CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

21-789

STATISTICAL REVIEW(S)

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	21-789 / 000
Drug Name:	Metronidazole Gel, 1%
Indication(s):	Roseacea
Applicant:	Dow Pharmaceutical Sciences
Date(s):	Received 8/30/2004, user fee (10 months) 6/30/2005
Review Priority:	Standard
Biometrics Division:	Division 3, HFD-725
Statistical Reviewer:	Steve Thomson, HFD-725
Concurring Reviewers:	Team Leader: Mohamed Alosh, Ph. D., HFD-725
Medical Division:	Dermatological and Dental Drug Products, HFD-540
Clinical Team:	Reviewer: J. Porres, M.D. Team Leader: M. Luke, M.D., Ph.D.
Project Manager:	K. Bhatt, HFD-540
Keywords:	Active control/noninferiority, adverse events, Bayesian Analysis Nonparametric tests, Cochran-Mantel-Haenszel, Confidence Interval

Table of Contents

1. EXECUTIVE SUMMARY

According to the Sponsor, metronidazole cream, 1% (Noritate™) is the only once-a-day application product available to patients for the treatment of rosacea. The Sponsor claims that a 1% gel formulation might be more attractive to patients. The single study reviewed here was designed by the Sponsor to support a 505(b)2 submission comparing this gel formulation to Noritate cream.

1.1 Conclusions and Recommendations

NDA 21-789 for Metronidazole Gel, 1%, for the treatment of rosacea, was submitted on August 30, 2004. The Sponsor provided the results of a single study comparing metronidazole gel, 1 %, to its vehicle and to its active comparator, Noritate cream, 1%, to assess the efficacy and safety of metronidazole gel. The protocol for the study specified that the primary efficacy measures were to be based on blinded investigator assessments of the signs and symptoms of rosacea, as measured by 1) the percent change from baseline in inflammatory lesion counts and 2) a 5-level Investigator's Global Severity Score. For the actual analysis, the 5-point Investigator's Global Severity Score was dichotomized so that scores of clear or almost clear (a score of 0 or 1) were interpreted as a success on this endpoint, other scores as a failure. The protocol further specified that demonstrations of non-inferiority of Metronidazole Gel, 1% relative to Noritate Cream, 1%, were to be based on a one-sided 97.5% confidence interval (C.I.) approach with a non-inferiority margin of 10% for both the percent reduction in inflammatory lesion count and the success rate in the dichotomized Investigator's Global Severity score.

The study results showed an overall reduction in lesion counts for all three treatment groups. In the Metronidazole gel group the mean percent reduction in inflammatory lesion counts was 21.6% (median 26.8%) at the Week 2 visit and increased to 50.7% (median 66.7%) in the Week 10, end of study, intent-to-treat (ITT) last-observation-carried-forward (LOCF) population. The mean percent reduction was 21.4% (median 26.3%) at Week 2 and 46.4% (median 58.3%) for the Week 10 ITT-LOCF analysis in the Noritate treatment group. In the gel vehicle group, the mean percent reduction was 14.1% (median 20.6%) at Week 2 and 32.6% (median 46.2%) for the Week 10 ITT-LOCF group. Whether based on the original data or the ranked data, an ANOVA of the Week 10, end of study, ITT LOCF population comparing Metronidazole gel to vehicle showed statistically significant differences in favor of Metronidazole gel ($p< 0.0001$ for both the percent change from baseline and the ranked percent change). However, the percent change from baseline is sensitive to outliers. One alternative to the rank transformed analysis is to adjust for outliers by to replacing values below -100%, with -100% (i.e., Winsorize the data). Both Winsorized and simple change from baseline gave similar results in terms of statistical significance to the percent change or the ranked per cent change described above.

Non-inferiority analyses are usually done using confidence intervals. Using the rank transformed analysis described in Appendix 1, the lower limit of the one-sided 97.5% confidence interval (C.I.) for the difference in percent change from baseline between Metronidazole gel and

Noritate in the ITT-LOCF population had a lower limit of 0.0%. Using the data in the original scale gave results in the ITT-LOCF data (lower bound of 97.5% C.I. of -2.18%) and Per Protocol (PP) data (lower bound 97.5% C.I. of -3.0%). Since all of these are above the agreed to boundary of -10.0%, we would conclude non-inferiority of Metronidazole gel to Noritate for the percent change from baseline. Appendix 10 has a preliminary Bayesian analysis of the non-inferiority of the primary endpoints.

Success rates for the dichotomized Investigator's Global Severity Score for the primary Week 10 ITT-LOCF analysis were comparable between Metronidazole Gel, 1% and Noritate Cream, 1%, 38.4% and 35.5%, respectively, versus a success rate for the vehicle gel group of 27.7%. In this population the Week 10 ITT-LOCF success rates for Metronidazole gel versus vehicle were statistically significantly better for both the ITT (p=0.0069) and the PP populations (p=0.0505). The lower limit of the one-sided 97.5% C.I. on the difference between success proportions was greater than -0.10 for the nominal Week 10 analysis in both the ITT-LOCF (lower 97.5% C.I. of -2.8%) and PP (lower 97.5% C.I. of -5.0%) populations. Thus one would conclude that non-inferiority was established in both populations.

Note that there does seem to be a small tendency for patients to apply more of Metronidazole gel than Noritate. This issue is addressed in Appendix 4.

1.2 Brief Overview of Clinical Studies

This study was designed as a multi-center, randomized, investigator-blinded, active and vehicle controlled, parallel group comparison in patients with rosacea. A total of 1299 patients in 54 centers in the United States were randomized 3:3:1 to either Metronidazole Gel, 1%, Noritate™ Cream, 1%, or vehicle gel. Patients were instructed to apply the study medication topically to the face once daily at bedtime. Duration of treatment was for 10 weeks with evaluations at baseline and at nominal Weeks 2, 4, 7, and 10. Efficacy assessments were based on blinded investigator assessments of the signs and symptoms of rosacea, as measured by the percent change from baseline in inflammatory lesion counts and a 5-point Investigator's Global Severity Score. The study was initiated on 6 June 2003, and terminated with the last patient on 12 February 2004.

1.3 Statistical Issues and Findings

1. The primary analyses specified in the protocol were based on the percent change from baseline in the Week 10 inflammatory lesion count and the dichotomized investigator global evaluation, including non-inferiority testing using the intent-to-treat (ITT) and the per protocol (PP) populations and superiority testing using the ITT population. Missing values in the ITT population were to be imputed using last observation carried forward (LOCF) techniques. The tests for demonstrating non-inferiority of Metronidazole Gel, 1%, relative to Noritate Cream, 1% were based on a one-sided 97.5% confidence interval (C.I.) approach with a non-inferiority margin of 10% for the percent reduction in inflammatory lesion count and the success rate in the dichotomized Investigator's Global Severity Score. For the ITT-LOCF population, at Week 10

both Metronidazole Gel and Noritate were shown to be statistically significantly better than Metronidazole Vehicle for each of following measures: the percent change from baseline, the ranked percent change from baseline, simple change from baseline, and the dichotomized investigator global evaluation.

2. The protocol specified that if the data for the percent reduction in inflammatory lesions were skewed, the analysis would be based on rank transformed data. Simulations conducted by this reviewer suggest that for smaller sample sizes than used here, ANOVA is quite robust to skewness, and the proposed rank transform is unneeded. In general, when feasible, for interpretability an analysis of data based on the original scale would be preferable to an analysis in a transformed scale. Further, note that to a first order approximation the percent reduction in lesion count resembles a log transform and thus would seem to be a correction for skewness. However, while there can only be a 100% reduction in lesion count, the increase (i.e., a negative decrease) is unbounded. In practice, this asymmetry tends to lead to outliers. The Division of Dermatological and Dental Drug Products has noted that the actual change from baseline seems less sensitive to outliers than the percent change and hence, is generally recommended for analysis instead of the percent change. An alternative approach to the analysis was to Winsorize the percent reduction to -100.0% to reduce the effect of the extreme outliers, which were all in the negative direction. Thus, analyses of the simple change from baseline, the percent change from baseline, the rank transform of the percent change, and results from the Winsorized percent change are all discussed here.

3. The Medical team expressed interest in assessing the effect of actual usage of the drug product. The approach used here models the response at the end of the study using the amount of drug product used. Nonlinearity of treatment was assessed by quadratic and cubic terms, along with corresponding homogeneity over treatment. These results for the primary endpoint in the ITT-LOCF population are presented in Appendix 4.

4. For the analysis of data from small centers (i.e., an investigator with less than 15 patients in either active arm, or less than five patients in the vehicle arm) patients were to be pooled with the center having the largest number of randomized patients to define "analysis centers." Then the data of the Investigator with the second smallest number of randomized patients was to be combined with that of the Investigator with the second largest number of randomized patients, and so on, for all centers that did not have a minimum number of patients per treatment arm. A table comparing the pooled analysis centers with the original investigator sites is given in Appendix 9.

5. The possible differential effect of centers was evaluated by assessing interaction terms. The Breslow-Day test for homogeneity of the odds ratio was to be used to assess consistency in overall success rate across analysis centers. Significant qualitative interactions were to be investigated to determine if pooling was justified. The protocol specified that the primary ANOVA analysis was not to include treatment by center interaction. Such interaction was to be tested in a secondary model, and if significant the model would be modified. But effectively that defines a pretest for interaction. However, note that the usual nominal significance levels of the

tests of models without the interaction term are not computed as conditional on this pretest. Largely for this reason, when there are many degrees of freedom in the data, this reviewer would be inclined to analyze the more general model retaining the interaction terms. However, the protocol indicates that the main analysis does not include the interaction terms, and the interaction analyses are to be supportive. So in the analyses presented here, interaction terms are not included. Conclusions would be the same with or without the interaction terms.

2. INTRODUCTION

2.1 Overview

According to the Sponsor "Metronidazole is already marketed in different formulations as cream, gel, and lotion, and has been proven effective in the treatment of rosacea. Currently, metronidazole cream, 1% (Noritate™) is the only once-a-day application product available to patients for the treatment of rosacea. The rationale for evaluating metronidazole in a 1% gel formulation is to provide affected individuals with a more cosmetically appealing alternative once-a-day formulation, as daily use of gels is generally preferred over cream formulations. This Phase 3 clinical study was undertaken in support of a 505(b)2 Registration Authorization in the United States."

A Pre-IND meeting was held with the Agency on 12 February 2001 to discuss the clinical development plan for the treatment of rosacea. Based on the discussion at this meeting, the Sponsor submitted the initial IND on 15 March 2002. The Sponsor was to conduct a single Phase 3 non-inferiority study using Noritate Cream as the reference-listed drug. On 9 August 2002 a proposed Phase 3 protocol was submitted for special protocol assessment.

Based on subsequent correspondence between the Sponsor and the Agency and upon Agency request, the Sponsor submitted a Statistical Analysis Plan (SAP) for the Phase 3 protocol to the Agency on 5 February 2003, and a Clinical Protocol Amendment, including the SAP, on 30 July 2003. This amended Phase 3 protocol included: a modified Evaluator's Global Severity Scale (static score) and Erythema Severity Scale, a study duration of 10 weeks similar to that of the comparator drug Noritate Cream, 1%, a three-arm investigator-blind design, the use of last observation carried forward (LOCF) for discontinued patients, use of a one-sided 97.5% confidence interval approach with a non-inferiority margin of 10%, changes in the number of study sites and the number of patients per arm as reflected by modified power calculations, and the use of a non-parametric ranking method. There was an issue about the length of the time on study, but the change in study duration (from 12 weeks) to 10 weeks was done at the recommendation of the Medical team since that was duration of its active comparator (review filed September 24, 2002).

The Sponsor states that the independent Investigator Review Board, which had jurisdiction over most of the investigational sites in the Phase 3 study, would not approve the use of Noritate Cream, 1% or Metronidazole Gel, 1%, in pregnant or nursing women. Thus, the

Sponsor decided not to pursue enrollment of pregnant or nursing women in the Phase 3 study (revision to protocol submitted 30 July 2003) as was previously recommended by the Agency.

In a review (filed 22 September 2003, after study initiation) the biostatistical reviewer noted that the expected success rate using the Investigators Global Severity Score was about 20% in the two active groups and 5% in the control group. The manner of assessing non-inferiority proposed by the Sponsor is slightly different than the method proposed by that reviewer, but it does imply that the two active groups would be assessed as equivalent on the IGSS as long as Metronidazole gel preserved as much as 50% of the effect of Noritate (33% by her calculations). However, the actual preservation levels achieved were considerably higher than 50%.

The study was initiated on 6 June 2003, and terminated with the last patient 12 February 2004.

2.2 Data Sources

Data for the pivotal study was downloaded from the FDA Electronic Data Room as 17 SAS transport files accessed through the following eCTD format document:

\\Cdsesub1\evsprod\N021789\0000\m5\53-clin-stud-rep\533-rep-human-pk-stud\5332-patienpk-init-tol-stud-rep\study-report-0215-r3-c-04-02\data\DEMO.xpt

In response to a request for additional data, the Sponsor later submitted a SAS data set: tube_wt.sas7bdat, with measures of actual usage (in grams).

3. STATISTICAL EVALUATION

Note that this Phase 3 clinical study was undertaken as part of a 505(b)2 submission.

3.1 Evaluation of Efficacy

Study Design and Endpoints

This study was conducted as a multi-center, randomized, investigator-blind, active- and vehicle controlled, parallel comparison involving patients with rosacea meeting specific inclusion/exclusion criteria. A total of 1299 patients were enrolled in the study. The randomization ratio was 3:3:1 (557 and 553 patients were randomized to Metronidazole Gel, 1% and Noritate™ Cream, 1%, respectively, and 189 patients were randomized to receive the gel vehicle). A total of 54 independent study centers enrolled between 1 and 63 patients each. Patients applied the study medication topically to the face once daily at bedtime. The duration of treatment was 10 weeks with evaluations at baseline and at nominal Weeks 2, 4, 7, and 10 (i.e., visits 1-5). Efficacy assessments were based on blinded Investigator assessments of the signs and symptoms of rosacea. The same Investigator was to perform all evaluations for the same

patient at each study visit. Investigators who performed evaluations participated in training and provided documentation of their qualifications to perform efficacy assessments to the Sponsor/CRO for review and approval.

Prior to the start of the study, a randomization list was supplied by the Sponsor. Kits containing five tubes of study drug were prepared for each patient number. Drug supplies for the entire multi-center trial were numbered sequentially. The drug supplies for Metronidazole Gel, 1%, Noritate Cream, 1%, and Metronidazole Gel Vehicle were packaged according to the randomization list in blocks of seven with a ratio of 3:3:1. Complete blocks of drug supplies were distributed to each of the investigational sites to maintain the randomization ratio within an investigational site. A unique drug kit number was associated with each drug supply kit, and this corresponded to the patient number. These numbers were assigned sequentially as patients entered the study at each investigational site. The Sponsor states that the randomization schedule remained blinded from those involved in the clinical conduct of the study, until a data base lock memo was issued.

It was noted during the study that the tubes of Noritate™ Cream, 1% in some of the study kits were due to expire during the enrollment period. The Sponsor indicates that investigators were notified and that such kits were replaced.

