CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

125526Orig1s000

STATISTICAL REVIEW(S)

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
CLINICAL OUTCOME ASSESSMENT

BLA / Sequence Number:	BLA 125-526 / Sequences 0000 and 0036
Drug Name:	Mepolizumab
Proposed Indication:	Severe eosinophilic asthma
Endpoints Assessed:	Saint George's Respiratory Questionnaire (SGRQ) Asthmas Control Questionnaire (ACQ-5)
Applicant:	GlaxoSmithKline
Date(s):	Received: November 4, 2014 PDUFA Due Date: November 4, 2015
Review Priority:	Standard
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	Robert Abugov, Ph.D.
Concurring Reviewer:	David Petullo, M.S.
Statistics Supervisor:	Thomas Permutt, Ph.D. (Division Director)
Medical Division:	Division of Pulmonary, Allergy, and Rheumatology Products
Clinical Team:	Sofia Chaudhry, M.D. (Medical Officer) Lydia Gilbert McClain, M.D. (Deputy Division Director) Badrul A Chowdhury, M.D., Ph.D. (Division Director)
Project Manager:	Nina Phuong Ton

Keywords: BLA review, Clinical Studies

Contents

Tables

1 EXECUTIVE SUMMARY

Unplanned analyses of three randomized, placebo-controlled, blinded, parallel arm trials (studies 75, 88, and 97) were conducted to examine improvements by mepolizumab compared to placebo for patient reported outcomes mean change from baseline SGRQ, SGRQ response rate, mean change from baseline ACQ-5, and ACQ-5 response rate.

Studies 75 and 88 examined change from baseline SGRQ. For mean change from baseline SGRQ , the statistical significance of differences between mepolizumab and placebo was uncertain in both studies. For SGRQ response rate, the differences between mepolizumab and placebo was of uncertain statistical significance in study 88 and not significant in study 75.

Studies 75, 88, and 97 examined change from baseline ACQ-5. For ACQ-5 mean change from baseline, the difference between mepolizumab and placebo was of uncertain statistical significance in studies 88 and 75, and was not statistically significant in study 97. For ACQ-5 response rate, the difference between mepolizumab and placebo was not statistically significant in any of the three studies.

For all four endpoints, in all studies examined, numerical improvements were in favor of mepolizumab rather than placebo.

2 INTRODUCTION

2.1 Overview

2.1.1 Drug Class and Indication

GlaxoSmithKline has proposed mepolizumab, a humanized interleukin-5 antibody, for add-on maintenance treatment, in patients aged 12 years and older, of severe eosinophilic asthma identified by blood eosinophilsat least 150 cells/mcL at initiation of treatment or blood eosinophils greater than or equal to 300 cells/mcL in the past 12 months.

2.1.2 Introductory Comments

A review for BLA 125526 (sequence 0000) was submitted to DARRTS on July 10, 2015. In that review, differences between mepolizumab and placebo for mean change from baseline SGRQ (ΔSGRQ) were not statistically significant, in one trial (study 88) because the ΔSGRQ was below a failed endpoint (asthma hospitalization rate) in the analysis hierarchy, and in a second trial (study 75) because ΔSGRQ was not included in the analysis hierarchy.

Nevertheless, there is renewed interest in patient reported outcomes within the Agency, and it does seem possible that, if the sponsor had been aware of that renewed interest, ΔSGRQ would have been included in the analysis hierarchy for study 75, and above failed endpoint asthma hospitalization rate in the analysis hierarchy for study 88. Had that been the case, it is possible that rigorous analyses would have demonstrated significant differences in ΔSGRQ between mepolizumab and placebo treated patients.

2.1.2.1 *Interpretation of Unplanned Analyses*

Given renewed interest in patient reported outcomes, this review undertakes a reanalysis of ΔSGRQ, even though it failed in the analysis hierarchy. The review also includes unplanned analyses for ΔACQ-5, proportion SGRQ responders, and proportion ACQ-5 responders.

Because the analyses provided are not preplanned, without proper control of type 1 error, the significance level at which the statistical tests are conducted will underestimate the false discovery rate.

A secondary goal of this review is to explore how to accurately depict unplanned analyses in statistical reviews. These explorations are intended as launching points for discussions within the Agency, but should not be construed as reflecting Agency policy. Such discussions are gaining urgency with potential application of court rulings regarding corporate freedom of speech to advertise 'true' effects. Unlike FDA, many scientific journals do not require pre-planned protocols, pre-specification of endpoints, or, indeed, any control of type 1 error when examining multiple endpoints. Whether sponsors mislead the public if they employ such 'normal' scientific standards rather than the arguably higher standards employed by FDA may be the subject of future court cases. There is, indeed, concern within the Agency that scientific standards may need to be lowered to correspond to some unspecified 'norm.'

When provided in a review or product label, statistical entities such as p-values and confidence limits are routinely interpreted as if they convey type 1 error and, when they do not, their provision may actually be misleading. There are at least four methods to improve clarity of language when describing results:

1. When multiplicity precludes discernment of type 1 error, reviewers could include confidence limits rather than p-values; this can alert the clinical team to potential lack of significance associated with multiplicity. However, unless the confidence limits are adjusted for multiplicity, their inclusion is subject to misinterpretation that upper and lower bounds represent a range of values within which type 1 error is controlled.

2. Indicate that provided values do not directly measure type 1 error by use of the term 'nominal.' For example, in the present case, differences between placebo and treatment from exploratory analyses of ΔSGRQ might be reported as 'nominally significant' because the 'nominal p-value' is less than the pre-planned family-wide level of statistical significance.

   The problem with this approach is that 'nominally significant' implies that the difference would have been significant if, somehow, different methods or hierarchies were prespecified to control type 1 error. However, the fact is that different methods or hierarchies were not prespecified, and correct prespecification of methods for control of

type 1 error are as integral to confirmatory statistical evaluations as prespecification of analyses and endpoints.

Therefore, no statistical conclusions should be drawn from 'nominally significant' results; each time they are provided by the statistical reviewer or the clinical team, the appropriate caveats need to be repeated; even then, the use of the term 'significant' in 'nominally significant' invites misinterpretation.

3. Interpretation of analyses without adequate control of type 1 error may be clarified by dispensing with the notation used for confirmatory analyses. For example, unadjusted p-values and 95% confidence intervals could be reported as 'U-values' and 95% UI respectively, and unadjusted significance levels could be reported as ξ rather than α. Interpretations of results associated with these quantities would differ from their adjusted counterparts (Table 1), and the statistics review would consistently reflect such differences.

Table 1. Interpretation of Analyses Adjusted or Not Adjusted to Control Type 1 Error

Result	**Statistical Implication**[a]	**Verbal Translation**[a]
$p > \alpha$	Type 1 error $> \alpha$	No significant difference
$p \leq \alpha$	Type 1 error $\leq \alpha$	Significant difference
$U \geq \xi$	Type 1 error $> \alpha$	No significant difference
$U < \xi$	Type 1 error uncertain	Statistical significance uncertain
$100(1-\alpha)\%$ CI	Coverage probability = $(1-\alpha)$	$100(1-\alpha)\%$ confidence interval
$100(1-\xi)\%$ UI	Coverage probability < $(1-\xi)$	$100(1-\alpha)\%$ is wider than $100(1-\xi)\%$ UI

[a]Interpretations based on fact that multiplicity adjustments always reduce type 1 error

4. Allocate some portion of α to unplanned endpoints. A simple way to achieve this without greatly altering outstanding protocols would be to acknowledge that negative consequences of type 1 error are greater for primary endpoints (approval of a drug which is not efficacious) than for secondary endpoints (lack of effectiveness for symptomatic or supportive endpoint).

   For example, as is currently the case, primary endpoints could be evaluated at the .05 level of significance; however all other endpoints could be evaluated at the familywide .075 level of significance, with α allocated among two sub-families: (i) control of type 1 error preplanned by the sponsor, (ii) control of type 1 error not preplanned by the sponsor. For endpoints in the first category overall type 1 error would be controlled at the 0.05 level of significance and, for endpoints in category (ii), an FDA default method, yet to be developed and made explicit for sponsors to understand, would limit the overall

type 1 error to .025, e.g., by using a Bonferroni adjustment in which endpoint is tested at the .025/(total number of secondary and exploratory endpoints measured) level of significance.

Results in this review will be presented according to alternative 3 above. Alternatives 1 and 2 confound rather than clarify statistical results, and alternative 4 is currently unsupported since FDA has no preplanned default method to control type 1 error among secondary and exploratory endpoints not addressed for multiplicity in the sponsor's protocol.

2.1.2.2 Interpretation of Tipping Point Analyses

Tipping point analyses are commonly employed to evaluate the sensitivity of results from primary analyses to changes in values of missing data. However, in this review, tipping point analyses are considered primary, representing multiple plausible values of the intent-to-treat estimand. Therefore, all plausible shifts from imputations based on observed values are considered of primary importance. An important ambiguity associated with current implementations is that the range of plausible shifts from observed is not generally preplanned.

2.2 Data Sources

Phase 3 study data, with additional datasets for SGRQ and ACQ-5, are located at:

\\cdsesub1\evsprod\BLA125526\0000\m5\datasets

and

\\cdsesub1\evsprod\BLA125526\0036\m5\datasets

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

Datasets, programs, and documentation provided by the applicant were adequate to evaluate the proposed claims.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

Study designs and endpoints included in the analysis hierarchies are outlined in Table 2 and Table 3, with results discussed in detail in the review for sequence 0000.

Table 2. Trial Design, Phase 2 Studies

Study[1]	Design	Population	Endpoints
97 (Trial 1)	M75 IV M250 IV M750 IV Pbo + SOC Parallel arm DB Pbo to W52	Asthma Age 12 to 65 years Controller medication ICS fluticasone ≥880 mcg/day ≥2 exacerbations past year Pre-bronch FEV1 <80% pred or PEF diurnal variability > 20% Eosinophils blood ≥ 300/mcL, or sputum ≥ 3% or Exhaled NO ≥ 50ppb or Loss of asthma control following ≤25% steroid reduct N 616 1:1:1:1 strat: maint OCS (Y, N)	*Primary:* Exacerbation rate *Secondary:* ΔPre-bronchodilator FEV1 at W52 AQLQ score at W52 Severe exacerbation rate ACQ-6 at W52

Source: Reviewer

[1]Trial number in parentheses cross references to label.

M75, M100 mepolizumab 75, 250, 750 mg once every four weeks, Pbo placebo, IV intravenous, SOC standard of care, DB double blind, ICS inhaled corticosteroids, FEV1 one second forced expiratory volume, PEF peak expiratory flow, NO nitric oxide, N sample size, strat stratifying variable, maint maintenance, AQLQ asthma quality of life questionnaire, ACQ asthma control questionnaire, W12, W24, W32, W52 weeks 12, 24, 32, and 52

Table 3. Trial Designs, Phase 3 Studies

Study	Design	Population	Endpoints
88 (Trial 2)	M75 IV M100 SC Pbo + SOC Parallel arm DB, DD Pbo to W32	Asthma Age ≥12 years Controller medication ICS[2] fluticasone ≥880 mcg/day (age≥18 yr) ≥440 mcg/day (age≥18 yr) ≥2 exacerbations past year Pre-bronch FEV1 <80% pred or FEV1:FVC < 0.80 Eosinophils blood ≥ 300/mcL past year, or blood ≥ 150/mcL screening N 576 1:1:1	*Primary:* Exacerbation rate *Secondary:* Severe exacerbation rate Hospitalization rate ΔTrough FEV1 W32 ΔSGRQ at W32

Source: Reviewer

DD double dummy,[1]Trial numbers in parentheses cross reference to label.

[2] or highest approved dose in investigator country

Table 3. Trial Designs, Phase 3 Studies (continued)

Study[1]	Design	Population	Endpoints
75	M100 SC	Asthma	*Primary:*
(Trial 3)	Pbo	Age ≥12 years	% Reduction OCS
		One or more failed controller meds	
	+ SOC	OCS 5 to 35 mg/day prednisone	
		ICS[1] fluticasone	
	Parallel arm	≥880 mcg/day (age≥18 yr)	
	DB	≥440 mcg/day (age≥18 yr)	
	Pbo to W24	Age ≥18	
		Pre-bronch FEV1 <80% pred	
		Age 12-17	
		Pre-bronch FEV1 <90% pred, or	
		FEV1:FVC < 0.80	
		Eosinophils	
		blood ≥ 300/mcL past year, or	
		blood ≥ 150/mcL baseline	
		N 135 1:1 strat: OCS (<5, ≥ 5 yr)	

[1] or highest approved dose in investigator country

3.2.2 Statistical Methodologies

3.2.2.1 *Studies 88 and 97*

In study 88 ΔSGRQ at week 32 was analyzed in the sponsor's primary analysis using analysis of covariance (ANCOVA), with independent factors treatment, region, baseline value, OCS usage at baseline (Y/N), and number of exacerbations in prior year. In addition, study 88 included as an independent factor prebronchodilator percent predicted FEV1 at baseline. SGRQ was not evaluated in Study 97.

In studies 88 and 97, ΔACQ-5 at week 32 was analyzed in the sponsor's primary analysis using mixed model repeated measures (MMRM) with independent factors treatment, OCS usage at baseline (Y/N), number of exacerbations in prior year, region, visit, baseline value, baseline by visit interaction, and visit by treatment interaction. In addition, study 88 included as an independent factor prebronchodilator percent predicted FEV1 at baseline.

Percent responders was analyzed using logistic regression, with a threshold for response equal to a decrease from baseline of at least 4.0 for SGRQ and 0.5 for ACQ-5.

Tipping point analyses were provided for ΔSGRQ and ΔACQ-5 using multiple imputation for missing data. Random draws were made from a normal distribution with mean equal to the tipping point mean and standard deviation based on observed data. Analyses of the completed datasets were performed using MMRM for change from baseline ACQ-5, ANCOVA for change from baseline SGRQ and logistic regression for percent responders. Results were combined across imputations using Rubin's method[1].

For study 97, patients who withdrew from the study prematurely were followed 8 to 24 weeks after the last dose of the investigational product. For study 88, the primary endpoint was measured up to 4 weeks after the final dose. Missing data was not imputed for the primary analyses.

All statistical analyses were on the intent-to-treat (ITT) population, defined as all randomized patients who received at least one dose of their randomized treatment.

3.2.2.2 *Study 75*

ΔSGRQ at week 24 was analyzed using ANCOVA with independent variables treatment, region, baseline value, number of years OCS use (< 5 years, ≥ 5 years), and baseline OCS dose.

ΔACQ-5 at week 24 was analyzed using MMRM with independent factors treatment, OCS usage at baseline (Y/N), region, visit, baseline value, baseline by visit interaction, and visit by treatment interaction.

Percent responders was analyzed using logistic regression, with a threshold for response equal to a decrease from baseline of at least 4.0 for SGRQ and 0.5 for ACQ-5.

Tipping point analyses were provided for ΔSGRQ and ΔACQ-5 using multiple imputation for missing data. Random draws were made from a normal distribution with mean equal to the tipping point mean and standard deviation based on observed data. Analyses of the completed datasets were performed using MMRM for change from baseline ACQ-5, ANCOVA for change from baseline SGRQ and logistic regression for percent responders. Results were combined across imputations using Rubin's method.

---

[1] Rubin, D.B. (1987) *Multiple Imputation for Nonresponse in Surveys*. J. Wiley & Sons, New York.

Patients who withdrew from the study prematurely were assigned to the lowest efficacy category for analysis.

All statistical analyses were on the intent-to-treat (ITT) population, defined as all randomized patients who received at least one dose of their randomized treatment.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

There were no obvious differences between treatments for demographic baseline characteristics in the submitted studies, as was noted in the prior statistical review. In studies 88 and 75, approximately 95% of patients completed the trial. In study 97, approximately 85% of patients completed the trial.

Table 4. Patient Disposition, Study 97

	N (%) of Patients			
	Pbo	**75 mg IV**	**250 mg IV**	**750 mg IV**
Randomized	155	153	152	156
Completed	127 (82)	129 (84)	131 (86)	133 (85)
Withdrawn	28 (18)	24 (16)	21 (14)	23 (15)
Adverse event [a]	6 (4)	5 (3)	8 (5)	9 (6)
Adverse event [b]	5 (3)	4 (3)	7 (5)	8 (5)
Lab abnormality [c]	1 (<1)	1 (<1)	1 (<1)	1 (<1)
Lack of efficacy	8 (5)	6 (4)	4 (3)	4 (3)
Protocol deviation	1 (<1)	1 (<1)	0 (0)	0 (0)
Lost to follow-up	1 (<1)	1 (<1)	4 (3)	0 (0)
Investigator discretion	1 (<1)	3 (2)	3 (2)	3 (2)
Withdrew consent	11 (7)	8 (5)	2 (1)	7 (4)

Source: CSR Table 5

a. adverse event leading to permanent discontinuation of investigational product or withdrawal from study
b. patients with 'Adverse event' as primary reason for withdrawal
c. patients with 'Subject reached protocol-defined stopping criteria' as primary reason for withdrawal and 'lab abnormality' as secondary reason for withdrawal

Table 5. Patient Disposition, Study 75

	Number (%) of Patients	
	Pbo	**100 mg SC**
Randomized	66	69
Completed	62 (94)	66 (96)
Withdrawn	4 (6)	3 (4)
Adverse event	3 (5)	3 (4)
Withdrew consent	1 (2)	0

Source: CSR Table 6

Table 6. Patient Disposition, Study 88

	Number (%) of Patients		
	Pbo	**75 mg IV**	**100 mg SC**
N	191	191	194
Completed	179 (94)	175 (92)	185 (95)
Withdrawn[a]	12 (6)	16 (8)	9 (5)
Withdrawal by Subject	5 (3)	9 (5)	4 (2)
Adverse event	4 (2)	0 (0)	1 (<1)
Lack of efficacy	1 (<1)	1 (<1)	2 (1)
Lost to Follow-up	0 (0)	2 (1)	2 (1)
Protocol deviation	0 (0)	3 (2)	0 (0)
Physician decision	2 (1)	1 (<1)	0 (0)

Source: CSR Table 3

a. Four patients were randomized and withdrawn without receiving any study medication and are not in the ITT population

3.2.4 Results and Conclusions

Because no differences between mepolizumab doses were seen in the prior statistical review, data from all mepolizumab doses were pooled for the analyses presented here.

3.2.4.1 Saint George's Respiratory Questionnaire (SGRQ)

In studies 88 and 75, mean reductions from baseline SGRQ seen in Table 7 were numerically greater among patients treated with mepolizumab (M) than among patients treated with placebo (Pbo), with statistical significance uncertain in both studies. Tipping point analyses conducted for study 88 (not shown) and study 75 (Table 8) confirmed lack of certainty regarding statistical signficance. Regarding the tipping point analyses for study 75, lack of statistical significance was seen only in cases where imputed values for patients randomized to mepolizumab were much worse than those for patients randomized to placebo, as noted in the highlighted area of Table 8.

