CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

20872

STATISTICAL REVIEW(S)

STATISTICAL REVIEW AND EVALUATION
CLINICAL STUDIES

Date: JUL 12 1999

NDA#:	20-872
Applicant:	Hoechst Marion Roussel, Inc. / Quintiles
Name of Drug:	Allegra (fexofenadine HCl)
Indications:	Seasonal allergic rhinitis once daily for adults (≥12 years) Seasonal allergic rhinitis twice daily for children (6-11 years) Chronic idiopathic urticaria twice daily for adults (≥12 years)
Documents Reviewed:	7/17/98 Volumes 1.1, 1.63-1.351; 9/29/98; 11/19/98; 2/16/99; 4/29/99 electronic data; 5/14/99; 5/24/99; 6/4/99; 6/18/99
Statistical Reviewer:	Barbara Elashoff, M.S.
Medical Input:	Alexandra Worobec, M.D.

Summary

- The sponsor submitted three studies to support a seasonal allergic rhinitis (SAR) indication in adults (administered using the once daily dosing regimen), two studies to support an SAR indication in children (twice daily dosing regimen), and three studies to support a chronic idiopathic urticaria (CIU) indication in adults (twice daily regimen).

SAR Adult, Once Daily

- The results of one of the three adult SAR studies (Study 81) provide statistical evidence of a small treatment effect for Allegra, administered once daily, using the primary endpoint, end-of-dosing interval Total Symptom Score. The Total Symptom Score (TSS) was the sum of four symptoms rated on a severity scale of 0 – 4, resulting in a TSS severity scale of 0 - 16. The difference between Allegra and placebo at the end-of-dosing interval in Study 81 was 0.3 TSS units for 120 mg QD, and 0.5 TSS units for 180 mg QD.

- The average improvement of the patients receiving the 180 mg QD dose in Study 81 was consistently greater than that of the 120 mg QD dose for all secondary endpoints.

- The results of the other two adult SAR studies (Studies 32 and 61) demonstrated statistical significance on the primary endpoint (TSS assessed over the previous 12 hours). However the results of these studies should be viewed with caution as the studies appear to have been of uneven quality, the study reports contained numerous errors (some corrected, some identified, but not corrected). The sponsor did not submit the electronic data for these studies, thus a thorough statistical review could not be performed.

- Corrected study reports and the electronic data for Studies 32 and 61 should be submitted for review.

SAR Pediatric, Twice Daily

- The results of the pediatric SAR studies were inconsistent, making it difficult to assess the efficacy of Allegra in the pediatric population. Study 77 provides statistically significant evidence of a treatment effect (0.8 – 1.0 TSS units assessed over the previous 12 hours) for all of the doses: 15, 30, and 60 mg. The treatment effect at the end-of-dosing interval (secondary endpoint) was between 0.6 – 0.8 TSS units. The lowest dose (15 mg) resulted in the greatest efficacy, whereas the middle dose (30 mg) was least efficacious. The results of the other pediatric study (Study 66) did not demonstrate a statistically significant difference (nor a favorable numerical difference) between the active treatments and placebo. On average, the symptoms of the placebo patients improved more than the symptoms of the patients who received Allegra.

CIU Adult, Twice Daily

- Allegra, given twice daily, demonstrated efficacy for CIU in adult patients. The results of two of the three adult CIU studies (Studies 39 and 67) provide statistically significant evidence of a treatment effect (0.3 - 0.7 units, on a severity scale of 0 - 4) for the twice daily regimen. The results of the third study (Study 19) provide statistically significant evidence of a treatment effect, however, there was a 59% placebo dropout rate, and the study differed substantially from the Studies 39 and 67 (dosing regimen was once daily, formula was capsules instead of tablets, and the patients did not assess symptoms at the end-of-dosing interval).

1 Introduction

The sponsor submitted the results from eight clinical trials in this NDA, three to support once daily dosing for the SAR indication, two to support a pediatric SAR indication, and three to support a twice daily dosing for the chronic urticaria indication. Critical features of these trials are summarized in the table below. Note that the sponsor submitted electronic data for five of the eight studies.

Table 1: Study Characteristics

Indication	Study Number	Dates	Location	Formulation	Doses	Total N Rand-omized	Electr-onic data?
SAR Once Daily Adult	**81**	8/97-11/97	US	Tablets	Placebo 120 QD 180 QD	864	√
	32	5/95-2/96	Europe, South Africa, Australia	Capsules; (Cet Tablets enclosed in Caps)	Placebo 120 QD 180 QD Cet 10mg QD	842	
	61	5/96-8/96	Europe	Capsules	Placebo 80 QD 120 QD	1415	
SAR Twice Daily Pediatric	**66**	8/97-11/97	US, Canada	Tablets	Placebo 15 BID 30 BID 60 BID	463	√
	77	8/97-11/97	US, Canada	Tablets	Placebo 15 BID 30 BID 60 BID	414	√
CIU Twice Daily Adult	**39**	11/96-6/97	US, Canada	Tablets	Placebo 20 BID 60 BID 120 BID 240 BID	468	√
	67	11/96-6/97	US, Canada	Tablets	Placebo 20 BID 60 BID 120 BID 240 BID	476	√
CIU Once Daily Adult	**19**	3/95-3/96	Europe	Capsules	Placebo 60 QD 120 QD 180 QD 240 QD	224	

2 Quality of Submission

The eight studies submitted to this NDA were of uneven quality. The studies the sponsor considered "pivotal" appeared to be adequate and well-controlled, and were of the best quality (SAR: Study 81; Pediatric SAR: Studies 66 and 77; CIU: Studies 39 and 67), whereas the other studies (SAR: Studies 32 and 61; CIU: Study 19) were of poor quality. Problems included large and differential dropout rates across treatment groups, treatment-by-center interactions, centers enrolling patients in more than one trial for the same indication, and an investigator convicted of fraud. The sponsor submitted electronic data for only five of the eight studies making it difficult to thoroughly investigate the dropout rates and interactions. Further complicating the review of this submission, two study reports included numerous errors (Studies 32 and 61). The sponsor found these errors after the reports were finalized for submission to the European Community, during a routine check for submission to the FDA.[1]

In the submission to the FDA, the sponsor provided an addendum to the study reports that listed the errors they had discovered and corrected. The list included the volume and page number of the text that was incorrect, along with the correct text. Many of these errors were means and p-values of the primary and secondary efficacy variables. The sponsor also inserted pages in the front of some appendices stating that all the information in the appendix was incorrect. It was difficult to perform a thorough statistical review of Studies 32 and 61 due to the problems with the study reports and the lack of electronic data.

[1] Teleconference with sponsor (April 30, 1999).

Thomas Edwards, an investigator in Studies 81 (adult SAR), 39 (adult CIU) and 66 (pediatric SAR), was convicted of fraud in 1998. The sponsor was notified and analyses for the primary endpoint without the data from Thomas Edwards were performed. The results did not change the conclusions from these studies.

The sponsor reported the data entry, audit and management process in the NDA. This information is rarely reported to the FDA. The process is summarized for each study in the appendix (Table A- 1). The sampling plans and tests in the audit processes differed across studies. It appears as though the audit process for Study 81 resulted in the least percentage of errors detected before unblinding (0.0%). Interestingly, however, Study 81 also appears to have had the greatest number of errors corrected after unblinding. The report of the audit processes provides some assurance that the validity of the data was a concern to the sponsor.

3 Seasonal Allergic Rhinitis (SAR) Adult

3.1 Introduction

Summary: Studies 81, 32 and 61 were conducted to study efficacy of the once daily dosing regimen for adult SAR patients. The sponsor considered Study 81 (conducted in the US) as "primary" evidence, and foreign Studies 32 and 61 as supportive. In general, the quality of the study conduct, data, and the study reports of Studies 32 and 61 are poor. Studies 32 and 61 had nine common investigative sites. These sites enrolled 100 patients in both studies, in subsequent spring pollen seasons in Europe. Other differences between Study 81 and Studies 32 and 61 include: use of capsules in the foreign studies (instead of tablets), reflective score as primary (instead of trough, instantaneous), treatment-by-center interactions for the instantaneous assessments, high and differential dropout rates (15%) for studies of short duration (1-2 weeks) and large numbers of randomized patients who either did not receive study medication or did not have any post-baseline measurement. Further, Study 61 had a short double-blind treatment period (1 week).

In addition to the problems with the study design, conduct, and results, the study reports for Studies 32 and 61 had numerous errors that were either fixed in an addendum, or identified, but not corrected. The sponsor did not submit electronic data for Studies 32 and 61, therefore this review will present the results from all the adult SAR studies, but focus on the results of Study 81.

Proposed Labeling Claims
The sponsor would like to make the following claims in the label:

Redacted 2

pages of trade

secret and/or

confidential

commercial

information

3.2 Results

3.2.1 Study Conduct

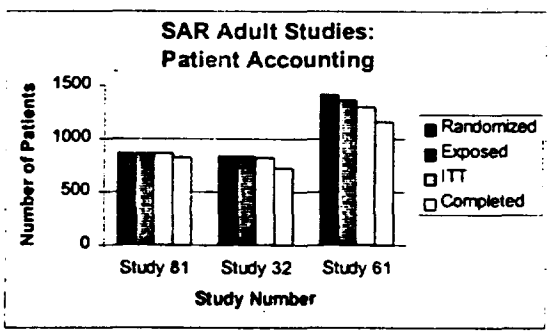
The studies were large (n>800) and had many centers (number of centers > 40). The sponsor defined Intent-to-Treat (ITT) as randomized patients who received study medication and had at least one post-baseline measurement. The ITT population in Study 81, as defined by the sponsor, excluded only 0.1% of the randomized patients who received at least 1 dose of the study medication ("exposed patients"). The sponsor-defined ITT populations in Studies 32 and 61 excluded 1.8% and 5.1% of the exposed patients, respectively. The smaller percentage of randomized patients included in the ITT population in Study 61 may have been due in part to the fact that patients were randomized prior to baseline. 119 randomized patients from Study 61 were not included in the ITT results. The effect that this missing data may have had on the results is not known.

Table 3: Numbers of Patients In Each Study

	Study 81 N (%)	Study 32 N (%)	Study 61 N (%)
Randomized	864	842	1415
Exposed	864 (100)	839 (99.6)	1366 (96.5)
ITT	863 (99.9)	824 (97.9)	1296 (91.6)
Completed	825 (95.5)	722 (85.7)	1161 (82.0)

Exposed = exposed to double-blind study medication
ITT = have both a baseline and at least one post-baseline measurement

Figure 4: Number of Patients by Study



Exposed = exposed to double-blind study medication
ITT = have both a baseline and at least one post-baseline measurement

Investigative Sites: The numbers of patients at the sites ranged from 1 to 36 in Study 81, from 0 to 165 in Study 32 and from 0 to 206 in Study 61, with most sites contributing between 1 and 35 patients.

Table 4: Distribution of Sizes of Sites

		Number of Sites with Sample Sizes of:			
	Total N Randomized	N=0	N=1-35	35 < N < 68	N>100
Study 81	864	0	45	1	0
Study 32	842	10	46	3	1 (n=165)
Study 61	1415	2	114	5	1 (n=206)

Studies 32 and 61 shared 9 common investigative sites. Two of these sites were the two largest sites in both studies. Martin Stern, of Midland Asthma and Allergy Research Association, enrolled 165 patients in Study 32 and 206 patients in Study 61. Peter Howarth, of Southamptom General Hospital in Great Britain, enrolled 68 patients in both studies. Recruitment in Great Britain for Study 32 occurred during the spring season of 1995

and for Study 61 in the spring season of 1996. The protocol for Study 61 did not discuss the eligibility of Study 32 participants. One-hundred patients enrolled in both studies (see Table 5 below). The sponsor did not indicate how many patients were *randomized* into both. (In Study 32, the number of patients enrolled did not equal the number of patients randomized.)

Table 5: Investigators in Studies 32 and 61

Investigator (Site # in Study 32/ Site # in Study 61)	Country	# Enrolled in Study 32	# Enrolled in Study 61	# Enrolled in Both	Exposed to Study Med in Both
Martin Stern (221/87)	U.K.	165	206	87	61
Peter Howarth (222/86)	U.K.	68	68	10	10
L.M. Adler (246/35)	Germ.	36	≤5		
F. Leynadier (236/220)	F/B	29	8		
F. Wessel (245/252)	F/B	27	11		
A. Sabbah (242/251)	F/B	≤6	≤18		
J. Bousquet (223/230)	F/B	7-8	12		
F. Favennec (231/250)	F/B	7-8	12	1	1
Martin Grosclaude (233/200)	F/B	7-8	36	2	2
Total				100	74

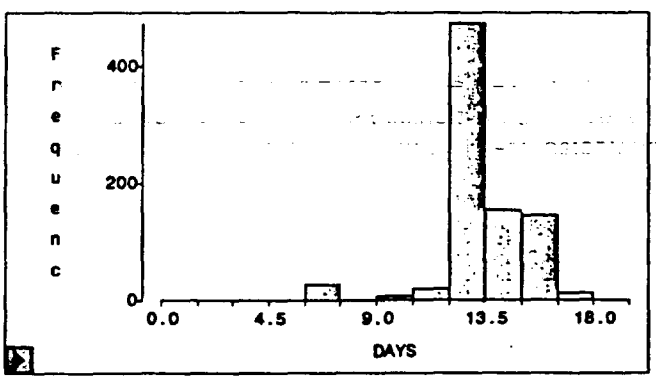
F/B: France/Belgium

Additionally, two patients in Study 81 also participated in Study 67, one of the chronic idiopathic urticaria studies.

The results of Study 32 and 61 are related due to the common patients and, therefore, should not be considered independent studies. The fact that these patients were allowed to be enrolled in both studies is indicative of poor study design (no exclusion criteria based on participation in previous study) and poor study conduct by the investigators.

Dropouts: The overall discontinuation rate in Study 81 was 4.5%, with all treatment groups yielding similar percentages. Sixty-three percent of the patients had fewer than 14 days of double-blind treatment. The patients with < 14 days were evenly distributed across treatment groups (placebo: 62%; 120 mg: 66%; 180 mg: 61%). Below are two histograms of the number of days the patients were on-study (difference between last visit and first visit, plus one) and the number of days the patients were on study during the double-blind treatment period (calculated similarly). Despite the skewed distribution, the mean and median number of days were similar (On-study: mean 20.6, median 21; Double-blind treatment period: mean 13.1, median 13).

Figure 5: Number of Days On Study During Double-Blind Treatment Period



The dropout rates in Studies 32 and 61 (14% and 15%, respectively) were much higher than the dropout rate observed in Study 81 (4.5%), based on the "exposed" population (not the randomized population). In Study 32, the dropout rate was highest for the placebo group (17%) and lowest for the cetirizine group (10%). In Study 61, the rate was highest for the placebo group (19%) and lowest for the 80 mg group (12%).

The 2-week average score was calculated using observed data only. This analysis assumes that if the patient had remained in the study, the subsequent symptoms would have equaled the average of the values prior to the dropout date.

Considering the fact that the percentage of patients who dropped out of Study 81 was low and the patients were evenly distributed across treatment groups, the missing data in Study 81 probably did not strongly influence the reported results. However, Studies 32 and 61 had large percentages of dropouts for studies of 1-2-weeks' duration and, therefore, the results from these studies are potentially less reliable.

The characteristics of the dropout patients and the completers may have been different. The sponsor did not provide this information and the reviewer could not investigate it without the electronic data. The large percentage and the differential percentages across treatment groups may have biased the results from these studies. The magnitude and direction of this bias is unknown.

3.2.2 Sample Size

The sample size calculations for Studies 81, 32 and 61 were based on results from the two studies submitted to the original Allegra NDA (Allegra 60 mg BID for the treatment of allergic rhinitis, NDA 20-625).

Table 6: Results from Studies from Original Allegra NDA (20-625)

TSS: sum of 4 symptoms (sneezing, rhinorrhea, itchy nose, itchy eye) rated on a scale of 0-4; TSS on a scale of 0-16				
		Time	TSS Trt Effect Size* 60 mg	TSS Standard Dev 60 mg
Study 23	Reflective	7 PM	1.07	1.7
		7 AM	1.11	1.9
N/group=141	Trough	7 PM	1.05	2.1
		7 AM	1.02	2.4
Study 24	Reflective	7 PM	0.66	1.6
		7 AM	0.52	1.8
N/group=138	Trough	7 PM	0.69	2.0
		7 AM	0.46	2.0

*Each of the four individual components of TSS were rated on a scale of zero to four. In general, the results of the individual symptom scores in Study 23 (submitted to the original NDA 20-62) demonstrated differences of 0.25 units. Since the TSS was a sum of 4 symptoms, the differences in TSS were, on average, about 1.0 units in Study 23. As will be seen, the current studies demonstrate statistically significant effects about half as large as that seen in Study 23. These differences were statistically significant due to the two- to three-fold increases in sample size.

Study 32 was powered to detect a difference of 0.70 units in average change of the 24-hour **reflective TSS** from baseline between placebo and a treatment, given a standard deviation of 2.4. A sample size of 200 patients per treatment group (total n=600) was determined to provide about 83% power. The study was planned to be conducted during the 1995 spring pollen season in Europe, but due to an unexpectedly rainy season, recruitment was low. Consequently, the study was extended to South Africa and Australia for the pollen season in October 1995/January 1996. A total of 842 patients were randomized (ITT n=824).

Study 61 was powered to detect a difference of 0.55 units in average change of the 24-hour **reflective TSS** given a standard deviation of 2.0. A sample size of 279 patients per treatment group (total n=837) was determined to provide about 90% power. Following the recruitment experience gained from Study 32, a "vigorous recruitment drive was initiated" for Study 61. This resulted in 1513 patients being screened. Of this number, 1415 patients were randomized (ITT n=1296) within a 4-month period (May-August 1996), well above the 837 necessary to provide 90% power.

Study 81 was powered to detect a difference of 0.55 units in average change of the 24-hour **trough instantaneous TSS** given a standard deviation of 2.2. A sample size of 250 patients per treatment group was determined to provide about 80% power. Due to the smaller than expected standard deviation in Study 81 (2.0 instead of 2.2), and the larger than planned sample size (863 instead of 750 ITT patients), effect sizes of 0.3 (about half of the 0.55 units the study was originally planned to detect) were statistically significant. Of the individual symptom scores, differences as low as 0.10 units (on a 0-4 scale) were statistically significant.

Table 7 below shows the planned versus actual sample sizes for each of the studies.

Table 7: SAR Studies' Planned versus Actual Sample Sizes

	Sample Size Calculations			Results		
Study	Delta	Planned N	Power	ITT N	Smallest Difference found to be Statistically Significant	
					TSS (0-16 scale)	Indiv. Sympt (0-4 scale)
81	0.55	750	80	863	0.30	0.10
32	0.70	600	83	824	0.58	0.17
61	0.55	837	90	1161	0.30	0.09

Enrolling more patients than necessary for 80-90% power gives a more precise estimate of the treatment effect, and is, in fact, beneficial in determining whether the study has met the objectives. However, in very large trials, very small differences can be statistically significant. A determination of the *clinical relevance* of the treatment difference should be made when assessing efficacy. Therefore, when reviewing the study results, the clinical relevance of the effect sizes should be considered in addition to the statistical significance of the results. It is generally accepted good clinical trial practice to stop enrolling patients when the planned sample size has been reached.

