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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	NDA 204275
Supplement #:	S0001
Drug Name:	Breo Ellipta (Fluticasone Furoate/Vilanterol Inhalation Powder)
Indication(s):	Treatment of asthma in adult and adolescent subjects (12 years of age and older)
Applicant:	GlaxoSmithKline
Date(s):	Received June 30, 2014 PDUFA due date: April 30, 2014
Review Priority:	Standard
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	Yu (Jade) Wang, PhD
Concurring Reviewers:	Ruthanna Davi, PhD
Medical Division:	Division of Pulmonary, Allergy, and Rheumatology Products
Clinical Team:	Tracy Kruzick, MD, Medical Reviewer Banu Karimi-Shah, MD, Medical Team Leader
Project Manager:	Angela Ramsey

Keywords:

Link to keywords:

http://intranetapps.fda.gov/scripts/ob_apps/ob/eWork/uploads/eWork/2009/Keywords-in-DFS.htm

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

GlaxoSmithKline has proposed Breo Ellipta, a combination of fluticasone furoate (FF) and vilanterol (VI), a once-daily inhaled corticosteroid (ICS)/long-acting beta-agonist (LABA) combination for the treatment of asthma in adult and adolescent patients (12 years and older). Breo Ellipta has been previously approved for the treatment of chronic obstructive pulmonary disease (COPD) in the United States. Arnuity Ellipta, which contains the FF component of Breo Ellipta has also been previously approved for treatment of asthma.

Effectiveness and safety of two different dosages were examined for this submission: FF/VI 100/25 mcg and FF/VI 200/25 mcg. The review focused on four lung function studies (12/24 weeks in duration) and one long term exacerbation study (24-76 weeks in duration).

Direct evaluation of the contribution of FF to the combination through a comparison of FF/VI to VI alone was not possible since administration of a LABA, such as VI, without co-administration of an inhaled corticosteroid, such as FF, is generally considered unacceptable in treating asthma. The effect of FF as a monocomponent was examined in the previous approving process of Arnuity Ellipta by comparing FF to placebo. It is not known whether the effect of FF over placebo will be consistent regardless of the presence or absence of VI.

The contribution of VI to the overall effectiveness of the combination for all major endpoints was directly examined. In support of VI's contribution, after 12 or 24 weeks of treatment, patients assigned to receive FF/VI 100/25 or FF/VI 200/25 consistently showed statistically greater improvement in weighted mean FEV_1 and trough FEV_1 than patients assigned to receive FF only. Also, in the 24 – 76 week long term exacerbation study, patients assigned to FF/VI 100 consistently showed statistically significant improvement in terms of time to first asthma exacerbation and incidence rate of exacerbation than did patients assigned to receive FF 100.

Subgroup analyses were conducted to investigate the level of consistency or heterogeneity of the treatment effect across subgroups of interest, especially the age subgroup, due to the pre-existing concerns regarding asthma-related serious adverse events associated with LABAs. Results of the lung function studies using the WM FEV1 and trough FEV1 endpoints within the adolescents are numerically but not statistically significantly unfavorable and the treatment-by-age interaction tests are not significant so that there is no evidence of a differing treatment effect between age groups. Results for the exacerbation endpoint send out a stronger signal of a differing treatment effect across age groups with the estimate of the treatment-by-age interaction in the analysis of the rate of exacerbations being borderline significant (p=0.055); however, this may be indicative of a qualitative (i.e., a reversal of the treatment effect from adults to adolescents) or only quantitative (i.e., only the magnitude of the difference between treatment groups varies across age groups but the benefit of the combination is positive in both subgroups) interaction. In addition, we acknowledge that the challenge in interpreting subgroup findings is whether to attribute to differences in treatment effect across subgroups to true heterogeneity or to random chance.

This submission supports effectiveness of FF/VI 100/25 and FF/VI 200/25 for once daily treatment of asthma in adult population (age 18 years and older). For the adolescent patient group, however, the contribution of VI to the effectiveness of the combination is questionable, particularly for the exacerbation endpoint. A large well-powered dedicated adolescent asthma exacerbation study would be necessary to fully understand the contribution of VI to the combination in this patient population.

2 INTRODUCTION

2.1 Overview

2.1.1 Background

Asthma is a common chronic lung disease caused by a combination of airway smooth-muscle constriction and inflammation of the bronchi. The disease is characterized by variable and recurring symptoms, reversible airflow obstruction and bronchospasm. Common symptoms include wheezing, coughing, chest tightness, and shortness of breath. Asthma is clinically classified into four severity levels: intermittent, mild persistent, moderate persistent and severe persistent (National Heart, Lung, and Blood Institute, 2007).

There is more than one way to categorize asthma drugs. Categorized by predominant effect to treat asthma, there are bronchodilators (relaxation of airway smooth muscle) and anti-inflammatory drugs (suppression of airway inflammation). Classified by role in the overall management of asthma, there are quick relief and long-term control drugs. Inhaled corticosteroids (ICS) are considered the most effective anti-inflammatory treatments for all severities of persistent asthma. The ICS dose is selected based on the severity of the patient's asthma. Long-acting beta-agonists (LABAs) are drugs that provide bronchodilation for 12 hours or longer. To achieve long term asthma control, combination therapy of ICS with LABA is preferred to increase dose of ICS. Current approved ICS/LABA combinations need to be administered twice daily (BID) (Fanta, 2009).

Fluticasone Furoate (FF) is a novel glucocorticoid for use as a once daily (QD) inhaled treatment for asthma. It is approved by the FDA for maintenance treatment of asthma as prophylactic therapy in patients aged 12 years and older. Vilanterol (VI) is an inhaled LABA being developed for use in combination with ICS as a QD maintenance treatment of asthma. FF/VI is a combination ICS/LABA proposed to be indicated for the once-daily treatment of asthma in patients 12 years of age and older. With this supplemental NDA submission, the applicant is seeking approval for two strengths of the combination drug, FF/VI 100/25 and 200/25.

2.1.2 History of Drug Development

The combination product FF/VI (100/25 mcg) was approved in 2013 for long-term, once-daily, maintenance treatment of airflow obstruction and for reducing exacerbations in patients with chronic obstructive pulmonary disease (COPD). Monotherapy FF was approved in 2014 for maintenance treatment of asthma as prophylactic therapy in patients 12 years of age and older.

There have been multiple meetings/discussions/correspondences regarding the clinical development program of the combination drug for the indication of asthma between the Agency and the applicant since the combination had been introduced to the Division under IND 077855. FF/VI as a combination therapy has also been discussed in the context of treatment of COPD. FF

and VI have also been discussed in the context of monotherapy products, and combination therapy with other categories of drug such as LAMA for treatment of asthma or COPD. Table 1 is a list of the relevant INDs and NDAs that have influenced the development program of the current application.

Table 1. List of relevant INDs and NDAs

Drug Product (Short)	Drug Product	Category	Indication	IND Number	NDA Number	Approval Date
FF	Arnuity Ellipta Fluticasone Furoate GW685698	ICS	Asthma and COPD	IND 070297	NDA 205625	20AUG2014
VI	Vilanterol GW642444	LABA	Asthma and COPD	IND 074696		
FF/VI	Fluticasone Furoate / Vilanterol	ICS/LABA	Asthma and COPD	IND 077855	NDA 204275	10MAY2013
FF/UMEC	Fluticasone Furoate / Umeclidinium	ICS/LAMA	Asthma	IND 112510		
UMEC/VI	Umeclidinium Bromide/ Vilanterol Trifenatate	LAMA/LABA	COPD	IND 106616	NDA 203975	18DEC2013

There are several topics that have been discussed extensively during the development of the above listed INDs and NDAs.

1. Dose Ranging / Dosing Interval
   While *the rationale for the new product rests in the applicant's desire to market a once-daily combination product*, from as early as the PIND70297 meeting on FEB042005, the Division has been emphasizing the necessity of the phase II programs to address dosing interval as well as dose. The Division recommended contrasting once-daily dosing of FF vs. twice-daily dosing in phase II trials.

2. Adequacy of the safety database (PIND70297 FEB042005)
   The applicant estimated that there will be approximately 2400 FF, 900 VI and 2000 FF/VI subjects at the time of NDA submission. The Division concurred with the adequacy based on recommendations in ICH Guidance E1A.

There were 23 phase 2 and 3 clinical trials conducted under the FF/VI asthma clinical development program. These trials provide efficacy and safety data for FF/VI, FF and VI and support the regulatory filing for FF/VI in patients with asthma. Eight of the 23 trials were phase 2 studies in subject with asthma and evaluated doses and dosing intervals for FF and VI and time of day administration for FF/VI. Fifteen phase 3 trials were conducted in adolescent and adult subjects with asthma, aged 12 years and older. Among them, there are seven completed phase 3 FF/VI efficacy and safety studies.

Three FF/VI trials with lung function tests as primary/co-primary efficacy endpoint (HZA106827, HZA116863, and HZA106829) are considered by the applicant as primary studies. Efficacy and safety results of three trials were included in section 14 of the applicant's draft labeling. Study HZA106837 is a supportive long term safety and efficacy trial that examined the drug effect on the risk of asthma exacerbation. Study HZA113091 is a trial comparing the 100/25 strength of FF/VI with Advair (FP/Salm 250/50 BID), another GSK FF/VI on the market. This review will focus on these five studies (Table 2) for the efficacy of FF/VI.

Table 2. List of all studies included in review

	Phase and Design	Treatment Period	Study Population	Randomized (Completed)	Treatment Arms	FPLV-LPLV
HZA 106827	Phase 3, multicenter, stratified, randomized, double-blind, placebo-controlled (with rescue medication), parallel group, efficacy study	12 Weeks	Asthmatics with FEV_1 40-90%	203(151)/ 205(185)/ 201(179)	Placebo FF 100 FF/VI 100/25	2010/08 – 2011/10
HZA 116863	Phase 3, multicenter, randomized, double-blind, stratified, parallel group study	12 Weeks	Asthmatics with FEV_1 40-80%	347(296)/ 346(314)/ 346(321)	FF100 FF/VI 100/25 FF/VI 200/25	2012/09 – 2013/10
HZA 106829	Phase 3, multicenter, stratified, randomized, double-blind, double-dummy, parallel group study.	24 Weeks	Asthmatics with FEV_1 40-90%	194(146)/ 197(169)/ 195(161)	FF200 FF/VI 200/25 FP 500 BID	2010/06 – 2011/10
HZA 113091	Phase 3, multicenter, randomized, double-blind, double dummy, parallel group efficacy and safety study	24 Weeks	Asthmatics with FEV_1 40-85%	403(358)/ 403(357)	FF/VI 100/25 FP/Salm 250/50 BID	2010/06 – 2011/07
HZA 106837	Long-Term, Phase 3, multicenter, stratified, randomized, double-blind, parallel group study	24-76 Weeks	Asthmatics with FEV_1 40-90%	1009(885)/ 1010(863)	FF/VI 100/25 FF100	2010/02 – 2011/09

Source: Reviewer
Abbreviations: FPFV=First Patient First Visit; LPLV=Last Patient Last Visit

2.2 Data Sources

Data were submitted by the applicant to the CDER electronic data room in SAS transport format. Protocols, Reporting and Analysis Plans, Study Reports, correspondence, and data listings were accessed under then network path \\cdsesub1\evsprod\NDA204275\204275.enx

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

The submitted datasets were of acceptable quality and were adequately documented. We were able to reproduce the results of all key primary and secondary analyses. The applicant reported that GCP issues were identified for two investigators ((b) (6) and (b) (6)). The data of affected studies 27 and 37 were re-analyzed excluding data from the two investigators. Results regarding the treatment effect were similar.

3.2 Evaluation of Efficacy

Section 3.2 describes the pre-specified study designs and analyses of the confirmatory studies in this clinical program. The reader is encouraged to also refer to section 4 for comments regarding possible differences in efficacy across subgroups and especially section 5 for comments regarding the sensitivity of these efficacy results to missing data.

FF/VI is a fixed combination of a previously approved asthma monotherapy FF (Arnuity Ellipta, FF 100 and 200 QD) and a new LABA ingredient VI which is not intended for approval as a monotherapy. Given the fact that FF is approved by the Agency for asthma, and the fact that monotherapy LABA without concurrent ICS is not considered an appropriate asthma treatment for safety reasons, this GSK asthma combination drug application relied mainly on the following primary efficacy studies (Table 3) by focusing on certain comparisons under each. To be specific, in the primary efficacy studies, no VI monotherapy arm was included due to safety concern. That is, the efficacy of the combination is only demonstrated via comparisons with FF monotherapy, active control and placebo.

Table 3. Primary Efficacy Studies Treatment Arms

Category	Study	Treatment Arms						
		Placebo	FF100QD	FF200QD	FF/VI 100/25	FF/VI 200/25	FP 500 BID	FP/S 250/50 BID
Lung Function	HZA 106827	✓	✓		✓			
	HZA 113091				✓			✓
	HZA 116863		✓		✓	✓		
	HZA 106829			✓		✓	✓	
Exacerbation	HZA 106837		✓		✓			

Consequently, in this review, limited by the lack of a mono VI arm in phase 3 development program and to enforce the FDA combination rule, we will focus on confirming the contribution of VI by validating the comparison results of FF/VI versus FF in each study. Also, the relative

effect between the two strengths of FF/VI, and the effect of FF/VI compared to an active control (Advair) will be examined.

In the four lung function studies listed in Table 3, the primary or co-primary efficacy endpoints used are: change from baseline in clinic visit trough FEV_1 at the end of treatment period (trough FEV_1) and weighted mean serial FEV_1 over 0–24 hours post-dose (WM FEV_1). The 24-hour serial FEV_1 included a pre-dose assessment within 5 minutes prior to dosing and post-dose assessments after 5, 15, 30 minutes and 1, 2, 3, 4, 5, 12, 16, 20, 23 and 24 hours. The weighted mean endpoint was derived as the average area under the curve using the trapezoidal rule, and dividing by the relevant time interval which in essence provides an estimate of the average height across time.

Measurements of symptoms, rescue drug use, and patient quality of life were also assessed as secondary efficacy endpoints including change from baseline in the percentage of rescue-free 24-hour periods, change from baseline in the percentage of symptom-free 24-hour periods, change from baseline in total Asthma Quality of Life Questionnaire (AQLQ +12) score, etc.

Under each of the comparisons of concern, we want to assess the expected drug effects by evaluating the appropriate endpoints. By consulting the medical review team assigned to this application, we developed our rationale on the priority of the lung function tests. Across the four lung function studies in Table 3, in the applicant's statistical analysis plans, weighted mean (0–24 h) FEV_1 was either a primary or co-primary endpoint; trough FEV_1 was either a co-primary or a secondary endpoint in both the lung function studies and the exacerbation study. While in general the FF and VI components both have effects on post-dose lung function improvement, as LABAs are expected to provide a rapid response after dosing, in this review the WM FEV_1 (0–24 h) is the preferred endpoint to measure the contribution of LABA over ICS; trough FEV_1 is a measure more suitable to demonstrate the background effect of ICS.

FEV_1 has not been formally demonstrated to be a surrogate for asthma exacerbation; however, in the regulatory setting it is often used as a surrogate measure of efficacy, and may reflect benefit on reduction in asthma exacerbation. As an exacerbation study is a long-term large study, asthma development programs do not usually have an exacerbation study to complement demonstrating effects on FEV_1 assessments. In this application, the applicant provided a dedicated long-term asthma exacerbation study which employed time to first asthma exacerbation (primary endpoint) and annualized rate of exacerbation (secondary endpoint) to measure drug effect on exacerbation. Aside from the importance of the efficacy aspect of the exacerbation data, there are also safety related concerns drawn from analyses of exacerbation data. Thus the exacerbation study became a main focus during our review process.

3.2.1 Study HZA 106827 (27)

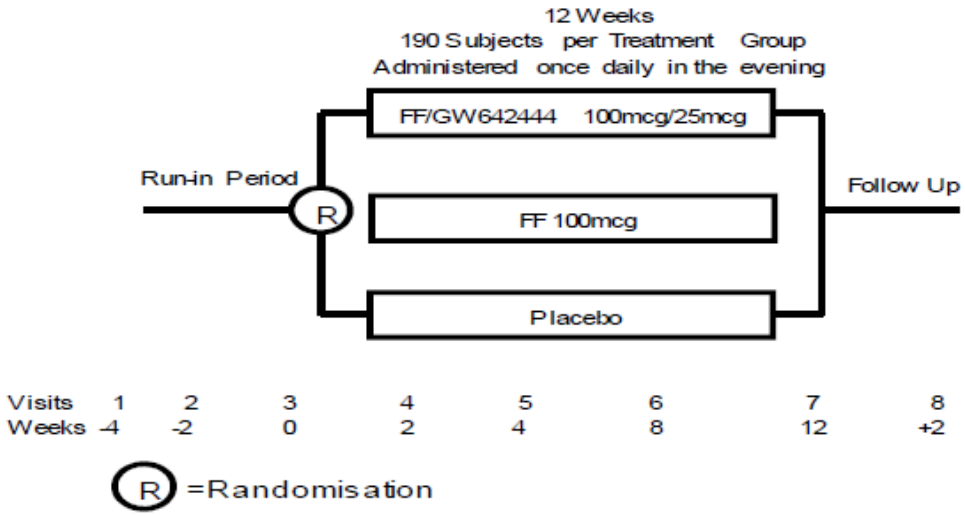
3.2.1.1 Study Design and Endpoints

Study 27 is a multi-center, stratified, randomized, double-blind, placebo-controlled (with rescue medication), parallel-group, efficacy study comparing the efficacy and safety of FF/VI 100 mcg/25 mcg and FF 100 mcg both administered QD in the evening in adolescent and adult subjects, 12 years of age and older, with persistent bronchial asthma over a 12-week treatment period. Among the primary studies, study 27 is the only placebo-controlled study. Chronologically speaking, study 27 was the first trial conducted comparing the efficacy and safety of FF/VI 100 mcg/25 mcg and FF 100 mcg.

Patients at least 12 years of age who had a diagnosis of asthma and had been taking at least 12 weeks of an ICS, with or without LABA and on a stable low- to mid-dose of ICS or a low dose combination of ICS and LABA for at least 4 weeks prior to screening were recruited to the study. Also, patients need to have a best pre-bronchodilator FEV_1 of 40% - 90% of the predicted normal value at screening.