Table 7. Mean Reductions from Baseline SGRQ, Studies 88 and 75, Observed Data

Study	Week	Treatment (N)		M – Pbo (95% UI)
		Pbo	M	U-Value
88	32	-9	-16	-7
		(177)	(356)	(-9, -4)
				<.0001
75	24	-3	-9	-6
		(61)	(65)	(-11, -1)
				.02

source: sequence 0000 Study 75 CSR Table 34 page 80, reviewer programs SGRQ Merged S88 2015 07 15.sas, SGRQ Merged Study 75 2015 07 15.sas

Table 8. Mean Reduction from Baseline SGRQ Tipping Point Analysis U-Values, Study 75

	Mepolizumab Change from Baseline among Withdrawers							
Placebo Change from Baseline for Withdrawers	*-16*	*-12*	*-8*	*-4*	*0*	*4*	*8*	*12*
-16	0.027	0.033	0.039	0.043	0.054	0.064	0.076	0.092
-12	0.020	0.024	0.027	0.033	0.040	0.049	0.057	0.072
-8	0.014	0.016	0.021	0.025	0.031	0.037	0.045	0.055
-4	0.011	0.012	0.014	0.019	0.022	0.028	0.035	0.043
0	0.008	0.009	0.011	0.014	0.017	0.022	0.026	0.034
4	0.006	0.007	0.008	0.010	0.013	0.016	0.021	0.025
8	0.004	0.005	0.006	0.008	0.010	0.012	0.016	0.020
12	0.003	0.004	0.005	0.006	0.008	0.010	0.012	0.016

Source: Table 60-15, sequence 0036

Regarding percent SGRQ responders, with non-responder imputation, statistical significance was ruled out for study 75 and uncertain for study 88 (Table 9). Tipping point analyses confirmed uncertainty of statistical significance for study 88.

Table 9. SGRQ Percent Responders[1], Studies 88 and 75, Non-Responder Imputation

Study	Week	Treatment (N)		M – Pbo Diff % OR (95% UI OR) U-Value
		Pbo	M	
88	32	55% (105/191)	69% (267/385)	14% 2.0 (1.4, 2.9) <.001
75	24	41% (27/66)	58% (40/69)	17% 1.9 (1.0, 3.8) 0.068

Source: sequence 0036 Study 88 Table 60.15, Study 75 Table 60.13
OR – odds ratio
1. SGRQ responder defined as patient with change from baseline ≤ -4.0

3.2.4.2 Change from Baseline ACQ-5

Compared to placebo, mepolizumab provided statistically uncertain reductions in ACQ-5 (Table 10) in all three studies. Tipping point analyses indicated that statistical significance was uncertain for studies 75 and 88. However, for study 97, statistical significance was rejected for a broad range of imputed values, including cases in which imputed values are equal among patients randomized to placebo and mepolizumab (highlighted values, Table 11).

Table 10. Mean Reduction from Baseline ACQ-5, Studies 97, 88 and 75, Observed Data

Study	Week	Treatment (N)		M – Pbo (95% UI)
		Pbo	M	Nominal U-Value
97	52	-0.58 (461)	-0.80 (155)	-0.22 (-0.43, -0.02) .03
88	32	-0.50 (385)	-0.93 (191)	-0.43 (-0.59, -0.26) <.0001
75	24	-0.09 (61)	-0.61 (65)	-0.52 (-0.87, -0.17) 0.004

source: reviewer program acq5 merged 2015 07 15.sas

Table 11. Mean Reduction from Baseline ACQ-5 Tipping Point Analysis U-Values, Study 97

	Mepolizumab Change from Baseline among Withdrawers						
Placebo Change from Baseline for Withdrawers	*-1.5*	*-1*	*-0.5*	*0*	*0.5*	*1*	*1.5*
-1.5	0.243	0.599	1.000	1.000	1.000	1.000	1.000
-1	0.069	0.203	0.523	0.959	1.000	1.000	1.000
-0.5	0.012	0.047	0.159	0.461	0.970	1.000	1.000
0	<0.001	0.005	0.025	0.136	0.421	0.809	1.000
0.5	<0.001	<0.001	0.004	0.025	0.103	0.307	0.727
1	<0.001	<0.001	<0.001	0.004	0.021	0.108	0.297
1.5	<0.001	<0.001	<0.001	<0.001	0.004	0.024	0.088

Source: Table 60.08, Sequence 0036

In studies 97 and 75, compared to placebo, mepolizumab did not provide statistically significant improvements in percent ACQ-5 response rate (Table 12). Although the statistical significance of the improvement in study 88 in was uncertain, tipping point point analyses provided U-values greater than 0.05 when imputed values were slightly worse for patients randomized to mepolizumab than among patients randomized to placebo (highlighted values, Table 12). Therefore benefits of mepolizumab over placebo for percent ACQ-5 response in study 88 are not statistically significant.

Table 12. ACQ-5 Percent Responders[1], Studies 97, 88 and 75, Non-Responder Imputation

Study	Week	Treatment (N)		M – Pbo Diff % OR (95% UI OR) Nominal U-Value
		Pbo	M	
97	52	50% (77/155)	48% (222/461)	2% 1.1 (0.71, 1.55) .8
88	32	45% (85/191)	52% (202/385)	7% 1.6 (1.1, 2.3) .02
75	24	29% (19/66)	42% (29/69)	13% 1.7 (-0.77, 3.83) .19

Source: sequence 0036 Study 97 Table 60.06, Study 88 Table 60.06, Study 75 Table 60.04
[1] ACQ-5 responder defined as patient with change from baseline ≤ -0.5

Table 13. ACQ Responder Rate Tipping Point Analysis U-Values, Study 88

	Mepolizumab Change from Baseline among Withdrawers						
Placebo Change from Baseline for Withdrawers	*-1.5*	*-1*	*-0.5*	*0*	*0.5*	*1*	*1.5*
-1.5	0.038	0.055	0.105	0.174	0.232	0.288	0.305
-1	0.026	0.038	0.070	0.115	0.172	0.197	0.219
-0.5	0.013	0.021	0.038	0.065	0.098	0.121	0.135
0	0.004	0.009	0.018	0.036	0.054	0.064	0.076
0.5	0.002	0.004	0.009	0.019	0.030	0.037	0.047
1	0.001	0.003	0.005	0.011	0.021	0.024	0.032
1.5	0.001	0.002	0.004	0.009	0.016	0.022	0.025

source: sequence 0036, Table 60.11

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

No subgroup analyses were conducted since this submission only included exploratory endpoints.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical issues

No outstanding statistical issues were seen noted.

5.2 Collective Evidence

Unplanned analyses examined effects of mepolizumab on SGRQ in two studies and on ACQ-5 in three studies. Compared to placebo, mepolizumab was associated with numerical improvements in response rate which were generally not statistically significant.

5.3 Conclusions and Recommendations

Unplanned analyses of three randomized, placebo-controlled, blinded, parallel arm trials (studies 75, 88, and 97) were conducted to examine improvements by mepolizumab compared to placebo for patient reported outcomes mean change from baseline SGRQ, SGRQ response rate, mean change from baseline ACQ-5, and ACQ-5 response rate.

Studies 75 and 88 examined change from baseline SGRQ. For mean change from baseline SGRQ , the statistical significance of differences between mepolizumab and placebo was uncertain in both studies. For SGRQ response rate, the differences between mepolizumab and placebo was of uncertain statistical significance in study 88 and not significant in study 75.

Studies 75, 88, and 97 examined change from baseline ACQ-5. For ACQ-5 mean change from baseline, the difference between mepolizumab and placebo was of uncertain statistical significance in studies 88 and 75, and was not statistically significant in study 97. For ACQ-5 response rate, the difference between mepolizumab and placebo was not statistically significant in any of the three studies.

For all four endpoints, in all studies examined, numerical improvements were in favor of mepolizumab rather than placebo.

5.4 Labeling Recommendations

If informational statements regarding effects of mepolizumab on SGRQ and ACQ-5 are needed, the following wording may be useful:

Studies 75 and 88 examined SGRQ response rate (reduction from baseline at least 4.0). Compared to placebo, the improvement by mepolizumab for was not statistically significant in study 75 and was of uncertain statistical significance in study 88.

Studies 75, 88, and 97 examined ACQ-5 response rate (reduction from baseline at least 0.5). Compared to placebo, the improvement by mepolizumab for was not statistically significant in any of the three trials.

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

/s/

ROBERT ABUGOV
09/22/2015

DAVID M PETULLO
09/22/2015
I concur.

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

Statistical Review

CLINICAL STUDY

BLA / Sequence Number:	BLA 125-526 / Seq 0000
Drug Name:	Mepolizumab
Proposed Indication:	Severe Eosinophilic Asthma
Applicant:	GlaxoSmithKline
Date(s):	Received: November 4, 2014 PDUFA Due Date: November 4, 2015
Review Priority:	Standard
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	Robert Abugov, Ph.D.
Concurring Reviewer:	David Petullo, M.S.
Statistics Supervisor:	Thomas Permutt, Ph.D. (Division Director)
Medical Division:	Division of Pulmonary, Allergy, and Rheumatology Products
Clinical Team:	Sofia Chaudhry, M.D. (Medical Officer) Lydia Gilbert McClain, M.D. (Deputy Division Director) Badrul A Chowdhury, M.D., Ph.D. (Division Director)
Project Manager:	Nina Phuong Ton

Keywords: BLA review, Clinical Studies

Contents

Tables

Figures

1 EXECUTIVE SUMMARY

Two randomized, placebo-controlled, blinded, parallel arm studies provide strong evidence that, compared to placebo, mepolizumab reduces the rate of clinical exacerbations in patients with severe asthma who have moderate to high blood eosinophil counts.

Interaction tests strongly indicate a positive association between screening blood eosinophil count and treatment effect for reduction of exacerbation rate. However, the evidence does not support the applicant's assertion that eosinophil counts above 300 per mcL during the year prior to treatment indicate that treatment will be effective.

A single randomized, placebo-controlled, blinded, parallel arm study provides evidence that treatment with mepolizumab reduces dependence on OCS for control of asthma.

Enrollment of adolescents and African Americans was not sufficient to demonstrate statistically significant treatment effects in these subgroups.

2 INTRODUCTION

2.1 Overview

2.1.1 Drug Class and Indication

GlaxoSmithKline proposes mepolizumab, a humanized interleukin-5 antibody, for add-on maintenance treatment, in patients aged 12 years and older, of severe eosinophilic asthma identified by blood eosinophils greater than or equal to 150 cells/mcL at initiation of treatment or blood eosinophils greater than or equal to 300 cells/mcL in the past 12 months.

2.1.2 History of Drug Development

The mepolizumab clinical development program for asthma was introduced to the Agency on December 20, 2005 under IND 6,971. Relevant communication between the Agency and the applicant are summarized below.

In response to questions submitted by the applicant (b) (4) , the Division recommended that the applicant conduct a phase 2 proof of concept study to identify the appropriate population for treatment, with establishment of appropriate biomarkers, including a comparison of safety and efficacy in patients with and without high sputum eosinophil counts. The Division also noted that the proposed (b) (4)

(b) (4). The Agency requested clarification regarding an unblinded interim analysis for sample size re-estimation because the time of assessment (e.g. 75% of minimum number of recruited patients) was not pre-specified and policies for maintenance of blinding were not submitted to the Agency for review.

In a meeting with the applicant held on April 21, 2009, the Division noted that discussion of phase 3 trials seemed premature considering that the applicant had not yet provided evidence which clearly identified a target patient population. The Division agreed with the proposed statistical analysis of exacerbations for study, MEA112997 (study 97) using a negative binomial regression model with dependent variables for treatment, baseline FEV_1, baseline number of exacerbations prior year, and log baseline sputum eosinophil differentials. Regarding multiplicity, the Division stated that, while the proposed Hochberg procedure would suffice for control of type 1 error, it may be too conservative for a dose ranging study. The Division also stated that study 97 could potentially serve as one of two required replicate adequate and well-controlled studies if the population of patients defined in the exclusion and inclusion criteria matched that of the target population.

In an end-of-phase 2 (EOP2) meeting held May 4, 2012 the Division noted that blood rather than sputum eosinophil levels are more suitable to identify the target population in most clinical settings. Therefore, identification of the target population in clinical practice may differ critically from that suggested from results based on sputum eosinophil counts in study 97. The Division also noted that proposed studies MEA115575 and (b) (4) (studies 75 and (b) (4) for reduction in oral steroid use for asthma use were (b) (4) . (b) (4) to consider inclusion of steroid sparing data on the label as secondary support for efficacy.

The Division also noted that, based on results from study 97, doses lower than the minimum dose examined in study 97, 75 mg intravenous (IV), may be effective, and that adequacy of dose ranging would be a review issue. Further, there were no efficacy data to support a 100 mg subcutaneous (SC) dose, and that bridging between the 75 mg IV and the 100 mg SC doses would be a review issue. The Division further noted that the clinical program would need to justify the proposed restriction of mepolizumab to the subset of severe asthma patients with eosinophilic inflammation.

On January 16, 2013, the Division conveyed to the applicant that the statistical analysis plans for studies 75 and 88 should detail plans to control type 1 error with multiple endpoints. Since the negative binomial model proposed to analyze exacerbations would assume that data is missing at random, sensitivity analyses should be conducted to examine the effects of missing data according to potential mechanisms of withdrawal. The Division also recommended that the applicant continue collecting exacerbation data after withdrawal from study treatment and to use that data in the analyses of treatment effect on exacerbation rate. In addition, the Division recommended that reasons for discontinuation be clearly documented and informative.

In a pre-BLA meeting held January 15, 2014, the Division reiterated, as in the EOP2 meeting, that bridging between the 75 mg IV and the 100 mg SC doses would be a review issue, and expressed concern that, in the absence of adequate bridging, the long-term database for the 100 mg SC dose would be inadequate for evaluation of safety. And again, as in the EOP2 meeting, the Division noted that the clinical program would need to justify the proposed restriction of mepolizumab to a subset of asthma patients. Regarding missing data for studies 97 and 88, the applicant assured the Agency that, although not all post-withdrawal data was collected, the discontinuation rate was low, less than 5%. The Division requested that any datasets submitted indicate when data was collected after patient withdrawal. The applicant also agreed to include SAS programs used for efficacy and safety analyses of studies 97, 88, 75.

2.2 Data Sources

Phase 3 study data, with corrected exacerbation datasets for study 88 submitted on April 3, 2015, are currently located at:

\\cdsesub1\evsprod\BLA125526\0000\m5\datasets

\\cdsesub1\evsprod\BLA125526\0018\m5\datasets

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

Datasets, programs, and documentation provided by the applicant were adequate to evaluate the proposed claims. Results from review analyses generally matched those in the submission.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

The mepolizumab development program for severe eosinophilic asthma included two phase 2 and two phase 3 randomized, parallel-arm, double-blind, double-dummy, placebo-controlled, standard of care add-on studies in asthma patients 12 years of age and older (Table 1 and Table 2). Trial SB-240563/006 (study 6) evaluated the effects of mepolizumab on morning peak expiratory flow rate (AM PEFR) by randomizing 362 patients in a 1:1:1 ratio to mepolizumab 250 mg (M250), mepolizumab 750 mg (M750), or placebo (Pbo) administered IV every four weeks (Q4W). Study 97 (trial 1 on proposed product label) evaluated the effects of mepolizumab on asthma exacerbation[1] rate by randomizing 616 patients in a 1:1:1:1 ratio to mepolizumab 75 mg (M75), M250, M750, or placebo administered IV Q4W. To support the to-be-marketed product, mepolizumab 100 mg administered SC (M100 SC) QW4, the applicant conducted study 88 (trial 2 on proposed product label). This study evaluated the effect of mepolizumab on exacerbations by randomizing 576 patients in a 1:1:1 ratio to M75 IV, M100 SC, or placebo administered QW4. A second phase 3 trial, study 75 (trial 3 on proposed product label), was conducted to evaluate the effect of the to-be-marketed product on percent reduction in oral corticosteroids (OCS) use, randomized 135 patients in a 1:1 ratio to M100 SC or placebo administered QW4.

Studies 75 and 88 restricted enrollment to patients with blood eosinophil counts ≥ 150 / mcL, and study 97 restricted enrollment to patients with symptoms of eosinophilic inflammation. Although no claims resulted from study 6, it was evaluated in the current review to explore the effect of mepolizumab on lung function.

[1] asthma exacerbation – worsening of asthma requiring, hospitalization, emergency department visits, the use of OCS at least double the existing maintenance dose for at least three days and/or, for study 88, a single intramuscular injection of corticosteroids.

Table 1. Trial Design, Phase 2 Studies

Study[1]	Design	Population	Endpoints
6	M250 IV M750 IV Pbo Parallel arm DB Pbo to W12	Asthma Age 18 to 55 years ICS beclomethasone equiv up to 1000 mcg / day 50% ≤ FEV1 ≤80% pred FEV1 reversibility ≥12% N 362 1:1:1	*Primary:* AM PEFR W12 domiciliary *Secondary:* ΔPre-dose FEV1 over 20 weeks Asthma summary symptom score Rescue medication Eosinophil count
97 (Trial 1)	M75 IV M250 IV M750 IV Pbo + SOC Parallel arm DB Pbo to W52	Asthma Age 12 to 65 years Controller medication ICS fluticasone ≥880 mcg/day ≥2 exacerbations past year Pre-bronch FEV1 <80% pred or PEF diurnal variability > 20% Eosinophils blood ≥ 300/mcL, or sputum ≥ 3% or Exhaled NO ≥ 50ppb or Loss of asthma control following ≤25% steroid reduct N 616 1:1:1:1 strat: maint OCS (Y, N)	*Primary:* Exacerbation rate *Secondary:* ΔPre-bronchodilator FEV1 at W52 AQLQ score at W52 Severe exacerbation rate ACQ-6 at W52 *Exploratory:* Screening blood eosinophil cutoff

Source: Reviewer

[1]Trial number in parentheses cross references to label.

SOC standard of care, DB double blind, DD double dummy, IAE investigator defined asthma exacerbation, PEFR peak expiratory flow rate, FEV1 one second forced expiratory volume, AQLQ asthma quality of life questionnaire, ACQ asthma control questionnaire, W12, W24, W32, W52 weeks 12, 24, 32, and 52

Table 2. Trial Design, Phase 3 Studies

Study[1]	Design	Population	Endpoints
88 (Trial 2)	M75 IV M100 SC Pbo + SOC Parallel arm DB, DD Pbo to W32	Asthma Age ≥12 years Controller medication ICS[2] fluticasone ≥880 mcg/day (age≥18 yr) ≥440 mcg/day (age≥18 yr) ≥2 exacerbations past year Pre-bronch FEV1 <80% pred or FEV1:FVC < 0.80 Eosinophils blood ≥ 300/mcL past year, or blood ≥ 150/mcL screening N 576 1:1:1	*Primary:* Exacerbation rate *Secondary:* Severe exacerbation rate Hospitalization rate ΔTrough FEV1 W32 ΔSGRQ at W32
75 (Trial 3)	M100 SC Pbo + SOC Parallel arm DB Pbo to W24	Asthma Age ≥12 years One or more failed controller meds OCS 5 to 35 mg/day prednisone ICS[2] fluticasone ≥880 mcg/day (age≥18 yr) ≥440 mcg/day (age≥18 yr) Age ≥18 Pre-bronch FEV1 <80% pred Age 12-17 Pre-bronch FEV1 <90% pred, or FEV1:FVC < 0.80 Eosinophils blood ≥ 300/mcL past year, or blood ≥ 150/mcL baseline N 135 1:1 strat: OCS (<5, ≥ 5 yr)	*Primary:* % Reduction OCS *'Supportive': W20 to W24* ≥ 50% Reduction OCS OCS ≤ 5 mg OCS discontinuation Median % reduction OCS *Exploratory:* Median OCS dose W24 ΔSGRQ

Source: Reviewer

[1]Trial numbers in parentheses cross reference to label.

