US 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
New Drug Application
Clinical Studies

NDA:	205-625
Drug Name:	Fluticasone Furoate Inhalation Powder
Indication:	Maintenance treatment of asthma as prophylactic therapy in patients 12 years of age and older
Applicant:	GlaxoSmithKline
Dates:	Received October 22, 2013 PDUFA due date: August 22, 2014
Review Priority:	Standard
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	Gregory Levin, PhD
Concurring Reviewers:	David Petullo, MS Thomas Permutt, 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:	Nina Ton
Keywords:	NDA review, clinical studies, missing data, surrogate outcomes, labeling

Reference ID: 3595633

Contents

1 Executive Summary **6**

2 Introduction **7**

2.1 Overview . . . 7

2.1.1 Background . . . 7

2.1.2 History of Drug Development . . . 7

2.1.3 Specific Studies Reviewed . . . 8

2.2 Data Sources . . . 9

3 Statistical Evaluation **10**

3.1 Data and Analysis Quality . . . 10

3.2 Study Design . . . 10

3.2.1 Study 27 . . . 10

3.2.2 Study 59 . . . 11

3.2.3 Additional Phase 3 Studies . . . 12

3.2.4 Phase 2 Studies . . . 13

3.3 Statistical Methodologies . . . 14

3.3.1 Primary and Secondary Analyses . . . 14

3.3.2 Evaluating the Potential Effect of Missing Data . . . 16

3.4 Evaluation of Efficacy . . . 19

3.4.1 Dose Selection . . . 19

3.4.2 Patient Disposition, Demographic, and Baseline Characteristics . . . 19

3.4.3 Results in Studies 27 and 59 . . . 25

3.4.4 Potential Effect of Missing Data . . . 32

3.4.5 Results in Additional Studies . . . 37

3.5 Evaluation of Safety . . . 41

4 Findings in Special/Subgroup Populations **41**

5 Summary and Conclusions **44**

5.1 Statistical Issues . . . 44

5.2 Collective Evidence . . . 47

5.3 Labeling Recommendations . . . 47

Appendix **49**

References **57**

Reference ID: 3595633

List of Tables

1 Overview of Key Phase 3 Studies 9
2 Baseline Characteristics in Study 27 21
3 Baseline Characteristics in Study 59 22
4 Patient Dropout, by Reason for Withdrawal, in Study 27 24
5 Patient Dropout, by Reason for Withdrawal, in Study 59 24
6 Analyses of the Co-Primary Endpoint Trough FEV_1 in Study 27: Differences in Mean Changes from Baseline to the Last Available Visit up to Week 12 27
7 Analyses of the Co-Primary Endpoint 0–24 Hour Postdose Weighted Mean FEV_1 in Study 27: Differences in Mean Changes from Baseline to Week 12 27
8 Analyses of the Primary Endpoint Trough FEV_1 in Study 59: Differences in Mean Changes from Baseline to the Last Available Visit up to Week 24 28
9 Analyses of Additional Supportive Endpoints in Studies 27 and 59: Differences in Mean Changes from Baseline During Adherence to Treatment 31
10 Exploring the Potential Effect of Missing Data: Results for the Primary Endpoint Trough FEV_1 with the Primary Analysis as Compared to a Multiple Imputation Supportive Analysis in Studies 27 and 59 34
11 Tipping Point Analysis in Study 27: Estimated Differences Between FF and Placebo in the Mean Change from Baseline to 12 Weeks in Trough FEV_1 (L) under Different Imputed Values for the Missing Mean Changes in Patients who Withdrew from the Study Early on Each Treatment Arm 35
12 Tipping Point Analysis in Study 59: Estimated Differences Between FF and Placebo in the Mean Change from Baseline to 24 Weeks in Trough FEV_1 (L) under Different Imputed Values for the Missing Mean Changes in Patients who Withdrew from the Study Early on Each Treatment Arm 36
13 Differences Between FF and Placebo in the Probability of Achieving Certain Threshold Changes in Trough FEV_1 in Studies 27 and 59 36
14 Analyses of Trough FEV_1 and Additional Endpoints in Study 29: Differences in Mean Changes from Baseline During Adherence to Treatment 38
15 Analyses of Trough FEV_1 and Additional Endpoints in Study 96: Differences in Mean Changes from Baseline During Adherence to Treatment 39
16 Baseline Characteristics in Study 29 49
17 Baseline Characteristics in Study 96 50
18 Patient Dropout, by Reason for Withdrawal, in Study 29 52
19 Patient Dropout, by Reason for Withdrawal, in Study 96 52
20 Baseline Characteristics in Study 27, Stratified According to Patient Withdrawal 55
21 Baseline Characteristics in Study 59, Stratified According to Patient Withdrawal 56

