U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	STN 103976 s5211
Drug Name:	Xolair (omalizumab)
Indication(s):	Treatment of adults and adolescents (12 years of age and above) with chronic idiopathic urticarial who remain symptomatic despite H1 antihistamine treatment
Applicant:	Genentech | Novartis
Date(s):	Letter Date: July 25, 2013
	PDUFA Due Date: May 25, 2014
Review Priority:	Standard
Biometrics Division:	Division of Biometrics 2
Statistical Reviewer:	Ruthanna Davi
Concurring Reviewers:	Joan Buenconsejo, Statistics Team Leader
Medical Division:	Division of Pulmonary Allergy Rheumatology Products
Clinical Team:	Sofia Chaudhry, Medical Reviewer
	Susan Limb, Medical Team Leader
Project Manager:	Colette Jackson

Keywords: phase 3

Table of Contents

1 EXECUTIVE SUMMARY

From a statistical perspective, studies Q4881g and Q4882g each demonstrate statistically significant effects on the primary efficacy endpoint, the change from baseline to week 12 in weekly itch severity score, for both the Xolair 300 mg and Xolair 150 mg groups. Similar demonstration of efficacy for the Xolair 75 mg group was not achieved. Conclusions regarding the comparisons of each Xolair dose group to placebo in terms of the secondary efficacy endpoints were generally consistent with and supportive of those of the primary efficacy endpoint. The demonstration of efficacy for Xolair 300 mg and Xolair 150 mg in terms of the primary efficacy endpoint are not sensitive to the methods applied for missing data. Statistical methods that appropriately account for the adaptive randomization were also supportive of these conclusions and in fact yielded nearly identical results to traditional statistical tests.
No meaningful statistically significant differences in the treatment effect in terms of the primary efficacy endpoint across gender, race, age, or baseline IGE level were identified.

2 INTRODUCTION

Xolair was FDA approved on June 20, 2003 for treatment of adults and adolescents (12 years of age and above) with moderate to severe persistent asthma who have a positive skin test or in vitro reactivity to a perennial aeroallergen and whose symptoms are inadequately controlled with inhaled corticosteroids.

The current submission provides data relevant to the use of Xolair for the treatment of adults and adolescents (12 years of age and above) with chronic idiopathic urticaria who remain symptomatic despite H1 antihistamine treatment.

2.1 Overview

In the current submission, the sponsor has provided the results of two phase 3 studies (titled and numbered as follows) with the intention of supporting the demonstration of efficacy of Xolair for treatment of adults and adolescents (12 years of age and above) with chronic idiopathic urticarial (CIU) who remain symptomatic despite H1 antihistamine treatment.

- A Phase III, Multicenter, Randomized, Double-blind, Placebo-controlled, Dose-ranging Study to Evaluate the Efficacy and Safety of Xolair (omalizumab) in Patients with Chronic Idiopathic Urticaria (CIU) Who Remain Symptomatic Despite Antihistamine Treatment (H1)" (Q4881g)
- A Phase III, Multicenter, Randomized, Double-blind, Dose-Randing, Placebo-controlled, Study to Evaluate the Efficacy, Response Duration and Safety of Xolair (omalizumab) in Patients with Chronic Idiopathic Urticaria (CIU) Who Remain Symptomatic Despite Antihistamine Treatment (H1)" (Q4882g)

Communication with the sponsor regarding these protocols and the development plan is documented under BB IND 101612 and occurred between 2008 and 2013. A Pre-IND meeting, an End-of-Phase 2 (EOP2) meeting, and a pre-BLA meeting were held April 8, 2008, May 7,

2010, and April 16, 2013, respectively. Additional written communication regarding the statistical analysis plans were also exchanged regarding this program in July and August of 2012. The key statistical agreements and recommendations made between the sponsor and FDA that are relevant to the review of studies Q4881g and Q4882g are summarized below.

- Discussion or written communication regarding the choice of the primary or co-primary efficacy endpoints occurred in connection with the pre-IND and EOP2 meetings as well as in a post-EOP2-meeting written communication. Agreement was reached among the sponsor and FDA that the itch intensity score (from administration of the Urticaria Activity Score (UAS7) instrument) could serve as a primary efficacy endpoint and the hives component of the UAS7 instrument would be considered a supportive endpoint. This agreement was implemented by the sponsor in studies Q4881g and Q4882g.
- Discussion or written communication regarding the methods for addressing missing data in the primary and secondary efficacy endpoints occurred in connection with the EOP2 and pre-BLA meetings. Although the sponsor initially proposed a last-observation-carried-forward (LOCF) approach, agreement was reached among the sponsor and FDA that a baseline-observation-carried-forward (BOCF) approach would be used. A BOCF approach is desirable in this setting in that patients who discontinue treatment (for lack of efficacy or unwillingness to tolerate some toxicity) represent a failure of the study treatment in that patient so that imputation of the baseline value (likely a relatively bad value) is appropriate. At the time of the pre-BLA meeting, the FDA noted this previous commitment but requested that since BOCF is a single imputation procedure, the sponsor should consider providing sensitivity analyses that adequately estimate the variance associated with the treatment effect (e.g., multiple imputation approach) but that do not perpetuate the treatment effect. As previously agreed, the sponsor utilized a BOCF approach as the primary approach to missing data in the current submission. Analyses of the primary efficacy endpoint utilizing a LOCF approach as well as utilizing a mixed-model-for-repeated-measures (MMRM) were provided by the sponsor as sensitivity analyses. From a theoretical statistical perspective, neither of these sensitivity analyses adequately captures the variance associated with the treatment effect while also not relying on assumptions that perpetuate the treatment effect. From a practical perspective; however, the differences between treatment groups in the primary efficacy endpoint in studies Q4881g and Q4882g are highly statistically significant when utilizing the pre-specified BOCF approach so that it is unlikely that introduction of a reasonable amount of variance associated with the treatment effect would change the qualitative conclusions regarding the significance of the treatment effect. (Refer to section 3.2.4 for further comment on missing data in studies Q4881g and Q4882g.)
- In response to the sponsor's request for review of the statistical analysis plans, the FDA noted that a dynamic randomization scheme was used to randomly assign treatments and requested re-randomization tests for the primary and secondary efficacy analysis. The sponsor agreed to this request and provided these analyses in the clinical study reports. (Refer to section 3.2.4 for comment on the re-randomization tests in studies Q4881g and Q4882g.)
- Also in response to the sponsor's request for review of the statistical analysis plans, the FDA noted that the hierarchical analyses planned for the secondary efficacy endpoints (that allow testing of the ordered secondary endpoints for each dose versus placebo when

the comparison of only that dose to placebo for the primary endpoint is significant) does not completely control the type I error since there are three doses being examined. In response, the sponsor agreed that the multiplicity plan for the secondary endpoints does not strongly control the overall type I error rate among the three doses; however, because it does strongly control the type I error rate within each dose, the sponsor continued to consider it a reasonable approach and implemented it in the current submission without modification. (Refer to section 3.2.4 for further comment on type I error control for the secondary endpoints in studies Q4881g and Q4882g.)