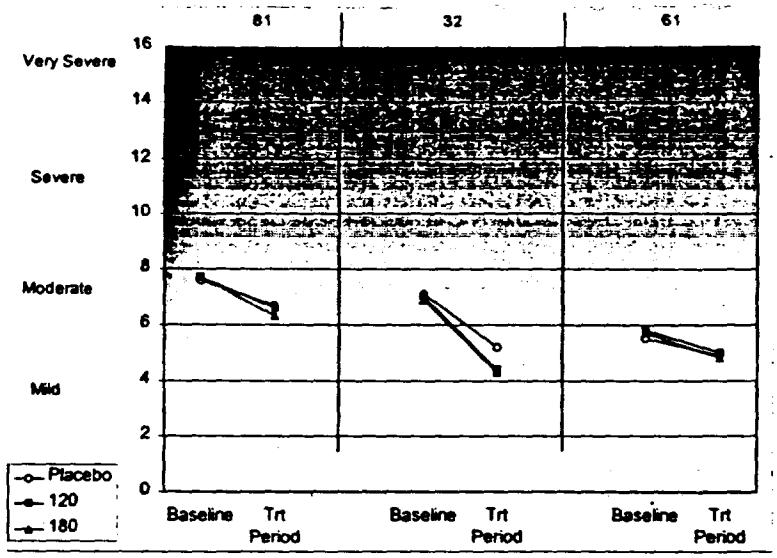
3.2.3 Demographics

Treatment groups in all three adult SAR studies were well balanced for demographic factors, with no statistically significant differences in gender, age, race, body weight, height or years since first SAR episode. The ITT population was predominantly female in Study 81 (65%), and more evenly distributed in Studies 32 (49% female) and 61 (44% female). The study population was overwhelmingly Caucasian in all three studies (Study 81: 88%, Study 32: 89%, Study 61: 94%). Patient ages ranged from 12-65 years in Study 81, 12-66 years in Study 32 and 12-70 years in Study 61. The mean age was 32 years in Studies 81 and 61, and 33 years in Study 32.

3.2.4 Primary & Secondary Efficacy Variables

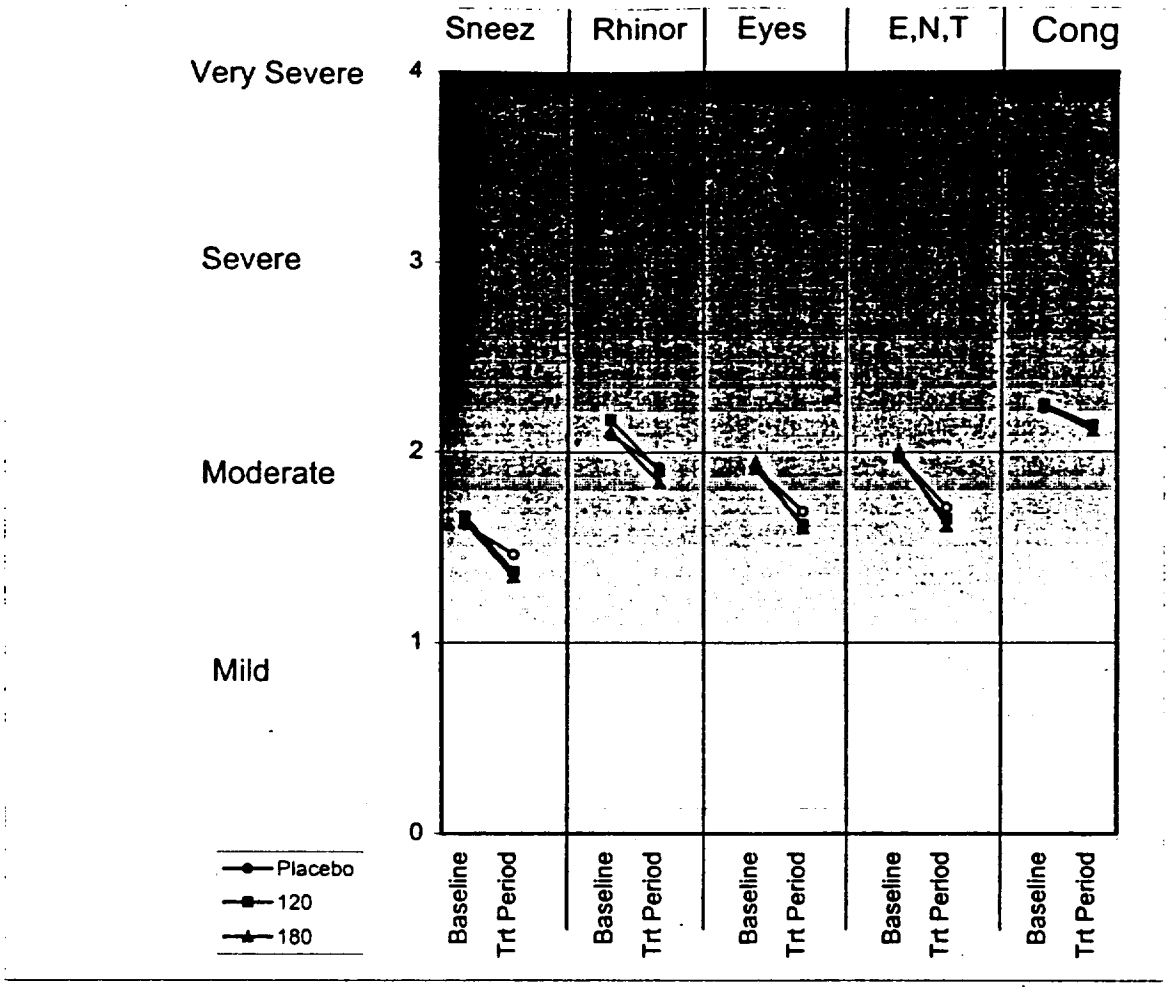
The SAR studies enrolled patients with moderate allergy symptoms. The ranges of the means in each of the studies at baseline and during the treatment period are plotted in Figure 6 below. Study 61 had the lowest baseline means (5.5 - 6.7, on the low end of moderate), whereas Study 81 had the highest baseline means (7.4 - 7.7, on the high end of moderate).

Figure 6: Average Baseline & Trt Period Instantaneous TSS Scores



The Study 81 individual symptom assessments are plotted in Figure 7 below. Patients appeared to be most affected by nasal congestion and least by sneezing at baseline. As will be seen, Allegra appeared to have the greatest effect on sneezing and the least on nasal congestion.

Figure 7: Study 81 Average Baseline & Trt Period Individual Instantaneous Symptom Scores



The baseline means were similar across treatment groups in all studies. A summary of the results of the primary and some secondary efficacy variables are presented in Tables 8 and 9 below.

Randomization was stratified by number of baseline period instantaneous TSS ≥ 5 in Study 81, by average 24-hour reflective TSS (≤ 8, >8) in Study 32, and by entry trough 24-hour reflective TSS scores (≤ 8, >8) in Study 61. The baseline scores appeared to be similar across treatment groups in all three studies (see Table 8 below).

Table 8: Adult SAR Studies' Descriptive Statistics of TSS

Means of changes from baseline are adjusted for investigative site and baseline value.
Standard errors ranged from 0.1 to 0.2 for all means in this table.

Assess	Study		Placebo	80 mg QD	120 mg QD	180 mg QD	Cetirizine
Instant	81	N	292		287	282	
		Baseline	7.6		7.7	7.7	
		Change	**-0.9**		**-1.1**	**-1.4**	
	32	N	193		202	192	200
		Baseline	7.1		7.0	6.9	7.1
		Change	**-1.9**		**-2.6**	**-2.6**	**-2.6**
	61	N	417	437	430		
		Baseline	5.5	5.8	5.7		
		Change	**-0.6**	**-0.8**	**-0.9**		
24-hr Ref	81	N	291		287	282	
		Baseline	7.4		7.5	7.4	
		Change	**-0.7**		**-1.3**	**-1.4**	
	32	N	201		211	202	207
		Baseline	7.3		7.2	7.4	7.3
		Change	**-1.9**		**-3.0**	**-3.3**	**-3.3**
	61	N	389	398	412		
		Baseline	6.5	6.6	6.7		
		Change	**-0.6**	**-1.7**	**-1.6**		

The mean changes from baseline for the placebo group in Study 81 were approximately 0.7 - 0.9 units. The mean changes in the active treatment groups in Study 81 were slightly greater (1.1 - 1.3 units). The mean changes in Study 32 were greater than those in Study 81 for both placebo and active treatment groups. Relative to Study 81, Study 61 had smaller mean changes from baseline for the placebo group and larger mean changes for the active treatment groups.

The primary efficacy variable for Study 81 was the 8 AM instantaneous assessment. The primary variable for Studies 32 and 61 was the 24-hour reflective assessment. These are shaded in Table 9 below. For Studies 81 and 32, in which the AM and PM reflective assessments were made separately, the results are presented for the reflective assessments separated by the first and second 12-hours of the dosing period (12-hr AM = second 12 hours, 12-hr PM = first 12 hours).

Table 9: Adult SAR TSS Results (Change from Baseline TSS)

Shaded cells indicate the primary efficacy variable for the study. Treatment differences in changes from baseline, confidence intervals and p-values are calculated from an ANCOVA model containing investigative site, treatment and baseline value. **Positive treatment differences indicate Allegra superiority.**

TSS: sum of 4 symptoms (sneezing, rhinorrhea, itchy nose, itchy eye) rated on a scale of 0-4; TSS on a scale of 0-16						
Assess	Study		80 mg QD	120 mg QD	180 mg QD	Cetirizine
Instant	81 n=864	Trt Diff (95% CI) p-value		**0.3** (0.0, 0.6) 0.0505	**0.5** (0.2, 0.8) 0.0016	
	32[1] n=842	Trt Diff (95% CI) p-value		**0.7** (0.1, 1.3) 0.0024	**1.1** (0.5, 1.7) 0.0001	**1.1** (0.5, 1.7) 0.0001
	61[2] n=1415	Trt Diff (95% CI) p-value	**0.2** (0.0, 0.4) 0.2133	**0.3** (0.1, 0.5) 0.0379		
24-hr Ref	81 n=864	Trt Diff (95% CI) p-value		**0.6** (0.3, 0.9) 0.0001	**0.7** (0.4, 1.0) 0.0001	
	32 n=842	Trt Diff (95% CI) p-value		**1.1** (0.7, 1.5) 0.0001	**1.4** (1.0, 1.8) 0.0001	**1.4** (1.0, 1.8) 0.0001
	61 n=1415	Trt Diff (95% CI) p-value	**1.0** (0.6, 1.4) 0.0001	**1.0** (0.6, 1.4) 0.0001		
12-hr AM Reflective	81 n=864	Trt Diff (95% CI) p-value		**0.5** (0.2, 0.8) 0.0012	**0.7** (0.4, 0.9) 0.0001	
	32 n=842	Trt Diff (95% CI) p-value		**0.9** (0.5, 1.3) 0.0001	**1.3** (0.9, 1.7) 0.0001	**1.2** (0.8, 1.6) 0.0001
12-hr PM Reflective	81 n=864	Trt Diff (95% CI) p-value		**0.8** (0.5, 1.1) 0.0001	**0.8** (0.5, 1.1) 0.0001	
	32 n=842	Trt Diff (95% CI) p-value		**1.2** (0.8, 1.6) 0.0001	**1.5** (1.1, 1.9) 0.0001	**1.5** (1.1, 1.9) 0.0001

1. The results presented here are from a model without the treatment-by-center interaction. The interaction was significant at the 0.10 level (p=0.0158). In a model with the interaction term, the treatment differences were smaller (120 mg QD: 0.6, p=0.0443; 180 mg QD: 0.7, p=0.0238; Cetirizine: 0.6, p=0.0457).
2. The results presented here are from a model without treatment-by-center interaction. The interaction was significant (p=0.0477). The sponsor did not report treatment differences for each treatment group in the model with the interaction term.

The sponsor's primary analysis (pre-specified) in all three studies was an analysis of covariance on the change from baseline TSS with baseline TSS as a covariate and the treatment group and investigative site as factors. (In Studies 32 and 61, the sponsor pooled sites according to geographic region and type of center, [allergy clinic or general practice] due to small numbers of patients at some sites.) The sponsor tested the treatment-by-baseline and treatment-by-center interaction terms at the 0.10 level. If the treatment-by-baseline term was significant at the 0.10 level, the sponsor included it in the final model and calculated the treatment effects using the average baseline scores.[2] The sponsor used the Step-Down procedure to protect the overall Type I error rate. All tests were conducted at the alpha-level of 0.05.

The results demonstrate that the treatment differences between drug and placebo were greatest for the 180 mg dose and greatest during the first 12-hours of the dosing period. In general, the treatment effects were smaller for the second 12-hours and still smaller for the last hour (8 AM instantaneous assessment) for all treatment groups.

Although all comparisons were statistically significant at the 0.05 level [or marginally significant (0.0505) in the case of the 8AM instantaneous comparison of the 120 mg dose in Study 81], the treatment effect sizes were relatively small for Studies 81 and 61. The largest treatment effect sizes were seen in Study 32, with effects ranging from 0.7 units for the instantaneous assessments to 1.5 units for the PM reflective assessment. The smallest treatment effects were seen in Study 81, with effects ranging from 0.3 units for the instantaneous assessments to 0.8 units for the PM reflective assessments. The 180 mg QD dose appeared to be more effective than the 120 mg QD dose and appeared to be similar in efficacy to the Cetirizine 10 mg QD dose.

For the instantaneous assessments, the studies gave somewhat conflicting results. Studies 81 and 61 appear to have shown very small differences in effect at the end of the dosing interval, whereas Study 32 demonstrated a 1.1 unit difference for the 180 mg dose. The results of the instantaneous assessments in Studies 32 and 61

[2] The sponsor did not specify a detailed algorithm for testing the interactions, as was done in the CIU studies.

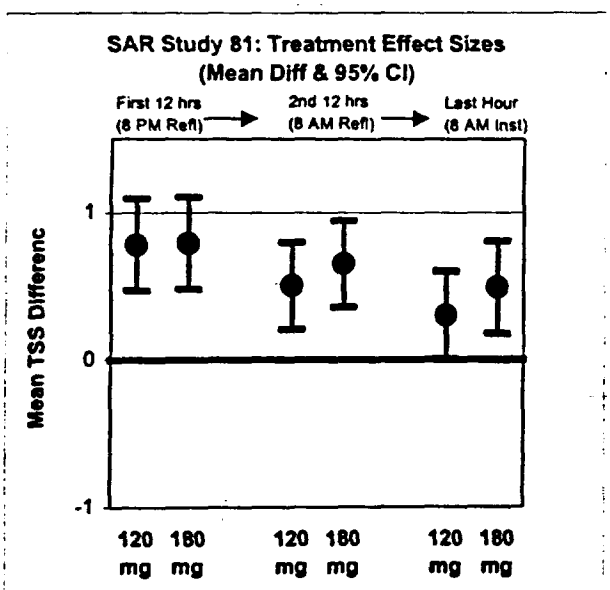
were complicated by a treatment-by-center interaction. In general, this interaction was present in the analyses of the instantaneous individual symptom scores as well. Most of the centers had about 10-30 patients, however, one center (Martin Stern) enrolled 165 patients in Study 32 and 206 patients in Study 61. As the sponsor did not submit the electronic data, this reviewer asked the sponsor to investigate the interaction in Study 32). The sponsor found that the interaction was "highly sensitive" to a relatively moderate-sized German center (n=24, Dr. Bostajovski). When this center was removed from the analysis, the interaction was no longer significant (p=0.2652) and the treatment effect differences for the instantaneous assessment were smaller (120 mg: 0.7 units; 180 mg: 0.9 units; Cetirizine: 1.0 units), submission June 21, 1999. All differences were still statistically significant. Note that a treatment-by-center interaction was also significant (p=0.0477) in the analysis of the instantaneous assessment in Study 61, but was not investigated.

In Figures 8 and 9 below, the treatment effects for the different assessments are presented in a time-line to better demonstrate the declining effects during the dosing period for Study 81, the US study. In this study, the effect size at the end of the dosing period for the 120 mg dose was about 38% of that during the first 12-hours of the dosing period (0.3 / 0.8 = 0.375). The effect size of the 180 mg dose at the end of the dosing period was about 63% of that during the first 12 hours. It is expected that the drug would not be as effective at the end of the dosing period. However, the effect sizes during all periods of the dosing interval should be considered when determining efficacy. (It should be noted that the differences in treatment effect could be due, in part, to the different times of day they reflect, i.e., daytime and nighttime.)

Figure 8: Study 81 Treatment Effects (Adjusted for Model Effects)

	1st 12-hrs of dosing period	2nd 12-hrs of dosing period	
120 mg	0.8	0.5	0.3
180 mg	0.8	0.7	0.5
			Last hr of dosing period

Figure 9: Study 81 Treatment Effect Sizes Over Dosing Period By Treatment Group



Weekly Analyses
The results of the analyses of Study 81 of each week are presented in the appendix, Table A-2. The analyses of each week demonstrated that Allegra had the greatest effect on symptoms during the first week. The treatment effects, if any, appeared to be less in the second week. The sponsor states that this phenomenon was "mainly due to improvement in placebo patients, narrowing the differences between treatment groups relative to those observed over Week 1." (Volume 1.1-P301) The purpose of including a placebo group in a study is to compare the mean responses on drug to the mean responses on placebo. The improvement of the placebo group during the second week is informative, in that it tells us that the changes from baseline on drug during the second week may not necessarily have been due to effects of the drug. If the scores during the second week were at the "floor" of the scale (i.e.: zero), the improvement in placebo would indeed diminish any possibility of the drug showing a treatment effect. However, the scores were not near zero. In fact, over half the patients were still experiencing allergy symptoms during the second week. Fifty-nine percent of the Allegra patients and 60% of the placebo patients were still symptomatic enough to be considered "allergy sufferers" in that they still met all inclusion criteria for entrance into the trial, see appendix Table A- 3.[3] Therefore, the lack of a treatment difference during the second week is potentially important.

The means of the daily symptom scores (using LOCF) graphed over time are provided in the appendix Figure A-1 for Study 81. The placebo patients appear to be improving during the second week, even after carrying the last observation forward for the small number of dropouts.

The sponsor's results of Study 32 by week are provided in the appendix Table A-4. The pattern seen in Study 81 of a decreasing treatment effect during the second week, due in part to an improving placebo group, was present in Study 32 as well.

Individual Symptom Analyses
Results of the individual symptom assessments are presented in appendix Table A-2 and Figure A-1. In general, the results showed that Allegra had the greatest effect on sneezing (0.12-0.16 units for difference between active doses and placebo in instantaneous scores), with minor effects on rhinorrhea (0.07-0.10 units), itchy, watery, red eyes (0.06-0.11 units), and itchy nose, mouth, throat and/or ears (0.05-0.12 units). In addition, during the first week, Allegra appeared to have a minor effect on congestion symptoms assessed over the first 12-hours of the dosing period.

Shift Table Analyses
Shift tables are another way of looking at the efficacy demonstrated in the study. Percentages of patients that shifted none, one or more categories were calculated to determine the rate of responses of the individual patients.

Table 10: Adult SAR Study 81 Shift Table

Study 81: Percent of Patients That Shifted Categories
Change from Baseline 2-Week Average Scores Rounded to Nearest Integer
TSS Divided into Categories of 4 units Each (ie: a change of 4 units = a shift of 1 category)

	Categories Shifted	8 AM Instantaneous			8 AM Reflective			8 PM Reflective		
		Placebo	120 mg	180 mg	Placebo	120 mg	180 mg	Placebo	120 mg	180 mg
TSS	-3	0	0	0	0	0	0	0	0	0
	-2	1	3	3	1	2	4	1	4	6
	-1	35	38	39	34	38	40	31	37	41
	0	54	55	51	53	53	49	53	53	43
	1	10	4	7	12	7	7	14	7	9
	2	0	0	0	1	0	0	0	0	0

As indicated in the mean scores, the shift table presentation also demonstrates the rather modest effect of the drug. The instantaneous symptom scores of 1% of the placebo group patients decreased two categories, whereas 3% of the 120 and 180 mg QD patients declined two categories. The instantaneous symptom scores of 10% of the placebo patients increased one category, whereas only 4% of the 120 mg QD group 7% of the

[3] The patients' scores from the last seven days of double-blind treatment were used in this calculation. Therefore, the analysis takes into account the small number of patients who did not complete the study.

180 mg group increased one category. The differences were greater for the reflective scores. In general, the greatest differences in percentages of patients with relief of symptoms were seen between the 180 mg QD and placebo groups. The results of these analyses for each of the individual symptoms are provided in the appendix, Table A-5. The results of the individual symptoms were similar to those of the TSS.