The study includes three periods: Run-in Period (RI), Double-blind Treatment Period (DB), and Follow-up Period (FU). The RI period was provided for completion of baseline safety evaluations including the 12-lead ECG performed at Visit 2 and to obtain baseline measures of asthma status.

Figure 1. Study Schematic (Study 27)



Source: The applicant's study 27 CSR Figure 1

At the randomization visit, patients were stratified according to their concurrent asthma medication (ICS or ICS/LABA) and randomly assigned to receive 12 weeks of treatment with FF/VI 100 mcg/25 mcg, FF100 mcg, or placebo.

The primary time-point is the end of week 12 measurement. In the applicant's rationale, co-primary endpoints were used to assess bronchodilator effect on pulmonary functions; change

from baseline in clinic visit trough FEV_1 was measured in all patients to demonstrate the bronchodilator effect of FF/VI and FF out to 24 hours; weighted mean 24-hour FEV_1 was performed in approximately 60% of the population to demonstrate the profile of bronchodilation over the 24-hour period.

Expected drug effect on asthma symptoms and rescue drug use were assessed by secondary efficacy endpoints.

3.2.1.2 Statistical Methodologies

For one of the co-primary efficacy endpoints, change from baseline in trough FEV_1 at week 12, the primary analysis population was the Intent-To-Treat (ITT) population, defined as all subjects randomized to treatment and who received at least one dose of study medication, with missing data imputed using Last Observation Carried Forward (LOCF) method. The analysis model is an ANCOVA model with the change from baseline in trough FEV_1 as the dependent variable, treatment as the independent variable, while controlling for the effects of the covariates, baseline FEV_1, region, sex, age, and treatment group. The adjusted means for each treatment and the estimated treatment differences for the treatment comparisons were presented together with 95% confidence intervals the differences and p-values.

Use of the LOCF imputation method can be problematic in that it may result in biased estimate of the treatment effect (in comparison to the estimated treatment effect in all randomized subjects regardless of adherence). . When the last observed value is not indicative of the status of the subject (regardless of adherence) at the planned measurement time the treatment effect obtained from a LOCF method should not be interpreted as a result representative of the planned measurement time point. Rather, it is an analysis estimating the treatment effect using the last available observation (LAO), a mean effect across subjects at various measurement time points.

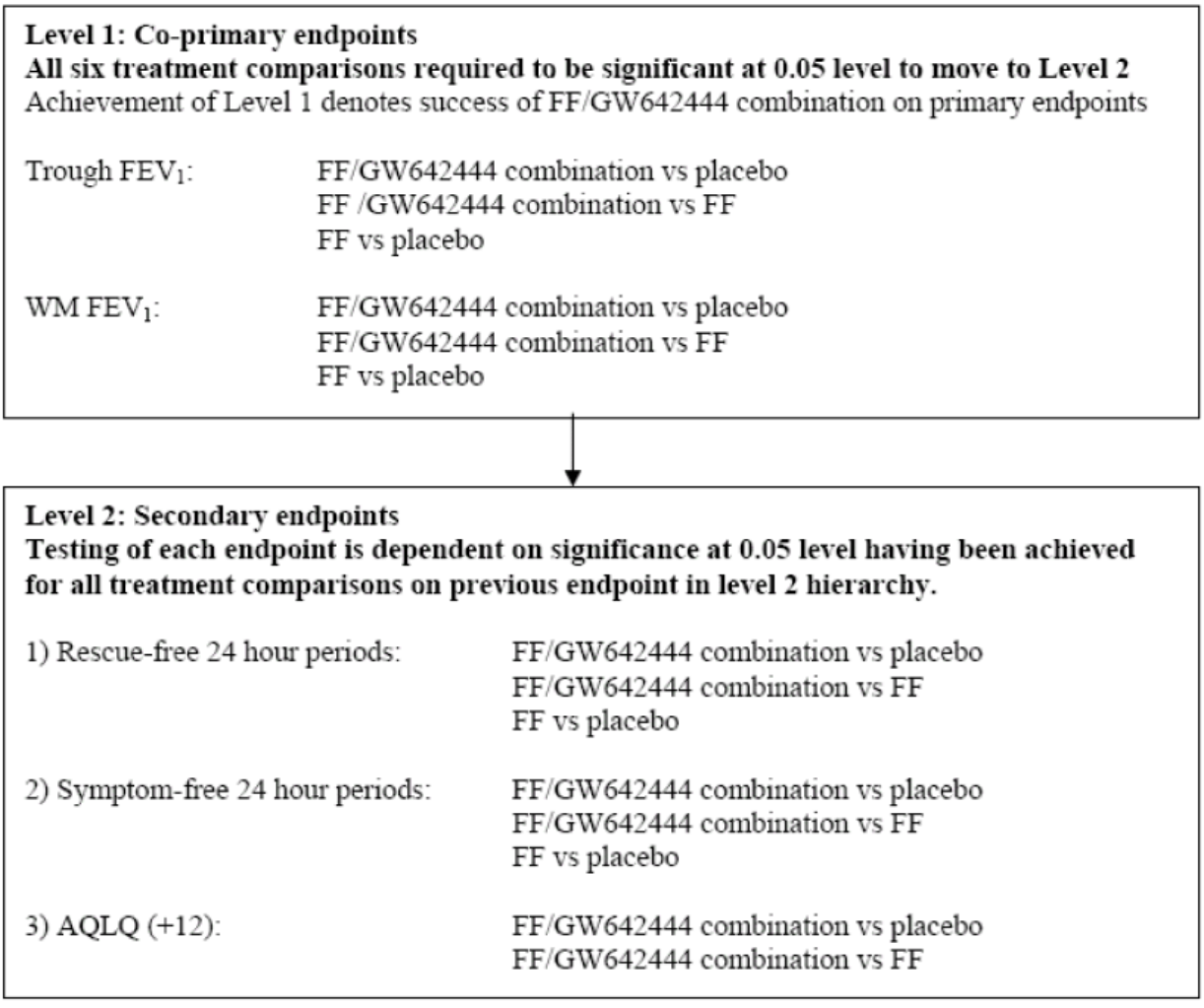
As a supportive analysis, a likelihood-based mixed-effects model for repeated measures (MMRM) was pre-specified and performed for trough FEV_1. The dependent variable was the change from baseline in trough FEV_1. The model included treatment, visit, treatment by visit interaction, region, sex and age as fixed effects; the baseline FEV_1, and visit by baseline interaction as covariates; and visit as a repeated measure. The trough FEV_1 data from four post-baseline visits (week 2, 4, 8 and 12) were included in the model. The model utilized an unstructured covariance matrix.

WM FEV_1 was analyzed using the pre-specified ANCOVA model with effects due to baseline FEV_1, region, sex, age and treatment group. The adjusted means for each treatment and the estimated treatment differences for the treatment comparisons were presented together with 95% confidence intervals for the differences and p-values (Results shown in Section 3.2.1.4.)

Secondary efficacy endpoints including change from baseline in the percentage of rescue-free/symptom-free 24-hour periods and change from baseline in total AQLQ score were also analyzed using the ANCOVA model similar with that for trough FEV_1.

This study involved three treatment arms, two co-primary endpoints and a set of secondary endpoints. The term co-primary is used in this setting to indicate that successful demonstration of a treatment effect (at a significance level of 0.05) is required for both endpoints to conclude that demonstration of efficacy has been achieved. To control family wise type I error across treatment comparisons and primary and key secondary endpoints, the applicant applied a step-down closed test procedure to handle this multiplicity issue. A two level testing hierarchy was used to the order of testing (Figure 2). All pairwise treatment comparisons for the co-primary endpoints are required to be significant at the 0.05 level before making any inference on the secondary endpoints. Within the second level of a pre-defined hierarchy of endpoints, inferential testing of a certain endpoint is allowed only if the previous tests are all significant at the 0.05 level.

Figure 2. Statistical Testing Strategy (Study 27)



Source: The applicant's study 27 RAP Figure 1

3.2.1.3 Patient Disposition, Demographic and Baseline Characteristics

Patient demographic and baseline characteristics are well balanced across the treatment arms (Table 4).

Table 4. Summary of Demographics (ITT, Study 27)

		Placebo (N=203)	FF 100 OD (N=205)	FF/VI 100/25 OD (N=201)	Total (N=609)
Sex	n	203	205	201	609
	F	111 (55%)	126 (61%)	116 (58%)	353 (58%)
	M	92 (45%)	79 (39%)	85 (42%)	256 (42%)
Age (yrs)	n	203	205	201	609
	< 18 years	33 (16%)	28 (14%)	21 (10%)	82 (13%)
	>=18 years to < 65 years	160 (79%)	161 (79%)	169 (84%)	490 (80%)
	>=65 years	10 (5%)	16 (8%)	11 (5%)	37 (6%)
U.S. Site	n	203	205	201	609
	No	138 (68%)	134 (65%)	141 (70%)	413 (68%)
	Yes	65 (32%)	71 (35%)	60 (30%)	196 (32%)
Race	n	203	205	201	609
	African American/African Heritage	14 (7%)	16 (8%)	13 (6%)	43 (7%)
	Asian*	19 (9%)	16 (8%)	16 (8%)	51 (8%)
	Other**	1 (<1%)	2 (1%)	0	3 (<1%)
	White***	169 (83%)	171 (83%)	172 (86%)	512 (84%)

Source: Reviewer.
* Asian includes Central/South Asian Heritage, Asian - East Asian Heritage, Asian - Japanese Heritage, and South East Asian Heritage.
**Other includes American Indian or Alaskan Native, Native Hawaiian or Other Pacific Islander, and Mixed Race.
***White includes Arabic/North African Heritage and White/Caucasian/European Heritage

Early study treatment discontinuation (and therefore early study withdrawal as the protocol did not distinguish between the two) was more common in the placebo group at approximately 26% as opposed to the FF 100 mcg and FF/VI 100 mcg/25 mcg groups at approximately 10% and 11%, respectively. (Table 5). The most common reasons for early withdrawal and the apparent cause for the disproportionate results between the placebo and FF 100 mcg or FF/VI100 mcg/25 mcg groups was "lack of efficacy". The reader is referred to section 5.1.1 of this document for further discussion of the impact of missing data on the efficacy analyses.

Table 5. Summary of Patient Disposition (Study 27)

	Placebo	FF 100 QD	FF/VI 100/25 QD	Total
Screened*				1110
Screen failure				379
Run-in failure				120
Randomized	203	205	202	610
Intent-to-Treat (ITT)	203	205	201	609
Per Protocol (PP)	181 (89%)	184 (90%)	181 (90%)	546 (90%)
Completed	151 (74%)	185 (90%)	179 (89%)	
Primary withdrawn	52 (26%)	20 (10%)	22 (11%)	94 (15%)
Adverse event	1 (<1%)	0	2 (<1%)	3 (<1%)
Lack of efficacy	32 (16%)	6 (3%)	7 (3%)	45 (7%)
Protocol deviation	7 (3%)	0	2 (<1%)	9 (1%)
Lost to follow-up	0	1 (<1%)	2 (<1%)	3 (<1%)
Investigator discretion	6 (3%)	7 (3%)	6 (3%)	19 (3%)
Withdraw consent	6 (3%)	6 (3%)	3 (1%)	15 (2%)

Source: The applicant's study 27 CSR Tables 5.1 – 5.4.
*: All subjects screened and for whom a record exists on the study database.
Note: All the proportions are calculated with respect to the corresponding ITT population number.

3.2.1.4 Results and Conclusions

Results from the primary analyses of study 27 are shown in Table 6. While there are three treatment arms in study 27, as FF 100 OD was approved earlier for asthma, we only focus on the comparison between FF/VI 100/25 OD and FF 100 OD for this application. Neither of the co-primary efficacy endpoints reached statistical significance while both showed numerical trend favoring the combination arm. The estimated trough FEV_1 effect is 0.04 L (95% CI: -0.05, 0.12; p=0.405). The estimated WM FEV_1 effect size is 0.17 L (95% CI: -0.01, 0.24; p=0.06).

Trough FEV_1 data were collected on weeks 0, 2, 4, 8 and 12. Trough FEV_1 treatment comparison results at week 12 from a supportive repeated measure analysis (MMRM) adjusted for baseline FEV_1, region, sex, age, treatment, visit, visit by baseline FEV_1 interaction and visit by treatment interaction also give similar results. The estimated trough FEV_1 effect size is 0.04 (95% CI: -0.04, 0.13; p=0.29).

Table 6. Analyses of the Co-Primary Efficacy Endpoints: Trough FEV_1 and WM FEV_1 at Week 12 (Study 27, LOCF, ITT)

	Trough FEV_1			Weighted Mean FEV_1		
	Placebo N=193	FF 100 N=203	FF/VI 100/25 N=200	Placebo N=95	FF 100 N=106	FF/VI 100/25 N=108
LS Mean Change from Baseline (L)	0.20	0.33	0.37	0.33	0.46	0.26
Difference vs. Placebo, CI, p-value		0.14 (0.05, 0.22) 0.002	0.17 (0.09, 0.26) <0.001		0.19 (0.06, 0.31) 0.003	0.30 (0.18, 0.43) <0.001
Difference vs. FF 100, CI, p-value			0.04 (-0.05, 0.12) 0.405			0.17 (-0.01, 0.24) 0.060

Source: Reviewer
Abbreviations: CI = Confidence interval, LS = Least square.
Note: Analysis performed using ANCOVA with covariates of baseline FEV_1, region, sex, age and treatment.

Chronologically, study 27 was the first phase 3 study conducted to compare FF/VI with FF. The sponsor indicated that because this trial failed to show statistical significance of the combination therapy over ICS monotherapy, the size and design of subsequent trials were modified and more appropriately powered for demonstrating a difference between treatment groups.

3.2.2 Study HZA 116863 (63)

3.2.2.1 Study Design and Endpoints

Study 27 failed to demonstrate the contribution of VI to the co-primary endpoints of trough FEV1 and WM FEV_1, as FF/VI 100/25 showed a numerical but not statistically significant benefit over FF 100 at week12. Study 63 was conducted subsequently with similar design targeting a group of patients with more severe baseline asthma symptoms by including an additional higher dose arm FF/VI 200/25. The size of study 63 was approximately 1.5 times that of study 27.

Study 63 was a multi-center, stratified, randomized, double-blind, active-controlled, parallel group, efficacy study. The treatment period is also 12 weeks. The primary objective of the study was to compare the efficacy and safety of FF/VI 100/25 QD with FF 100 QD in adult and adolescent subjects >=12 years of age with moderate to severe, persistent bronchial asthma over 12 weeks. The secondary objective of the study is to assess the relative numerical efficacy of the two strengths of combination FF/VI 200/25 and FF/VI 100/25. Achieving statistically significant differences between dose levels was not expected or planned.

The primary efficacy endpoint is WM FEV_1 over 0–24 hours post-dose at the end of the 12-week treatment period. There are two powered secondary efficacy endpoints: change from baseline

trough FEV_1 at the end of the 12 week treatment period and change from baseline in the percentage of rescue-free 24 periods during the treatment period.

3.2.2.2 Statistical Methodologies

The statistical analysis methods for WM FEV_1 and trough FEV_1 were the same as in study 27.

In order to account for multiplicity across the key efficacy endpoints, a step-down closed testing procedure was applied (Figure 3). Again, the primary treatment comparison was required to be significant at the 0.05 level for the primary endpoint in order to perform inferential testing on the secondary endpoints. Inference for a test in the pre-defined hierarchy of secondary endpoints was dependent upon statistical significance having been achieved for the previous comparison in the hierarchy of secondary endpoints.

Figure 3. Statistical Testing Strategy (Study 63)

Testing of each endpoint is dependent on significance at the 0.05 level having been achieved on the previous endpoint in the hierarchy.

Primary Efficacy Endpoint	
1) WM FEV_1:	FF/VI 100mcg/25mcg vs. FF 100mcg
Secondary Efficacy Endpoints	
2) Trough FEV_1:	FF/VI 100mcg/25mcg vs. FF 100mcg
3) Rescue-free 24 hour periods:	FF/VI 100mcg/25mcg vs. FF 100mcg
4) Symptom-free 24 hour periods:	FF/VI 100mcg/25mcg vs. FF 100mcg
5) AM PEF:	FF/VI 100mcg/25mcg vs. FF 100mcg
6) PM PEF:	FF/VI 100mcg/25mcg vs. FF 100mcg

Source: The applicant's study 63 RAP Figure 1

3.2.2.3 Patient Disposition, Demographic and Baseline Characteristics

Patient demographic and baseline characteristics were well balanced across the treatment arms (Table 7).

Table 7. Summary of Demographics (ITT, Study 63)

		FF 100 OD (N=347)	FF/VI 100/25 OD (N=346)	FF/VI 200/25 OD (N=346)	Total (N=1039)
Sex	n	347	346	346	1039
	F	199 (57%)	205 (59%)	224 (65%)	628 (60%)
	M	148 (43%)	141 (41%)	122 (35%)	411 (40%)
Age (yrs)	n	347	346	346	1039
	< 18 years	26 (7%)	23 (7%)	16 (5%)	65 (6%)
	>=18 years to < 65 years	288 (83%)	284 (82%)	296 (86%)	868 (84%)
	>=65 years	33 (10%)	39 (11%)	34 (10%)	106 (10%)
U.S. Site	n	347	346	346	1039
	No	262 (76%)	270 (78%)	258 (75%)	790 (76%)
	Yes	85 (24%)	76 (22%)	88 (25%)	249 (24%)
Race	n	347	346	345	1038
	African American/African Heritage	26 (7%)	20 (6%)	28 (8%)	74 (7%)
	Asian*	4 (1%)	2 (1%)	2 (1%)	8 (1%)
	Other**	12 (3%)	17 (5%)	15 (4%)	44 (4%)
	White***	305 (88%)	307 (89%)	300 (87%)	912 (88%)

Source: Reviewer.
* Asian includes Central/South Asian Heritage, Asian - East Asian Heritage, Asian - Japanese Heritage, and South East Asian Heritage.
**Other includes American Indian or Alaskan Native, Native Hawaiian or Other Pacific Islander, and Mixed Race.
***White includes Arabic/North African Heritage and White/Caucasian/European Heritage.