Efficacy measures:

The primary endpoints specified in the protocol:

1. Percent change from baseline in inflammatory lesion counts.

This is defined by the Sponsor as % change = (baseline_count - observed_count) / baseline_count.

2. Dichotomized global severity score

Dichotomized	Grade	Score	Clinical Description
Success	Clear	0	No inflammatory lesions, at most mild erythema
	Almost Clear	1	Very mild erythema, very few small papules/pustules
Failure	Mild	2	Mild erythema, several small papules/pustules
	Moderate	3	Moderate erythema, several papules/pustules, up to two nodules
	Severe	4	Severe erythema, numerous papules/pustules, may be several nodules

The week 10 ITT-LOCF values of these endpoints are the primary endpoints. However, further analyses of percent change from baseline measures has clarified that they can be quite sensitive to outliers in the data. The Sponsor's response is to analyze the ranks of the percent

change from baseline measures (although the given reason was skewness in the data). Besides the analysis on ranks, as discussed earlier, an analysis of the Winsorized means is also provided.

Erythema was graded in five different regions of the face, including the forehead, each cheek, nose, and chin. This erythema was measured on the following scale:

Erythema Severity Score

Score	Clinical Description
0	None: no redness present
1	Very mild: Slight pinkness
2	Mild: pink to light red
3	Moderate: definite redness, easily recognized
4	Severe: marked erythema, fiery red

The following were defined as secondary efficacy measures:

1. raw inflammatory lesion counts at baseline and each post-baseline visit,
2. reduction from baseline in inflammatory lesion counts at each post-baseline visit,
3. percent reduction from baseline in inflammatory lesion counts at Weeks 2, 4, and 7
4. Investigator's Global Severity Score at each post-baseline visit,
5. raw combined (across five regions) erythema severity scores at baseline and each post-baseline visit,
6. change from baseline in combined erythema severity scores at each post-baseline visit,
7. worst (across five regions) erythema severity scores at each post-baseline visit.

The erythema scale is an ordinal scale, and such scales have no natural arithmetic zero, and possibly have unequal intervals between scores. Then unit differences between points on the scale are likely not to be commensurable and each nominal unit difference measures a different quantity. Thus, from a rigid measurement theory perspective, change from baseline scales or similar constructs are not appropriate summaries. However, as a rough statistic these operations are often performed and seem to lead to interpretable results, particularly with well chosen ordinal scales.

Summaries of the results for each of the secondary efficacy measurements above are provided in the appendices. The Sponsor's protocol and statistical analysis plan specified that these were to be assessed with descriptive statistics. Appendix 2 includes results for measures 1-3 above. Appendix 3 has summarizes results for measure 4. Appendix 5 summarizes results for measures 5-7.

Safety was evaluated by the Cutaneous Signs and Symptoms scores (dryness, scaling, pruritus, and stinging/burning) at baseline and all post-baseline visits, and by the incidence of adverse events (AEs) reported. Detailed descriptions and summaries for each of these results are given in Appendices 6 and 7.

Patient Disposition, Demographic and Baseline Characteristics

Originally the Sponsor planned to recruit 552 patients in each of the Metronidazole Gel and Noritate Cream groups, and 184 in the Metronidazole Gel vehicle group. The following tables summarize the baseline characteristics and disposition of patients entered in the study:

Table 1. Patient Demographics and Disposition

	Metronidazole Gel	Noritate Cream	Metronidazole Vehicle
N	557	553	189
Gender Male	149 (26.8%)	143 (25.9%)	48 (25.4%)
Female	408 (73.2%)	410 (74.1%)	141 (74.6%)
Race White	484 (86.9%)	489 (88.4%)	164 (86.8%)
Black	6 (1.1%)	8 (1.4%)	1 (0.5%)
Asian	3 (0.5%)	1 (0.2%)	3 (1.6%)
Hispanic	64 (11.5%)	55 (9.9%)	21 (11.1%)
Age 18-64	491 (88.2%)	491 (88.8%)	170 (90.0%)
65+	66 (11.8%)	62 (11.2%)	19 (10.0%)
Completed	500	481	162
Withdrew	57 (10.2%)	72 (13.0%)	27 (14.3%)
Adverse Event	11 (2.0%)	12 (2.2%)	5 (2.6%)
Lack of Efficacy	0	2 (0.4%)	2 (1.1%)
Patient Request	15 (2.7%)	21 (3.8%)	8 (4.2%)
Protocol Violation	9 (1.6%)	9 (1.6%)	2 (1.1%)
Lost to follow-up	18 (3.2%)	26 (4.7%)	10 (5.3%)
Pregnancy	3 (0.5%)	0	0
Other	1 (0.2%)	2 (0.4%)	0

Note that among the patients 73.8% were female, 87.5% were white, and 88.7% were aged 18-64. The most common reason for withdrawal was simple loss to follow-up (4.2% of all patients).

Statistical Methodologies

The analysis specified by the protocol called for tests demonstrating non-inferiority of Metronidazole Gel, 1% relative to Noritate™ Cream, 1% were to be based on a one-sided 97.5% confidence interval (C.I.) approach with a non-inferiority margin of 10% for the percent reduction in inflammatory lesion count and the success rate in the dichotomized Investigator's Global Severity Score. Thus, non-inferiority with regard to lesion counts was to be established if the lower limit of the one-sided 97.5% C.I. for the percent reduction from baseline in inflammatory lesion counts (Metronidazole Gel, 1% minus Noritate™ Cream, 1%) at Week 10 LOCF was greater than -10%. Non-inferiority with regard to success rate for the dichotomized Investigator's Global Severity Score was established if the lower limit of the one-sided 97.5% C.I. for success rate (Metronidazole Gel, 1% minus Noritate™ Cream, 1%) at Week 10 LOCF was greater than -10%. Analyses were conducted on the Week 10 data with last observation carried forward (LOCF) imputation of dropouts in the ITT population and the provided Week 10 EOS PP (end-of-study per protocol) population. Superiority of Metronidazole Gel, 1% over

Metronidazole Gel Vehicle was based on two-sided tests of the percent reduction from baseline in inflammatory lesion counts and the dichotomized Investigator's Global Severity Score at Week 10 LOCF conducted on the ITT population.

The protocol specified that if the data for the percent reduction in inflammatory lesions were skewed, the analysis would be based on rank transformed data. Simulations conducted by this reviewer suggest that for considerably smaller sample sizes than used here, ANOVA is quite robust to skewness, and the rank transform is unneeded. This reviewer would argue that an analysis of data based on the original scale would be preferred, and should take precedence over the rank transformed analyses. However, since the latter are specified in the protocol they are included in this review.

Results and Conclusions:

The following table summarizes the results for the inflammatory lesion counts. The median, mean, and standard deviation of the number of lesions, the change from baseline in the number of lesions (i.e., the absolute reduction), and the percent reduction in the number of lesions are presented for each treatment group. For both the absolute reduction and the percentage change from baseline the p-values for the test of differences between each active treatment group and vehicle are presented using original scale or ranked data, plus confidence intervals on the differences in success rates in the active treatment groups.

Table 2. Week 10 ITT-LOCF Inflammatory Lesion Counts

Treatment / Score, Measure	Metronidazole Gel, 1%	Noritate Cream, 1%	Metronidazole Vehicle
Actual Lesion Count			
Median	5.0	6.0	8.0
Mean (Std Dev)	8.9 (11.3)	9.2 (9.3)	12.8 (14.1)
Absolute Reduction in Lesion Count			
Median	8.0	8.0	7.0
Mean (Std Dev)	9.4 (11.7)	8.9 (10.3)	5.6 (12.0)
p-value Superiority over Vehicle	<0.0001	0.0002	
% Reduction in Lesion Count			
Median	66.7	58.3	46.2
Mean (Std Dev)	50.7 (51.2)	46.4 (56.5)	32.6 (63.0)
p-value Superiority over Vehicle			
% Reduction Original Scale	<0.0001	0.0017	
% Reduction Ranked Data	<0.0001	0.0033	
Non-inferiority Confidence Interval			
% Reduction Original Scale	(-2.18,10.04)		
% Reduction Ranked Data	(0.0,7.69)		

For both the absolute reduction and the percentage change from baseline, the p-values for the test of differences between each active treatment group and vehicle were statistically

significant (all p = 0.0033). The confidence interval for the difference between Metronidazole and Noritate for the percent change from baseline in the ITT-LOCF population was (-2.18, 10.04). In the PP population the corresponding confidence interval was (0.0, 7.69). Results in the Winsorized population were similar. Since all the intervals on the percent change are above -10% we conclude that non-inferiority is established. Note that for the percent reduction in lesion count, the test of non-inferiority using ranks actually is superceded by the test of superiority of Metronidazole over Noritate using the ranked scores (p=0.0375).

Investigator's Global Severity Score

The following table summarizes the results on the Investigator's Global Severity Score, giving the frequencies of each response for each treatment group and the p-values for the tests of differences in the dichotomized success rate between each active treatment group and vehicle, plus a confidence interval on the difference in success rates in the active treatment groups.

Table 3. Week 10 ITT-LOCF Investigator's Global Severity Scores

Treatment / Score, Measure	Metronidazole Gel, 1%	Noritate Cream, 1%	Metronidazole Vehicle
0 = Clear	30 (5.4 %)	30 (5.4 %)	10 (5.3 %)
1 = Almost Clear	184 (33.0 %)	166 (30.0 %)	42 (22.2 %)
2 = Mild	174 (31.2 %)	173 (31.3 %)	53 (28.0 %)
3 = Moderate	159 (28.5 %)	178 (32.2 %)	77 (40.7 %)
4 = Severe	10 (1.8 %)	6 (1.1 %)	7 (3.7 %)
Total	557	553	189
Dichotomized Success:	214 (38.4 %)	196 (35.4 %)	52 (27.5 %)
P-value superiority test	0.0078	0.0491	
Non-inferiority Conf. Interval	(-2.8%,7.9%)		

Note that in the Week 10 ITT-LOCF population the test of differences in success rates between vehicle (27.5%) and each of Metronidazole (38.4%) and Noritate (35.4%) were statistically significant (p = 0.0069 and p = 0.0352, respectively). The confidence interval for the difference between Metronidazole and Noritate in success percentages in the ITT-LOCF population was (-2.8%, 7.9%). In the PP population the corresponding confidence interval was (-5.0%, 6.8%). Since each of these is above the -10.0% bound, in each circumstance we can say that non-inferiority was established.

Sponsor's Analysis

The following table summarizes the Sponsor's results on the Primary ITT-LOCF endpoints. Note that results on the percent change from baseline are based on ranked data. The actual means and medians on the percent change and the actual success proportions agree with the results provided above.

For the success proportions the Sponsor used ridit scores in the CMH test and a simple unweighted comparison of proportions, with a continuity correction, for the non-inferiority confidence intervals. This reviewer would prefer to use the equivalent test for both the superiority and non-inferiority comparison. Since ridit scores seem harder to justify with a non-inferiority comparison, in the FDA reviewer's analysis explicit tables scores were used in the superiority comparisons in success proportions. The non-inferiority computations of success proportions in the FDA reviewer's analyses are also stratified on analysis center. This seems to be the source of the slight differences in the FDA analyses and the Sponsor's analyses of the success proportions. But note that conclusions about efficacy are consistent.

Table 4. Week 10 ITT-LOCF Sponsor's Analyses on Primary Endpoints

Treatment / Score, Measure	Metronidazole Gel, 1%	Noritate Cream, 1%	Metronidazole Vehicle
% Reduction in Lesion Count			
Median	66.7	58.3	46.2
Mean	50.7	46.4	32.6
p-value Superiority over Vehicle % Reduction Ranked Data	<0.0001	0.0028	
Non-inferiority Confidence Interval % Reduction Ranked Data	(0.0, upper[1])		
Dichotomized Success on Severity Score:	214 (38.4 %)	196 (35.4 %)	52 (27.5 %)
P-value superiority test over vehicle	0.0060	0.0331	
Non-inferiority Conf. Interval	(-2.9%, upper[1])		

[1] - Since the interest is primarily in non-inferiority, the Sponsor did not present the upper bounds.

The Statistical analysis plan specified that Visit 2 is targeted at week 2, with days ranging from 2 to 21. Visits 3, 4, and 5 are targeted at weeks 4, 7, and 10, with day ranges from 22-39, 40-60, and 61-84 respectively. The Sponsor's original protocol seemed to specify a much smaller window in days for each visit, but the day ranges in the Statistical Analysis Plan seemed to be more appropriate. Finally, the Sponsor's SAS data sets included a VISIT variable that differed slightly from both the protocol and the statistical analysis plan. These differences had virtually no effect on results from the baseline and the ITT-LOCF analyses, but did have some effect on the reported significance levels at the Week 10 and the Per Protocol Week 10 results.

However, none of these discrepancies affect final conclusions.

3.2 Evaluation of Safety

Appendix 6 displays descriptions of the profiles over time of the local irritation measures: dryness, scaling, pruritis, and stinging/burning. The Investigator was to assess these local irritation measures by direct evaluation (dryness and scaling) or by querying the patient (pruritis and stinging/burning). For all four endpoints the local irritation measures were better in both the

Metronidazole gel and the Noritate cream than in the gel vehicle, though the differences were not always statistically significant.

Using the classification into the general types of adverse events provided by the Sponsor, Appendix 7 displays by treatment group, the number of adverse events reported in that type or class, the number of subjects experiencing an adverse event in that class, and the percentage of the safety population experiencing that class of event. Note that overall, event rates seem to be fairly consistent across treatment groups.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

This study was not powered to detect differences in subgroups. Therefore, no statistical tests for treatment differences within subgroups are provided. However, profiles over time in the primary endpoints may be helpful in describing results. Tables 5-10 below display these profiles. To limit the number of entries only baseline, Visit 3 (nominal week 4), Week 10 EOS, and at the end of study (EOS-LOCF) are shown. At Baseline lesion counts are actual counts. At later time points, both the reduction (i.e., change from baseline), and the percent change from baseline are provided. "P50" denotes the median score, and is a robust estimate of centrality. "Std" denotes the standard deviation. Metronidazole treatment is denoted by "Met", while Noritate and Vehicle are denoted by "Nor" and "Veh", respectively.