2.2 Data Sources

The study report, protocol, and statistical analysis plan for studies Q4881g and Q4882g were utilized in the review of this submission. The following data sets were submitted electronically and utilized in the review of this submission.

\\cdsesub1\bla\ectd_submissions\stn103976\0348\m5\datasets\q4881g\analysis\pat.xpt
\\cdsesub1\bla\ectd_submissions\stn103976\0348\m5\datasets\q4881g\analysis\pateff.xpt
\\cdsesub1\bla\ectd_submissions\stn103976\0348\m5\datasets\q4882g\analysis\pat.xpt
\\cdsesub1\bla\ectd_submissions\stn103976\0348\m5\datasets\q4882g\analysis\pateff.xpt

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

The quality and integrity of the submitted data (i.e. study reports, protocol, statistical analysis plan, and electronic data sets) were adequate for review.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

Studies Q4881g and Q4882g were similarly designed and were multicenter, randomized, double-blind, parallel-group, dose-ranging, and placebo-controlled studies in patients aged 12 to 75 years with chronic idiopathic urticaria (CIU) who remained symptomatic despite standard-dosed H1 antihistamine treatment. The primary objective of each of the studies was to assess the efficacy of Xolair compared with placebo in patients with refractory CIU receiving concomitant H1 antihistamine therapy.

For each study, eligible subjects were patients aged 12 to 75 years with chronic idiopathic urticaria (CIU) who remained symptomatic despite standard-dosed H1 antihistamine treatment. Subjects were required to have had a clinic-established urticarial activity score (UAS) ≥4 based on the 12 hours prior to either day -14 or day -7, used an approved dose of an H1 antihistamine for treatment of CIU at day -7 and for at least 3 consecutive days immediately prior to day -14, and demonstrated willingness and ability to complete the electronic symptom diary twice daily throughout the two week screening period. At baseline (day 1) subjects were randomly assigned

(in a 1:1:1:1 ratio) using a hierarchical dynamic randomization scheme (described below) to one of the following treatment groups. Randomization was stratified by baseline weekly itch severity score, baseline weight, and study site. For the first 12 weeks of the double-blind treatment period, the time of the primary efficacy assessment, subjects were required to maintain stable doses of their pre-randomization H1 antihistamine treatment.

- Placebo subcutaneous injection every 4 weeks during the 24-week for study Q4881g and 12-week for study Q4882g double blind treatment period
- Xolair 75 mg subcutaneous injection every 4 weeks during the 24-week for study Q4881g and 12-week for study Q4882g double blind treatment period
- Xolair 150 mg subcutaneous injection every 4 weeks during the 24-week for study Q4881g and 12-week for study Q4882g double blind treatment period
- Xolair 300 mg subcutaneous injection every 4 weeks during the 24-week for study Q4881g and 12-week for study Q4882g double blind treatment period

Treatment randomization was performed by using an interactive voice response system (IVRS). In order to assure relatively even treatment balance overall and within the stratification factors, subject allocation to a treatment group was performed using a biased-coin assignment. The desired balance between treatment groups was 1:1:1:1 for each Xolair dose and placebo. The treatment-balancing algorithm utilized the following in hierarchial order: overall balance (imbalance threshold 4), baseline weekly itch score (<13 versus ≥13 with imbalance threshold of 3), baseline body weight (<80 kg versus ≥80 kg with imbalance threshold of 3) and center (imbalance threshold 1).

The primary efficacy endpoint was the change from baseline in the weekly itch severity score (a component of the UAS7) at week 12. Itch severity was to be recorded twice daily (morning and evening) on a scale of 0 (none) to 3 (severe). The daily itch severity score is the average of the morning and evening scores. When either the morning or evening score is missing, the non-missing itch severity score for that day will be used as the daily itch severity score and when both the morning and evening itch scores are missing, the daily itch score will be considered missing. The weekly itch severity score is the sum of the daily itch severity over that week so that the range for the weekly itch severity score is from 0 to 21. If there are less than 7 but at least 4 non-missing daily itch severity scores available, the weekly itch severity score is the prorated average of those scores. If there are less than 4 non-missing itch severity scores, the weekly itch severity score is considered missing for that week.

The secondary efficacy endpoints were

- Change from baseline in UAS7 at week 12
  The UAS7 weekly score is defined as the sum, across seven days, of the daily averages of morning and evening scores of a composite score of the severity of the number of hives (scale of 0 (none) to 3 (severe)) and the intensity of the itch (scale of 0 (none) to 3 (intense)). The range of the daily averages is from 0 to 6 so that the range for the weekly UAS7 scores is from 0 to 42. Missing data is imputed in an analogous way to the primary efficacy endpoint.
- Change from baseline in the weekly number of hives score at week 12

The weekly number of hives score is defined as the sum, across seven days, of the daily averages of morning and evening scores of the number of hives (scale of 0 (none) to 3 (>12)). Thus the range for the weekly UAS7 scores is from 0 to 21. Missing data is imputed in an analogous way to the primary efficacy endpoint.

- Time to weekly itch severity score minimally important difference response by week 12
  Weekly itch severity score minimally important difference response is defined as a reduction from baseline in weekly itch severity score of ≥5 points.
- Proportion of patients with UAS7 ≤ 6 at week 12
  Week 12 UAS7 is defined as above and then dichotomized at a threshold of 6. Subjects missing week 12 UAS7 score are classified as non-responders.
- Proportion of weekly itch severity score minimally important difference responders at week 12
  Weekly itch severity score is defined as above and then dichotomized at a threshold of 5. Subjects missing week 12 itch severity score are classified as non-responders.
- Change from baseline in weekly size of the largest hive score
  The weekly size of the largest hives score is defined as the sum, across seven days, of the daily averages of morning and evening scores of the size of the largest hive (scale of 0 (none) to 3 (>2.5 cm)). Thus the range for the weekly UAS7 scores is from 0 to 21. Missing data is imputed in an analogous way to the primary efficacy endpoint.
- Change from baseline in health-related quality-of-life as measured by the Dermatology Life Quality Index(DLQI) at week 12
  The DLQI is a 10-item dermatology-specific health-related quality of life measure. Patients rate their dermatology symptoms as well as the impact of their skin condition on various aspects of their lives over the last week. The DLQI is calculated by summing the score for each question resulting in a minimum of 0 and a maximum of 30. The higher the score, the more quality of life is impaired.
- Proportion of angioedema-free days from week 4 to week 12 of therapy
  The occurrence of angioedema is recorded once daily in the evening. The proportion of angioedema-free days from week 4 to week 12 is defined as the number of days for which the subject indicated a "no" response divided by the total number of days with a non-missing entry.
- Proportion of complete responders at week 12 (pre-specified as a secondary efficacy endpoint in study Q4881g only)
  Week 12 UAS7 is defined as above. Subjects will be classified as a complete responder when the week 12 UAS7 score is 0. Subjects missing week 12 UAS7 score are classified as non-responders.

The primary efficacy endpoint and the secondary efficacy endpoints were derived from data collected via the Urticaria Patient Daily Diary with an electronic handheld device. Subjects were instructed to complete this electronic diary twice a day for the duration of the study.