Early study treatment discontinuation and therefore early study withdrawal as the protocol did not distinguish between the two was slightly more common in the FF 100 QD group at approximately 15% as opposed to 9% and 7% in the FF 100/25 QD and FF/IV 200/25 QD groups, respectively. (Table 8). The most common reasons for early withdrawal and the apparent cause for the disproportionate results between the single ingredient and combination groups was "lack of efficacy." The reader is referred to section 5.1.1 of this document for further discussion of the impact of missing data on the efficacy analyses.

Table 8. Summary of Patient Disposition (Study 63)

	FF 100 QD	FF/VI 100/25 QD	FF/VI 200/25 QD	Total
Screened*				2019
Screen failure				523
Run-in failure				456
Randomized	347	346	346	1039
Intent-to-Treat (ITT)	347	346	346	1039
Per Protocol (PP)				
Completed	296 (85%)	314 (91%)	321 (93%)	931 (90%)
Primary withdrawn	51 (15%)	32 (9%)	25 (7%)	108 (10%)
Adverse event	4 (1%)	3 (<1%)	3 (<1%)	10 (<1%)
Lack of efficacy	33 (10%)	13 (4%)	11 (3%)	57 (5%)
Protocol deviation	2 (<1%)	3 (<1%)	0	5 (<1%)
Subject reached protocol-defined stopping criteria	1 (<1%)	0	1 (<1%)	2 (<1%)
Lost to follow-up	0	1 (<1%)	1 (<1%)	2 (<1%)
Investigator discretion	3 (<1%)	4 (1%)	4 (1%)	11 (1%)
Withdraw consent	8 (2%)	8 (2%)	5 (1%)	21 (2%)

Source: The applicant's study 63 CSR Tables 5.1 – 5.4.
*: All subjects screened and for whom a record exists on the study database.
Note: All the proportions are calculated with respect to the corresponding ITT population number.

3.2.2.4 Results and Conclusions

Results from the primary analysis of study 63 are shown in Table 9. The primary comparison of WM FEV_1 at the end of week 12 between FF/VI 100/25 OD and FF 100 OD is statistically significant with an estimated effect size of 0.11 L (95% CI: 0.05, 0.17; p<0.001). The relative efficacy of higher dose FF/VI 200/25 versus FF/VI 100/25 on WM FEV_1 is not significant while numerically better with an estimated effect size 0.02 L (95% CI: -0.04, 0.09; p=0.4).

Table 9. Analyses of the Primary Efficacy Endpoint: WM FEV_1 at Week 12 (Study 63, LOCF, ITT)

	Weighted Mean FEV1 at Week 12		
	FF 100 N=347	FF/VI 100/25 N=346	FF/VI 200/25 N=346
LS Mean Change from Baseline	0.37	0.47	0.50
Difference vs. FF100, CI, p-value		0.11 (0.05, 0.17) <0.001	
Difference vs. FF/VI 100/25, CI, p-value			0.02 (-0.04, 0.09) 0.443

Source: Reviewer
Abbreviations: CI = Confidence interval, LS = Least square.
Note: Analysis performed using ANCOVA with covariates of baseline FEV_1, region, sex, age and treatment.

At the design stage, study 63 was powered to detect a treatment difference of 0.135 L in WM FEV_1 between FF/VI 100/25 and FF 100, assuming a common standard deviation of 0.415 L. While the effect size observed for the difference between FF/VI 100/25 and FF 100 in study 63 is consistent with that of study 27,, study 63 improves the reliability of the previous finding of numerical trend in study 27 in for the primary comparison by providing statistical significance for this comparison. This study established the efficacy of the FF/VI combination therapy over monotherapy FF at the lower strength of FF. For the secondary objective to assess the relative efficacy of FF/VI 200/25 versus FF/VI 100/25, there is no evidence that the higher dose is better in this subject population.

3.2.3 Study HZA 113729 (29)

3.2.3.1 Study Design and Endpoints

Study 29 was a multi-center, stratified, randomized, double-blind, double-dummy, active-controlled (with rescue medication), parallel group study to compare the efficacy and safety of FF/VI 200/25 OD, FF/VI 200 OD and Fluticasone Propionate (FP) 500 BID. The trial blinding is double-dummy due to the presence of an FP arm the administration of which is different from the study drug. The treatment duration is 24 weeks which is twice that of studies 27 and 63 involving the lower strength combination.

The primary objective of study 29 was to compare the efficacy and safety of FF/VI 200/25 to FF 200 in adolescent and adult subjects 12 years of age and older with persistent bronchial asthma over a 24 week treatment period. The secondary objective of this study was to compare the efficacy of FF 200 with FP 500 BID. As FF200 monotherapy is approved in another submission, our focus here is to validate the results of the primary comparison.

With approximately 200 patients per treatment group, the study was designed to have 95% power to detect a treatment difference of 0.15 L in change from baseline in trough FEV_1 between the FF/VI combination and FF alone. Serial FEV_1 was measured on a pre-specified subset of subjects. In the subset of 60% of all randomized subjects who completed the treatment period, this study has 96% power to detect a treatment difference of 0.175 L in WM FEV_1 between FF/VI and FF alone.

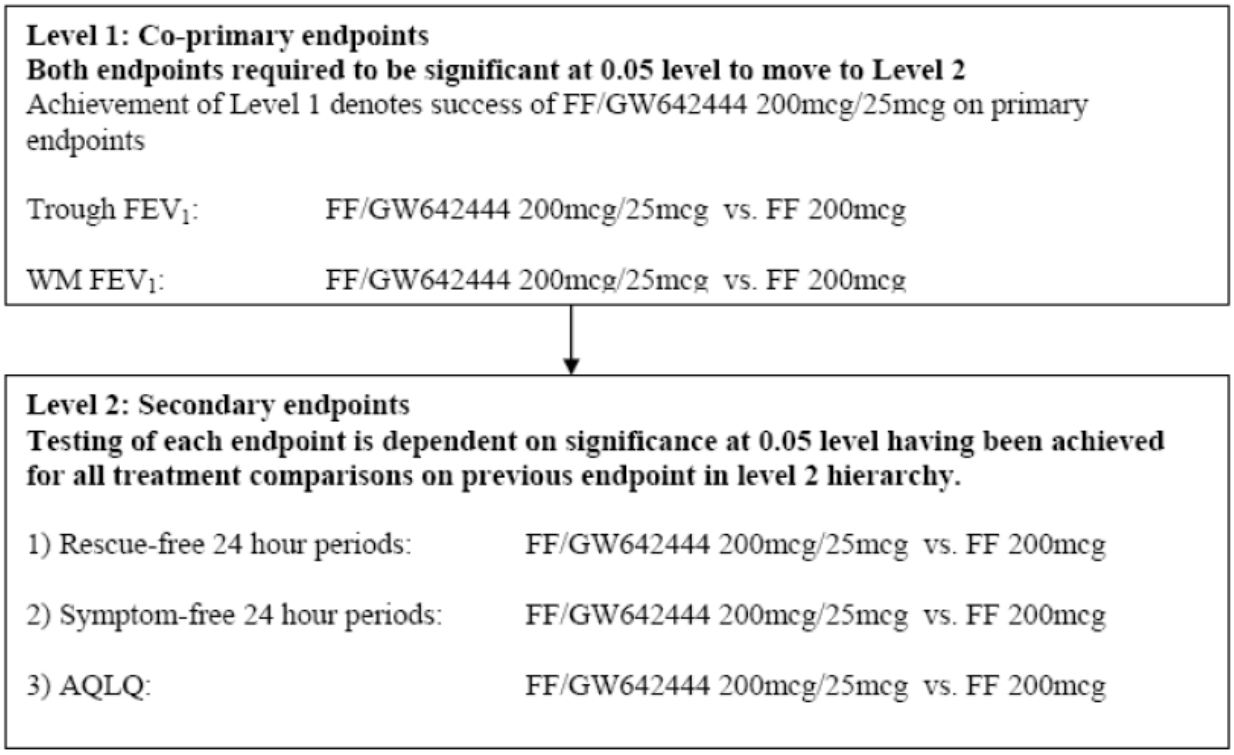
The co-primary efficacy endpoints are trough FEV_1 and WM FEV_1 at the end of 24 weeks. The term co-primary is used in this setting to indicate that successful demonstration of a treatment effect (at a significance level of 0.05) is required for both endpoints to conclude that demonstration of efficacy has been achieved. The powered secondary efficacy endpoint is change from baseline in the percentage of rescue-free 24-hour periods during the 24-week treatment period.

3.2.3.2 Statistical Methodologies

The statistical analysis methods for the co-primary efficacy endpoints were the same as in study 27.

Again, in order to account for multiplicity across the key endpoints, a step-down closed testing procedure was applied with the testing strategy shown in Figure 4.

Figure 4. Statistical Testing Strategy (Study 29)



Source: The applicant's study 29 RAP Figure 1

3.2.3.3 Patient Disposition, Demographic and Baseline Characteristics

Patient demographic and baseline characteristics were well balanced across the treatment arms (Table 10). Within the study, approximately 87% of patients were from 18 to 65 years of age. Approximately 4% were between 12 to 17 years of age and 9% were greater or equal to 65 years of age. Approximately 59% were female and approximately 84% were white.

Table 10. Summary of Demographics (ITT, Study 29)

		FF 200 OD (N=194)	FF/VI 200/25 OD (N=197)	FP 500 BD (N=195)	Total (N=586)
Sex	n	194	197	195	586
	F	113 (58%)	116 (59%)	116 (59%)	345 (59%)
	M	81 (42%)	81 (41%)	79 (41%)	241 (41%)
Age (yrs)	n	194	197	195	586
	< 18 years	7 (4%)	8 (4%)	8 (4%)	23 (4%)
	>=18 years to < 65 years	173 (89%)	167 (85%)	171 (88%)	511 (87%)
	>=65 years	14 (7%)	22 (11%)	16 (8%)	52 (9%)
U.S. Site	n	194	197	195	586
	No	146 (75%)	149 (76%)	148 (76%)	443 (76%)
	Yes	48 (25%)	48 (24%)	47 (24%)	143 (24%)
Race	n	194	197	195	586
	African American/African Heritage	16 (8%)	16 (8%)	19 (10%)	51 (9%)
	Asian*	12 (6%)	15 (8%)	13 (7%)	40 (7%)
	Other**	1 (1%)	1 (1%)	1 (1%)	3 (1%)
	White***	165 (85%)	165 (84%)	162 (83%)	492 (84%)

Source: Reviewer.
* Asian includes Central/South Asian Heritage, Asian - East Asian Heritage, Asian - Japanese Heritage, and South East Asian Heritage.
**Other includes American Indian or Alaskan Native, Native Hawaiian or Other Pacific Islander, and Mixed Race.
***White includes Arabic/North African Heritage and White/Caucasian/European Heritage.

Early study treatment discontinuation and therefore early study withdrawal as the protocol did not distinguish between the two was more common in the FF 200 QD group at approximately 25% as opposed to 14% and 17% in the FF 200/25 QD and PF 500 BD groups, respectively. (Table 11). The most common reasons for early withdrawal were "adverse event", "lack of efficacy" and "withdrawal of consent". The reader is referred to section 5.1.1 of this document for further discussion of the impact of missing data on the efficacy analyses.

Table 11. Summary of Patient Disposition (Study 29)

	FF 200 QD	FF/VI 200/25 QD	FP 500 BD	Total
Screened*				1206
Screen failure				478
Run-in failure				141
Randomized	194	197	196	587
Intent-to-Treat (ITT)	194	197	195	586
Per Protocol (PP)	175	172	168	515
Completed	146 (75%)	169 (86%)	161 (83%)	476 (81%)
Primary withdrawn	48 (25%)	28 (14%)	34 (17%)	110 (19%)
Adverse event	3 (2%)	7 (4%)	2 (1%)	45 (8%)
Lack of efficacy	21 (11%)	6 (3%)	18 (9%)	57 (5%)
Protocol deviation	5 (3%)	3 (2%)	5 (3%)	13 (2%)
Lost to follow-up	0	1 (<1%)	1 (<1%)	2 (<1%)
Investigator discretion	2 (1%)	0 (<1%)	1 (<1%)	3 (<1%)
Withdraw consent	13 (7%)	4 (2%)	7 (4%)	24 (4%)

Source: The applicant's study 29 CSR Tables 5.1 – 5.4.
*: All subjects screened and for whom a record exists on the study database.
Note: All the proportions are calculated with respect to the corresponding ITT population number.

3.2.3.4 Results and Conclusions

Results from the primary analyses of study 29 are shown in Table 12. While there are three treatment arms in study 29, as FF 200 OD was approved earlier for asthma, we only focus on the comparison between FF/VI 200/25 OD and FF 200 OD for this application. Both of the co-primary efficacy endpoints reached statistical significance. The estimated trough FEV_1 effect size is 0.19 L (95% CI: 0.11, 0.28; $p<0.001$). The estimated WM FEV_1 effect size is 0.14 L (95% CI: 0.001, 0.27; $p=0.048$).

Trough FEV_1 data were collected on weeks 0, 2, 4, 8, 12, 16, 20 and 24. Trough FEV_1 treatment comparison results at week 24 from a supportive repeated measure analysis (MMRM) adjusted for baseline FEV_1, region, sex, age, treatment, visit, visit by baseline FEV_1 interaction and visit by treatment interaction also give similar results. The estimated trough FEV_1 effect size is 0.16 (95% CI: 0.08, 0.25; $p<0.001$).

Table 12. Analyses of the Co-Primary Efficacy Endpoints: Trough FEV_1 and WM FEV_1 at Week 24 (Study 29, LOCF, ITT)

	Trough FEV_1			Weighted Mean FEV_1		
	FF 200 N=186	FF/VI 200/25 N=187	FP 500 BID N=190	FF 200 N=83	FF/VI 200/25 N=89	FP 500 BID N=86
LS Mean Change from Baseline (L)	0.20	0.39	0.18	0.33	0.46	0.26
Difference vs. FF200, CI, p-value		0.19 (0.11, 0.28) <0.001			0.14 (0.001, 0.27) 0.048	
Difference vs. FP 500 BID, CI, p-value	0.02 (-0.07, 0.10) 0.676	0.21 (0.13, 0.29) <0.001		0.07 (-0.07, 0.21) 0.316	0.21 (0.07, 0.34) 0.003	

Source: Reviewer
Abbreviations: CI = Confidence interval, LS = Least square.
Note: Analysis performed using ANCOVA with covariates of baseline FEV_1, region, sex, age and treatment.

With the both the co-primary efficacy endpoints showing significant improvement of FF/VI 200/25 over FF 200, this study established the efficacy of the FF/VI combination therapy over monotherapy FF at the higher strength of FF.

3.2.4 Study HZA 113091 (91)

3.2.4.1 Study Design and Endpoints

Study 91 was a multi-center, randomized, double-blind, double-dummy, parallel group study. The primary objective of the study is to compare the efficacy of FF/VI 100/25 QD with efficacy of Advair (FP/Salm 250/50 BID) in subjects 12 year of age and older with persistent bronchial asthma over a 24 week treatment period.

The primary efficacy endpoint is WM FEV_1 over 0–24 hours post-dose at the end of 24 week treatment period. Among the secondary efficacy endpoints, individual serial FEV_1 assessments at the end of the 24 week treatment period, including 12-hour and 24-hour post-dose trough values will be validated in this review.

3.2.4.2 Statistical Methodologies

The statistical analysis methods for the primary efficacy endpoint WM FEV_1 was the same as in study 27. The individual serial FEV_1 assessments at week 24 were analyzed separately (by their planned time) using an ANCOVA model allowing for the effects due to baseline FEV_1, region,

sex, age and treatment group. The adjusted means for each treatment and the estimated treatment differences were presented together with 95% confidence intervals the mean differences for each planned time. P-values were presented for the 12- and 24- hour post-dose values.

3.2.4.3 Patient Disposition, Demographic and Baseline Characteristics

Patient demographic and baseline characteristics were well balanced across the treatment arms (Table 13). Within the study, approximately 82% of patients were from 18 to 65 years of age. Approximately 9% were between 12 to 17 years of age and 9% were greater or equal to 65 years of age. Approximately 61% were female and approximately 59% were white and 31% were Asian.

Table 13. Summary of Demographics (ITT, Study 91)

		FF/VI 100/25 QD (N=403)	FP/Salm 250/50 BD (N=403)	Total (N=806)
Sex	n	403	403	806
	F	244 (61%)	245 (61%)	489 (61%)
	M	159 (39%)	158 (39%)	317 (39%)
Age (yrs)	n	403	403	806
	< 18 years	31 (8%)	41 (10%)	72 (9%)
	>=18 years to < 65 years	341 (85%)	323 (80%)	664 (82%)
	>=65 years	31 (8%)	39 (10%)	70 (9%)
U.S. Site	n	403	403	806
	No	276 (68%)	288 (71%)	564 (70%)
	Yes	127 (32%)	115 (29%)	242 (30%)
Race	n	403	403	806
	African American/African Heritage	36 (9%)	43 (11%)	79 (10%)
	Asian*	123 (31%)	125 (31%)	248 (31%)
	Other**	2 (<1%)	3 (<1%)	5 (<1%)
	White***	242 (60%)	232 (58%)	474 (59%)

Source: Reviewer.
* Asian includes Central/South Asian Heritage, Asian - East Asian Heritage, Asian - Japanese Heritage, and South East Asian Heritage.
**Other includes American Indian or Alaskan Native, Native Hawaiian or Other Pacific Islander, and Mixed Race.
***White includes Arabic/North African Heritage and White/Caucasian/European Heritage.

Early study treatment discontinuation and therefore early study withdrawal as the protocol did not distinguish between the two was balanced across the FF/VI 100/25 QD and FP/Salm 250/50 BD groups at approximately 11%. (Table 14). The most common reason for early withdrawal

was "lack of efficacy". The reader is referred to section 5.1.1 of this document for further discussion of the impact of missing data on the efficacy analyses.