Table 5. Change from Baseline in Lesion Counts by Gender

		Visit											
		Baseline			Week 4			Week 10			EOS LOCF		
		Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh
Male													
Reduction	N	149	143	48	143	132	43	140	121	39	149	143	48
	P50	15.0	15.0	16.0	6.0	7.0	5.0	9.0	8.0	8.0	9.0	8.0	6.0
	Mean	19.0	18.9	18.7	8.6	8.2	3.6	11.2	11.6	6.4	11.0	11.1	5.3
	Std	11.3	10.6	9.3	9.7	9.4	8.9	13.3	11.0	10.0	13.4	11.0	11.1
% Change	from Baseline												
	P50	.	.	.	45.5	49.4	28.6	73.5	70.0	44.4	72.2	63.6	35.7
	Mean	.	.	.	40.3	42.3	19.3	56.3	57.2	34.7	52.7	50.1	25.6
	Std	.	.	.	37.4	41.0	53.3	45.5	43.4	54.1	47.9	46.1	56.6
Female													
Reduction	N	408	410	141	383	391	130	366	364	121	408	410	141
	P50	15.0	15.0	15.0	6.0	6.0	5.0	9.0	9.0	8.0	9.0	9.0	7.0
	Mean	18.1	17.8	18.3	6.9	5.8	3.9	9.9	9.5	7.4	9.5	9.3	6.5
	Std	9.9	9.8	10.2	8.7	8.7	8.8	10.6	9.3	10.8	11.2	10.7	12.8
% Change	from baseline												
	P50	.	.	.	46.2	40.0	34.3	70.0	62.5	63.6	65.7	58.0	50.0
	Mean	.	.	.	38.8	32.1	24.3	54.9	51.7	43.7	50.0	45.1	34.9
	Std	.	.	.	40.7	45.1	47.8	45.5	45.2	52.1	52.4	59.7	65.1

Note that in Table 5, for both genders, the trends in mean and median scores for Metronodazole and Noritate over vehicle seem to be very similar within each gender, however there is weak evidence that mean differences from baseline are slightly higher in males than in

females. Further mean scores in the two active groups seem to be somewhat lower in females than in males. Again, these are only descriptions. Since the study was not planned to detect differences in subgroups, no real conclusion is possible.

Table 6. Success on Investigator's Global Severity Score by Gender

		Visit								
		Week 4			Week 10			EOS LOCF		
		Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh
Male	N	143	132	43	140	121	39	149	143	48
	Success %	14.0	12.1	9.3	43.6	43.8	23.1	40.3	37.8	18.8
Females	N	383	391	130	366	363	120	408	409	140
	Success %	13.6	11.3	7.7	40.4	38.3	35.0	37.7	34.7	30.7

Note that in Table 6, at the end of the study the differential effect of Metronidazole and Noritate over vehicle appears to be higher in males than in females, primarily due to a lower success rate in the vehicle group among males.

Patients were dichotomized into two "race" groups, Caucasian versus non-Caucasian (i.e., "Other"). Results by these race groups of the primary endpoints are presented in Tables 7 and 8 below.

Table 7. Change from Baseline in Lesion Counts by Race

		Visit											
		Baseline			Week 4			Week 10			EOS LOCF		
		Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh
Caucasian													
Reduction	N	484	489	164	458	467	150	440	431	139	484	489	164
	P50	15.0	15.0	15.0	6.0	6.0	5.0	9.0	9.0	8.0	9.0	9.0	7.0
	Mean	18.2	17.9	17.5	7.4	6.7	3.7	10.2	10.3	6.8	9.8	10.1	6.0
	Std	10.1	9.8	9.0	9.1	8.7	8.4	11.1	9.7	9.6	11.5	10.8	11.9
% Change from baseline													
	P50	.	.	.	47.2	44.4	33.3	71.4	66.7	60.0	66.7	62.5	47.1
	Mean	.	.	.	40.0	36.4	25.0	55.4	54.8	43.4	51.1	48.0	34.5
	Std	.	.	.	40.1	43.8	48.0	44.5	44.7	51.1	50.7	57.9	62.9
Other													
Reduction	N	73	64	25	68	56	23	66	54	21	73	64	25
	P50	15.0	18.0	21.0	6.0	4.5	7.0	11.0	7.0	8.0	9.0	7.0	4.0
	Mean	19.1	20.0	24.4	6.9	3.8	4.7	10.7	7.7	9.2	10.2	7.5	7.5
	Std	11.0	11.4	13.8	8.7	10.0	11.5	13.2	10.6	15.8	13.9	10.2	15.1
% Change from baseline													
	P50	.	.	.	41.5	25.0	18.9	68.3	48.9	41.2	65.2	35.4	33.3
	Mean	.	.	.	33.9	20.8	10.3	54.1	39.3	28.6	48.2	33.9	20.1
	Std	.	.	.	37.7	45.9	55.3	51.9	43.5	61.0	55.2	42.2	63.7

Most subjects were Caucasian (please see Table 1). Both treatments seem to be superior to vehicle in both race groups. Note that Noritate (and vehicle) seem to be less effective in the "Other" race group than in the Caucasian group, but, otherwise there seems to be no particular trends.

Table 8. Success on Investigator's Global Severity Score by Race

		Visit								
		Week 4			Week 10			EOS LOCF		
		Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh
Caucasian	N	458	467	150	440	430	138	484	488	163
Success	%	14.0	11.6	9.3	40.0	40.5	32.6	37.4	36.5	28.2
Other	N	68	56	23	66	54	21	73	64	25
Success	%	11.8	10.7	0.0	50.0	33.3	28.6	45.2	28.1	24.0

Again, Noritate seems to be less effective in the non-Caucasian "Other" group than in the Caucasian group. It may be worthwhile to reiterate that these are only descriptions, not conclusions.

Tables 9 and 10 below present results on the primary endpoints for each of the age ranges, in years, 18-45, 46-64, and 65+.

Table 9. Change from Baseline in Lesion Counts by Age Group

					Visit								
Age Group		Baseline			Week 4			Week 10			EOS LOCF		
		Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh
18-45													
Reduction	N	250	232	81	229	208	72	217	195	65	250	232	81
	P50	15.0	17.0	16.0	6.0	6.0	5.0	9.0	9.0	7.0	9.0	9.0	7.0
	Mean	19.6	19.9	20.9	7.9	6.0	2.6	10.5	9.7	7.4	10.0	9.9	7.2
	Std	11.0	10.5	11.2	9.8	9.4	10.0	13.1	9.8	12.2	13.2	11.8	12.4
% Change from baseline													
	P50	.	.	.	44.4	38.0	23.4	68.2	59.3	46.7	63.1	51.2	41.2
	Mean	.	.	.	36.2	29.5	15.1	52.3	46.9	36.9	45.2	37.9	29.7
	Std	.	.	.	39.2	44.6	52.8	47.1	49.4	53.6	56.6	67.8	55.1
46-64													
Reduction	N	241	259	89	233	254	82	224	232	77	241	259	89
	P50	15.0	13.0	13.0	6.0	6.0	5.5	9.0	8.5	7.0	9.0	8.0	7.0
	Mean	17.6	16.9	16.9	6.9	6.4	4.8	9.9	10.3	6.8	9.6	9.7	5.8
	Std	9.5	9.7	8.9	8.8	8.5	8.1	10.3	9.9	9.8	11.0	10.1	10.1
% Change from baseline													
	P50	.	.	.	47.1	43.8	38.0	70.7	69.1	55.6	66.7	63.6	44.4
	Mean	.	.	.	40.1	36.2	27.3	55.7	57.0	40.8	53.1	51.7	34.4
	Std	.	.	.	40.4	44.9	46.7	46.1	40.8	51.4	48.1	46.7	52.2
65+													
Reduction	N	66	62	19	64	61	19	65	58	18	66	62	19
	P50	11.0	13.0	13.0	7.0	6.0	5.0	8.0	9.0	8.0	8.0	8.5	8.0
	Mean	16.1	16.5	15.0	7.2	7.5	4.4	10.5	9.6	7.7	10.6	9.7	3.6
	Std	9.5	7.8	6.4	7.0	8.7	6.3	9.0	9.5	8.4	9.0	9.5	19.8
% Change from baseline													
	P50	.	.	.	53.8	50.0	39.1	77.8	70.4	78.9	77.5	68.3	77.8
	Mean	.	.	.	46.5	46.1	35.0	63.7	58.2	61.0	62.8	55.5	35.8
	Std	.	.	.	39.0	38.4	41.5	36.3	41.8	51.8	36.9	41.8	120.8

In Table 9.0 above it seems that among the oldest group of patients there appears to be a difference between means and medians. Using medians the patients aged 65+ seem to show no difference among the three treatment groups, but do show some difference in means between the actives and vehicle. This is consistent with the larger variances in the vehicle groups among older patients. Otherwise there seems to be a general consistency in medians across age groups,

coupled a trend in increasing means. Again, this is just a descriptive observation, not a conclusion.

Table 10. Success on Investigator's Global Severity Score by Age Group

Age Group		Visit: Week 4			Week 10			EOS LOCF		
		Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh
18-45	N	229	208	72	217	194	64	250	231	80
	Success %	11.8	7.2	6.9	37.8	33.0	25.0	33.6	27.7	21.3
46-64	N	233	254	82	224	232	77	241	259	89
	Success %	14.2	13.8	7.3	41.1	43.5	29.9	39.4	40.5	25.8
65+	N	64	61	19	65	58	18	66	62	19
	Success %	18.8	16.4	15.8	53.8	46.6	66.7	53.0	43.5	63.2

Descriptively, there seems to be a trend for increasing success rate over increasing age, with the highest success rate in vehicle group in patients aged above 65! However, this latter group is relatively small, and among younger patients the success rate is higher in the Metronidazole and Noritate groups than in the gel vehicle groups.

4.2 Other Special/Subgroup Populations

As part of the analysis of the effect of dose, results in the Week 10 ITT-LOCF population were broken down by whether the usage totaled less than 50 grams / week or more than 50 grams / week. These results are summarized in Appendix 4.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

1. The protocol specified that the primary analysis for the inflammatory lesions should be based on the percent reduction in inflammatory lesion counts. The protocol further specified that if the percent reduction was skewed, the analysis would be based on rank transformed data. Simulations conducted by this reviewer suggest that for the sample sizes used here, ANOVA is quite robust to skewness, and the rank transform is unneeded. For interpretability, an analysis of data based on the original scale is preferable. Further while there can only be a 100% reduction in lesion count, the increase (i.e., a negative decrease) is unbounded. In practice, this asymmetry seems to lead to outliers, particularly when one of more of the treatments is of low efficacy (e.g. a true placebo). The Division of Dermatological and Dental Drug Products has noted that the actual change from baseline seems less sensitive to outliers than the percent change and hence, may be preferable for analysis. An alternative analysis was to Winsorize the percent reduction to -100.0% to reduce the effect of the extreme outliers, which were all in the negative direction. Thus, analyses of the simple change from baseline, the percent change from baseline, the rank transform of the percent change, and a Winsorized percent change were all conducted here.

2. The primary analyses specified in the protocol were based on the percent change from baseline in the Week 10 inflammatory lesion count and the dichotomized investigator global

evaluation, including non-inferiority testing in the ITT-LOCF and the per protocol (PP) populations and superiority testing using the ITT-LOCF population. For both the percent change from baseline in inflammatory lesion counts and the dichotomized Investigator Global Severity Score (IGSS) the protocol and SAP specified a -10% lower boundary for non-inferiority. The percent change, the ranked per cent change, and the IGSS all demonstrated non-inferiority using this boundary. Statistically significant superiority of the Metronidazole group over vehicle was demonstrated for the percent change, the Winsorized per cent change, the ranked per change, and the absolute change for both the ITT-LOCF, Week 10 completers, and the PP populations. Most of these analyses are presented in Appendices 2 and 3.

3. The Medical team expressed interest in assessing the effect of actual usage of the drug product. The approach used here models the response at the end of the study using the amount of drug product used. Nonlinearity of treatment was assessed by quadratic and cubic terms, along with corresponding homogeneity over treatment. These results for the primary endpoint in the ITT-LOCF population are presented in Appendix 4. Note that there is strong evidence of the effect of dose. However, there is no strong evidence of any heterogeneity over treatment. That is, there is no evidence of a difference in effect of the Metronidazole and Noritate doses.

4. For the analysis of data from small centers (i.e., an investigator with less than 15 patients in either active arm, or less than five patients in the vehicle arm) patients were to be pooled with the center having the largest number of randomized patients to define "analysis centers." Then the data of the Investigator with the second smallest number of randomized patients was to be combined with that of the Investigator with the second largest number of randomized patients, and so on, for all centers that did not have a minimum number of patients per treatment arm. A table comparing the pooled analysis centers with the original investigator sites is given in Appendix 8.

5. The possible differential effect of centers was evaluated by assessing interaction terms. The Breslow-Day test for homogeneity of the odds ratio was to be used to assess consistency in overall success rate across analysis centers. Significant qualitative interactions were to be investigated to determine if pooling was justified. The protocol specified that the primary ANOVA analysis was not to include treatment by center interaction, unless it was statistically significant. Such interaction was to be tested in a secondary model. Logically however, the parameter space of models with interaction span the parameter space of the main effects only models and thus arguably testing of these models should precede the testing of the main effects only models. Note that the usual nominal significance levels of tests of models without the interaction term are not computed as conditional on lack of a significant interaction test. Largely for this reason, this reviewer would be inclined to keep the interaction terms in the model. However, the protocol indicates that the main analysis does not include the interaction terms, and the interaction analyses are to be supportive. So in the analyses presented here, interaction terms are not included.

5.2 Conclusions and Recommendations

NDA 21-789 for Metronidazole Gel, 1%, for the treatment of rosacea, was submitted on August 30, 2004. The Sponsor provided the results of a single study comparing metronidazole gel, 1 %, to its vehicle and to its active comparator, Noritate cream, 1%, to assess the efficacy and safety of Metronidazole gel. The protocol for the study specified that the primary efficacy measures were to be based on blinded investigator assessments of the signs and symptoms of rosacea, as measured by 1) the percent change from baseline in inflammatory lesion counts and 2) a 5-level Investigator's Global Severity Score. For the actual analysis, the 5-point Investigator's Global Severity Score was dichotomized so that scores of clear or almost clear (a score of 0 or 1) were interpreted as a success on this endpoint, other scores as a failure. The protocol further specified that demonstrations of non-inferiority of Metronidazole Gel, 1% relative to Noritate Cream, 1%, were to be based on a one-sided 97.5% confidence interval (C.I.) approach with a non-inferiority margin of 10% for both the percent reduction in inflammatory lesion count and the success rate in the dichotomized Investigator's Global Severity score.

The Study results showed an overall reduction in lesion counts for all three treatment groups. In the Metronidazole gel group the median percent reduction in inflammatory lesion counts was 26.8% at visit 2 and increased to 66.7% in the visit 5, end of study, intent-to-treat (ITT) last-observation-carried-forward (LOCF) population. The median percent reduction was 26.3% at visit 2 and 58.3% for the visit 5 EOS-ITT-LOCF analysis in the Noritate treatment group. In the gel vehicle group, the median percent reduction was 20.6% at Week 2 and 46.2% for the Week 10 ITT-LOCF group. Whether based on the original data or the ranked data, an ANOVA of the Week 10, end of study, ITT LOCF population comparing Metronidazole gel to vehicle showed statistically significant differences in favor of Metronidazole gel ($p< 0.0001$ for both the percent change from baseline and the ranked percent change). The percent change from baseline is sensitive to outliers. One way to adjust for outliers is to replace values below -100%, with -100% (i.e., Winsorize the data). Both Winsorized and simple change from baseline gave similar results to the percent change or the ranked per cent change described above.

Using the rank transformed analysis described in Appendix 1, the lower limit of the one-sided 97.5% confidence interval (C.I.) for the difference in percent change from baseline between Metronidazole and Noritate in the ITT-LOCF population had a lower limit of 0.0%. Using the data in the original gave results in the ITT-LOCF data (lower bound of 97.5% C.I. of -2.18%) and Per Protocol (PP) data (lower bound 97.5% C.I. of -3.0%). Since all of these are above the agreed to boundary of -10.0%, we would conclude non-inferiority of Metronidazole to Noritate.

