedDRA Version 9.1) and where necessary adverse events were recoded accordingly.

The applicant did not utilize a specific dermal classification system to characterize the severity of the rash. The clinical investigators categorized the dermal findings as mild, moderate, or severe to determine if AE had worsened or not.

The applicant also submitted the results of three Dermal Sensitivity Studies to evaluate:
1) Cumulative irritation potential (Study VOSG-PN- 108)
2) Skin sensitizing potential (Study VOSG-PN-111)
3) UVA and UVB phototoxicity potential (VOSG-PE-112)

The sponsor states that the dermal sensitivity studies were negative.

The Applicant confirmed (communication dated May 19, 2007) that studies VOSG-PN-108, VOSG-PN-111 and VOSG-PE-112 were performed with the to-be-marketed formulation of diclofenac sodium topical gel, 1%.

The consult specifically poses the following:
1) Evaluate the results of the 3 Dermal Sensitivity Studies, and comment on the applicant's interpretation of the findings.

2) Discuss whether the scheme for grading dermatologic adverse events and the way those events were evaluated was acceptable for a topical product.

3) Make suggestions regarding appropriate labeling of the dermatologic safety findings for this product.

Materials Reviewed
Dermal safety study data and dermatologic safety findings included in the Summary of Clinical Safety submitted to NDA 22-122.

Review

The Applicant submitted data from the following three dermal safety studies conducted in healthy human subjects: VOSG-PN-111 (cumulative irritation and sensitization), VOSG-PN-108 (cumulative irritation), and VOSG-PE-112 (phototoxcitiy study) as noted in Applicant's Table 9-1 below. Assessment of photocontact allergic potential was not included in the NDA submission.

Table 9-1 Special studies conducted in support of DSG, 1%

Table 9-1 Special studies conducted in support of DSG, 1%

Study No. (country)	Study title	Number of subjects enrolled / completed	Outcome
VOSG-PN-108 (USA)	Study of the skin cumulative irritation potential of diclofenac sodium gel, 1%, when applied topically to normal, healthy volunteers (21-day cumulative irritation test)	42/36	No evidence of cumulative skin irritation potential
VOSG-PN-111 (USA)	Study of the skin sensitization potential of diclofenac sodium gel, 1%, when applied topically to healthy volunteers	260/233	No evidence of sensitization potential; no cumulative irritation potential
VOSG-PE-112 (UK)	Assessment of photo-toxicity potential of Diclofenac Sodium Topical Gel, 1%, after single cutaneous application and UV exposure in healthy volunteers	35/35	No evidence of clinically relevant photo-toxicity potential

According to the Applicant (Summary of Clinical Safety, pg. 128) none of the studies identified any evidence that the drug had potential to produce irritation, sensitization or photo-toxicity.

Reviewer comment:

The line listings do not support the Applicant's conclusion in that none of the studies identified "any evidence" that the drug had potential to produce irritation, sensitization or photo-toxicity. See comments for the studies that follow.

VOSG-PN-111 Title: "Study of the skin sensitization potential of diclofenac sodium gel, 1%, when applied topically to healthy volunteers". Study period: First subject enrolled: 11-May-2004 Last subject completed: 23-Jul-2004. A total of 260 subjects were enrolled, and 233 subjects completed all phases of the study.

The primary objective of this study was to determine the potential of diclofenac sodium gel, 1%, to cause sensitization by repeated topical occlusive application to the skin of healthy human volunteers. The secondary objective of this study was to evaluate the potential of diclofenac sodium gel, 1%, to cause cutaneous irritation by repeated topical occlusive applications to the skin of healthy human volunteers.

This was an evaluator-blind, randomized, multiple-application, two-center, repeated insult patch test study to evaluate the skin sensitization potential of diclofenac sodium gel, 1%, in comparison to that of the diclofenac sodium gel vehicle and a blank patch (negative control). All subjects were patched with the same 3 test materials. The test materials were randomized and administered to assigned sites on the subject's back.

Study personnel applied all test materials 3 times per week for 3 weeks (9 applications). The study consisted of the following phases: screening and enrollment (Days –7 to 1), induction, which included patch application, patch removal, and site evaluation (Days 1 to 22), rest (Days 23 to 35), and challenge, which was one patch application with evaluations for sensitization at 48 hours + 30 minutes, 72 hours, and 96 hours after patch application (Days 36 to 40). A rechallenge was conducted if needed, i.e., if there was a "+" grade at any challenge patch site at 72 hours or 96 hours.

