

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:

20-784

STATISTICAL REVIEW(S)

7 Page(s) Withheld

✓ Trade Secret / Confidential

 Draft Labeling

 Deliberative Process

**STATISTICAL REVIEW AND EVALUATION
CLINICAL STUDIES**

NDA #: 20-784 Date: DEC 17 1997

Applicant: Rhone-Poulenc Rorer Pharmaceuticals, Inc.

Name of Drug: Nasacort HFA Nasal — (triamcinolone acetate) 55mcg Actuation

Indication: Seasonal and perennial allergic rhinitis symptoms in adults and children

Documents Reviewed: 17-Dec-96: Volumes 1.1, 1.21 - 49 [including, 2 diskettes in Vol. 1.21: RG5029T-311 Demographic Data (DEM0311.502) and Daily Diary Data (EFFW311.ZIP)]

Statistical Reviewer: Stephen E. Wilson, Dr. P.H.

Medical Input: Cheung Kwan, M.D. (HFD-570)

Key Words: therapeutic comparability, CFC Switch, two one-sided 90% confidence interval, allergic rhinitis, double-delta

Summary

The sponsor has submitted the results from a clinical program “comprised of therapeutic comparability, long-term safety, and a pharmacokinetic study” in support of proposed approval for a reformulation of its Nasacort Nasal — with a non-CFC propellant, (HFA-134a) for the treatment of seasonal and perennial allergic rhinitis. A statistical review of the sponsor’s data for Study RG 5029T-311, designed in accordance with the Division of Pulmonary Drug Product’s “Points to Consider: Clinical Development Programs for New Nasal Spray Formulations,” demonstrates that this new formulation [Nasacort HFA Nasal — (triamcinolone acetate) 55mcg Actuation] is “comparable” to the currently marketed formulation and, therefore, supports approval for the seasonal/perennial allergic rhinitis claim.

CONTENTS

1. BACKGROUND	3
2. ANALYSIS PLAN: STUDY 311	4
2.1 PRIMARY EFFICACY VARIABLE(S)	4
2.2 SECONDARY EFFICACY VARIABLES / ADDITIONAL DATA	5
2.3 BASELINE	5
2.4 SAMPLE SIZE	5
2.5 ANALYTICAL METHODOLOGY	6
2.5.1 Pairwise comparisons: within formulations	7
2.5.2 Comparability	7
2.5.3 Pairwise comparisons: across formulations	8
2.5.4 Additional analyses	8
3. RESULTS: STUDY 311	9
3.1 PATIENT POPULATION: BASELINE CHARACTERISTICS	9
3.2 STUDY CONDUCT: COMPLIANCE / DROPOUTS / EVALUABILITY	10
3.3 TREATMENT COMPARISONS	11
3.3.1 Data and Program Files	12
3.3.2 Two Week Mean 24-Hour Reflective Symptom Scores	12
3.3.3 Comparability	14
3.3.4 Pairwise Comparisons: Across Formulations – the Double-Delta Approach	16
3.3.5 “Snap-Shot” Symptom Scores	16
3.3.6 Day 1 - 4 Symptom Scores: Reflective and Snap-Shot	19
3.3.7 Other Secondary Endpoint Analyses	22
3.3.8 Exploratory/Subgroup Analyses	22
4. CONCLUSIONS	24

1. Background

Study Design -- Study 311

An important element of the CFC-switch program selected by the sponsor is the demonstration that comparability between the new and reference formulations be demonstrated by “a comparison of the dose-response curves of these two formulations in a single efficacy trial” and that the “doses studied should encompass the highest and the lowest to-be-marketed doses of the new formulation.” Accordingly the sponsor designed and conducted a study entitled “A Randomized, Placebo-Controlled, Double-Blind, Double-Dummy, Parallel Group, Therapeutic Equivalence Comparison of Nasacort Nasal Inhaler and Nasacort 134a Nasal — in Patients with Seasonal (Fall Ragweed) Allergic Rhinitis” – Protocol No. RG-5029T-311 (or Study 311). In the proposed labeling the sponsor recommends that “dosing be started at 220 mcg (2 sprays in each nostril) once a day in patients age 6 years of age and older

Important features of this trial are summarized in Figure 1, below.

Figure 1. Study 311: Study Design

Indication	No. of Centers (location)	Start Date	Duration of Treatment (frequency)	Visits	Total Daily Dose	Patients Enrolled	Endpoint(s)
Seasonal Allergic Rhinitis	16 (USA)	8/30/94	2 weeks (once daily: morning)	4	P-12	780	Reflective and “Snapshot” Nasal Index: Nasal Stuffiness; Nasal Discharge; and Sneezing
				Screening	14 mcg		
				Baseline	110 mcg		
				Week 1	440 mcg		
				Week 2	HFA-134a		
					14 mcg		
					110 mcg		
					440 mcg		

Review Notes:

In this review -- “HFA 134a,” “134a,” and “HFA” all refer to the proposed, to-be-marketed product – Nasacort HFA 134a (triamcinolone acetonide) Nasal — and P-12 and CFC refer to the currently marketed reference product, Nasacort P12 Nasal Inhaler.

2. Analysis Plan: Study 311

2.1 Primary Efficacy Variable(s)

As specified in the protocol, the sponsor described the primary variables for the study to include the mean change from baseline of [ref. NDA 20-784, vol. 1.23, p. 8-3-130]:

1. Nasal Index (sum of nasal discharge, nasal stuffiness, and sneezing scores);
2. Nasal Stuffiness;
3. Nasal Discharge; and
4. Sneezing.

Using diaries, patients assessed their symptoms twice daily, before bedtime (PM scores) and in the morning prior to dosing (AM scores). These were 12-hour “reflective” scores, based on a four point rating scale [0 = symptom absent; 1 = mild (present, but not annoying); 2 = moderate (present and annoying..., but does not interfere with sleep or daily living); and 3 = severe (interferes with / or unable to carry out activities of daily living or sleep)]. Thus, the Nasal Index (NI), a combined score, based on the three nasal symptoms, ranged from 0 - 9. The PM scores represent the first 12 hours post-dosing, while the AM scores are reflective of the next 12 hours [ref. NDA 20-784, vol. 1.23, p. 8-3-161].