[2] or highest approved dose in investigator country

Studies 6, 75, 88, and 97 were randomized, parallel-arm, double-blind, placebo-controlled multinational trials (Table 1 and Table 2). Studies 6 and 97 were conducted in patients at least 18 years of age, while studies 88 and 75 were conducted in patients at least 12 years of age.

Study 75 was double-dummy because different arms required different modes of administration that were visible to investigators and patients. Inclusion criteria, study treatments, primary endpoints, and secondary endpoints for each study are detailed in Table 1 and Table 2.

In studies 88 and 97, patients remained on their current asthma medications. However, in study 75, OCS reduction was undertaken during a five-week run-in period to ensure that patients would enter randomized treatment on the lowest OCS dose that would manage their current symptoms. This was established by using an increase from initial ACQ-5 greater or equal to 0.5 as an indicator to terminate dose reduction and return to the previous dose.

The post-randomization OCS dose among patients in study 75 followed a predefined schedule of reduction unless at least one of the following held:

1. Mean AM PEF was < 80% of the baseline stability limit
2. Mean asthma-related night time awakenings >50% increase over the baseline period (per night), >150% of the baseline mean
3. Rescue medication use requiring 4 or more puffs/day above the mean baseline value for any 2 consecutive days in the prior week, or 12 puffs or more on any one day in the prior week
4. Change in ACQ-5 $\geq$ 0.5 from the prior month OCS dose assessment
5. Symptoms of adrenal insufficiency

3.2.2 Statistical Methodologies

3.2.2.1 Study 6

Analysis of study 6 will focus on bronchodilation, evaluated as the change from baseline in pre-dose FEV1 (Δ pre-dose FEV1) at week 12 using an analysis of covariance (ANCOVA) with independent factors treatment, region, and treatment by region interaction.

3.2.2.2 Study 97

Analysis of the primary endpoint, exacerbation rate, first evaluated a linear trend test for decrease in exacerbation rate as a function of mepolizumab dose. Then, if the trend was significant at the 0.05 level, each dose was tested against placebo.

Exacerbation rates were analyzed using a generalized linear model with negative binomial distribution having independent factors treatment, OCS usage at baseline, region, number of exacerbations in year prior to study, and baseline disease severity (% predicted FEV1). The planned offset variable was logarithm of time on treatment.

Type 1 error across these comparisons was to be controlled at the 0.05 level by a truncated Hochberg procedure. First, tests of individual doses against placebo were conducted only if the overall linear trend test across doses (including placebo) was statistically significant. Comparisons to placebo for each of the three doses were then conducted in the following ordered hierarchy provided on page 23 of the applicant's Reporting and Analysis Plan:

1. Rate of exacerbations
2. FEV1 pre-bronchodilator at week 52, AQLQ at week 52
3. AQLQ at week 52
4. Rate of exacerbations requiring hospitalizations or emergency department visits
5. ACQ-6 at week 52.

The applicant's plan for control of type 1 error for the primary endpoints is questionable because the Hochberg test is only guaranteed to control familywide type 1 error for more than two doses if the effects of the doses are independent. In the present case, however, independence between doses seems unlikely.

Secondary endpoints, annual rate of investigator defined asthma exacerbations (IAE) and severe exacerbations (requiring hospitalization or emergency department visits), were to be analyzed using the negative binomial regression described above for the primary endpoint.

Time to first exacerbation and IAE were to be compared between treatment groups using a Cox proportional hazards model with independent factors treatment, OCS usage at baseline, region, number of exacerbations in year prior to study, and baseline disease severity (% predicted FEV1).

Trough FEV1 and post-bronchodilator FEV1 were analyzed using mixed models repeated measures (MMRM) with independent factors treatment, OCS usage at baseline, region, baseline FEV1, visit, and visit by baseline FEV1 interaction, and visit by treatment interaction.

ACQ and AQLQ were analyzed using MMRM with independent factors treatment, OCS usage at baseline, region, baseline value, and visit. For ACQ, additional terms included visit by baseline ACQ interaction and visit by treatment interaction.

Patients who withdrew from the study prematurely were followed up 8 to 24 weeks after the last dose of the investigational product. Missing data was not imputed for the primary analyses. To examine the robustness of the results with respect to patient withdrawal, tipping point analyses were conducted in which exacerbation rate after withdrawal varied between one and five exacerbations per year.

All statistical analyses were on the intent-to-treat (ITT) population, defined as all randomized patients who received at least one dose of their randomized treatment.

3.2.2.3 Study 88

Exacerbation rates were analyzed using a generalized linear model with negative binomial distribution with independent factors treatment, OCS usage at baseline, region, number of exacerbations in year prior to study, and baseline disease severity (% predicted FEV1). The planned offset variable was logarithm of time on treatment. For calculation of marginal treatment outcomes, class variables OCS usage at baseline and region were weighted according to frequency in the sampled population.

Secondary endpoints annual rate of severe exacerbations (requiring hospitalization and/or emergency department visits), were to be analyzed using the negative binomial regression described above for the primary endpoint.

Trough FEV1 was analyzed using MMRM with independent factors treatment, baseline OCS usage (Y/N), region, baseline FEV1, visit, and visit by baseline FEV1 interaction, and visit by treatment interaction.

SGRQ was analyzed using ANCOVA with independent factors treatment, baseline OCS usage (Y/N), region, baseline percent predicted FEV1, number of exacerbations in prior year, and baseline SGRQ.

Planned comparisons were M75 IV versus placebo and M100 SC versus placebo. Type 1 error was controlled over multiple endpoints using a truncated Hochberg procedure conducted at the one-sided 0.025 level of significance. Significance for an endpoint was declared if both tests were significant at the unadjusted 0.025 level or if at least one test was significant at the unadjusted .0125 level. If both of the tests for an endpoint were significant at the one-sided unadjusted .025 level, then the next endpoint in the defined hierarchy was tested. The endpoint hierarchy was defined on page 19 of the Reporting Analysis Plan as the primary endpoint followed by secondary endpoints in the order listed in Table 2. The gamma parameter for the Hochberg procedure was 1.

Contrary to Division recommendations, data collection did not continue beyond four weeks after patients withdrew from treatment. To examine the robustness of treatment results to patient withdrawal from treatment, tipping point analyses were conducted in which exacerbation rate after withdrawal varied between 1 and 5 exacerbations per year.

All statistical analyses were on the intent-to-treat (ITT) population, defined as all randomized patients who received at least one dose of their randomized treatment.

3.2.2.4 Study 75

Comparison of M100 SC and placebo for percent reduction of daily prednisone dose while maintaining asthma control was analyzed using a proportional odds model with the following categories of percent reduction: 0%, >0% to <50%, 50% to <75%, 75% to 90%, and 90% to 100%. The model included independent variables treatment, number of years on OCS (< 5 years, ≥ 5 years), region, and baseline OCS dose.

OCS dose reduction of at least 50%, dose reduction to ≤ 5 mg / day, and reduction in OCS dose (Y/N) was to be analyzed using logistic regression, with independent factors the same as in the primary analysis.

Median percent reduction in OCS dose at week 24 was to be analyzed using the Mann-Whitney U test (also known as the Wilcoxon rank-sum test) adjusted by randomization stratum OCS use (<5 years, ≥ 5 years). Patients who withdrew prematurely or who did not maintain asthma control between weeks 20 and 24 were assigned a rank corresponding to a worse percent reduction than seen in any other patient.

For primary and secondary endpoints, a patient was to be defined as having achieved asthma control between weeks 20 and 24 if they did not have an exacerbation during this period.

No adjustments were made for the analyses of the secondary endpoints as the applicant considered these analyses to be sensitivity analyses of the primary endpoint.

Contrary to Division recommendations, data collection was not continued on patients who withdrew from treatment. Instead, patients who withdrew prematurely or who did not maintain asthma control between weeks 20 and 24 were assigned a rank corresponding to the worst OCS percent reduction category. To examine the robustness of treatment results to patient withdrawal from treatment, tipping point analyses were conducted in which all mepolizumab patients who withdrew were considered treatment failures. Placebo patients who withdrew without evidence of loss of asthma control were all assigned to categories of percent reduction ranging from 0% to between 90% and 100%.

All statistical analyses were on the intent-to-treat (ITT) population, defined as all randomized patients who received at least one dose of their randomized treatment.

Change from baseline Saint George's Respiratory Questionnaire (SGRQ) at week 24 was evaluated as an exploratory endpoint using ANCOVA with independent variables treatment, baseline SGRQ, region, number of years OCS use (< 5 years, ≥ 5 years), and baseline OCS dose.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

There were no obvious differences between treatments for baseline characteristics in the submitted studies (Appendix A; Table 27, Table 28, Table 29, and Table 30). Since withdrawal rates were similar regardless of treatment,patterns of patient disposition did not contradict efficacy of mepolizumab (Table 3, Table 4, Table 5, and Table 6).

Table 3. Patient Disposition, Study 6

	Pbo N (%)	**250 mg IV N (%)**	**75 mg IV N (%)**
Randomized*	126	120	116
Completed	119 (94)	110 (92)	112 (97)
Withdrawn	7 (6)	10 (8)	4 (3)
Adverse Event	5 (4)	4 (3.3)	1 (1)
Lack of Efficacy	0 (0)	0 (0)	0 (0)
Protocol Deviation	1 (1)	1 (1)	1 (1)

Source: CSR Tables 3 and 13.3.1

* Center 003 excluded - reason given is audit by FDA during investigation

Table 4. Patient Disposition, Study 97

	N (%) of Patients			
	Pbo	**75 mg IV**	**250 mg IV**	**750 mg IV**
Randomized	155	153	152	156
Completed	127 (82)	129 (84)	131 (86)	133 (85)
Withdrawn	28 (18)	24 (16)	21 (14)	23 (15)
Adverse event [a]	6 (4)	5 (3)	8 (5)	9 (6)
Adverse event [b]	5 (3)	4 (3)	7 (5)	8 (5)
Lab abnormality [c]	1 (<1)	1 (<1)	1 (<1)	1 (<1)
Lack of efficacy	8 (5)	6 (4)	4 (3)	4 (3)
Protocol deviation	1 (<1)	1 (<1)	0 (0)	0 (0)
Lost to follow-up	1 (<1)	1 (<1)	4 (3)	0 (0)
Investigator discretion	1 (<1)	3 (2)	3 (2)	3 (2)
Withdrew consent	11 (7)	8 (5)	2 (1)	7 (4)

Source: CSR Table 5

a. adverse event leading to permanent discontinuation of investigational product or withdrawal from study
b. patients with 'Adverse event' as primary reason for withdrawal
c. patients with 'Subject reached protocol-defined stopping criteria' as primary reason for withdrawal and 'lab abnormality' as secondary reason for withdrawal

Table 5. Patient Disposition, Study 75

	Number (%) of Patients	
	Pbo	**100 mg SC**
Randomized	66	69
Completed	62 (94)	66 (96)
Withdrawn	4 (6)	3 (4)
Adverse event	3 (5)	3 (4)
Withdrew consent	1 (2)	0

Source: CSR Table 6

Table 6. Patient Disposition, Study 88

	Number (%) of Patients		
	Pbo	**75 mg IV**	**100 mg SC**
N	191	191	194
Completed	179 (94)	175 (92)	185 (95)
Withdrawn[a]	12 (6)	16 (8)	9 (5)
Withdrawal by Subject	5 (3)	9 (5)	4 (2)
Adverse event	4 (2)	0 (0)	1 (<1)
Lack of efficacy	1 (<1)	1 (<1)	2 (1)
Lost to Follow-up	0 (0)	2 (1)	2 (1)
Protocol deviation	0 (0)	3 (2)	0 (0)
Physician decision	2 (1)	1 (<1)	0 (0)

Source: CSR Table 3

a. Four patients were randomized and withdrawn without receiving any study medication and are not in the ITT population

3.2.4 Results and Conclusions

3.2.4.1 Primary Endpoint: Exacerbation Rate

Compared to placebo, mean rate of all exacerbations in study 97 was significantly reduced among patients administered mepolizumab (Table 7), with point estimates for reductions in exacerbation rate compared to placebo ranging from 0.9 to 1.2 exacerbations per year. Further, point estimates for reductions in exacerbation rate did not suggest that any additional benefits were provided by higher doses (Table 7), and application of the primary analysis model to compare different mepolizumab doses showed no statistically significant differences between doses (Table 8). In Table 7 and similar tables, p-values were omitted when they were non-significant after the Hochberg adjustment for multiplicity. For example, in the analysis hierarchy, exacerbations due to hospitalization and/or emergency room visits fell below the failed endpoint change in FEV1. Similarly, because rate of hospitalizations was not included in the analysis hierarchy, the improvement for M750 compared to placebo for rate of hospitalizations was only nominally significant.

Table 7. Exacerbation Rates, Study 97

Criteria	Pbo n=155	75 mg IV n=153	250 mg IV n=152	750 mg IV n=156
All				
Exac / yr	2.4	1.2	1.5	1.2
Risk ratio, p-value		0.5 (<.0001)*	0.6 (.0006)*	0.5 (<.0001)*
95% CI		(0.4, 0.7)	(0.5, 0.8)	(0.4, 0.6)
Hosp+Emrgncy Dept				
Exac / yr	0.4	0.17	0.25	0.22
Risk Ratio, p-value		0.4	0.6	0.5
95% CI		(0.2, 0.8)	(0.3, 1.1)	(0.3, 1)
Hosp Only				
Exac / yr	0.2	0.1	0.1	0.07
Risk Ratio, p-value		0.6	0.7	0.4
95% CI		(0.3, 1.3)	(0.3, 1.4)	(0.2, 0.9)

source: CSR Tables 10 and 23, reviewer program exac studies 88 97 2015 06 03.sas

* statistically significant effect

Table 8. Exacerbation Rates, Differences Between Mepolizumab Doses, Study 97

Criteria	750 - 250 mg IV	750 - 75 mg IV	750- 75 mg IV
All			
Risk ratio, p-value	0.79	0.93	1.17
95% CI	(0.6, 1.1)	(0.7, 1.3)	(0.9, 1.6)
Hosp+Emrgncy Dept			
Risk Ratio, p-value	0.9	1.31	1.46
95% CI	(0.4, 1.8)	(0.6, 2.8)	(0.7, 3.1)
Hosp Only			
Risk Ratio, p-value	0.57	0.61	1.08
95% CI	(0.2, 1.4)	(0.2, 1.6)	(0.5, 2.5)

source: CSR Table 13, reviewer program exac studies 88 97 2015 06 03.sas

Similarly, rate of all exacerbations in study 88 was significantly reduced among patients administered mepolizumab compared to those administered placebo (Table 9). Point estimates for reduction in exacerbation rate ranged from 0.8 to 0.9 exacerbations per year. There was not a statistically significant difference between the M75 and M100SC doses. Again, p-values which were non-significant after Hochberg adjustment were omitted from the table.

Table 9. Exacerbation Rates, Study 88

Criteria	Pbo n=191	75 mg IV n=191	100 mg SC n=194	100 mg SC – 75 mg IV
All				
Exac / yr	1.7	0.9	0.8	
Risk ratio, p-value		0.5 (<.0001)*	0.5 (<.0001)*	0.9
95% CI		(0.4, 0.7)	(0.4, 0.6)	(0.6, 1.2)
Hosp+Emrgncy Dept				
Exac / yr	0.20	0.1	0.1	
Risk Ratio, p-value		0.7	0.4 (.03)*	0.6
95% CI		(0.3, 1.4)	(0.2, 0.8)	(0.3, 1.3)
Hosp Only				
Exac / yr	0.10	0.1	0.0	
Risk Ratio, p-value		0.6	0.3	0.5
95% CI		(0.2, 1.7)	(0.1, 0.9)	(0.2, 1.6)

source: CSR Tables 13, 29, 30, 40, response errata Table 3.016 reviewer program exac studies 88 97 2015 06 03.sas
* statistically significant effect

The estimated mean exacerbation rates presented in Table 9 weighted classes within region and OCS use at baseline by their proportions in the sampled population. A method more often seen in past regulatory submissions, however, weights such classes equally, and it therefore seems worthwhile to examine whether the outcome depends on weighting method. As in the proportionally weighted analyses, equal weighting analyses also demonstrated statistically significant improvements in exacerbation rate compared to placebo among patients administered mepolizumab (Table 10).

Table 10. Exacerbation Rates, with Equal Weighting of Class Variables, Study 88

Criteria	Pbo n=191	75 mg IV n=191	100 mg SC n=194	100 mg SC – 75 mg IV
All				
Exac / yr	2.1	1.1	1.0	
Risk ratio, p-value		0.5 (<.0001)*	0.5 (<.0001)*	0.9
95% CI		(0.4, 0.7)	(0.4, 0.6)	(0.6, 1.2)
Hosp+Emrgncy Dept				
Exac / yr	0.3	0.2	0.1	
Risk Ratio, p-value		0.7	0.4 (.03)*	0.6
95% CI		(0.3, 1.4)	(0.2, 0.8)	(0.3, 1.3)
Hosp Only				
Exac / yr	0.1	0.1	0.0	
Risk Ratio, p-value		0.6	0.3	0.5
95% CI		(0.2, 1.7)	(0.1, 0.9)	(0.2, 1.6)

source: reviewer program exac studies 88 97 2015 05 21.sas
* statistically significant effect

Tipping point sensitivity analyses indicated that results for exacerbations are robust in the face of missing data (Appendix 6.2).

In summary, there is strong evidence that mepolizumab reduces exacerbation rate among patients who experience exacerbations despite ongoing use of inhaled steroids plus controller medications such as LABA, leukotriene receptor antagonists, or theophylline.

3.2.4.2 Change from Baseline FEV_1

In study 6, change from baseline FEV1 (ΔFEV1) was considered an exploratory endpoint because there was no treatment effect for the primary efficacy variable, AM PEFR at Week 12 (CSR Table 23). In study 6, confidence limits for ΔFEV1 overlapped between placebo and all three mepolizumab doses (Table 11).