3.2.2 Statistical Methodologies

The protocol specified that the efficacy analyses were to be performed using the modified-intent-to-treat (mITT) population defined as all randomized subjects who received at least one dose of

study drug. Subjects who discontinued from study treatment or took excluded therapy were to be considered missing for purposes of the efficacy analyses.

The primary efficacy endpoint, the change from baseline at week 12 in the weekly itch severity score, was to be compared between each of the Xolair dose and placebo groups using the protocol-specified analysis of covariance (ANCOVA) controlling for baseline weekly itch severity score (<13 vs. ≥13), and baseline weight (<80 kd vs. ≥80kg). Missing week 12 weekly itch severity scores were imputed by the pre-specified method of carrying forward the baseline weekly itch severity score. In pre-submission communications and since BOCF is a single imputation procedure, the FDA requested that the sponsor consider sensitivity analyses that adequately estimate the variance associated with the treatment effect (e.g., multiple imputation approach) but that do not perpetuate the treatment effect. Analyses of the primary efficacy endpoint utilizing a LOCF approach as well as utilizing MMRM (fitting all observed weekly itch severity scores from baseline to week 12 controlling for baseline weekly itch severity score (<13 vs. ≥13) and baseline weight (<80 kg vs. ≥80 kg) for each Xolair dose versus placebo comparison separately) were provided by the sponsor as pre-specified sensitivity analyses. From a theoretical statistical perspective, neither of these sensitivity analyses adequately captures the variance associated with the treatment effect while also not relying on assumptions that perpetuate the treatment effect. (Refer to section 3.2.4 for further comment on missing data in studies Q4881g and Q4882g.) In response to an FDA pre-submission request and to account for the use of a hierarchical randomization scheme, a sensitivity analysis on the primary efficacy endpoints utilizing a re-randomization test was provided by the sponsor. (Refer to section 3.2.4 for comment on the re-randomization tests in studies Q4881g and Q4882g.)

Table 1 provides the statistical procedures utilized for analyzing the secondary efficacy endpoints. In addition, in response to an FDA pre-submission request, the sponsor provided re-randomization tests for each of these comparisons.

Table 1 Statistical Analysis of Secondary Endpoints*

Secondary Endpoint	Statistical Test	Baseline Covariate / Stratification Variables	Summary of Handling of Missing Data (Imputation Method)
Change from baseline in UAS7 at week 12	ANCOVA	Baseline UAS7 (categorized by median) and weight (categorized by 80 kg)	Baseline-observation-carried-forward
Change from baseline in weekly number of hives score at week 12	ANCOVA	Baseline weekly number of hives score (categorized by median) and weight (categorized by 80 kg)	Baseline-observation-carried-forward
Time to MID response in weekly itch severity score by week 12	Cox proportional hazards model	Baseline weekly itch severity score (categorized by 13) and weight (categorized by 80 kg)	Censored at date of last non-missing weekly itch severity score in the absence of MID response
Proportion of patients with UAS7≤6 at week 12	Cochran-Mantel-Haenszel	Baseline UAS7 (categorized by median) and weight (categorized by 80 kg)	Classified as non-responder
Proportion of weekly itch severity score MID responders at week 12	Cochran-Mantel-Haenszel	Baseline weekly itch severity score (categorized by 13) and weight (categorized by 80 kg)	Classified as non-responder
Change from baseline in weekly size of largest hive score at week 12	ANCOVA	Baseline weekly size of largest hive (categorized by median) and weight (categorized by 80 kg)	Baseline-observation-carried-forward
Change from baseline in DLQI at week 12	ANCOVA	Baseline Dermatology Life Quality Index (stratified by median) and weight stratified by 80 kg)	Baseline-observation-carried-forward
Proportion of angioedema-free days from week 4 to week 12 of therapy	Van Elteren's test	Presence of angioedema at baseline (yes/no) and weight (stratified by 80 kg)	No imputation
Proportion of complete responders (UAS7=0) at week 12 (Q4881g only)	Cochran-Mantel-Haenszel	Baseline UAS7 (categorized by median) and weight (categorized by 80 kg)	Classified as non-responder

*Source: Adapted from Table 4 of Clinical Study Reports for Studies Q4881g and Q4882g

To maintain an overall type I error rate of 0.05 (two-sided) for the primary efficacy endpoint across the three Xolair dose levels, the testing of the primary efficacy endpoint was to be conducted in the following order, proceeding to the next step only when the previous is statistically significant with α=0.05 (two-sided): (1.) Xolair 300 mg to placebo, (2.) Xolair 150 mg to placebo, and (3.) Xolair 75 mg to placebo. A hierarchical analysis of the secondary efficacy endpoints (in the order listed in Table 1) was to be performed for each dose group found to be statistically significant different from placebo in the primary efficacy endpoint. All tests of the secondary efficacy endpoints were to be conducted using a significance level of 0.05 (two-sided). Note that the hierarchical testing of the secondary efficacy endpoints is independent between different dose levels. In pre-submission communications, the FDA noted that the

hierarchical analyses planned for the secondary efficacy endpoints does not completely control the type I error since there are three doses being examined. In response, while the sponsor agreed that the type I error for the secondary efficacy endpoints would not be strongly controlled among the three doses; because the approach does strongly control the type I error rate within each dose, the sponsor continued to consider this a reasonable approach. (Refer to section 3.2.4 for further comment on type I error control for the secondary endpoints in studies Q4881g and Q4882g.)

According to the sponsor, the sample size for studies Q4881g and Q4882g were determined primarily based on safety and regulatory considerations. For purposes of demonstration of efficacy, 300 patients (randomized 1:1:1:1 among treatment groups) were expected to provide approximately 98% power to detect a difference in the treatment effect in the primary efficacy endpoint with a two-sided 0.05 significance level (assuming a mean change from baseline in the primary efficacy endpoint of 9 points and 3.5 points for the Xolair and placebo groups, respectively, with a common standard deviation of 6 points, all assumptions which were largely confirmed by studies Q4881g and Q4882g). In such a setting, a careful understanding of a "highly significant" p-value is needed. In general, with respect to a comparison between treatment groups, a highly significant p-value may be a result of the magnitude of the true difference between treatment groups, the level of variability in the efficacy measure, and/or the number of subjects studied. While it may seem natural to assume that a highly significant p-value is an indication that the magnitude of the treatment effect is large, this may or may not be the case. Rather the p-value is a measure of the certainty of the finding. With studies Q4881g and Q4882g, the certainty of the finding is great since the number of subjects studied is more than what would have normally been required, to achieve 80% power, for example. Estimation of the treatment effect is correspondingly precise. However, the magnitude of the treatment effect associated with a highly significant p-value is not necessarily large. The reader should avoid inaccurate interpretation of the p-value and rely on the point estimate for the difference between treatment groups and the corresponding confidence interval for estimation of the magnitude of the treatment effect.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

As described in Table 2, 319 and 323 subjects were randomly assigned in a 1:1:1:1 ratio to receive placebo, Xolair 75 mg, Xolair 150 mg, and Xolari 300 mg in studies Q4881g and Q4882g, respectively. One subject in each study did not receive study treatment and therefore was not included in the mITT group. Early study treatment discontinuation was most common in the placebo group and ranged from 10% to 24% across treatment groups in study Q4881g. The most frequent reasons for early study treatment discontinuation in study Q4881g were adverse event and disease progression. As might be expected due to the shorter treatment period associated with study Q4882g, early study treatment discontinuation was less frequent in study Q4882g than Q4881g and ranged from 3% to 10% across treatment groups. The data in Table 2 reflect treatment discontinuation rates throughout the studies and do not account for the timing of the primary and secondary efficacy evaluations at week 12 so that the importance of these events may not be directly relevant to the demonstration of efficacy.