Reviewer’s Comment. The sponsor did not select a single primary variable for this study, choosing to describe four variables as “primary.” However, it’s not clear what is meant by “primary” in a study of treatment “comparability.” If a variable is “primary,” does this mean that any results for that variable that fall outside pre-specified boundaries are indicative of a lack of comparability? Given the conditions of the CFC Switch Program, this would be an overly conservative approach. As sample size considerations were based on differences in Nasal Index scores, this reviewer will focus on an evaluation of NI to assess the strength of the statistical / quantitative evidence provided by the sponsor. The nasal scores comprising this index and a number of other variables collected and analyzed by the sponsor will be used to assess support for conclusions based on the analysis of the NI.

This study evolved during its planning stages – three amendments to the protocol document these changes. Based on “recommendations received from the Food and Drug Administration,” the sponsor modified the original plan for measuring 24-hour reflective scores in a protocol amendment dated August 15, 1994. The breakdown into two 12 hour periods for reflective symptom measurement, provides for more refined assessment of symptoms. In addition, in Amendment #3, dated August 19, 1994 the sponsor added an additional “snap-shot” symptom assessment for the one hour period prior to dosing – attempting to capture the end-of dosing-period experience. This amendment only applied to the study activities 10 of the 16 investigators. [ref. NDA 20-784, vol. 1.23, p. 8-3-145-152]. These modifications appear to have been carefully implemented by the sponsor, and the results based on these data, in general, tend to strengthen study conclusions.

2.2 Secondary Efficacy Variables / Additional Data

Mean change from baseline of Nasal Itchiness and Total Eye Symptoms were included in the study as secondary variables [ref. NDA 20-784, vol. 1.23, p. 8-3-130]. In addition, the sponsor collected data describing patient and physician global evaluations (five point scales: 0 = greatly improved; ... 4 = greatly worsened), adverse events, physical examinations, laboratory values, compliance, prior/concomitant therapies, outdoor air exposure, and pollen counts [ref. NDA 20-784, vol. 1.23, p. 8-3-155-172].

2.3 Baseline

As described in the sponsor's "Revised Statistical Analysis Plan" dated 9/1/94, mean baseline values were derived from symptom scores recorded during the 4 days prior, plus the morning of Visit 2 – a total of 9 ratings (five AM and four PM) preceding randomization. [ref. NDA 20-784, vol. 1.24, p. 8-4-5]

Reviewer's Comment: Prospectively, the sponsor defined baseline as including data from four days prior to randomization. However, documenting the data submitted to the Agency, the sponsor describes "Baseline Week" as the "average over Days -3 to -1" [ref. NDA 20-784, vol. 1.21, p. 8-viii]. This is a somewhat troubling discrepancy – it is assumed, for the purposes of this review, that this post-hoc modification did not strongly influence the reported results of the analyses.

2.4 Sample Size

Per protocol, sample size calculations were based on a "two one-sided 90% confidence interval method of establishing equivalence of two means." After examining data from previous studies demonstrating estimates of 0.8 to 1.0 in the detectable differences between effective and sub-optimal doses, the sponsor powered the study to show that "the mean change from baseline in the nasal index does not differ by more than 0.75." [ref. NDA 20-784, vol. 1.23, p. 8-3-129]. Based on a protocol amendment dated 8/15/94, the sponsor intended to "have a 80% probability of showing that the mean change from baseline in the nasal index does not differ by more than 0.75, approximately 98 patients per group is sufficient for analysis." [ref. NDA 20-784, vol. 1.23, p. 8-3-147]

Reviewer's Comment. Though not described in the protocol, the assumption of a mean change from baseline difference of not greater than 0.75, as projected from earlier studies, was an attempt by the sponsor to provide assurance that the treatment efficacy would not vary by more than 30%. Though the planned size of this trial changed with each protocol amendment, as there is "no gold standard" for what is meant by

“comparability” between CFC and HFA products, this study, with more than 100 patients per treatment group was sufficiently large to provide a good “comparison between the dose response curves of the new formulation and the previously approved reference product...by visual inspection.” (A primary requirement for “comparability” -- described in the Division of Pulmonary Drug Product’s guidance entitled “Rationales Behind Various Issues Pertaining to the Nasal Spray Switch Points-to-Consider Document” [p. 4]). Based on the sponsor’s statistical approach, the efficacy of the HFA product could be 30% worse than the reference product and still be considered “therapeutically equivalent.” This 30% difference does not appear to be a generally recognized standard, but it is a reasonable approach to justify sample size for this type of study.

2.5 Analytical Methodology

This study was designed to demonstrate that Nasacort HFA-134a is safe, effective and “therapeutically equivalent” to Nasacort P12 Nasal Inhaler, a CFC propellant. As stated above, “primary endpoints” included mean change from baseline in the nasal index and each of its components (Nasal Stuffiness; Nasal Discharge; and Sneezing). Planned analyses included statistical tests of :

pairwise comparisons *within* each Nasacort formulation (each dose group against all others);

comparability (described by the sponsor as “Therapeutic/Comparability Equivalence”); and

and

pairwise comparisons *across* Nasacort formulations [for each dose studied, described by the sponsor to be a “double delta” approach, i.e., (CFC-placebo) - (HFA-placebo)].

The planned analyses of reflective mean changes included the following comparisons: daily, (each of the first four days -- to examine onset); AM, PM and 24 hour; treatment week; and overall treatment period. In addition, the sponsor analyzed snap-shot means for the first four days of treatment, treatment week and overall treatment period.

Reviewer’s Comment: The plethora of analyses provided by the sponsor facilitates an examination of a large number of aspects regarding the “comparability,” of the tested products. However, with the large number of p-values produced describing these comparisons, the reader should use appropriate caution in interpreting the strength of conclusions based on any specific secondary and post hoc results.

2.5.1 Pairwise comparisons: within formulations

As an analysis considered by the sponsor to be of “secondary importance,” the “Revised Statistical Analysis Plan” describes two-way ANOVA tests of mean 24-hour symptom change from baseline scores for each P12 and HFA dose against a placebo (two-sided, $\alpha = 0.05$, center and treatment effects, no interaction). The sponsor prospectively proposed to combine data from the P12 and HFA placebo treatment groups after testing for a difference (nasal index mean change from baseline, two sample t-test, two-sided, $\alpha = 0.05$). As the sponsor planned 6 contrasts, the analysis plan called for a Bonferroni adjustment for these multiple comparisons [ref. NDA 20-784, vol. 1.24, p. 8-4-147].