Health Outcomes Results
The sponsor would like to make claims about the benefits of Allegra on Quality of Life endpoints based on Study 81, however, data from this study were not supportive of these claims. Quality of Life was measured in this study, not as a secondary efficacy variable, in the strict sense, but as a primary variable in a "companion study". Investigators and patients in Study 81 were asked if they wanted to participate in this companion study by filling out three extra questionnaires at each visit. If the investigator agreed, then all patients at the investigative site were eligible. If the patient also agreed, then the patient was enrolled. All investigative sites and all patients agreed to participate in this "companion study".

Patients completed self-administered quality of life (QOL) questionnaires at each of the four visits. Generic QOL was measured using the Medical Outcomes Study 36-Item Short Form Health Survey (SF-36). Disease specific QOL was measured using the Rhinoconjunctivitis QOL Questionnaire (RQLQ). Performance impairment was evaluated using the Work Productivity and Activity Impairment questionnaire (WPAI). The Division of Drug Advertising, Marketing and Communication (DDMAC) recognizes the RQLQ as a validated instrument for determining QOL (see Ms. Fisher's review, June 15, 1999). The SF-36 and WPAI are not considered validated instruments by DDMAC. The sponsor described the QOL results in a separate study report from the main study report for safety and efficacy. In this separate study report, the sponsor states that the "primary outcome measure" was the change from baseline in the overall QOL score generated by the RQLQ over the 2-week double-blind treatment period. However, it is not a primary outcome measure, because it is not independent from the primary efficacy variable. A multiple comparisons adjustment should be made; however, it is unknown what adjustment is appropriate due to the Step-Down Procedure the sponsor used to protect the overall Type I error rate for the primary efficacy variable.

The 120 mg QD group reported statistically significantly greater improvement than the placebo group for 4 of the 7 domains in the RQLQ, while the 180 mg QD group reported statistically significantly greater improvement for all 7 domains. A treatment-by-investigator interaction was statistically significant at the 0.25 level in the primary analysis ($p=0.0395$).[4] The sponsor did not examine the source of the interaction, stating that it was of no concern because the treatment effect was statistically significant in the presence of the interaction.[5] It is this reviewer's opinion that the true treatment effect cannot be estimated in the presence of an interaction. The interaction demonstrates that the treatment differences were different at different sites, therefore, the sites cannot be analyzed together and no overall treatment effect can be calculated. The interaction potentially compromised the inferences of the QOL results.

Neither of the other two SAR studies (32 and 61) assessed QOL.

3.2.5 Special Populations

Summary: The treatment effect was examined across subgroups to assess consistency of response. In general, different treatment effects were seen across subgroups of age, baseline, and race, indicating that the treatment effect was not robust. Some of these interactions were of concern, others included too few patients in the subgroups to be conclusive. The most concerning subgroup interaction was baseline. In previous Allegra studies (submitted to the original NDA 20-625), the treatment effect in the primary endpoint (reflective scores) seemed to be pronounced among the high baseline patients. The opposite occurred in one of the current studies, Study 81. Among the high baseline patients, the placebo group was superior to the 180 mg QD group. The treatment effect for the end-of-dosing interval appeared to be carried by the low baseline patients (<7.4 TSS units at baseline). The sponsor did not provide baseline subgroup analyses for Studies 32 and 61. Since this reviewer did not have the electronic data for these studies, the baseline interaction seen in Study 81 could not be confirmed (or ruled out) in Studies 32 and 61.

In all the analyses of Study 81, the instantaneous score was used as the dependent variable; in Studies 32 and 61, the primary variable, 24-hr reflective score, was used as the dependent variable. All of the analyses presented below were performed by the sponsor using Dr. Edwards' data (in Study 81); this reviewer analyzed

[4] Treatment interactions in this review are tested at the conservative level of 0.25 for exploratory purposes. In cases where the interaction p-value is <0.10, the assessment of the overall treatment effect is difficult.
[5] Electronic data was not submitted for the QOL endpoints, therefore, this reviewer could not examine the problem in more detail.

the data excluding patients from Dr. Edward's center and the results were similar. The analyses of Studies 32 and 61 presented below were also performed by the sponsor. In some cases, the subgroups are defined differently for the different studies.

Study 81: The consistency of treatment effect on 8AM instantaneous TSS in Study 81 was evaluated in subgroups of patients defined by investigative site and five demographic or baseline characteristics: age, gender, race, body weight and level of baseline SAR symptoms. For the latter characteristics, patients were distributed into one of 2 categories, depending on whether their baseline 8 AM instantaneous TSS was less than or equal to ("low") or greater than ("high") the median baseline TSS of the intent-to-treat population (baseline TSS median: 7.4 units).

The analysis of **age category**-by-treatment interaction had a p-value of 0.0581, indicating different treatment effects for different age groups. This was driven primarily by a relatively weak response to treatment in patients 12 to 16 years of age in the 180 mg QD fexofenadine HCl group. Examination of treatment effect (difference with placebo) by age group reveals that the ≥40 years group benefited most from Allegra, and patients <16 years old the least.

Table 11: Study 81 Age Subgroups

Age		N	Change from Baseline	Difference From Placebo
< 16 years	Placebo	31	-0.96	
	120 mg QD	31	-1.01	**0.07**
	180 mg QD	37	-0.38	**-0.58**
16 - <40 years	Placebo	181	-1.02	
	120 mg QD	172	-1.23	**0.21**
	180 mg QD	157	-1.55	**0.53**
≥ 40 years	Placebo	80	-0.50	
	120 mg QD	84	-1.13	**0.63**
	180 mg QD	88	-1.44	**0.94**

Treatment effect varied significantly with the level of pretreatment symptoms (p=0.0256), with higher **baseline values** associated with greater responses. This was most striking in the placebo group, which had a mean change in TSS of only –0.28 in the "Low Baseline" group, compared to –1.50 in the "High Baseline" group. High baseline values were associated with greater response for all treatment groups. However, high baseline values were also associated with lower treatment differences. Patients with high baseline values did not benefit at all from 120 mg QD and benefited less from 180 mg QD than patients with low baseline values. In fact, among the high baseline patients, placebo was numerically superior to the 120 mg QD dose. It appears as though the drug did not alleviate symptoms (more than placebo) at the end of the dosing interval when the symptoms were severe. By the last hour of the dosing period, the treatment effect for the most severe symptoms was very small for the 180 mg QD (0.25 TSS units) and had disappeared entirely for the 120 mg QD group. The results of the same analysis on the AM reflective scores were similar but not as striking (p-value for treatment-by-baseline interaction term = 0.2007). The results of the PM scores did not show any difference in treatment differences between baseline subgroups (p=0.3497). These findings are consistent with the results of the primary analysis, that is, the treatment effect appears to be the most robust during the first 12-hours of the dosing interval and least during the last hour.

Table 12: Study 81 Baseline Subgroups Instantaneous TSS

Baseline		N	Change from Baseline	Difference From Placebo
Low (<7.4 units)	Placebo	156	-0.28	
	120 mg QD	140	-0.99	**0.71**
	180 mg QD	144	-0.98	**0.70**
High (>7.4 units)	Placebo	136	-1.50	
	120 mg QD	147	-1.36	**-0.14**
	180 mg QD	138	-1.75	**0.25**

Treatment effect varied significantly by race (interaction p=0.2113), with Caucasian patients associated with greater treatment effects. The placebo group was numerically superior to the 180 mg QD group among the

non-Caucasian patients. This treatment difference favoring placebo (-0.31) was almost as large as the treatment difference favoring Allegra between the placebo and 120 mg QD Caucasian groups (0.33 units). However, the non-Caucasian groups were too small to make conclusive inferences.

Table 13: Study 81 Race Subgroups Instantaneous TSS

Baseline		N	Change from Baseline	Difference From Placebo
Caucasian	Placebo	256	-0.82	
	120 mg QD	248	-1.15	**0.33**
	180 mg QD	257	-1.40	**0.58**
Other	Placebo	36	-1.23	
	120 mg QD	39	-1.32	**0.09**
	180 mg QD	25	-0.92	**-0.31**

There were no notable differences in treatment effects by **gender** or **weight** subgroups in Study 81.

Study 32: The consistency of treatment effect on 24-hour reflective TSS was evaluated in subgroups of patients defined by investigative site, country, age, gender and race. The statistically significant investigator-by-treatment interaction evident in the analysis of the instantaneous scores was not present in the analysis of the reflective scores. The p-values of all the interactions with treatment (site, country, age, gender and race) for the reflective scores were ≥0.3565. This indicates that the decreases in 24-hour reflective TSS observed across the four treatment groups were not significantly different among the subgroups defined by these factors.

Since this study included two different pollen seasons (spring in Europe, fall/winter in South Africa and Australia), it is interesting to compare the results across countries (see Table 14 below). The changes from baseline appeared to be similar across countries. The treatment effects were largest in Australia (n=18).

Table 14: Study 32 Responses Across Countries (and Pollen Seasons)

(The standard errors in this table range from 0.3-0.4 for all means except for those in Australia, which range from 1.1-1.3)

Season	Country	Treatment	N	Change from Baseline	Difference From Placebo
Spring 1995	United Kingdom	Placebo	56	-0.3	
		120 mg	58	-1.4	**1.1**
		180 mg	58	-1.9	**1.6**
		Cetirizine	59	-2.2	**1.9**
	France/Belgium	Placebo	44	-1.5	
		120 mg	43	-2.6	**1.1**
		180 mg	40	-2.2	**0.7**
		Cetirizine	44	-3.1	**1.6**
	Germany	Placebo	39	-2.5	
		120 mg	43	-3.3	**0.8**
		180 mg	42	-4.3	**1.8**
		Cetirizine	44	-3.6	**1.1**
Fall/Winter 1995-96	South Africa	Placebo	58	-2.0	
		120 mg	63	-2.9	**0.9**
		180 mg	57	-3.3	**1.3**
		Cetirizine	55	-2.7	**0.7**
	Australia	Placebo	4	-1.3	
		120 mg	4	-3.9	**2.6**
		180 mg	5	-2.8	**1.5**
		Cetirizine	5	-4.6	**3.3**

Study 61: The consistency of treatment effect on 24-hour reflective TSS was evaluated in subgroups of patients defined by investigative site, type of practice (allergy or general), country, region within countries, age, gender, and race.

The treatment-by-**type of practice** interaction was significant (p=0.0949). This was due, in part, to the absence of a mean placebo response among the allergy clinics.

Table 15: Study 61 Type of Practice Subgroups Reflective TSS

Baseline		N	Change from Baseline	Difference From Placebo
General Practice	Placebo	309	-0.8	
	80 mg QD	328	-1.7	**0.9**
	120 mg QD	328	-1.7	**0.9**
Allergy Clinic	Placebo	113	0.1	
	120 mg QD	110	-1.5	**1.6**
	180 mg QD	108	-1.5	**1.6**

Treatment effect for the 24-hour reflective score varied by **age category** in this study also. The patients <18 years old on Allegra had almost no benefit over that seen in the placebo patients.

Table 16: Study 61 Age Subgroups Reflective TSS

Age		N	Change from Baseline	Difference with Placebo
< 18 years	Placebo	39	-1.4	
	80 mg QD	51	-1.6	**0.2**
	120 mg QD	54	-1.4	**0.0**
18 - 39 years	Placebo	273	-0.5	
	80 mg QD	270	-1.7	**1.2**
	120 mg QD	272	-1.7	**1.2**
≥ 40 years	Placebo	110	-0.7	
	80 mg QD	117	-1.8	**1.1**
	120 mg QD	110	-1.8	**1.1**

Treatment effect for the 24-hour reflective score varied across **countries** also. Patients in France had almost no benefit from the 120 mg QD dose (0.2 TSS units). The 80 mg QD dose in France performed slightly better than the 120 mg QD dose. The opposite was true in Ireland. In the UK, the doses performed equally well.

Table 17: Study 61 Country Subgroups Reflective TSS

Country		N	Change from Baseline	Difference with Placebo
U.K.	Placebo	324	-0.4	
	80 mg QD	329	-1.6	**1.2**
	120 mg QD	322	-1.6	**1.2**
Ireland	Placebo	30	-0.7	
	80 mg QD	37	-2.0	**1.3**
	120 mg QD	40	-2.5	**1.8**
France	Placebo	68	-1.5	
	80 mg QD	72	-1.9	**0.4**
	120 mg QD	74	-1.7	**0.2**

Treatment effect varied across **race** as well (Caucasian and other). Among the non-Caucasian patients, the placebo group performed superior to the 120 mg QD group. The difference between the 80 mg QD group favored Allegra, but was small, at 0.5 units, half of the treatment effect seen with Caucasians. The sponsor did not submit electronic data for this study, therefore, the race interaction seen in Study 81 on the instantaneous scores could not be confirmed on the instantaneous scores in this study.

Table 18: Study 61 Race Subgroups Reflective TSS

Baseline		N	Change from Baseline	Difference with Placebo
Caucasian	Placebo	402	-0.6	
	80 mg QD	410	-1.7	**1.1**
	120 mg QD	411	-1.7	**1.1**
Other	Placebo	20	-1.4	
	80 mg QD	28	-1.9	**0.5**
	120 mg QD	25	-1.3	**-0.1**

There were no differences in treatment effects between genders, across individual sites, or across different regions within countries.

3.2.6 Safety

Safety evaluations included clinical laboratory panels, physical examinations, and adverse event reporting. The patients were not queried about adverse events, but spontaneously reported them on their diary cards.

There was no obvious relationship between treatment group and patients reporting at least one adverse event in any of the adult SAR studies (see Table 19 below).

Table 19: Adult SAR Studies Adverse Events

	Study 81	Study 32	Study 61
Placebo	88 (30%)	82 (39%)	151 (34%)
80 mg QD			141 (31%)
120 mg QD	86 (30%)	70 (33%)	149 (32%)
180 mg QD	86 (30%)	84 (40%)	
Cetirizine		92 (44%)	

The system organ classes within which most adverse events were reported in both treatment groups were the respiratory (Study 81) and central and peripheral nervous system (Studies 32 and 61). The adverse event reported most frequently by patients in each of the treatment groups in both studies was headache (Study 81: 28%; Study 32: 13%; Study 61: 12%). No dose-related trends in adverse events were seen.

3.3 Conclusions

The sponsor has demonstrated a small benefit of Allegra when administered once daily. The statistical significance of the treatment effects are indicative of a difference in response to treatments (between placebo and active doses), however, the differences are small. Using an identical primary endpoint TSS scale of 0-16, the original Allegra approval (twice daily administration) was based on results from one adequate and well-controlled study with a one unit difference (0.25 units on individual symptoms) between treatment groups for both reflective and instantaneous scores and three additional adequate and well-controlled studies with differences in the range of 0.3 - 0.7 units.

The current submission includes one adequate and well-controlled study (Study 81) with a statistically significant treatment effect of 0.5 units in instantaneous TSS and about 0.7 units in reflective TSS for the 180 mg QD dose. The 120 mg QD dose in this study demonstrated smaller treatment effects. In general, the treatment effect was largest during the first 12 hours of the dosing period, and during the first week. The results showed that Allegra had the greatest effect on sneezing (0.12-0.16 units for difference between active doses and placebo in instantaneous scores), with minor effects on rhinorrhea (0.07-0.10 units), itchy, watery, red eyes (0.06-0.11 units), and itchy nose, mouth, throat and/or ears (0.05-0.12 units).

As indicated by the mean scores, a shift table presentation of the results also demonstrated the rather modest effect of the drug. The instantaneous symptom scores of 1% of the placebo group patients decreased 2 categories, whereas 3% of the 120 and 180 mg QD patients shifted 2 categories.

Two additional studies (Studies 32 and 61) were based on different study designs and used capsules instead of tablets. Both studies had large discontinuation rates. The two studies had 100 patients in common. Study 61 demonstrated very small treatment differences with placebo for the instantaneous scores, was only 7-10 days in length and the inferences from this study were problematic due to a treatment-by-center interaction. The study reports for both of these studies (32 and 61) had numerous errors that the sponsor did not correct before submission, making the review of these studies difficult. The sponsor did not submit electronic data for these studies, severely limiting the statistical review of the results from these trials.

In general, in Studies 81 and 61, the treatment effect was not robust across different subgroups of patients.

The Quality of Life results in Study 81 do not demonstrate statistically significant differences between Allegra and placebo due to an investigator-by-treatment interaction. Further, the results were not independent of the efficacy endpoints, and, therefore, need to be adjusted for multiple comparisons.

4 Seasonal Allergic Rhinitis (SAR) Pediatric

4.1 Introduction

Summary

Studies 66 and 77 assessed the safety and efficacy of Allegra in the pediatric population and estimated the lowest effective dose for children. The two studies had identical designs. The studies did not recruit enough patients, therefore the sponsor contacted FDA with the request to combine studies. FDA agreed, but cautioned the sponsor that the combined study might yield statistically significant p-values for treatment differences that were not "clinically relevant". Study 66 did not demonstrate differences between treatment groups (the placebo group was numerically superior to the Allegra groups). Study 77, however, demonstrated (larger than expected) statistically significant treatment differences between Allegra and placebo (Allegra was superior). The results from the two studies were discordant due to a large placebo response (relative to Allegra mean responses) at the Canadian sites in Study 66. The reason for such a large response could not be explained by baseline, pollen counts, age, protocol violations, outliers, or possibility of a drug supply packaging error. It is highly questionable as to whether or not these studies can be combined for analysis.

In addition to the inconsistent effects observed across studies, the treatment effects within studies appeared to be inconsistent. Further, the responses of the different dose groups were not dose-ordered in either study. The overall conclusion of efficacy based on the two studies is difficult to make due to these inconsistencies.

Label

Redacted 1

pages of trade secret and/or confidential commercial information

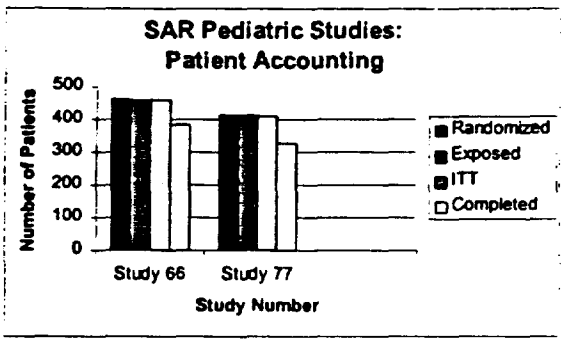
4.2 Results

4.2.1 Study Conduct

One-hundred percent and 99.3% of patients exposed to double-blind study medication were included in the ITT analyses of Studies 66 and 77, respectively.

One patient was randomized into both studies at two different centers (Study 66:852-0005, 15 mg BID; Study 77: 917-0009, 60 mg BID) and completed 14 days of double-blind treatment in both studies. The patient entered Study 77 seven days after completing Study 66. This patient was counted twice in the pooled analysis. The sponsor performed analyses excluding this patient and the results were similar.

Figure 11: Numbers of Patients in SAR Pediatric Studies



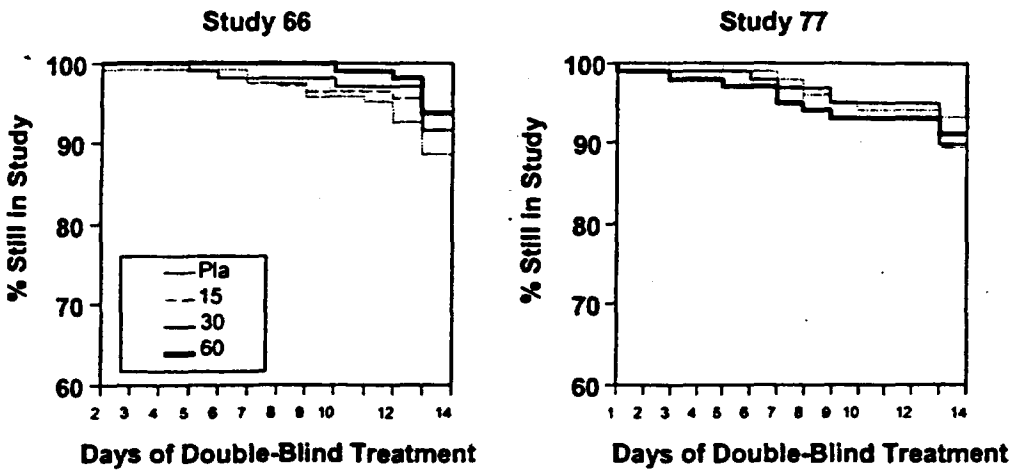
Dropouts: The overall discontinuation rates (based on the exposed population) were similar across studies. However, the rates in Study 66 were different across treatment groups, whereas they were similar across treatment groups in Study 77.