Table 2: Subject Disposition (ITT)

	Study Q4881g				Study Q4882g			
	Placebo	Xolair 75 mg	Xolair 150 mg	Xolair 300 mg	Placebo	Xolair 75 mg	Xolair 150 mg	Xolair 300 mg
Subjects Randomized	80	78	80	81	79	82	83	79
mITT	**80 (100%)**	**77 (99%)**	**80 (100%)**	**81 (100%)**	**79 (100%)**	**82 (100%)**	**82 (99%)**	**79 (100%)**
Did not receive study drug		1 (1%)					1 (1%)	
Early Study Trt. Disc.	**19 (24%)**	**10 (13%)**	**16 (20%)**	**8 (10%)**	**3 (4%)**	**8 (10%)**	**5 (6%)**	**2 (3%)**
Reason for Early Study Trt. Disc.								
Adverse Event	7 (9%)	2 (3%)	4 (5%)	2 (3%)	0	3 (4%)	2 (3%)	1 (1%)
Lost to follow-up	1 (1%)	0	0	0	1 (1%)	0	1 (1%)	0
Physician decision	0	2 (3%)	2 (3%)	1 (1%)	0	1 (1%)	1 (1%)	0
Pt/legal guardian dec.	1 (1%)	3 (4%)	5 (6%)	3 (4%)	1 (1%)	1 (1%)	0	1 (1%)
Disease progression	10 (13%)	3 (4%)	5 (6%)	2 (3%)	1 (1%)	3 (4%)	1 (1%)	0

Source: Adapted from Table 6 of Clinical Study Reports for Studies Q4881g and Q4882g

The double-blind treatment period was 24 and 12 weeks for studies Q4881g and Q4882g, respectively. However, the primary efficacy evaluation was at 12 weeks for each study. Table 3 displays the proportion of subject with sufficiently complete primary and secondary efficacy data at week 12 (so that imputation was not necessary) for the mITT group.

Table 3: Analysis Groups / Reason for Incomplete Week 12 Efficacy Data (mITT)

	Study Q4881g				Study Q4882g			
	Placebo N=80	Xolair 75 mg N=77	Xolair 150 mg N=80	Xolair 300 mg N=81	Placebo N=79	Xolair 75 mg N=82	Xolair 150 mg N=82	Xolair 300 mg N=79
Had complete primary and secondary efficacy data at wk 12	**64 (80%)**	**66 (86%)**	**64 (80%)**	**73 (90%)**	**69 (87%)**	**70 (85%)**	**73 (89%)**	**74 (94%)**
Discontinued from trt	14 (18%)	7 (9%)	11 (14%)	5 (6%)	3 (4%)	8 (10%)	5 (6%)	2 (3%)
Took excluded meds (and did not discontinue from trt)	1 (1%)	2 (3%)	3 (4%)	1 (1%)	3 (4%)	3 (4%)	4 (5%)	3 (4%)
Less than 4 days of diary records for week1 2 (and did not take excluded meds or discontinue from trt)	1 (1%)	2 (3%)	2 (3%)	2 (3%)	4 (5%)	1 (1%)	0 (0%)	0 (0%)

Source: Adapted from Tables 7 and 27of Clinical Study Reports for Studies Q4881g and Q4882g

Among randomized subjects, approximately 80% to 90% of subjects in study Q4881g and 85% to 94% of subjects in study Q4882g had complete week 12 primary and secondary efficacy data. The reasons for missing information at week 12 included premature discontinuation from study treatment, took protocol-specified excluded medication, and insufficient diary data recorded in week 12. These exceptions were fairly balanced across treatment groups within each study and therefore are not expected to have overly influenced the assessment of efficacy.

Demographic and baseline characteristics by treatment group for studies Q4881g and Q4882g are described in Table 4. As would be expected because of the random treatment assignment, these factors were generally well-balanced across treatment groups.

Table 4: Subject Demographics and Baseline Characteristics (mITT)

		Study Q4881g				Study Q4882g			
		Placebo N=80	Xolair 75 mg N=77	Xolair 150 mg N=80	Xolair 300 mg N=81	Placebo N=79	Xolair 75 mg N=82	Xolair 150 mg N=82	Xolair 300 mg N=79
Age (years)	Median	37.5	41.0	43.0	42.0	43.1	36.0	43.0	43.0
	Range	13-74	13-72	12-68	14-72	17-73	14-75	14-72	15-75
Gender [n (%)]	Male	28 (35%)	22 (29%)	16 (20%)	21 (26%)	24 (30%)	21 (26%)	17 (21%)	16 (20%)
	Female	52 (65%)	55 (71%)	64 (80%)	60 (74%)	55 (70%)	61 (74%)	65 (79%)	63 (80%)
Ethnicity [n(%)]	Hispanic or Latino	7 (9%)	5 (7%)	6 (8%)	3 (4%)	6 (8%)	9 (11%)	8 (10%)	3 (4%)
	Not Hispanic or Latino	71 (89%)	71 (92%)	74 (93%)	78 (96%)	73 (92%)	73 (89%)	74 (90%)	74 (94%)
	Not available	2 (3%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (3%)
Race [n(%)]	American Indian or Alaska Native	0 (0%)	0 (0%)	1 (1%)	1 (1%)	0 (0%)	0 (0%)	1 (1%)	0 (0%)
	Asian	3 (4%)	4 (5%)	6 (8%)	1 (1%)	2 (3%)	4 (5%)	1 (1%)	2 (3%)
	Black	10 (13%)	9 (12%)	9 (11%)	5 (6%)	4 (5%)	12 (15%)	5 (6%)	7 (9%)
	Nt Hawaiian Pac Islndr	NA	NA	NA	NA	1 (1%)	0 (0%)	0 (0%)	0 (0%)
	White	64 (80%)	62 (81%)	63 (79%)	74 (91%)	70 (89%)	64 (78%)	70 (85%)	68 (86%)
	> 1 race indicated	NA	NA	NA	NA	0 (0%)	0 (0%)	2 (2%)	1 (1%)
	Not available	3 (4%)	2 (3%)	1 (1%)	0 (0%)	2 (3%)	2 (2%)	3 (4%)	1 (1%)
Weight [n(%)]	<80kg	35 (44%)	38 (49%)	40 (50%)	45 (56%)	41 (52%)	43 (52%)	41 (50%)	41 (52%)
	≥80 kg	45 (56%)	39 (51%)	40 (50%)	36 (44%)	38 (48%)	39 (48%)	41 (50%)	38 (48%)
BMI	Median	27.9	28.4	29.0	27.2	28.0	28.4	28.2	28.0
	Range	19-47	18-49	16-54	20-52	18-56	19-50	18-54	18-48
Duration of CIU (years)	Median	3.7	3.8	4.3	3.2	3.3	2.5	3.9	3.5
	Range	0.5-48.2	0.5-50.5	0.5-44.4	0.5-35.4	0.6-66.4	0.5-41.9	0.6-44.5	0.5-36.0
# previous CIU meds	Median	4.5	4.0	4.0	4.0	3.0	4.0	4.0	4.0
	Range	1-13	1-13	1-18	1-10	1-13	1-9	1-17	1-11
Previous systemic steroids for CIU	Yes	31 (39%)	41 (53%)	32 (40%)	36 (44%)				
Positive CU index test	Yes	25 (31%)	18 (23%)	16 (20%)	21 (26%)	23 (29%)	26 (32%)	27 (33%)	18 (23%)
Total IgE level (IU/mL)	Median	92	91	71	86	76	88	70	94
	Range	1-1010	1-2030	1-5000	1-2330	1-966	1-1320	1-1450	5-1040
In-clinic UAS	Median	5	5	5	5	5	6	5	5
	Range	4-6	4-6	4-6	4-6	4-6	2-6	4-6	4-6
UAS7	Median	32	32	31	32	32	32	31	29
	Range	16-42	17-42	16-42	20-42	17-42	17-42	17-42	17-42
Weekly itch severity score	<13	26 (33%)	28 (36%)	26 (33%)	28 (35%)	34 (43%)	34 (42%)	36 (44%)	37 (47%)
	≥13	54 (68%)	49 (64%)	54 (68%)	53 (65%)	45 (57%)	48 (59%)	46 (56%)	42 (53%)
Weekly number of hives score	Median	18.3	19.0	17.0	18.5	18.0	17.5	18.5	16.0
	Range	5-21	7.5-21	4.5-21	8.5-21.0	6-21	8-21	7-21	7-21
Presence of angioedema	Yes	44 (55%)	35 (46%)	38 (48%)	34 (42%)	30 (38%)	31 (38%)	38 (46%)	32 (41%)
Level of thyroperoxidase antibody	High (>34.99 U/mL)	12 (15%)	16 (21%)	10 (13%)	9 (11%)	10 (13%)	17 (21%)	17 (21%)	11 (15%)
	Normal (≤34.99 U/mL)	67 (85%)	58 (78%)	70 (88%)	72 (89%)	67 (87%)	65 (79%)	65 (79%)	64 (85%)