Reviewer's Comment: This analysis, testing the efficacy of each treatment group against placebo is an important element in establishing the validity of the study. The Bonferroni adjusted t-test described by the sponsor is the most conservative approach available to adjust for multiple comparisons. The proposal to combine placebo treatment groups (P12 and HFA) relies on the assumption that the vehicles are comparable – even though the sponsor chose to test this difference at $\alpha = 0.10$, given the study's planned sample sizes this does not serve as a rigorous assessment of a potential treatment difference in the vehicles.

2.5.2 Comparability

The “Revised Statistical Analysis Plan” for the study, dated 9/1/94, describes the following “primary comparisons” for the determination of “Therapeutic/Comparability Equivalence”:

P12, 110 mcg vs. HFA, 110 mcg; and
P12, 440 mcg vs. HFA, 440 mcg.

As “secondary comparisons” this plan called for an additional assessment of the 14 mcg doses of the CFC and HFA products. The sponsor stated that the “two one-sided test procedure will be used to evaluate equivalence, comparable.” [ref. NDA 20-784, vol. 1.24, p. 8-4-7]. With the design of this study the sponsor prospectively stated that goal was to demonstrate that the “difference of the treatment groups are within 30% of each other.” [ref. NDA 20-784, vol. 1.24, p. 8-4-7]

Reviewer's Comment: As the guidance for the CFC Switch Program only requires a “visual inspection” of dose response curve “comparability,” the sponsor is not required to establish a statistical test to assess “therapeutic equivalence.” It appears, based on these “primary comparisons” described, that the sponsor planned to emphasize

comparisons of P12 and HFA at specific doses, not an overview of similarity in dose response curves. This is a reasonable methodology for these two “comparability” comparisons (though the cited reference to the work by Huque, et al. (“Establishing therapeutic equivalence with clinical endpoints,” American Statistical Association 1990. Proceedings of the Biopharmaceutical Section, 91-97) outlines an approach, but does not describe the multiple comparison testing conditions of this particular application. In providing guidance for the testing of “therapeutic equivalence,” it appears that the Agency is generally satisfied to see results for just one prospective comparison). Though the selected 30% difference is not a well established, “acceptable” limit for a “comparability” claim, it is likely, given planned sample sizes, that for the confidence intervals of the treatment effects to stay within this range, the observed effect sizes would necessarily need to be of similar magnitudes.

2.5.3 Pairwise comparisons: across formulations

For comparisons across formulations the sponsor chose to test two-way ANOVA contrasts on the NI “for every dose level (110 mcg, 440): Nasacort P12 minus Placebo) minus (Nasacort 132a minus Placebo)” [ref. NDA 20-784, vol. 1.24, p. 8-4-7]. In the reported results these tests are referred to as “double delta contrasts.”

Reviewer's Comment: Though it could be argued that this analytical approach would, perhaps, serve as an assessment of differences in the shapes of the dose response curves of the two formulations, the sponsor did not provide for a direct evaluation of a given dose of one formulation against a dose the other formulation (e.g., 110mcg, P12 vs. HFA).

2.5.4 Additional analyses

2.5.4.1 Combining placebo groups

The sponsor planned to combine the two placebo groups in the study if a two sample t-test of the observed difference in change from baseline of the 24-hour nasal index score over the two-week double blind period failed to achieve significance at the “conservative” level of 10% [ref. NDA 20-784, vol. 1.24, p. 8-4-7].

Reviewer's Comment: While this procedure of combining data from the two placebo groups may have enhanced the power of subsequent statistical testing, it depended heavily on the assumption that there was no difference in the efficacy of the two vehicles (P12 and HFA). Given the size of the placebo groups, it was not likely that this study would provide a sensitive test of any potential (real) difference between the two placebo groups.

2.5.4.2 Interaction.

To justify pooling the results from the different centers, the sponsor planned to test treatment by center interaction “using a conservative level of 15%.” [ref. NDA 20-784, vol. 1.24, p. 8-4-8]

3. Results: Study 311

3.1 Patient Population: Baseline Characteristics

As shown in Table 1, below, Study 311 included 780 subjects randomized to eight treatment groups (two placebo groups were combined in the sponsor’s analysis). Approximately 55% of these patients were male, about 84% were Caucasians and the mean age was 36 years (ranging from 18 to 83). The baseline nasal index scores for the treatment groups ranged from 6.4 to 7.1)

Table 1. Study 311: Study Population / Baseline Characteristics.

Treatment	N	Sex		Race		Baseline Nasal Index
		%M	%F	%Cauc.	%Other	
Total	780	55	45	84	16	N/A.
P12 Nasal Inhaler						
14mcg	113	56	44	86	14	7.0
110mcg	115	54	46	89	11	6.7
440mcg	108	56	44	81	19	6.7
HFA-134a Nasal Inhaler						
14mcg	113	59	41	77	23	7.1
110mcg	107	54	46	89	11	6.4
440mcg	113	53	47	83	17	6.8
All Placebo	111	54	46	84	16	6.7

Sources: NDA 20-784, Vol. 1.23, p. 8-3-39 and p.8-3-69

Reviewer's Comment: In general, baseline characteristics and scores appear to be well-balanced across treatment groups. The sponsor states that the “reflective rhinitis symptoms were similar across treatments for all variables,” but did not provide the

results of statistical tests of these scores/characteristics.

However, in examining baseline NI scores, the HFA-134a 110mcg treatment group stands out as the group with the lowest score. This group also had the lowest baseline mean “snap-shot” score. These low scores are a potential indication that, on-average, patients in this treatment group were not as sick as those in the other groups – potentially confounding results for treatment comparisons described below.

3.2 Study Conduct: Compliance / Dropouts / Evaluability

The sponsor reported that 48 (6.2%) of the 780 patients who entered the study dropped-out before the two week blinded portion of the study was completed. The largest number of drop-outs 19 (40% of all of the dropouts) were due to the ineffectiveness of the treatment (14 of these coming from placebo and 14mcg treatment groups).