Table 14. Summary of Patient Disposition (Study 91)

	FF/VI 100/25 QD	FP/Salm 250/50 BD	Total
Screened*			1564
Screen failure			623
Run-in failure			135
Randomized	403	403	806
Intent-to-Treat (ITT)	403	403	806
Per Protocol (PP)	361	380	741
Completed	358 (89%)	357 (89%)	715 (89%)
Primary withdrawn	45 (11%)	46 (11%)	91 (11%)
Adverse event	6 (1%)	8 (2%)	14 (2%)
Lack of efficacy	20 (5%)	11 (3%)	31 (4%)
Protocol deviation	7 (2%)	10 (2%)	17 (2%)
Lost to follow-up	5 (1%)	7 (1%)	12 (1%)
Investigator discretion	0	1 (<1%)	1 (<1%)
Withdraw consent	7 (2%)	9 (2%)	16 (2%)

Source: The applicant's study 91 CSR Tables 5.1 – 5.4.
*: All subjects screened and for whom a record exists on the study database.
Note: All the proportions are calculated with respect to the corresponding ITT population number.

3.2.4.4 Results and Conclusions

Results from the primary analysis of study 91 are shown in Table 15. The comparison of WM FEV_1 at the end of week 24 between FF/VI 100/25 OD and Advair is not statistically significant; however, FF/VI 100/25 QD is numerically worse than Advair with an estimated effect size of -0.04 L (95% CI: -0.09, 0.02; p=0.162). As evidence by the limits of the 95% confidence interval, we conclude that the true change from baseline in WM FEV_1 for FF/VI 100/25 is as much as 0.02 L better or 0.09 L worse than that of Advair.

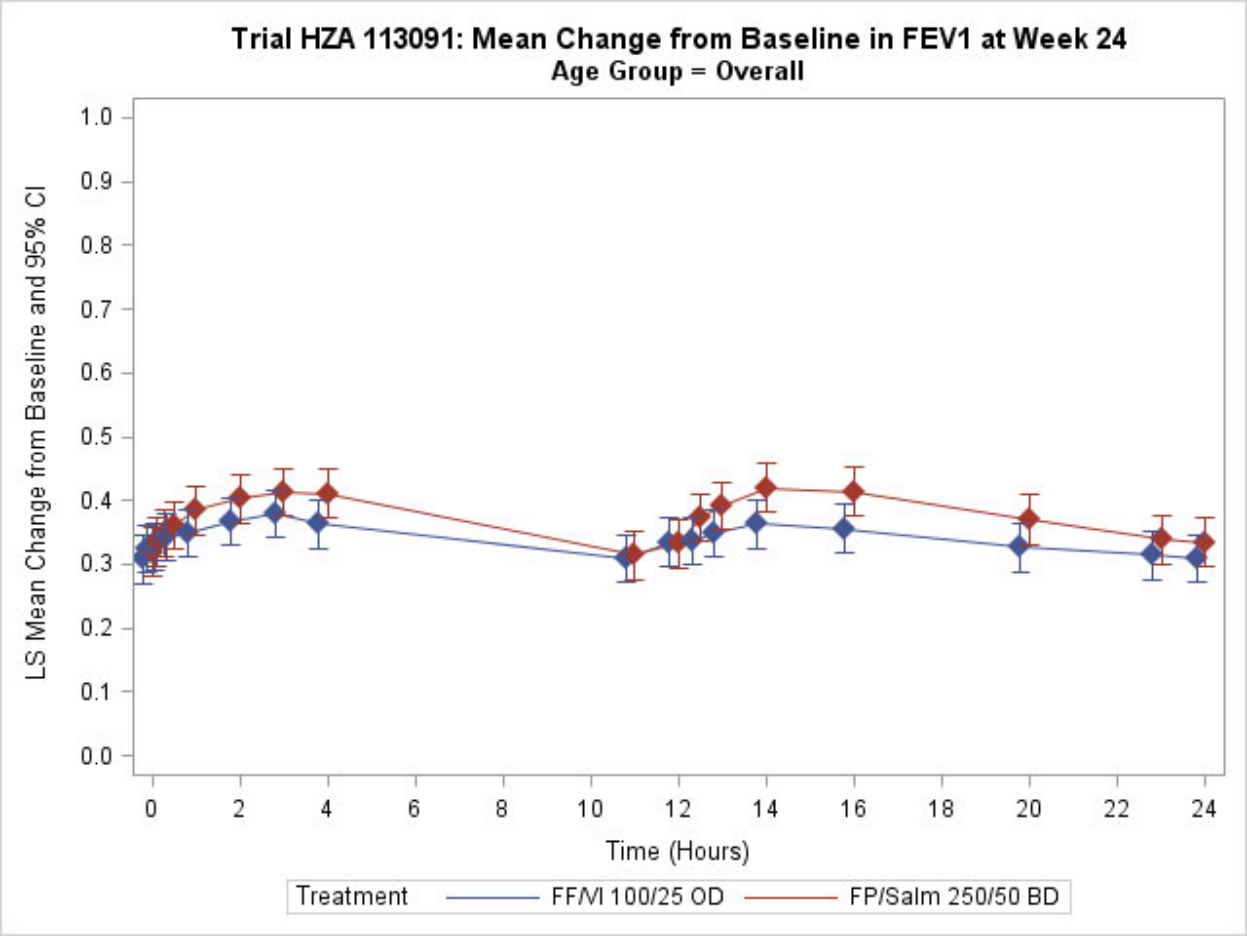
Table 15. Analyses of the Primary Efficacy Endpoint: WM FEV_1 at Week 24 (Study 91, LOCF, ITT)

	Weighted Mean FEV1 at Week 12	
	FF/VI 100/25 N=352	FP/Salm 250/50 BID N=347
LS Mean Change from Baseline (L)	0.34	0.38
Difference vs. FP/Salm 250/50 BID, CI, p-value	-0.04 (-0.09, 0.02) 0.162	

Source: Reviewer
Abbreviations: CI = Confidence interval, LS = Least square.
Note: Analysis performed using ANCOVA with covariates of baseline FEV_1, region, sex, age and treatment.

Results from the secondary analysis of serial FEV1 at week 24 is shown in Figure 5. The estimated mean changes from baseline at each time point illustrates numerically separated profiles with Advair performing numerically better than FF/VI from 0 to 24 hours; however, many of these differences are not associated with nominal p-values less than 0.05. In general, we conclude the two drugs effects in terms of FEV_1 are not meaningfully different. Serial FEV_1 results are consistent with the WM FEV_1 results.

Figure 5. Mean Change from Baseline in Serial FEV_1 at Week 24 (Study 91, ITT)



Source: Reviewer
Note: Analysis performed using ANCOVA with covariates of baseline FEV_1, region, sex, age and treatment. Analysis was performed separately for each planned time point. Serial FEV_1 measure were taken at pre-dose, 5, 15, and 30 minutes and 1, 2, 3, 4, 11, 12, 12.5, 13, 14, 16, 20, 23 and 24 hours.

3.2.5 Study HZA 106837 (37)

Results from studies 27, 91, 63 and 29 have demonstrated that once daily FF/VI combination therapy significantly improve pulmonary function compared with FF monotherapy in terms of the primary and key secondary endpoints of WM FEV1 and trough FEV1.

To demonstrate the anticipated efficacy of the FF/VI combination over single ingredient FF in terms of exacerbation prevention, study 37 was designed utilizing the primary endpoint, time to first severe asthma exacerbation.

The definition of severe asthma exacerbation in this application followed the ATS/ERS Statement. A severe asthma exacerbation is defined as deterioration of asthma requiring the use of systemic corticosteroids (tablets, suspension, or injection) for at least 3 days or an in-patient hospitalization or emergency department visit due to asthma that required systemic corticosteroids

3.2.5.1 Study Design and Endpoints

Study 37 was a multi-center, randomized, double-blind, parallel group, active control study.

Due to the nature of the primary endpoint being time to the event of a subject experiencing the first asthma exacerbation, the study design is event driven. That is, the duration of the treatment period was designed to be variable. It was dependent on the number of events (number of subjects with one or more severe asthma exacerbations) that occur. For any one subject, treatment duration will vary but will not exceed 76 weeks.

The duration of the study is event-driven while event being defined as a subject with one or more severe asthma exacerbation. The study continued until the required total number of events has occurred. There were several assumptions used in deriving the approximate number of events. Previous studies on ICS/LABA combination therapy versus ICS alone have given the following reference estimates: Rates of severe asthma exacerbation: a 17% - 45% improvement in rates of combination therapy over ICS alone; 20% in the FF arm would have severe asthma exacerbation in 1 year; 10% attrition Rate; a certain enrollment rate. The study was planned to end when 330 events have occurred.

The secondary efficacy endpoints are: rate of severe asthma exacerbations per subject per year and change from baseline at Week 36 in evening pre-dose trough FEV1. There are other efficacy endpoints in the protocol, but we will focus on the exacerbation event related endpoints in our review since it is the most clinically important outcome while the others typically are used as surrogates for an effect on exacerbation.

3.2.5.2 Statistical Methodologies

The time to first severe asthma exacerbation was analyzed using a Cox proportional hazards regression model, adjusting for baseline FEV_1, sex, age, region and treatment. Subjects who did not experience an exacerbation during the treatment period were treated as censored observations at the time of treatment discontinuation, whether premature or at the end of the study. Subjects lost to follow-up were censored at their last visit. The estimated hazard ratio was presented together with a 95% confidence interval and p-value. Cox proportional hazards regression

modeling relies on an assumption that censoring is non-informative meaning that the censoring should not be related to the probability of an exacerbation occurring. This assumption may seem plausible for the cases censored by the administrative end of the study; however, this assumption is not so probable in the case of subjects censored by early discontinuation of study treatment.

The rate of severe asthma exacerbations per subject per year over the treatment period was analyzed using a negative binomial regression model with log time on treatment as an offset variable. The model included adjustment for effects due to baseline FEV_1, sex, age, region and treatment. The adjusted mean rates per year, treatment ratio, and confidence interval for the ratio were presented. While the negative binomial regression model does adjust for the time on treatment in calculating the mean exacerbation rate, estimates of the treatment effect may be inappropriately influenced by premature study discontinuation if observed data are not representative of post-discontinuation data.

3.2.5.3 Patient Disposition, Demographic and Baseline Characteristics

Patient demographic and baseline characteristics were well balanced across the treatment arms (Table 16).

Table 16. Summary of Demographics (ITT, Study 37)

		FF 100 QD (N=1010)	FF/VI 100/25 QD (N=1009)	Total (N=2019)
Sex	n	1010	1009	2019
	F	689 (68%)	661 (66%)	1350 (67%)
	M	321 (32%)	348 (34%)	669 (33%)
Age (yrs)	n	1010	1009	2019
	< 18 years	130 (13%)	151 (15%)	281 (14%)
	>=18 years to < 65 years	809 (80%)	788 (78%)	1597 (79%)
	>=65 years	71 (7%)	70 (7%)	141 (7%)
U.S. Site	n	1010	1009	2019
	No	824 (82%)	822 (81%)	1646 (82%)
	Yes	186 (18%)	187 (19%)	373 (18%)
Race	n	1010	1009	2019
	African American/African Heritage	47 (5%)	40 (4%)	87 (4%)
	Asian*	110 (11%)	112 (11%)	222 (11%)
	Other**	110 (11%)	117 (12%)	227 (11%)
	White***	743 (74%)	740 (73%)	1483 (73%)

Source: Reviewer.
* Asian includes Central/South Asian Heritage, Asian - East Asian Heritage, Asian - Japanese Heritage, and South East Asian Heritage.
**Other includes American Indian or Alaskan Native, Native Hawaiian or Other Pacific Islander, and Mixed Race.
***White includes Arabic/North African Heritage and White/Caucasian/European Heritage.

Early study treatment discontinuation and therefore early study withdrawal as the protocol did not distinguish between the two was slightly more common in the FF 100 QD group at approximately 15% as opposed to 12% in the FF 100/25 QD groups. (Table 17). The most common reasons for early withdrawal were "withdraw consent", "adverse event", "lack of efficacy" and "protocol deviation". The reader is referred to section 5.1.1 of this document for further discussion of the impact of missing data on the efficacy analyses.

Table 17. Summary of Patient Disposition (Study 37)

	FF 100 QD	FF/VI 100/25 QD	Total
Screened*			2668
Screen failure			485
Run-in failure			162
Randomized	1011	1009	2020
Intent-to-Treat (ITT)	1010	1009	2019
Per Protocol (PP)	903 (89%)	889 (88%)	1792 (89%)
Completed	863 (85%)	885 (88%)	1748 (87%)
Primary withdrawn	147 (15%)	124 (12%)	271 (13%)
Adverse event	19 (2%)	15 (1%)	34 (2%)
Lack of efficacy	22 (2%)	13 (1%)	35 (2%)
Protocol deviation	26 (2%)	17 (2%)	43 (2%)
Subject reached protocol-defined stopping criteria	0	1 (<1%)	1 (<1%)
Study closed/terminated	7 (<1%)	8 (<1%)	15 (<1%)
Lost to follow-up	11 (1%)	9 (<1%)	20 (<1%)
Investigator discretion	9 (<1%)	6 (<1%)	15 (<1%)
Withdraw consent	53 (5%)	55 (5%)	108 (5%)

Source: The applicant's study 37 CSR Tables 5.1 – 5.4.
*: All subjects screened and for whom a record exists on the study database.
Note: All the proportions are calculated with respect to the corresponding ITT population number.

3.2.5.4 Results and Conclusions

Results from the time to event analysis are shown in Table 18. The hazard ratio of FF/VI 100/25 over FF 100 is statistically significant with an estimated effect size of 80% (95% CI: 64%, 99%; p=0.04). That is, FF/VI 100/25 treated patient will have exacerbation more slowly compared with patients treated with FF 100, and that a patient in the FF/VI 100/25 group who has not yet experienced exacerbation by a certain time has 80% the chance of having exacerbation at the next point in time compared with someone in the FF 100 group.

Table 18. Cox Proportional Hazard Analysis of Time to First Severe Asthma Exacerbation (Study 37, ITT)

	FF 100 N=1010	FF/VI 100/25 N=1009
Number of patients with at least 1 event (n) %	186 (18%)	154 (15%)
FF/VI 100/25 vs. FF 100		
Hazard ratio		0.80
95% CI		(0.64, 0.99)
p-value		0.04

Source: Reviewer
Abbreviations: CI = Confidence interval
Note: Analysis performed using Cox proportional hazard regression model with covariates of baseline FEV_1, region, sex, age and treatment.

Results from analysis of the incidence rate of asthma exacerbation are shown in Table 19. The mean severe asthma exacerbation rate was 0.19 and 0.14 per subject per year for the FF100 and FF/VI 100/25 group, respectively. The incidence rate ratio of FF/VI 100/25 over FF 100 is statistically significant with an estimated effect size of 75.5% (95% CI: 60.3%, 94.5%; p=0.014). That is, FF/VI 100/25 treated patient will experience exacerbation less frequent compared with patients treated with FF 100. Time to event and incidence rate are two measures of incidence of exacerbation. Results from analyses of the two endpoints are consistent with each other.

Table 19. Analysis of Rate of Severe Asthma Exacerbations per Subject per Year (Negative Binomial Model) (Study 37, ITT)

	FF 100	FF/VI 100/25
N	1010	1009
Mean severe asthma exacerbation rate	0.19	0.14
FF/VI 100/25 vs. FF 100 Ratio		0.755
95% CI		(0.603, 0.945)
p-value		0.014
% Reduction in Rate		25%
95% CI		(5%, 40%)

Source: Reviewer
Abbreviations: CI = Confidence interval
Note: Analysis performed using negative binomial regression model with covariates of baseline FEV_1, region, sex, age and treatment, log time on treatment as an offset variable.

3.3 Evaluation of Safety

The reader is referred to the Medical Review by Dr. Tracy Kruzick for an evaluation of the safety of FF/VI in asthmatic patients and a Meta-Analysis of Asthma-Related Serious Adverse Events by Dr. Janelle Charles.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

As is routine, this section provides subgroup analyses by gender, race, age, and geographical region. In light of the pre-existing suggestion of an increased risk of asthma-related serious adverse events associated with long-acting-beta-agonists, especially in children ages 12 to 17 years old, subgroup analyses by age are of particular importance to inform the benefit-risk of the combination product in adolescents.

First, as the four lung functions studies are different with respect to treatment duration, comparator arms, we could not perform an overall pooled subgroup analysis across the four studies. Instead, the only common lung function treatment comparison of FF/VI 100/25 vs. FF 100 is covered at week 12 in studies 27 and 63. WM FEV_1 and trough FEV_1 subgroup analyses were conducted on the pooled data of these two studies.

Second, subgroup analysis of $WMFEV_1$ and trough FEV_1 from the four lung function studies (studies 27, 63, 29 and 91) are provided. In summary, no statistically significant differences in the treatment effect across gender, race, age, and geographical region were identified in that no significant treatment-by-subgroup-interactions were found.

In addition, subgroup analyses for both the time to first severe asthma exacerbation and mean rate of severe asthma exacerbation in the long term exacerbation study (study 37) are included. Numerically negative but not statistically significant trends for the comparison of FF/VI 100/25 versus FF 100 were found for adolescents in each type of analysis of exacerbations.

4.1.1 Statistical Method for Subgroup Analyses

For the pooled subgroup analysis on WM FEV_1 or trough FEV_1, an interaction analysis was performed first with an ANCOVA model by including Study, Treatment, Age group, Study by Treatment interaction, Study by Age group interaction, Treatment by Age group interaction and the three way interaction Treatment by Age group by Study. The significance of the interaction between Treatment and Age group is first tested. Subsequently, by age group analysis were performed with Study and Treatment in the model to estimate the by age group treatment effects between FF/VI 100/25 and FF 100.

For each individual study subgroup analysis, the model was adapted from the pre-specified primary efficacy analysis model. For the pulmonary function efficacy endpoints: weighted mean FEV_1 and trough FEV_1, an interaction analysis was performed with an ANCOVA model by including Treatment, Baseline FEV_1 score, Age, Region, Sex, Subgrouping Variable and Subgrouping Variable-by-Treatment interaction as covariates. When a covariate in the model is the subgrouping variable, it is replaced with the categorical version of itself when needed. A by-subgroup ANCOVA model was conducted to examine the treatment effects under each

subgroup. Under each subgroup, least square mean estimate and confidence interval of the difference between the FF/VI combination therapy and FF single ingredient arm was presented using a Forest plot.