Criteria for evaluation:
Sensitization: Sensitization was evaluated during the challenge phase using the International Contact Dermatitis Research Group scale. A crescendo evolution of intensity or the presence of a score of "+"at any time was suggestive of sensitization and required a rechallenge. If a score of "++" or greater occurred during the challenge and the reaction was considered a sensitization reaction, the subject was discontinued from further participation in the study; the subject was not rechallenged.

Irritation: Irritation was assessed after each induction application. Reactions to the test products including effects on superficial layers of the skin were scored from 0 (no evidence of irritation) to 7 (strong reaction spreading beyond test site). Superficial effects included glazing, peeling, fissures, and petechial erosions.

Safety: Safety was assessed through the monitoring and recording of all adverse events (AEs) andserious AEs. Adverse event information was recorded on an AE Case Report Form as each incidentwas noted (whether volunteered by the subject, discovered by investigator questioning, or detectedthrough physical examination).

Statistical methods: Key data collected during the study was summarized for each treatment using summary statistics. There were two populations. Population I included only those subjects who completed 8 evaluations during the induction phase and 3 evaluations during the challenge phase. Population II included all subjects who received the study products and was also used for safety evaluation.

The sensitization potential was assessed only for subjects who completed the induction and challenge phases (Population I) and those subjects who showed evidence of sensitization, even though they hadnot completed all visits (ie, score of "++" or greater during challenge). The diagnosis of sensitization was a judgment of the investigator and was based on observations at challenge and rechallenge, and on the patterns of reactivity during induction. The rate of sensitization was summarized for each test product based on the results of the challenge phase and the rechallenge phase: any crescendo reaction with a score of at least "++," whether at challenge or at rechallenge, was considered a sensitization reaction.

Irritation potential was assessed for both populations (Populations I and II). The irritation rate was determined for each level of irritation score as the percentage of subjects with any irritation score ≥ 1, ≥ 2, or ≥ 3. The McNemar's test was used to compare differences among treatment groups at each evaluation and on an overall basis across all evaluations.

The Frequency Index (FI) was calculated for each subject in order to interpret irritation responses based on response frequency and treats each individual irritation score as a distinct threshold.

Results

For sensitization potential the findings indicates that based on conditions of the study for the Population I, none of the 233 subjects demonstrated evidence of sensitization to the study products tested (diclofenac sodium gel, 1%, the vehicle gel or the blank patch). One subject demonstrated erythema and edema (+) in response to the vehicle gel at the 48-hour challenge evaluation. The subsequent responses for this subject were minimal (weak and questionable reaction = ? score) at 72 and 96 hours challenge evaluations. When reevaluated during the rechallenge phase, the subject showed no evidence of sensitization.

Statistical analysis of the irritation potential of the irritation rates for Population I indicated that irritation scores ≥ 3 were associated with each of the 3 treatments; however, diclofenac sodium gel, 1% was not different from its vehicle gel or the blank patch ($p \geq 0.2500$).

During the induction phase (irritation potential assessment), seven subjects exhibited a score of 3 (strong erythema or erythema and papules) resulting in the move of patch applications to a naïve adjacent site. For 2 of the 7 subjects, the score of 3 was associated with all of the 3 treatment conditions including the blank patch.

Adverse Events: Fourteen subjects experienced 14 treatment-emergent AEs. The most commonly reported AE was headache. Six subjects discontinued the study due to AEs. All AEs leading to treatment discontinuation were of moderate severity and none were serious. There were no treatment related AEs.

Sensitization potential evaluation

During the challenge phase (sensitization potential evaluation), Subject# 240 demonstrated a possible positive response that required a rechallenge. The subject exhibited erythema and edema (+) for the vehicle gel at the time of the 48 hour challenge evaluation; however, there was no sensitization at the rechallenge evaluation, indicating that the sensitization potential of diclofenac sodium gel is very weak.

Reviewer comment:

Based on line listings, this reviewer concurs with the Applicant's statement (Section 11, Discussion and overall conclusions, pg. 38, Novartis Consumer Health, Inc., Clinical Study Report 9-November-2004 Study No. VOSG-PN-111) that based on conditions of this study, data indicates that the sensitization potential signal of diclofenac sodium gel, 1%, and its vehicle gel are weak.

During the induction phase of the study where cumulative irritancy was assessed, seven subjects exhibited strong erythema or erythema and papules (a score of 3) resulting in the moving the patch applications to naïve adjacent sites. For 5 of the 7 subjects, the score of 3 was associated with active and vehicle only (excluding the blank patch).

VOSG-PN-108 Study of the skin cumulative irritation potential of diclofenac sodium gel, 1%, when applied topically to normal, healthy volunteers (21-day cumulative irritation test). This was an evaluator-blinded, randomized, multiple-application, single-center study to test the cumulative irritation potential of diclofenac sodium gel, 1%, its vehicle, sodium lauryl sulfate (SLS), 0.1% (positive irritant control), and a blank patch (negative irritant control). According to the Applicant, the Lanman method was used.