Table 21: Pediatric SAR Studies' Discontinuation Rates

	Placebo	15 mg	30 mg	60 mg	Total
Study 66	8 (6.5%)	5 (4.2%)	2 (1.9%)	1 (0.9%)	16 (3.5%)
Study 77	5 (4.8%)	5 (4.7%)	5 (5.0%)	4 (3.9%)	19 (4.6%)

This disparity warranted further examination of the dropout rates in light of the different results from the two studies. Kaplan-Meier curves of the two studies (Figure 12, below) depict the rates of withdrawal between treatment groups in Study 66. The withdrawal of patients in Study 77 occurred earlier than in Study 66. The 60 mg groups and the placebo groups appeared to have slightly different patterns of withdrawal in the two different studies. These small differences in withdrawal rates may have contributed to the differences in results of the primary endpoint between studies. The sponsor investigated the observed rates of protocol violations between the two studies and did not find any significant differences. It is not clear if the sponsor investigated dropout rate and dropout time within treatment groups between the two studies.

Figure 12: Discontinuation Rates Across Treatment Groups



4.2.2 Sample Size

With 150 patients per treatment group and an alpha-level of 0.05, both studies had 80% power to detect a difference in mean change in 7 PM reflective score of 0.72 units assuming a standard deviation of 2.2 units. The studies could not recruit enough patients to satisfy their planned sample sizes, therefore, the sponsor, prior to breaking study blinds, contacted FDA with the request to combine the results of the two studies. FDA agreed, but cautioned that the combined study might yield statistically significant p-values for treatment differences that were not "clinically relevant". However, the results of the two trials were not consistent. Study 66 demonstrated a numerically superior placebo response in comparison with the Allegra dose groups, and Study 77 demonstrated statistically significant treatment differences larger than the 0.72 units the studies had originally been powered to detect.

4.2.3 Demographics

The treatment groups in both studies were well balanced for demographic factors, with no statistically significant differences in gender, age, race, body weight or height. The ITT population was predominantly male in both studies (59%), and overwhelmingly Caucasian (86%). The patients ranged in age from 5-12, with an average age in both studies of 9 years. Thirty-two 5- and 6-year old patients (18 in Study 66 and 14 in Study 77) were exposed to Allegra at doses greater than 15 mg BID. Thirty-one of these patients completed at least 14 days of double-blind treatment. One patient discontinued on the third day.

Study 66 had a greater percentage of Canadians than did Study 77. As will be seen, the mean changes from baseline at the Canadian sites were greater than those at the US sites in both studies, regardless of treatment group.

Table 22: Demographics Across Treatment Groups and Studies

Study	Demographic Factor	Placebo	15 mg	30 mg	60 mg	Total
66		N=119	N=113	N=105	N=107	N=444
	Mean Age	9.2	9.0	9.0	9.1	9.1
	Mean Weight	35.8	36.2	34.5	35.3	35.5
	Mean Height	138.9	137.4	136.4	137.6	137.6
	Mean TSS baseline	8.3	8.0	8.0	7.8	8.0
	% Males	60	63	62	50	59
	% Caucasian	83	89	86	86	86
	% US	83	85	85	86	85
77		N=105	N=105	N=100	N=101	N=411
	Mean Age	9.3	9.3	9.2	9.0	9.2
	Mean Weight	37.4	36.9	35.7	33.4	35.9
	Mean Height	139.0	139.8	138.8	136.9	138.7
	Mean TSS baseline	7.9	7.5	7.6	7.7	7.7
	% Males	62	63	56	55	59
	% Caucasian	81	87	87	88	86
	% US	90	90	94	91	91

4.2.4 Primary & Secondary Efficacy Variables

The primary endpoint was change from baseline in the 7 PM reflective scores recorded by the caregivers. Table 23 describes the study results for the baseline, 2-week average and change from baseline scores for the PM reflective scores and the AM instantaneous. The results of Study 81 are provided for comparison.

Table 23: SAR Pediatric Studies Descriptive Statistics

PM Reflective Assessments

			Baseline				2-Week Average				Change from Baseline			
Study	Dose	N	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
66	0	119	8.3	2.5	3.5	14.2	6.7	3.2	1.2	14.2	-1.5	2.6	-9.1	6.5
	15	113	8.0	2.6	2.2	15.7	6.8	3.1	0.3	15.5	-1.1	2.3	-6.2	8.2
	30	105	8.0	2.4	2.2	15.3	6.5	3.0	0.4	13.4	-1.5	2.7	-7.8	7.7
	60	107	7.8	2.4	4.0	13.5	6.5	3.1	1.0	13.9	-1.3	2.3	-7.7	4.7
77	0	105	7.9	2.5	1.8	14.8	7.0	2.5	1.9	13.3	-0.8	2.6	-9.7	4.2
	15	105	7.5	2.5	2.0	15.2	5.9	2.8	0.8	14.2	-1.6	2.2	-8.1	4.4
	30	100	7.6	2.3	4.2	14.2	6.2	3.0	0.6	13.9	-1.4	2.3	-8.8	5.2
	60	101	7.7	2.3	4.0	15.2	6.1	2.8	0.3	12.5	-1.6	2.9	-11.4	4.2
81	0	285	7.4	2.1	1.8	12.0	6.7	2.2	0.3	12.0	-0.6	2.1	-6.2	8.2
	120	279	7.4	2.1	0.0	12.0	5.9	2.4	0.2	11.2	-1.4	2.0	-7.0	3.4
	180	273	7.4	2.1	0.0	12.6	6.0	2.5	0.3	12.8	-1.4	2.4	-8.2	5.8

AM Instantaneous Assessments

			Baseline				2-Week Average				Change from Baseline			
Study	Dose	N	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
66	0	119	7.3	3.1	1.7	13.4	6.4	3.1	0.7	14.0	-1.0	2.6	-8.7	7.0
	15	113	7.3	3.2	1.5	16.0	6.4	3.3	0.2	15.6	-0.8	2.5	-7.7	7.9
	30	104	7.1	2.8	2.1	13.5	6.3	3.0	0.3	13.5	-0.8	2.8	-10.2	7.7
	60	105	7.0	3.1	1.6	14.3	6.3	3.3	0.6	13.4	-0.7	2.1	-8.8	5.1
77	0	105	6.9	2.9	1.0	14.9	6.6	2.6	1.0	13.3	-0.3	2.4	-8.5	5.6
	15	105	6.4	3.0	0.3	14.8	5.4	2.7	0.0	12.8	-0.9	2.3	-6.5	4.8
	30	99	6.5	2.5	2.0	14.0	5.8	3.0	0.0	14.0	-0.8	2.2	-5.5	8.6
	60	101	6.7	2.9	0.7	15.6	5.7	2.7	0.3	13.1	-0.9	2.7	-11.3	7.2
81	0	285	7.6	1.8	4.4	12.0	6.8	2.2	0.3	12.1	-0.8	2.0	-6.5	5.3
	120	279	7.8	1.7	4.6	12.0	6.5	2.4	0.7	12.0	-1.2	1.8	-6.9	3.6
	180	273	7.7	1.8	4.0	12.0	6.4	2.4	1.0	13.2	-1.4	2.1	-8.6	3.5

Recall that the baseline requirements and the symptom score scale in the pediatric studies were different from those in the adult SAR study. The highest score in the pediatric studies (4: very severe) had a somewhat milder definition "symptom prevented normal daily activity or sleep" than did the score of "4" in the adult studies, "symptom is so severe as to warrant an immediate visit to the physician". To examine if these differences resulted in differences in the change from baseline scores, the means and standard deviations of the pediatric data were compared to the adult SAR data. It appears as though the standard deviations of the change from baseline scores were slightly greater in the pediatric studies. Among all of the active dose groups, the mean changes in the AM instantaneous assessment were similar across the two pediatric studies, and smaller than those seen in the adult SAR study. In this assessment, the placebo group in Study 66 had a larger response than any of the active dose groups in the two pediatric studies.

The distribution of scores during the baseline and treatment periods is provided in Table 24 below. In the pediatric studies, approximately 4-8% of the scores were coded "very severe" (=4). Only 0.2-0.3% of the scores in the adult study equaled 4. This is primarily due to the ineligibility of patients with any assessment equal to "very severe" during the baseline period in the adult study. (Thus, the adult scale is essentially a 4-point scale while the pediatric is a 5-point scale.) The frequency of instantaneous scores coded "None" (= 0) appeared to be greater in the pediatric studies than in the adult study.

Table 24: Distribution of Scores for Each Symptom Across Pediatric and Adult SAR Studies

Instantaneous AM

Study		Sneezing		Rhinorrhea		Itchy Eyes		Itchy E/N/T	
		Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
66	0	2996	32.1	1320	14.1	2355	25.2	1741	18.6
	1	2156	23.1	1843	19.7	2452	26.2	2249	24.1
	2	2262	24.2	2983	31.9	2585	27.7	2982	31.9
	3	1457	15.6	2417	25.9	1519	16.3	1839	19.7
	4	475	5.1	785	8.4	434	4.6	536	5.7
77	0	3064	36.1	1373	16.2	2577	30.4	1783	21.0
	1	1837	21.6	1668	19.7	2308	27.2	2088	24.6
	2	2045	24.1	2815	33.3	2115	24.9	2667	31.4
	3	1224	14.4	1977	23.4	1076	12.7	1547	18.2
	4	315	3.7	631	7.5	408	4.8	401	4.7
81	0	3717	20.9	1203	6.8	1979	11.1	1737	9.8
	1	4951	27.9	3673	20.7	4746	26.7	4539	25.6
	2	6102	34.4	7372	41.5	6942	39.1	7397	41.7
	3	2943	16.6	5447	30.7	4044	22.8	4038	22.8
	4	36	0.2	54	0.3	38	0.2	38	0.2

Reflective PM

Study		Sneezing		Rhinorrhea		Itchy Eyes		Itchy E/N/T	
		Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
66	0	1866	19.7	1026	10.9	2073	21.9	1481	15.7
	1	2472	26.2	1881	19.9	2370	25.1	2054	21.7
	2	2910	30.8	3231	34.2	2763	29.2	3287	34.8
	3	1756	18.6	2587	27.4	1741	18.4	2065	21.9
	4	446	4.7	724	7.7	502	5.3	563	6.0
77	0	1831	21.5	1101	13.0	2227	26.2	1421	16.7
	1	2085	24.5	1569	18.5	2232	26.2	1912	22.5
	2	2763	32.5	3019	35.6	2336	27.5	2932	34.5
	3	1458	17.1	2186	25.7	1273	15.0	1801	21.2
	4	370	4.3	615	7.2	436	5.1	441	5.2
81	0	3000	17.8	1642	9.7	2373	14.1	2014	11.9
	1	5475	32.5	4588	27.2	4975	29.5	4832	28.7
	2	5818	34.5	6613	39.2	6182	36.7	6581	39.0
	3	2531	15.0	3961	23.5	3285	19.5	3394	20.1
	4	30	0.2	50	0.3	39	0.2	33	0.2

To investigate differences, the mean baseline scores and changes from baseline for the primary efficacy variable (7 PM reflective TSS) pooled, by study and by country are presented in Table 25 below.

Table 25: Mean changes in 7 PM 12-hour reflective TSS: US versus Canadian sites*

	Pooled Protocol		*Protocol 066*		*Protocol 077*	
	Mean Baseline (n)	*Mean Change From Baseline*	*Mean Baseline (n)*	*Mean Change From Baseline*	*Mean Baseline (n)*	*Mean Change From Baseline*
All Study Sites						
Placebo	8.04 (229)	-1.21	8.19 (124)	-1.51	7.87 (105)	-0.85
Fex 15 mg	7.72 (223)	-1.38	7.90 (118)	-1.17	7.50 (105)	-1.62
Fex 30 mg	7.78 (208)	-1.42	7.96 (108)	-1.43	7.58 (100)	-1.41
Fex 60 mg	7.71 (212)	-1.42	7.75 (111)	-1.26	7.66 (101)	-1.59
Canadian Sites Only						
Placebo	7.20 (30)	**-2.14**	7.19 (20)	**-2.65**	7.22 (10)	**-1.12**
Fex 15 mg	7.96 (28)	-1.92	7.96 (17)	-1.63	7.96 (11)	-2.36
Fex 30 mg	7.95 (22)	-1.33	7.61 (16)	-1.15	8.86 (6)	-1.82
Fex 60 mg	7.38 (24)	-1.45	6.84 (15)	-0.96	8.29 (9)	-2.27
US Sites Only						
Placebo	8.17 (199)	**-1.07**	8.38 (104)	**-1.29**	7.94 (95)	**-0.82**
Fex 15 mg	7.68 (195)	-1.31	7.90 (101)	-1.09	7.45 (94)	-1.54
Fex 30 mg	7.76 (186)	-1.43	8.03 (92)	-1.48	7.50 (94)	-1.38
Fex 60 mg	7.75 (188)	-1.42	7.89 (96)	-1.31	7.60 (92)	-1.53
* Raw means.						

The results of the raw means demonstrate a difference in responses across studies and across countries. The placebo group in Study 77 had a small response (-0.85 units) relative to the active dose groups (15 mg: -1.62; 30 mg: -1.41; 60 mg :-1.59), whereas the placebo group in Study 66 had a large response (-1.51 units) relative to the active dose groups (15 mg: -1.17; 30 mg: -1.43; 60 mg :-1.26). The baseline means were different across

treatment groups in Study 66, with the placebo group having the greatest baseline mean. This baseline imbalance at first glance may appear to explain the large responses, however, the high baseline scores were associated with the US sites in Study 66. The placebo patients at the Canadian sites had the lowest baseline mean and the greatest change from baseline of all the treatment groups. Thus, the baseline imbalance did not explain the large placebo response in Study 66.

The results from the Canadian sites appeared to be different from the US sites. The mean changes from baseline in the Canadian sites were larger (in general) across all treatment groups, than those in the US sites.[6] Recall that Study 66 had a greater percentage of Canadian patients than did Study 77.

As noted above, prior to unblinding, the sponsor proposed to pool the studies due to smaller than projected enrollment. However, the results from Studies 66 and 77 were very different. In general, decisions about pooling studies are controversial. The sponsor concluded that the studies were "poolable", referencing a special issue of *Statistics in Medicine* devoted to meta-analysis. The sponsor stated that, "when differences cannot be identified statistically through careful examination, then pooling of the studies is still valid..." Upon request by this reviewer, the sponsor performed an analysis on the data from both studies using the primary analysis dependent variable (7 PM reflective TSS change from baseline), baseline TSS as a covariate, and treatment, study and a treatment-by-study interaction as factors. The results demonstrated a statistically significant (at the 0.25 level) treatment-by-study interaction (p=0.0432).[7]

It is highly questionable whether or not it is proper to combine these studies because of the observed treatment-by-study interaction. The sponsor's argument to pool the studies assumes all possible differences in population or study conduct can be investigated and identified. There are always elements of a trial that are not measured or recorded. In this case, for some unknown reason, the two study populations behaved differently. The observed statistical difference on the primary endpoint, with no discernable explanation, is enough to demonstrate that the two studies were, indeed, different, and should not be pooled.

The results of the primary analysis are presented in Table 26 below. For comparison, the results are presented for the pooled data, each study and the US sites separately.

Table 26: Pediatric Studies Least Squares Means, Treatment Differences & P-Values

7 PM Reflective

All Models included Baseline, Treatment and Site as factors (unless otherwise stated)

Positive differences indicate Allegra superiority.

	Pooled			Study 66			Study 77		
	LS Mean Change	Diff with placebo	p-value	LS Mean Change	Diff with placebo	p-value	LS Mean Change	Diff with placebo	p-value
All Sites									
Placebo	-1.21			-1.59			-0.84		
15 mg	-1.49	**0.28**	0.2197	-1.30	**-0.23**	0.3559	-1.83	**0.99**	0.0023
30 mg	-1.54	**0.33**	0.1585	-1.53	**-0.06**	0.8470	-1.65	**0.81**	0.0138
60 mg	-1.55	**0.34**	0.1416	-1.44	**-0.15**	0.6442	-1.73	**0.89**	0.0064
US Only									
Placebo	-0.98			-1.36			-0.65		
15 mg	-1.42	**0.44**	0.0637*	-1.26	**-0.10**	0.7619	-1.69	**1.04**	0.0026
30 mg	-1.55	**0.57**	0.0178*	-1.63	**0.27**	0.4441	-1.52	**0.87**	0.0112
60 mg	-1.53	**0.55**	0.0221*	-1.47	**0.11**	0.7577	-1.61	**0.96**	0.0057

* Model included Baseline-by-treatment interaction.

The sponsor's primary analysis (pre-specified) was an analysis of covariance on the change from baseline TSS with baseline TSS as a covariate and the treatment group and investigative site as factors. The sponsor tested the treatment-by-baseline and treatment-by-center interaction terms at the 0.10 level. If the treatment-by-baseline term was significant at the 0.10 level, the sponsor included it in the final model and calculated the treatment effects using the average baseline scores. The sponsor used the Step-Down procedure to protect the overall Type I error rate. All tests were conducted at the alpha-level of 0.05.

[6] It should be noted that Allegra was approved OTC in Canada before the studies began, (S8-V1.225-p321).
[7] Submission dated November 19, 1998.

In the sponsor's pre-specified pooled analysis the active dose groups were numerically greater than placebo, but not statistically superior. The differences were less than half of the 0.72 unit difference the individual studies were originally powered to detect. However, as noted above, this reviewer believes that it is not appropriate to pool these studies.

Examining the results of the studies separately demonstrates the inconsistency of results. In Study 66, the placebo group had a greater response than did any of the active treatments. These numerical differences were not statistically significant. In contrast, in Study 77, the results demonstrated statistically significant differences, favoring Allegra, for all dose groups, with the greatest effect seen for the 15 mg dose. The 30 mg group had the smallest treatment effect. The superior response of the 15 mg dose group was consistent across all secondary TSS assessments (7 AM reflective and 7 AM and PM instantaneous) and most individual symptom scores, see Tables A-6 - A-9 in the appendix.

4.2.5 Analyses to Determine Differences Between Studies

The treatment-by-study interaction effect in the pooled dataset, Study 66/77, is indicative of a difference between the two studies (p=0.0432). When this reviewer tested the interaction at a conservative alpha-level=0.25, it was robust to changes in the patient population (with and without patients from the Thomas Edwards' site, with and without patients with low or high baseline scores), time point (weeks 1, 2 or both) and model assumptions (parametric or non-parametric). The differences in results could be caused by one or more factors, including demographic differences, baseline severity differences, or differences in study conduct. The sponsor performed numerous analyses to try to identify and/or examine differences between the studies. The sponsor submitted a summary table of these analyses (provided in the appendix, Table A-10). The objectives of these analyses were:

- To identify factors which could explain differing levels of treatment effect
- To identify unbalanced baseline characteristics across treatment groups and across studies
- To determine if age accounts for some of the large placebo effect in Study 66 and/or different efficacy results in Study 66 and Study 77
- To determine if level of baseline symptoms accounts for some of the large placebo effect in Study 066 and/or different efficacy results in Study 66 and Study 77
- To compare the randomization rate and assess its correlation with treatment difference
- To assess the correlation between level of pollen count and treatment
- To compare the treatment effect adjusted for all baseline characteristics and randomization failure rate
- To examine whether level of protocol violations differed between Study 66 and Study 77
- To evaluate the impact of a site which reported an unusually large placebo response (mean reduction from baseline =-6.1)
- To moderate the impact of outliers and potential wrong model assumption
- To determine if study drug was packaged correctly
- To examine if treatment effect was consistent between Canadian and US study sites

The sponsor reported the following:

> "For Canadian sites, placebo effect was twice of that for fexofenadine in Protocol 066 (placebo: -2.65, Fex: –1.63, -1.15, and –0.96 for 15, 30 and 60 mg, respectively). For the US sites, although the placebo effect in Protocol 066 was still relatively larger than expected, fexofenadine effect in Protocol 066 was only slightly smaller than that in Protocol 077. Without Canadian sites, fexofenadine 30 mg and 60 mg were statistically significantly and 15 mg was marginally superior to placebo (p=.0814, .0276 and .0364 for 15, 30 and 60 mg respectively) in the pooled dataset. It is clear that the overall treatment comparison was largely weakened by the large placebo effect in Canadian sites of Protocol 066."