*Small amount (<5%) of missing data for certain endpoints ignored in calculations.
Source: Adapted from Clinical Study Reports for studies Q4881g and Q4882g, Tables 8 and 9

3.2.4 Results and Conclusions

The pre-specified primary efficacy analysis, as provided by the sponsor is shown in Table 5. The primary efficacy endpoint, the change from baseline at week 12 in the weekly itch severity score, was compared between each of the Xolair dose and placebo groups using the protocol-specified analysis of covariance (ANCOVA) controlling for baseline weekly itch severity score (<13 vs. ≥13), and baseline weight (<80 kd vs. ≥80kg). Subjects who discontinued from study treatment, took excluded therapy, or had insufficient week 12 diary data (see Table 3) were considered missing for purposes of the efficacy analyses. Missing week 12 weekly itch severity scores were imputed by the pre-specified BOCF method.

The decreases from baseline to week 12 in the mean weekly itch severity score were larger in the Xolair groups than placebo in study Q4881g. Each of these comparisons, beginning with comparison of the Xolair 300 mg group to placebo, were statistically significant, allowing, according to the pre-specified multiplicity plan, inferential hypothesis testing thru and including the comparison of Xolair 75 mg to placebo for study Q4881g. A priori estimates of the power associated with these comparisons were estimated at approximately 98% so that while the highly significant p-values are desirable in the sense that they demonstrate with great precision that the true treatment effect is beyond chance, they are not necessarily indicative of a large treatment effect. Rather the confidence intervals for the difference between treatment group means afford such estimates and should be the focus of the evaluation of the treatment effect size. The magnitude of the treatment effect was numerically larger for the Xolair 300 mg group than the Xolair 150 mg and Xolair 75 mg groups. The true treatment effect over placebo is estimated from study Q4881g, with 95% confidence, to be as small as 4.1 units and as large as 7.5 units for the Xolair 300 mg group and as small as 1.2 units and as large as 4.7 units for the Xolair 150 mg and 75 mg groups.

In study Q4882g, the decrease from baseline to week 12 in the mean weekly itch severity score was statistically significant larger for the Xolair 300 mg group than placebo, allowing, according to the pre-specified multiplicity plan, inferential hypothesis testing to continue to the Xolair 150 mg to placebo comparison. The decrease from baseline to week 12 in the mean weekly itch severity score was again statistically significant larger for the Xolair 150 mg group than placebo, allowing inferential hypothesis testing to continue to the Xolair 75 mg to placebo comparison; however the comparison of Xolair 75 mg to placebo was not statistically significant. Similarly to study Q4881g, the confidence intervals for the difference between treatment group means should be the focus of the evaluation of the effect sizes for the Xolair groups. The magnitude of the treatment effect was numerically larger for the Xolair 300 mg group than the Xolair 150 mg and Xolair 75 mg groups. The true treatment effect over placebo is estimated from study Q4882g, with 95% confidence, to be as small as 3.1 units and as large as 6.5 units for the Xolair 300 mg group and as small as 1.2 units and as large as 4.9 units for the Xolair 150 mg groups. Consistent with the statistically insignificant p-value, the 95% confidence interval for the difference between the Xolair 75 mg group and placebo included zero as a plausible value for the true treatment effect at that dose.

Table 5: Primary Efficacy Analysis: Change from Baseline in Weekly Itch Severity Score at Week 12 (mITT)

	Study Q4881g				Study Q4882g			
	Placebo N=80	Xolair 75 mg N=77	Xolair 150 mg N=80	Xolair 300 mg N=81	Placebo N=79	Xolair 75 mg N=82	Xolair 150 mg N=82	Xolair 300 mg N=79
Mean Chg from Baseline in Weekly Itch Severity Score	-3.6	(b) (4)	-6.7	-9.4	-5.1	(b) (4)	8.1	9.8
LS Mean Diff from Placebo			-3.0	-5.8			-3.0	-4.8
95% Confidence Interval			(-4.7, -1.2)	(-7.5, -4.1)			(-4.9, 1.2)	(-6.5, 3.1)
p-value			0.001	<0.0001			0.001	<0.0001

Source: Adapted from Clinical Study Reports for studies Q4881g and Q4882g, Table 12

In pre-submission communications and since BOCF is a single imputation procedure, the FDA requested that to assess the impact of missing data the sponsor consider sensitivity analyses that

adequately estimate the variance associated with the treatment effect (e.g., multiple imputation approach) but that do not perpetuate the treatment effect. Analyses of the primary efficacy endpoint utilizing a LOCF approach as well as utilizing MMRM (fitting all observed weekly itch severity scores from baseline to week 12 controlling for baseline weekly itch severity score (<13 vs. ≥13) and baseline weight (<80 kg vs. ≥80 kg) for each Xolair dose versus placebo comparison separately) were provided by the sponsor as pre-specified sensitivity analyses and are displayed in Tables 6 and 7, respectively. While results of these sensitivity analyses are largely consistent with the results of the primary BOCF analysis and are supportive of the efficacy of Xolair 300 mg and Xolar 150 mg, from a theoretical statistical perspective, neither of these sensitivity analyses adequately captures the variance associated with the treatment effect while also not relying on assumptions that perpetuate the treatment effect.