Success rates for the dichotomized Investigator's Global Severity Score for the primary Week 10 ITT-LOCF analysis were comparable between Metronidazole Gel, 1% and Noritate™ Cream, 1%, 38.4% and 35.5%, respectively, versus a success rate for vehicle gel group of 27.7%. In this population the Week 10 ITT-LOCF success rates for Metronidazole versus vehicle were statistically significantly better for both the ITT ($p=0.0069$) and the PP populations ($p=0.0505$). The lower limit of the one-sided 97.5% C.I. on the difference between success proportions was

greater than -0.10 for the Week 10 analysis in both the ITT-LOCF (lower 97.5% C.I. of -2.8%) and PP (lower 97.5% C.I. of -5.0%) populations. Thus one would conclude that non-inferiority was established in both populations.

APPENDICES

Appendix 1: Nonparametric Confidence Intervals

The Sponsor cites a non-parametric confidence interval method for assessing non-inferiority. First, one adds a value "delta" to the percent reduction from baseline in inflammatory lesion count for the Metronidazole Gel, 1% treatment group. Next, the data from the two active treatment groups are pooled, ranked, and then analyzed by a two-way analysis of variance, with factors of treatment and analysis center. The added delta value is adjusted by iteration so that the p-value of the treatment effect in ANOVA is equal to the desired level of significance, $\alpha/2$.

That the procedure works without the rank transform is apparent. For example, suppose we want to compare two groups with respective independent samples $X_{1,1}, \ldots, X_{1,n1}$ and $X_{2,1}, \ldots, X_{2,n2}$. The usual comparison is based on the difference in means (ΣX_{1i} /n1 - ΣX_{2i} /n2). Since we assume samples are independent, variances within a sample are invariant to the addition of a single constant to each observation. So, for any d, Var (Σ (X_{1i} - d) / n1 - Σ X_{2i} / n2) = Var (Σ X_{1i} / n1 - Σ X_{2i} / n2), where the last expression is labeled s^2. Note s is the standard error of the difference, adjusted for d or not. If we choose d so that (Σ (X_{1i} -d) / n1 - Σ X_{2i} / n2) /s = $Z_{\alpha/2}$, as suggested by the procedure above, then d = Σ (X_{1i} / n1 - Σ X_{2i} / n2) - $Z_{\alpha/2}$ s. That is, d is the lower bound of a 100(1-a)% confidence interval. Note that the added delta value was equal in magnitude to the lower confidence bound, but of the opposite sign (i.e., a positive delta value with a negative lower confidence bound).

The Sponsor's claim is that the analogous procedure works with the ranked data. That is, denoting the rank transform by R(.), one solves for d , such that (Σ R(X_{1i} -d)/n1 - Σ R(X_{2i})/n2) /s = $Z_{\alpha/2}$. One problem is that unlike the case for the observed data above, the relation between the ranked values and the observed values is not one to one. That is, unless there is some j such that X_{1i} -d = X_{1j}, then there exists a neighborhood around X_{1i} -d where the ranks are constant. That is, there is no simple inverse from the significance level to the d bound. That means that the coverage probabilities vary.

Despite the lack of normality and the heterogeneity in variances for ranks, the proposed procedure may be adequate. In fact, preliminary simulations by this reviewer seem to suggest that rank transformed ANOVA's preserve Type I error for highly skewed data (though actually no better than with the untransformed data). However, there have been references in the statistical literature that have indicated that rank transformed analyses can be anti-conservative[1]. This suggests that the proposed nonparametric procedure requires some justification, not included in the submission.

[1] For example: Akritas, M.J. (1990) ,The Rank Transform Method in Some Two-Factor Designs, *Journal of the American Statistical Association*, 85, 73-78, and Hendrick, T.C. & Routou, 0.(2001), An Investigation of the Rank Transform in Multiple Regression, *Computational Statistics and Data Analysis*, 38, 203-215.

Appendix 2: Inflammatory Lesion Counts

The following table displays summary statistics by treatment for the actual lesion counts, the reduction from baseline, and the percent reduction from baseline. At Week 0, baseline, only the actual counts are displayed. The summary statistics are the number of cases, N, the median score, the mean score, and the standard deviation (denoted "Std"). The rows labeled lesions provide summary statistics for the unadjusted lesion counts. Summaries in the ITT population are presented first (Table A.2.1), followed by similar results in the PP population (Table A.2.2).

Table A.2.1. Inflammatory Lesion Counts (ITT Population)

		Treatment								
		Metronidazole Gel, 1%			Noritate Cream, 1%			Vehicle		
Week		N	Median	Mean(Std)	N	Median	Mean(Std)	N	Median	Mean(Std)
0	lesions	557	15.0	18.3(10.3)	553	15.0	18.1(10.0)	189	15.0	18.4(10.0)
2	lesions	535	11.0	13.9(10.1)	515	11.0	13.8 (9.9)	178	13.0	16.0(13.3)
	reduct	535	4.0	4.3 (8.2)	515	4.0	4.1 (8.1)	178	3.0	2.4 (9.9)
	% chng	535	26.8	21.6(42.7)	515	26.3	21.4(43.3)	178	20.6	14.1(52.4)
4	lesions	526	9.0	11.2 (9.7)	523	9.0	11.4 (9.9)	173	11.0	14.4(12.7)
	reduct	526	6.0	7.4 (9.0)	523	6.0	6.4 (8.9)	173	5.0	3.9 (8.8)
	% chng	526	46.2	39.2(39.8)	523	43.5	34.7(44.3)	173	33.3	23.1(49.1)
7	lesions	508	6.0	9.1 (9.7)	498	7.0	9.5(10.1)	162	9.0	12.3(11.1)
	reduct	508	8.0	9.4(10.2)	498	8.0	8.3(10.0)	162	7.0	5.9 (9.2)
	% chng	508	60.0	49.7(46.3)	498	58.3	45.0(54.2)	162	44.9	33.6(48.3)
10	lesions	506	5.0	8.2(10.6)	485	6.0	7.9 (8.6)	160	7.0	11.0(12.2)
	reduct	506	9.0	10.2(11.4)	485	9.0	10.0 (9.8)	160	8.0	7.1(10.6)
	% chng	506	71.1	55.3(45.4)	485	63.6	53.1(44.8)	160	56.5	41.5(52.5)
10 LOCF	lesions	557	5.0	8.9(11.3)	553	6.0	9.2 (9.3)	189	8.0	12.8(14.1)
	reduct	557	8.0	9.4(11.7)	553	8.0	8.9(10.3)	189	7.0	5.6(12.0)
	% chng	557	66.7	50.7(51.2)	553	58.3	46.4(56.5)	189	46.2	32.6(63.0)

The Week 10 LOCF values above are the primary values of interest. At week 10 LOCF the median reduction in the number of lesions was 8 for both metronidazole gel and Noritate and 7 for vehicle. Due to extreme outliers in the vehicle group, i.e., patients who got worse, mean differences are larger, particularly with the difference in percent change from baseline.

Recall that for the analysis small centers were pooled with larger centers. For the ITT population the interactions of treatment and grouped investigator were investigated and found to be statistically non-significant (all $p = 0.3446$ for the percent change and $p = 0.3259$ for the ranked percent change). However, the percent change from baseline measures were highly skewed. For such skewed data, the Sponsor proposed an analysis of the rank transformed values. Note that simulations by this reviewer on both exponential data and on data defined by the absolute values of normal data cubed, both with mean structure similar to the ones observed here, all showed excellent preservation of Type I error under the null hypothesis of no treatment

differences. Since one would prefer to analyze data in the original scale, the need for rank transformation seems debatable. However, since the protocol did call for this procedure, and since it is not clearly inappropriate, both the simple percent change from baseline and the corresponding ranked percent change in responses were analyzed using the following linear model:

Response = pooled_center + treatment

The model for the absolute reduction included an interaction term.

Significance levels of tests of Superiority: Metronidazole

Versus	Endpoint	Week	p-value
Vehicle	% change	10	0.0003
		10 ITT-LOCF	<0.0001
	Rank % change	10	0.0005
		10 ITT-LOCF	<0.0001
	Absolute Reduction	10	0.0006
		10 ITT-LOCF	<0.0001

Significance levels of tests of Superiority: Noritate

Versus	Endpoint	Week	p-value
Vehicle	% change	10	0.0017
		10 ITT-LOCF	0.0017
	Rank % change	10	0.0133
		10 ITT LOCF	0.0033
	Absolute Reduction	10	0.0012
		10 ITT-LOCF	0.0002

Thus for all three endpoints, both Metronidazole Gel and Noritate were shown to be statistically significantly better than vehicle.

Significance levels of tests of Superiority: Metronidazole (deleting vehicle as per protocol/SAP)

Versus	Endpoint	Week	p-value
Noritate	% change	10 LOCF	0.2078
	Rank % change	10 LOCF	0.0375
	Absolute Reduction	10 LOCF	0.5231

Note that the confidence interval "test" of the non-inferiority in ranked percent change from baseline at Week 10 in the ITT-LOCF population is actually superceded by the test of superiority of Metronidazole over Noritate (p=0.0375).

To test non-inferiority of the per cent change from baseline, using the least squares means, we compute the intervals about the estimates. For the percent change from baseline these

are based on least squares means, with vehicle treatment deleted from the model. The rank transformed values utilize the algorithm described in Appendix 1.

Bounds for confidence intervals assessing non-inferiority: Metronidazole

Versus	Endpoint	Week	Interval
Noritate	% Change	Week 10	(-3.42, 7.16)
		10 ITT-LOCF	(-2.18, 10.04)
	Rank % change	Week 10	(-0.37, 6.25)
		10 ITT-LOCF	(0.0, 7.69)

Since these intervals are above -10% we would conclude non-inferiority of Metronidazole to Noritate. For the absolute reduction no particular non-inferiority bounds were specified, and this comparison was not made (but see Appendix 10 for one approach).

Corresponding Results on the Per Protocol population are given below:

Table A.2.2. Inflammatory Lesion Counts (Per Protocol Population)

		Treatment: Metronidazole Gel, 1%			Noritate Cream, 1%			Vehicle		
Week		N	Median	Mean(Std)	N	Median	Mean(Std)	N	Median	Mean(Std)
0	lesions	480	15.0	18.3(10.0)	479	15.0	18.1(10.0)	158	15.0	18.2 (9.9)
2	lesions	473	11.0	13.9 (9.9)	465	11.0	13.7 (9.9)	155	13.0	15.7(13.3)
	reduct	473	4.0	4.6 (7.8)	465	4.0	4.2 (8.3)	155	4.0	2.5 (8.2)
	% chng	473	27.3	22.5(38.9)	465	26.7	21.8(43.8)	155	22.2	15.8(41.1)
4	lesions	475	9.0	10.9 (9.0)	480	9.0	11.6(10.1)	153	10.0	14.1(12.2)
	reduct	475	6.0	7.5 (8.8)	480	6.0	6.5 (9.1)	153	5.0	4.1 (8.7)
	% chng	475	46.2	39.4(39.3)	480	43.8	34.9(45.1)	153	35.3	24.4(48.7)
7	lesions	465	6.0	8.9 (9.1)	465	6.0	9.5(10.3)	150	9.0	12.3(11.3)
	reduct	465	8.0	9.5(10.0)	465	8.0	8.5(10.1)	150	7.0	5.8 (9.2)
	% chng	465	60.0	50.1(46.0)	465	60.0	46.2(54.3)	150	45.5	33.9(48.7)
10	lesions	481	5.0	8.0 (9.7)	473	6.0	8.0 (8.0)	153	7.0	11.2(12.4)
	reduct	481	9.0	10.3(11.1)	473	9.0	10.1 (9.9)	153	8.0	7.0(10.6)
	% chng	481	71.4	55.4(45.2)	473	65.0	53.1(45.1)	153	55.6	41.0(53.0)

Results of tests in the per protocol population are given below:

Significance levels of tests of Superiority: Metronidazole

Versus	Endpoint	Week	p-value
Vehicle	% change	10 PP	0.0002
	Rank % change	10 PP	0.0004
	Absolute Reduction	10 PP	0.0004

Significance levels of tests of Superiority: Noritate

Versus	Endpoint	Week	p-value
Vehicle	% change	10 PP	0.0010
	Rank % change	10 PP	0.0018
	Absolute Reduction	10 PP	0.0002

So again, for all endpoints, both Metronidazole Gel and Noritate were shown to be statistically significantly better than vehicle.

Bounds for confidence intervals assessing non-inferiority: Metronidazole

Versus	Endpoint	Week	Interval
Noritate	% Change	10 PP	(-3.00, 5.31)
	Rank % Change	10 PP	(-0.76, 6.21)

Since both intervals are well above the -10% lower bound, again we would conclude non-inferiority of Metronidazole to Noritate.

Appendix 3: Investigator Global Severity Score

Observed Scores:

The following displays the overall profiles of the cases with data and the Week 10 ITT-LOCF over time:

Table A.3.1. Investigator's Global Severity Score over time

		Base			Week 2			Week 4			Week 7		
		Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh
0=Clear	n	.	.	.	2	3	.	2	2	.	10	7	1
	%	.	.	.	0.4	0.6	.	0.4	0.4	.	2.0	1.4	0.6
1=Almost	n	.	.	.	22	25	4	70	59	14	116	108	25
Clear	%	.	.	.	4.1	4.9	2.3	13.7	11.7	8.4	23.0	22.3	15.4
2=Mild	n	1	.	.	161	146	45	199	192	58	209	199	59
	%	0.2	.	.	30.3	28.3	26.3	38.9	38.0	34.7	41.5	41.1	36.4
3=Moderate	n	556	553	189	342	337	120	237	248	92	163	163	74
	%	99.8	100.0	100.0	64.4	65.4	70.2	46.3	49.1	55.1	32.3	33.7	45.7
4=Severe	n	.	.	.	4	4	2	4	4	3	6	7	3
	%	.	.	.	0.8	0.8	1.2	0.8	0.8	1.8	1.2	1.4	1.9
All		557	553	189	531	515	171	512	505	167	504	484	162

		Week 10			10 LOCF		
		Met	Nor	Veh	Met	Nor	Veh
0=Clear	n	30	30	10	30	30	10
	%	5.6	5.8	5.7	5.4	5.4	5.3
1=Almost	n	184	165	42	184	166	42
Clear	%	34.6	31.7	24.0	33.0	30.1	22.3
2=Mild	n	169	170	52	174	172	52
	%	31.8	32.7	29.7	31.2	31.2	27.7
3=Moderate	n	140	149	65	160	178	77
	%	26.3	28.7	37.1	28.7	32.2	41.0
4=Severe	n	9	6	6	9	6	7
	%	1.7	1.2	3.4	1.6	1.1	3.7
All		532	520	175	557	552	188

Note the overall similarity in response profiles for Metronidazole Gel and Noritate Cream over time, as well as the level of superiority of both over the gel vehicle. Using the consecutive integer scores, the CMH tests of superiority of Metronidazole and Noritate over vehicle were statistically significant ($p = 0.0005$ and $p = 0.0042$, respectively) in the ITT-LOCF tables.