For time to first severe asthma exacerbation, for each subgrouping variable, an interaction analysis was performed with a Cox proportional hazard model by including Treatment, Baseline FEV_1 score, Age, Region, Sex, Subgrouping Variable and Subgrouping Variable * Treatment interaction as covariates. This is an adaptation of the pre-specified primary efficacy analysis model. Under each subgroup, an estimate of the hazard ratio of the FF/VI combination therapy over the FF single ingredient arm and confidence interval of the hazard ratio was presented using a Forest plot.

For the incidence rate (annualized rate) of severe asthma exacerbation, an interaction analysis was performed with a negative binomial regression model. The response variable to be explained by the model is the total number of on treatment severe asthma exacerbations. The model to test the interaction includes Treatment, Baseline FEV_1 score, Age, Region, Sex, Subgrouping Variable and Subgrouping Variable * Treatment interaction as covariates. Subsequently, a by subgroup analysis was conducted for each subgroup. The ratio of the exacerbation rate from the negative binomial analysis for FF/VI 100/25 versus FF 100 under each subgroup was reported, together with the CI with a Forest plot.

Subgroup analysis results are presented in this section by study. The subgroup analyses results are presented in this section with tables for age group analyses. For subgroup analyses by region, sex and race, estimated effect together with confidence intervals and sample sizes are presented together with the forest plots.

Across the endpoints and studies, there is no significant interaction between subgroups and treatment. Lack of a significant treatment-by-subgroup interaction should not be interpreted as evidence that no interaction exists. Definite conclusions cannot be drawn due to limitations such as small sample size in some of the subgroups.

4.1.2 Subgroup Analyses Results

4.1.2.1 Pooled Studies 27 and 63

Pooled subgroup analyses of weighted mean FEV_1 and trough FEV_1 by age group are provided in Table 20 and Table 21. There is no significant interaction between treatment and age groups for either WM FEV_1 or trough FEV_1.

Table 20. Pooled Studies HZA 106827 and 106863 Change from Baseline WM FEV1 at Week 12 (LOCF, ITT, Subgroup Analysis by Age Group)

	Adolescents (12-17)		Adult (>=18)	
	FF100 N=40	FF/VI 100/25 N=35	FF100 N=354	FF/VI 100/25 N=385
LS Mean Change from Baseline	0.82	0.75	0.34	0.46
Difference vs FF 100		-0.06		0.12
CI		(-0.33, 0.21),		(0.06, 0.18)
P-value		0.65		<0.001

Source: Reviewer
Note: Interaction test: p-value=0.12, there is no statistically significant difference in terms of the treatment difference between the adolescent group and the adult group.

Table 21. Pooled Studies HZA 106827 and 106863 Change from Baseline Trough FEV1 at Week 12 (LOCF, ITT, Subgroup Analysis by Age Group)

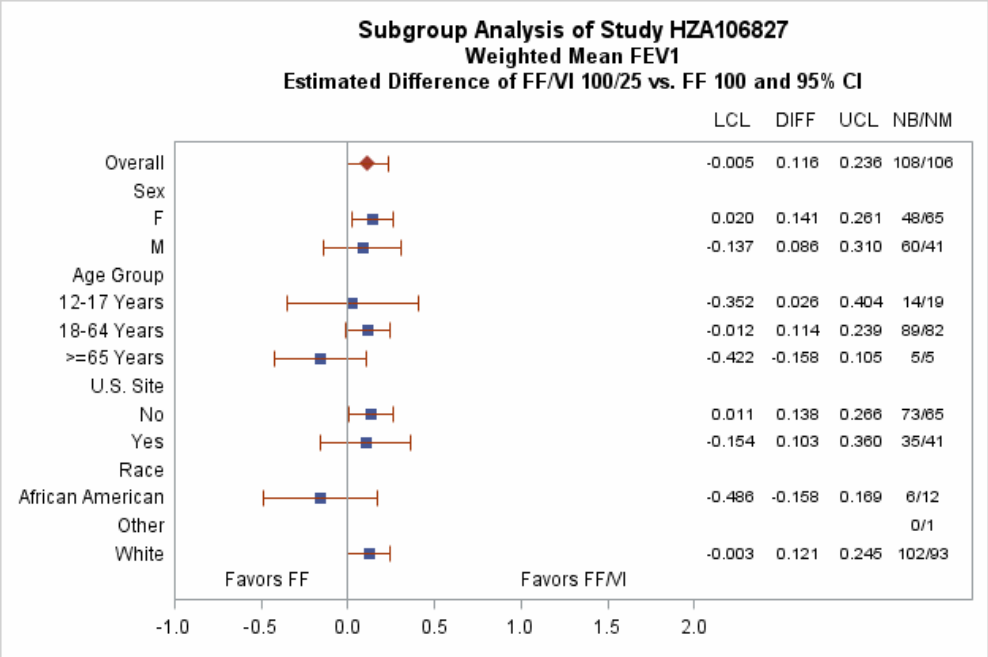
	Adolescents (12-17)		Adult (>=18)	
	FF100 N=51	FF/VI 100/25 N=42	FF100 N=488	FF/VI 100/25 N=492
LS Mean Change from Baseline	0.71	0.66	0.323	0.393
Difference vs FF 100		-0.05		0.07
CI		(-0.30, 0.20),		(0.01, 0.13)
P-value		0.72		0.01

Source: Reviewer
Note: Interaction test: p-value=0.26, there is no statistically significant difference in terms of the treatment difference between the adolescent group and the adult group.

4.1.2.2 Study 27

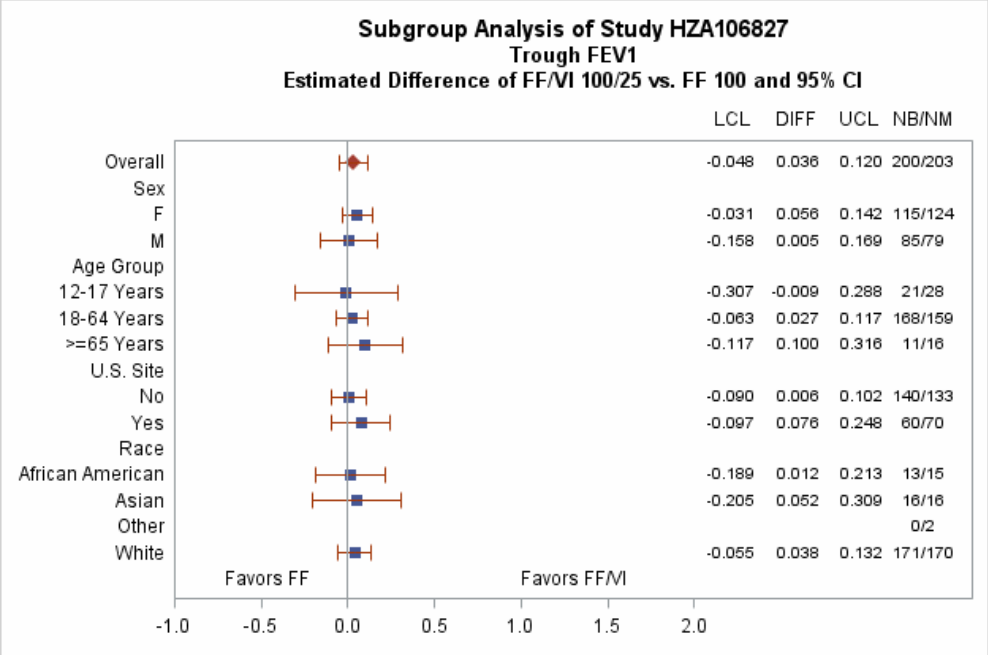
Subgroup analyses of weighted mean FEV_1 and trough FEV_1 by gender, race, age, and geographical region for study 27 are provided in Figures 6 and 7 and Tables 20 and 21, respectively.

Figure 6. Subgroup Analysis of Weighted Mean FEV_1 (Study HZA106827)



Source: Reviewer.
Abbreviations: DIFF: Estimated mean difference between the FF/VI and FF under each subgroup. LCL: Lower limit of the confidence interval of the mean difference; UCL: Upper limit of the confidence interval of the mean difference; NB: Number of subjects under FF/VI; NM: Number of subjects under FF.

Figure 7. Subgroup Analysis of Trough FEV_1 (Study HZA106827)



Source: Reviewer.
Abbreviations: DIFF: Estimated mean difference between the FF/VI and FF under each subgroup. LCL: Lower limit of the confidence interval of the mean difference; UCL: Upper limit of the confidence interval of the mean difference; NB: Number of subjects under FF/VI; NM: Number of subjects under FF.

Table 22. Study HZA 106827 Change from Baseline WM FEV1 at Week 12 (LOCF, ITT, Subgroup Analysis by Age Group)

	Adolescents (12-17)			Adults (>=18)		
	Placebo N=24	FF100 N=19	FF/VI 100/25 N=14	Placebo N=71	FF100 N=87	FF/VI 100/25 N=94
LS Mean Change from Baseline	0.442	0.648	0.675	0.184	0.343	0.452
Difference vs. Placebo		0.207			0.159	
CI		(-0.116, 0.530)			(0.015, 0.303)	
P-value		0.205			0.03	
Difference vs. FF100			0.027			0.109
CI			(-0.347, 0.4)			(-0.027, 0.244)
p-value			0.887			0.116

Source: Reviewer
Note: Interaction test: p-value=0.798, there is no statistically significant difference in terms of the treatment difference between the adolescent group and the adult group.

Table 23. Study HZA 106827 Change from Baseline Trough FEV1 at Week 12 (LOCF, ITT, Subgroup Analysis by Age Group)

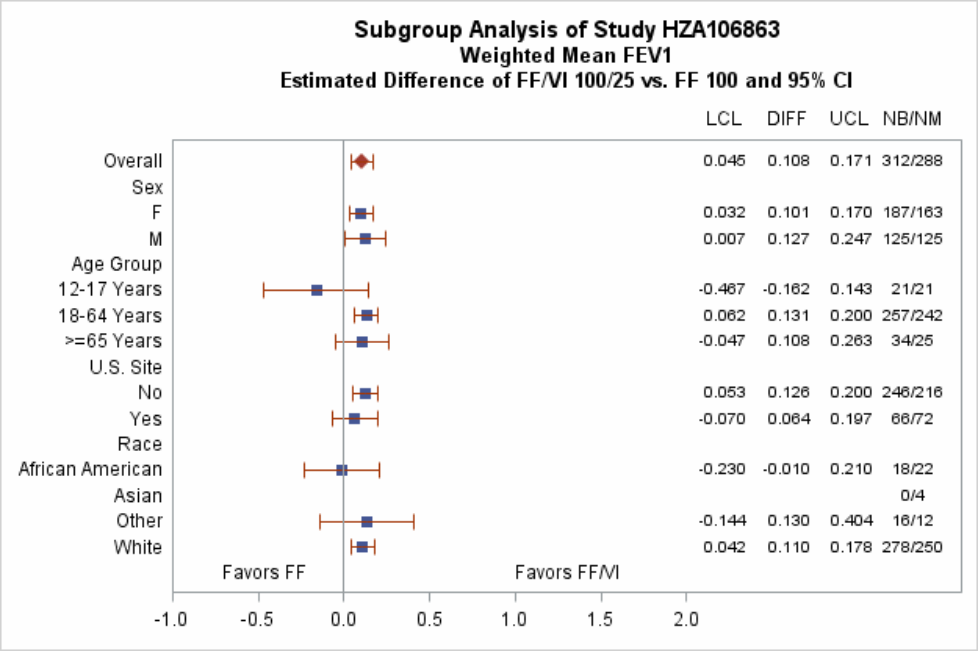
	Adolescents (12-17)			Adults (>=18)		
	Placebo N=33	FF100 N=28	FF/VI 100/25 N=21	Placebo N=160	FF100 N=175	FF/VI 100/25 N=179
LS Mean Change from Baseline	0.365	0.520	0.526	0.197	0.292	0.327
Difference vs. Placebo		0.155			0.095	
CI		(-0.105, 0.415)			(0.001, 0.189)	
P-value		0.238			0.048	
Difference vs. FF100			0.006			0.036
CI			(-0.286, 0.300)			(-0.555, 0.127)
p-value			0.968			0.442

Source: Reviewer
Note: Interaction test: p-value=0.6254, there is no statistically significant difference in terms of the treatment differences between the adolescent group and the adult group.

4.1.2.3 Study 63

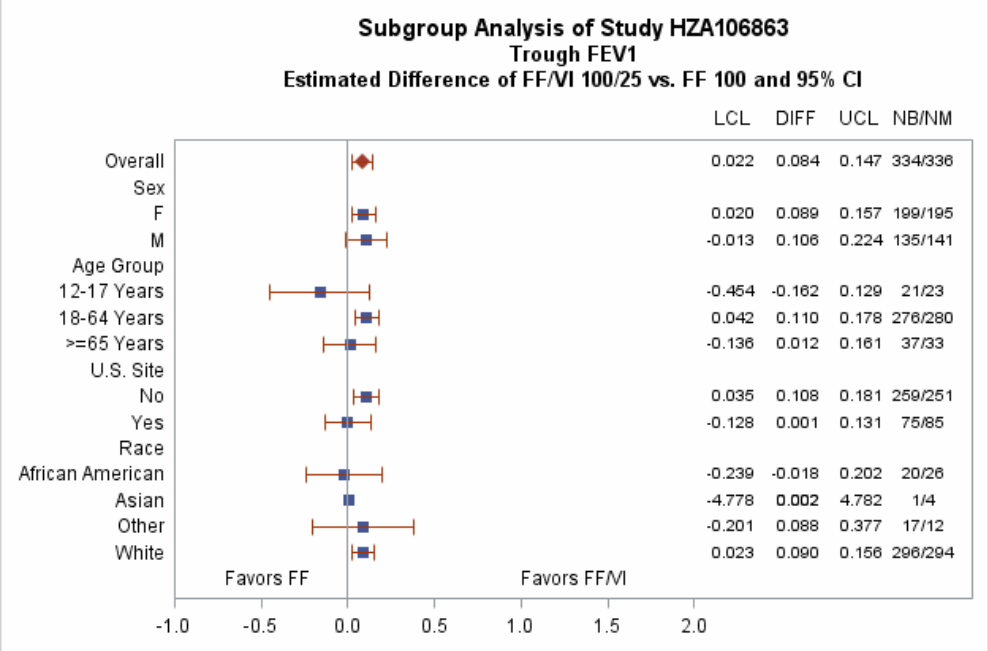
Subgroup analyses of weighted mean FEV_1 and trough FEV_1 by gender, race, age, and geographical region for study 63 are provided in Figures 8 and 9 and Tables 22 and 23, respectively.

Figure 8. Subgroup Analysis of Weighted Mean FEV_1 (Study HZA106863)



Source: Reviewer.
Abbreviations: DIFF: Estimated mean difference between the FF/VI and FF under each subgroup. LCL: Lower limit of the confidence interval of the mean difference; UCL: Upper limit of the confidence interval of the mean difference; NB: Number of subjects under FF/VI; NM: Number of subjects under FF.

Figure 9. Subgroup Analysis of Trough FEV_1 (Study HZA106827)



Source: Reviewer.
Abbreviations: DIFF: Estimated mean difference between the FF/VI and FF under each subgroup. LCL: Lower limit of the confidence interval of the mean difference; UCL: Upper limit of the confidence interval of the mean difference; NB: Number of subjects under FF/VI; NM: Number of subjects under FF.

Table 24. Study HZA 106863 Change from Baseline WM FEV_1 at Week 12 (LOCF, ITT, Subgroup Analysis by Age Group)

	Adolescents (12-17)			Adults (>=18)		
	FF 100 N=21	FF/VI 100/25 N=21	FF/VI 200/25 N=13	FF 100 N=267	FF/VI 100/25 N=291	FF/VI 200/25 N=299
LS Mean Change from Baseline	0.961	0.77	0.985	0.343	0.447	0.463
Difference vs. FF100		-0.190	0.024		0.104	
CI		(-0.496, 0.115)	(-0.335, 0.383)		(0.036, 0.171)	
P-value		0.215	0.892		0.003	
Difference vs. FF/VI 100/25			0.215			0.017
CI			(-0.155, 0.584)			(-0.049, 0.082)
p-value			0.248			0.619

Source: Reviewer
Note: Interaction test: p-value=0.122, there is no statistically significant difference in terms of the treatment difference between the adolescent group and the adult group.

Table 25. Study HZA 106863 Change from Baseline Trough FEV_1 at Week 12 (LOCF, ITT, Subgroup Analysis by Age Group)

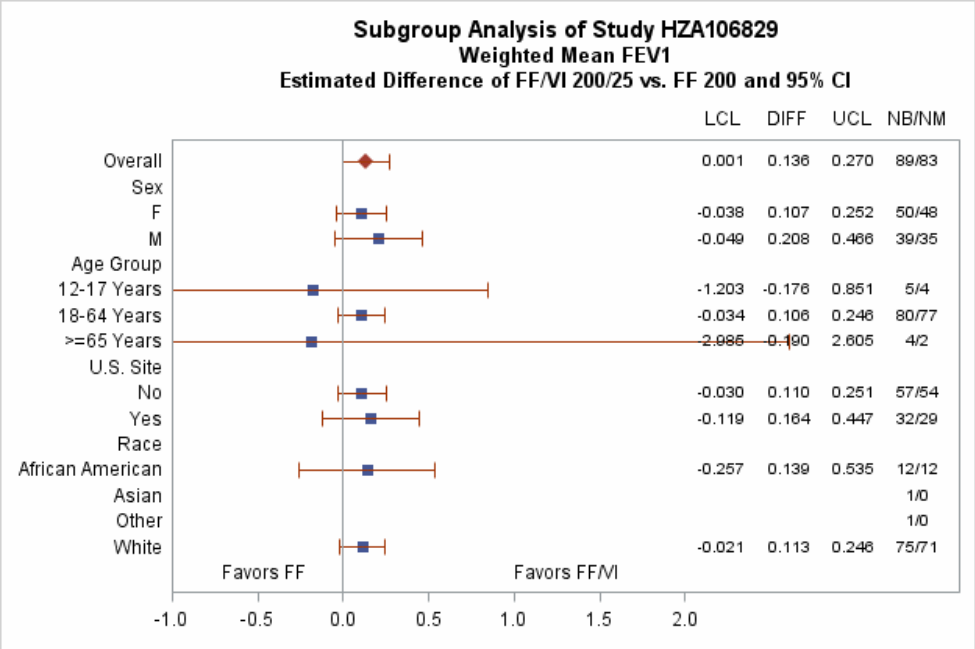
	Adolescents (12-17)			Adults (>=18)		
	FF 100 N=23	FF/VI 100/25 N=21	FF/VI 200/25 N=14	FF 100 N=313	FF/VI 100/25 N=313	FF/VI 200/25 N=323
LS Mean Change from Baseline	0.954	0.758	0.854	0.341	0.425	0.414
Difference vs. FF100		-0.196	-0.100		0.084	
CI		(-0.498, 0.105)	(-0.446, 0.246)		(0.018, 0.150)	
P-value		0.197	0.563		0.012	
Difference vs. FF/VI 100/25			0.096			-0.011
CI			(-0.262, 0.455)			(-0.076, 0.055)
p-value			0.591			0.752

Source: Reviewer
Note: Interaction test: p-value=0.170, there is no statistically significant difference in terms of the treatment difference between the adolescent group and the adult group.