Table 11. Exploratory Analysis, ΔFEV1 at Week 12, Study 6.

	Δ Pre-Dose FEV1 (N)		
	Pbo	**250 mg IV**	**M750 mg IV**
N =	(129)	(121)	(118)
FEV1 (mL)	138	88	89
Diff from Pbo		-51	-50
95% CI		(-162, 60)	(-160, 60)

source: reviewer program FEV S06 biomarker 2015 02 20, CSR Table 25

In study 97, M75, M250, and M750 were not significantly different from placebo for ΔFEV1 at Week 52 (Table 12).

Table 12. ΔFEV1 at Week 52, Preplanned Analysis, Study 97

Week	Δ Pre-Dose FEV1 (N)			
	Pbo	**75 mg IV**	**250 mg IV**	**750 mg IV**
N=	(127)	(129)	(129)	(132)
FEV1 (mL)	60	121	140	115
Diff from Pbo		61	81	56
P-Value		(.23)	(.11)	(.27)
95% CI		(-38, 161)	(-19, 180)	(-43, 155)

source: CSR Study 97 page 549, reviewer program fev study 97 2015 05 22.sas

In studies 88 and 75, ΔFEV1 was evaluated only as an exploratory endpoint because it was either below a failed endpoint in the analysis hierarchy (study 88, asthma hospitalization rate) or not prespecified in the analysis hierarchy (study 75). Nominal confidence limits suggest an effect which exceeded placebo in study 88 (Table 13) but not in study 75 (Table 14). For study 75, the applicant claimed that the average difference over the treatment period between placebo and M100SC, (FEV1 measured at weeks 4, 8, 12, 16, 20, and 24) was significant (p = .03); however that claim of significance was from an exploratory analysis, without any control of type 1 error.

Table 13. Exploratory Analysis, ΔFEV1 at Week 32, Study 88

	Pbo	**75 mg IV**	**100 mg SC**
N =	(191)	(191)	(194)
ΔFEV1	86	186	184
Trt-Pbo		100	98
95% CI		(14, 187)	(12, 184)

source reviewer program fev study 88 2015 02 10.sas, study 88 CSR Table 40,

Table 14. Exploratory Analysis, ΔFEV1 at Week 24, Study 75

	Pbo	**100 mg SC**
	N = (191)	(194)
ΔFEV1	-4	110
Trt-Pbo		114
95% CI		(-44, 273)

source reviewer program fev study 75 2015 02 11, study 75 CSR Table 6.38

Although point estimates in study 6 for differences between mepolizumab and placebo suggest a negative effect of mepolizumab on ΔFEV1 compared to placebo (Table 11), point estimates from the other studies, enriched for severe asthma and high blood eosinophil count, favor mepolizumab (Table 12, Table 13, and Table 14).

In summary, while patterns for improvement of ΔFEV1 among severe asthma patients with high eosinophil count do not contradict results for exacerbations, the impact of mepolizumab on ΔFEV1 was not significant in any of the trials. The conditions, if any, under which mepolizumab can be relied upon to act as a bronchodilator remain undefined.

3.2.4.3 Primary Endpoint: Reduction of Oral Corticosteroid (OCS) Use

In study 75 patients administered mepolizumab (M100SC) rather than placebo experienced significantly increased odds of greater average percent reduction from baseline OCS dose while maintaining asthma control during weeks 20 to 24 (Table 15).

Table 15. Percent Reduction OCS, Weeks 20 to 24.

% Reduction OCS from Baseline	**Treatment N (%)**		**Odds Ratio (95% CI)**	**P-Value**
	Pbo (N=66)	**100 mg SC (N=69)**		
90% - 100%	7 (11%)	16 (23%)		
75% - <90%	5 (8%)	12 (17%)		
50% - <75%	10 (15%)	9 (13%)		
>0% - <50%	7 (11%)	7 (10%)		
No change or any increase or lack of asthma control or withdrawal from treatment	37 (56%)	25 (36%)		
Statistical Analysis			2.39 (1.25, 4.56)	.009

source: reviewer program ocs perc s75 2015 05 21.sas, CSR Table 16

Missing data was not an issue for this primary endpoint; tipping point analyses indicate that the results are robust (Appendix 6.2).

The applicant proposed for inclusion on the product label (b) (4) .

3.2.4.4 Exploratory Analyses: Saint George's Respiratory Questionnaire (SGRQ)

In the analysis hierarchy for study 88, SGRQ was tested after a failed endpoint, asthma hospitalization rate, and in study 75, SGRQ was analyzed without control of type 1 error. Therefore SGRQ was evaluated in studies 75 and 88 as an exploratory endpoint. From the analyses of SGRQ data in both studies, 95% confidence limits for the difference between mepolizumab and placebo exclude zero, and therefore appear to suggest a statistically significant effect of mepolizumab on change from baseline SGRQ (Table 17 and Table 18). However, actual confidence limits, with preplanned adjustments for multiplicity, would have been wider and may have included zero.

Table 17. Exploratory Analyses. Change from Baseline SGRQ at Week 32, Study 88

	Pbo	**75 mg IV**	**100 mg SC**
	N= (177)	(172)	(184)
ΔSGRQ	-9	-15	-16
Trt-Pbo		-6	-7
95% CI		(-10, -3)	(-10, -4)

source reviewer program sgrq s 88 2015 02 23.sas, study 88 CSR Table 40

Table 18. Exploratory Analyses. Change from Baseline SGRQ at Week 24, Study 75

	Pbo	100 mg SC
N=	(61)	(65)
ΔSGRQ	-3	-9
Trt-Pbo		-6
95% CI		(-11, -3)

source reviewer program sgrq study 75 2015 02 23, study 75 CSR Table 34.

3.3 Evaluation of Safety

Safety evaluations for this submission were conducted by the Medical Reviewer, Sofia Chaudhry, M.D. and are provided in her review.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

Effects of population subgroups on efficacy as measured by exacerbation rate in studies 97 and 88 and OCS reduction in study 75 were examined by adding the relevant subgroup and treatment by subgroup interaction to the primary analysis model, with results evaluated at the nominal 0.05 level of significance.

4.1 Gender, Race, Age, and Geographic Region

With the exception of gender in study 97 (Table 19), no statistically significant impacts of subgroups on treatment efficacy were seen in studies 75, 88, and 97 for race, age (12-17, 18-64, ≥65 years), or geographic region (North America, elsewhere). However, sample sizes were often inadequate for a thorough evaluation. A summary of the sample size for each subgroup is shown in Table 19.

Table 19. Sample Sizes, for Particular Demographics, Studies 97, 88, and 75

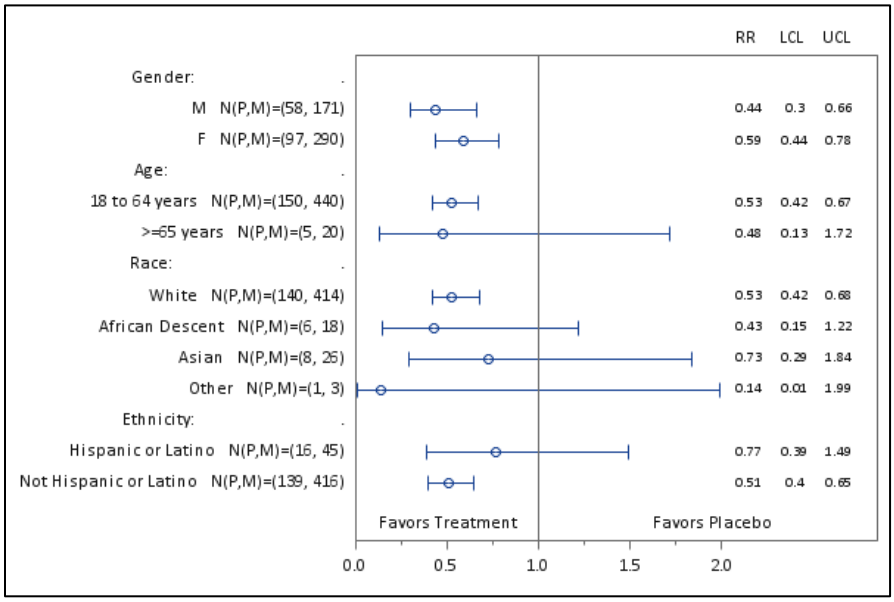
Category	Study		
	97	**88**	**75**
Randomized	616	576	135
African Descent	24 (4%)	16 (3%)	0 (0%)
American of African Descent	22 (4%)	14 (2%)	0 (0%)
Asian	34 (6%)	105 (18%)	3 (2%)
Hispanic	61 (10%)	51 (9%)	5 (4%)
12 to 17 years old	1 (0%)	25 (4%)	2 (1%)

source: reviewer programs exac study 97 misssubgr 2015 05 04.sas, exac study 88 misssubgr 2015 05 04.sas, OCS MissSubgr S75 2015 05 04.sas

Whether there is adequate representation in this study for patients of African descent and children from 12 to 17 years of age is a concern. Each study is discussed separately below.

Study 97. The 95% CIs for exacerbation rate ratios (mepolizumab/placebo) for subgroups age, gender, and race are shown in Figure 1.

Figure 1. Exacerbation Rate Ratios, by Gender, Age, Race, and Ethnicity, Study 97



source: reviewer program exac forest plots misssubgr s97 2015 05 06.sas

The treatment by sex interaction in study 97 was statistically significant (p=.01) but was not qualitative; that is, point estimates of effect indicated that, compared to placebo, mepolizumab reduced exacerbation rate in both sexes (Table 20).

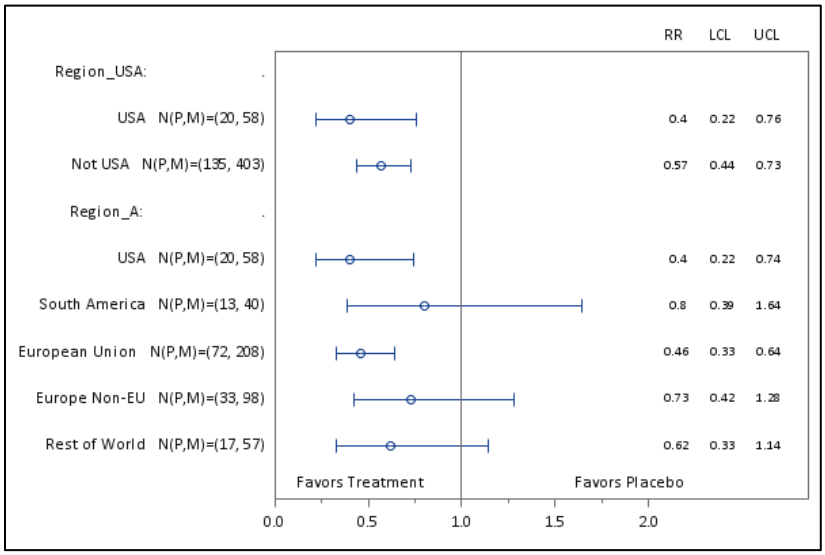
Table 20. Sex by Treatment Interaction for Exacerbation Rate, Study 97

	Treatment				Difference from Placebo		
Sex	**Pbo**	**M75**	**M250**	**M750**	**75 mg IV**	**250 mg IV**	**750 mg IV**
F	2.29 (97)	1.27 (104)	1.78 (93)	1.02 (93)	-1.02	-0.51	-1.27
M	2.09 (58)	0.83 (49)	0.78 (59)	1.26 (63)	-1.27	-1.31	-0.83

source: reviewer program exac study 97 subgr 2015 02 24.sas

Lack of statistically significant differences in treatment effect between geographic regions was graphically confirmed using forest plots for study 97 (Figure 2).

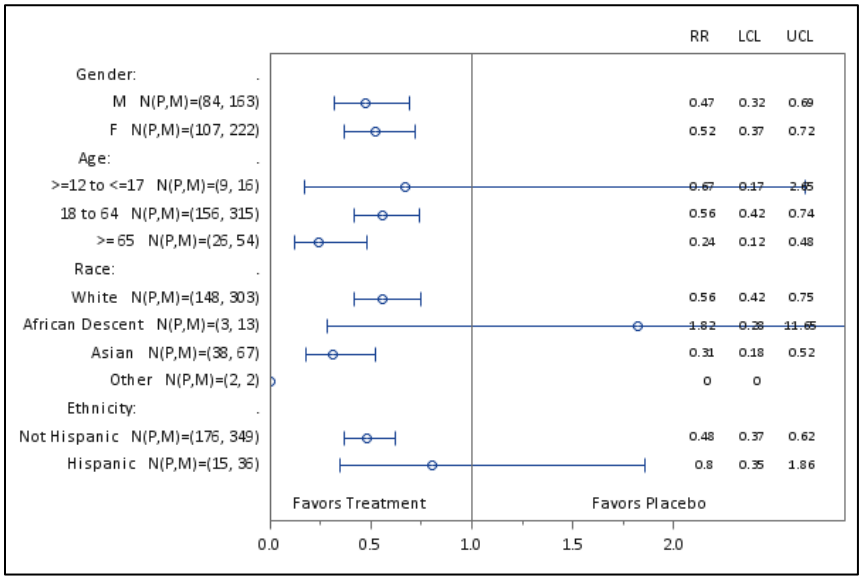
Figure 2. Exacerbation Rate Ratios by Region, Study 97



source: Exac Forest Plots Region S97 2015 06 02.sas

Study 88. Point estimates and 95% confidence limits for exacerbation rate ratios (mepolizumab/placebo) for the subgroups age, gender, and race are shown in Figure 3. For patients of African descent, the point estimate for rate ratio exceeded unity. Confidence limits for patients of African descent, however, did not exclude beneficial effects of this drug.

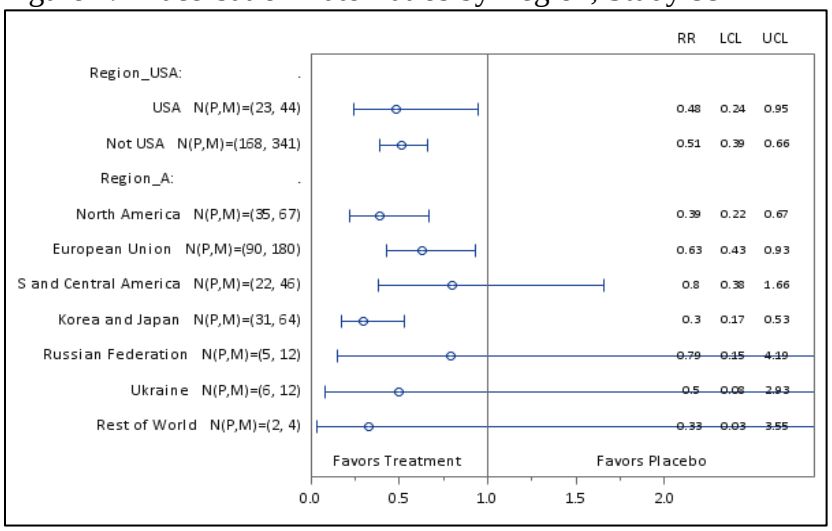
Figure 3. Exacerbation Rate Ratios, by Gender, Age, Race, and Ethnicity. Study 88



source: reviewer program exac forest plots misssubgr s88 2015 05 06.sas

Lack of statistically significant differences in treatment effect between geographic regions was graphically confirmed using forest plots for study 88 (Figure 4).

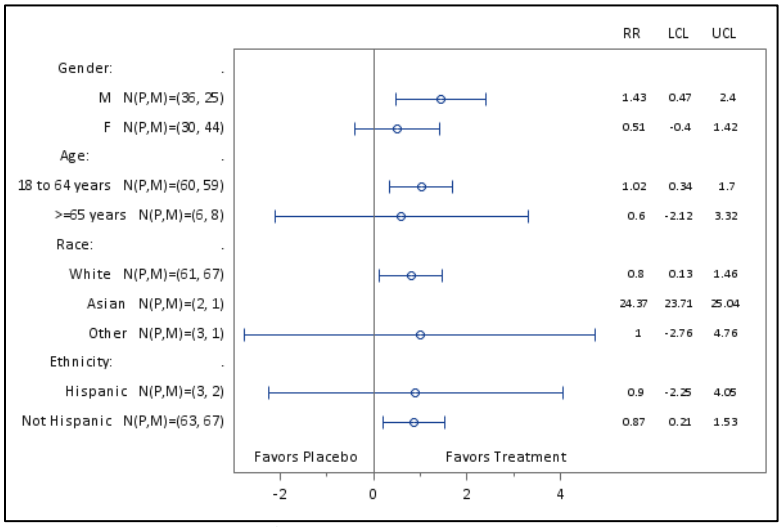
Figure 4. Exacerbation Rate Ratios by Region, Study 88



source: Exac Forest Plots Region S88 2015 06 02.sas

Study 75. Treatment effects for OCS reduction are shown for the subgroups age, gender, and race in Figure 5. Lack of data precluded analyses for patients who were of African descent or who were 12 to 17 years old.

Figure 5. OCS Reduction Log Odds Ratios, by Gender, Age, Race, and Ethnicity.



source: reviewer program exac forest plots misssubgr s75 2015 05 06.sas

For OCS reduction in study 75, the treatment by age (<40, ≥40 years) interaction was statistically significant (p=.0009), with the odds ratio indicating that OCS reduction by mepolizumab was successful only among patients who were at least 40 years old (Table 21). Further, point estimates for the odds ratio suggest that the interaction may be qualitative, with mepolizumab increasing requirements for OCS among patients younger than 40 years of age (Table 21).

Table 21. Age by Treatment Interaction for OCS Reduction, Study 75

Age	OR	P-Value
< 40	0.25	.074
≥ 40	4.35	.0002

source: reviewer program ocs subgr s74 2015 03 13.sas

Consistent with a qualitative age by treatment interaction for OCS reduction, among patients who were less than 40 years old, the percent of patients who experienced 90% to 100% OCS reduction was highest among those randomized to placebo, while the number of patients experiencing no improvement was highest among patients randomized to mepolizumab (Table 22). The opposite pattern held among patients who were at least 40 years old; the percent of patients who experienced 90% to 100% OCS reduction was highest among those randomized to treatment, while the number of patients experiencing no improvement was highest among patients randomized to placebo (Table 23).