List of Figures

1 Applicant's Strategy to Control Type I Error Across Multiple Comparisons in Study 27 (Source: Applicant's Study Report) 15
2 Patient Withdrawal over Time in Study 27 (Source: Reviewer) 23
3 Patient Withdrawal over Time in Study 59 (Source: Reviewer) 23
4 Mean Change from Baseline in Trough FEV_1 over Time in Study 27 Based on Observed Data. Error Bars Represent Plus or Minus One Standard Error. (Source: Reviewer) 29
5 Mean Change from Baseline in Trough FEV_1 over Time in Study 59 Based on Observed Data. Error Bars Represent Plus or Minus One Standard Error. (Source: Reviewer) 29
6 Empirical Distribution Function for Change from Baseline in Trough FEV_1 at 12 Weeks in Study 27 (Source: Reviewer) 30
7 Empirical Distribution Function for Change from Baseline in Trough FEV_1 at 24 Weeks in Study 59 (Source: Reviewer) 30
8 Trough FEV_1 over Time in Dropouts Prior to Withdrawal in Study 27 (Source: Reviewer) 34
9 Trough FEV_1 over Time in Dropouts Prior to Withdrawal in Study 59 (Source: Reviewer) 35
10 Estimated Difference Between FF 200 and FF 100, Stratified by Selected Subgroups, in Study 96. Solid Vertical Line Represents Estimated Treatment Effect in Overall Population, and Dashed Vertical Line Represents No Difference. (Source: Reviewer) 40
11 Estimated Treatment Effect of FF, Stratified by Selected Subgroups, in Study 27. Solid Vertical Line Represents Estimated Treatment Effect in Overall Population, and Dashed Vertical Line Represents No Difference. (Source: Reviewer) 42
12 Estimated Treatment Effect of FF, Stratified by Selected Subgroups, in Study 59. Solid Vertical Line Represents Estimated Treatment Effect in Overall Population, and Dashed Vertical Line Represents No Difference. (Source: Reviewer) 43
13 Patient Withdrawal over Time in Study 29 (Source: Reviewer) 51
14 Patient Withdrawal over Time in Study 96 (Source: Reviewer) 51
15 Mean Change from Baseline in Trough FEV_1 over Time in Study 29 Based on Observed Data. Error Bars Represent Plus or Minus One Standard Error. (Source: Reviewer) 53
16 Mean Change from Baseline in Trough FEV_1 over Time in Study 96 Based on Observed Data. Error Bars Represent Plus or Minus One Standard Error. (Source: Reviewer) 53

17 Empirical Distribution Function for Change from Baseline in Trough FEV_1 at 24 Weeks in Study 29 (Source: Reviewer) 54
18 Empirical Distribution Function for Change from Baseline in Trough FEV_1 at 24 Weeks in Study 96 (Source: Reviewer) 54

1 Executive Summary

This review considers the once-daily inhaled corticosteroid (ICS) fluticasone furoate (FF) for maintenance treatment of asthma as prophylactic therapy in patients 12 years of age and older. We primarily focus on two phase 3, randomized, double-blind, parallel-group, placebo-controlled trials that evaluated the efficacy of FF with respect to pulmonary function at 12 or 24 weeks. Patients in these studies had persistent asthma and had been using an ICS, with or without a long-acting beta$_2$-adrenergic agonist (LABA). Concomitant LABA therapy was prohibited during the studies and patients were provided as-needed salbutamol/albuterol for rescue treatment of asthma symptoms.

There was statistical evidence of benefit for FF 100 mcg with respect to the primary endpoints in phase 3 Studies HZA106827 and FFA112059. The estimated treatment effects on trough FEV_1, which should be interpreted as differences from placebo in mean changes from baseline to the last visit during adherence, were 0.14 L (95% confidence interval [CI]: 0.05, 0.22) and 0.15 L (95% CI: 0.04, 0.26), respectively. In an additional randomized, double-blind, parallel-group clinical trial (Study FFA114496) comparing FF 100 and 200 mcg, there were trends toward slightly greater FEV_1 improvement with the higher 200 mcg dose (estimated difference: 0.08 L, 95% CI: -0.04, 0.19 L).