The LOCF analysis is a single imputation approach so that the variance of the treatment effect may be underestimated. The MMRM analysis more appropriately estimates the variance of the treatment effect; however, this analysis relies on an assumption that the missing data are *missing at random* or in other words, that the unobserved data is similar to observed data thus perpetuating the treatment effect found in the observed data by assuming the same to be true in the unobserved data. We acknowledge that statistical methods that address missing data and adequately capture the variance associated with the treatment effect while also not relying on assumptions that perpetuate the treatment effect are not well-developed or easily accessible at this time. Even the absence of such analyses, in this case and in the opinion of this reviewer, the highly statistically significantly treatment effects associated with the BOCF approach (Table 5) are sufficient demonstration that a positive treatment effect in terms of the change from baseline in the weekly itch severity score at week 12 for the Xolair 300 mg and Xolair 150 mg groups relative to placebo exists despite these statistical limitations. We believe that, first, the BOCF approach is likely a fair estimation of patient-level efficacy in the sense that it applies a presumably undesirable efficacy measure (i.e., the baseline score) to subjects who are unable or unwilling to continue receiving treatment in exchange for the efficacy that is being received. Second, the possibility of an underestimation of the variance of the treatment effect by utilizing a single imputation procedure is of less concern in this case due to the highly statistically significant differences between treatment groups so that while it is not readily apparent from a theoretical statistical perspective how the variance should be appropriately inflated, introduction of any reasonable additional variance is unlikely to alter the qualitative conclusions regarding the existence of a positive treatment effect for the Xolair 300 mg and Xolair 150 mg doses.

Table 6: Sensitivity (LOCF) Efficacy Analysis: Change from Baseline in Weekly Itch Severity Score at Week 12 (mITT)

	Study Q4881g				Study Q4882g			
	Placebo N=80	Xolair 75 mg N=77	Xolair 150 mg N=80	Xolair 300 mg N=81	Placebo N=79	Xolair 75 mg N=82	Xolair 150 mg N=82	Xolair 300 mg N=79
Mean Chg from Baseline in Weekly Itch Severity Score	-4.3	(b) (4)	-7.5	-10.19	-5.5	(b) (4)	-8.2	-10.1
LS Mean Diff from Placebo			-3.2	-6.0			-2.8	-4.9
95% Confidence Interval			(-5.0, -1.3)	(-7.5, -4.4)			(-4.6, -1.0)	(-6.5, -3.3)
p-value			0.0008	<0.0001			0.003	<0.0001

Source: Adapted from Clinical Study Reports for studies Q4881g and Q4882g, Table

Table 7: Sensitivity (MMRM) Efficacy Analysis: Change from Baseline in Weekly Itch Severity Score at Week 12 (mITT)

	Study Q4881g				Study Q4882g			
	Placebo N=80	Xolair 75 mg N=77	Xolair 150 mg N=80	Xolair 300 mg N=81	Placebo N=79	Xolair 75 mg N=82	Xolair 150 mg N=82	Xolair 300 mg N=79
LS Mean Diff from Placebo		(b) (4)	-3.3	-5.8		(b) (4)	-2.8	-4.8
95% Confidence Interval			(-5.2, -1.5)	(-7.4, -4.2)			(-4.7, -1.0)	(-6.5, -3.2)
p-value			0.0004	<0.0001			0.003	<0.0001

Source: Adapted from Clinical Study Reports for studies Q4881g and Q4882g, Table

In response to an FDA pre-submission request and to account for the use of a hierarchical randomization scheme, a sensitivity analysis on the primary efficacy endpoints utilizing a re-randomization test was provided by the sponsor. Results of these analyses for the primary efficacy endpoint were nearly identical to the pre-specified analysis displayed in Table 5. The statistical significance associated with the comparisons between each Xolair group and placebo remained unchanged from the pre-specified analysis at (b) (4), p=0.001, and p<0.0001 for the Xolair 75 mg, Xolair 150 mg and Xolair 300 mg comparisons in study Q4881g, respectively. Similarly, (b) (4), p=0.001, and p<0.0001 for the Xolair 75 mg, Xolair 150 mg and Xolair 300 mg comparisons in study Q4882g.

Since the Xolair to placebo group comparisons for the primary efficacy endpoint were statistically significant for all dose groups in both studies except the Xolair 75 mg to placebo comparison in study Q4882g, according to the pre-specified multiplicity plan, inferential statistical analysis may continue to the first secondary efficacy endpoint for those doses. Also according to the pre-specified multiplicity plan, inferential testing of the following hierarchical secondary efficacy endpoints may continue as long as evaluations of the previous secondary efficacy endpoints are statistically significant for that Xolair dose compared to placebo. The FDA had previously communicated with the sponsor regarding the control of type I error for the secondary efficacy endpoints. The FDA noted and the sponsor agreed that the hierarchical analyses planned for the secondary efficacy endpoints (that allow testing of the ordered secondary endpoints for each dose versus placebo when the comparison of only that dose to placebo for the primary endpoint is significant) does not completely control the type I error since there are three doses being examined. Despite their agreement with this concern, the sponsor elected to continue with this pre-specified multiplicity plan. From a statistical perspective, the type I error associated with falsely declaring statistical significance for at least one endpoint for at least one dose is greater than 0.05, however, applying a post-hoc Bonferroni correction for the three dose groups (i.e., testing hierarchically within each dose group at $\alpha=0.05/3=0.17$) does not alter the conclusions regarding the secondary efficacy endpoint for the Xolair 300 mg to placebo comparisons in either study. Comparison of the Xolair 150 mg group to placebo for four secondary endpoints in study Q4881g and one endpoint in study Q4882 that were previously considered significant under the pre-specified multiplicity plan would not be considered statistically significant when applying the conservative Bonferroni approach. In the opinion of this reviewer, in appreciation of the relatively consistent results even under the conservative Bonferroni approach and the clear dose-response displayed in the secondary efficacy endpoints, it unlikely that conclusions regarding the efficacy of Xolair in general will be inaccurate based on a single or at most a small number of falsely significant results. From a practical perspective,

the technical inadequacies of the pre-specified multiplicity plan are unlikely to have adversely altered the overall interpretation of efficacy of each Xolair dose.

The pre-specified statistical analyses of the secondary efficacy endpoints are shown in Table 8. Comparisons that are considered statistically significant (according to the pre-specified multiplicity plan) and according to the outcome of the analyses are shaded. Statistically significant benefits over placebo in terms of every secondary efficacy endpoint for both studies were observed for the Xolair 300 mg group. Similar results are observed for the Xolair 150 mg group over placebo with lack of statistical significance in three and two cases in studies Q4881g and Q4882g. Statistically significant differences from placebo in the secondary efficacy endpoints for the Xolair 75 mg group were sparse and the efficacy of Xolair at that dose is not supported.