Table 22. Percent Reduction OCS, Age < 40

% Reduction OCS from Baseline	Treatment N (%)		Odds Ratio	P-Value
	Pbo (N=12)	100 mg SC (N=18)		
90% - 100%	4 (33%)	1 (6%)		
75% - <90%	0 (0%)	3 (17%)		
50% - <75%	4 (33%)	3 (17%)		
>0% - <50%	1 (8%)	1 (6%)		
No change or any increase or lack of asthma control or withdrawal from treatment	3 (25%)	10 (56%)		
Statistical Analysis			0.25	.074

source: reviewer program ocs primary s75 by age 2015 03 16.sas

Table 23. Percent Reduction OCS, Age ≥ 40

% Reduction OCS from Baseline	Treatment N (%)		Odds Ratio	P-Value
	Pbo (N=54)	100 mg SC (N=51)		
90% - 100%	3 (6%)	15 (29%)		
75% - <90%	5 (9%)	9 (18%)		
50% - <75%	6 (11%)	6 (12%)		
>0% - <50%	6 (11%)	6 (12%)		
No change or any increase or lack of asthma control or withdrawal from treatment	34 (63%)	15 (29%)		
Statistical Analysis			4.35	.0002

source: reviewer program ocs primary s75 by age 2015 03 16.sas

In summary, sample sizes were often not adequate to evaluate effects of mepolizumab among patients of African descent and among 12 to 17 year olds. In studies 97 and 75, it could not be determined whether treatment had any effect among 12 to 17 year olds, and in study 75, no comparisons between treatment and placebo were available for patients of African descent. In study 88, the point estimate actually indicated a negative effect of treatment on patients of African descent, however the confidence interval did not rule out treatment benefits.

In addition, evidence from a single study suggests that benefits of mepolizumab for OCS reduction may be restricted to patients who are at least 40 years old.

4.2 Blood Eosinophil Count as an Effect Modifier

4.2.1 Statistical Methods

4.2.1.1 The Use of Cutpoints

The applicant's proposed indication identifies patients with severe eosinophilic asthma based blood eosinophil counts, either ≥ 150 cells/mcL at initiation of treatment, or ≥ 300 cells/mcL in the past 12 months. However, imposition of cut points, as proposed above, on a continuous or integer biomarker, such as blood eosinophil count, may greatly reduce statistical power to detect interactions if information regarding interactions is lost when replacing continuous or integer variables with categories. Such loss of information regarding interactions is perhaps responsible for the lack in study 97 of statistically significant differences in rate ratios between patients who did meet and who did not meet the blood eosinophil count criteria proposed for inclusion in the label indication (b) (4).

(b) (4)

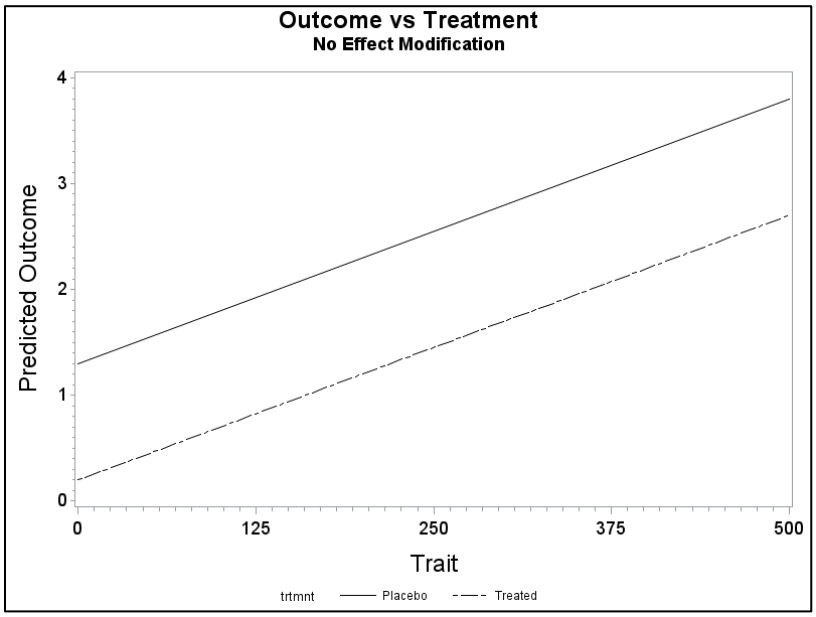
4.2.1.2 Methodological Notes: Analyses for Effect Modification

Before establishment of a trait as a diagnostic, complementary, or predictive biomarker, appropriate to help determine whether a particular patient should or should not receive a drug, the trait should be examined to determine whether it modifies the treatment effect. If the trait does modify treatment effect, a cut point may be required if there is a clear change in the balance of benefit to risk important for appropriate prescribing of the drug.

Statistical tests to determine whether a trait is an effect modifier are consistent with a simple geometric approach. For example, in Figure 6, the difference between placebo and treatment is constant over the range of the trait, and the trait is therefore not an effect modifier.

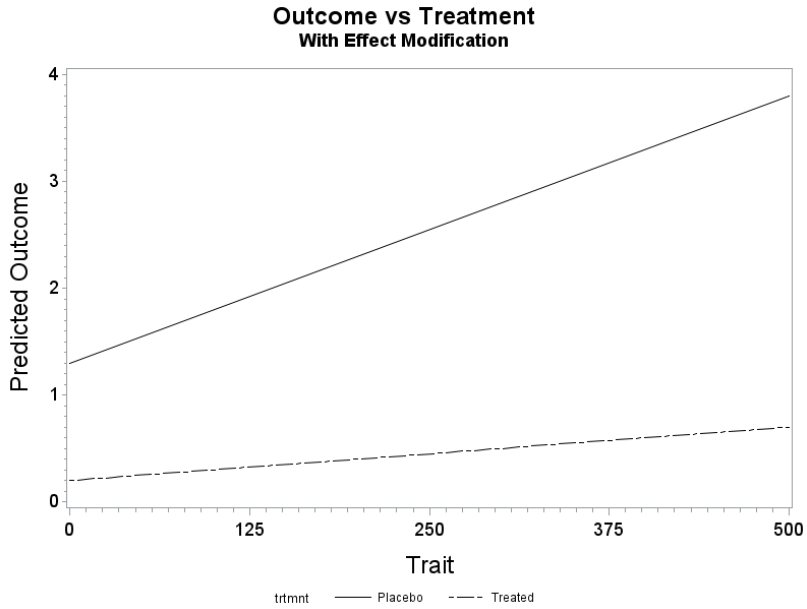
In particular, from Figure 4, if the slope of predicted outcome as a function of trait value is the same in both arms, the trait is not an effect modifier. Instead, differences in outcome between treatment and placebo are generated only by differences in their y-intercepts.

Figure 6. Treatment Effect, Not Modified by Trait



In contrast, when the slopes of outcome with respect to the trait differ between treatment and placebo, the treatment effect, i.e. the difference between treatment and placebo, depends on the value of the trait, and the trait is an effect modifier, as in Figure 7.

Figure 7. Treatment Effect, With Effect Modifier[2]



source: reviewer program idealized fits 04 03.sas

In summary, to evaluate whether or not a trait an effect modifier, we test whether the slope of outcome as a function of the trait value differs between placebo and treatment. This test is accomplished by including in the statistical model terms for the trait and the trait-by-treatment interaction, with effect modification indicated if the trait by treatment interaction is statistically significant.

To avoid wasting statistical power in the context of the current submission, no categories are imposed on continuous or integer valued traits while testing for effect modification. Instead, the continuous or integer values of the trait are used, without any reliance on cut-points.

Results from interaction tests may be examined in this review using graphics such as forest plots, and such graphics may use cut-points to categorize effect modifiers (e.g., Figure 10). However, when the statistical interaction tests are on potential biomarkers originally measured as continuous or integer variables, the categories in the graphs are only visual aids to help to help understand the meaning of the statistical interaction tests.

[2] *after* Wang, Sue-Jane. Biomarker as a classifier in pharmacogenomics clinical trials: a tribute to the 30th anniversary of PSI. Pharmaceutical Statistics 6:283-296.

4.2.2 Design of Study 88 Enrichment Criteria

The applicant analyzed data from study 97 to detect potential effect modifiers, and saw positive associations between treatment effect and two variables, blood eosinophil count and number of exacerbations in the prior year. These two variables were incorporated into enrollment criteria for enrichment of study 88.

Reanalysis of the data from study 97 corroborates the applicant's results. For the reanalysis, I pooled the mepolizumab doses into a single mepolizumab treatment arm, and added terms for each potential effect modifier and its interaction with treatment to the primary analysis model for exacerbation rate. The results, given in Table 25, are consistent with the applicant's analyses, and indicate that blood eosinophil count and number of exacerbations in prior year are effect modifiers which are potentially useful as enrichment criteria for study 88.

Table 25. Treatment by Potential Effect Modifier Interactions, Screening Analyses Study 97

Potential Biomarker	Nominal P-Value
log screening blood eosinophil count	.04
log # exacerbations in prior year	.02
screening exhaled nitric oxide (ppb)	.17
baseline pre-bronchodilator FEV1	.16
screening FEV1 percent reversibility	.06
baseline ACQ-6	.13
sputum eosinophil differential count	.50
loss of control post ≤ 25% OCS reduction	.20

source: reviewer programs exac study 97 Biom 2015 05 28.sas

Blood eosinophil count and number of exacerbations in the prior year were log transformed for the analyses in Table 25, as those analyses are based on normally distributed covariates. Their distributions were strongly skewed to the right, as shown in Figure 8 and Figure 9.

Figure 8. Screening Blood Eosinophil Count Among Enrolled Patients, Study 97

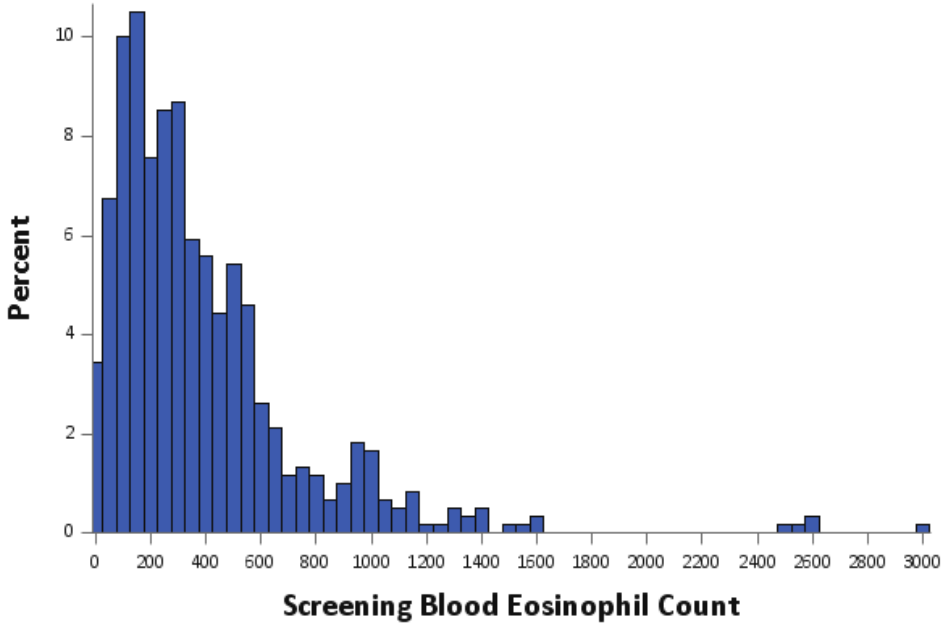
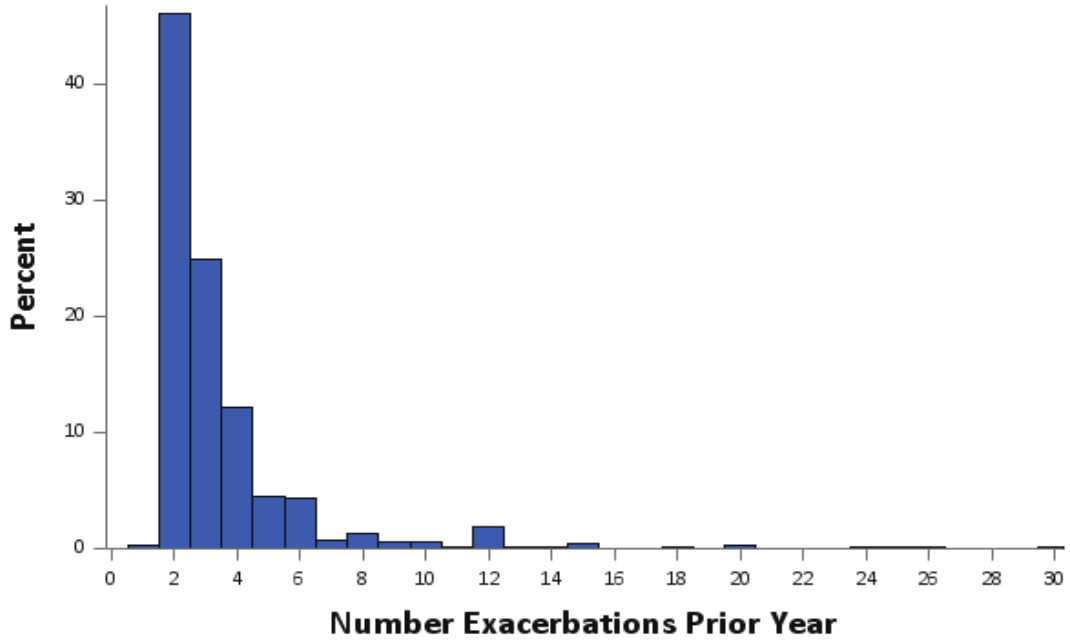


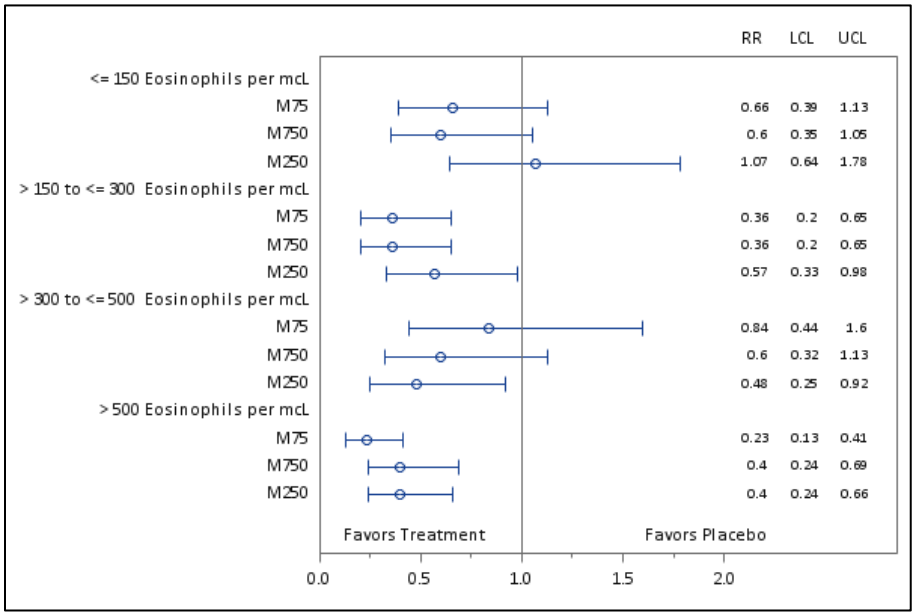
Figure 9. Number of Exacerbations in Prior Year Among Enrolled Patients, Study 97



source: reviewer programs bleos histogram s97.sas

Regarding screening blood eosinophil count, effects on exacerbation rates were significant in study 97 for all three mepolizumab doses among patients with more than 500 screening blood eosinophils per mcL, but were not significant when screening blood eosinophil counts were less than 150 per mcL (Figure 10).

Figure 10. Exacerbation Rate Ratios, by Screening Blood Eosinophil Count, Study 97

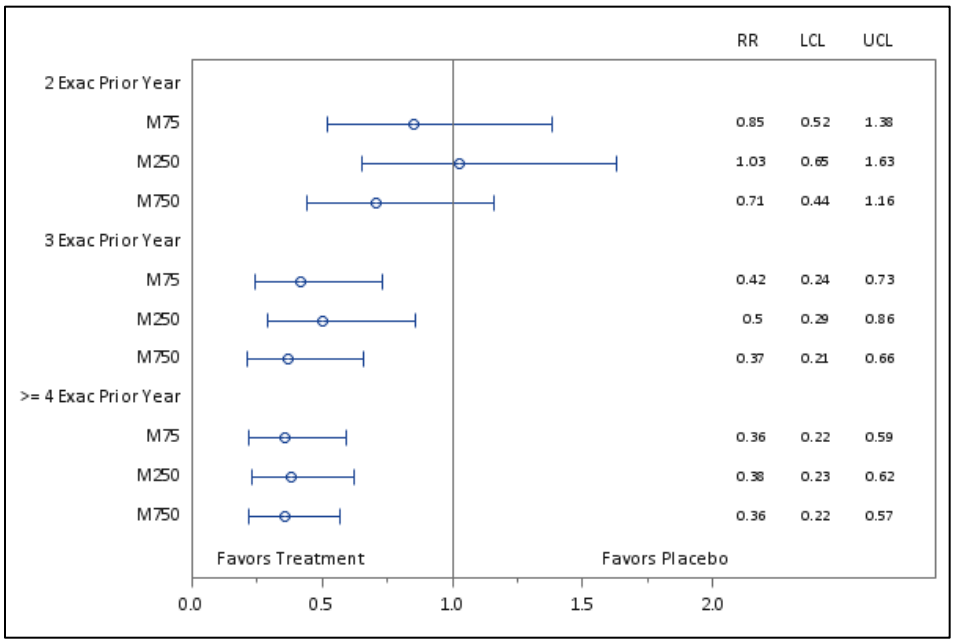


Source: reviewer program Exac Forest Plots Subgr S97 2015 06 17.sas

Similarly, a positive association was seen between treatment effect and number of exacerbations in the prior year, with statistically significant treatment effects seen only among patients with 3 or more exacerbations in the year prior to the trial (Figure 11).

Graphs for the other, non-significant interactions tests in Table 25, are provided in Appendix 6.3.

Figure 11. Exacerbation Rate Ratios, by Number Exacerbations in Prior Year, Study 97



Source: reviewer program Exac Forest Plots Subgr S97 2015 06 17.sas

From the exploratory analyses of study 97, the applicant decided to limit enrollment in subsequent study 88 to patients with two or more exacerbations in the prior year and at least 150 blood eosinophils per mcL.

However, without supportive analyses from clinical data, the applicant decided to also enroll patients with at least one blood eosinophil count in the past year $\geq$ 300 cells per mcL. In addition to lack of supportive analyses, this expanded enrollment criterion was measured without careful considerations of measurement methodology typically required for variables to be included on the product label. In particular, for any given patient, demonstration of at least 300 eosinophils per mcL in the past year could have depended on the number of times blood draws were counted; however the number of such draws was neither controlled nor recorded. Further, blood counts could have been conducted using different platforms, and without standardization of counts to those from the Coulter LH750 used in the confirmatory trials, such eosinophil counts could not be considered indicative of any particular percentile in the reference range. And finally, as alluded to on page 142 of the applicant's clinical study report, and confirmed during the June 11, 2015 Advisory Committee meeting for this drug, such historical blood counts were often patient reported rather than derived from medical records.

4.2.3 Effect Modification in Study 88

As in study 97, screening blood eosinophil count and number of exacerbations in the prior year were log transformed for the analyses in Table 26, as those analyses are based on normally distributed covariates. The distributions were skewed to the right (Figure 12 and Figure 13).