Study period: First subject enrolled: 15-Mar-2004 Last subject completed: 05-Apr-2004. A total of 42 subjects were enrolled into the study, and 36 subjects completed all aspects of the study.

All subjects received the same 4 treatments. Treatments were randomized and applied to assigned patch sites on each subject's back. Starting on a Monday, there were 15 consecutive patch applications with subsequent removal and evaluation over 21 days. Irritation was assessed 24 hours after each weekday patch application except for Friday when the patch was removed and irritation was evaluated on the following Monday, 72 hours after application.

Reactions to the test products and effects on superficial layers of the skin were scored on a scale from 0 (no evidence of irritation) to 7 (strong reaction spreading beyond test site) with annotations for superficial effects that included glazing, fissures, exudates, and petechial erosions. Patch applications were terminated if a score of 3 or greater was observed, or at the discretion of the Investigator. Population I comprised subjects who completed all evaluations or discontinued applications due to limiting reactions (i.e., an irritation score of ≥ 3). Population II included all subjects with at least one application of investigational treatment. Overall safety was assessed by monitoring any treatment emergent AEs.

Study VOSG-PN-108 Results

A total of 42 subjects were randomized and all received treatment. Six subjects discontinued the study prematurely: 3 withdrew consent, 1 was lost to follow-up, 1 missed an evaluation visit and 1 was withdrawn at the Investigator's discretion following severe tape reaction. Thirty-six (36) subjects completed all aspects of the study and represent the population whose results are presented in the following paragraphs.
A second population was composed of all 42 subjects and was the population evaluated for safety. In the final study report, all the irritation analyses were performed on both populations and no important differences in the outcomes were observed between them.

The demographic characteristics of the 36 subjects who completed all aspects of the study were as follows. Their mean age was 44 years (range: 28-66); the sex distribution was 14% (5/36) male and 86% (31/36) female. With respect to ethnicity, 58% (21/36) were White, 17% (6/36) were Black and 25% (9/36) were Other.

Irritation evaluations: In all analyses, diclofenac sodium gel, 1% was similar in irritation potential to its vehicle and significantly less irritating than the positive control (SLS, 0.1%).

Adverse events: One subject reported 1 mild AE (body ache) during the study. The AE resolved the same day and was considered by the Investigator to be unrelated to study medication. No other AEsor serious adverse events (SAEs) occurred during the study.

Applicant's Conclusions: The cumulative irritation effect of diclofenac sodium gel, 1%, applied under occlusion on the skin over 21 days, as assessed by irritation rate, mean irritation score, cumulative irritation score, frequency indices and time to irritation was minimal when compared to that of SLS (positive irritant control) and only slightly higher than that of the blank patch (negative irritant control).

Reviewer comment:
This Reviewer concurs with the Applicant's conclusion for Study VOSG-PN-108.

Study VOSG-PE-112: Title of study: "Assessment of photo-toxicity potential of Diclofenac Sodium Topical Gel, 1%, after single cutaneous application and UV exposure in healthy volunteers" Study period: First subject enrolled: 17-Feb-2005 and the last subject completed: 11-Mar-2005. Thirty-five (35) subjects (33 female and 2 male, aged 18-60 years) completed the study and analyzed.

Objectives: Primary objective: To assess the photo-toxicity potential of Diclofenac Sodium Topical Gel, 1%, in 30 healthy male and female volunteers after single application and UV exposure. Secondary objectives: To evaluate overall safety and tolerability of Diclofenac Sodium Topical Gel, 1%.

Methodology: This was a single-center, single-application, randomized, double-blind, within-subject comparison study of active and vehicle vs. blank controls (area treated but not irradiated and area not treated but irradiated) . Enrolled subjects were to be healthy Caucasian males or females 18 to 60 years of age with skin photo-type II (always burns easily, tans minimally) or III (burns minimally, tans gradually and uniformly light brown) according the Fitzpatrick classification of skin phototypes.

Following the determination of individual minimal erythema dose (MED), three sets of 3 areas were delineated on the skin of the subjects' backs. For each set, one area was treated with Diclofenac Sodium Topical Gel, 1% , another one with vehicle gel , and the last one remained untreated. The order of these treatments was allocated according to a pre-established randomization list. All drugs were directly applied into Finn Chambers that were fixed on the skin with medical paper tape (empty Finn Chambers for untreated control areas).