4.1.2.4 Study 29

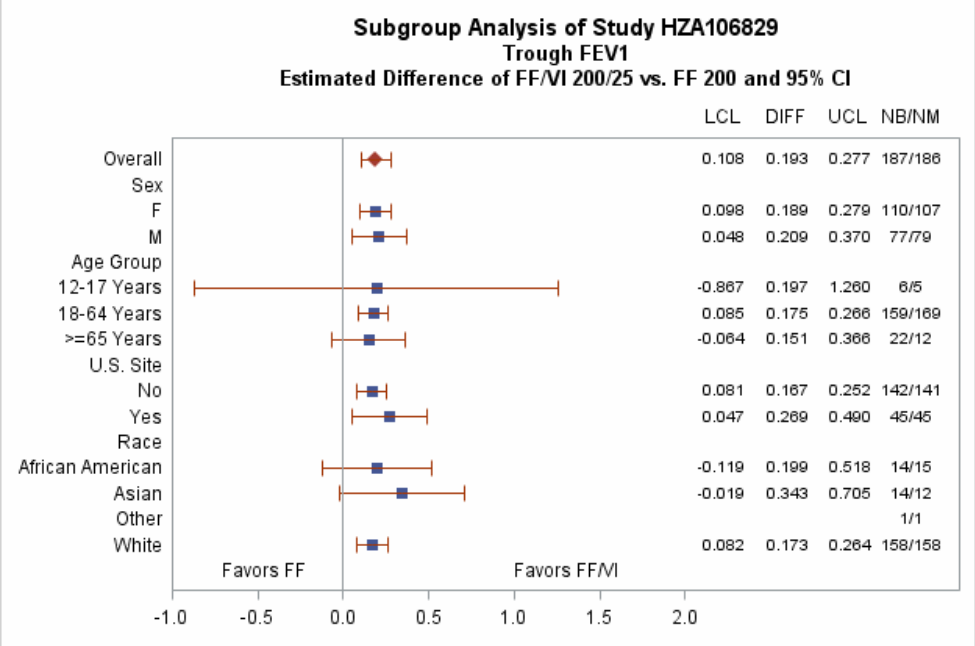
Subgroup analyses of weighted mean FEV_1 and trough FEV_1 by gender, race, age, and geographical region for study 29 are provided in Figures 10 and 11 and Tables 24 and 25, respectively.

Figure 10. Subgroup Analysis of Weighted Mean FEV_1 (Study HZA106829)



Source: Reviewer.
Abbreviations: DIFF: Estimated mean difference between the FF/VI and FF under each subgroup. LCL: Lower limit of the confidence interval of the mean difference; UCL: Upper limit of the confidence interval of the mean difference; NB: Number of subjects under FF/VI; NM: Number of subjects under FF.

Figure 11. Subgroup Analysis of Trough FEV_1 (Study HZA106829)



Source: Reviewer.
Abbreviations: DIFF: Estimated mean difference between the FF/VI and FF under each subgroup. LCL: Lower limit of the confidence interval of the mean difference; UCL: Upper limit of the confidence interval of the mean difference; NB: Number of subjects under FF/VI; NM: Number of subjects under FF.

Table 26. Study HZA 106829 Change from Baseline WM FEV_1 at Week 24 (LOCF, ITT, Subgroup Analysis by Age Group)

	Adolescents (12-17)			Adults (>=18)		
	FP 500 N=5	FF 200 N=4	FF/VI 200/25 N=5	FP 500 N=81	FF 200 N=79	FF/VI 200/25 N=84
LS Mean Change from Baseline	1.084	0.695	0.644	0.197	0.345	0.428
Difference vs. FP 500		-0.390	-0.441		0.148	0.231
CI		(-1.212, 0.433)	(-1.382, 0.500)		(-0.005, 0.302)	(0.082, 0.38)
P-value		0.306	0.312		0.058	0.003
Difference vs. FF200			-0.051			0.083
CI			(-0.993, 0.891)			(-0.068, 0.234)
p-value			0.903			0.279

Source: Reviewer
Interaction test: p-value=0.27, there is no statistically significant difference in terms of the treatment difference between the adolescent group and the adult group.

Table 27. Study HZA 106829 Change from Baseline Trough FEV_1 at Week 24 (LOCF, ITT, Subgroup Analysis by Age Group)

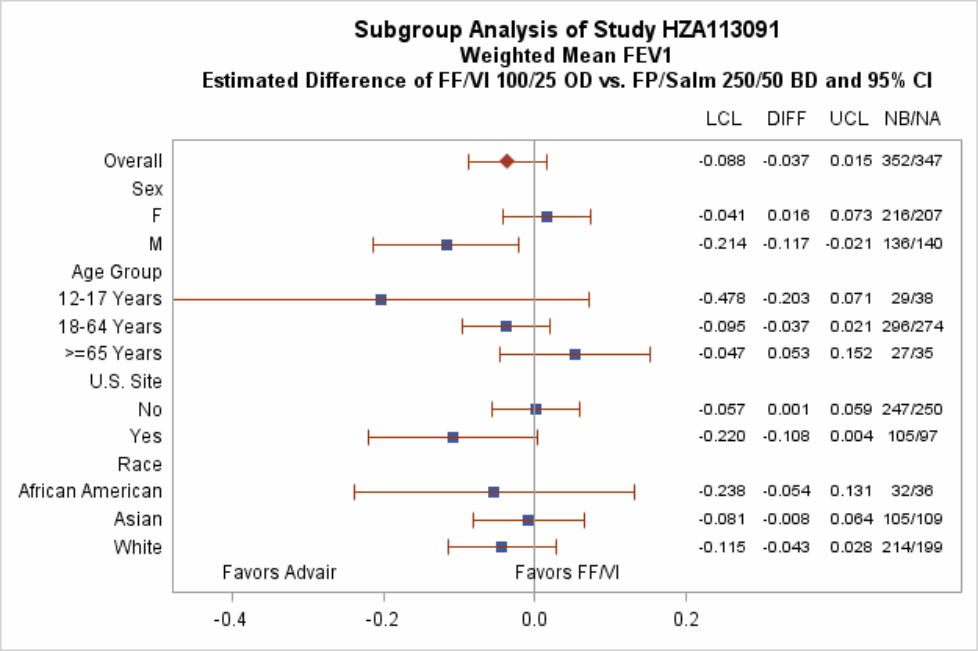
	Adolescents (12-17)			Adults (>=18)		
	FP 500 N=8	FF 200 N=5	FF/VI 200/25 N=6	FP 500 N=182	FF 200 N=181	FF/VI 200/25 N=181
LS Mean Change from Baseline	0.648	0.836	1.043	0.151	0.205	0.364
Difference vs. FP 500		0.198	0.405		0.054	0.214
CI		(-0.693, 1.090)	(-0.452, 1.262)		(-0.035, 0.144)	(0.125, 0.303)
P-value		0.639	0.326		0.235	<0.0001
Difference vs. FF200			0.207			0.160
CI			(-0.773, 1.186)			(0.07, 0.249)
p-value			0.656			0.0005

Source: Reviewer
Interaction test: p-value=0.710, there is no statistically significant difference in terms of the treatment difference between the adolescent group and the adult group.

4.1.2.5 Study 91

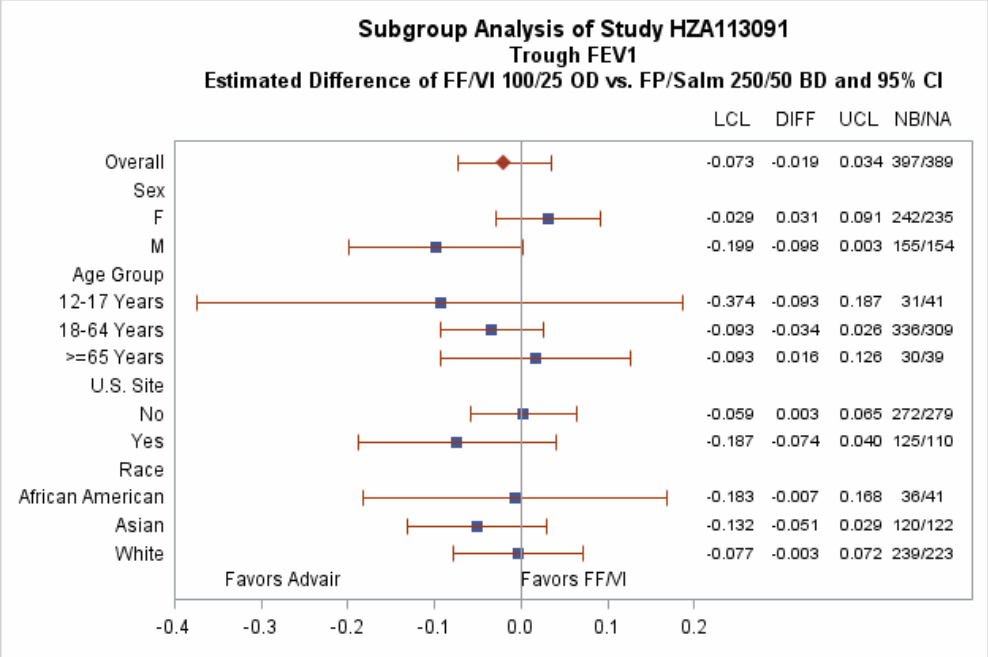
Subgroup analyses of weighted mean FEV_1 and trough FEV_1 by gender, race, age, and geographical region for study 91 are provided in Figures 12 and 13 and Tables 26 and 27, respectively.

Figure 12. Subgroup Analysis of Weighted Mean FEV_1 (Study HZA113091)



Source: Reviewer.
Abbreviations: DIFF: Estimated mean difference between the FF/VI and FF under each subgroup. LCL: Lower limit of the confidence interval of the mean difference; UCL: Upper limit of the confidence interval of the mean difference; NB: Number of subjects under FF/VI; NA: Number of subjects under Advair.

Figure 13. Subgroup Analysis of Trough FEV_1 (Study HZA106829)



Source: Reviewer.
Abbreviations: DIFF: Estimated mean difference between the FF/VI and FF under each subgroup. LCL: Lower limit of the confidence interval of the mean difference; UCL: Upper limit of the confidence interval of the mean difference; NB: Number of subjects under FF/VI; NA: Number of subjects under Advair.

Table 28. Study HZA 113091 Change from Baseline WM FEV_1 at Week 24 (LOCF, ITT, Subgroup Analysis by Age Group)

	Adolescents (12-17)		Adults (>=18)		Overall	
	Fp/Salm 250/50 BID N=38	FF/VI 100/25 N=29	Fp/Salm 250/50 BID N=309	FF/VI 100/25 N=323	Fp/Salm 250/50 BID N=309	FF/VI 100/25 N=323
LS Mean Change from Baseline	0.691	0.488	0.347	0.320	0.377	0.341
Difference vs. Fp/Salm 250/50 BID		-0.203		-0.027		-0.037
CI		(-0.478, 0.071)		(-0.080, 0.026)		(-0.088, 0.015)
P-value		0.144		0.315		0.162

Source: Reviewer
Interaction test: p-value=0.178, there is no statistically significant difference in terms of the treatment difference between the adolescent group and the adult group.

Table 29. Study HZA 113091 Change from Baseline Trough FEV_1 at Week 24 (LOCF, ITT, Subgroup Analysis by Age Group)

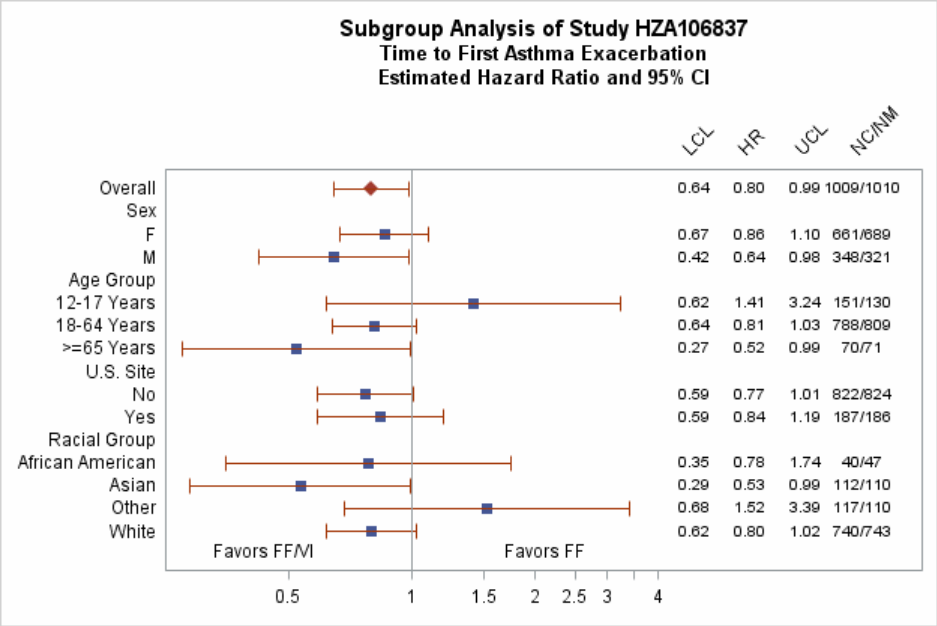
	Adolescents (12-17)		Adults (>=18)	
	Fp/Salm 250/50 BID N=41	FF/VI 100/25 N=31	Fp/Salm 250/50 BID N=348	FF/VI 100/25 N=366
LS Mean Change from Baseline	0.526	0.432	0.284	0.258
Difference vs. Fp/Salm 250/50 BID		-0.093		-0.026
CI		(-0.374, 0.187)		(-0.081, 0.029)
P-value		0.508		0.348

Source: Reviewer
Interaction test: p-value=0.896, there is no statistically significant difference in terms of the treatment difference between the adolescent group and the adult group.

4.1.2.6 Study 37

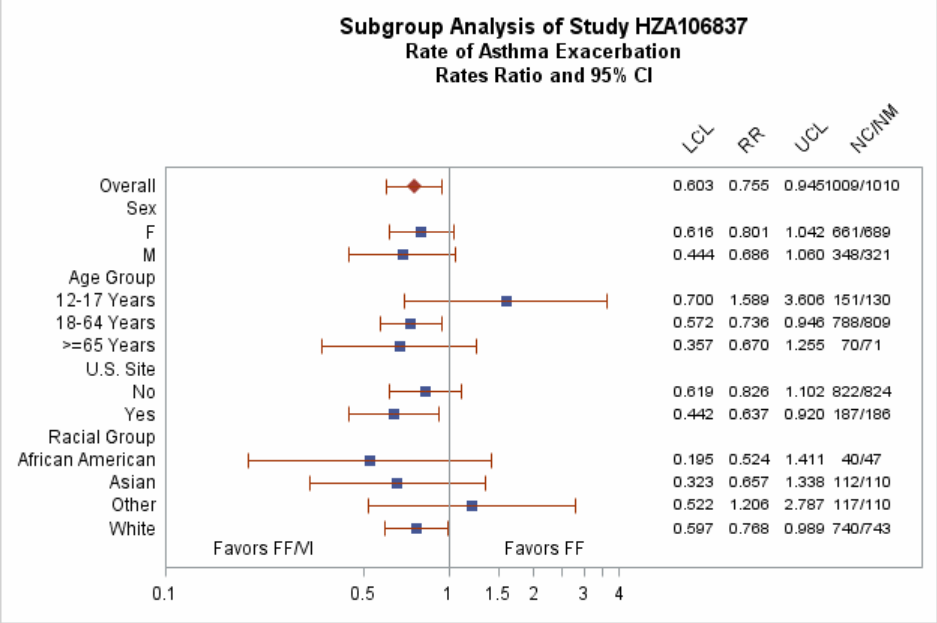
Subgroup analyses of time to first severe exacerbation and the severe exacerbation rate by gender, race, age, and geographical region for study 37 are provided in Figures 14 and 15 and Tables 28 and 29, respectively.

Figure 14. Subgroup Analysis of Time to First Asthma Exacerbation (Study 106837)



Source: Reviewer.
Abbreviations: HR: Estimated hazard ratio from analysis with a Cox Proportion hazard model of time to first asthma exacerbation between the FF/VI and FF; LCL: Lower limit of the confidence interval of the hazard ratio; UCL: Upper limit of the confidence interval of the hazard ratio; NC/NM: Number of patients under the FF/VI (Combination) arm versus number of patients under the FF (Mono Therapy) arm.

Figure 15. Subgroup Analysis of Incidence Risk Ratio of Asthma Exacerbation (Study HZA 106837)



Source: Reviewer.
Abbreviations: RR: Incidence Rate Ratio of asthma exacerbations between the FF/VI and the FF arm from analysis with a negative binomial regression model under each subgroup; LCL: Lower limit of the confidence interval of the rates ratio; UCL: Upper limit of the confidence interval of the rates ratio; NC/NM: Number of patients under the FF/VI (Combination) arm versus number of patients under the FF (Mono Therapy) arm.