Figure 12. Screening Blood Eosinophil Count Among Enrolled Patients, Study 88

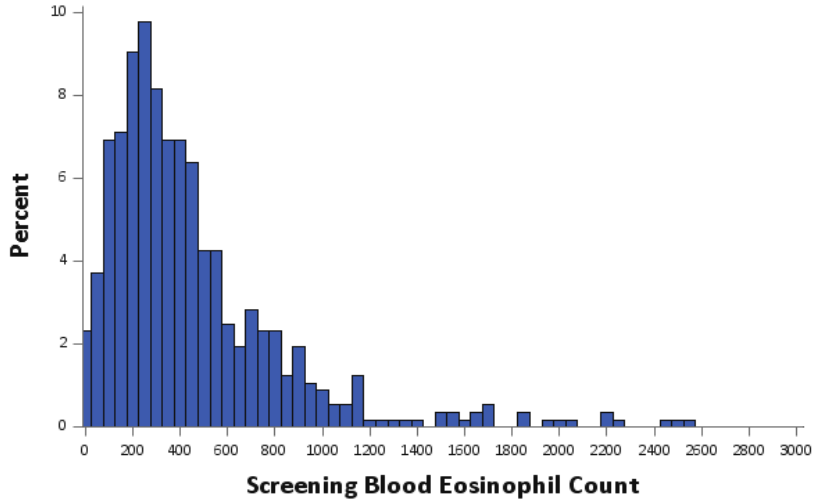
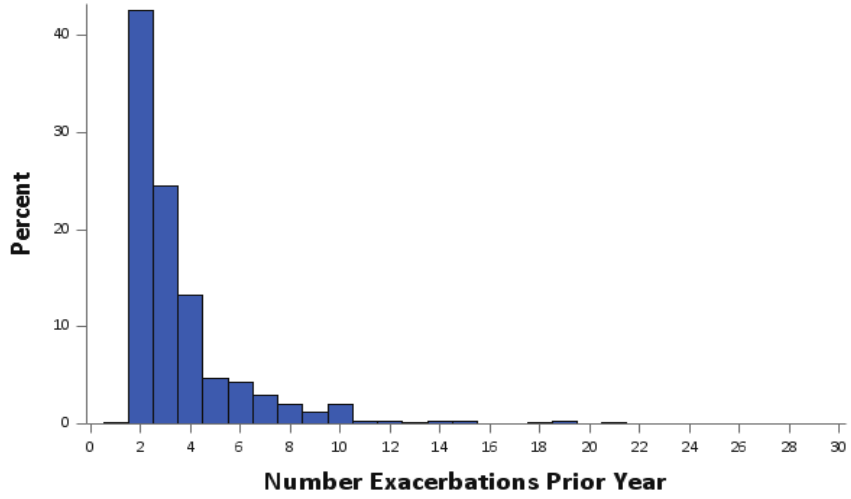


Figure 13. Number of Exacerbations in Prior Year Among Enrolled Patients, Study 88



source: reviewer programs bleos histogram s97.sas, bleos histogram s88.sas

Analysis of the potential effect modifiers noted in study 97 revealed a significant interaction between treatment and screening blood eosinophil count (p=.03,). However, the interaction between treatment and number of exacerbations in the prior year was not significant (p=0.7, Table 26).

Table 26. Treatment by Potential Effect Modifier Interactions, Study 88

Potential Biomarker	Nominal P-Value
log screening blood eosinophil count	.03
log # exacerbations in prior year	.7

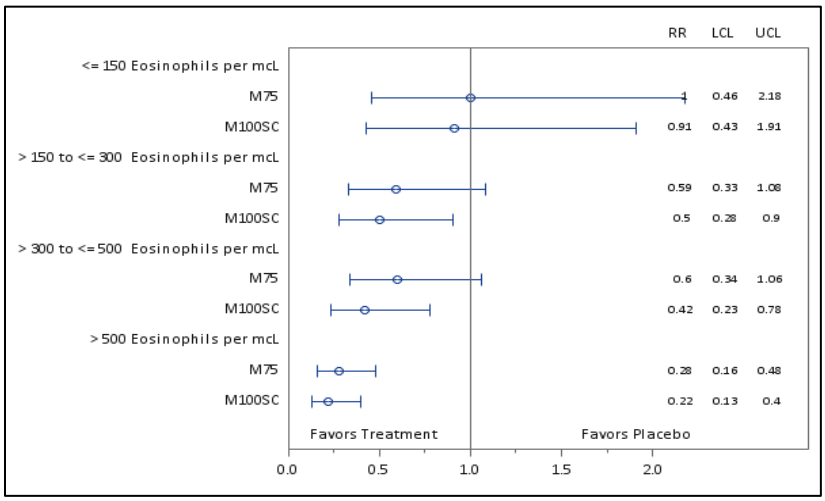
source: reviewer programs program exac study 88 Biom 2015 04 17.sas

As indicated by the significant treatment interaction, the association in study 88 between screening blood eosinophil count and effect of mepolizumab on exacerbation rate is positive, Figure 14.

The additional inclusion criterion allowing enrollment if blood eosinophil count was ≥ 300 cells/mcL in the past year did not contribute to population enrichment. In particular, for patients enrolled with screening blood eosinophil counts ≤ 150 cells/mcL, who were enrolled solely because historical blood eosinophil counts were ≥ 300 cell/mcL, the point estimated rate ratios indicated nearly no treatment effect (Figure 14).

Given the lack of underlying supportive analyses from study 97 and the lack of standardization in measurement of blood eosinophil counts in the prior year discussed in Section 4.2.2, it is perhaps not surprising that patients who were enrolled solely because they met the ≥ 300 cell/mcL criterion, those in study 88 with fewer than 150 blood eosinophils per mcL at screening, did not show any treatment effect (Figure 14).

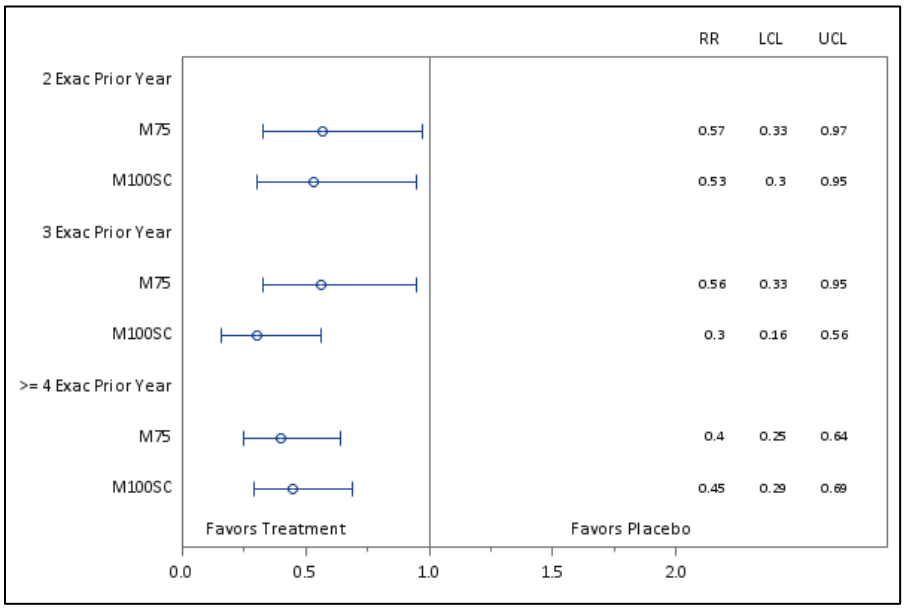
Figure 14. Exacerbation Rate Ratios, by Screening Blood Eosinophil Count, Study 88



Source: reviewer program Exac Forest Plots Subgr S88 2015 04 06.sas

As expected from the lack of statistical significance in Table 26, number of exacerbation in the prior year was not associated with any trend in treatment effect (Figure 15).

Figure 15. Exacerbation Rate Ratios, by Number Exacerbations in Prior Year, Study 88



Source: reviewer program Exac Forest Plots Subgr S88 2015 04 06.sas

The analyses described in Table 25 and Table 26 pooled the various mepolizumab doses into a single mepolizumab arm before comparison to placebo. While such pooling may increase the power of the interaction test, studies 88 and 97 were conducted in part to discern potential differences between doses, and pooling the doses will obscure any differences between them. Therefore, for dose ranging trials 88 and 97, additional analyses was conducted in which doses were not pooled (Appendix 6.4).

In Appendix 6.4, loss of statistical power to detect interactions, associated with lack of pooling, seemed to obtain in study 88. However, with more precise measurements of blood eosinophil count, provided by analyzing blood eosinophil averaged at baseline and screening, effect modifications associated with blood eosinophil count were clearly evident (e.g., Figure 24, Appendix 6.4).

As a final note regarding effect modification, analyses were also performed in studies 6, 75, and 88 to examine the effect of blood eosinophil count on another endpoint, ΔFEV1. Baseline blood eosinophil count was not seen to significantly affect ΔFEV1 in any of the studies; with log baseline blood eosinophil count and log baseline blood eosinophil count by treatment interaction added to the preplanned model, the log baseline blood eosinophil count by treatment interaction was not significant for study 6 (p=.8), study 75 (p=.4), study 88 (p=.999), or study 97 (p=.3).

In summary, studies 88 and 97 provide evidence of a positive association between blood eosinophil count and the effect of mepolizumab on exacerbation rate.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical issues

Two issues were identified in this review. The clinical study reports and the proposed label consistently ignored measures taken to control type 1 error in the face of multiple endpoints and comparisons. Such lack of control, however, has been addressed in this review.

Additionally, although screening blood eosinophil count was seen to be an effect modifier, its precision may be enhanced by using the average of multiple measurements, taken at least one week apart. The literature further suggests that specifying the time of day at which measurements are taken may improve assessment of patient eosinophil status.

5.2 Collective Evidence

There is strong evidence that when compared to placebo, mepolizumab reduces the rate of clinical exacerbations in patients with severe asthma. There is also strong evidence of a positive association between treatment effect and blood eosinophil count. The submission provides no evidence that historical eosinophil counts influence treatment effect.

The available evidence does not support a blood eosinophil count cutoff for prescribing mepolizumab. In particular, because the medical officer's review does not indicate that mepolizumab poses any safety issues, there are no values of eosinophil count for which the expected risks of mepolizumab treatment exceed expected benefits. Although expected treatment benefits may be low among patients with low eosinophil count, such patients are experiencing serious disease despite standard of care and, with lack of alternative treatments and lack of demonstrable risks, mepolizumab may be considered a treatment option.

A single study provides evidence that treatment with mepolizumab reduces dependence on OCS for control of asthma. Results from the study suggest that effectiveness for OCS reduction may be dependent on age, with significant benefits of treatment limited to patients at least 40 years old.

5.3 Conclusions and Recommendations

Two randomized, placebo-controlled, blinded, parallel arm studies provide strong evidence that, compared to placebo, mepolizumab reduces the rate of clinical exacerbations in patients with severe asthma who have moderate to high blood eosinophil counts.

Interaction tests strongly indicate a positive association between screening blood eosinophil count and treatment effect for reduction of exacerbation rate. However, the evidence does not support the applicant's assertion that eosinophil counts above 300 per mcL during the year prior to treatment indicate that treatment will be effective.

A single randomized, placebo-controlled, blinded, parallel arm study provides evidence that treatment with mepolizumab reduces dependence on OCS for control of asthma.

Enrollment of adolescents and African Americans was not sufficient to demonstrate statistically significant treatment effects in these subgroups.

5.4 Labeling Recommendations

I recommend the following changes in the label:

(i) removal of proposed blood eosinophil count cut points from the indication, since there is no clear cut point for which the risk associated with mepolizumab exceeds the benefits,

(ii) [redacted] (b) (4)

(iii) inclusion of language in the indication to inform physicians of the positive association between blood eosinophil count and treatment effect,

(iv) denoting cut points, whether included as text or graphics such as forest plots, in terms of percentage of reference range or in terms of manual counts to at least partially account for differences between measuring platforms used in clinical practice,

(v) inclusion of absolute reductions in exacerbation rates, to inform physicians of absolute benefits associated with prescribing treatment,

(vi) removal of ≥ 300 eosinophils per mcL as an inclusion criterion for study 88, since it was not really measured among candidates for inclusion in that trial, and because neither exploratory nor confirmatory analyses support its use as an indicator for treatment benefit.

(vii) [redacted] (b) (4), and

(viii) inclusion of forest plots showing effects of screening blood eosinophil count on effectiveness for reduction of exacerbations,

Proposed product labeling should also be reevaluated for potential:

(ix) [redacted] (b) (4)

(x) [redacted] (b) (4)

(xi) [redacted] (b) (4)

6 Appendices

6.1 Baseline Demographic Characteristics

Table 27. Baseline Demographics, Study 6

Category	Pbo	250 mg IV	750 mg IV
Randomized	126	120	116
Age (mean)	37	36	36
Male (%)	38%	43%	52%
Race (%)			
White	84%	89%	81%
Black	14%	6%	16%
Asian	2%	0%	2%
Other	0%	5%	1%
Weight (mean kg)	74	75	75
Height (cm)	168	170	171
Blood Eos Count	404	344	342

Source: CSR Table 8, reviewer program fev s06 2015 02 20

Table 28. Baseline Demographics, Study 97

Category	Pbo	75 mg IV	250 mg IV	750 mg IV
Randomized	155	153	152	156
Age (mean)	46	50	49	49
Male (%)	37%	32%	39%	40%
Race (%)				
White	90%	91%	89%	90%
Black	4%	3%	5%	3%
Asian	5%	6%	5%	6%
Other	<1%	0%	1%	<1%
Weight (mean kg)	78	78	79	81
Height (cm)	167	165	167	168
Blood Eos Count	421	370	398	364

Source: CSR Table 8, reviewer program Exac Forest Plots Subgr S97 2015 04 06.sas

Table 29. Baseline Demographics, Study 88

Category	Pbo	75 mg IV	100 mg SC
Randomized	191	191	194
Age (mean)	49	50	51
Male (%)	44%	45%	40%
Race (%)			
White	77%	79&	78%
Black	2%	3%	4%
Asian	20%	17%	18%
Other	1%	<1%	<1%
Weight (mean kg)	75	77	73
Height (cm)	165	166	165
Blood Eos Count	460	419	456

Source: CSR Tables 6 and 5.11, reviewer program exac forest plots subgr s88 2015 04 17

Table 30. Baseline Demographics, Study 75

Category	Pbo	100 mg SC
Randomized	66	69
Age (mean)	50	50
Male (%)	55%	36%
Race (%)		
White	92%	97%
Black	0%	0%
Asian	3%	3%
Other	6%	1%
Weight (mean kg)	87	79
Height (cm)	172	169
Blood Eos Count	347	413

Source: CSR Tables 9 and 5.12, reviewer program OCS Perc S 75 2015 05 21.sas

6.2 Tipping Point Analyses

For exacerbations, tipping point sensitivity analyses were conducted in which mean exacerbation rates ranged in a 'tipping point grid' from 1 to 5 exacerbations per year in increments of 0.5 on placebo and treatment arms separately. The grid included the possibility that missing data from the mepolizumab arms had worse outcomes than that from the placebo arms.

To complete each data record, a random draw was made from the negative binomial model from the primary analysis, with the expected value fixed according mean exacerbation rate on the tipping point grid described in the immediately preceding paragraph. For each patient, the assumed exacerbation rate after withdrawal did not depend on exacerbation rate prior to withdrawal. The data was then analyzed using the primary analysis model, with results combined across imputations using Rubin's method.

In addition to examining individual treatment arms against placebo, to help understand overall tipping points, analyses were also provided which compared the combined mepolizumab arms against placebo.

For study 97 combined mepolizumab doses versus placebo and M750 versus placebo, no p-values greater than .05 were seen on the tipping point grids for any combination of post withdrawal rates (Figure 16). For M75 versus placebo, p-values greater than .05 were seen only for an assumed post-withdrawal mepolizumab rate of 5 per year and an assumed post-withdrawal placebo rate of 1 per year. For M250 versus placebo, p-values greater than .05 were seen for post withdrawal mepolizumab rates of 4 per year or greater with a placebo rate less than or equal to 2.5 per year.

Even in the worst case, seen for M250, the combinations of rates yielding p-values greater than .05 requires that patients withdrawn from the mepolizumab arm have at least 1.5 more exacerbations per year than withdrawn patients from the placebo arm. Because this scenario seems unlikely, I conclude that the results showing mepolizumab superior to placebo are robust in the face of patient withdrawal.

Similar tipping point analyses were conducted for exacerbation rates in study 88 and for OCS withdrawal in study 75. For study 88, the tipping point grid was the same as that for study 97; no combination of exacerbation rates yielded p-values greater than 0.05 for the mepolizumab arms combined or for either of the mepolizumab dosage arms alone versus placebo. For study 75, a p-value greater than .05 (equal to .051) was seen only when all missing patients from the mepolizumab arm were in the worst category, 'no decrease in OCS,' and all missing patients from the placebo arm were in the best category, '90 to 100% OCS reduction.' Again, this scenario in which efficacy is completely reversed seems unlikely, and I therefor conclude that results for the primary analyses in studies 75 and 88 are robust in the face of missing data.