Table 2. Study 311: Discontinued Patients by Reason and Treatment Group

Reason	Total	P12				HFA			
		Placebo	14mcg	110mcg	440mcg	Placebo	14mcg	110mcg	440mcg
n	780	54	113	115	108	57	113	107	113
Ineffective	19	3	6	1	1	4	1	1	2
Adverse Event	4	0	0	1	0	1	2	0	0
Other	25	3	4	5	2	0	3	4	4
Total	48	6	10	7	3	5	6	5	6

Source: NDA 20-784, Vol. 1.23, p. 8-3-36

Nearly 98% of the study subjects were characterized as “evaluable” by the sponsor. In general, the percentages of evaluable patients were well-balanced across treatment groups. The HFA 440mcg group had the lowest percentage with 94.7%.

Table 3. Study 311: Evaluable patient populations by treatment group

Population	Total	P12			HFA			
		Placebo	14mcg	110mcg	440mcg	14mcg	110mcg	440mcg
All Treated	780	111	113	115	108	113	107	113
All Treated w/ Data	775	110	113	114	108	113	105	111
Evaluable	762	109	112	112	106	111	105	107
% Evaluable	97.7	98.2	99.1	97.4	98.1	98.2	98.1	94.7

Sources: NDA 20-784, Vol. 1.23, p. 8-3-36 and 8-3-38

Reviewer's Comments: These results indicate that the study was successfully implemented. As there were low numbers of drop-outs, and these drop-outs were fairly, evenly balanced, it is not anticipated that this factor will have a large effect on the analysis of study results. As expected, indicative of efficacy, there were somewhat larger numbers of dropouts for non-effect in the placebo and "sub-therapeutic" 14mcg treatment groups.

In addition to assessing the evaluability of patients, the sponsor also determined whether or not data for particular days were evaluable. However, the analyses for the "All Treated" population include all treated patients and data for all days recorded. Results for these different populations and different sets of data are, in general, similar. Results described in this review, unless otherwise noted, are for the All Treated population.

Regarding compliance: the sponsor states that "In general, the patients followed the recommended dosing regimens...throughout the study." The reported high percentages of patients considered "evaluable" by the sponsor tend to support this contention (only 5 of the 13 patients were classified as "not evaluable" based on use of disallowed therapy). No additional data were cited in the study report to verify compliance..

3.3 Treatment Comparisons

Reviewer's Comments: The sponsor provided a large number of statistical comparisons to evaluate the effectiveness of the treatments. In assessing the strength of evidence from Study 311, this review primarily focuses on the analysis of Nasal Index (NI) 24-hour reflective and "snap-shot" scores to assess the strength of evidence supporting the sponsor's claim that the HFA and CFC products are "comparable."

3.3.1 Data and Program Files

The sponsor submitted two diskettes (DEMO311 and EFFW311) "including SAS datasets which were used to generate the efficacy for the RG 5029T-311 study." [ref. NDA 20-784, Vol. 1.23, p. 8-3-36]. The sponsor described these files by including proc contents of the files, an "outline" of the SAS datasets, and a description of "useful" variables. The sponsor documented the analyses programs by providing SAS programming language code in Sections B-D of the "Revised Statistical Analysis Plan" dated 3/14/95." [ref. NDA 20-784, Vol. 1.23, p. 8-4-9 through 8-4-16].

Reviewer's Comment: This reviewer was able to verify to his satisfaction that the data and program files submitted by the sponsor support the analytical results presented in this NDA.

3.3.2 Two Week Mean 24-Hour Reflective Symptom Scores

Study results describing two week mean 24-Hour Reflective Symptom Scores for Study 311 are presented in Figure 2 and Table 4. These 24-hour scores, as described above, were identified prospectively by the sponsor as primary endpoints for this study. Scores for HFA and CFC treatment groups ranged from about 2.0 to 2.8. For doses above 14mcg (the sponsor's "sub-therapeutic" dose) the CFC treatment scores were higher than those recorded for the HFA product – with an effect difference of 0.26 for 110mcg treatment groups and 0.20 for the 440mcg groups.

To establish the validity of the trial, controlling for multiple comparisons, the sponsor tested the within treatment contrasts of the treatments vs. a combined CFC/HFA control group (see below). As the results in Table 4 demonstrate, as anticipated, all treatments (even the so-called sub-therapeutic 14mcg dose) were statistically superior to the placebo.

The sponsor also provided unadjusted p-value assessments (unadjusted for multiple comparisons) describing differences between dose groups within each formulation. The p-values for CFC dosage contrasts between both higher doses (110mcg and 440mcg) and 14mcg were less than 0.05, however only with the comparison of the highest dose of the HFA product vs. the 14mcg dose does the p-value comparison fall below 0.05.

Figure 2. 24-hour Reflective Nasal Index (NI) Mean Change from Baseline by Treatment Group and Dose

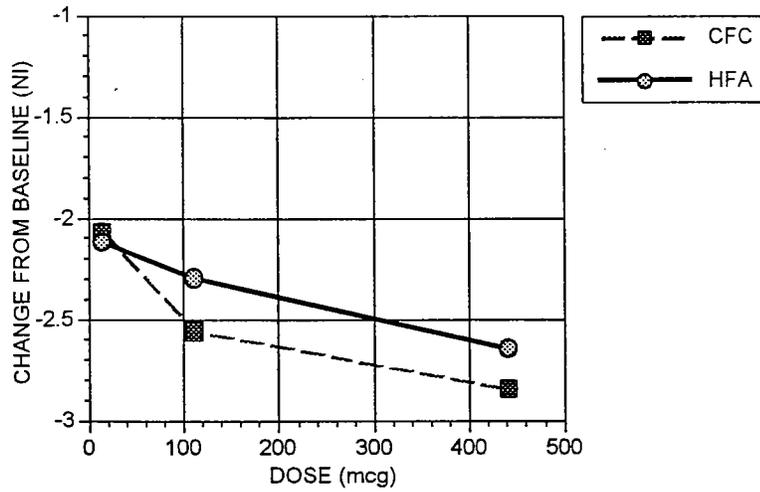


Table 4. Change from Baseline in 24-Hour Reflective Mean Nasal Index Symptom Score by Treatment (All Treated Patients, Two Week Means)^a