After 24 hours, Finn chambers were removed and two of the three sets were irradiated: one with 20 J/cm2 UV A (320-400 nm) and another one with 0.75 MED UV B (280-320 nm); one set remained nonirradiated. Cutaneous readings by two independent assessors

were performed before treatment application, as well as 15-30 min after Finn chamber removal and at 10 min, 24 hours and 48 hours after UV exposure. The following erythema intensity scores were used: 0 = no erythema, 0.5 = barely visible, 1 = mild, 2 = moderate, 3 = severe. Having non-integer scales is not optimal study design. Future studies should not have ½ scores for erythema.

Photo-irritation intensity for each subject was defined as the highest photo-toxic reaction Score (PtRS) observed at any time after irradiation (10 min, 24 hours or 48 hours after UV exposure). Since all subjects were given all treatments and irradiations, McNemar's test was used for all pair-wise comparisons. The probabilities (p values) given are two-sided exact values.

Duration of treatment: Single drug application for 24 h (Day 1 → Day 2)
Criteria for evaluation:
Photo-toxicity potential: Cutaneous macroscopic evaluation of treated and non-treated irradiated and non-irradiated areas; determination of erythema reaction score (ERS), photo-toxic reaction score (PtRS) and photo-irritation intensity in individual subjects. Any ERS greater than zero was considered to reflect an erythema reaction. Any erythema occurring on the treated and irradiated areas was considered a result of a photo-toxic reaction if its corresponding ERS was higher than that observed on treated, non-irradiated and on non-treated, irradiated control areas (i.e. both differences >0). In this case, a PtRS was determined (defined as the lower numerical value of the two differences). Photo-irritation intensity for each subject was defined as the highest PtRS observed at any time after irradiation (10 min, 24 h or 48 h after UV exposure).

Safety: Assessment of adverse events (AEs; including pregnancies) and serious AEs (SAEs); physical examination & vital signs; investigator-rated overall tolerability

Statistical methods: The definition of photo-toxic reactions and photo-irritation intensities after irradiation was based on scored erythema assessments (as described above). As all subjects were given all treatments and irradiations, the McNemar test was used for all pair-wise comparisons, in recognition of the within-subject nature of the comparisons. The probabilities (p-values) given are two sided exact values. The SAS® (version 8.2) was used to perform the analyses. No interim analyses were planned or performed.

Results for Study VOSG-PE-112:
A total of 35 subjects were enrolled in order to have 30 subjects evaluable. All completed the study and all were included in the population analyzed. Of the 35 subjects, 2 (6%) were male and 33 (94%) were female. Their median age was 38.0 years (range: 18-60) and all were Caucasian.

Photo-toxicity potential: All observed skin erythema were rated as 'barely visible' or 'mild' erythema, and none as 'moderate' or 'severe'. No erythema was seen at the readings at 48 h. The highest frequency of skin erythema was observed after UVA irradiation, but there were no statistically significant differences in the number or severity

of erythema reactions between the different study treatments, including the non-treated blank control. A limited number of at most 'barely visible' erythema reactions were seen with UVB, and both their number and severity were lower than those observed without UV irradiation.

Eight (8) subjects out of 35 exhibited possible photo-toxic reactions with Diclofenac Sodium Topical Gel, 1% (6 with UVA and 2 with UVB). All were seen at 10 min or 24 h after irradiation and no subject reacted twice in time. No photo-toxic reaction was seen at 48 h readings. All reactions had the lowest possible photo-irritation intensity of 0.5. All possible photo-toxic reactions observed were weak and transient. There were overall slightly more possible reactions with UVA than with UV B. However, there was no statistical evidence that photo-toxic reactions were more likely under active treatment than under vehicle treatment.

Safety: There were few AEs (nine in total) and all were judged to be not drug-related. There were no SAEs and no deaths. All pregnancy tests were negative.

Conclusion
From the data of the present trial, there is no evidence for a clinically relevant photo-toxicity potential of Diclofenac Sodium Topical Gel, 1%, in healthy volunteers, after a single dose applied for 24 h (under occlusion) and followed by single UVA or UVB irradiation.

Reviewer comment:
The Applicant's summary that the obtained results indicate an absence of a clinically relevant photo-toxicity potential of Diclofenac Sodium Topical Gel, 1% is consistent with the data.

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

/s/

Brenda Vaughan
6/11/2007 11:00:55 AM
MEDICAL OFFICER

Markham Luke
6/13/2007 05:04:45 PM
MEDICAL OFFICER
Consult #968 for the Division of Anesthesia, Analgesia, and Rheumatology regarding Dermal Safety studies. Concur with Dr. Vaughan's recommendations.

Susan Walker
6/13/2007 09:56:52 PM
DIRECTOR