Further, the sponsor found that the older patients had lower baseline values, which corresponded to smaller changes from baseline. The sponsor examined the age distribution across treatment groups in both studies and found that age was well balanced for the four treatment groups for the Study 66 Canadian sites. In fact, the baseline TSS was actually lower for the placebo group in these sites than in the 15 and 30 mg dose groups. Since lower baseline values were associated with smaller changes, baseline TSS did not account for the large placebo effect observed in the Study 66 Canadian sites.

4.2.6 Special Populations

Summary: The treatment effect was examined across subgroups to assess consistency (see Table 27). For both studies, the consistency of treatment effect across subgroups of patients was assessed using change from baseline 7 PM reflective TSS, the primary efficacy variable. These subgroup analyses were conducted on the primary efficacy endpoint using an ANCOVA model with treatment and subgroup factor as factors, and baseline TSS as a covariate. An interaction between treatment and the subgroup was included in these models as well. Factors used in the subgroup analyses were baseline, age, weight, height, gender, and country. Treatment interactions were tested at the conservative alpha=0.25 level for exploratory purposes.

The only interaction of note was an unusually strong statistically significant gender interaction. This interaction was present in both studies, however it did not manifest itself similarly across the two studies. The mean, weight, height and age across gender and dose subgroups were similar and did not help to explain the disparate results across genders. These analyses demonstrate the inconsistency of results from these two studies, complicating interpretation and the overall conclusions of efficacy in the pediatric population.

Table 27: Results of ANCOVAs with Interaction Terms*

	Study 66 (excludes Edwards' data) p-values		Study 77 p-values	
Variable	**Main Factor (without interaction)**	**Main Factor / Interaction**	**Main Factor (without interaction)**	**Main Factor / Interaction**
Baseline	0.0001 (more severe pts had greater changes)	0.0001 / 0.5530	0.0001 (more severe pts had greater changes)	0.0001 / 0.0019 (treatment effect greater & more robust among patients with lower baseline scores)
Age (yrs)	0.3640	0.4169 / 0.2667	0.0838 (younger pts had greater changes from baseline)	0.0637 / 0.7108
Weight (kg)	0.8053	0.8769 / 0.8683	0.8595	0.9036 / 0.3241
Height (kg)	0.9166	0.8768 / 0.7686	0.2190	0.1662 / 0.1325 (taller pts had positive dose-response; opposite with shorter pts)
Gender (see Fig 13)	0.3946 (males & females had similar overall responses)	0.4722 / 0.0111 (females had positive dose-response; placebo outperformed all dose groups among males)	0.2689 (males & females had similar overall responses)	0.2591 / 0.0351 (females had negative dose-response; males had positive dose response)
Country	0.1111	0.1698 / 0.1030 (placebo patients outperformed the 15 mg dose in US whereas placebo outperformed all the dose groups in Canada)	0.2700	0.3438 / 0.9474

* The model used to test the interactions included change from baseline 7 PM reflective TSS as the dependent variable and baseline TSS as a covariate and treatment as a factor. Investigative site was not included because it was not independent of weight or height. The mean weight within the sites ranged from 23 kg (site #873) to 49 kg (site #906). The mean height within the sites ranged from 125 cm (site #873) to 147 cm (site #919). Analyses of variance (using weight, height or age as dependent variables) demonstrated statistically or marginally statistically significant differences across sites (p-value of overall F-tests: weight p=0.0336; height p=0.0503; age p=0.0804).

Baseline
In both studies, the main baseline factor was significant, and revealed that more severe patients demonstrated greater changes from baseline. Similar to Study 81, a treatment-by-baseline interaction in Study 77 showed results different from those seen in the original Allegra studies submitted to NDA 20-625. The patients with lower baseline scores appeared to receive the most benefit from Allegra (as compared to placebo).

Age, Weight, Height
The main factor age was marginally significant in Study 77. Recall that the sponsor found that younger patients also had higher baselines. The marginally significant factor of age in a model with baseline indicates that even after adjusting for baseline, younger patients had greater changes from baseline.

Weight was not related to change from baseline or differences in change from baseline in either study.

The treatment-by-height interaction was significant at the 0.25 level (p=0.1325) in Study 77. The interaction was due to a difference in dose-response. The taller patients had a positive dose response, whereas the shorter patients had a negative dose response.

Gender
As stated above, there were significant treatment-by-gender interactions in both studies (Study 66: p=0.0111; Study 77:0.0351) that could not be explained by demographic factors (see Table 28). The mean weight, height and age appeared to be similar across subgroups.

In Study 77, the male patients demonstrated an increase in treatment effect with increasing dose (Figure 13). Among the female patients, those who received 15 mg dose demonstrated a large response, (-2.77 units) as compared to the 30 and 60 mg dose groups (-0.50 and -0.25 units, respectively). The mean response of the 15 mg dose group females was not affected by outliers. The demographic characteristics of the female patients in the 15 mg dose group are listed in Table 29. The patients with greater changes from baseline do not appear to be different (in terms of weight, height, age or baseline TSS) from the patients with small (or positive) changes from baseline.

The mean changes from baseline in Study 66 also differed across genders. The females demonstrated a dose-response between the lowest dose and the upper two doses. The 15 mg dose, the dose that showed the greatest response among the females in Study 77, showed no treatment effect (as compared to placebo) in Study 66. The upper two doses demonstrated differences with placebo comparable to those seen in Study 81, the adult SAR study, (30 mg: 0.90 units, p=0.1098; 60 mg: 0.86 units, p=0.0956). In contrast, the males in Study 66 in the placebo group demonstrated statistically significant and marginally statistically significant benefit over the Allegra dose groups (15 mg: p=0.0803, 30 mg: p=0.0327; 60 mg: 0.0603). These p-values are based on post-hoc analyses of a subset of patients, therefore they should be viewed with caution, as descriptive results illustrating potential differences in results between genders.

Figure 13: Inconsistent Gender-by-Treatment Interactions

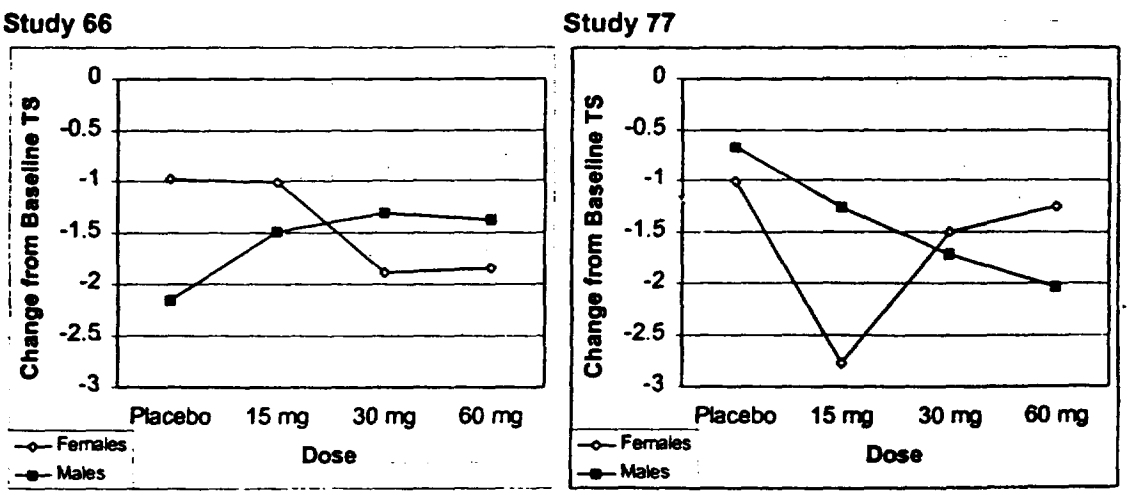


Table 28 :Mean Values (Age, Weight, Height, Baseline TSS & Change TSS) Across Dose, Gender and Study

Dose	Gender	Study	n	Age (yrs)	Weight (kg)	Height (cm)	Baseline TSS	Change
Placebo	F	66	48	9.4	37.4	139.7	8.75	-0.85
Placebo	F	77	40	9.1	35.8	137.4	7.99	-1.25
Placebo	M	66	71	9.0	34.8	138.4	7.92	-1.95
Placebo	M	77	65	9.4	38.3	140.0	7.80	-0.60
15 mg	F	66	42	9.1	36.1	138.8	7.91	-0.81
15 mg	F	77	39	9.4	37.1	140.5	7.57	-2.41
15 mg	M	66	71	8.9	36.3	136.6	8.04	-1.35
15 mg	M	77	66	9.2	36.8	139.5	7.47	-1.16
30 mg	F	66	40	8.8	32.8	134.5	8.44	-1.98
30 mg	F	77	44	9.2	35.8	139.7	7.44	-1.28
30 mg	M	66	65	9.2	35.5	137.6	7.70	-1.18
30 mg	M	77	56	9.2	35.6	138.0	7.69	-1.51
60 mg	F	66	53	8.8	32.9	135.6	7.85	-1.55
60 mg	F	77	45	8.9	32.2	135.7	7.73	-1.29
60 mg	M	66	54	9.3	37.7	139.5	7.71	-1.06
60 mg	M	77	56	9.0	34.4	137.8	7.61	-1.84

Note: the mean change from baseline of the 15 mg dose group females, highlighted above, was a major component of the treatment-by-gender interaction.

Table 29: Study 77 Females who received 15 mg dose

Patients with changes ≤ –4.0 units are highlighted.

Patient	Country	Center	Weight (kg)	Height (cm)	Age (yrs)	Baseline TSS	Change
0898.0003	**USA**	**898**	**52**	**155**	**11**	**9.8**	**-6.3**
0899.0005	USA	899	25	119	7	2.0	-0.1
0900.0005	USA	900	44	149	11	6.7	-1.1
0900.0015	USA	900	35	147	9	7.5	-1.6
0900.0025	**USA**	**900**	**41**	**143**	**9**	**9.2**	**-8.1**
0900.0027	USA	900	20	122	7	6.8	-0.9
0901.0015	USA	901	44	152	11	7.5	-0.7
0902.0004	USA	902	21	113	7	5.5	-2.5
0902.0013	USA	902	29	133	11	9.8	-3.0
0902.0023	USA	902	34	142	9	4.7	-2.8
0903.0001	USA	903	59	165	11	8.6	0.2
0904.0016	USA	904	19	118	7	5.4	-2.7
0905.0004	**USA**	**905**	**21**	**119**	**6**	**5.7**	**-4.7**
0905.0016	USA	905	41	149	10	9.8	-3.5
0907.0006	USA	907	23	127	8	6.7	-3.0
0907.0020	USA	907	65	160	11	10.5	0.2
0907.0033	USA	907	56	155	11	6.7	-2.9
0908.0039	USA	908	43	135	9	5.5	-3.0
0911.0006	**USA**	**911**	**23**	**117**	**8**	**12.0**	**-4.9**
0911.0012	USA	911	35	137	7	9.3	1.3
0911.0015	USA	911	25	125	8	5.0	-2.8
0911.0016	USA	911	43	151	12	10.6	-3.5
0912.0004	USA	912	40	142	10	4.8	4.4
0913.0018	USA	913	35	141	10	6.2	0.0
0914.0025	USA	914	25	127	9	7.8	-3.7
0914.0026	USA	914	43	149	10	7.8	-2.9
0915.0001	USA	915	32	152	8	10.7	-1.8
0915.0019	USA	915	42	149	10	10.7	-3.8
0916.0026	USA	916	28	136	9	9.0	-2.8
0919.0012	USA	919	42	147	11	8.5	-3.8
0921.0004	USA	921	35	147	11	6.0	-1.9
0922.0005	USA	922	40	150	10	7.0	-1.9
0922.0011	USA	922	45	155	11	7.3	-3.5
0922.0012	USA	922	54	155	10	9.0	-2.1
0923.0008	CAN	923	47	155	11	11.2	-2.0
0925.0014	CAN	925	25	133	10	4.5	-0.3
0925.0027	**CAN**	**925**	**35**	**125**	**5**	**7.0**	**-4.8**
0925.0028	**CAN**	**925**	**31**	**133**	**10**	**6.0**	**-5.1**
0926.0003	CAN	926	49	149	10	6.7	-1.7
		Mean:	37.13	140.46	9.36	7.57	-2.41

Country
The main effect for country was not significant. However, the overall mean changes from baseline appear to be slightly larger for the Canadian sites (Study 66: -1.7; Study 77: -1.9) than the US sites (Study 66: -1.3; Study 77: -1.3). The averages age, weight and height of the Canadian patients appeared to be similar to those of the US patients in both studies. Average baseline TSS was 7.4 and 8.0 units in the Canadian sites in Studies 66 and 77, respectively, and 8.1 and 7.6 units for the US sites in Studies 66 and 77, respectively. None of the demographic or baseline factors explained the differences in mean TSS change from baseline seen between the Canadian and US sites.

The treatment-by-country interaction was significant (at the 0.25 level) in Study 66 (p=0.1030). As seen in Table 25 in Section 4.2.4 above, the interaction was due in part to the large placebo response among the Canadian sites (-2.65 units). The sponsor investigated this large response (relative to the Allegra groups) thoroughly and could not explain it (using any variables collected in the study).

4.2.7 Safety

Safety evaluations included clinical laboratory panels, physical examinations, and adverse event reporting. The patients were not queried about adverse events, but spontaneously reported them on their diary cards.

The sponsor presented the data for the two studies combined. There was no obvious relationship between treatment group and number of patients reporting at least one adverse event in the two pooled pediatric SAR studies.

Table 30: Number and Percent of Patients Experiencing At Least 1 Adverse Event in the Pooled Pediatric Studies (66 & 77)

	Total N	# (%)
Placebo	229	83 (36.2)
15 mg BID	224	79 (35.3)
30 mg BID	209	77 (36.8)
60 mg BID	213	74 (34.7)

4.3 Conclusions

Studies 66 and 77 were SAR studies performed in children ages 5-12 to determine the efficacy and safety of Allegra in the pediatric population. As evidenced by previous studies for other drugs submitted to this division, it is difficult to demonstrate the efficacy of a drug for SAR in pediatric populations. Accordingly, the Allegra pediatric studies have also produced results that are difficult to interpret. Two studies, with identical designs, and seemingly identical patient populations (with the exception of a slightly larger percent of Canadian patients in Study 66), have provided incompatible results. Study 66 showed a numerical superiority for placebo, whereas Study 77 demonstrated a statistically significant treatment effect for all dose groups. Of the Allegra dose groups, the 30 mg performed the best in Study 66 and the worst in Study 77, with no demonstration of dose response in either study.

The review of the results from subgroups revealed several complex relationships. The changes from baseline and differences in changes from baseline between placebo and the Allegra dose groups were related to baseline, age, gender and country. These differences were not exhibited similarly across studies. The inconsistency of results within and across studies (and countries), and the lack of a dose response in both studies, make it difficult to draw an overall conclusion regarding the efficacy of Allegra in the pediatric population.

5 Chronic Idiopathic Urticaria (CIU) Adult

5.1 Introduction

Summary

The sponsor submitted two adequate and well-controlled studies (Studies 39 and 67) to support the chronic idiopathic urticaria (CIU) indication for a twice daily dosing regimen. A third study (Study 19) used capsules instead of tablets and used a once daily dosing regimen. The sponsor submitted electronic data for Studies 39 and 67, but not Study 19. Thirty-four percent of patients in Study 19 dropped out, and this discontinuation rate was very different across treatment groups (Placebo: 59%, Allegra groups: 21%-32%). The large and differential dropout rates in Study 19 may have biased the results, therefore, the clinical review team (medical and statistics) focused review on the results of Studies 39 and 67. Studies 39 and 67 provide strong evidence that Allegra improves the symptoms of chronic urticaria.

Label

The sponsor would like to make the following claims in the label regarding the efficacy of Allegra in treating the symptoms of chronic idiopathic urticaria:

•

•

Redacted 1

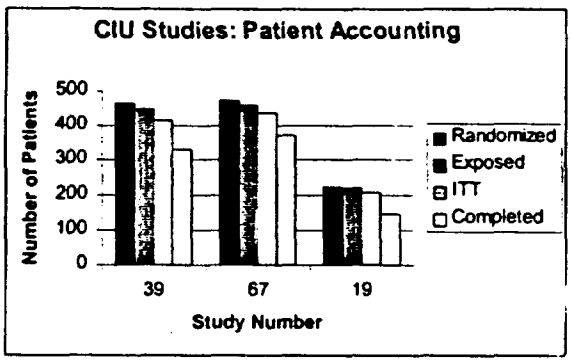
pages of trade secret and/or confidential commercial information

5.2 *Results*

5.2.1 Study Conduct

Studies 39 and 67 were large (n>400), multi-center trials (number of centers > 35). The sponsor defined Intent-to-Treat (ITT) as randomized patients who received study medication and had at least 1 post-baseline measurement. The ITT populations in Studies 39, 67, and 19 as defined by the sponsor, excluded 7%, 5%, and 7%, respectively, of the randomized patients who received at least 1 dose of the study medication ("exposed patients"). The effect that this missing data may have had on the results is not known.

Figure 16: CIU Number of Patients



Exposed = exposed to double-blind study medication
ITT = have both a baseline and at least one post-baseline measurement

Investigative Sites: Studies 39 and 67 had 37 and 39 investigative sites, respectively, with between 2 and 35 patients each.

Dropouts: Based on the "exposed" population, 26% (116/448) of patients from Study 39, 19% (88/461) of patients from Study 67, and 34% (76/222) of the patients from Study 19 discontinued treatment before completion of the study. The quality of a study is determined by many factors, including the number and percent of dropouts. The rate of dropout was different for different treatment groups (see Table 32 below). The placebo group had the greatest percentage of dropouts in all three studies. There is no "acceptable" percentage of dropouts for a study, however, when the percent exceeds 20-25%, the results of a study are likely to be unreliable and the confidence in the estimates obtained from the analyses is low.

Table 32: Studies 39 & 67 Number and Percent of Dropouts by Treatment Group

(Using All Randomized Patients Exposed to Study Medication)

	Placebo	20 mg	60 mg	120 mg	240 mg	Total
Study 39	32 (36%)	26 (28%)	21 (21%)	17 (21%)	20 (23%)	116 (26%)
Study 67	27 (29%)	23 (24%)	16 (18%)	11 (12%)	11 (13%)	88 (19%)
Study 19	30 (59%)	14 (32%)	8 (21%)	16 (32%)	8 (21%)	76 (34%)

* The sponsor's numbers for "Exposed Patients" in Table 9, (S8-V1.170-P66) for Study 39 were not consistent with the total number of randomized patients minus the numbers of patients listed as "Elected to discontinue prior to exposure" in Table 12, S8-V1.170-P68. Similarly, for Study 67, the sponsor's numbers in Table 9, S8-V1.189-P65 were not consistent with the numbers in Table 12, S8-V1.189-P67. The sponsor was requested to clarify the discrepancy and did so in a subsequent submission (6-15-99). The numbers in this table reflect the sponsors revised numbers.

The percent of dropouts in Study 19 was greater than that of the other two studies. This was not due to the longer duration of treatment (6 weeks instead of 4 weeks), in that most of the placebo dropouts in Study 19 discontinued by the end of Week 2 (37%) and by the end of the fourth week, 53% of the placebo patients had dropped out. The rates of dropout due to "lack of therapeutic effect " in Study 19 were different across treatment groups as well (placebo: 33%, combined Allegra treatment groups: 13%). Missing data that is related to the primary endpoint is problematic, especially when the quantity of missing data varies across treatment groups. As stated above, there is no "acceptable" percentage of dropouts for a study, however, the 59% withdrawal rate and the differential rates across treatment groups in Study 19 was unacceptably high. The

magnitude or direction of the bias introduced by these missing data cannot be determined, therefore, the primary analysis results of Study 19 will not be presented in this review.