Figure 16. Tipping Point Analysis for Exacerbation Rate, Study 97

Mepolizumab all doses vs. placebo

Placebo rate post withdrawal	Mepolizumab rate post-withdrawal								
	1	1.5	2	2.5	3	3.5	4	4.5	5
1	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	0.004	0.01	0.021
1.5	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	0.005	0.012
2	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.001	0.003	0.007
2.5	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	0.004
3	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002
3.5	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.001
4	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
4.5	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
5	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Mepolizumab 75mg IV vs. placebo

	Mepolizumab rate post-withdrawal								
	1	1.5	2	2.5	3	3.5	4	4.5	5
1	<0.001	<0.001	<0.001	0.002	0.004	0.01	0.02	0.037	0.061
1.5	<0.001	<0.001	<0.001	<0.001	0.003	0.006	0.013	0.025	0.042
2	<0.001	<0.001	<0.001	<0.001	0.001	0.004	0.008	0.017	0.032
2.5	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	0.005	0.012	0.021
3	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	0.004	0.007	0.015
3.5	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	0.005	0.01
4	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	0.004	0.007
4.5	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.001	0.002	0.005
5	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	0.004

Mepolizumab 250mg IV vs. placebo

	Mepolizumab rate post-withdrawal								
	1	1.5	2	2.5	3	3.5	4	4.5	5
1	0.002	0.004	0.008	0.015	0.028	0.047	0.074	0.111	0.156
1.5	0.001	0.002	0.005	0.01	0.018	0.031	0.052	0.079	0.113
2	<0.001	0.001	0.003	0.006	0.011	0.021	0.034	0.055	0.082
2.5	<0.001	<0.001	0.002	0.003	0.008	0.013	0.024	0.038	0.059
3	<0.001	<0.001	<0.001	0.002	0.005	0.009	0.016	0.027	0.044
3.5	<0.001	<0.001	<0.001	0.001	0.003	0.006	0.011	0.019	0.031
4	<0.001	<0.001	<0.001	<0.001	0.002	0.004	0.008	0.014	0.022
4.5	<0.001	<0.001	<0.001	<0.001	0.001	0.003	0.005	0.01	0.017
5	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	0.004	0.007	0.011

Figure 16 (continued)

Mepolizumab 750mg IV vs. placebo										
		Mepolizumab rate post-withdrawal								
		1	1.5	2	2.5	3	3.5	4	4.5	5
	1	<0.001	<0.001	<0.001	<0.001	0.002	0.004	0.01	0.021	0.039
	1.5	<0.001	<0.001	<0.001	<0.001	<0.001	0.003	0.006	0.014	0.026
	2	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	0.004	0.009	0.018
	2.5	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.003	0.006	0.012
	3	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	0.004	0.008
	3.5	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.001	0.003	0.006
	4	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	0.004
	4.5	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.001	0.003
	5	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002

source: BLA 125526 sequence 0007 response to information request

6.3 Screening of Study 97. Potential Effect Modifiers Found Not Significant

Figure 17. Exacerbation Rate Ratios, by Screening Exhaled NO, Study 97

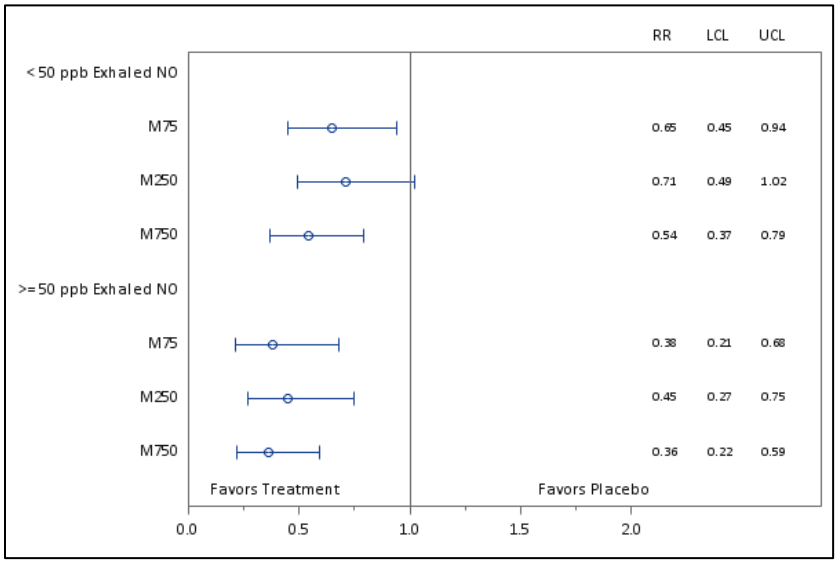


Figure 18. Exacerbation Rate Ratios, by Baseline Pre-bronchodilator FEV1, Study 97

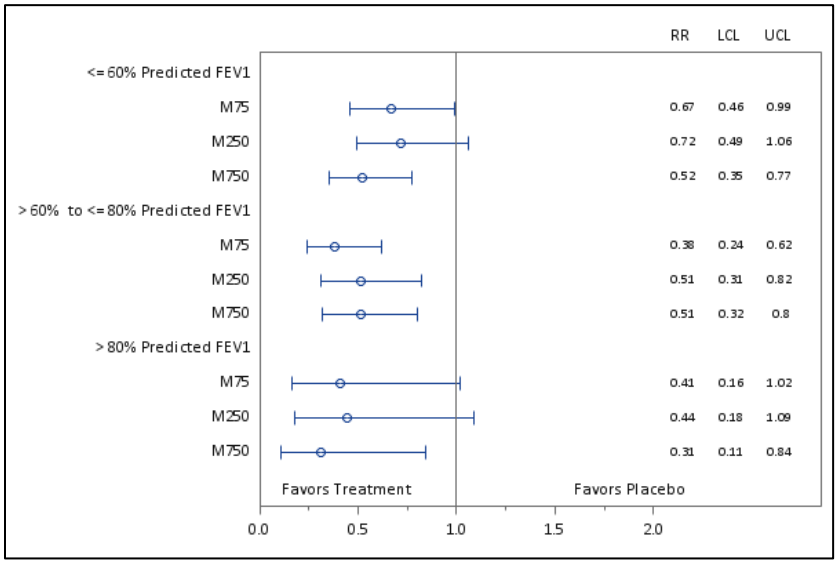


Figure 19. Exacerbation Rate Ratios, by Screening FEV1 Reversibility, Study 97

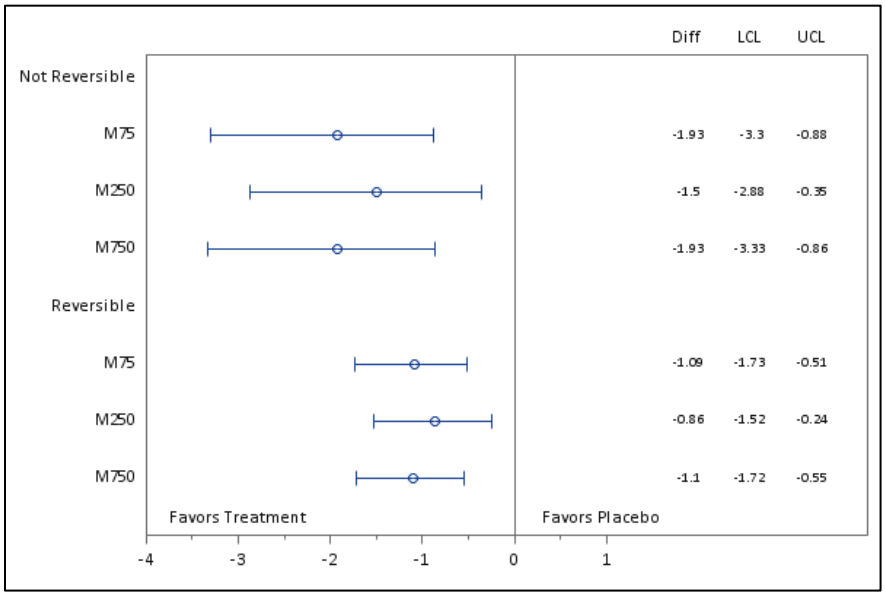


Figure 20. Exacerbation Rate Ratios, by Screening ACQ-6, Study 97

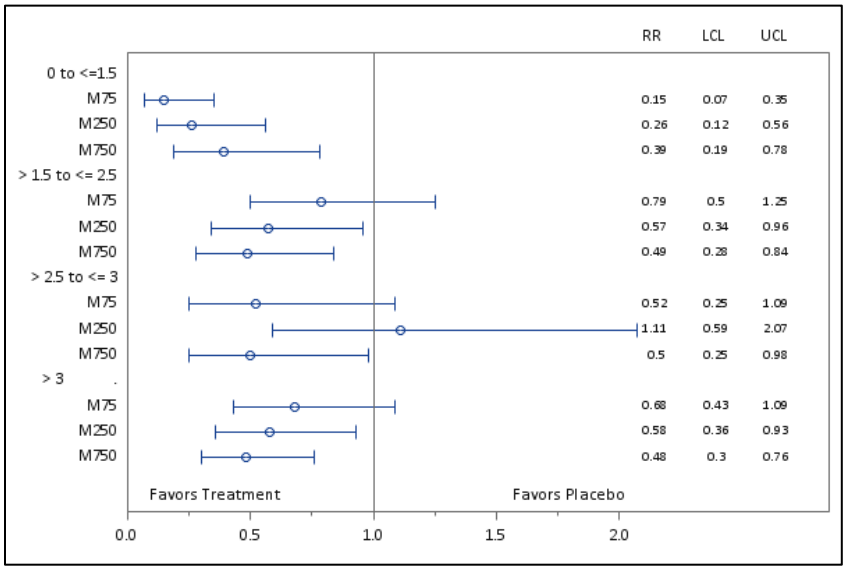


Figure 21. Exacerbation Rate Ratios, by Screening Sputum Eosinophil Count, Study 97

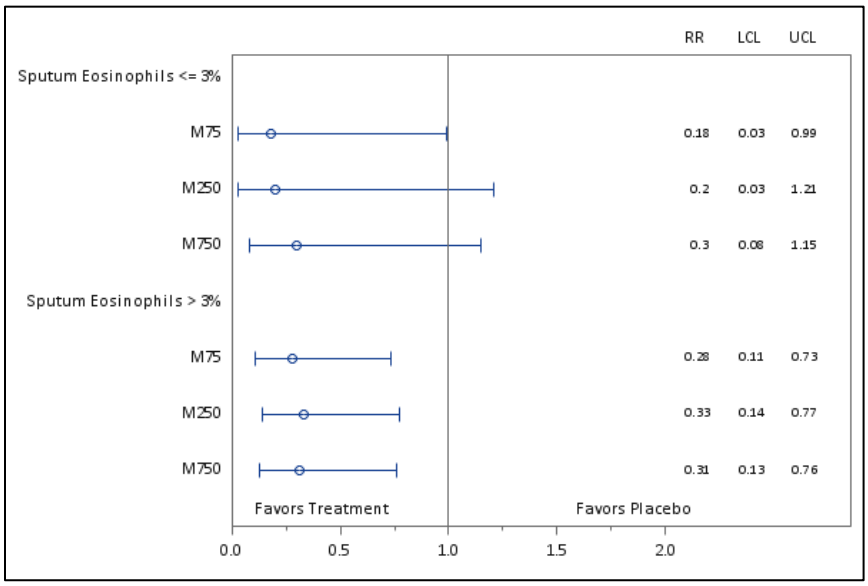
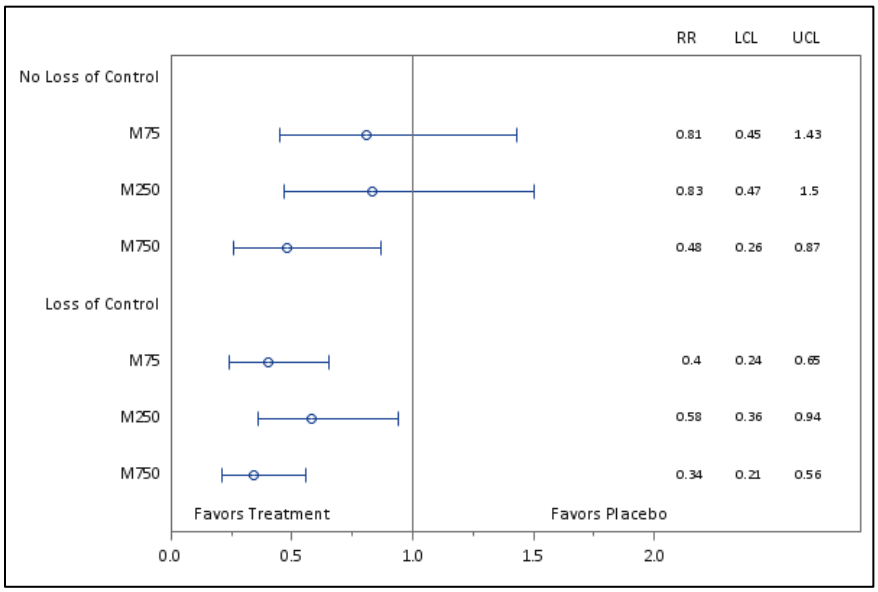


Figure 22. Exacerbation Rate Ratios, by Loss of Asthma Control at Screening with ≤ 25% Reduction in Steroid Dose, Study 97



Source: reviewer program Exac Forest Plots Subgr S97 2015 06 05.sas

6.4 Blood Eosinophils as Effect Modifier: Analysis by Unmerged Mepolizumab Treatments

Figure 7 shows that a proper test for whether a proposed trait is an effect modifier will evaluate whether the outcome slopes with respect to the trait differ between placebo and treatment. As discussed in Section 4.2.1.2, this test is accomplished by including in the statistical model terms for trait by trait treatment interaction; a significant trait by treatment term demonstrates a difference between the slopes which indicates effect modification.

For individual treatment arms, interpretation of results from an analysis which includes trait and trait by treatment interactions is straightforward. Consider a regression with three independent variables, treatment (x), trait (z), and trait by treatment interaction (xz). Then for each patient's response (y) the regression model is:

$$y = \alpha + \beta x + \gamma xz + \varepsilon$$

which rearranges to

$$y = a + x(\beta + \gamma z) + \varepsilon$$

The effect of treatment x may therefore be considered a composite variable which can be further analyzed by evaluating: β, defined as the effect of treatment x when trait value z equals zero, and γ, the additional effect of treatment x per unit increase in the value of the biomarker.

If γ is different in different treatment arms, then trait z is an effect modifier.

Regarding the y-intercept, most statistical packages provide values for each treatment which, by default, are at the reference values for each of the covariates. Such intercepts are not necessarily representative of the sample. Therefore, in this review, covariates were recoded to provide intercepts obtained as the predicted value of the covariate means.

For tests of effect modification we add the effect modifier and a term for its interaction with treatment to the applicant's primary analysis model.

For a model in which mepolizumab treatments are not merged or averaged for comparison to placebo, the interaction between screening blood eosinophils and treatment is significant in studies 97 but not in study 88 (Table 31).

Table 31. Treatment by Log Screening Blood Eosinophil Count Interaction Tests for Exacerbations

Study	P-Value for Interaction with Treatment
97	.047
88	.058

A similar lack of statistical significance in study 88 holds for the treatment by baseline eosinophil count (Table 32).

Table 32. Treatment by Log Baseline Blood Eosinophil Count Interaction Tests for Exacerbations

Study	P-Value for Interaction with Treatment
97	.004
88	.12

Although the log screening and baseline blood eosinophil by treatment interactions was not statistically significant for study 88 (Table 31 and Table 32), its significance for study 97 in Table 31 and Table 32 and its near significance in Table 31 suggests that the interaction may, in fact, not be spurious.

Given that the exacerbation data underlying Table 31 and Table 32 was the same, differences between these two tables regarding interaction tests must be the direct result of differences between baseline and screening eosinophil counts. That such differences are plausible is supported by numerous authors, who have noted large within patient variability in measured blood eosinophil count.[3,4,5,6] Within day and between week differences are commonly cited as being important sources of variation. However, regardless of whether caused by imprecision in measurement or by instability in patient blood eosinophil levels, such variations may obscure assessment of patient status, and may therefore obscure statistically significant impacts of eosinophil count on treatment effect.

[3] Rudd, F. 1947. Acta Psychiatrica et Neurologica. Supplement XL.
[4] Acland JD, and AH Gould. 1956. J Physiol. 133:456–466.
[5] Spector, SL and RA Tan. 2012. Journal of Asthma. 49(8): 807-810.
[6] Tatai K, and S Ogawa, 1951. Japan J. Physiol. 1: 328-33

As expected from the above-mentioned literature and differences between baseline and screening eosinophil count p-values for interactions with treatment (Table 31 and Table 32), patient eosinophil counts varied widely between screening and baseline. For example, Table 33 provides, for study 97, probabilities of change in eosinophil count categories[7] from screening to baseline. For example, among enrolled patients who had low eosinophil counts at screening, 56% were still low at baseline, 24% became medium low at baseline, 10% became medium high at baseline, and 11% became high eosinophil count patients at baseline. From the diagonal, the probability of remaining in the same quartile from baseline to screening ranged from a low of 43% to a high of 69%.

Transition probabilities were similar for study 88 (Table 34), in which baseline measurements were taken one week after screening rather from one to six weeks after screening as in study 97.

Table 33. Transition Matrix* for Blood Eosinophil Count, from Screening to Baseline, Study 97

	BL	**BML**	**BMH**	**BH**
SL	0.56	0.24	0.10	0.11
SML	0.29	0.49	0.13	0.09
SMH	0.14	0.18	0.43	0.26
SH	0.07	0.05	0.19	0.69

Source: reviewer program markov eosin count s97 2015 04 20.sas
*For study 97, L, ML, MH, H are low medium-low, medium high, and high eosinophil quartiles, bounded by 0, 150, 290, and 500 eosinophils per microliter

Table 34. Transition Matrix* for Blood Eosinophil Count, from Screening to Baseline, Study 88

	BL	**BML**	**BMH**	**BH**
SL	0.63	0.19	0.12	0.06
SML	0.34	0.45	0.14	0.07
SMH	0.17	0.13	0.43	0.27
SH	0.06	0.08	0.20	0.65

Source: reviewer program markov eosin count s97 2015 04 20.sas
*For study 88, L, ML, MH, H are low medium-low, medium high, and high eosinophil quartiles, bounded by 0, 205, 345, and 560 eosinophils per microliter

[7] categories based on quartiles at screening and baseline among patients who were randomized to treatment

Transition rates between quartiles from screening to baseline, with approximately 50% of patients leaving their original quartile, as described in Table 33 and Table 34, indicate high variability. It therefore seems likely that precision for the assessment of overall patient eosinophil status may be improved by using the average of multiple eosinophil counts rather than just a single eosinophil count.

Compared to single screening or baseline measurements, exploratory analyses using the average of screening and baseline measurements suggest that the low, but not statistically significant, p-values for interactions of blood eosinophil count with treatment seen for study 88 in Table 31and Table 32 were not spurious. It instead seems likely that effect modification was obscured by variability in eosinophil count which, in turn, impeded evaluation of patient status. In particular, the interaction of treatment with log of averaged screening and baseline eosinophil count interaction term was of clear nominal significance, with the p-value for study 88 equal to .004 (Table 35).

Table 35. Exploratory Analysis. Treatment by Log Average Blood Eosinophil Count Interaction Tests for Exacerbation Rate

Study	P-Value for Interaction with Treatment
97	.006
88	.004

source: reviewer programs exac study 97 Biom 2015 06 04.sas and program exac study 88 Biom 2015 06 04.sas

Evaluation of the intercept and slope parameters for the log exacerbation rate as a function of log average blood eosinophil count in study 97 suggested that intercepts for exacerbation rate were significantly lowered compared to placebo for all three mepolizumab doses, and that slopes were lower compared to placebo for the M75 and M250 doses.

Table 36. Exploratory Analysis. Exacerbation Parameters for Log Average Eosinophil Count by Treatment Interaction, Study 97

Study	Parameter	Difference from Pbo (P-Value)*			
		Pbo	75 mg IV	250 mg IV	750 mg IV
97	Intercept (β)	0.789	-1.16	-1.07	-0.898
		(.0001)	(<.0001)	(<.0001)	(.0003)
	Slope (γ)	0.221	-0.365	-0.423	-0.156
		(.049)	(.014)	(.002)	(.3)

source: reviewer program Exac study 97 Biom 2015 04 03.sas

For study 88, intercepts and slopes for log exacerbation rate as a function of log average eosinophil count were significantly lowered compared to placebo for both mepolizumab doses (Table 37).