Treatment	N	Mean Baseline	Adjusted Change from Baseline (%)	Level of Significance / p-Value ^b vs.		
				Placebo	14mcg	110mcg
14mcg CFC	113	6.99	-2.06 (-29.5)	**		
110mcg CFC	114	6.65	-2.55 (-38.3)	**	0.04	
440mcg CFC	108	6.73	-2.84 (-42.2)	**	<0.01	0.24
14mcg HFA	113	7.06	-2.11 (-29.9)	**		
110mcg HFA	105	6.41	-2.29 (-35.7)	**	0.47	
440mcg HFA	111	6.78	-2.64 (-38.9)	**	0.03	0.16
Placebo	109	6.75	-1.39 (-20.6)			

- a p-values are computed from t-tests for a two way analysis of variance model with treatment and center as main effects and no interaction term
- b p-values for comparisons with placebo are adjusted for multiple comparisons, descriptive p-values presented for comparisons between active treatments are not adjusted for multiple comparisons.
- ** For comparisons of active treatments to placebo, significant, $p < 0.05$, 2-tailed, with adjustment for Bonferroni adjustment 6 multiple comparisons

Sources: NDA 20-784, Vol. 1.23, p. 8-3-69 and Vol. 1.24, p. 8-4-60, and Vol. 1.24, p 8-4-74

Reviewer's Comments: With the assumption that it is acceptable to combine placebo groups (see discussion below), these results strongly demonstrate that, as expected, the CFC and HFA treatment groups were significantly better than placebo in the trial. This conclusion is strengthened by the prospectively planned application of the Bonferroni adjustment for multiple comparisons employed by the sponsor (a conservative approach to this analysis).

In general, the response of the groups to the active treatments (to some extent satisfying the CFC Switch Guidance) appears to be similar, each formulation demonstrating a clear dose response with a similar increase. However, with only 3 points it is somewhat difficult to compare the shapes of the dose response curves. For example, in comparison to the HFA, the effect of CFC product rose noticeably more quickly between the 14mcg and 110mcg doses.

It is also evident that for the 110mcg and 440mcg treatment groups, the CFC product scores were greater than those observed for the HFA – the dose response curves appear to parallel one another. The sponsor, perhaps not anticipating differences in the formulations, did not plan for the direct between treatment statistical testing of the groups (i.e., there are no direct comparisons of CFC and HFA 110mcg treatment groups in the study report.)

3.3.3 Comparability

To statistically test the “comparability” of the HFA and CFC products (between formulation comparisons) the sponsor used a “two one-sided 90% confidence interval ratio” approach. The results of these analyses are presented in Table 5. Reflecting the point estimates described above, the ratios were between 0.91 (for the 110mcg CFC/HFA ratio) to 1.03 (for the 440mcg CFC/HFA ratio). The lower bounds for the two one-sided 90% confidence intervals ranged from 0.75 (110mcg) to 0.82 (14mcg) – successfully meeting the sponsor’s pre-specified “comparability” conditions (i.e., the lower bounds of the CI’s needed to be greater than 0.70 for the treatments to be declared “comparable.”)

Table 5. Change from Baseline in Reflective 24-Hour Nasal Index Symptom Score by Treatment (Two One-sided 90% Confidence Interval Ratios, All Treated Patients, Two Week Means)

Treatment	P12		HFA-134a		Ratio HFA/P12	Two one-sided 90% CI of Ratio
	N	Mean Change	N	Mean Change		
14 mcg	113	-2.04	113	-2.09	1.03	(0.82 1.23)
110 mcg	114	-2.54	105	-2.30	0.91	(0.75 1.07)
440 mcg	108	-2.84	111	-2.65	0.93	(0.78 1.09)

Source: NDA 20-784, Vol. 1.23, p. 8-3-47 [Note: referenced table in submission (Vol. 1.24, Table 8.1.1) did not include a description of results for analysis of Nasal Index].

Reviewer's Comment: In the absence of established guidance regarding the assessment of "comparability", the sponsor took a reasonable approach in statistically testing the "comparability" for treatment groups at given doses. A similar approach was recommended for the assessment of therapeutic equivalence by Huque, et al. However, the sponsor's design calls for a number of tests without planned adjustment for the multiple comparisons. Huque and his colleagues describe a methodology for a single comparison of test drug and comparator – there is no allowance made for comparisons at a number of dose levels. In this case, all three of the ratio CI estimates fell within the prescribed limits – the sponsor did not discuss how to make a "comparability" assessment if any of these dose comparisons fell outside the pre-specified limit? (In fact a number of the tests for individual symptoms failed this comparability test procedure). Additionally, as noted above, no agreement has been made by clinicians that formulations of this type can vary by 30% and still be called "comparable."

The sponsor's statistical approach, though reasonable, does not appear to be good fit with the intent of the CFC switch program – where guidance dictates that comparability be assessed by a comparison of dose response curves, not at given, specific doses. The sponsor's approach presupposes that equivalent doses of the two drugs have equivalent effects – following the notion that going into the trial the correct doses for the "switch" have been well established. In spite of the statistical testing procedures endorsed by the sponsor, this did not appear to be the case in this trial, in that the measurements for the HFA were consistently lower than the CFC at "therapeutic" doses (110mcg and 440mcg). If the estimated dose response curve described in Figure 2 is accurate, then it appears that the 100mcg dose of CFC would produce an effect that is equal to about 300mcg of the HFA product. Given this visual assessment of the dose response curves, it might be possible to describe these two products as "comparable" within the limits of the CFC Switch Program, but the two products (at the doses selected for the trial) should not be considered "therapeutically equivalent." If these differences are real, it is important that patients and physicians be made aware of these results.

3.3.4 Pairwise Comparisons: Across Formulations – the Double-Delta Approach

The sponsor reported that the “double delta differences, which are defined as (P12 dose i minus P12 dose j) minus (HFA-134a dose i minus HFA-134a dose j), ranged from -0.31 and 0.35 for the nasal index and were not found to be significantly different from zero. Therefore, the magnitude of the difference in response to efficacy between doses within a formulation was the same across formulations for the nasal index. Similar results were obtained for nasal stuffiness, nasal discharge and sneezing” [NDA 20-784, Vol. 1.23, 8-3-86].