The primary analysis the sponsor used in Studies 39 and 67 was based on 4-week average scores, calculated using the observed data. This analysis assumes that the patient's average symptoms while s/he was on treatment would have continued with no change had the patient remained in the study (a form of imputation). The sponsor also performed analyses assuming the patient's symptoms on the last day of treatment would have continued with no change had the patient remained in the study (LOCF analysis).

5.2.2 Sample Size

The sponsor powered Studies 39 and 67 based on the results from Study 19. The difference in mean pruritus scores (average effect over 6 weeks) in Study 19 between fexofenadine and placebo was 0.48 units with a standard deviation of 0.68. Using a two-sided, alpha=0.05 level test, a sample size of 75 patients per group provides 80% power for detecting an underlying difference of 0.32 units, provided that the population standard deviation is no greater than 0.68. In contrast to the SAR studies, the results of Studies 39 and 67 demonstrated greater treatment effects than expected (the mean pruritus differences ranged from 0.4 - 0.7 units).

5.2.3 Demographics

Treatment groups were well-balanced for all demographic factors in both studies (39 and 67). There were no statistically significant differences among treatment groups for gender, age, race, body weight or height in either study. The ITT population was predominantly female in both studies (Study 39: 70%; Study 67: 74%) and overwhelmingly Caucasian (Study 39: 88%; Study 67: 90%). Patient ages ranged from 12-70 (Study 39) and 12-68 years (Study 67). However, almost all patients (98% and 97%) were 16 years of age and older. The mean age in both studies was 39 years.

5.2.4 Primary and Secondary Efficacy Variables

The primary variable in both Studies 39 and 67 was mean pruritus score. Patients assessed the pruritus at approximately 7AM and 7PM every day for the symptoms over the previous 12-hours. No instantaneous (or trough) assessments were made in these studies. The primary endpoint was the change from baseline in average of the AM and PM pruritus scores over the four-week period. Secondary endpoints included:

- AM Pruritus
- PM Pruritus
- AM number of wheals (scored categorically, described on page 36)
- PM number of wheals
- Average of AM & PM number of wheals
- TSS (sum of average AM/PM pruritus and wheals scores)
- Interference with sleep
- Interference with daily activities
- Investigator's assessment of number of wheals (actual number of wheals, not a score)
- Investigator's assessment of longest diameter of wheals of average (scored categorically)
- Investigator's assessment of intensity of erythema on average
- Investigator's assessment of extent of skin area involved
- Quality of life instruments
- Dermatology Life Quality Index (DLQI): domains were symptoms/feelings, daily activities, leisure, work/school, personal relations, and treatment
- Work Productivity and Activity Impairment (WPAI): domains were work productivity, classroom productivity, and regular activity

Tools for the measurement of wheals and the extent of skin area involvement were provided to each investigator.

The results of the primary endpoint, mean pruritus, and the secondary endpoint, mean number of wheals, demonstrated improvement for all groups, including placebo. The scores of pruritus symptoms were, in general, lower than those of mean number of wheals, at baseline and during the treatment period (see Figure 17 below).

Figure 17: CIU Mean Baseline and Treatment Period (Average of 4 weeks) Scores

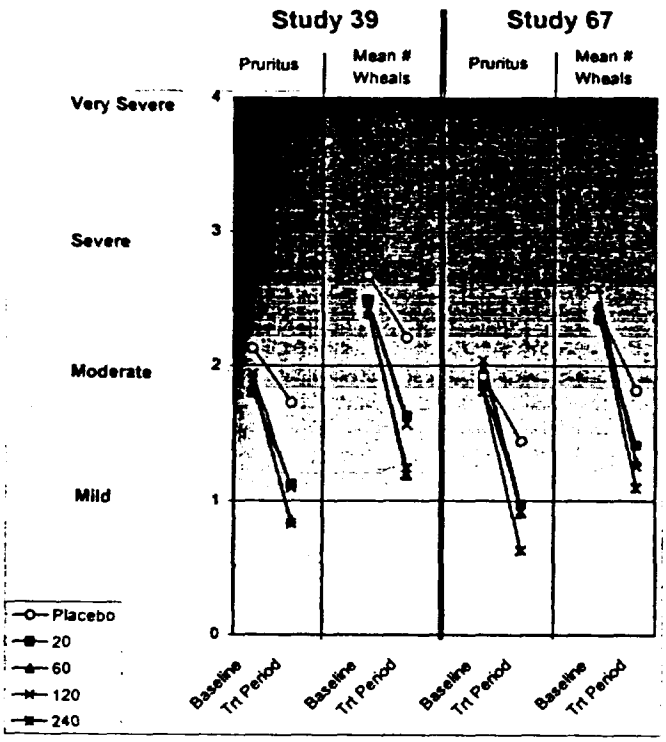


Table 33: CIU Studies' Pruritus, Number of Wheals and TSS Descriptive Results

(Change is the average of the four-week treatment period minus the baseline average)
(Standard errors = 0.1 for all means)

Assess	Study		Placebo	20 mg BID	60 mg BID	120 mg BID	240 mg BID
Mean Pruritus Score (AM + PM)	39	N	79	90	90	77	82
		Baseline	2.13	1.80	1.86	1.94	1.91
		Change	-0.40	-0.68	-1.00	-0.84	-1.08
	67	N	90	91	86	89	83
		Baseline	1.92	1.85	1.98	2.04	1.81
Scale: 0-4		Change	-0.47	-0.88	-1.07	-1.07	-1.18
Mean # of Wheals Score (AM + PM)	39	N	79	89	87	77	82
		Baseline	2.68	2.49	2.39	2.47	2.42
		Change	-0.47	-0.86	-1.20	-0.90	-1.17
	67	N	88	91	84	88	82
		Baseline	2.43	2.35	2.45	2.58	2.40
Scale: 0-4		Change	-0.61	-0.93	-1.15	-1.32	-1.30
Mean TSS (AM + PM)	39	N	79	89	87	77	82
		Baseline	4.81	4.30	4.26	4.42	4.33
		Change	-0.89	-1.53	-2.19	-1.75	-2.26
	67	N	88	91	84	88	82
		Baseline	4.36	4.20	4.42	4.63	4.20
Scale: 0-4		Change	-1.06	-1.81	-2.24	-2.39	-2.47

Studies 39 and 67 had very similar results, with two minor differences. The 20 mg and 120 mg groups in Study 67 performed slightly better than those in Study 39. In general, each increase in dose in Study 67 resulted in a greater mean change from baseline.

Table 34: CIU Studies Primary Analyses Results

Treatment differences in changes from baseline, confidence intervals and p-values are calculated from an ANCOVA model containing investigative site, treatment and baseline value for Study 67. The same dependent variable and factors, including a atment-by-baseline interaction, were used for Study 39 (all endpoints). Positive differences indicate Allegra superiority.

Assessment	Study		20 mg BID	60 mg BID	120 mg BID	240 mg BID
Mean Pruritus (AM + PM)	39	Trt Diff (95% CI)	0.3 (0.1, 0.5)	0.6 (0.4, 0.8)	0.4 (0.2, 0.7)	0.7 (0.5, 0.9)
		p-value	0.0098	0.0001	0.0001	0.0001
	67	Trt Diff (95% CI)	0.4 (0.2, 0.6)	0.6 (0.4, 0.8)	0.6 (0.4, 0.8)	0.7 (0.5, 0.9)
Scale: 0-4		p-value	0.0001	0.0001	0.0001	0.0001
Mean # Wheals (AM + PM)	39	Trt Diff (95% CI)	0.4 (0.1, 0.7)	0.7 (0.4, 1.0)	0.4 (0.1, 0.7)	0.7 (0.4, 1.0)
		p-value	0.0115	0.0001	0.0068	0.0001
	67	Trt Diff (95% CI)	0.3 (0.0, 0.6)	0.5 (0.3, 0.8)	0.7 (0.4, 1.0)	0.7 (0.4, 1.0)
Scale: 0-4		p-value	0.0238	0.0002	0.0001	0.0001
Mean TSS (AM + PM)	39	Trt Diff (95% CI)	0.6 (0.1, 1.1)	1.3 (0.8, 1.8)	0.9 (0.3, 1.4)	1.4 (0.9, 1.9)
		p-value	0.0109	0.0001	0.0011	0.0001
	67	Trt Diff (95% CI)	0.7 (0.3, 1.2)	1.2 (0.7, 1.6)	1.3 (0.9, 1.8)	1.4 (1.0, 1.9)
Scale: 0-8		p-value	0.0010	0.0001	0.0001	0.0001

The sponsor's primary analysis (pre-specified) was an analysis of covariance on the change from baseline mean pruritus score with the baseline pruritus score as a covariate and the treatment group and pooled investigative site as factors. (The sponsor pooled sites due to small numbers of patients at some sites.) The sponsor tested the treatment-by-baseline and treatment-by-center interaction terms at the 0.10 level. If the term was significant at the 0.10 level, the sponsor included it in the final model.[8] The only interaction that was significant at the 0.10 level was treatment-by-baseline in Study 67. (The p-values of both treatment-by-site interactions were >0.25, a more conservative alpha-level.)

[8] The algorithm for testing the interactions was as follows: put both interactions in model simultaneously; if p-values of both < 0.10 then keep both; if one >0.10, drop it from model and test other interaction; if the remaining interaction was <0.1, keep it; if not, the final model included only baseline, treatment and pooled investigative site.

Although all comparisons were statistically significant at the 0.05 level, the treatment effect sizes were relatively small for the 20 mg dose (0.3 – 0.4 units). The results demonstrate that the treatment differences between drug and placebo were consistently greatest for the 240 mg dose (0.7). A consistent increase in treatment effect with increasing dose was found in Study 67, but not Study 39.

Weekly Analyses

The results of the analyses of each week (using LOCF, with the same model the sponsor used excluding the treatment-by-baseline interaction) are presented in the appendix (Tables A-11 and A-12) for the mean pruritus and mean number of wheals scores. In Study 39, statistical significance on the pruritus endpoint was maintained through Week 4 for the 60 and 240 mg dose groups The placebo groups in Study 39 continued to improve during the four weeks of the study (even with the last observation carried forward for dropouts). The mean symptom scores of the active treatment groups also declined through Week 4, but less so than did the placebo group. Thus, the treatment differences decreased with time in Study 39. The treatment differences in Study 67 were robust (and statistically significant) for all dose groups through the fourth week.

AM vs. PM Analyses

The results of the analyses of the AM and PM Pruritus and Number of Wheals measures are provided in the appendix, Tables A-11 and A-12. The treatment effect sizes of the AM Pruritus Scores were slightly larger than those of the PM Pruritus Scores. With the exception of the 60 mg dose, the treatment effect sizes of the AM Number of Wheals scores were also larger than the PM Number of Wheals scores. The AM scores assessed the 12-hours prior to the 7AM dose. Therefore, it appears that a greater benefit was demonstrated for the nighttime symptoms than the daytime symptoms.

Shift Table Analyses

Percentages of patients that shifted none, one or more categories were calculated to determine the magnitude of responses of the individual patients.

Table 35: CIU Shift Table

Percent of Patients That Shifted Categories
Change from Baseline 2-Week Average Scores Rounded to Nearest Integer
TSS Divided into Categories of 2 units Each (ie: a change of 2 units = a shift of 1 category)

	Categories	Study 39					Study 67				
	Shifted	Placebo	20 mg	60 mg	120 mg	240 mg	Placebo	20 mg	60 mg	120 mg	240 mg
TSS	-4	0	0	1	1	4	0	0	2	5	0
	-3	1	9	9	10	9	1	9	10	11	16
	-2	23	18	32	26	29	15	26	31	32	27
	-1	32	39	36	35	37	45	37	36	34	45
	0	27	24	14	16	13	23	20	17	14	10
	1	16	10	7	10	9	16	8	5	5	2
	2	1	0	1	1	0	0	0	0	0	0
Mean	-4	0	0	0	0	0	0	0	1	2	0
Pruritus	-3	1	2	8	5	9	1	3	6	6	4
	-2	10	17	23	21	20	8	20	30	24	27
	-1	35	37	39	38	39	40	41	35	44	43
	0	42	38	24	31	32	43	31	24	22	24
	1	11	7	4	5	1	8	4	3	2	2
	2	0	0	1	0	0	0	1	0	0	0
Mean #	-4	1	0	2	3	4	0	1	2	10	4
of Wheals	-3	1	11	10	12	12	6	8	8	8	13
	-2	13	17	18	13	22	7	21	27	24	23
	-1	38	29	39	32	34	40	31	33	34	30
	0	37	38	23	30	20	41	26	24	20	27
	1	8	4	6	6	7	7	3	4	3	2
	2	3	0	1	4	1	0	0	1	0	0

The results of these analyses are consistent with the results of the primary analysis. The treatment differences are large for both Mean Pruritus Scores and Mean Number of Wheals Scores. Adding up the percentages of atients whose symptoms decreased more than 1 category, in Study 39 the differences between the 60 mg oup and placebo were 20% for Pruritus and 15% for Wheals. The differences between the 60 mg group and placebo in Study 67 were even more striking (28% for Pruritus and 24% for Wheals). These differences in

percentages seem remarkable when compared to those seen in the adult SAR studies. The differences with placebo in percentages of patients whose symptoms decreased more than 1 category in the SAR studies ranged between 1 and 12, with most differences between 7 and 9.

The increase in effect with increasing dose is most pronounced in the TSS of Study 67. The percentages of patients with no improvement at all (or increased severity of symptoms) for the 4 dose groups in order from lowest to highest strength were 28, 22, 19 and 12, respectively.

Other Patient Assessments

The results of other secondary efficacy measures in Studies 39 and 67 were concordant with results of the primary efficacy measure. Mean differences (using LOCF) in interference with sleep scores were about a half a unit for the 60 – 240 mg dose groups while difference in interference with daily activities for these dose groups were slightly larger (0.69 – 0.74 units), see Appendix Tables A-11 and A-12. The differences for the 20 mg dose group were 0.35 and 0.34 units for the sleep and daily activities scores respectively. All dose groups demonstrated changes from baseline that were statistically significantly greater than placebo changes for the sleep and daily activities measures.

The sponsor states that the "investigator assessment results did not correlate well with the patients' assessments and contributed little to the evaluation of treatment efficacy" (VOL 1.1, P319). This reviewer feels that the investigator assessment results were fairly consistent with the patients' mean pruritus and mean number of wheals scores in that

- in Study 39, the 120 mg dose group did not perform as well as the 60 and 240 mg dose groups in Study 39, and
- in Study 67 the 240 mg dose group demonstrated greater changes from baseline than the other treatment groups.

Health Outcomes Results

Health outcomes [quality of life (QOL), work/classroom productivity] data were collected in Studies 39 and 67. The impact of urticaria on the patient's health outcomes was determined from the results of self-administered questionnaires completed by the patient at the end of each visit. The QOL instrument was the Dermatology Life Quality Index (DLQI). The productivity instrument was the Work Productivity and Activity Impairment (WPAI) questionnaire. DDMAC does not recognize either of these instruments as validated in the chronic idiopathic urticaria disease area (see Ms. Fisher's review, June 15, 1999). The sponsor defined three "primary outcome parameters" in both studies as follows:

1. average change from baseline in overall DLQI;
2. average change from baseline percent of work/classroom productivity; and
3. average change from baseline percent of work/classroom time missed.

The latter two "primary" parameters each have two parts and are actually four primary parameters:

1. average change from baseline percent of work productivity;
2. average change from baseline percent of classroom productivity;
3. average change from baseline percent of work time missed.
4. average change from baseline percent of classroom time missed.

Secondary endpoints focused on specific domains of the instruments (DLQI symptoms/feelings, DLQI daily activities, DLQI leisure, DLQI work/school, DLQI personal relations, DLQI treatment, WPAI work productivity, WPAI classroom productivity, and WPAI regular activity).

Some of the questions on the DLQI questionnaire were related to the questions of the primary efficacy endpoint, pruritus. The scores of 2 (moderate) and 3 (severe) of the pruritus endpoint were:

- moderate: annoying and troublesome; may have interfered somewhat with normal daily activity and/or sleep
- severe: very annoying and troublesome; substantially interfered with normal daily activity and/or sleep)

Three of the 10 questions on the DLQI questionnaire were related to normal daily activity:

- Over the last week, how much has your skin interfered with you going shopping or looking after your home or garden?
- Over the last week, how much has your skin affected any social or leisure activities?
- Over the last week, how much has your skin made it difficult for you to do any sport?

The sponsor conducted all statistical tests at the alpha=0.05 level. No corrections were made for multiple comparisons. The sponsor states that the dose decision is based on the primary efficacy analysis. Once the dose decision had been made, the sponsor planned to make statistical decisions concerning the health outcomes only with the indicated dose, and not all dose groups. Therefore, the sponsor argued, no multiple comparisons adjustments were necessary. It is necessary to adjust the alpha-level because these tests were secondary to the primary efficacy endpoint in the studies: mean pruritus score. In addition, the sponsor did not choose one primary outcome measure, but three (there are five if the WPAI scores are divided into two groups: work and classroom).

The sponsor did not submit the health outcomes data electronically, therefore, this reviewer could not perform a thorough statistical review of these results.

In both Studies 39 and 67, the numbers of patients who had completed a baseline and at least one follow-up questionnaire were fewer than the ITT numbers used for the primary efficacy variable.

Table 36: CIU Quality of Life Patient Numbers

Study	N Rand-omized	N Ex-posed	N (% of exposed) used in Prim. Eff. Var	N (% of exposed) used in DLQI	N (% of exposed) used in WPAI work/class
39	468	449	418 (93)	**403 (90)**	**369 (82)** work : 311 (69) class: 58 (13)
67	476	461	439 (95)	**423 (92)**	**359 (78)** work: 294 (64) class: 65 (14)

For Studies 39 and 67, respectively, only 90 and 92% of the patients who were randomized and received study medication were included in the analysis of the health outcomes data. The mean baseline total symptom score of the patients included in the health outcomes dataset in Study 39 was 5 units, greater than that of the patient population included in the primary efficacy analysis. The mean baseline TSS of every treatment group of the patients in the primary efficacy analysis was < 5 units (Placebo: 4.8; 20 mg: 4.3; 60 mg: 4.3; 120 mg: 4.42; 240 mg: 4.3). From the limited information presented in the study report, it appears as though the patients in the health outcomes ITT dataset in Study 39 had more severe symptoms than those in the primary efficacy variable dataset. The impact that the excluded patients (with less severe symptoms) may have had on the analysis is impossible to quantify, however, it can be postulated, from the results of the primary efficacy variable, that the treatment effect of Allegra on health outcomes would be more pronounced in the more severe patients. Therefore, the exclusion of the less severe patients would yield a greater treatment effect than if all patients exposed to drug had been included in the analysis. The baseline TSS values of the health outcomes ITT dataset in Study 67 were not provided.

The results of the three or five primary outcome measures were consistent across studies. In both studies, the overall DLQI change from baseline score was statistically significant between all Allegra groups and placebo. Additionally, in both studies the WPAI percent of work productivity was statistically significant for all doses (with the exception of the 20 mg dose in Study 39). The "primary" outcome measures that did not demonstrate statistical significance for any dose group in either study were: WPAI percent of classroom productivity, the average change from baseline percent of classroom time missed, and the average change from baseline percent of work time missed (with the exception of the 240 mg dose group in Study 67). Since only two out of five co-primary outcome endpoints were statistically significant between Allegra groups and placebo, and the sponsor did not adjust for multiple comparisons, the sponsor has not proven that Allegra improves health outcomes more than placebo.

5.2.5 Special Populations

Summary: For both studies, the consistency of treatment effect across subgroups of patients was assessed using change from baseline MPS, the primary efficacy variable. These subgroup analyses were conducted on the primary efficacy endpoint using an ANCOVA model with treatment, center, and subgroup factor as factors, and baseline MPS as a covariate. An interaction between treatment and the subgroup was included in these models as well. Factors used in the subgroup analyses were investigative site, age (<16 years, 16-<40 years,

≥40 years), weight (<60 kg, 60-<90 kg, and ≥90 kg), race (Caucasian, non-Caucasian), gender, and baseline subgroup (≤2 units, >2 units).