Dichotomized Success Rates

For the primary analysis, the Investigator's Global Severity Score is dichotomized so that clear or almost clear define "success." A response of mild or worse is defined as a "failure." Note that for superiority tests the ITT-LOCF population is used. As a sensitivity analysis of the success rate in the dichotomized Investigator's Global Severity Score in the ITT population the Sponsor proposed a modification of LOCF where any subject with no valid observation at week 10 was scored as a failure. This is labeled as LOCF2 below.

Table A.3.1. Success Rate in Dichotomized Investigator's Global Severity Score (ITT-LOCF Population)

	Treatment								
	Metronidazole Gel			Noritate Cream			Vehicle		
Nominal	success			success			success		
week	N	n	%	N	n	%	N	n	%
0	557	0	0.0	553	0	0.0	189	0	0.0
2	535	26	4.9	515	28	5.4	178	4	2.2
4	526	72	13.7	523	60	11.5	173	14	8.1
7	508	126	24.8	484	118	23.7	162	26	16.0
10	506	209	41.3	520	192	39.7	159	51	32.1
10 LOCF	557	214	38.4	552	196	35.5	189	52	27.5
10 LOCF2	557	185	33.2	553	177	32.0	190	45	23.7

The following display the results of tests of superiority over vehicle for both active treatments in the ITT-LOCF population.

Significance levels of tests of Superiority: Metronidazole

Versus	Endpoint	Week	p-value
Vehicle	Success	10 ITT LOCF	0.0069
		10	0.0526
		10 LOCF 2	0.0220

Significance levels of tests of Superiority: Noritate

Versus	Endpoint	Week	p-value
Vehicle	Success	10 ITT LOCF	0.0352
		10	0.1062
		10 LOCF 2	0.0895

Note that Metronidazole is better than vehicle in all three populations, although the differences in the two LOCF populations are statistically significant. Although the effects of Noritate are similar, but only the comparison in the LOCF population is statistically significant.

One approach to assess non-inferiority of Metronidazole gel to Noritate cream using the success proportions on the dichotomized Investigator Global Severity scores is to ignore the stratification variable (analysis center) and compute a 95% simple binomial confidence interval about the test proportion - active control proportion. This is denoted "Simple C.I." below. An alternative to the simple difference in proportions is to use the weighted combination used in the calculation of the Mantel-Haenszel statistic. Note that this is the same contrast used in the computation of SAS Type 2 sums of squares and adjusts for unequal numbers of observations per stratum. This is denoted "Weighted C.I." below.

Bounds for confidence intervals assessing non-inferiority: Metronidazole

Versus	Endpoint	Week	Simple C.I.	Weighted C.I.
Noritate	Success	10 ITT LOCF	(-2.7%, 8.6%)	(-2.8%, 7.9%)
		10 ITT LOCF2	(-4.3%, 6.7%)	(-5.5%, 6.6%)

At week 10 in the ITT population using LOCF, both confidence intervals are above -0.10, consistent with the non-inferiority of Metronidazole gel to Noritate.

Table A.3.2. Success Rate in Dichotomized Investigator's Global Severity Score (Per Protocol Population)

	Treatment								
	Metronidazole Gel			Noritate Cream			Vehicle		
		success			success			success	
week	N	n	%	N	n	%	N	n	%
0	480	0	0.0	479	0	0.0	158	0	0.0
2	473	22	4.7	465	26	5.6	155	3	1.9
4	475	61	12.8	480	58	12.1	153	14	9.2
7	465	113	24.3	465	113	24.3	150	24	16.0
10	481	199	41.4	473	189	40.0	153	49	32.0

The following display the results of CMH tests comparing Metronidazole and Noritate to vehicle.

Significance levels of tests of Superiority: Metronidazole

Versus	Endpoint	Week	p-value
Vehicle	Success	10 PP	0.0505

Significance levels of tests of Superiority: Noritate

Versus	Endpoint	Week	p-value
Vehicle	Success	10 PP	0.0870

Thus, as assessed by the dichotomized Investigator's Global Severity Score both Metronidazole and Noritate are statistically significantly better than vehicle.

Bounds for confidence intervals assessing non-inferiority: Metronidazole

Versus	Endpoint	Week	Simple C.I.	Weighted C.I.
Noritate	Success	10	(-4.8, 7.6)	(-5.0, 6.8)

At week 10 in the PP population the test of differences between Metronidazole and vehicle was not quite statistically significant (p = 0.0505). However, since the lower bound of the interval about the success differences -5.0 is greater than -10.0, we would accept the hypothesis that Metronidazole is noninferior to Noritate.

Appendix 4: Analysis of Usage/Dosage

At baseline each patient received two tubes of study medication. These were to be returned at the Week 4 visit, and a further three tubes were to be dispensed. These tubes were supposed to be returned at the end of the study. The following table summarized actual usage. Note that total usage is defined as the difference in grams between total weights of the dispensed tubes minus the total weights of the returned tubes. Daily usage is defined as the ratio of total usage to study duration in days (i.e., number of days between first and last use, inclusive). It is worthwhile to note that all analyses in this Appendix post hoc, not even implied by the protocol or statistical analysis plan. Following the Sponsor, the few subjects with zero differences are deleted. Unlike the Sponsor, the one subject with a negative difference is also deleted. .

Table A.4.1. Total Medication Usage/Daily Medication Usage

	Treatment Metronidazole	Noritate Cream	Vehicle
Total Usage (g)			
N	543	531	181
Mean (Std Dev)	53.20 (41.52)	39.20 (28.64)	49.86 (42.87)
Median	39.48	30.82	35.76
Min, Max	1.05, 207.90	1.08, 140.01	0.46, 223.94
Daily Usage (g/day)			
Mean (Std Dev)	0.82 (0.75)	0.59 (0.44)	0.84 (1.65)
Median	0.59	0.45	0.53
Min, Max	0.06, 8.73	0.08, 3.44	0.03, 21.55

The maximum values in the vehicle group are due to a single subject with only data at baseline, but whose reported usage still totals 21.55 grams. This is almost certainly a data error, but does reflect the data provided to the FDA, and is left uncorrected. Note that the vehicle group is ignored in the modeling below.

When analyzing dose one can work with daily usage or with total usage. For assessing levels of actual exposure daily usage would seem to be the more useful. However, when analyzing overall effect in the ITT-LOCF population which includes responses measured at different lengths on study, the total usage would be the more useful measure. Both are reported below.

In a previous pharmacokinetic study, the Sponsor defined usage of Metronidazole at 1 gram/day as "maximal use." The following tables, table A.4.2 and A.4.3, display the numbers of subjects and the corresponding percentages within each treatment group at various dose levels, both for total usage, and for daily usage x 70, where 70 days s the planned duration of treatment. This measure is used to make the subgroups in the daily usage table (Table A.4.3) more comparable to the subgroups defined in the total usage below (Table A.4.2).

Table A.4.2. Counts and Percentages of the Subjects Classified by Total Medication Use.

	Treatment					
	Metronidazole		Noritate Cream		Vehicle	
Total Usage (g)	n	%	n	%	n	%
50 =	331	61.0%	392	73.8%	120	66.3%
> 50	212	39.0%	139	26.2%	61	33.7%
> 70	140	25.8%	78	14.7%	40	22.1%
> 90	98	18.0%	48	9.0%	26	14.4%
>110	63	11.6%	17	3.2%	20	11.0%
>130	40	7.4%	5	0.9%	13	7.2%
>140	30	5.5%	1	0.2%	11	6.1%

Table A.4.3. Percentages of the Subjects Classified by Daily Medication Usage.

		Treatment:					
	Actual	Metronidazole		Noritate Cream		Vehicle	
Daily Usage X 70	Daily (g) Usage	n	%	n	%	n	%
0-50 g	0-0.71 g	324	59.7%	390	73.4%	117	64.6%
> 50 g	>0.71 g	219	40.3%	141	26.6%	64	35.4%
> 70 g	>1.00 g	151	27.8%	86	16.2%	42	23.2%
> 90 g	>1.29 g	100	18.4%	48	9.0%	25	13.8%
>110 g	>1.57 g	65	12.0%	17	3.2%	20	11.1%
>130 g	>1.86 g	44	8.1%	8	1.5%	14	7.7%
>140 g	>2.00 g	35	6.4%	4	0.8%	12	6.6%

At the November 6, 2002, meeting the Sponsor noted that the weekly dosage of the test material was not expected to exceed 5 grams/week. Of the Metronidazole patients above, 219 (40%) had usage greater than 5 grams/week, 151 (28%) had usage greater than 1 gram/day, 35 (6%) had usage greater than 2 grams/day, and 5 (0.9 %) had usage greater than 3 grams/day.

The table below summarizes the results for the percent change from baseline in inflammatory lesion counts in the ITT-LOCF population:

Table A.4.4. Summary of Week 10 ITT-LOCF Change From Baseline Measures in Inflammatory Lesion Measures Broken Down by Daily (x70) Medication Usage

Daily	Metronidazole			Noritate Cream			Vehicle		
Usage x70	Median	Mean	(Std)	Median	Mean	(Std)	Median	Mean	(Std)
Absolute Reduction from Baseline									
50=	9	9.3	10.7	8.0	8.8	10.3	8.0	7.3	11.1
>50	8	10.1	13.1	9.0	10.5	10.5	6.0	3.3	13.8
>70	8	10.2	11.5	9.0	10.3	11.1	6.0	4.0	10.2
>90	10	11.8	10.8	9.5	10.9	10.4	7.0	4.8	10.3
>110	9	11.5	10.8	13.0	13.9	12.9	7.5	7.0	9.5
>130	9	12.2	11.4	4.5	12.1	17.2	7.0	7.3	6.6
>140	11	13.1	11.8	0.0	2.0	4.0	7.5	8.1	6.8

Table A.4.4. (cont.) Summary of Week 10 ITT-LOCF Change From Baseline Measures in Inflammatory Lesion Measures Broken Down by Daily (x70) Medication Usage

Daily	Metronidazole			Noritate Cream			Vehicle		
Usage x70	Median	Mean	(Std)	Median	Mean	(Std)	Median	Mean	(Std)
% Change from Baseline									
50=	69.6	52.9	48.1	60.0	45.6	60.8	55.6	40.0	55.0
>50	66.7	50.5	55.8	69.6	54.6	44.6	41.0	23.6	77.5
>70	66.7	52.0	45.4	70.8	53.0	51.2	36.7	25.9	48.3
>90	75.0	58.1	40.9	70.3	57.6	37.4	45.5	32.9	44.8
>110	66.7	55.6	43.8	71.0	58.2	37.9	49.4	42.3	37.3
>130	75.4	57.8	39.4	29.5	38.4	41.0	47.2	40.9	31.8
>140	80.0	61.4	40.8	0.0	12.5	25.0	56.3	44.0	32.1

The table above was requested by the Medical Team. As noted earlier for actual analysis of the ITT-LOCF population the total usage, as used in Table A.4.5 is arguably the better measure of usage.

Table A.4.5. Summary of Week 10 ITT-LOCF Change From Baseline Measures in Inflammatory Lesion Measures Broken Down by Total Medication Usage

Total	Metronidazole			Noritate Cream			Vehicle		
Usage	Median	Mean	(Std)	Median	Mean	(Std)	Median	Mean	(Std)
Absolute Reduction from Baseline									
50=0	8.0	8.6	10.9	8.0	8.7	10.2	7.0	5.9	13.3
>50	9.0	11.1	12.7	9.0	10.9	10.7	7.0	5.8	10.0
>70	9.0	11.2	11.6	9.5	10.9	10.6	6.0	4.7	9.6
>90	10.0	12.3	10.8	11.0	12.6	10.8	7.0	5.1	10.2
>110	10.0	12.4	10.6	13.0	15.9	11.3	7.5	7.1	9.4
>130	11.5	14.4	11.4	17.0	21.0	17.5	7.0	7.8	6.5
>140	13.0	15.1	11.6	49.0	49.0	.	7.0	8.1	7.1
% Change from Baseline									
50=0	66.7	49.0	54.9	58.0	45.2	59.4	48.8	32.4	69.9
>50	72.8	56.3	44.8	70.0	55.9	48.5	48.6	37.2	50.4
>70	71.1	55.2	43.9	75.0	59.1	48.5	40.7	29.9	46.2
>90	76.6	60.6	40.2	75.7	64.6	33.9	47.2	34.8	44.4
>110	76.2	60.7	41.9	81.3	73.7	26.5	49.4	42.7	36.9
>130	78.7	65.4	34.9	71.0	59.3	35.5	48.9	44.1	30.8
>140	85.4	70.5	36.1	100.0	100.0	.	63.6	45.5	33.3

As might be expected, except for the vehicle group, there seems to be a general trend of increasing treatment effect over dose. The daily use, high usage Noritate groups (usage > 130 g) do not follow this trend, although the total usage, high dose Noritate groups do. The baseline mean number of lesions in the Noritate group was 31.25 in the >130 daily group and 34.25 in the >140 group. This compares to 20.6 and 19.3 in the corresponding Metronidazole and vehicle treatments, respectively, in the >130 groups, and 21.6 and 20.8 in the >140 groups. The range of lesion counts in the >140 group is 16-54 with a mean reduction of only 2 lesions, versus a mean reduction of 13.1 in the Metronidazole group. So the lack of a general reduction in the high daily

dose Noritate group seems to be due to a small number of recalcitrant subjects who use a large amounts of the medication over a relatively short period in the study.

The following table displays the overall number of subjects ("N"), the number of successes ("n"), and the percentage of successes ("%") on the Investigator Global Severity Score using the severity score for various levels of usage.

Table A.4.6. Summary of Week 10 ITT-LOCF Success on Investigator Global Severity Broken Down by Daily Usage (x 70) / Total Usage

Daily Usage(g)	Metronidazole			Noritate Cream			Vehicle		
x 70	N	n	%	N	n	%	N	n	%
=50 g	324	124	38.3	387	137	35.4	114	34	29.8
> 50	219	90	41.1	141	59	41.8	64	18	28.1
> 70	151	67	44.4	86	36	41.9	42	9	21.4
> 90	100	51	51.0	48	19	39.6	25	8	32.0
>110	65	29	44.6	17	7	41.2	20	7	35.0
>130	44	19	43.2	8	2	25.0	14	5	35.7
>140	35	17	48.6	4	0	0.0	12	4	33.3

Total Usage(g)	Metronidazole			Noritate Cream			Vehicle		
	N	n	%	N	n	%	N	n	%
= 50 g	324	124	38.3	389	137	35.2	114	34	29.8
>50	219	90	41.1	141	59	41.8	66	18	27.3
>70	152	67	44.1	86	36	41.9	42	9	21.4
>90	101	51	50.5	48	19	39.6	25	8	32.0
>110	66	29	43.9	17	7	41.2	20	7	35.0
>130	45	19	42.2	8	2	25.0	14	5	35.7
>140	36	17	47.2	4	0	0.0	12	4	33.3

Again, roughly up to 100 grams, the Noritate success rates generally improve, and then more or less stabilize. The pattern for Noritate reflects the recalcitrant subjects.