Table 8: Pre-specified Secondary Efficacy Analyses (mITT)

	Study Q4881g				Study Q4882g			
	Placebo N=80	Xolair 75 mg N=77	Xolair 150 mg N=80	Xolair 300 mg N=81	Placebo N=79	Xolair 75 mg N=82	Xolair 150 mg N=82	Xolair 300 mg N=79
Mean Chg from Baseline to Week 12 in UAS7 (BOCF)	-8.0	(b) (4)	-14.4	-20.8	-10.4	(b) (4)	-17.9	-21.7
LS Mean Diff from Placebo			-6.5	-12.8			-7.7	-12.4
95% Confidence Interval			(-10.3, -2.8)	(-16.4, -9.2)			(-11.5, -3.9)	(-16.1, -8.7)
p-value			0.0008	<0.0001			0.0001	<0.0001
Mean Chg from Baseline to Week 12 in Weekly Number of Hives Score (BOCF)	-4.4		-7.8	-11.4	-5.2		-9.8	-12.0
LS Mean Diff from Placebo			-3.4	-6.9			-4.5	-7.1
95% Confidence Interval			(-5.6, -1.3)	(-9.1, -4.8)			(-6.7, -2.4)	(-9.3, -4.9)
p-value			0.002	<0.0001			<0.0001	<0.0001
Median Time (in weeks) to MID (reduction of ≥ 5 pts) Response in Weekly Itch Severity Score by Week 12	4.0		2.0	1.0	4.0		2.0	1.0
Hazrd Ratio versus placebo			1.5	2.3			1.6	2.1
95% Confidence Interval			(1.0, 2.1)	(1.6, 3.4)			(1.1, 2.3)	(1.5, 3.0)
p-value			0.03	<0.0001			0.01	<0.0001
Number and Proportion of Patients with UAS7≤6 at Week 12 (non-responder imputation)	9 (11%)		32 (40%)	42 (52%)	15 (19%)		35 (43%)	52 (66%)
Diff in prop (vs. placebo)			29%	42%			24%	37%
p-value			<0.0001	<0.0001			0.001	<0.0001
Number and Proportion of Weekly Itch Severity Score MID (reduction of ≥ 5 pts) Responders at Week 12 (non-responder imputation)	29 (36%)		45 (56%)	61 (75%)	38 (48%)		57 (70%)	62 (79%)
Diff in prop (vs. placebo)			20%	39%			22%	31%
p-value			0.02	<0.0001			0.005	<0.0001
Mean Chg from Baseline to Week 12 in Weekly Size of Largest Hive Score (BOCF)	-3.9		-7.0	-9.8	-4.0		-7.8	-11.0
LS Mean Diff from Placebo			-3.2	-5.7			-3.8	-7.2
95% Confidence Interval			(-5.1, -1.3)	(-7.6, 3.9)			(-5.6, -1.9)	(-9.0, -5.3)
p-value			0.001	<0.0001			<0.0001	<0.0001
Mean Chg from Baseline to Week 12 in Overall DLQI at Week 12 (Observed Data)	-6.1		-8.0	-10.3	-6.1		-8.3	-10.2
LS Mean Diff from Placebo			-1.3	-4.1			-2.5	-3.8
95% Confidence Interval			(-3.5, 0.8)	(-6.0, -2.2)			(-4.6, -0.4)	(-5.9, -1.7)
p-value			0.2	<0.0001			0.02	0.0004
Mean Prop of Angioedema Free Days from Week 4 to Week 12 (pts missing>40% of days excluded)	88%		90%	96%	89%		92%	96%
Mean Diff from Placebo			2%	4%			3%	7%
p-value (Wilcoxon test)			0.2	<0.0001			0.09	<0.0001
Number and Proportion of Complete Responders (UAS7=0) at Week 12 (non-responder imputation)	7 (9%)		12 (15%)	29 (36%)	4 (5%)		18 (22%)	35 (44%)
Diff in prop (vs. placebo)			6%	27%			17%	39%
p-value			0.2	<0.0001			0.002	<0.0001*

Source: Adapted from Clinical Study Reports for studies Q4881g and Q4882g, Tables 13 thru 20

*Proportion of Complete Responders was a pre-specified secondary efficacy endpoint in study Q4881g only. Results for study Q4882g are included because of the clinical importance of this endpoint designated by the FDA clinical team.

3.3 Evaluation of Safety

During the course of this review, no safety endpoints were identified as requiring more rigorous statistical evaluation. The reader is referred to the medical review of this application for an evaluation of the safety of Xolair.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

No meaningful statistically significant differences in the treatment effect in terms of the primary efficacy endpoint across gender, race, or age categories were identified (for gender p = 0.5, 0.1, and 0.8 for the subgroup-by-treatment interaction for the Xolair 75 mg, 150 mg, and 300 mg groups, respectively, in study Q4881g and p = (b) (4) 0.6, and 0.6 for the subgroup-by-treatment interaction for the Xolair 75 mg, 150 mg, and 300 mg groups, respectively, in study Q4882g, for race p = (b) (4), NE, and NE for the subgroup-by-treatment interaction for the Xolair 75 mg, 150 mg, and 300 mg groups, respectively, in study Q4881g and p = (b) (4) , NE, and NE for the subgroup-by-treatment interaction for the Xolair 75 mg, 150 mg, and 300 mg groups, respectively, in study Q4882g, for age p = (b) (4), 0.7, and 0.7 for the subgroup-by-treatment interaction for the Xolair 75 mg, 150 mg, and 300 mg groups, respectively, in study Q4881g and p = (b) (4), 0.01, and 0.2 for the subgroup-by-treatment interaction for the Xolair 75 mg, 150 mg, and 300 mg groups, respectively, in study Q4882g).

Nevertheless analysis of the primary efficacy endpoint, the change from baseline to week 12 in the weekly itch severity score (BOCF), is presented stratified by gender, age, and race in Table 9. The results indicate that the treatment effects of Xolair 300 mg and Xolair 150 mg over placebo are present and relatively consistent across these strata.