In general, the treatment effect was consistent across subgroups of patients; i.e. the treatment-by-subgroup interactions were not statistically significant (p≥0.25). However, the numbers of non-Caucasian patients and patients less than 16 years old were small, making the results of these subgroups of patients inconclusive. One interesting finding was that the main effect of age was statistically significant in Study 39 and marginally statistically significant in Study 61. The older patients scores did not decrease during the treatment period as much as the younger patients' scores did. This phenomenon was seen in all treatment groups, including placebo.

Study 39: The baseline-by-treatment interaction was significant at the 0.25 level (p=0.0190). The treatment effect of the higher doses (60-240 mg) was greater in the high baseline patients. This is in contrast to the effect on SAR symptoms seen in the U.S. adult study (Study 81) and one of the pediatric studies (77) in which the treatment effect was only evident in the low baseline patients.

Table 37: Study 39 Baseline Subgroups

Baseline		N	Change from Baseline	Difference with Placebo
Low (≤2 units)	Placebo	44	-0.11	
	20 mg BID	62	-0.51	**0.40**
	60 mg BID	62	-0.57	**0.46**
	120 mg BID	50	-0.41	**0.30**
	240 mg BID	55	-0.68	**0.58**
High (>2 units)	Placebo	35	-0.76	
	20 mg BID	28	-0.98	**0.22**
	60 mg BID	28	-1.86	**1.10**
	120 mg QD	27	-1.68	**0.92**
	240 mg QD	27	-1.86	**1.10**

The main effect of age was statistically significant at the 0.05 level (p=0.0114), indicating a greater effect for patients under the age of 40. However, the treatment *differences* were similar across all age groups.

Table 38: Study 39 Age Subgroups

Age		N	Change from Baseline	Difference with Placebo
< 16 years	Placebo	5	-0.52	
	20 mg BID	2	-0.82	**0.30**
	60 mg BID	1	-1.34	**0.82**
	120 mg BID	0	—	
	240 mg BID	2	-0.72	**0.20**
16- <40 years	Placebo	39	-0.38	
	20 mg BID	51	-0.72	**0.34**
	60 mg BID	46	-1.16	**0.78**
	120 mg BID	36	-1.04	**0.66**
	240 mg BID	45	-1.17	**0.79**
≥ 40 years	Placebo	35	-0.31	
	20 mg BID	37	-0.66	**0.35**
	60 mg BID	43	-0.79	**0.48**
	120 mg BID	41	-0.66	**0.35**
	240 mg BID	35	-0.99	**0.68**

Study 67: The baseline-by-treatment interaction was significant at the 0.25 level (p=0.1080). Similar to Study 39, the treatment effect was greater in the high baseline patients. This is in contrast to the effect on SAR symptoms seen in the adult SAR study (Study 81) in which the treatment effect was only evident in the low baseline patients.

Table 39: Study 67 Baseline Subgroups

Baseline		N	Change from Baseline	Difference with Placebo
Low (≤2 units)	Placebo	62	-0.23	
	20 mg BID	61	-0.55	**0.32**
	60 mg BID	52	-0.69	**0.46**
	120 mg BID	51	-0.74	**0.51**
	240 mg BID	64	-0.82	**0.59**
High (>2 units)	Placebo	30	-0.87	
	20 mg BID	34	-1.47	**0.60**
	60 mg BID	35	-1.79	**0.92**
	120 mg QD	40	-1.83	**0.95**
	240 mg QD	20	-1.89	**1.02**

The main effect of age was marginally statistically significant at the 0.05 level (p=0.0673), indicating a greater effect for patients under the age of 40. Similar to Study 39, the treatment *differences* were similar across all age groups, with the exception of the patients less than 16 years. The small treatment differences were due to the large placebo response. Since the number of patients <16 years was so small, the small treatment differences in this subgroup are inconclusive.

Table 40: Study 67 Age Subgroups

Age		N	Change from Baseline	Difference with Placebo
< 16 years	Placebo	2	-1.02	
	20 mg BID	3	-0.65	**-0.37**
	60 mg BID	2	-0.82	**-0.20**
	120 mg BID	2	-1.20	**0.18**
	240 mg BID	3	-1.26	**0.24**
16- <40 years	Placebo	42	-0.50	
	20 mg BID	46	-0.90	**0.40**
	60 mg BID	49	-1.11	**0.61**
	120 mg BID	44	-1.16	**0.66**
	240 mg BID	37	-1.36	**0.86**
≥ 40 years	Placebo	46	-0.42	
	20 mg BID	42	-0.87	**0.45**
	60 mg BID	35	-1.01	**0.59**
	120 mg BID	43	-0.96	**0.54**
	240 mg BID	43	-1.02	**0.60**

5.2.6 Safety

Safety evaluations included clinical laboratory panels, physical examinations, and adverse event reporting. The patients were not queried about adverse events, but spontaneously reported them on their diary cards.

There was no obvious relationship between treatment group and number of patients reporting adverse events in either of the two studies (see Table 41 below). The most notable finding was that Study 67 had slightly larger percentages of patients reporting at least one adverse event compared to Study 39.

Table 41: Number and Percent of Patients with At Least One Adverse Event

	Study 39	Study 67

Placebo	44 (52%)	61 (66%)
20 mg	57 (62%)	67 (71%)
60 mg	57 (59%)	51 (57%)
120 mg	48 (61%)	60 (65%)
240 mg	52 (62%)	50 (59%)

The system organ classes within which most adverse events were reported in both treatment groups were neurologic and respiratory (Studies 39 and 67) and gastrointestinal (Study 39 only). The adverse event reported most frequently by patients in each of the treatment groups in both studies was headache (Study 39: 25%; Study 67: 19%). No dose-related trends in adverse events were seen.

5.3 Conclusions

The sponsor submitted three studies to support a chronic idiopathic urticaria indication. Study 19, the first of the three studies, was performed in Europe and used a once-daily dosing regimen. Valid inferences from this 6-week study cannot be made due to high and different discontinuation rates across treatment groups (placebo: 59%; 20 and 120 mg: 32%; 60 and 240 mg: 21%).

Studies 39 and 67 (using a twice daily dosing regimen) were only 4 weeks long and had smaller dropout rates. The results of these studies demonstrated statistically significant treatment effects for all dose groups using the primary endpoint 12-hour reflective mean pruritus score (0-4 scale). The treatment effects ranged from 0.3 units for the 20 mg dose to 0.7 units for the 240 mg dose. There seemed to be a slight increase of benefit with increasing dose in Study 67. In contrast, in Study 39, the 120 mg dose did not perform as well as the 60 mg dose, resulting in a lack of dose response.

The treatment effect in Study 39 decreased from Week 1 through Week 4 due to an improving placebo group (using both observed data and last observation carried forward). In contrast, the results of the LOCF analyses of Study 67 demonstrate robust treatment effects through Week 4.

The studies did not assess symptoms at the end-of-dosing interval.

The treatment effects were consistent across subgroups of patients stratified by age, gender, race and weight. In general, patients with greater severity of pruritus symptoms at baseline received greater benefit from Allegra.

The sponsor defined three "primary" Quality of Life (QOL) endpoints in this study (actually five separate endpoints - due to classroom versus work related questions). The patients who responded to the QOL questionnaires were a subset of the patients included in the primary efficacy analyses. In Study 39, this subset of patients had greater pruritus and number of wheals scores at baseline than the primary efficacy ITT population of patients. Since the patients with more severe pruritus symptoms at baseline appeared to receive greater benefit of Allegra for the symptoms of pruritus, it is reasonable to assume that the missing QOL data of the less severe patients may have biased the QOL results in favor of Allegra. (The sponsor did not provide the baseline pruritus or wheal information for the QOL ITT subset of patients in Study 67.) In both studies, with the exception of the 240 mg dose group in Study 67, the results of only 2 of the 5 "primary" QOL endpoints were statistically significantly different between Allegra and placebo. Therefore, the studies have not demonstrated that Allegra improves the QOL endpoints that the sponsor lists in the proposed label.

6 Appendix

6.1 General

Table A- . .a Audit Procedures

	SAR-81	SAR-32	SAR-61	SAR-66/77	CIU-39	CIU-67
CRF data	Electronic sent to HMR	Paper sent to Parexel	Paper sent to HMR	Paper sent to Parexel	Paper sent to Parexel	Paper sent to Parexel
Data Entry	Single data entry by investigative site[9]	Double-entry at Parexel	Double-entry at HMR	Double-entry at Parexel	Double-entry at Parexel	Double-entry at Parexel
Validation checks	Checks & changes made prior to receipt of data in-house	Run in-house after data entry	Run in-house after data entry	Checks & changes made as data was entered	Checks & changes made as data was entered	Checks & changes made as data was entered
Audit & Error Rate	7 patients randomly chosen and 100% verification performed on the 10,476 data fields; no errors found (0.0%) [10]; trt assign-ment numbers were verified on 10% of pts at each site, 0 errors found	All patients chosen and 100% of critical items verified[11]; 5% or 1 pt/site was chosen and 100% verification performed; no systematic errors found; overall error rate=0.115%	10% of all data audited by comparing final SAS data listings with study books; 100% audit of safety data; 10% of patients audited for primary eff. var.; 100% audit of trt assignment numbers (no error rates were listed in the NDA for this study)	1st 3 pts[12]; when 25% of CRF data were on-line, a 100% verifica-tion for 8 randomly chosen pts (not previously audited); repeated again after 50% & 75% of data on-line; when 100% of data on-line, 10% of pts had 100% CRF audit; error rate was 0.02% for each study	1st 3 pts[13]; when 25% of CRF data were on-line, a 100% verifica-tion for 7 randomly chosen pts (not previously audited); repeated again after 50% & 75% of data on-line; 10% of pts had 100% CRF verific-ation; error rates were 0.1 - 0.39% [14], [8], [9]	1st 3 pts[3]; when 25% of CRF data were on-line, a 100% verifica-tion for 7 randomly chosen pts (not previously audited); repeated again after 50% & 75% of data on-line; 10% of pts had 100% CRF verification; error rates were 0.1 - 1.7% [15], [16], [17]
Changes after unblinding	Database was re-opened three times to correct several diff. kinds of errors[18]	Re-opened after unblinding (17 changes made to diary & demography data)	Re-opened after unblinding due to late return of queries from investigator[19]	No changes	No changes	No changes

[9] The Sponsor trained each site in the Remote Study Management system (entering the data into a notebook computer). User manuals were provided to each site as a reference for use throughout the study.

[10] In addition, the transfer of data from external sources was checked: 1 patient randomly chosen was verified (hard copy of lab report compared to database line listing received from Covance); no errors were found. The transfer of data from database to SAS datasets was checked as well: 1 patient randomly chosen - one systematic error found in the previous medications dosage.

[11] Critical items included: adverse events, serious adverse events, patient final evaluation.

[12] The first 3 complete patients in-house (for each study) were manually reviewed 100% comparing CRF to the database.

[13] The first 3 "clean" patients in house were manually reviewed 100% comparing the CRF to the database to ensure consistency and appropriate completion of the CRF. "Clean" patients were defined as patients for whom all Screening Visit and Visit 1 data were in-house and on-line, with all data validation checks and data issues resolved and all corrections documented.

[14] The sponsor pre-specified that if the error rate was <0.1%, no further verification was performed. If error rate ≥ 0.1%, then 100% verification of all data fields for the failed database table was performed. The following tables (with error rates in parentheses) had to be 100% verified: Adverse Events (0.2%) and Laboratory (0.39%).

[15] The sponsor pre-specified that if the error rate was <0.1%, no further verification was performed. If error rate ≥ 0.1%, then 100% verification of all data fields for the failed database table would be performed. The following tables (with error rates in parentheses) had to be 100% verified: Additional Inclusion (1.7%), Adverse Events (0.24%), Medical History (0.33%), Pregnancy (0.14%) and Study Medication Compliance (error rate 0.13%).

[16] After the database lock (but before unblinding), the sponsor found blank fields in the Adverse Event database tables. The sponsor investigated this problem and found queries that had never been reviewed (due to a status coding system problem). A detailed audit was completed to determine which queries had not been reviewed and the impact on the accuracy and validity of the database. All discrepant data points were reviewed and appropriate corrections were made to the database.

[17] Further, after incorporation of external data; disposition codes were assigned to each patient prior to database finalization. A 100% verification of the protocol violation codes and disposition classifications was performed against the database to ensure accuracy of the disposition code assignment.

[18] Errors included: adverse events, demographics, pk bar codes, concomitant medications, diary dates, missing values, study termination reasons, and treatment assignment numbers.

[19] The majority of these final amendments were of the following type: reason for withdrawal, coding of adverse events, relationship between concomitant medication and adverse events, diary card dates and day numbers. These amendments did not affect the primary efficacy analysis.

6.2 *Seasonal Allergic Rhinitis Adult*

Figure A- 1: Individual 8 AM Instantaneous Symptom Scores over time (LOCF)

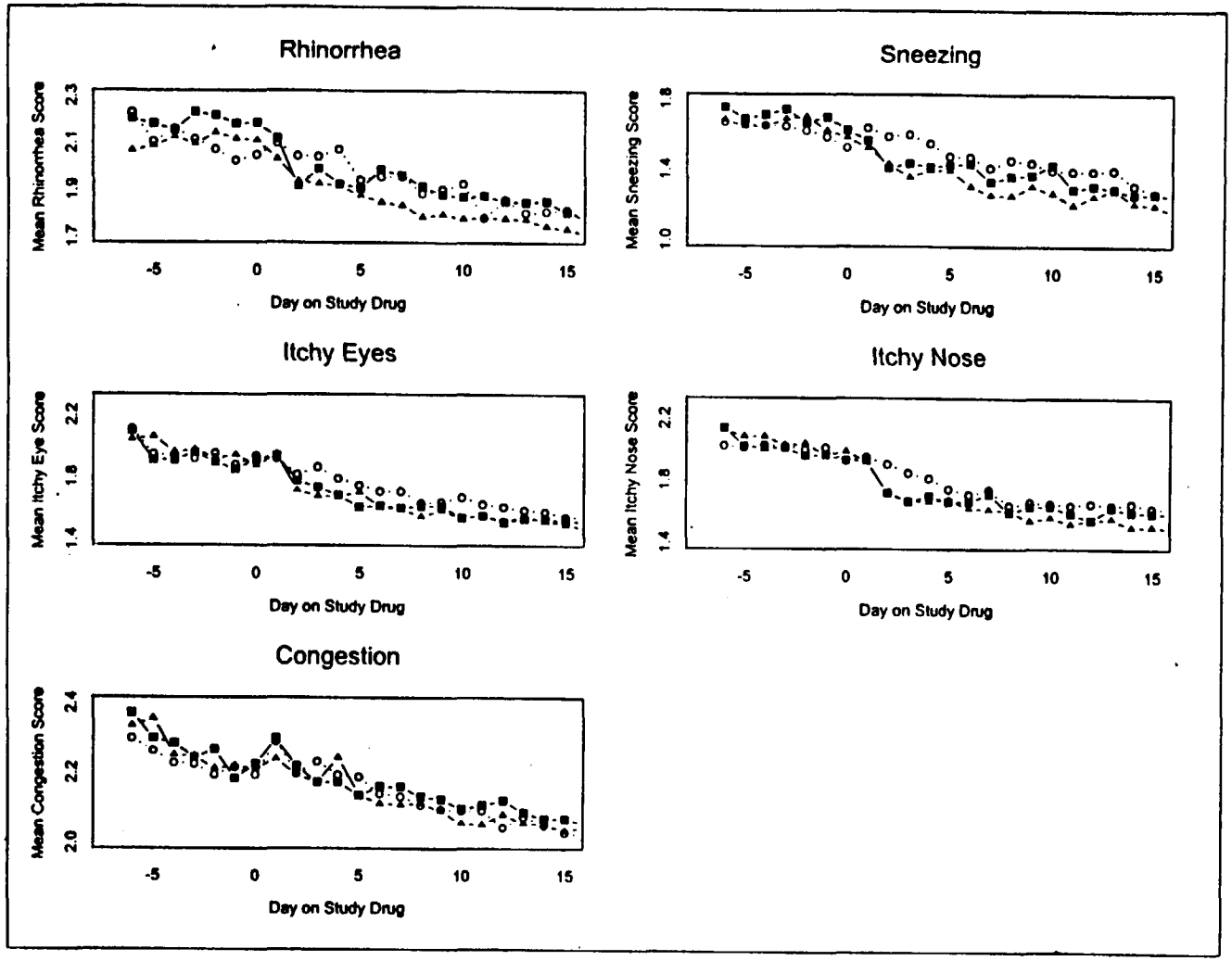


Table A- 2: Study 81. Secondary Efficacy Variable Results

Means for Baseline, Week 1, Week 2 and Weeks 1 & 2 averages are unadjusted (raw) means. Treatment differences in change from baseline means and the associated p-values are from an ANCOVA model containing investigative site, treatment and baseline value.

Symptom		Timepoint	8 AM Instantaneous						8 AM 12-hour Reflective						8 PM 12-hour Reflective					
			Placebo n=293	Fexofenadine120 mg n=288			Fexofenadine180 mg n=283		Placebo n=293	Fexofenadine 120 mg n=288			Fexofenadine 180 mg n=283			Placebo n=293	Fexofenadine 120 mg n=288			Fexofenadine180 mg n=283	
			MEANS	MEANS	DIFF	P-VAL	MEANS	DIFF P-VAL	MEANS	MEAN	DIFF	P-VAL	MEANS	DIFF	P-VAL	MEANS	MEANS	DIFF	P-VAL	MEANS	DIFF P-VAL
TSS		Baseline	7.61	7.72			7.69		7.57	7.66			7.57			7.32	7.34			7.30	
	Change	Week 1	-0.58	-1.01	0.39	0.0121	-1.16	0.55 0.0004	-0.36	-1.01	0.61	0.0001	-1.08	0.72	0.0000	-0.30	-1.32	0.99	0.0000	-1.21	0.91 0.0000
	From	Week 2	-1.16	-1.41	0.20	0.2849	-1.58	0.40 0.0313	-0.98	-1.40	0.37	0.0484	-1.51	0.54	0.0042	-1.00	-1.53	0.51	0.0085	-1.60	0.63 0.0014
	Baseline	Wks 1-2 Avg	-0.83	-1.17	0.30	0.0505	-1.35	0.49 0.0016	-0.62	-1.17	0.50	0.0012	-1.28	0.65	0.0000	-0.60	-1.40	0.78	0.0000	-1.39	0.79 0.0000
Sneezing		Baseline	1.62	1.66			1.63		1.67	1.70			1.66			1.65	1.70			1.67	
	Change	Week 1	-0.10	-0.26	0.14	0.0049	-0.28	0.17 0.0005	-0.08	-0.26	0.17	0.0003	-0.28	0.20	0.0000	-0.03	-0.38	0.33	0.0000	-0.33	0.29 0.0000
	From	Week 2	-0.23	-0.34	0.10	0.0880	-0.36	0.14 0.0158	-0.19	-0.35	0.16	0.0051	-0.37	0.19	0.0006	-0.20	-0.43	0.21	0.0002	-0.41	0.20 0.0003
	Baseline	Wks 1-2 Avg	-0.16	-0.29	0.12	0.0117	-0.32	0.16 0.0008	-0.12	-0.30	0.16	0.0004	-0.32	0.20	0.0000	-0.11	-0.40	0.27	0.0000	-0.37	0.25 0.0000
Rhinorrhea		Baseline	2.09	2.17			2.11		2.04	2.11			2.05			1.92	1.94			1.92	
	Change	Week 1	-0.10	-0.23	0.10	0.0410	-0.23	0.12 0.0123	-0.05	-0.21	0.13	0.0061	-0.20	0.15	0.0019	-0.05	-0.29	0.23	0.0000	-0.23	0.18 0.0004
	From	Week 2	-0.26	-0.33	0.03	0.5468	-0.33	0.07 0.2403	-0.20	-0.32	0.09	0.0897	-0.31	0.11	0.0394	-0.19	-0.32	0.11	0.0531	-0.32	0.13 0.0294
	Baseline	Wks 1-2 Avg	-0.17	-0.27	0.07	0.1351	-0.27	0.10 0.0333	-0.11	-0.25	0.11	0.0147	-0.25	0.14	0.0030	-0.11	-0.30	0.17	0.0003	-0.27	0.15 0.0016
Itchy, Watery		Baseline	1.92	1.91			1.95		1.90	1.89			1.88			1.84	1.82			1.83	
Red Eyes	Change	Week 1	-0.16	-0.24	0.07	0.1270	-0.30	0.13 0.0061	-0.11	-0.24	0.13	0.0045	-0.27	0.17	0.0004	-0.11	-0.31	0.21	0.0000	-0.31	0.21 0.0000
	From	Week 2	-0.32	-0.38	0.05	0.3836	-0.42	0.09 0.1210	-0.31	-0.37	0.06	0.3137	-0.39	0.09	0.1285	-0.30	-0.41	0.11	0.0591	-0.44	0.15 0.0119
	Baseline	Wks 1-2 Avg	-0.23	-0.29	0.06	0.2188	-0.35	0.11 0.0194	-0.19	-0.29	0.10	0.0331	-0.33	0.14	0.0034	-0.18	-0.35	0.17	0.0005	-0.37	0.19 0.0001
Itchy Nose,		Baseline	1.98	1.97			2.01		1.95	1.96			1.98			1.91	1.88			1.88	
Mouth, Throat	Change	Week 1	-0.21	-0.29	0.08	0.1166	-0.35	0.13 0.0071	-0.12	-0.30	0.17	0.0007	-0.33	0.20	0.0000	-0.12	-0.33	0.21	0.0000	-0.34	0.23 0.0000
and/or Ears	From	Week 2	-0.35	-0.36	0.01	0.9024	-0.47	0.11 0.0627	-0.29	-0.35	0.05	0.3576	-0.44	0.14	0.0159	-0.30	-0.38	0.08	0.1706	-0.44	0.15 0.0109
	Baseline	Wks 1-2 Avg	-0.27	-0.32	0.05	0.3217	-0.40	0.12 0.0106	-0.19	-0.32	0.12	0.0143	-0.38	0.18	0.0002	-0.20	-0.35	0.16	0.0014	-0.39	0.20 0.0001
Congestion		Baseline	2.23	2.25			2.24		2.24	2.23			2.23			2.02	2.02			2.02	
	Change	Week 1	-0.06	-0.08	0.02	0.6984	-0.09	0.02 0.6705	-0.04	-0.07	0.03	0.4828	-0.07	0.03	0.4438	0.00	-0.17	0.16	0.0006	-0.13	0.11 0.0158
	From	Week 2	-0.17	-0.16	-0.02	0.6642	-0.19	0.01 0.8687	-0.13	-0.16	0.02	0.6799	-0.20	0.06	0.2205	-0.10	-0.20	0.10	0.0675	-0.19	0.09 0.0845
	Baseline	Wks 1-2 Avg	-0.10	-0.11	0.00	0.9762	-0.13	0.01 0.7915	-0.07	-0.10	0.03	0.4861	-0.13	0.05	0.2359	-0.04	-0.18	0.13	0.0034	-0.15	0.10 0.0217