Using the linear model for the percent change from baseline described in Appendix 2, among those Metronidazole patients whose total usage was less than or equal to 50 grams, but with no restrictions on the Noritate or Metronidazole groups, Metronidazole was statistically significantly better than gel vehicle ($p = 0.0037$). Ignoring the vehicle (because of possibly heterogeneous variances) a 95% confidence interval about the difference in means between Metronidazole, with total usage less than or equal to 50 grams, and Noritate, with no restrictions on usage was (-7.11, 7.88). Since this is above the -10% bound, this shows non-inferiority of the Metronidazole usage restricted group to the Noritate unrestricted usage group.

Despite the tables displayed above, the analyses presented here compare the results of the restricted Metronidazole groups to the unrestricted Noritate and vehicle groups (i.e., the Noritate and vehicle groups above are pooled). The results for Success on the Investigator Global Severity Score are less clear than the results on lesion counts presented above. Although the success rate in the usage restricted Metronidazole group is somewhat higher than in the unrestricted vehicle group, 35.8% versus 28.9%, this difference is not statistically significant

(CMH p= 0.1696). However, the comparison to Noritate is consistent with non-inferiority. That is, the weighted 95% confidence interval about the difference in success percentages was (-0.8, 9.5), uniformly above the -10.0% bound. (However, note again that the appropriateness of this -10.0% bound has been questioned, though a much smaller bound would have still been consistent with non-inferiority).

The Medical team indicated that an analysis of the per protocol population using a total of 50 grams would be of interest. The following table displays the per cent change from baseline and the simple change from baseline for the Per Protocol population restricted to those patients with total usage of at most 50 grams/week and those with more than 50 grams/week:

Table A.4.7. Summary of Week 10 Per Protocol Measures with Daily Usage (x 70)

	Metronidazole			Noritate Cream			Vehicle		
	Median	Mean	(Std)	Median	Mean	(Std)	Median	Mean	(Std)
= 50 g Inflammatory Lesions:									
% Change	71.4	55.0	46.2	62.5	47.5	62.1	60.0	42.4	53.3
Reduction	9.0	9.8	10.4	8.0	9.2	10.5	8.0	7.6	10.9
>50 g									
% Change	69.6	52.3	56.2	70.7	58.5	43.1	44.4	22.4	82.3
Reduction	9.0	10.7	12.6	9.5	11.3	10.6	6.0	3.2	14.5

Success on Investigators Global Evaluation:									
	N	n	%	N	n	%	N	n	%
=50 g	285	113	39.6%	351	132	37.6%	103	32	31.1%
>50	192	83	43.2%	128	57	44.5%	55	17	30.9%

For percent change from baseline in this restricted population, Metronidazole gel was statistically significantly better than vehicle, but while Noritate was better than vehicle gel, with the usual level, the difference was not quite statistically significant ($p = 0.0242$ and $p = 0.0595$, respectively). For the absolute reduction, results for Metronidazole and Noritate were less statistically significant ($p = 0.1062$ and $p = 0.0881$, respectively). The 95% confidence interval about the difference in percent change between Metronidazole and Noritate was (-4.6, 9.04). Since the lower bound was above -10%, one could conclude that non-inferiority was established. Using the success proportion neither Metronidazole nor Noritate was statistically significantly better than vehicle, but Noritate was not ($p = 0.0866$ and $p = 0.1495$, respectively). The 95% confidence interval about the difference between success rates was (-4.6%, 9.5%), also consistent with the hypothesis of non-inferiority. Again, the results of the statistical analyses are not quite consistent with those given above, since for these analyses the restriction in dose applies to all three treatments, not just Metronidazole.

Note there is a statistically significant difference in both total usage and daily usage between Metronidazole and Noritate. One topic of interest would be to see if the dose effect of Metronidazole and Noritate differs. For pooled center(i), i=1,...,32 and treat(j), j=1,2, for the Metronidazole and Noritate treatments respectively we model:

Mean Response = β_0 + pooled_center(i) + β_1 * baseline + β_2 * usage + treat(j) + treat(j)*usage + β_3 * usage * usage + treat(j)*usage*usage + β_4 * usage * usage * usage.

The β_2, β_3, and β_4 coefficients provide linear and nonlinear effects of dose, in this case total usage. The treat(j) term provides an assessment of the homogeneity of the intercept, while the treat(j)*usage and treat(j)*usage*usage terms assess the homogeneity of the linear and quadratic effect of dose.

Using this to model the effects on the percent change from baseline gives the following ANOVA table:

Source	DF	Sequential SSQ	Mean Square	F Value	Pr > F
Analysis center	31	321066.6096	10356.9874	3.99	<.0001
Baseline score	1	6331.8569	6331.8569	2.44	0.1188
Usage	1	68487.7195	68487.7195	26.37	<.0001
Treatment	1	157.8008	157.8008	0.06	0.8054
Usage*treatment	1	6180.6850	6180.6850	2.38	0.1232
Usage*Usage	1	16241.8268	16241.8268	6.25	0.0125
Usage*Usage*treatment	1	2227.6626	2227.6626	0.86	0.3546
Usage*Usage*Usage	1	22474.3765	22474.3765	8.65	0.0033

Note there is strong evidence of the effect of total usage, both linear and non-linear effects. However, there is no strong evidence of any heterogeneity over treatment in the effect of this dose. That is, there is no evidence of a difference in effect of the Metronidazole and Noritate doses. Results for the Winsorized percent change and the absolute reduction were similar.

Modeling the success proportion with a logistic regression model incorporating the mean structure above, gives the following table:

Effect	DF	Chi-Square	Wald Pr > ChiSq
Analysis center	31	111.2391	<.0001
Baseline score	1	10.8746	0.0010
Usage	1	4.1266	0.0422
Treatment	1	0.0737	0.7860
Usage*treatment	1	0.0494	0.8240
Usage*Usage	1	1.2296	0.2675
Usage*Usage*treatment	1	0.0076	0.9307
Usage*Usage*Usage	1	0.2959	0.5865

The Hosmer and Lemeshow Goodness-of-Fit Test was highly non-significant (p = 0.8236), very consistent with fit. Note that for this analysis there is no evidence of nonlinearity in effect or a differential effect of usage. In particular, the dose effect seems to be linear, and again, there is no evidence of a difference in effect between the Metronidazole and Noritate doses.

Appendix 5: Erythema Summary Scores

Erythema scores are based on the following values:

Score	Clinical Description
0	None: no redness present
1	Very mild: Slight pinkness
2	Mild: pink to light red
3	Moderate: definite redness,easily recognized
4	Severe: marked erythema, fiery red

Recall that these are secondary analyses, as discussed in section 3.1 of the report and are only to be displayed descriptively.

Table A.5.1 Mean Profiles in Erythema Scores and Change from Baseline in Erythema Score

	Visit Number											
	Base			Week 2			Week 4			Week 7		
	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh
N	557	553	189	536	515	178	526	523	173	508	498	162
Erythema Score												
Mean	2.4	2.4	2.4	2.1	2.1	2.2	2.0	2.0	2.0	1.8	1.8	1.8
Std Dev	0.5	0.5	0.5	0.6	0.6	0.6	0.7	0.7	0.7	0.7	0.7	0.7
Change in Erythema Score												
Mean	.	.	.	-0.2	-0.3	-0.2	-0.4	-0.4	-0.4	-0.6	-0.6	-0.5
Std Dev	.	.	.	0.5	0.5	0.5	0.6	0.6	0.6	0.7	0.6	0.6

	Visit Number					
	Week 10			Week 10 LOCF		
	Met	Nor	Veh	Met	Nor	Veh
N	506	485	160	557	551	187
Erythema Score						
Mean	1.6	1.6	1.7	1.7	1.7	1.8
Std Dev	0.8	0.8	0.8	0.8	0.8	0.8
Change in Erythema Score						
Mean	-0.8	-0.8	-0.6	-0.7	-0.7	-0.6
Std Dev	0.8	0.7	0.7	0.8	0.7	0.7

Note the similarities over time in both endpoints for the three different treatment groups. All three treatments are associated with a decrease in erythema, however any treatment effect is confounded with regression effects and any secular trend. But there is no evidence of any treatment differences.

The following table displays the worst erythma score over the five facial sites. Note that this endpoint would be expected to and does seem to show much more variation than the mean score in Table A.5.1.

Table A.5.2 Profiles of Worst Erythema Scores

Worst Erythema Score		Visit Number											
		Base			Week 2			Week 4			Week 7		
		Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh
0=None	n	.	.	.	1	.	.	.	1	.	1	2	.
	%	.	.	.	0.2	.	.	.	0.2	.	0.2	0.4	.
1=Very Mild	n	5	2	.	19	17	3	42	39	8	76	65	15
	%	0.9	0.4	.	3.5	3.3	1.7	8.0	7.5	4.6	15.0	13.1	9.3
2=Mild	n	16	20	7	108	106	30	139	153	51	155	187	47
	%	2.9	3.6	3.7	20.1	20.6	16.9	26.4	29.3	29.5	30.5	37.6	29.0
3=Moderate	n	487	483	164	373	368	130	325	311	104	260	226	90
	%	87.4	87.3	86.8	69.6	71.5	73.0	61.8	59.5	60.1	51.2	45.4	55.6
4=Severe	n	49	48	18	35	24	15	20	19	10	16	18	10
	%	8.8	8.7	9.5	6.5	4.7	8.4	3.8	3.6	5.8	3.1	3.6	6.2
All		557	553	189	536	515	178	526	523	173	508	498	162

Worst Erythema Score		Visit Number					
		Week 10			Week 10 LOCF		
		Met	Nor	Veh	Met	Nor	Veh
0=None	n	11	9	2	12	9	2
	%	2.2	1.9	1.3	2.2	1.6	1.1
1=Very Mild	n	108	92	21	109	94	23
	%	21.3	19.0	13.1	19.6	17.1	12.3
2=Mild	n	159	181	56	172	196	59
	%	31.4	37.3	35.0	30.9	35.6	31.6
3=Moderate	n	217	188	74	246	230	92
	%	42.9	38.8	46.3	44.2	41.7	49.2
4=Severe	n	11	15	7	18	22	11
	%	2.2	3.1	4.4	3.2	4.0	5.9
All		506	485	160	557	551	187

Again, the percentages in each response category for the three treatment profiles seem to be similar over time, though far less similar than the mean scores.

Appendix 6: Adverse Events: Cutaneous Signs and Symptoms

The Investigator was to assess these local irritation measures by direct evaluation (dryness and scaling) or by querying the patient (pruritis and stinging/burning).

Dryness

Score	Clinical Description
0	None: No dryness
1	Mild: Slight but definite roughness
2	Moderate: Moderate roughness
3	Severe: Marked roughness

The following table shows the overall responses over time.

Table A.6.1 Results for Dryness over Visit

Dryness		Base			Week 2			Week 4			Week 7		
		Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh
0=None	n	302	301	105	326	344	101	346	370	103	347	379	90
	%	54.2	54.5	55.6	60.9	66.8	56.7	65.8	70.7	59.9	68.4	76.1	55.6
1=Mild	n	191	191	53	164	139	60	152	127	54	145	95	54
	%	34.3	34.6	28.0	30.7	27.0	33.7	28.9	24.3	31.4	28.6	19.1	33.3
2=Moderate	n	59	58	30	42	30	15	28	24	15	14	24	17
	%	10.6	10.5	15.9	7.9	5.8	8.4	5.3	4.6	8.7	2.8	4.8	10.5
3=Severe	n	5	2	1	3	2	2	.	2	.	1	.	1
	%	0.9	0.4	0.5	0.6	0.4	1.1	.	0.4	.	0.2	.	0.6
	All	557	552	189	535	515	178	526	523	172	507	498	162

Dryness		Week 10			10 LOCF		
		Met	Nor	Veh	Met	Nor	Veh
0=None	n	361	369	92	396	409	106
	%	71.3	76.1	57.9	71.1	74.0	56.4
1=Mild	n	118	104	53	129	122	64
	%	23.3	21.4	33.3	23.2	22.1	34.0
2=Moderate	n	27	12	14	30	21	17
	%	5.3	2.5	8.8	5.4	3.8	9.0
3=Severe	n	.	.	.	2	1	1
	%	.	.	.	0.4	0.2	0.5
	All	506	485	159	557	553	188

CMH tests of differences between both Metronidazole Gel and Noritate cream versus gel vehicle, were both statistically significant ($p = 0.0004$ and $p < 0.0001$, respectively).

Scaling

Score	Clinical Description
0	None: No scaling
1	Mild: Barely perceptible shedding, noticeable only on light scratching or rubbing
2	Moderate: Obvious but not profuse scaling
3	Severe: Heavy scale production

Table A.6.2 Results for Scaling over Visit

		Visit											
		Base			Week 2			Week 4			Week 7		
Scaling		Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh
0=None	n	345	342	116	368	368	115	395	393	114	388	401	101
	%	61.9	62.0	61.4	68.8	71.5	64.6	75.1	75.1	65.9	76.5	80.5	62.3
1=Mild	n	162	161	54	127	123	45	99	106	43	104	74	42
	%	29.1	29.2	28.6	23.7	23.9	25.3	18.8	20.3	24.9	20.5	14.9	25.9
2=Moderate	n	49	48	18	39	21	17	31	23	15	14	23	17
	%	8.8	8.7	9.5	7.3	4.1	9.6	5.9	4.4	8.7	2.8	4.6	10.5
3=Severe	n	1	1	1	1	3	1	1	1	1	1	.	2
	%	0.2	0.2	0.5	0.2	0.6	0.6	0.2	0.2	0.6	0.2	.	1.2
	All	557	552	189	535	515	178	526	523	173	507	498	162

		Week 10			10 LOCF		
Scaling		Met	Nor	Veh	Met	Nor	Veh
0=None	n	398	391	112	439	433	130
	%	78.7	80.6	70.4	78.8	78.3	69.1
1=Mild	n	88	81	36	95	97	42
	%	17.4	16.7	22.6	17.1	17.5	22.3
2=Moderate	n	20	13	11	20	23	16
	%	4.0	2.7	6.9	3.6	4.2	8.5
3=Severe	n	.	.	.	3	.	.
	%	.	.	.	0.5	.	.
All	n	506	485	159	557	553	188

CMH tests of differences between both Metronidazole Gel and Noritate cream versus gel vehicle, were both statistically significant (both p = 0.0045).