We consider FEV_1 to be a surrogate endpoint because it does not directly measure how a patient functions or feels in daily life, or how long a patient survives. The claim of effectiveness based on the primary analyses thus relies on the conclusion that the treatment effect on FEV_1 will reliably predict a treatment effect on a clinically meaningful endpoint. Therefore, we also gave importance to analyses of the following secondary endpoints that might be considered to directly measure how patients function or feel: percent rescue-free 24-hour periods, percent symptom-free 24-hour periods, Asthma Quality of Life Questionnaire for 12 years and older (AQLQ [+12]) score, and Asthma Control Test (ACT) score. The observed trends toward benefit for these endpoints increased confidence that the treatment effect on FEV_1 is likely to predict improvements in how asthma patients function, feel, or survive.

There were substantial missing data at the end of the study in the two placebo-controlled phase 3 clinical trials, with overall dropout rates of 15% and 26%. The last available observation estimand evaluated in the primary analyses may not be meaningful for all patients because it assigns positive outcomes to patients who showed an early FEV_1 improvement but could not tolerate or adhere to the therapy long term. Therefore, we gave importance to supportive analyses evaluating alternative estimands. These analyses generally supported the effectiveness of FF. However, estimates of the treatment effect on the mean change in trough FEV_1 at the end of the study, regardless of adherence to assigned therapy, were approximately 20–30% smaller

than estimates from the primary analyses.

2 Introduction

2.1 Overview

2.1.1 Background

Asthma is a common chronic lung disease that causes airway inflammation and narrowing. Symptoms may include wheezing, chest tightness, shortness of breath, and coughing. Asthma occurs in people of all ages, often beginning during childhood. The diagnosis of asthma is typically based on symptom patterns and lung function testing (by spirometry).

Treatment options include but are not limited to inhaled short-acting beta$_2$-adrenergic agonists (SABAs) for relief of acute symptoms, as well as inhaled corticosteroids (ICSs) and long-acting beta$_2$-adrenergic agonists (LABAs) for long-term maintenance therapy to help prevent symptoms. Inhaled corticosteroids aim to control asthma symptoms and decrease the frequency and severity of asthma exacerbations by reducing airway inflammation. Most available ICSs, including fluticasone propionate (FP), beclomethasone, and budesonide, are administered twice daily. The availability of a safe and effective once-daily (OD) ICS might help improve adherence and lead to better health outcomes.

This review considers the evaluation of the once-daily inhaled corticosteroid fluticasone furoate (FF) for maintenance treatment of asthma as prophylactic therapy in patients 12 years of age and older. FF is administered from a Dry Powder Inhaler (DPI), and two strengths are proposed for the treatment of asthma: 100 and 200 mcg.

2.1.2 History of Drug Development

The applicant has submitted results from the following five phase 3 clinical trials to support the the regulatory approval of FF for maintenance treatment of asthma as prophylactic therapy in patients 12 years of age and older: Studies HZA106827, FFA112059, HZA106829, FFA114496, and HZA106837 (which we will refer to by the last two numbers). The clinical development program for FF was introduced to the Division of Pulmonary, Allergy, and Rheumatology Products under IND 70,297. A related GlaxoSmithKline (GSK) combination product, Breo Ellipta, consisting of FF and the LABA vilanterol (VI), 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). FF is also being developed by

Reference ID: 3595633

GSK as the ICS component in combination products with VI for asthma, with the anticholinergic umeclidinium bromide (UMEC) for asthma, and with both UMEC and VI for COPD.

Important meetings and correspondence with the applicant during drug development that might be relevant to this review are summarized below. An end-of-phase 2 meeting to discuss the development of FF was held on March 16, 2011. The Division considered the design, duration, and endpoints of Study 59 to be appropriate to evaluate the efficacy of FF 100 mcg in mild to moderate persistent asthma patients. To justify the need for both the 100 mcg and 200 mcg doses, the Division noted that a clinically relevant numerical separation in the primary outcome measure of Study 96, with supportive data from other endpoints, would be expected.

A pre-NDA meeting was held on February 11, 2013. The Division indicated that the degree of support for the added benefit of FF 200 over FF 100 would be a review issue. The sponsor noted that FF 100 had been evaluated in asthma patients receiving only SABAs in a single study, the 8-week dose-ranging Study FFA109687 (with phase 3 studies of FF 50 also in this population). The Division stated that proceeding with NDA submission, despite no phase 3 studies in the SABA-only population, was reasonable, and that the acceptability of the phase 2 data for labeling would be a review issue.