Table 9: Primary Efficacy Analysis: Change from Baseline in Weekly Itch Severity Score at Week 12 by Gender, Race, and Age (mITT)

	Study Q4881g				Study Q4882g			
	Placebo N=80	Xolair 75 mg N=77	Xolair 150 mg N=80	Xolair 300 mg N=81	Placebo N=79	Xolair 75 mg N=82	Xolair 150 mg N=82	Xolair 300 mg N=79
Males								
Sample Size	28	(b) (4)	16	21	24	(b) (4)	17	16
LS Mean Diff from Placebo			-0.6	-5.7			-3.9	-3.9
95% Confidence Interval			(-4.0, 2.8)	(-9.2, 2.1)			(-7.2, -0.4)	(-7.4, -0.4)
p-value			0.7	0.003			0.03	0.03
Females								
Sample Size	52		64	60	55		65	63
LS Mean Diff from Placebo			-3.9	-5.9			-2.8	-4.9
95% Confidence Interval			(-6.0, -1.8)	(-7.9, -3.9)			(-4.9, -0.6)	(-6.9, -2.9)
p-value			0.0005	<0.0001			0.01	<0.0001
Age<18								
Sample Size	4		7	2	2		2	2
LS Mean Diff from Placebo			-1.6	NE			0.3	-0.8
95% Confidence Interval			(-7.9, 4.8)	NE			(-16, 17)	(-17, 16)
p-value			0.6	NE			0.9	0.7
Ages 18 to 64								
Sample Size	71		70	76	74		77	70
LS Mean Diff from Placebo			-3.0	-5.8			-3.2	-5.5
95% Confidence Interval			(-4.9, -1.1)	(-7.6, -4.0)			(-5.1, -1.4)	(-7.2, -3.7)
p-value			0.002	<0.0001			0.0008	<0.0001
Age≥65								
Sample Size	5		3	3	3		3	7
LS Mean Diff from Placebo			-5.1	-10.3			-9.4	1.8
95% Confidence Interval			(-18, 7.5)	(-17.6, 3)			(-28.9, 10)	(-7.2, 11)
p-value			0.3	0.02			0.2	0.6
White								
Sample Size	64		63	74	70		70	68
LS Mean Diff from Placebo			-3.8	-6.0			-2.4	-5.2
95% Confidence Interval			(-5.8, -1.8)	(-7.9, -4.2)			(-4.4, -0.5)	(-6.9, -3.5)
p-value			0.0002	<0.0001			0.01	<0.0001
Black or African-American								
Sample Size	10	(b) (4)	9	5	4	(b) (4)	5	7
LS Mean Diff from Placebo			1.7	-3.8			-3.7	-0.8
95% Confidence Interval			(-2.3, 5.8)	(-9.5, 1.9)			(-16, 8.5)	(-14, 12.5)
p-value			0.4	0.2			0.5	0.9
Other Races								
Sample Size	6		8	2	5		7	4
LS Mean Diff from Placebo			0.05	-6.7			-11.3	0.5
95% Confidence Interval			(-7.1, 7.2)	(-19, 5.5)			(-19.6, -3)	(-13.9, 14.8)
p-value			0.9889	0.2			0.01	0.9

Source: Adapted from Clinical Study studies Q4881g and Q4882g, Post-text Tables 14.2/38 thru 14.2/40 for Study Q4881g and 14.2/35 thru 14.2/37 for study Q4882g

4.2 Other Special/Subgroup Populations

At the request of the FDA clinical team, differences in the treatment effect by baseline IGE level were considered. No difference in the treatment effect for any Xolair dose was observed in either study (p = (b) (4) 0.1, and 0.3 for the subgroup-by-treatment interaction for the Xolair 75 mg, 150 mg, and 300 mg groups in study Q4881g and p = (b) (4), 0.08, and 0.7 for the subgroup-by-treatment interaction for the Xolair 75 mg, 150 mg, and 300 mg groups in study Q4882g).

Nevertheless, analysis of the primary efficacy endpoint, the change from baseline in the weekly itch severity score to week 12 by baseline IGE (dichotomized at 80 IU/mL) is presented in Table 10. The results indicate that the treatment effects of Xolair 300 mg and Xolair 150 mg over placebo are present and relatively consistent across the baseline IGE level.

Table 10: Subgroup Efficacy Analysis: Change from Baseline in Weekly Itch Severity Score at Week 12 by Baseline IGE (mITT)

	Study Q4881g				Study Q4882g			
	Placebo N=80	Xolair 75 mg N=77	Xolair 150 mg N=80	Xolair 300 mg N=81	Placebo N=79	Xolair 75 mg N=82	Xolair 150 mg N=82	Xolair 300 mg N=79
Baseline IGE < 80 IU/mL (or missing)								
Sample Size	37	(b) (4)	46	40	42	(b) (4)	46	41
LS Mean Diff from Placebo			-5.0	-6.6			-2.0	-4.6
95% Confidence Interval			(-7.4, -2.5)	(-9.0, -4.1)			(-4.6, 0.5)	(-7.0, -2.1)
p-value			0.0001	<0.0001			0.1	0.0003
Baseline IGE ≥ 80 IU/mL								
Sample Size	43	(b) (4)	34	41	37		36	38
LS Mean Diff from Placebo			-0.9	-5.2			-4.3	-5.3
95% Confidence Interval			(-3.5, 1.7)	(-7.6, -2.9)			(-6.9, -1.7)	(-7.7, -2.9)
p-value			0.5	<0.0001			0.002	<0.0001

Source: FDA Analyses

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

During the course of this review, the following statistical issues were identified and resolved. Each issue is further described in the context of the referenced sections.

- The sample sizes for studies Q4881g and Q4882g were determined primarily based on safety and regulatory considerations. This resulted in an unnecessarily large sample size for purposes of efficacy. The reader should note that a highly significant p-value may be a result of the magnitude of the true difference between treatment groups, the level of variability in the efficacy measure, and/or the number of subjects studied. So that a highly significant p-value is not necessarily an indication that the magnitude of the treatment effect is large. Over interpretation of the p-value in this sense should be avoided. The point estimate and the corresponding confidence interval for the difference between treatment groups are the most appropriate means for estimation of the magnitude of the treatment effect. (Refer to sections 2.1 and 3.2.4)

- Pre-specified methods for missing data in the primary efficacy endpoint were not ideal because they did not simultaneously adequately estimate the variance associated with the treatment effect without perpetuating the treatment effect (Refer to sections 2.1 and 3.2.4)
- Dynamic randomization requires use of re-randomization tests (Refer to sections 2.1and 3.2.4)
- Within dose-level hierarchical analyses planned for the secondary efficacy endpoints do not completely control the type I error since there are three doses being examined (Refer to sections 2.1 and 3.2.4)

5.2 Collective Evidence

Studies Q4881g and Q4882g were generally consistent in findings and have been previously presented side-by-side; therefore, no formal statistical assessment of collective evidence across is studies is provided in this review and the reader is referred to section 5.3 for the conclusions and recommendations resulting from the review of study Q4881g and Q4882g.

5.3 Conclusions and Recommendations

From a statistical perspective, studies Q4881g and Q4882g each demonstrate statistically significant effects on the primary efficacy endpoint, the change from baseline to week 12 in weekly itch severity score, for both the Xolair 300 mg and Xolair 150 mg groups. Similar demonstration of efficacy for the Xolair 75 mg group was not achieved. Conclusions regarding the comparisons of each Xolair dose group to placebo in terms of the secondary efficacy endpoints were generally consistent with and supportive of those of the primary efficacy endpoint. The demonstration of efficacy for Xolair 300 mg and Xolair 150 mg in terms of the primary efficacy endpoint are not sensitive to the methods applied for missing data. Statistical methods that appropriately account for the adaptive randomization were also supportive of these conclusions and in fact yielded nearly identical results to traditional statistical tests.
No meaningful statistically significant differences in the treatment effect in terms of the primary efficacy endpoint across gender, race, age, or baseline IGE level were identified.

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

/s/

RUTHANNA C DAVI
02/10/2014

JOAN K BUENCONSEJO
02/10/2014
I concur.