Reviewer's Comment: This “double-delta” approach appears to be an attempt to statistically assess differences in the shape of the dose responses of the formulations. Though reasonable, it is not apparent that this is a sensitive / powerful means of testing differences between the curves. If the goal is to provide a useful statistical approach to the assessment of “comparability” or “similarity” of dose response curves – it would, perhaps be more useful to apply more sensitive methodology, one that better reflects the results of the study. For example, a more discriminating test might have been to model these data longitudinally (e.g., beginning with the 14mcg dose, the NI data for the HFA product, appear linear, while the CFC results, for the same doses, appear to be better represented by a quadratic).

3.3.5 “Snap-Shot” Symptom Scores

“Snap-Shot” scores, describing symptoms for the one hour period prior to dosing were collected at 10 of the 16 clinics for about two thirds of the treated patients (512 subjects). These snap-shot scores assess end-of-dosing-interval efficacy. Results are presented in Figures 3 and 4 and Tables 6 and 7 (Figure 3 and Table 6 describe results for mean data over the entire two weeks of double-blind treatment, while Figure 4 and Table 7 display results for Day 12).

Average change-from-baseline, snap-shot scores for the two week treatment period for the CFC and HFA treatment groups ranged from -1.08 to -2.07 – generally lower improvement scores than those recorded for the 24 reflective endpoint. As with the 24 hour reflective scores described above, the 110mcg HFA treatment group, in a comparison with the 110mcg P12 treatment group, demonstrated noticeably less improvement. However, with this secondary endpoint, the 440mcg HFA product had a numerically larger mean decrease in NI than did the CFC product.

Differences between the groups in snap-shot results for Day 12, near the end of the double-blind treatment period, were more pronounced, with 110mcg HFA product providing less efficacy than even the “sub-therapeutic” 14mcg dose (Figure 4).

Figure 3. Change From Baseline, “Snap-Shot” Nasal Index (NI) Symptom Score by Treatment (All Treated, Two Week Means)

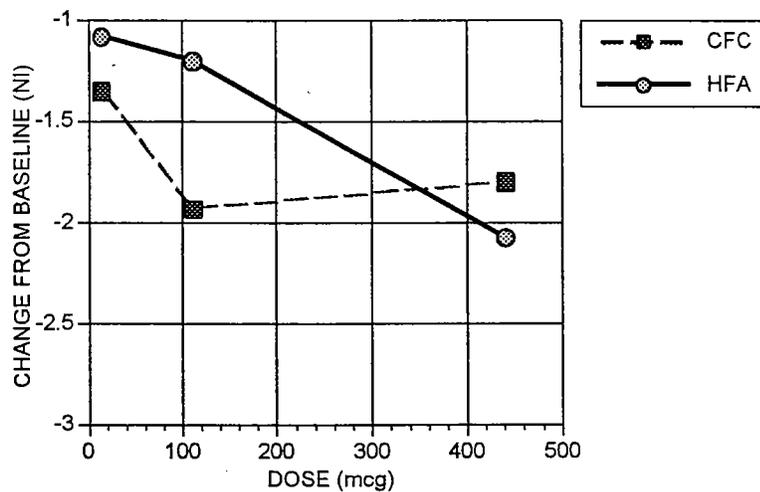


Table 6. Change from Baseline in “Snap Shot” Nasal Index Symptom Score by Treatment (All Treated, Two Week Means)^a

Treatment	N	Mean Baseline	Adjusted Change from Baseline (%)	Level of Significance / p-Value ^b vs.		
				Placebo	14mcg	110mcg
14 mcg CFC	75	6.45	-1.35 (-20.9)	1.00		
110 mcg CFC	77	6.18	-1.93 (-31.2)	**	0.08	
440 mcg CFC	72	5.87	-1.80 (-30.7)	0.06	0.17	0.70
14 mcg HFA	73	6.46	-1.08 (-16.7)	1.00		
110 mcg HFA	68	5.49	-1.20 (-21.9)	1.00	0.71	
440mcg HFA	76	6.21	-2.07 (-33.3)	**	<0.01	<0.01
Placebo	71	6.17	-0.94 (-20.6)			

- a p-values are computed from t-tests for a two way analysis of variance model with treatment and center as main effects and no interaction term
- b p-values for comparisons with placebo are adjusted for multiple comparisons, descriptive p-values presented for comparisons between active treatments are not adjusted for multiple comparisons.
- ** For comparisons of active treatments to placebo, significant, $p < 0.05$, 2-tailed, with adjustment for Bonferroni adjustment, 6 multiple comparisons

Sources: NDA 20-784, Vol. 1.23, p. 8-3-81 and Vol. 1.24, p. 8-4-80, and Vol. 1.24, p. 8-4-86

Figure 4. Change from Baseline in “Snap-Shot” Nasal Index (NI) by Treatment Group and Dose – Day 12

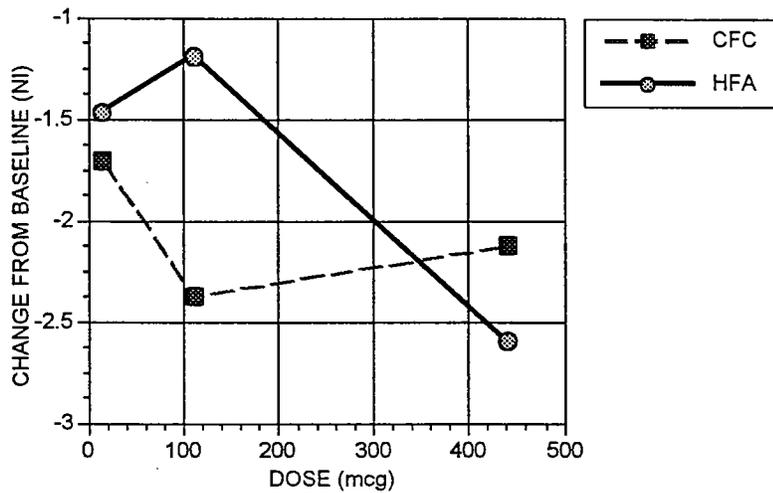


Table 7. Change from Baseline in “Snap Shot” Nasal Index Symptom Score (All Treated, Day 12 Means)^a