FDA submitted an information request as part of the 74-day letter regarding the potential effect of missing data on the reliability of efficacy results. FDA requested additional sensitivity analyses that did not rely on the assumption that observed treatment effects before withdrawal would be preserved after patients stopped taking the therapy. The applicant responded with results based on additional sensitivity analyses (see 3.3 for more details).

2.1.3 Specific Studies Reviewed

Our evaluation of the effectiveness of FF 100 focuses on the placebo-controlled Studies 27 and 59. Study 27 was a 12-week, randomized, double-blind, parallel-group, placebo-controlled clinical trial of FF 100 and FF/VI 100/25. Study 59 was a 24-week, randomized, double-blind, parallel-group, placebo- and active-controlled trial of FF 100, with FP 250 BD as an active control. We also discuss results from Studies 29 and 96. Study 29 was a 24-week, randomized, double-blind, parallel-group, active-controlled trial of FF 200 and FF/VI 200/25, with FP 500 BD as an active control. Study 96 was a 24-week, randomized, double-blind, parallel-group trial of FF 100 and 200, with no control group. We do not discuss results from Study 37 – a randomized, double-blind, parallel-group trial to compare FF 100 and FF/VI 100/25 with respect to the risk of severe asthma exacerbations, with no control group – because it only allows an evaluation of the contribution of vilanterol to the FF/VI combination. A summary of the four phase 3 studies that are the focus of this review is provided in Table 1.

Finally, we briefly comment on several phase 2 studies used to support the dose selection of fluticasone furoate. Studies FFA20001 and FFA106783 evaluated morning versus evening dosing, Study FFA112202 evaluated dosing frequency, and Studies FFA109684, FFA109685, and FFA109687 compared different once-daily FF doses. Results are also available from two additional placebo-controlled phase 3 studies that failed to demonstrate the efficacy of the lower 50 mcg dose of FF.

Direct within-study comparisons of the safety and effectiveness of FF 100 and 200 are possible from phase 3 Study 96, in addition to the 8-week phase 2 dose-ranging Studies FFA109685 and FFA109687.

Table 1: Overview of Key Phase 3 Studies

Study	Design	Treatment Arms	Number Subjects	Date[†]
HZA106827	12-week, randomized, double-blind, parallel-group, placebo-controlled	FF 100 OD FF/VI 100/25 OD Placebo	205 201 203	08/2010 – 10/2011
FFA112059	24-week, randomized, double-blind, parallel-group, placebo-controlled	FF 100 OD FP 250 BD Placebo	114 114 115	06/2010 – 01/2012
HZA106829	24-week, randomized, double-blind, parallel-group, active-controlled	FF 200 OD FP 500 BD FF/VI 200/25 OD	194 195 197	06/2010 – 10/2011
FFA114496	24-week, randomized, double-blind, parallel-group	FF 100 OD FF 200 OD	119 119	09/2011 – 10/2012

Source: Reviewer

[†] Dates correspond to the start and end of the study.

Abbreviations: FF = fluticasone furoate; FP = fluticasone propionate; VI = vilanterol; OD = once-daily; BD = twice-daily

2.2 Data Sources

Data were submitted by the applicant to the CDER electronic data room in SAS transport format. Protocols, correspondence, data listings, and study reports were accessed under the network path \\CDSESUB1\EVSPROD\NDA205625\205625.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. During audits, the applicant identified one site in each of Studies 59, 29, and 96 with Good Clinical Practices (GCP) issues. The applicant also expressed concerns about FEV_1 data quality for one site in Study 96. Results were similar when excluding each of these four sites.

3.2 Study Design

3.2.1 Study 27

Study 27 was a 12-week, randomized, double-blind, parallel-group, placebo-controlled clinical trial of FF 100 and FF/VI 100/25. The study consisted of patients at least 12 years of age who had a diagnosis of asthma for at least 12 weeks and had been using an ICS, with or without a LABA, for at least 12 weeks (with a stable ICS dose for at least 4 weeks) prior to screening. Patients had a best pre-bronchodilator FEV_1 of 40–90% of predicted, and had to demonstrate at least a 12% and 200 mL reversibility of FEV_1 within 10–40 minutes following 2–4 inhalations of salbutamol/albuterol. Patients were randomized 1:1:1 to FF/VI 100/25, FF 100, or placebo, and randomization was stratified by current asthma medication (ICS or ICS/LABA). All treatments were administered by oral inhalation once daily in the evening. There was a 4-week run-in period to establish eligibility, assess compliance with the Daily Diary and medication, and measure baseline characteristics, followed by a 12-week double-blind treatment period (with clinic visits at Weeks 2, 4, 8, and 12). Patients had to maintain a stable ICS dose through the run-in period and then discontinue ICS use after the morning dose (or the prior evening dose if taken once-daily in the evening) on the day of the baseline visit. Concomitant LABA therapy was prohibited during the run-in and double-blind periods. Patients were provided as-needed salbutamol/albuterol as rescue medication for treatment of asthma symptoms.