Pruritis

Score	Clinical Description
0	None: No itching
1	Mild: Slight itching, but not really bothersome
2	Moderate: Definite itching that is somewhat bothersome: without loss of sleep
3	Severe: Intense itching that has caused pronounced discomfort; sleep interrupted and excoriation of the skin from scratching may be present

Table A.6.3 Results for Pruritis over Visit

		Visit											
		Base			Week 2			Week 4			Week 7		
Pruritis		Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh
0=None	n	352	346	122	420	400	140	439	418	136	426	418	125
	%	63.2	62.7	64.6	78.5	77.7	78.7	83.5	79.9	78.6	83.9	83.9	77.2
1=Mild	n	148	133	52	91	79	24	73	78	32	68	65	32
	%	26.6	24.1	27.5	17.0	15.3	13.5	13.9	14.9	18.5	13.4	13.1	19.8
2=Moderate	n	57	70	14	24	30	13	13	25	5	13	14	5
	%	10.2	12.7	7.4	4.5	5.8	7.3	2.5	4.8	2.9	2.6	2.8	3.1
3=Severe	n	.	3	1	.	6	1	1	2	.	1	1	.
	%	.	0.5	0.5	.	1.2	0.6	0.2	0.4	.	0.2	0.2	.
All	n	557	552	189	535	515	178	526	523	173	508	498	162

Table A.6.3 (cont.) Results for Pruritis over Visit

		Visit					
Pruritis		Week 10			10 LOCF		
		Met	Nor	Veh	Met	Nor	Veh
0=None	n	440	421	135	480	467	152
	%	87.0	86.8	84.9	86.2	84.4	80.9
1=Mild	n	56	53	20	65	61	28
	%	11.1	10.9	12.6	11.7	11.0	14.9
2=Moderate	n	7	11	4	10	21	7
	%	1.4	2.3	2.5	1.8	3.8	3.7
3=Severe	n	3	.	.	2	4	1
	%	0.6	.	.	0.4	0.7	0.5
All	n	506	485	159	557	553	188

The CMH test of differences between Metronidazole Gel and gel vehicle was close to the usual statistical significance level (p = 0.0547), while the corresponding test of differences between Noritate cream and gel vehicle was not statistically significant (p = 0.5162).

Stinging/Burning

Score	Clinical Description
0	None: No stinging/burning
1	Mild: Slight warm, tingling sensation; not really bothersome
2	Moderate: Definite warm; tingling sensation that is somewhat bothersome
3	Severe: Hot, tingling sensation that has caused definite discomfort

Table A.6.4 Results for Stinging/Burning over Visit

		Visit											
Stinging/burning		Base			Week 2			Week 4			Week 7		
		Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh
0=None	n	386	380	129	449	432	143	465	461	145	458	448	137
	%	69.3	68.8	68.3	83.9	83.9	80.3	88.4	88.1	83.8	90.2	90.0	84.6
1=Mild	n	109	107	36	66	63	21	51	44	25	43	37	21
	%	19.6	19.4	19.0	12.3	12.2	11.8	9.7	8.4	14.5	8.5	7.4	13.0
2=Moderate	n	51	58	23	16	17	12	7	15	3	5	12	4
	%	9.2	10.5	12.2	3.0	3.3	6.7	1.3	2.9	1.7	1.0	2.4	2.5
3=Severe	n	11	7	1	4	3	2	3	3	.	2	1	.
	%	2.0	1.3	0.5	0.7	0.6	1.1	0.6	0.6	.	0.4	0.2	.
All	n	557	552	189	535	515	178	526	523	173	508	498	162

Stinging/burning		Week 10			10 LOCF		
		Met	Nor	Veh	Met	Nor	Veh
0=None	n	465	448	144	505	499	161
	%	91.9	92.4	90.6	90.7	90.2	85.6
1=Mild	n	37	33	14	43	39	22
	%	7.3	6.8	8.8	7.7	7.1	11.7
2=Moderate	n	2	3	1	6	11	4
	%	0.4	0.6	0.6	1.1	2.0	2.1
3=Severe	n	2	1	.	3	4	1
	%	0.4	0.2	.	0.5	0.7	0.5
All	n	506	485	159	557	553	188

Again, the CMH test of differences between Metronidazole Gel and gel vehicle was reasonably close to the usual statistical significance level (p = 0.0848), while the corresponding test of

differences between Noritate cream and gel vehicle was not statistically significant (p = 0.2689).

Appendix 7: Adverse Events: Overall Summary

Using the classification provided by the Sponsor, for each treatment the following table provides an estimate of the number of adverse events in that type or class reported, the number of subjects experiencing event in that class, and the percentage of the safety population experiencing that class of event.

Table A.7.1 Summary of Different Classes of Adverse Event

	Metronidazole			Noritate			Vehicle		
AE Type	# ae	# subj	%	# ae	# subj	%	# ae	# subj	%
CARDIAC DISORDERS	2	2	0.4%	1	1	0.2%	0	0	0.0%
EAR AND LABYRINTH DISORDERS	2	2	0.4%	3	3	0.5%	2	2	1.1%
EYE DISORDERS	8	8	1.4%	7	7	1.3%	2	2	1.1%
GASTROINTESTINAL DISORDERS	19	14	2.5%	18	14	2.5%	5	3	1.6%
GENERAL DISORDERS AND ADMINISTRATION SITE CONDITIONS	1	1	0.2%	9	8	1.4%	1	1	0.5%
HEPATOBILIARY DISORDERS	0	0	0.0%	1	1	0.2%	1	1	0.5%
IMMUNE SYSTEM DISORDERS	10	8	1.4%	3	3	0.5%	1	1	0.5%
INFECTIONS AND INFESTATIONS	90	76	13.6%	78	71	12.9%	32	28	14.8%
INJURY, POISONING AND PROCEDURAL COMPLICATIONS	18	17	3.1%	27	25	4.5%	0	0	0.0%
INVESTIGATIONS	1	1	0.2%	2	2	0.4%	2	2	1.1%
METABOLISM AND NUTRITION DISORDERS	4	4	0.7%	4	4	0.7%	0	0	0.0%
MUSCULOSKELETAL AND CONNECTIVE TISSUE DISORDERS	22	19	3.4%	9	9	1.6%	6	5	2.6%
NEOPLASMS BENIGN, MALIGNANT AND UNSPECIFIED (INCL CYSTS AND POLYPS)	5	4	0.7%	12	6	1.1%	2	2	1.1%
NERVOUS SYSTEM DISORDERS	23	18	3.2%	29	19	3.4%	7	3	1.6%
PSYCHIATRIC DISORDERS	10	10	1.8%	5	5	0.9%	1	1	0.5%
RENAL AND URINARY DISORDERS	2	2	0.4%	0	0	0.0%	1	1	0.5%
REPRODUCTIVE SYSTEM AND BREAST DISORDERS	2	2	0.4%	6	5	0.9%	1	1	0.5%
RESPIRATORY, THORACIC AND MEDIASTINAL DISORDERS	24	22	3.9%	19	16	2.9%	5	5	2.6%
SKIN AND SUBCUTANEOUS TISSUE DISORDERS	55	36	6.5%	62	35	6.3%	14	12	6.3%
SURGICAL AND MEDICAL PROCEDURES	1	1	0.2%	1	1	0.2%	0	0	0.0%
VASCULAR DISORDERS	10	8	1.4%	1	1	0.2%	1	1	0.5%

Note that overall, event rates seem to be fairly consistent across treatment groups. This was conformed by a formal statistical test. In particular, using the Fisher Exact test to test for differences between the actives and the controls, after adjusting for multiplicity using the bootstrap adjustment (with 20,000 iterations) in SAS® PROC MULTTEST, no differences between Metronidazole and vehicle control were statistically significant.

Appendix 8: Correspondence Between Original Investigator Sites and Analysis Centers

For the analysis of data from small centers (i.e., an investigator/site with less than 15 patients in either active arm, or less than five patients in the vehicle arm) patients were to be pooled with the center having the largest number of randomized patients. Then the data of the investigator with the second smallest number of randomized patients was to be combined with that of the investigator with the second largest number of randomized patients, and so on, for all investigators who did not have a minimum number of patients per treatment arm. The following table displays the correspondence between the original sites (i.e., investigators) and the pooled analysis sites along with the number of subjects at each measurement time point.

Table A.8.1. Correspondence Between Investigators and Pooled Analysis Center, with Number of Subjects per Time Point

		Visit Number																	
		Baseline			2			3			4			5			EOS LOCF		
Analysis Center	Site	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh
1	0001	15	15	5	15	13	5	15	16	4	15	15	4	15	15	4	15	15	5
3	0002	8	7	3	8	6	2	6	7	2	5	8	1	6	7	1	8	7	3
	0003	11	11	4	11	9	4	11	13	3	9	10	3	13	9	3	11	11	4
4	0004	10	11	3	10	9	2	8	10	3	7	10	3	9	9	3	10	11	3
	0064	8	8	3	7	8	3	7	8	3	7	8	2	7	8	4	8	8	3
6	0006	15	13	5	15	11	5	15	13	5	15	12	5	14	12	5	15	13	5
	0041	1	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1
7	0007	14	15	5	13	15	5	13	14	5	14	15	5	14	14	5	14	15	5
	0033	2	1	.	2	1	.	2	1	.	2	1	.	2	1	.	2	1	.
9	0009	12	10	4	12	10	4	12	10	4	12	10	4	12	9	4	12	10	4
	0020	8	7	2	8	7	2	7	7	2	7	8	2	7	6	2	8	7	2
11	0011	11	12	4	10	12	3	11	12	2	11	11	2	10	10	2	11	12	4
	0054	5	6	2	5	6	2	5	6	1	5	6	1	5	6	1	5	6	2
12	0012	16	15	5	16	15	5	16	14	5	16	14	5	15	14	5	16	15	5
13	0013	10	9	3	10	8	2	10	8	3	9	8	2	8	8	1	10	9	3
	0037	9	8	3	9	8	2	9	7	2	8	6	2	8	5	2	9	8	3
15	0015	10	11	4	10	10	4	9	10	3	9	10	3	9	10	3	10	11	4
	0026	7	9	3	7	7	3	7	9	3	6	8	2	6	8	2	7	9	3
16	0016	27	27	9	26	23	9	26	22	9	25	20	9	24	20	9	27	27	9
18	0018	12	10	4	11	8	4	12	8	4	12	8	4	11	8	3	12	10	4
	0031	6	7	2	7	7	2	5	7	2	5	7	2	5	6	1	6	7	2
19	0008	9	9	3	8	9	2	8	9	2	8	9	2	8	9	2	9	9	3
	0019	9	9	3	9	9	3	9	9	3	9	9	3	9	9	3	9	9	3

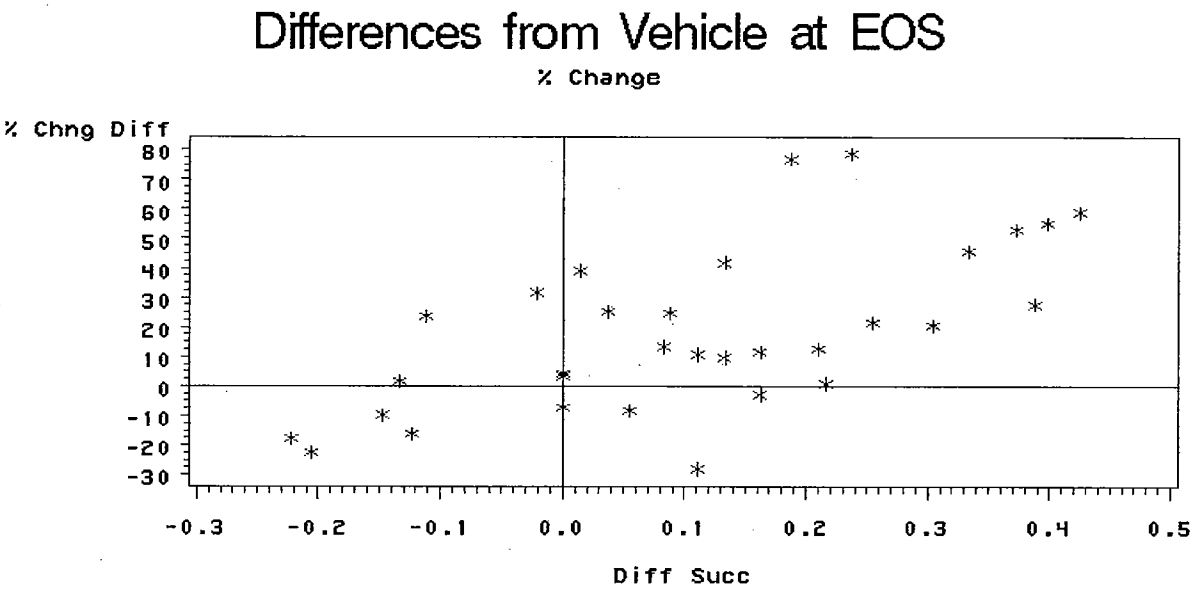
Table A.8.1. (cont.) Correspondence Between Investigators and Pooled Analysis Center, with Number of Subjects per Time Point

		Visit Number																		
		Baseline			2			3			4			5			EOS LOCF		
Analysis Center	Site	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh	Met	Nor	Veh
21	0014	4	6	2	4	6	2	4	5	2	4	5	2	4	5	2	4	6	2
	0021	12	12	4	11	11	4	11	11	3	12	11	3	10	10	3	12	12	4
27	0025	2	1	1	2	1	1	2	1	1	2	1	1	2	1	1	2	1	1
	0027	14	15	4	11	14	3	11	13	3	11	13	3	8	13	3	14	15	4
28	0028	12	11	4	11	10	4	11	10	4	11	10	4	11	10	4	12	11	4
	0046	5	5	2	5	5	2	5	5	2	5	5	2	5	5	1	5	5	2
35	0035	15	15	5	14	12	5	14	12	4	14	13	4	14	12	4	15	15	5
36	0024	5	4	2	5	4	2	5	4	2	5	3	2	5	3	2	5	4	2
	0036	12	14	4	11	13	4	13	14	4	12	12	4	12	13	4	12	14	4
40	0034	8	8	3	8	8	3	8	8	3	7	7	2	7	7	2	8	8	3
	0040	10	11	3	10	11	3	9	10	3	9	11	3	9	11	3	10	11	3
45	0045	20	19	7	20	18	7	19	16	5	19	16	4	19	16	4	20	19	7
47	0047	11	10	4	10	9	3	10	10	4	9	10	4	10	9	4	11	10	4
	0049	8	7	3	8	7	3	8	7	3	8	7	3	7	7	3	8	7	3
50	0050	16	18	5	16	16	5	14	16	4	14	14	4	15	15	4	16	18	5
52	0052	15	16	5	15	15	4	15	16	5	15	16	4	13	15	4	15	16	5
53	0042	5	4	2	4	4	2	4	3	2	4	3	2	4	3	2	5	4	2
	0053	12	12	4	12	12	4	12	12	4	12	11	4	12	11	4	12	12	4
55	0055	15	14	5	13	12	4	15	16	6	14	14	5	16	14	5	15	14	5
	0061	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
56	0044	7	8	3	7	8	3	7	8	3	7	8	3	7	7	3	7	8	3
	0056	12	10	3	12	10	3	11	10	3	11	9	3	10	9	3	12	10	3
58	0032	5	4	2	5	4	2	5	4	2	5	4	2	5	4	2	5	4	2
	0058	12	12	4	12	12	4	12	12	3	12	10	3	12	11	3	12	12	4
59	0059	17	15	5	16	14	4	16	14	4	14	14	4	16	13	4	17	15	5
60	0060	19	19	6	19	19	6	18	18	6	18	18	5	18	18	5	19	19	6
62	0029	9	8	3	9	7	3	7	8	3	6	7	2	7	7	3	9	8	3
	0062	9	10	3	8	10	3	8	9	3	7	8	2	7	7	2	9	10	3
63	0063	15	15	5	14	15	5	14	15	5	12	14	5	12	12	4	15	15	5
65	0030	3	3	1	3	2	1	3	2	1	3	1	1	2	1	1	3	3	1
	0065	13	14	5	10	13	5	11	13	5	9	9	5	9	12	5	13	14	5

Recall that analyses were stratified on or adjusted for analysis centers.