Treatment	N	Mean Baseline	Adjusted Change from Baseline (%)	Level of Significance / p-Value ^b vs.		
				Placebo	14mcg	110mcg
14 mcg CFC	65	6.44	-1.70 (-26.4)	0.22		
110 mcg CFC	72	6.27	-2.37 (-37.8)	0.01	0.12	
440 mcg CFC	69	5.94	-2.12 (-35.7)	0.03	0.33	0.55
14 mcg HFA	67	6.47	-1.46 (-22.6)	0.49		

110 mcg HFA	65	5.37	-1.18 (-22.0)	0.97	0.51	
440mcg HFA	72	6.24	-2.59 (-41.5)	<0.01	0.01	<0.01
Placebo	64	6.15	-1.16 (-18.9)			

- a p-values are computed from t-tests for a two way analysis of variance model with treatment and center as main effects and no interaction term
- b p-values for comparisons with placebo are adjusted for multiple comparisons, descriptive p-values presented for comparisons between active treatments are not adjusted for multiple comparisons.

Source: Fax dated 10/27/97 from RPR

Reviewer's Comments: The Snap-shot symptom scores collected on a subset of the patients tended to support and strengthen the results reported for the pre-defined primary 24-hour reflective scores. In particular, the 110mcg HFA treatment group appeared to be an somewhat of an "outlier," resulting in a distinctively different dose response curve for the HFA product compared to the relationship exhibited by the CFC product.

3.3.6 Day 1 - 4 Symptom Scores: Reflective and Snap-Shot

Results for Days 1 - 4, 24-Hour and Snap-Shot Symptom Scores are displayed in Figures 5 and 6, and Table 8. Both reflective and snap-shot scores for Days 1 - 4, demonstrate that patients receiving HFA and CFC products generally improved over the first four days of the double-blinded portion of the study. From the first day the results for the 24-Hour Reflective Scores reflect the superiority of all of the active treatments, while snap-shot scores for the lowest dose (14mcg) were not well differentiated from placebo. In general these scores appear to show a pattern of dose response, with higher doses outperforming lower doses of a given formulation. As noted above for the two week means, 24-hour reflective scores demonstrate the superior performance of the 440mcg CFC product. By Day 4 it appears that the treatments were approximate 80 - 90 % of the average for the two week treatment period.

Table 8 includes (Bonferroni-adjusted) p-values describing these results. In a dose-by-dose comparison the CFC product generally outperformed the HFA product (e.g., the 440mcg dose of the CFC product demonstrated consistent statistically significant results from Day 1, while the results for the highest dose of the HFA product, with symptom scores consistently below the CFC, were statistically significant starting on Day 2).

Though the snap-shot scores appear to support the results for the 24-hour reflective

scores, they tended to be less discriminating. For example, the 440mcg CFC product was not clearly superior for this endpoint -- the snap-shot scores for the 110mcg and 440mcg CFC product and the 440mcg HFA treatment were very similar over the first four days of the study.

Figure 5. Change From Baseline in 24-Hour Reflective Nasal Index Symptom Score by Treatment and by Day (All Treated, Days 1 - 4)

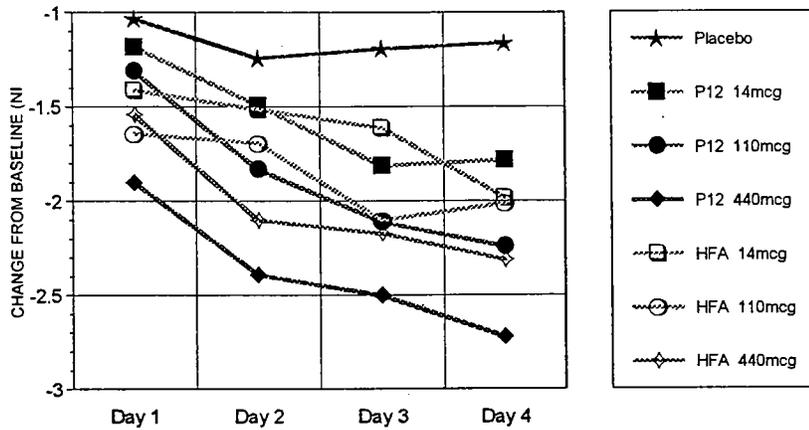


Figure 6. Change From Baseline in Snap-Shot Nasal Index Symptom Score by Treatment and by Day (All Treated, Days 1 - 4)

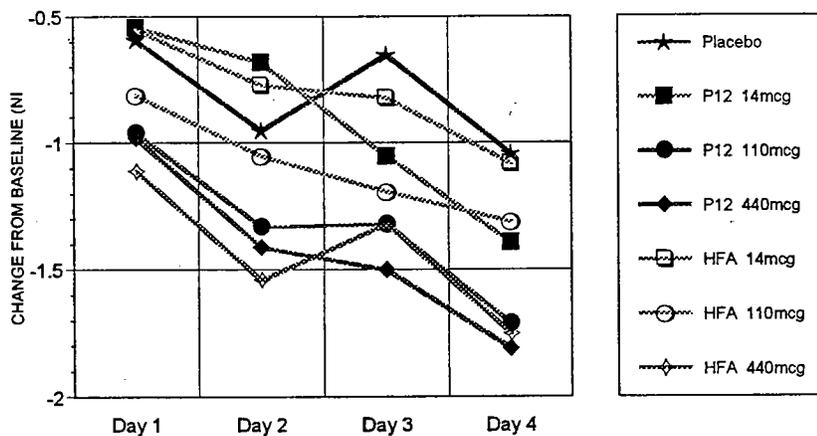


Table 8. Change from Baseline in 24-Hour Reflective and “Snap Shot” Nasal Index Symptom Score by Treatment (All Treated, Days 1 - 4 Means)