Withdrawal *from the treatment* was equivalent to withdrawal *from the study* because patients who stopped taking the therapy early were not followed up for safety and efficacy assessment for the remainder of the 12-week treatment period. Possible protocol-specified reasons for withdrawal included but were not limited to adverse event, loss to follow-up, protocol violation, lack of efficacy, non-compliance, and abnormal laboratory results. If possible, an early withdrawal visit was conducted within 24 hours of the patient stopping medication. The many potential reasons for stopping treatment, combined with the fact that the applicant did not continue to

Reference ID: 3595633

collect information on patients who stopped therapy early, led to substantial missing efficacy and safety data (see 5.1 for further discussion).

The co-primary efficacy endpoints were the mean change from baseline in trough (pre-dose) FEV_1 at Week 12 and the mean change from baseline in the weighted mean serial FEV_1 over the 24 hours after dosing at Week 12. The term co-primary meant that statistical significance needed to be achieved on both endpoints for the trial to provide support for the efficacy of FF. FEV_1 was measured by spirometry in the evening. The calculation of 0–24 hour weighted mean serial FEV_1 was based on assessments at 5, 15, and 30 minutes, and 1, 2, 3, 4, 5, 12, 16, 20, 23, and 24 hours after dosing. Secondary efficacy endpoints included the mean change from baseline in the percentage of rescue-free 24-hour periods during the 12-week treatment period, the mean change from baseline in the percentage of symptom-free 24-hour periods during the 12-week treatment period, the mean change from baseline in the total Asthma Quality of Life Questionnaire for 12 years and older [AQLQ (+12)] score at Week 12, and the number of withdrawals due to lack of efficacy during the 12-week treatment period. Several additional exploratory endpoints were also evaluated. The assessment of the extent of asthma symptoms and of the number of inhalations of rescue salbutamol/albuterol in a 24-hour period were based on day-time and night-time reports from patients in an electronic diary. A sample size of 570 patients was planned to provide approximately 83% power across all primary comparisons, assuming a standard deviation (SD) of 405 mL and 5% missing data for trough FEV_1, and a SD of 325 mL and 15% missing data for weighted mean FEV_1.

To preserve the blind against FF/VI 100/25, FF was administered using a double-strip inhaler, where the first strip contained FF 100 mcg blended with lactose, and the second strip contained (b) (4) The double-strip FF inhaler is different than the single-strip inhaler that was used in Study 59 and is proposed for approval. There were some differences between the single- and double-strip inhalers in the quantity of delivered fine particle mass. Discussion of this issue can be found in the Medical Review by Dr. Tracy Kruzick.

3.2.2 Study 59

Study 59 was a 24-week, randomized, double-blind, double-dummy, parallel-group, placebo- and active-controlled trial of FF 100, with FP 250 BD as an active control. Many aspects of the population and design were similar to those of Study 27. A protocol amendment effective January 28, 2011 relaxed the entry criteria (largely based on FEV_1 and past ICS use). After the amendment, the study enrolled asthma patients at least 12 years of age on a stable ICS dose who had 40–90% predicted FEV_1 and at least a 12% and 200 mL FEV_1 reversibility. Patients were randomized 1:1:1 to FP 250 BD, FF 100, or placebo administered by oral inhalation once daily in the evening. There was a 4-week run-in period, followed by a 24-week double-blind

treatment period, with visits at Weeks 2, 4, 8, 12, 16, 20, and 24. Patients had to maintain a stable ICS dose through the run-in period and then discontinue ICS use after the morning dose (or the prior evening dose if taken once-daily in the evening) on the day of the baseline visit. Concomitant LABA therapy was prohibited during the run-in and double-blind periods, and patients were provided as-needed salbutamol/albuterol for recue medication. As in Study 27, withdrawal from the treatment was equivalent to withdrawal from the study, and there were many possible reasons for discontinuation of study therapy. This led to substantial missing data in efficacy and safety analyses.