Appendix 9: Consistency of Primary Endpoints

The following displays for each pooled analysis center the difference in mean percent change from baseline between the Metronidazole gel group and the vehicle group in the ITT-LOCF population plotted against the corresponding difference in success proportions on the investigator global severity.



Note that the upper right and lower left quadrant correspond to centers with an inconsistency between the percent change form baseline in lesion counts and success rates. In particular, the upper left quadrant corresponds to cases where the mean percent change from baseline is higher in the Metronidazole group than in the vehicle group, but the proportion of successes is lower. The lower right quadrant corresponds to cases where the mean percent change from baseline is lower in the Metronidazole group than in the vehicle group, but the success proportion is higher. Still, the endpoints seem roughly track each other and tend to be consistent. Note that most centers are in the upper right quadrant, consistent with the superiority of Metronidazole gel over vehicle.

Appendix 10: A Bayesian Look at the Non-inferiority Issues

One approach to assessing non-inferiority of a test drug to a reference drug is to assess how much of the relative effect of the reference drug over the vehicle is preserved by the test drug. With Metronidazole gel as the test drug, Noritate as the reference, and the vehicle gel as placebo, one measure of this relative efficacy can be stated as $(\mu_{Met} - \mu_{Nor}) / (\mu_{Nor} - \mu_{Veh})$. Note that, assuming Notitate is more effective than vehicle, if this ratio is greater than zero, then $\mu_{Met} > \mu_{Nor}$, i.e., Metronidazole gel is actually superior to Noritate. The probability that this is true and the probability of exceeding some small negative bounds would be of interest.

For this analysis we model the expected reduction in inflammatory lesion counts in the visit 5, ITT-LOCF population as a function of the baseline score, a random center effect, and a treatment effect. The treatment effect is defined by dummy variables β_4 and β_5, denoting metronidazole and Noritate respectively. Then, conditional on the covariates, for a subject j in center i:

$\mu_{Met}(i,j) = \beta_1 + \beta_2$*baseline(i,j) + β_3*center(i) + β_4
$\mu_{Nor}(i,j) = \beta_1 + \beta_2$*baseline(i,j) + β_3*center(i) + β_5
$\mu_{Veh}(i,j) = \beta_1 + \beta_2$*baseline(i,j) + β_3*center(i),

where baseline(i,j) is lognormal(μ,τ) and center(i) is distributed Normal(0,1). Because of the randomization the expected baseline scores are identical within each center. Thus, unconditionally (i.e., in a population average context):

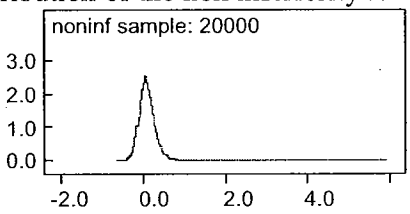
$(\mu_{Met} - \mu_{Nor}) / (\mu_{Nor} - \mu_{Veh}) = (\beta_4 - \beta_5) / \beta_5$

Besides assessing the distribution of this ratio, it of interest to assess the probability that this ratio is greater than 0, greater than -0.1, and greater than -0.2 since these estimate the superiority of Metronidazole gel and the probability of preserving at least 90% and 80%, respectively, of the treatment effect of Noritate.

Note recent recommendations are that the gamma priors traditionally used for variance parameters in such models are far too informative. Hence, uniform priors are placed on the standard deviation parameters in the program below. These result in the following MCMC estimates:

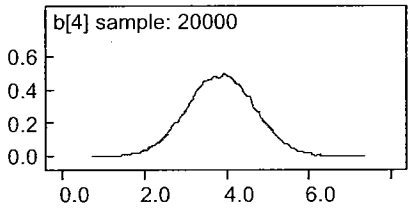
node	mean	sd	MC error	2.5%	median	97.5%	start	sample
b[1]	-308.5	84.26	7.102	-440.1	-288.1	-171.5	5001	20000
b[2]	0.5084	0.02746	5.252E-4	0.4548	0.5083	0.5621	5001	20000
b[3]	211.7	27.47	2.312	156.4	216.5	243.5	5001	20000
b[4]	3.861	0.8046	0.01634	2.285	3.862	5.439	5001	20000
b[5]	3.577	0.8051	0.01692	1.992	3.575	5.134	5001	20000
low[1]	0.6906	0.4622	0.003169	0.0	1.0	1.0	5001	20000
low[2]	0.8791	0.326	0.00249	0.0	1.0	1.0	5001	20000
low[3]	0.9685	0.1747	0.001282	0.0	1.0	1.0	5001	20000
noninf	0.1003	0.1995	0.001743	-0.2148	0.07937	0.5326	5001	20000
sigma	9.547	0.1898	0.001466	9.188	9.543	9.931	5001	20000
sigma.b	154.7	30.58	1.286	88.85	159.3	198.0	5001	20000
sigma.base	0.4889	0.009653	6.194E-5	0.4706	0.4887	0.5083	5001	20000

The distribution of the non-inferiority ratio is as estimated as follows:

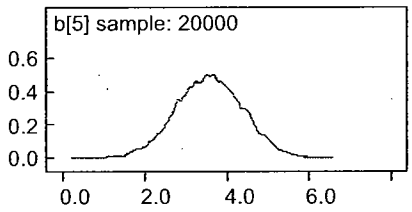


From the summary display of parameter estimates above the estimated probability that the non-inferiority ratio is greater than 0 is 0.6906, while the estimated probability that the ratio is above -10% is 0.8791, and that the ratio is above -20% is 0.9466. That is, we estimate the probability that Metronidazole is actually superior to Noritate as 0.691, the probability that Metronidazole preserves at least 90% of the effect of Noritate over vehicle as 0.879, and the probability it preserves at least 80% of the relative effect as 0.947.

The dummy variable corresponding to Metronidazole provides an estimate of the differential effect of Metronidazole over vehicle:



Similarly the dummy variable corresponding to Noritate is an estimate of the differential effect of Noritate over vehicle:



For both parameters, the lower bounds of the approximate 97.5% credible interval about $ß_4$ and $ß_5$ are 2.285 and 1.992 respectively, both well above zero. In a hypothesis testing framework this would be very consistent with the hypothesis that these parameters are above 0.0, i.e., that both Metronidazole gel and Noritate are better than vehicle.

Note this is only a partial analysis. Due to time constraints there was no assessment of fit or of alternative models. The only convergence diagnostics on the MCMC iterations were Brooks-Gelman-Rubin plots with three different sets of starting values, trace plots, and the autocorrelation function. But all seem quite consistent with convergence of the MCMC iterations. The final conclusion does seem to be consistent with the claim of non-inferiority in change from baseline in lesion counts.

The following WINBUGS 1.4 program was used to derive the estimates above.

```
   model {
   for ( i in 1:N ) {
      reduct[i] ~ dnorm(mu[i],tau)
      base[i]~dlnorm(mu.base,tau.base)
      mu[i] <- b[1] + b[2]*base[i] + b[3]*ctr[inv[i]] + b[4]*(d1[i]) + b[5]*(d2[i]);
      d1[i] <- equals(trt[i],1);
      d2[i] <- equals(trt[i],2);
                    }
   ctr[1] ~ dnorm(0,1)I(0,)
   for (m in 2:nc ) {
       ctr[m]~dnorm(0,1)
                          }
   for (k in 1:5){
         b[k]~dnorm(0.0,tau.b)
                    }
   noninf <- (b[4]-b[5])/ b[5];
   low[1]  <- step(noninf)
   low[2] <- step(noninf + 0.1)
   low[3] <- step(noninf + 0.2)
   tau         <-1/(sigma*sigma)
   tau.base<- 1/(sigma.base*sigma.base)
   tau.b <- 1/(sigma.b*sigma.b)
   mu.base~dunif(0.001,200)
   sigma.base~dunif(0.001,200)
   sigma.b~dunif(0.001,200)
   sigma~dunif(0.001,200)
   }
data
  list(N=1299,nc=32)
  inv[ ]  trt[ ] base[ ] reduct[ ]
    1     2       18       -17
    1     1       16        -5
    1     2       21        19
    1     1       13         8
        - data -
   32     0       33         3
   32     1        9         0
   32     2       20        15
   32     2       12         7
END
```

A similar model was specified for the success proportions at visit 5 in the ITT-LOCF population. That is, as in the model above, we can denote the probability of success for the jth subject in center i with treatment 'XXX' as $p_{XXX}(i,j)$, with expectation p_{XXX}, we can model this probability with a logisitic model as follows:

$\text{logit}(p_{Met}(i,j)) = ß_1 + ß_2*\text{baseline}(i,j) + ß_3*\text{center}(i) + ß_4$
$\text{logit}(p_{Nor}(i,j)) = ß_1 + ß_2*\text{baseline}(i,j) + ß_3*\text{center}(i) + ß_5$
$\text{logit}(p_{Veh}(i,j)) = ß_1 + ß_2*\text{baseline}(i,j) + ß_3*\text{center}(i),$

where, as above, the baseline(i,j) is lognormal(μ,τ) and center(i) is distributed Normal(0,1). Because of the randomization the population of baseline scores is identical across treatment groups, so expectations should taken over all subjects.

Again, one measure of the relative effect of Metronidazole over the reference drug, Noritate, can be expressed as in terms of success proportions as noninf = $(p_{Met} - p_{Nor}) / (p_{Nor} - p_{Veh})$. Using the program below, gives the following MCMC estimates:

Node	mean	sd	MC error	2.5%	median	97.5%	start	sample
b[1]	-0.9384	0.217	0.004151	-1.37	-0.9359	-0.5216	1501	58500
b[2]	-0.02902	0.0067	3.024E-5	-0.04235	-0.02897	-0.01601	1501	58500
b[3]	0.2372	0.6756	0.03255	-0.8802	0.6225	0.9431	1501	58500
b[4]	0.4285	0.193	0.002877	0.05968	0.4253	0.8122	1501	58500
b[5]	0.3016	0.192	0.00286	-0.06535	0.2987	0.6858	1501	58500
low[1]	0.7837	0.4117	0.002408	0.0	1.0	1.0	1501	58500
low[2]	0.8416	0.3651	0.002396	0.0	1.0	1.0	1501	58500
low[3]	0.8832	0.3211	0.002433	0.0	1.0	1.0	1501	58500
noninf	1.219	86.43	0.3546	-3.923	0.3697	6.227	1501	58500
prob0	0.2962	0.03186	4.985E-4	0.2357	0.2959	0.3594	1501	58500
prob1	0.3815	0.01927	8.145E-5	0.3439	0.3813	0.4196	1501	58500
prob2	0.3551	0.01907	7.671E-5	0.318	0.3549	0.3927	1501	58500
sigma.b	0.8265	0.4858	0.004507	0.3387	0.7114	2.012	1501	58500
sigma.base	0.4889	0.00961	4.224E-5	0.4705	0.4888	0.5081	1501	58500

Note that b[1]-b[5] denote $ß_1$ - $ß_5$, respectively. The non-inferiority ratio above is denoted noninf. The estimated probability that the non-inferiority ratio is above 0 is given in nodel low[1], above -10% is low[2], and above -20% is low[3]. Using this non-inferiority ratio, the estimated probability that the improvement of the gel over Noritate is greater than the superiority of Noritate over vehicle is 0.784. That is, the probability that metronodazole preserves at least 100% of the effect of Noritate is 0.784. Similarly, the estimated probability that metronidazole preserves at least 90% of the effect of Noritate over vehicle is 0.842, and probability 0.883 that it preserves at least 80%.

The following WINBUGS 1.4 program was used to derive these estimates:

```
model{
 for ( i in 1:N ) {
   success[i] ~ dbern(p[i]);
   base[i]~dlnorm(mu.base,tau.base);
   mu0[i] <- b[1] + b[2]*(base[i]-mean(base[ ])) + b[3]*ctr[inv[i]];
   mu[i] <- mu0[i] + b[4]*(d1[i]) + b[5]*(d2[i]);
   logit(p[i]) <- mu[i];
   d1[i] <- equals(trt[i],1);
   d2[i] <- equals(trt[i],2);
   p0[i]   <- (exp(mu0[i]))/(1 + exp(mu0[i]));
   p1[i]   <- (exp(mu0[i] + b[4]))/(1 + exp(mu0[i] + b[4]));
   p2[i]   <- (exp(mu0[i] + b[5]))/(1 + exp(mu0[i] + b[5]));
```

```
      ninf[i]  <- (p1[i] - p2[i]) / (p2[i] - p0[i]);
                          }
   ctr[1] ~ dnorm(0,1)I(0,)
   for (m in 2:nc ) {
       ctr[m]~dnorm(0,1)
                    }
   for (k in 1:5){
         b[k]~dnorm(0.0,tau.b)
                 }
   prob0<- mean(p0[ ]);
   prob1<- mean(p1[ ]);
   prob2<- mean(p2[ ]);
   noninf <- mean(ninf[ ]);
   noninf2 <- (prob1 - prob2) / (prob2-prob0);
   low[1] <- step(noninf2);
   low[2] <- step(noninf2 + 0.1);
   low[3] <- step(noninf2 + 0.2);
   tau.base<- 1/(sigma.base*sigma.base);
   tau.b <-   1/(sigma.b*sigma.b);
   mu.base~ dunif(0.001,200);
   sigma.base~dunif(0.001,200);
   sigma.b~dunif(0.001,200)
       }
 inits
    list(b=c(0,0,0,0,0),  sigma.b=1,sigma.base=1,mu.base=10 )
    list(b=c(2,1,0,-1,-2),sigma.b=5,sigma.base=2,mu.base=3 )
    list(b=c(-2,-2,2,2,2),sigma.b=4,sigma.base=4,mu.base=5 )
data
    list(N=1299,nc=32)
 inv[ ]  trt[ ] base[ ] success[ ]
   1    2       18    0
   1    1       16    0
   1    2       21    1
   1    1       13    1
   1    1       11    1
     - data -
  32    0       33    0
  32    1        9    0
  32    2       20    1
  32    2       12    0
END
```

SIGNATURES/DISTRIBUTION LIST

Primary Statistical Reviewer: Steve Thomson
Date: 29 April 2005

Concurring Reviewer:

Statistical Team Leader: Mohamed Alosh, Ph.D., HFD-725

cc:
HFD-540/Joseph Porres, M.D..
HFD-540/Kalyani Bhatt.
HFD-540/Jill Lindstrom, M.D..
HFD-540/Markham Luke, M.D., Ph.D.
HFD-540/Jon Wilkin, M.D.
HFD-725/Steve Thomson
HFD-725/ Mohamed Alosh, Ph.D.
HFD-725/Daphe Lin, Ph.D.
HFD-725/Mohammed Huque, Ph.D.,
HFD-725/Charles. Anello, Ph.D.

c:\Data\Profiles\MyDocuments~\N21789MetronidazoleDow.doc

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

/s/

Steven Thomson
4/29/05 11:06:21 AM
BIOMETRICS

Mohamed Alosh
4/29/05 12:34:56 PM
BIOMETRICS
Concur with review