Table A- 3: Percent of Patients considered "Allergy Sufferers" in Second Week of Double-Blind Therapy

Using the 8 AM Instantaneous Assessments

	Placebo n=293	120 mg QD n=288	180 mg QD n=283
# (%) of patients with ≥2 symptoms with scores of ≥2 on 5/7 days (of the last 7 days on study)	178 (61)	180 (63)	164 (58)
# (%) of patients with TSS≥5 on 5/7 days (of the last 7 days on study)	204 (70)	198 (69)	184 (65)
# (%) of patients who met both these criteria	176 (60)	173 (60)	163 (58)

(If the patient was on study only 6 days, 5 of the six days had to meet the criteria. If the patient was on study less than 6 days, all the days had to meet the criteria.)

Using the 8 AM Reflective Assessments

	Placebo n=293	120 mg QD n=288	180 mg QD n=283
# (%) of patients with ≥2 symptoms with scores of ≥2 on 5/7 days (of the last 7 days on study)	184 (63)	180 (63)	161 (57)
# (%) of patients with TSS≥5 on 5/7 days (of the last 7 days on study)	207 (71)	190 (66)	187 (66)
# (%) of patients who met both these criteria	181 (62)	172 (60)	159 (56)

Using the 8 PM Reflective Assessments

	Placebo n=293	120 mg QD n=288	180 mg QD n=283
# (%) of patients with ≥2 symptoms with scores of ≥2 on 5/7 days (of the last 7 days on study)	169 (58)	146 (51)	141 (50)
# (%) of patients with TSS≥5 on 5/7 days (of the last 7 days on study)	194 (66)	171 (59)	168 (59)
# (%) of patients who met both these criteria	164 (56)	141 (49)	140 (49)

Figure A- 2: Barchart of Percent of Patients Who Still Met Entrance Criteria At End of Study By Treatment Group

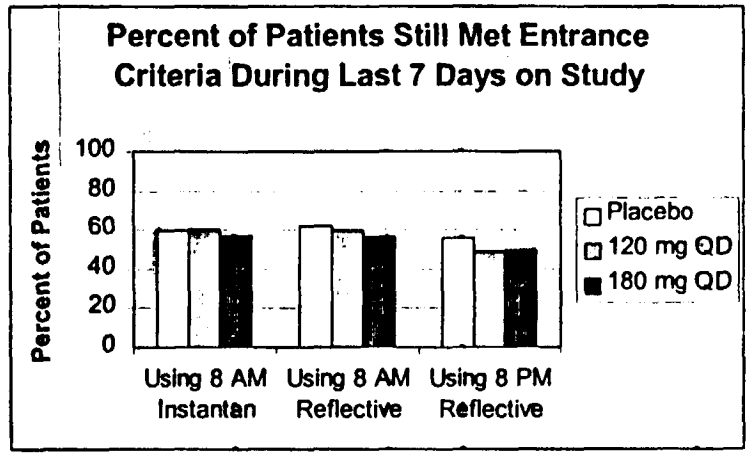


Table A-4: Study 32 Secondary Efficacy Variable Results for 24-hour Reflective

Study 32: Secondary Efficacy Variable Results (Sponsor's results)

Means for Baseline, Week 1, Week 2 and Weeks 1 & 2 averages are unadjusted (raw) means. Treatment differences in change from baseline means and the associated p-values are from an ANCOVA model containing investigative site, treatment and baseline value.

			24-hour Reflective													
Symptom		Timepoint	Placebo		Fexofenadine 120 mg				Fexofenadine 180 mg				Cetirizine			
			N	MEANS	N	MEANS	DIFF	P-VAL	N	MEANS	DIFF	P-VAL	N	MEANS	DIFF	P-VAL
TSS		Baseline	201	7.34	211	7.24			202	7.36			207	7.33		
	Change	Week 1	201	-1.21	210	-2.30	1.15	<0.0001	201	-2.63	1.46	<0.0001	207	-2.77	1.58	<0.0001
	From	Week 2	179	-2.08	193	-2.86	0.84	0.0021	184	-3.32	1.24	<0.0001	192	-3.10	1.02	0.0002
	Baselin	Wks 1-2 Avg	201	-1.51	211	-2.51	1.06	<0.0001	202	-2.89	1.41	<0.0001	207	-2.89	1.41	<0.0001
Sneezing		Baseline	201	1.79	211	1.78			202	1.78			207	1.77		
	Change	Week 1	201	-0.31	210	-0.59	0.29	<0.0001	201	-0.67	0.37	<0.0001	207	-0.74	0.44	<0.0001
	From	Week 2	179	-0.49	193	-0.70	0.20	0.0056	184	-0.82	0.33	<0.0001	192	-0.76	0.27	0.0002
	Baselin	Wks 1-2 Avg	201	-0.37	211	-0.63	0.26	<0.0001	202	-0.73	0.37	<0.0001	207	-0.74	0.38	<0.0001
Rhinorrhea		Baseline	201	1.92	211	1.90			202	1.96			207	1.89		
	Change	Week 1	201	-0.27	210	-0.54	0.28	<0.0001	201	-0.62	0.34	<0.0001	207	-0.66	0.40	<0.0001
	From	Week 2	179	-0.49	193	-0.67	0.19	0.0106	184	-0.81	0.30	0.0001	192	-0.76	0.30	0.0001
	Baselin	Wks 1-2 Avg	201	-0.34	211	-0.59	0.26	0.0001	202	-0.69	0.33	<0.0001	207	-0.70	0.37	<0.0001
Itchy, Watery		Baseline	201	1.75	211	1.76			202	1.76			207	1.79		
Red Eyes	Change	Week 1	201	-0.26	210	-0.55	0.28	<0.0001	201	-0.64	0.38	<0.0001	207	-0.65	0.38	<0.0001
	From	Week 2	179	-0.49	193	-0.72	0.24	0.0022	184	-0.77	0.29	0.0002	192	-0.75	0.25	0.0015
	Baselin	Wks 1-2 Avg	201	-0.34	211	-0.61	0.27	<0.0001	202	-0.68	0.35	<0.0001	207	-0.68	0.33	<0.0001
Itchy Nose,		Baseline	201	1.88	211	1.80			202	1.86			207	1.87		
Palate, and/or	Change	Week 1	201	-0.35	210	-0.62	0.31	<0.0001	201	-0.70	0.37	<0.0001	207	-0.71	0.36	<0.0001
Throat	From	Week 2	179	-0.60	193	-0.77	0.21	0.0050	184	-0.91	0.32	<0.0001	192	-0.83	0.23	0.0029
	Baselin	Wks 1-2 Avg	201	-0.45	211	-0.68	0.28	<0.0001	202	-0.79	0.36	<0.0001	207	-0.75	0.32	<0.0001
Congestion		Baseline	201	1.78	211	1.70			200	1.77			206	1.76		
	Change	Week 1	201	-0.15	210	-0.30	0.18	0.0028	199	-0.30	0.16	0.0083	205	-0.30	0.15	0.0114
	From	Week 2	179	-0.29	193	-0.38	0.14	0.0564	182	-0.45	0.18	0.0150	191	-0.41	0.13	0.0695
	Baselin	Wks 1-2 Avg	201	-0.20	211	-0.34	0.17	0.0052	200	-0.36	0.16	0.0076	206	-0.33	0.14	0.0199

Table A- 5: Study 81 Shift Table for Individual Symptom Scores

Study 81: Percent of Patients That Shifted Categories
Change from Baseline 2-Week Average Scores Rounded to Nearest Integer
TSS Divided into Categories of 4 units Each (ie: a change of 4 units = a shift of 1 category)

	Categories Shifted	8 AM Instantaneous			8 AM Reflective			8 PM Reflective		
		Placebo	120 mg	180 mg	Placebo	120 mg	180 mg	Placebo	120 mg	180 mg
TSS	-3	0	0	0	0	0	0	0	0	0
	-2	1	3	3	1	2	4	1	4	6
	-1	35	38	39	34	38	40	31	37	41
	0	54	55	51	53	53	49	53	53	43
	1	10	4	7	12	7	7	14	7	9
	2	0	0	0	1	0	0	0	0	0
Sneezing	-3	0	0	0	0	0	0	0	0	0
	-2	2	3	5	1	3	6	1	5	4
	-1	27	32	31	24	33	30	26	35	35
	0	57	59	56	60	56	55	57	53	52
	1	13	6	8	14	8	10	14	7	10
	2	1	0	0	1	0	0	1	0	0
Rhinorrhea	-3	0	0	0	0	0	0	0	0	0
	-2	1	3	4	0	0	0	1	4	6
	-1	27	25	31	1	3	2	25	29	27
	0	60	65	57	26	27	29	58	59	57
	1	11	6	9	60	61	60	14	7	10
	2	1	0	0	2	0	0	1	1	1
Itchy, Watery,	-3	0	0	0	0	0	0	0	0	0
Red Eyes	-2	2	3	4	2	3	5	3	4	6
	-1	31	33	33	30	31	32	29	32	35
	0	57	57	55	57	60	56	56	55	50
	1	10	6	7	10	6	7	11	9	8
	2	0	0	0	1	0	0	1	0	1
Itchy Nose,	-3	0	0	1	0	0	1	0	0	1
Mouth, Throat,	-2	3	2	4	2	3	5	3	6	6
and/or Ears	-1	32	35	37	31	31	33	30	30	36
	0	55	56	51	55	58	55	54	57	46
	1	9	7	6	12	7	6	12	6	11
	2	0	0	0	0	0	0	1	0	0
Congestion	-3	0	0	0	0	0	0	0	0	0
	-2	1	1	1	1	1	1	1	1	3
	-1	21	18	23	20	17	22	22	22	24
	0	68	71	63	68	72	65	61	66	60
	1	10	9	12	11	10	11	16	10	11
	2	1	0	1	0	0	1	0	0	2

6.3 Seasonal Allergic Rhinitis Pediatric

Table A- 6: Study 77 Primary & Secondary Endpoints 7 PM Reflective Scores

Symptom		Timepoint	Placebo n=105 MEAN	15 mg BID n=105 MEAN	DIFF	P-VAL	30 mg BID n=100 MEAN	DIFF	P-VAL	60 mg BID n=101 MEAN	DIFF	P-VAL
TSS		Baseline	7.87	7.50			7.58			7.66		
	Change	Week 1	-0.38	-1.38	**1.13**	0.0005	-1.11	**0.91**	0.0054	-1.32	**1.04**	0.0014
	From	Week 2	-1.25	-1.89	**0.98**	0.0121	-1.64	**0.68**	0.0789	-1.86	**0.82**	0.0366
	Baseline	Wks 1-2 Avg	-0.85	-1.62	**0.99**	0.0023	-1.41	**0.81**	0.0138	-1.59	**0.89**	0.0064
Sneezing		Baseline	1.83	1.82			1.81			1.75		
	Change	Week 1	-0.03	-0.41	**0.39**	0.0002	-0.34	**0.34**	0.0010	-0.26	**0.27**	0.0079
	From	Week 2	-0.26	-0.49	**0.28**	0.0183	-0.46	**0.25**	0.0414	-0.42	**0.23**	0.0614
	Baseline	Wks 1-2 Avg	-0.15	-0.46	**0.33**	0.0012	-0.41	**0.30**	0.0035	-0.34	**0.24**	0.0170
Rhinorrhea		Baseline	2.24	2.05			2.12			2.09		
	Change	Week 1	-0.09	-0.22	**0.22**	0.0370	-0.12	**0.10**	0.3259	-0.26	**0.25**	0.0189
	From	Week 2	-0.31	-0.30	**0.12**	0.3185	-0.22	**0.01**	0.9515	-0.37	**0.15**	0.2199
	Baseline	Wks 1-2 Avg	-0.21	-0.26	**0.16**	0.1310	-0.18	**0.06**	0.5466	-0.32	**0.20**	0.0581
Itchy, Watery Red Eyes		Baseline	1.80	1.66			1.56			1.80		
	Change	Week 1	-0.15	-0.31	**0.22**	0.0358	-0.25	**0.21**	0.0415	-0.43	**0.28**	0.0067
	From	Week 2	-0.39	-0.47	**0.18**	0.1383	-0.40	**0.16**	0.1792	-0.57	**0.21**	0.0746
	Baseline	Wks 1-2 Avg	-0.28	-0.38	**0.19**	0.0665	-0.33	**0.19**	0.0613	-0.50	**0.24**	0.0209
Itchy Nose, Mouth, Throat and/or Ears		Baseline	2.00	1.98			2.09			2.01		
	Change	Week 1	-0.11	-0.43	**0.33**	0.0008	-0.40	**0.28**	0.0045	-0.37	**0.26**	0.0075
	From	Week 2	-0.28	-0.64	**0.39**	0.0010	-0.56	**0.27**	0.0222	-0.49	**0.23**	0.0569
	Baseline	Wks 1-2 Avg	-0.21	-0.53	**0.33**	0.0006	-0.49	**0.27**	0.0051	-0.44	**0.24**	0.0155
Nasal Congestion		Baseline	2.28	2.16			2.19			2.23		
	Change	Week 1	-0.07	-0.14	**0.11**	0.2213	-0.10	**0.10**	0.3128	-0.04	**0.01**	0.9316
	From	Week 2	-0.21	-0.23	**0.06**	0.5844	-0.24	**0.11**	0.3412	-0.22	**0.04**	0.7086
	Baseline	Wks 1-2 Avg	-0.16	-0.17	**0.06**	0.5103	-0.18	**0.10**	0.2944	-0.13	**0.02**	0.8691

Table A- 7: Study 77 Primary & Secondary Endpoints 7 AM Instantaneous Scores

Instantaneous 7 AM Symptom Assessment	Timepoint	Placebo n=105 MEAN	15 mg BID n=105 MEAN	15 mg BID n=105 DIFF	15 mg BID n=105 P-VAL	30 mg BID n=98 MEAN	30 mg BID n=98 DIFF	30 mg BID n=98 P-VAL	60 mg BID n=101 MEAN	60 mg BID n=101 DIFF	60 mg BID n=101 P-VAL
TSS	Baseline	6.94	6.36			6.54			6.65		
	Change	-0.29	-0.93	**0.92**	0.0020	-0.78	**0.77**	0.0106	-0.90	**0.76**	0.0114
Sneezing	Baseline	1.45	1.32			1.34			1.37		
	Change	0.00	-0.23	**0.29**	0.0022	-0.10	**0.17**	0.0773	-0.17	**0.21**	0.0295
Rhinorrhea	Baseline	2.00	1.82			1.96			1.93		
	Change	-0.03	-0.09	**0.15**	0.1254	-0.10	**0.11**	0.2616	-0.14	**0.14**	0.1335
Itchy Watery, Red Eyes	Baseline	1.65	1.47			1.41			1.52		
	Change	-0.15	-0.27	**0.20**	0.0362	-0.26	**0.23**	0.0212	-0.25	**0.15**	0.1260
Itchy Nose, Mouth, Throat &/or Ears	Baseline	1.84	1.75			1.82			1.83		
	Change	-0.10	-0.34	**0.29**	0.0032	-0.33	**0.27**	0.0063	-0.35	**0.26**	0.0079
Nasal Congestion	Baseline	2.25	2.22			2.17			2.34		
	Change	-0.09	-0.18	**0.09**	0.3076	-0.11	**0.07**	0.4287	-0.17	**0.05**	0.6177

Table A- 8: Study 77 Primary & Secondary Endpoints 7 PM Instantaneous Scores

Instantaneous 7 PM Symptom Assessment	Timepoint	Placebo n=105 MEAN	15 mg BID n=105 MEAN	15 mg BID n=105 DIFF	15 mg BID n=105 P-VAL	30 mg BID n=98 MEAN	30 mg BID n=98 DIFF	30 mg BID n=98 P-VAL	60 mg BID n=101 MEAN	60 mg BID n=101 DIFF	60 mg BID n=101 P-VAL
TSS	Baseline	6.86	6.10			6.11			6.74		
	Change	-0.50	-0.88	**0.77**	0.0168	-0.70	**0.64**	0.0484	-1.23	**0.82**	0.0116
Sneezing	Baseline	1.46	1.25			1.21			1.34		
	Change	-0.07	-0.19	**0.23**	0.0200	-0.12	**0.20**	0.0521	-0.24	**0.24**	0.0147
Rhinorrhea	Baseline	1.91	1.74			1.74			1.88		
	Change	-0.09	-0.13	**0.13**	0.2087	-0.05	**0.08**	0.4617	-0.26	**0.19**	0.0631
Itchy Watery, Red Eyes	Baseline	1.67	1.42			1.36			1.62		
	Change	-0.22	-0.26	**0.16**	0.1220	-0.25	**0.18**	0.0904	-0.33	**0.13**	0.2192
Itchy Nose, Mouth, Throat &/or Ears	Baseline	1.83	1.68			1.80			1.90		
	Change	-0.11	-0.31	**0.25**	0.0083	-0.28	**0.20**	0.0377	-0.40	**0.26**	0.0075
Nasal Congestion	Baseline	2.09	1.92			1.90			2.11		
	Change	-0.07	-0.03	**0.04**	0.6821	-0.04	**0.08**	0.4013	-0.12	**0.05**	0.6038