Day	Treatment	N	Reflective		N	Snap-shot		
			Mean Change from Baseline	p-Value (placebo comparison)		Mean Change from Baseline	p-Value (placebo comparison)	
1	Placebo	108	-1.03		70	-0.59		
	P12	14mcg	112	-1.18	1.0	74	-0.54	0.84
		110mcg	113	-1.31	1.0	77	-0.96	0.32
		440mcg	107	-1.90	0.00*	72	-0.98	0.30
	HFA	14mcg	109	-1.41	0.83	72	-0.55	0.85
		110mcg	104	-1.64	0.12	68	-0.81	0.59
		440mcg	108	-1.54	0.29	76	-1.11	0.14
2	Placebo	105	-1.24		70	-0.95		
	P12	14mcg	112	-1.49	1.00	75	-0.68	0.46
		110mcg	113	-1.83	0.18	77	-1.33	0.33
		440mcg	107	-2.39	0.00*	71	-1.41	0.25
	HFA	14mcg	112	-1.51	1.00	73	-0.77	0.58
		110mcg	105	-1.69	0.65	68	-1.05	0.86
		440mcg	110	-2.10	0.01*	75	-1.54	0.12
3	Placebo	106	-1.19		69	-0.65		
	P12	14mcg	112	-1.81	0.17	75	-1.05	0.29
		110mcg	114	-2.11	0.01*	77	-1.32	0.08
		440mcg	107	-2.50	0.00*	72	-1.50	0.03*
	HFA	14mcg	111	-1.61	0.79	72	-0.82	0.71
		110mcg	105	-2.10	0.01*	68	-1.19	0.19
		440mcg	106	-2.17	0.00*	76	-1.32	0.08
4	Placebo	105	-1.42		68	-1.04		
	P12	14mcg	111	-1.78	1.00	73	-1.39	0.41
		110mcg	113	-2.24	0.03*	77	-1.71	0.09
		440mcg	107	-2.72	0.00*	72	-1.81	0.07
	HFA	14mcg	111	-1.98	0.32	73	-1.08	0.98
		110mcg	105	-2.01	0.35	68	-1.31	0.59
		440mcg	109	-2.31	0.02*	76	-1.75	0.07

* p < 0.05

Note: P-values for reflective scores, Bonferroni adjusted for 6 comparisons; p-values for snapshot scores have not been adjusted for multiple comparisons.

Sources: Reflective Scores – NDA 20784, Vol. 1.24, p. 8-4-220, Snap Shot Scores – fax dated October 27, 1997

Reviewer's Comment: These Day 1 - 4 results appear to indicate that the 440mcg dose of the CFC product, with adjusted statistically significant results on the first day, appears to work faster and better than does the HFA product at the highest dose. It's not apparent as to whether this difference is demonstrative of a lack of "comparability."

3.3.7 Other Secondary Endpoint Analyses

Within formulation analysis of patient and physician global assessments generally support the conclusions that 14mcg, 110mcg and 440mcg doses of HFA and CFC were efficacious, and that 110mcg and 440mcg doses were superior to the "sub-therapeutic" 14mcg dose of each product. In describing these results, the sponsor states, "both formulations had statistically significant improvement when compared to the combined placebo treatment group for the patients' and physicians' global evaluations. When adjusted for multiple comparisons were made, statistical significances were maintained for the 110mcg and 440mcg groups." [ref. NDA 20-784, Vol. 1.23, p. 8-3-88]

3.3.8 Exploratory/Subgroup Analyses

3.3.8.1 Gender

The study population consisted of more males (55%) than females (45%). The sponsor provided a brief description of 24 hour reflective scores by treatment group for males and females and observed that "no substantial clinical differences on efficacy were noted between sexes." [ref. NDA 20784, Vol. 1.49, p. 8-29-78].

Reviewer's Comment: It is interesting to note, in contrast to the sponsor's observation (that there appeared to be "no substantial clinical differences") that gender specific dose response curves for males and females were visually quite different – female data demonstrated no observed difference in efficacy for 14mcg and 110mcg doses of both HFA and CFC products, while male data evidenced much clearer dose response patterns. Though potentially noteworthy, it is not clear that the observed variation in results has a physiological rationale.

3.3.8.2 Age Group

3.3.8.2.1 Pediatric

Though the sponsor has requested an indication for treatment of children 6 years of age and older, no children were included in Study 311 (age range: 18 - 83).

3.3.8.2.2 Geriatric

Less than 4% of the patients in the study were 60 years of age or greater. The sponsor has provided 24 hour reflective scores for two age groups (18-59, 60+), and observes that "results for both age groups were similar, no patterns emerged which suggested a difference in treatment response among the two age groups." [ref. NDA 20784, Vol. 1.49, p. 8-29-80].

Reviewer's Comment: With only 29 patients who were 60+ years of age, these data do not support — similarity/comparability in response patterns of the geriatric population.

**APPEARS THIS WAY
ON ORIGINAL**

4. Conclusions

With Study 311 meeting the specifications included in the "*Rationales Behind Various Issues Pertaining to the Nasal Spray Switch Points-to-Consider Document*," the sponsor has provided evidence from one adequate and well-controlled study that its Nasacort HFA Nasal Inhaler is effective in treating seasonal allergic rhinitis and is "comparable" to its currently marketed CFC product. All tested doses (14mcg, 110mcg and 440mcg) of both the HFA and CFC products in this two week study were significantly better than placebo in treating combined symptoms included in the Nasal Index combined symptom score (NI -- nasal stuffiness; nasal discharge; and sneezing). Demonstrating "comparability," it is evident by visual inspection that the dose response curves for the two formulations, as described above, are roughly comparable.

As noted above, the comparison of the dose response curves was made somewhat more difficult by the observed results for the 110mcg HFA treatment group. It is possible that these somewhat discrepant observations were the result of the random selection of a less severe group of patients for this treatment group.

Though the CFC and HFA products are "comparable," in the sense that they appear to improve symptoms in a similar pattern over a similar range, it is also apparent from this study that the HFA doses used in the trial may not be as efficacious as the currently marketed CFC products at the same doses (e.g., as described above, the CFC 400mcg product was consistently more effective than HFA 400mcg formulation in terms of the primary endpoint – the 24-hour Reflective Nasal Index).



Stephen E. Wilson
Mathematical Statistician

Concur: S. E. Nevius *SEN 12-17-97*

cc:

Orig. NDA 20-784
HFD-570 / Division File
HFD-570 / SBarnes, PHonig, Jjenkins
HFD-715 / Chron
HFD-715 / SWilson, SENevius