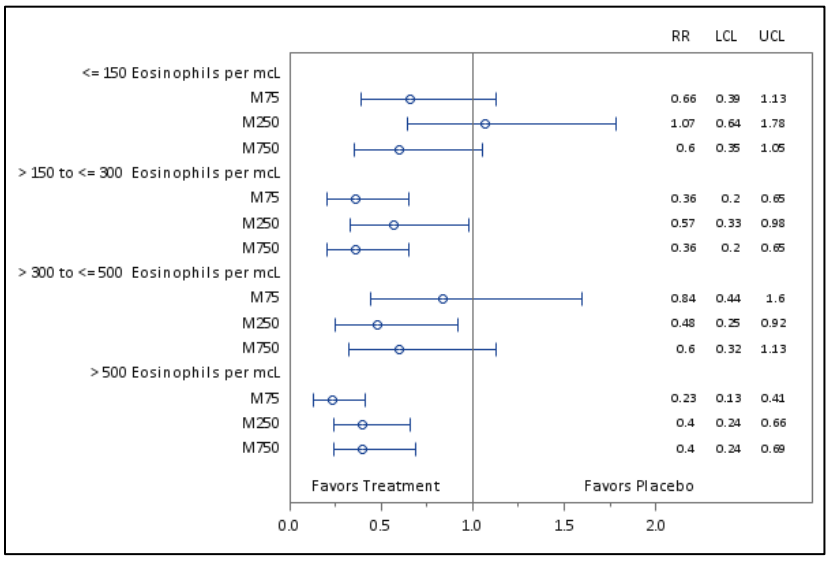
Table 37. Exploratory Analysis. Exacerbation Parameters for Log Average Eosinophil Count by Treatment Interaction, Study 88

Study	Parameter		Difference from Pbo (P-Value)*	
		Pbo	75 mg IV	100 mg SC
88	Intercept (β)	0.170 (.4)	-1.226 (<.0001)	-1.423 (<.0001)
	Slope (γ)	0.169 (.20)	-0.489 (.02)	-0.588 (.002)

source: reviewer program Exac study 88 Biom 2015 04 17.sas

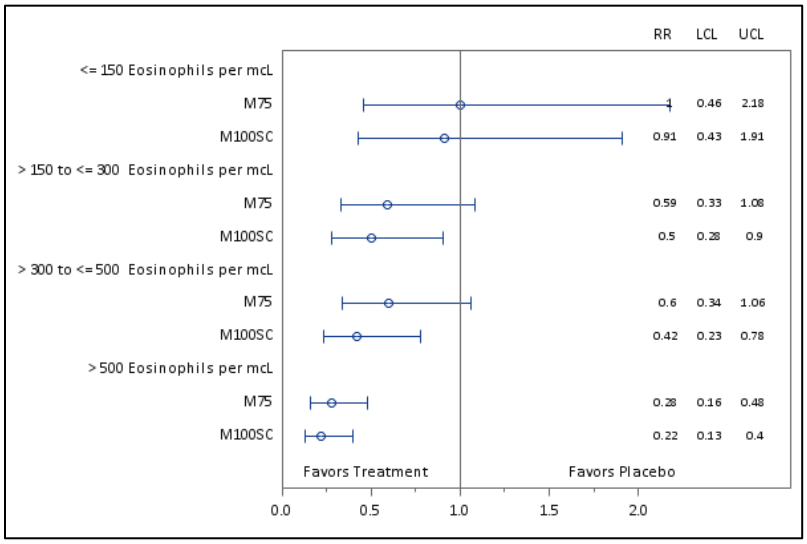
Forest plots detailing reductions in exacerbation rate according to average eosinophil count in studies 97 and 88 are provided in Figure 23 and Figure 24 respectively.

Figure 23. Exacerbation Rate Ratios, by Average Blood Eosinophil Count, Study 97



Source: reviewer program Exac Forest Plots Subgr S97 2015 06 17.sas

Figure 24. Exacerbation Rate Ratios, by Average Blood Eosinophil Count, Study 88



Source: reviewer program Exac Forest Plots Subgr S88 2015 06 17.sas

In summary, exploratory analyses indicate that blood eosinophil count modifies the effect of treatment on exacerbation rate for M75 (Table 36 and Table 37), M100 SC (Table 37), and M250 (Table 36). Compared to a single screening or baseline estimate of blood eosinophil count, the analyses also suggest that the average of screening and baseline blood eosinophil counts provides a more precise estimate of patient eosinophil status than either measurement alone.

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

/s/

ROBERT ABUGOV
07/10/2015

DAVID M PETULLO
07/10/2015
I concur.

THOMAS J PERMUTT
07/10/2015
I concur.

STATISTICAL REVIEW AND EVALUATION

Biometrics Division: VI

BLA NO.:	125526
DATE RECEIVED BY OB:	February 05, 2015
DRUG NAME:	Mepolizumab
DOSAGE FORM:	lyophilized powder for reconstitution
INDICATION:	Add-on maintenance treatment for patients 12 years of age and older with severe eosinophilic asthma as identified by certain blood eosinophil levels.
ROUTE OF ADMINISTRATION	subcutaneous
STRENGTHS	100 mg/vial
SPONSOR:	GSK
REVIEW FINIHSED:	April 10, 2015
STATISTICS REVIEWER:	Xiaoyu (Cassie) Dong, Ph.D.
PROJECT MANAGER:	Melinda Bauerlien

Concur: _____

Meiyu Shen, Team Leader, Mathematical Statistician, CDER/OTS/OB/DB VI

Yi Tsong, Division Director, DBVI, CDER/OTS/OB/DB VI

Distribution: BLA 125526

CDER/OTS/OB/DB VI/ Yi Tsong
CDER/OTS/OB/DB VI/ Meiyu Shen
CDER/OTS/OB/ Lillian Patrician
CDER/OBP/ Dr. Marjorie Shapiro
CDER/OBP/Melinda Bauerlien

TABLE OF CONTENTS

I. FDA REVIEWER'S RECOMMENDATIONS AND COMMENTS

In summary, we recommend the following for the sponsor regarding their qualification and characterization studies of the small scale model:

- Increase the sample sizes used in the qualification studies;
- When the observed variance at the small scale is less than the observed variance at the full scale, we recommend conducting the one-sided F-test to test if the unknown true variance of the small scale is less than the variance at the full-scale.
- We recommend the statistical equivalence testing to assess the equivalence in means.
- We also recommend providing scatter-plots of the individual data for each tested attribute for a side-by-side comparison between the small-scale and the full-scale.
- Provided more information of the multivariate model development as well as the final fitted models in Process Characterization Studies;
- Provide more information on the Bayesian predictive model development.

Our detailed statistical comments for each study are provided below.

In S.2.6. entitled "*Manufacturing Process Development_Mepolizumab*", the sponsor conducted the qualification of the small-scale model to evaluate the equivalency in performance at the small scale and at the full-scale. The purpose of this study is to qualify the scale-down model's ability to predict the performance at the full-scale. In particular, the sponsor conducted the statistical F-test to assess the similarity in variance, and conducted the statistical t-test to evaluate the similarity in the means between the small-scale and the full-scale.

We have the following comments regarding GSK's F-test to assess the similarity in variance.

- The F-test is to test if the variances of two groups are significantly different. Although having a significant result (p-value < 0.05) indicates the variance of the two scales is statistically different, having a non-significant value (p-value > 0.05) is not a sufficient evidence to demonstrate equivalency. Considering one of the purposes of the qualification study is to demonstrate that "*the variance of an attribute response at small-scale must be equal to or less than the variance of that attribute at full scale*", the F-test is not a suitable statistical approach to achieve such a study goal.
- The sample sizes used for the F-test are insufficient for some studies. Specifically, the sample sizes of the studies listed in the table below are too small to have a reliable statistical testing result. We recommend GSK increasing the sample sizes.

Table 1: List of GSK's Qualification Studies with Insufficient Sample Sizes to Assess the Equivalency in Variances between Small Scale and Full-scale.

Source	Study Title	Sample Size	Scale
Table 84	(b) (4)		
Table 87			
Table 88			
Table 89			
Table 90			
Table 91			

- When the observed sample variance of the small scale is less than the variance at the full-scale, the one-sided F-test should be performed to test if the unknown true variance of the small scale is less than the variance at the full-scale.
- We also recommend providing scatter-plots of the individual data for each attributes tested for a side-by-side comparison between the small-scale and the full-scale.

We have the following comments regarding GSK's *t*-test to assess the equivalency in mean values between the small-scale and the full-scale.

- Similarly as the F-test, the *t*-test is to test if the mean values of two groups are significantly different. Although having a significant result (p-value < 0.05) indicates the mean values are statistically different, having a non-significant value (p-value > 0.05) is not a sufficient evidence to demonstrate equivalency. Considering another purpose of the qualification study is to demonstrate that "*the means for each attribute must be statistically equivalent between scales*", the *t*-test is not a suitable statistical approach to achieve the study purchase. Instead, we recommend using the statistical equivalence testing. That is, the means of the two scales is statistical equivalent if the 90% confidence interval of the mean different is completely covered by the equivalence acceptance criteria (EAC). EAC should be determined based on scientific knowledge or historical data.
- In general, the sample sizes in the qualification study are small (as presented in Table 1). Thus, it is likely to have a non-significant result (p-value > 0.05) using the t-test, which is a significant-test-based approach with a known limitation of showing non-significant results with small sizes or with large variability. Thus, the test results from the –test can be misleading with insufficient number of data points. We recommend GSK increasing the sample sizes.

We have the following comments regarding GSK's Process Characterization Studies to assess the impact of the process parameters on the process performance and product attributes.

- (b) (4)
- It is not clear how the Bayesian predictive model was developed in the PAR studies for (b) (4) process parameters. GSK should provide the detailed Bayesian model they applied and the prior distribution. A model validation study should also be done to evaluate if the proposed model is acceptable.

II. INTRODUCTION

On February 05, 2015, Office of Biotechnology Products (OBP) requested the CMC statistics team in Office of Biostatistics to evaluate "*the statistical analysis of the small-scale models and the probability predictions from the multivariate studies. The information would be in Module 3.2.S.2.6 .4.3 (mostly pages 207 -262 and then there is a statistical approach described on page 327 for small scale resin reuse studies*".

Our comments regarding the sponsor's statistical approaches of the qualification and characterization studies for small-scale are provided in Section I. The sponsor's approaches are summarized in Section III.

III. GSK'S STATISTICAL APPROACHES

III.1 QUALIFICATION OF THE SMALL-SCALE MODEL

In S.2.6. entitled "*Manufacturing Process Development_Mepolizumab*", the sponsor conducted the qualification of the small-scale model to evaluate the equivalency in performance at the small scale and at the full-scale. The purpose of the qualification study is to qualify the scale-down model's ability to predict the performance at the full-scale.

In particular, the sponsor conducted the following two statistical tests for each stage:

(b) (4)

(b) (4)

The sponsor's testing results for the F-test and t-test for each stage are provided in the tables below. Please refer to Section I for our comments on the study designs.

(b) (4)

9 Page(s) have been Withheld in Full as b4 (CCI/TS) immediately following this page

---

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

/s/

XIAOYU DONG
04/10/2015

YI TSONG
04/10/2015

MEIYU SHEN
04/10/2015

STATISTICAL REVIEW AND EVALUATION

Biometrics Division: VI

BLA NO.:	125526
DATE RECEIVED BY OB:	January 16, 2015
DRUG NAME:	Mepolizumab
DOSAGE FORM:	lyophilized powder for reconstitution
INDICATION:	Add-on maintenance treatment for patients 12 years of age and older with severe eosinophilic asthma as identified by certain blood eosinophil levels.
ROUTE OF ADMINISTRATION	subcutaneous
STRENGTHS	100 mg/vial
SPONSOR:	GSK
REVIEW FINIHSED:	April 07, 2015
STATISTICS REVIEWER:	Xiaoyu Dong, Ph.D.
PROJECT MANAGER:	Melinda Bauerlien

Concur: _____

Meiyu Shen, Team Leader, Mathematical Statistician, CDER/OTS/OB/DB VI

Yi Tsong, Ph.D., Division Director, DBVI, CDER/OTS/OB/DB VI

Distribution: BLA 125526

CDER/OTS/OB/DB VI/ Yi Tsong
CDER/OTS/OB/DB VI/ Meiyu Shen
CDER/OTS/OB/ Lillian Patrician
CDER/OBP/ Dr. Marjorie Shapiro
CDER/OBP/Melinda Bauerlien

TABLE OF CONTENTS

I. FDA REVIEWER'S STATISTICAL COMMENTS

Comments on GSK's evaluation of the shelf life for (b) (4)/MDP1:

In your evaluation of the shelf life for both Mepolizumab Bulk Drug Substance (DS) and Drug Product (DP), you applied the linear regression model to the real-time stability data and concluded that the proposed shelf life of (b) (4) is supported. Please clarify if you compute the confidence interval of the mean regression line to determine the shelf life. If not, please apply the Q1E approach to support your proposed shelf life. That is, compute the 95% confidence limits (CL) of the mean regression line and compare the CLs against the specification limits using your primary stability batches. By doing this, the data variability can be appropriately taken into account.

Comments on GSK's (b) (4) approach for the comparability studies:

We also have the following comments regarding your comparability studies to support the (b) (4). In your study, you compared the 60 months stability data of (b) (4)/MDP1 with the 12 months stability data of (b) (4)/MDP2 at various storage conditions using a (b) (4) statistical approach.

(b) (4) Thus, the shelf life (b) (4) obtained from (b) (4)/MDP1 stability data can be bridged to (b) (4)/MDP2. However, the above proposed approach/decision rule has the following statistical deficiencies:

- (b) (4)
-
-

(b) (4) Our statistical comments regarding your (b) (4) are summarized below:

(b) (4)

In summary, even if (b) (4)/MDP2 and (b) (4)/MDP1 have exactly the same stability data from month 0 to month 12, the conclusion we can draw from the statistical analysis would be very limited because there is no data beyond the 12th month for (b) (4)/MDP2. Thus, a comprehensive scientific justification and the knowledge of the manufacturing process are critical to evaluate the shelf life for (b) (4)/MDP2. We also recommend submitting more data from (b) (4)/MDP2.

II. INTRODUCTION

On January 16, 2015, Office of Biotechnology Products (OBP) requested the CMC statistics team in Office of Biostatistics to "review statistical analysis of mepolizumab DS stability in 3.2.S.7.1 Section 5.4 pages 48-58 and DP stability in 3.2.P.8.1 Section 5.4 pages 72-83".

This review concentrates on the evaluation of GSK's stability analysis on the real-time stability data of (b) (4)/MDP1, as well as the comparability studies between (b) (4)/MDP2 and (b) (4)/MDP1. Our comments regarding the sponsor's statistical approaches of the above studies are provided in Section I. The sponsor's approaches are summarized in Section III.

III. GSK'S STATISTICAL APPROACHES

III.1 STABILITY ANALYSIS OF LONG TERM DATA FOR (b) (4)

For bulk drug substances (BDS) using (b) (4) process, the 60 months stability data at the long-term storage conditions (70 °C ± 10°C) at the (b) (4) scales were analyzed by GSK. The sponsor performed the linear regression model to assess the trend. Specially, the sponsor concluded no changes over time if the p-value of the estimated slope is non-significant (< 0.05).

The sponsor's stability plots and associate regression liners are provided below. In these figures, Batches T04L001, T04L006, T04L008, T04M001, T04H001, T04H006 and T04H012 were manufactured at the (b) (4) scale. Batches 240563-0XB0-C03, 240563-0ZA0-C04 and 240563-0ZB1-C06 were manufactured at the (b) (4) scale. Based on the results (no change and all results are with the specification limits), GSK concluded that a shelf life of (b) (4) is supported for (b) (4) drug substances using (b) (4) process.

3 Page(s) have been Withhheld in Full as b4 (CCI/TS) immediately following this page

III.2 COMPRARABOLLITY OF BDS USING (b) (4)

To assess if the stability data are comparable between (b) (4) so that the shelf life obtained using (b) (4) can be applied to (b) (4) GSK proposed a (b) (4) approach. In this study, the 60 months stability data of (b) (4) was compared with the 12 months stability data of (b) (4) at various storage conditions.

The above (b) (4) approach is summarized in the sponsor's chart below.

III.3 STABILITY ANALYSIS OF LONG TERM DATA FOR MDP1

For Mepolizumab for Injection 250 mg/vial using MDP1 process, the 60 months stability data at the long-term storage conditions (b) (4) was analyzed. The sponsor performed the linear regression model to assess the trend. Specially, the sponsor concluded no changes over time if the p-value of the estimated slope is non-significant (< 0.05).

The sponsor's analysis results are provided in the table below. Their stability plots of the long-term data and the associate regression liners are also provided. Based on the results (no change and all results are with the specification limits), GSK concluded that a shelf life of (b) (4) is supported for the drug product using MDP1 process.

Table 8 Summary of Statistical Analysis of MDP1 Data

(b) (4)

4 Page(s) have been Withheld in Full as b4 (CCI/TS) immediately following this page

III.4 COMPRARABOLLITY OF BDS USING MDP1 AND MDP2

The comparability study between MDP1 and MDP2 has a very similar study design as the study conducted between (b) (4). Again, GSK proposed a (b) (4) statistical approach. In this study, the 60 months stability data of MDP1 were compared with the 12 months stability data of MDP2 at various storage conditions.

The above (b) (4) is summarized in the sponsor's chart below.

Figure 73 Statistical Approach for Drug Product Stability Comparability

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

/s/

XIAOYU DONG
04/09/2015

MEIYU SHEN
04/09/2015

YI TSONG
04/10/2015

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

BLA Number: 125526 **Applicant: GlaxoSmithKline** **Stamp Date: 11/4/2014**

Drug Name: Mepolizumab **NDA/BLA Type: Standard**

On **initial** overview of the NDA/BLA application for RTF:

	Content Parameter	Yes	No	NA	Comments
1	Index is sufficient to locate necessary reports, tables, data, etc.	x			
2	ISS, ISE, and complete study reports are available (including original protocols, subsequent amendments, etc.)	x			
3	Safety and efficacy were investigated for gender, racial, and geriatric subgroups investigated (if applicable).	x			
4	Data sets in EDR are accessible and do they conform to applicable guidances (e.g., existence of define.pdf file for data sets).	x			

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE? Yes

If the NDA/BLA is not fileable from the statistical perspective, state the reasons and provide comments to be sent to the Applicant.

Please identify and list any potential review issues to be forwarded to the Applicant for the 74-day letter.

Content Parameter (possible review concerns for 74-day letter)	Yes	No	NA	Comment
Designs utilized are appropriate for the indications requested.	x			
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.	x			
Interim analyses (if present) were pre-specified in the protocol and appropriate adjustments in significance level made. DSMB meeting minutes and data are available.			x	
Appropriate references for novel statistical methodology (if present) are included.			x	
Safety data organized to permit analyses across clinical trials in the NDA/BLA.	x			
Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.		x		IR submitted prior to 74 day letter

---

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

/s/

ROBERT ABUGOV
12/29/2014

GREGORY P LEVIN
12/29/2014