Table 30. Study HZA 106837 Cox Proportional Hazard Analysis of Time to First Severe Asthma Exacerbation (ITT Population, Subgroup Analysis by Age Group)

	Adolescents (12-17)		Adults (>=18)	
	FF 100 N=130	FF/VI 100/25 N=151	FF 100 N=880	FF/VI 100/25 N=858
Number of patients with at least 1 event (n) %	9 (7%)	15 (10%)	177 (20%)	139 (16%)
Adjusted Probability of 1+ Severe Asthma Exacerbations by 52 Weeks (%)	8.7	12.0	16.8	13.1
95% CI	(3.0, 14.0)	(6.0, 17.6)	(14.2, 19.4)	(10.8, 15.4)
FF/VI 100/25 vs. FF 100				
Hazard ratio		1.405		0.764
95% CI		(0.614, 3.213)		(0.612, 0.955)
p-value		0.42		0.018

Source: Reviewer
Interaction test: p-value=0.16, there is no statistically significant difference in terms of the treatment hazard ratios between the adolescent group and the adult group.

Table 31. Study HZA 106837 Analysis of Rate of Severe Asthma Exacerbations per Subject per Year (Negative Binomial Model) (ITT Population, Subgroup Analysis by Age Group)

	Adolescents (12-17)		Adult (>=18)	
	FF 100 N=130	FF/VI 100/25 N=151	FF 100 N=880	FF/VI 100/25 N=858
n	130	151	880	858
Mean severe asthma exacerbation rate	0.08	0.13	0.31	0.21
FF/VI 100/25 vs. FF 100 Ratio		1.62		0.68
95% CI		(0.68, 3.86)		(0.54, 0.87)
p-value		0.27		0.002

Source: Reviewer
Interaction analysis performed with a negative binomial regression model. The model to test interaction includes Treatment, Age group and age group * treatment interaction as covariates. Interaction test: p = 0.055

4.1.3 Discussion of the Findings

Compared with findings from age subgroup analyses in lung function studies (where results within the adolescents are numerically but not statistically significantly unfavorable and the treatment-by-age interaction tests are not significant so that there is no evidence of a differing treatment effect between age groups), results for the exacerbation endpoint send out a stronger signal of a differing treatment effect across age groups with the estimate of the treatment-by-age interaction in the analysis of the rate of exacerbations being borderline significant (p=0.055). This borderline significant interaction indicates that the treatment effect differs across age groups; however, this difference may be qualitative (i.e., a reversal of the treatment effect from adults to adolescents) or only quantitative (i.e., only the magnitude of the difference between treatment groups varies across age groups but the benefit of the combination is positive in both subgroups). While we note these seemingly different effects of the treatment across age groups, we acknowledge that the challenge in interpreting subgroup findings is whether to differences in

treatment effect across subgroups to true heterogeneity or to random chance. When a large number of subgroups are evaluated, even when the study drug is truly effective and the effect is homogenous across subgroups, the probability that at least one of the evaluated subgroups will have an observed treatment effect in the opposite direction due to chance alone can be large.

In a scenario in which efficacy is demonstrated in the overall population but there appear to be a subgroup that did not benefit, according to the Subgroup Analysis Working Document of the Office of Biostatistics, such a product might be approved for the whole population studied, or it might be approved only for the favorable subgroup. Typically there would be a biologically plausible explanation for why the medical product would be effective only within the adult subgroup. Additional external data may be needed to support the conclusion. For such trials, the actual risk is for an increased Type II error (failing to establish efficacy for a subgroup when in fact the product is effective in the subgroup).

4.2 Other Special/Subgroup Populations

No other subgroups were analyzed.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

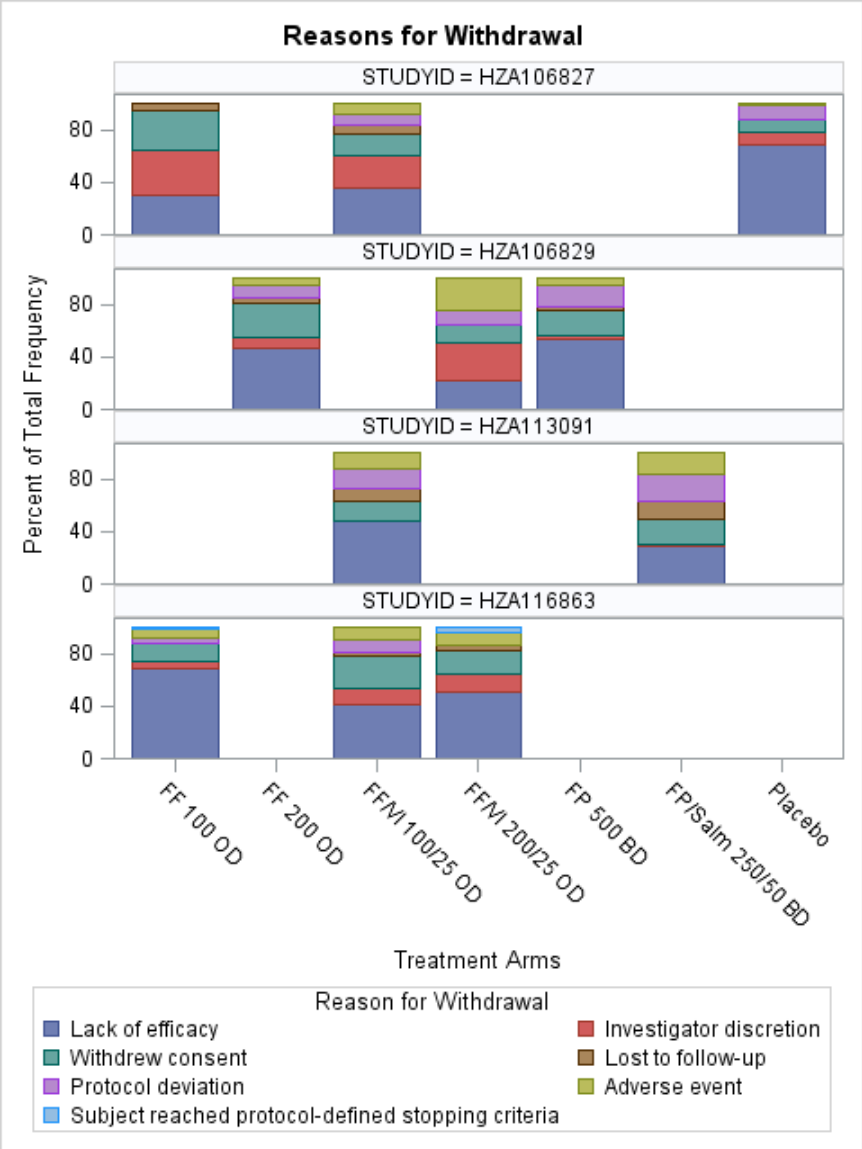
5.1.1 Missing Data and Sensitivity Analysis

5.1.1.1 WM FEV1 and Trough FEV1 (Studies 27, 63, 29, and 91)

As described in section 3.2, for studies 27, 63, 29, and 91, the primary analyses of WM FEV1 and trough FEV1 were conducted, as pre-specified, on the ITT population in which the missing data were imputed with the LOCF method. Use of the LOCF imputation method can be problematic in that it may result in biased and overly precise estimates of the treatment effect (in comparison to the treatment effect in all randomized subjects regardless of adherence). When the last observed value is not indicative of the status of the subject (regardless of adherence) at the planned measurement time the treatment effect obtained from a LOCF method should not be interpreted as a result representative of the planned measurement time point. Rather, it is an analysis estimating the treatment effect using the last available observation (LAO), a mean effect at various measurement time points.

Supportive analyses on trough FEV_1 were also conducted utilizing a mixed model for repeated measures (MMRM) analysis on all available data which generally gives estimators with relatively small bias and controls type I error rates under the condition of missing completely at random (MCAR) or missing at random (MAR). However, these assumptions are likely not plausible in the setting of this study since premature study discontinuation may be related to treatment assignment. With imbalance between treatment arms for dropout rates due to lack of efficacy and adverse events (Table 5, Table 8, Table 11, Table 14, and Table 17), the strong assumption of MCAR and MAR are not likely to be valid. Also, as illustrated by Figure 16, generally speaking, there is a trend of higher dropout rates due to lack of efficacy as dosage going to the direction of low level of monotherapy or placebo. Results of the MMRM analysis are not useful as an estimate of the effect of the treatment at the scheduled final measurement time point.

Figure 16. Reasons for Withdrawal



Source: Reviewer.

A sensitivity analysis based on the assumption that the expected value of the measure for those who drop out is not better than that for those who complete the study was conducted to evaluate the robustness of the primary analysis findings to missing data. The cumulative responder rate imputation is an adjustment to the usual test for a difference between treatments that allows for the inclusion of the probable effect of the dropouts; it provides a bound on the test for efficacy of the treatment.

There are different ways to impute missing data under the cumulative responder rate approach by assigning different response values to the dropouts. Considering that across the studies, significant proportion of the early withdrawals was due to "lack of efficacy" and "withdrew consent", by replacing all change from baseline trough FEV_1 values that are smaller than zero

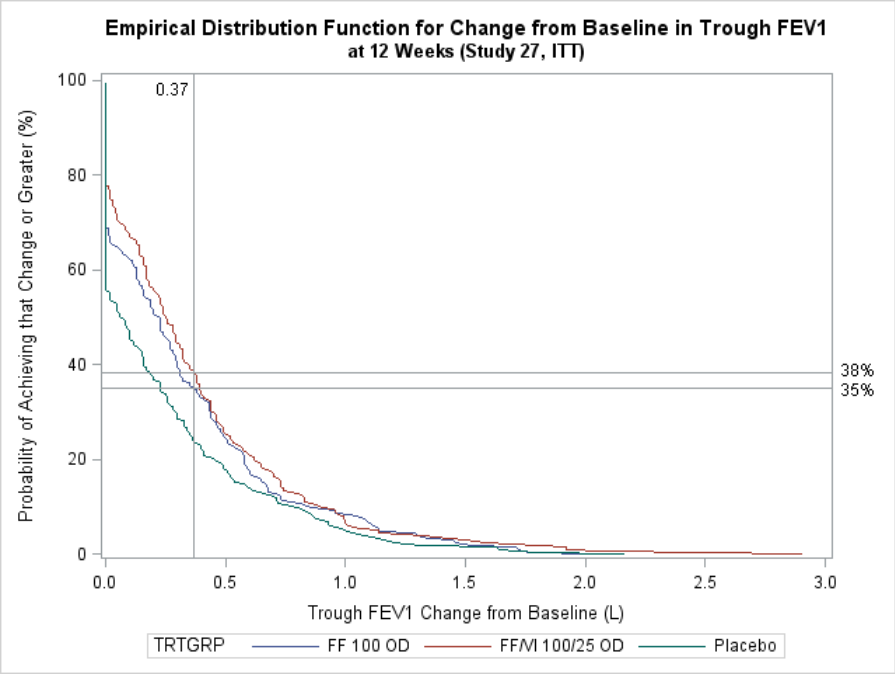
and imputing all missing change from baseline values with 0, we treat dropouts and subjects with negative effects equally as no effect (a value of 0).

Figure 17 to Figure 20 provide continuous responder curves (i.e., empirical distribution functions) for studies 27, 63, 29 and 91, respectively. These presentations are developed as follows. Each patient is classified as having been successfully or unsuccessfully treated according to whether or not the patient reached a certain threshold for the change from baseline in trough FEV_1 at the study primary time-point (week 12 or 24). This dichotomization of the change from baseline in trough FEV_1 is repeated across a range of possible thresholds, in this case from 0 to the maximum change from baseline value under each study. Patients with missing or negative change from baseline trough FEV_1 data at the primary time-point are classified as unsuccessfully treated for all thresholds. In the continuous responder curve, the x-axis displays the thresholds required to classify a patient as a successfully treated patient. The y-axis represents the proportion of ITT patients who achieved the corresponding threshold. For example, using study 27, in which the estimated least square mean change from baseline trough FEV_1 at week 12 is 0.37 L for the FF/VI 100/25 arm. In Figure 17, at the vertical reference line of a change from baseline in trough FEV_1 of 0.37 L, the continuous responder plot illustrates that 38% of FF/VI 100/25 patients had FEV_1 improved by at least 0.37 L while 35% of FF 100 patients experienced such a change.

As shown in all the figures, there is an initial dramatic drop from 100% to approximately 80% or below in the y-axis, corresponding to the sum of proportion of patients who dropped out since patients with missing data or negative change from baseline values were classified as unsuccessfully treated for all thresholds. Generally, across the studies, no-effects were more frequent in the placebo or FF monotherapy groups compared to the FF/VI combination group. Also evident from the figures is that there is little separation between the treatment groups in studies 27 and 91, while there is significant separation between curves corresponding to FF and FF/VI arms in studies 63 and 29.

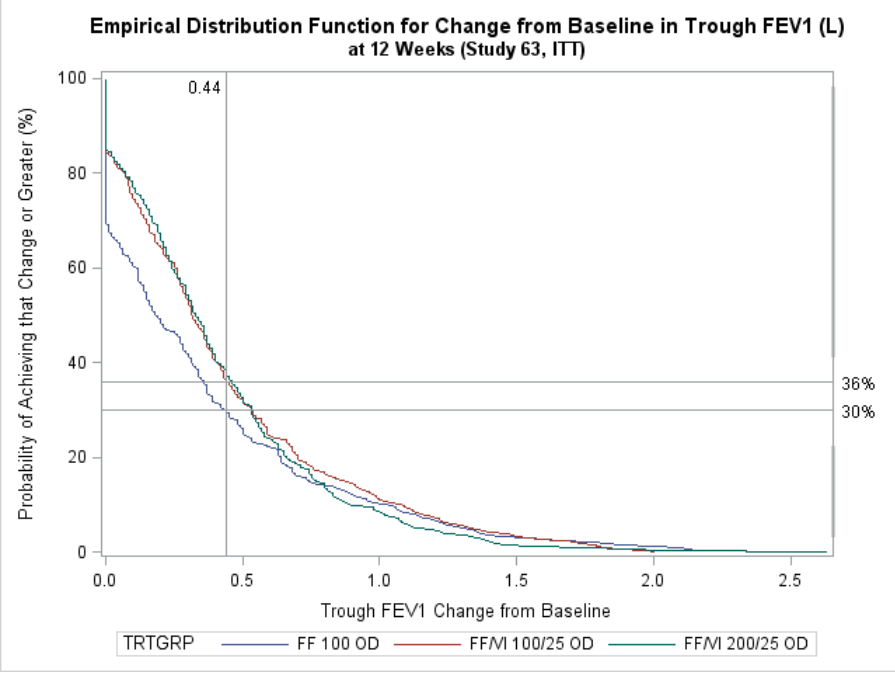
Then a corresponding rank sum statistic of Mann-Whitney-Wilcoxon test is calculated on the modified data. That is, output of the cumulative responder rate approach is in the form of an empirical distribution function plot and a corresponding p-value of the separation of the two distributions. Results from the Mann-Whitney-Wilcoxon tests, in the footnotes under Figure 17 to Figure 20 are consistent with the LOCF ANCOVA results and provide reassurance that the overall conclusions that FF/VI is more effective than FF alone in terms of trough FEV_1 are reliable despite missing data.

Figure 17. Empirical Distribution Function for Change from Baseline in Trough FEV_1 (Study 27, ITT)



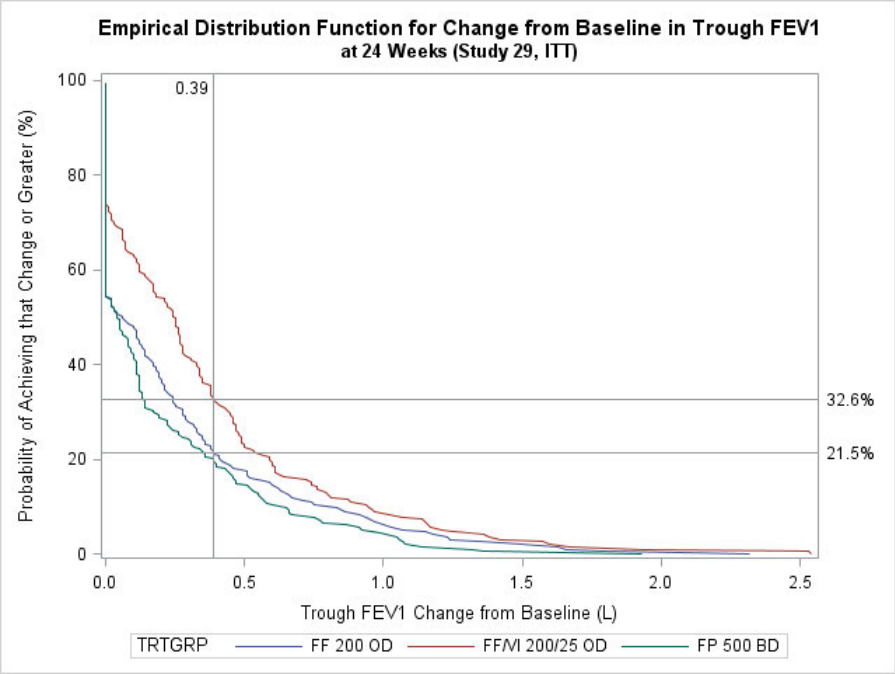
Source: Reviewer
Mann-Whitney-Wilcoxon Two-Sample Test between FF/VI 100/25 OD and FF 100 OD: p = 0.19.

Figure 18. Empirical Distribution Function for Change from Baseline in Trough FEV_1 (Study 63, ITT)



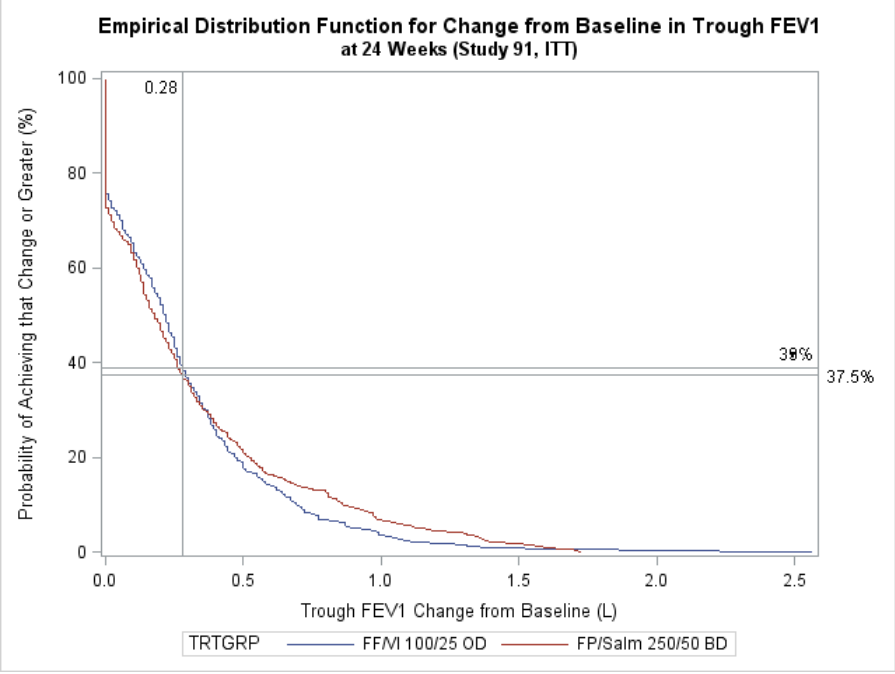
Source: Reviewer
Mann-Whitney-Wilcoxon Two-Sample Test between FF/VI 100/25 OD and FF 100 OD: p < 0.001.

Figure 19. Empirical Distribution Function for Change from Baseline in Trough FEV_1 (Study 29, ITT)



Source: Reviewer
Mann-Whitney-Wilcoxon Two-Sample Test between FF/VI 200/25 OD and FF 200 OD: $p < 0.001$.

Figure 20. Empirical Distribution Function for Change from Baseline in Trough FEV_1 (Study 91, ITT)



Source: Reviewer
Mann-Whitney-Wilcoxon Two-Sample Test between FF/VI 100/25 OD and FP/Salm 250/50 BD: $p = 0.64$.

5.1.1.2 Severe Exacerbation (Study 37)

The time to first severe asthma exacerbation was the primary efficacy analysis for study 37 and was analyzed, as pre-specified, using a Cox proportional hazards regression model, adjusting for baseline FEV_1, sex, age, region and treatment. Subjects who did not experience an exacerbation during the treatment period were treated as censored observations at the time of treatment discontinuation, whether premature or at the end of the study. Subjects lost to follow-up were censored at their last visit. Cox proportional hazards regression modeling relies on an assumption that censoring is non-informative meaning that the censoring should not be related to the probability of an exacerbation occurring. This assumption may seem plausible for the cases censored by the administrative end of the study; however, this assumption is not so probable in the case of subjects censored by early discontinuation of study treatment.

The rate of severe asthma exacerbations per subject per year over the treatment period, a secondary efficacy analysis, was analyzed, as pre-specified, using a negative binomial regression model with log time on treatment as an offset variable. In a negative binomial regression model, the mean count of incidence is a multiplication of the time on treatment and the rate per person-year of time on treatment. By applying the log transformation on both sides of the equation, the relationship between the log(mean incidence count) and the linear predictor is offset by the amount of log(time on treatment). The model included adjustment for effects due to baseline FEV_1, sex, age, region and treatment. The adjusted mean rates per year, treatment ratio, and confidence interval for the ratio were presented. While the negative binomial regression model does adjust for the time on treatment in calculating the mean exacerbation rate, estimates of the treatment effect may be inappropriately influenced by premature study discontinuation if observed data are not representative of post-discontinuation data.

As such, it is desirable to understand whether the significant treatment effect on time to first exacerbation represented in Table 18 and mean exacerbation rate represented in Table 19 are reliable despite missing data. Early study treatment discontinuation and therefore early study withdrawal as the protocol did not distinguish between the two was slightly more common in the FF 100 QD group at approximately 15% as opposed to 12% in the FF 100/25 QD groups. (Table 17). Early study withdrawal before experiencing a first exacerbation was slightly lower and not different between treatment groups at approximately 10%. The most common reasons for early withdrawal were “withdraw consent,” “adverse event,” “lack of efficacy” and “protocol deviation.” Study designs that require continued follow-up of patients even after early study treatment discontinuation are desirable in this setting so that an estimate of the effect of the investigational product in all patients randomized regardless of adherence can be obtained. In the absence of such data, we conducted a tipping point analyses to assess how extreme the off–study-treatment unobserved data would have to have been to negate the treatment effects estimated from the observed data. If the tipping point is so extreme that it is not clinically plausible, one may conclude that demonstration of efficacy is reliable despite the missing data.

Tipping point analyses and the associated statistical methods (including means to randomly generate unobserved data under varying sets of assumptions) are currently developing in the literature and in certain areas are not yet widely explored and accepted. An attractive feature of

the Cox model that is used in the primary analysis of time to first exacerbation is that it does not require any assumptions about the form of the baseline hazard rate. However, this feature also makes it challenging to impute missing survival times without additional parametric assumptions. That being said, we were able to implement a tipping point approach in the context of the negative binomial evaluation of severe exacerbation rates, where imputation of missing exacerbation counts post-discontinuation could be based on a negative binomial distribution.

Tables 30 and 31 provide estimates of the treatment effect (risk ratio and 95% confidence interval) and p-values associated with a test of whether the risk ratio differs from one for FF/VI 100/25 relative to FF 100 in the severe exacerbation rate, respectively. These analyses incorporate both observed data and imputed data. Imputed data are generated with varying assumptions about the rate of events for each treatment group (from 0.1 to 0.3 mean exacerbation rate) in patients who withdrew from the study early during the time for which they should have been observed but were not. Results shaded in aqua indicate analyses based on assumptions that require equal post-discontinuation exacerbation rates in each treatment group. Areas below those cells assume that the post-discontinuation rates in the FF/VI arm would have been lower than that of the FF arm. Areas above the diagonal represent the cases where the unobserved exacerbation rates for the FF/VI arm are assumed to be higher than that of the FF arm. Pink shaded regions include the cases where the assumptions regarding the post-discontinuation data are sufficient to "tip" the analysis of the risk ratio for the mean exacerbation rate (including observed and unobserved imputed data) so that the result numerically favoring the FF/VI group is no longer associated with a p-value less than 0.05.

In order for the hypothesis test to fail to demonstrate an advantage of FF/VI 100/25 over FF 100, the mean rate of severe exacerbation would need to be at least 67% larger in the FF/VI 100/25 dropouts than in the FF 100 dropouts at approximately 0.30 and 0.18 events per subject per year, respectively. As a point of reference, the mean exacerbation rates in the observed data were 0.14 and 0.19 for the FF/VI and FF groups, respectively. The post-discontinuation mean exacerbation rate for FF/VI patients would have to be more than twice as high as the observed mean exacerbation rate for FF/VI patients (0.30/0.14=2.14) while the post-discontinuation mean exacerbation rate for the FF patients would have to remain similar to the observed exacerbation rate for FF patients (0.18/0.19=0.95) to reach the tipping point for the test of the risk ratio being equal to one. Given the similar proportions of patients and distributions of reasons for early withdrawal on the two treatment arms, an assumption of such large differences between the outcomes in dropouts on the two arms seems implausible. Therefore, these tipping point analyses largely support the findings of the key efficacy analyses of the observed data presented in Section 3.2.5.4, Table 19.

Table 32. Estimate of the Treatment Effect (Risk Ratio and 95% Confidence Interval) for FF/VI 100/25 versus FF 100 under Varying Assumptions about the Rate of Exacerbations on Each Treatment Arm in Patients who Withdrew from the Study Early During the Time for which they Should Have Been Observed but Were Not

		FF/VI 100/25 Assumed Exacerbation Rate Post-Withdrawal										
		0.10	**0.12**	**0.14**	**0.16**	**0.18**	**0.20**	**0.22**	**0.24**	**0.26**	**0.28**	**0.30**
FF 100 Assumed Exacerbation Rate Post-Withdrawal	**0.10**	0.77 (0.62, 0.95)	0.77 (0.62, 0.96)	0.78 (0.63, 0.97)	0.79 (0.63, 0.98)	0.79 (0.64, 0.98)	0.80 (0.64, 0.99)	0.80 (0.64, 1.00)	0.81 (0.65, 1.00)	0.81 (0.65, 1.01)	0.82 (0.67, 1.01)	0.83 (0.67, 1.03)
	0.12	0.76 (0.62, 0.94)	0.77 (0.62, 0.96)	0.78 (0.62, 0.96)	0.78 (0.63, 0.97)	0.79 (0.63, 0.98)	0.79 (0.64, 0.98)	0.80 (0.64, 0.99)	0.80 (0.64, 1.00)	0.81 (0.65, 1.01)	0.82 (0.66, 1.01)	0.82 (0.66, 1.02)
	0.14	0.76 (0.61, 0.93)	0.76 (0.62, 0.95)	0.77 (0.62, 0.96)	0.78 (0.62, 0.97)	0.78 (0.63, 0.97)	0.79 (0.64, 0.98)	0.79 (0.64, 0.99)	0.80 (0.64, 0.99)	0.80 (0.65, 1.00)	0.81 (0.65, 1.01)	0.82 (0.66, 1.02)
	0.16	0.75 (0.61, 0.94)	0.76 (0.61, 0.94)	0.77 (0.62, 0.95)	0.77 (0.62, 0.96)	0.78 (0.62, 0.97)	0.78 (0.63, 0.97)	0.79 (0.63, 0.98)	0.79 (0.64, 0.99)	0.80 (0.64, 0.99)	0.81 (0.65, 1.00)	0.81 (0.65, 1.01)
	0.18	0.75 (0.60, 0.93)	0.75 (0.61, 0.94)	0.76 (0.61, 0.95)	0.77 (0.62, 0.95)	0.77 (0.62, 0.96)	0.78 (0.62, 0.97)	0.78 (0.63, 0.97)	0.79 (0.63, 0.98)	0.79 (0.64, 0.99)	0.80 (0.64, 1.00)	0.81 (0.65, 1.00)
	0.20	0.74 (0.60, 0.92)	0.75 (0.60, 0.93)	0.76 (0.61, 0.94)	0.76 (0.61, 0.95)	0.77 (0.62, 0.95)	0.77 (0.62, 0.96)	0.78 (0.62, 0.97)	0.78 (0.63, 0.97)	0.79 (0.64, 0.98)	0.80 (0.64, 0.99)	0.80 (0.64, 1.00)
	0.22	0.74 (0.59, 0.92)	0.74 (0.60, 0.92)	0.75 (0.61, 0.93)	0.76 (0.61, 0.94)	0.76 (0.61, 0.95)	0.77 (0.62, 0.95)	0.77 (0.62, 0.96)	0.78 (0.62, 0.97)	0.78 (0.63, 0.97)	0.79 (0.64, 0.98)	0.80 (0.64, 0.99)
	0.24	0.73 (0.59, 0.91)	0.74 (0.59, 0.92)	0.74 (0.60, 0.92)	0.75 (0.60, 0.93)	0.76 (0.61, 0.94)	0.76 (0.61, 0.94)	0.77 (0.62, 0.95)	0.77 (0.62, 0.96)	0.78 (0.62, 0.96)	0.78 (0.63, 0.97)	0.79 (0.64, 0.98)
	0.26	0.73 (0.58, 0.90)	0.73 (0.59, 0.91)	0.74 (0.59, 0.92)	0.74 (0.60, 0.93)	0.75 (0.60, 0.93)	0.75 (0.61, 0.94)	0.76 (0.61, 0.95)	0.76 (0.62, 0.95)	0.77 (0.62, 0.96)	0.78 (0.63, 0.97)	0.78 (0.63, 0.97)
	0.28	0.72 (0.58, 0.90)	0.73 (0.58, 0.90)	0.73 (0.59, 0.91)	0.74 (0.59, 0.92)	0.74 (0.60, 0.93)	0.75 (0.60, 0.93)	0.75 (0.61, 0.94)	0.76 (0.61, 0.95)	0.77 (0.62, 0.95)	0.77 (0.62, 0.96)	0.78 (0.62, 0.97)
	0.30	0.72 (0.58, 0.89)	0.72 (0.59, 0.89)	0.73 (0.59, 0.91)	0.74 (0.59, 0.91)	0.74 (0.59, 0.92)	0.74 (0.60, 0.93)	0.75 (0.60, 0.93)	0.75 (0.61, 0.94)	0.76 (0.61, 0.95)	0.77 (0.62, 0.96)	0.77 (0.62, 0.96)

Note: Confidence intervals including one are considered representative of a nonsignificant treatment effect. Shading represents intervals that include one. This is based on the calculated limit of the confidence interval rather than the figure shown in this table which is rounded to the nearest one-hundredth.

Table 33. P-values Associated with the Test of Whether the Risk Ratio Differs from One FF/VI 100/25 versus FF 100 under Varying Assumptions about the Rate of Exacerbations on Each Treatment Arm in Patients who Withdrew from the Study Early During the Time for which they Should Have Been Observed but Were Not

		FF/VI 100/25 Assumed Exacerbation Rate Post-Withdrawal										
		0.10	**0.12**	**0.14**	**0.16**	**0.18**	**0.20**	**0.22**	**0.24**	**0.26**	**0.28**	**0.30**
FF 100 Assumed Exacerbation Rate Post-Withdrawal	**0.10**	0.015	0.018	0.025	0.030	0.035	0.040	0.049	0.055	0.064	0.065	0.088
	0.12	0.013	0.018	0.022	0.027	0.031	0.036	0.043	0.049	0.057	0.068	0.079
	0.14	0.009	0.014	0.019	0.023	0.027	0.029	0.038	0.043	0.050	0.059	0.070
	0.16	0.011	0.011	0.016	0.020	0.023	0.024	0.032	0.037	0.043	0.052	0.061
	0.18	0.009	0.011	0.014	0.017	0.020	0.023	0.028	0.032	0.038	0.045	0.053
	0.20	0.008	0.010	0.012	0.014	0.017	0.020	0.024	0.027	0.030	0.039	0.046
	0.22	0.006	0.008	0.009	0.012	0.014	0.016	0.018	0.023	0.027	0.033	0.039
	0.24	0.005	0.006	0.008	0.009	0.010	0.012	0.016	0.019	0.022	0.027	0.033
	0.26	0.004	0.005	0.006	0.008	0.009	0.011	0.014	0.016	0.019	0.024	0.028
	0.28	0.003	0.004	0.005	0.007	0.008	0.009	0.012	0.014	0.017	0.020	0.025
	0.30	0.003	0.003	0.005	0.006	0.007	0.008	0.010	0.012	0.014	0.018	0.022

5.1.2 Subgroup Analysis

See the discussion under section 4.1.

5.2 Collective Evidence

As summarized in Table 34, effectiveness and safety of two different dosages were examined: FF/VI 100/25 mcg and FF/VI 200/25 mcg. The review focused on four lung function studies (12/24 weeks in duration) and one long term exacerbation study (24-76 weeks in duration).

Table 34. Summary of the Key Efficacy Test Results

Effect	Treatment Groups		Study	Lung Function Parameters		Exacerbation Rate	
	Study Drug	Control		Trough FEV1	WM FEV1	Time to Event	Incidence Rate
FF/VI	FF/VI 100/25	Placebo	27	X	X		
		FP/SALM 250/50 BID	91	NW	NW		
Contribution of VI	FF/VI 100/25	FF 100	63	X	X	X	X
			37	X			
			27	NB	NB		
	FF/VI 200/25	FF 200	29	X	X		
FF 200/25 over 100/25	FF 100	Placebo	27	X	X		
	FF/VI 200/25	FF/VI 100/25	63	X	X		

Source: Reviewer.
Note: X: Statisticially significiant; NB: Numerically better; NW: Numerically worse.

5.3 Conclusions and Recommendations

Direct evaluation of the contribution of FF to the combination thru a comparison of FF/VI to VI alone was not possible since administration of a LABA, such as VI, without co-administration of an inhaled corticosteroid, such as FF, is generally considered unacceptable in treating asthma. The effect of FF as a monocomponent was examined in the previous approving process of Arnuity Ellipta by comparing FF to placebo. It is not known whether the effect of FF over placebo will be consistent regardless of the presence or absence of VI.

The contribution of VI to the overall effectiveness of the combination for all major endpoints was directly examined. In support of VI's contribution, after 12 or 24 weeks of treatment, patients assigned to receive FF/VI 100/25 or FF/VI 200/25 consistently showed statistically greater improvement in weighted mean FEV_1 and trough FEV_1 than patients assigned to receive FF only. Also, in the 24 – 76 week long term exacerbation study, patients assigned to FF/VI 100 consistently showed statistically significant improvement in terms of time to first asthma exacerbation and incidence rate of exacerbation than did patients assigned to receive FF 100.

Subgroup analyses were conducted to investigate the level of consistency or heterogeneity of the treatment effect across subgroups of interest, especially the age subgroup, due to the pre-existing concerns regarding asthma-related serious adverse events associated with LABAs. Results of the lung function studies using the WM FEV1 and trough FEV1 endpoints within the adolescents are numerically but not statistically significantly unfavorable and the treatment-by-age interaction tests are not significant so that there is no evidence of a differing treatment effect between age groups. Results for the exacerbation endpoint send out a stronger signal of a differing treatment effect across age groups with the estimate of the treatment-by-age interaction in the analysis of the rate of exacerbations being borderline significant (p=0.055); however, this may be indicative of a qualitative (i.e., a reversal of the treatment effect from adults to adolescents) or only quantitative (i.e., only the magnitude of the difference between treatment groups varies across age groups but the benefit of the combination is positive in both subgroups) interaction. In addition, we acknowledge that the challenge in interpreting subgroup findings is whether to differences in treatment effect across subgroups to true heterogeneity or to random chance.

This submission supports effectiveness of FF/VI 100/25 and FF/VI 200/25 for once daily treatment of asthma in adult population (age 18 years and older). For the adolescent patient group, however, the contribution of VI to the effectiveness of the combination is questionable, particularly for the exacerbation endpoint. A large well-powered dedicated adolescent asthma exacerbation study would be necessary to fully understand the contribution of VI to the combination in this patient population.

6 References

Fanta, C. H. (2009). Asthma. *The New England Journal of Medicine*, 1002-14.

National Heart, Lung, and Blood Institute. (2007). *Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma.*

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

/s/

YU WANG
04/14/2015

RUTHANNA C DAVI
04/14/2015