The primary efficacy endpoint was the mean change from baseline in trough FEV_1 at Week 24. Secondary efficacy endpoints included the mean change from baseline in the percentage of rescue-free 24-hour periods during the 24-week treatment period, the mean changes from baseline in daily trough morning and evening peak expiratory flow (PEF) averaged over the 24-week treatment period, the mean change from baseline in the percentage of symptom-free 24-hour periods during the 24-week treatment period, and the mean change from baseline in total AQLQ (+12) score at Week 24. Several additional exploratory endpoints were also evaluated. A sample size of 330 patients was planned to provide 94% power to detect a difference of 200 mL between FF and placebo in trough FEV_1, assuming a standard deviation of 405 mL and 5% missing data caused by withdrawal within the first 2 weeks.

It is important to note that Studies 27 and 59 consisted of asthma patients who had been on a stable dose of an inhaled corticosteroid product. Patients continued receiving that ICS until the day before or morning of initiation of the randomized treatment in these studies. Therefore, there is no phase 3 data to support the efficacy of FF in patients naive to ICS asthma treatment. Furthermore, these studies consisted of a randomized withdrawal-like design, although the randomized, experimental ICS product was different than the run-in ICS product. The only study that evaluated the efficacy of FF in asthma patients who had not been receiving ICS treatment was the phase 2 Study FFA109687.

3.2.3 Additional Phase 3 Studies

We evaluated two additional phase 3 studies: Studies 29 and 96. Study 29 was a 24-week, randomized, double-blind, double-dummy, parallel-group, active-controlled trial of FF 200 and FF/VI 200/25, with FP 500 BD as an active control. The study consisted of patients at least 12 years of age who had a diagnosis of asthma and had been using an ICS, with or without LABA, for at least 12 weeks. Patients had to be on a stable high ICS dose (FP 500 BD or equivalent) or a stable mid-dose ICS/LABA combination (SERETIDE/ADVAIR 250/50 BD or equivalent) for at least 4 weeks prior to the run-in. Patients were randomized 1:1:1 to FF/VI 200/25, FF 200, or FP 500 BD. Randomization was stratified by current medication (ICS or ICS/LABA).

There was a 4-week run-in period (during which the stable ICS dose was maintained), followed by a 24-week double-blind treatment period. Concomitant LABA therapy was prohibited, and patients were provided as-needed salbutamol/albuterol as rescue medication. There were many possible reasons for patients to stop treatment early, and these patients were withdrawn from the study. The co-primary endpoints were the mean change from baseline in trough FEV_1 at Week 24 and the mean change in weighted mean serial FEV_1 over 0–24 hours post-dose at Week 24 (calculated in a subset of patients). Secondary endpoints included the mean changes from baseline in the percentage of rescue-free 24-hour periods, percentage of symptom-free 24-hour periods, and total AQLQ (12+) score.

Study 96 was a 24-week, randomized, double-blind, parallel-group trial of FF 100 and 200 mcg once daily, with no control group. The study consisted of patients at least 12 years of age who had a diagnosis of asthma for at least 12 weeks and had been on a stable mid-to-high ICS dose for at least 4 weeks. Patients were stratified by baseline FEV_1 (40–65% versus >65% predicted) and randomized 1:1 to FF 100 or 200. There was a 4-week run-in period (during which the stable ICS dose was maintained), followed by a 24-week double-blind treatment period. Concomitant LABA therapy was prohibited, and patients were provided as-needed salbutamol/albuterol as rescue during the study. There were many possible reasons for patients to stop treatment early, and these patients were withdrawn from the study. The primary endpoint was the mean change from baseline in trough FEV_1 at Week 24. Secondary endpoints included the mean changes from baseline in the percentage of rescue-free 24-hour periods, percentage of symptom-free 24-hour periods, and daily evening and morning peak expiratory flows. The sample size of approximately 220 patients was chosen to ensure that the half-width of the 95% confidence interval for the mean difference between doses in FEV_1 was no greater than 110.1 mL.

3.2.4 Phase 2 Studies

The following phase 2 clinical trials were used to support the dose selection of FF: Studies FFA20001, FFA106783, FFA112202, FFA109684, FFA109685, and FFA109687. Study FFA20001 was a 28-day, randomized, double-blind, double-dummy, parallel-group, placebo-controlled trial to evaluate morning versus evening dosing. Patients were randomized to placebo, FF 100 in the morning (AM), FF 100 in the evening (PM), or FF 250 PM. Study FFA106783 was an 8-week, randomized, double-blind, parallel-group, placebo-controlled trial to evaluate the timing and frequency of dosing. Patients were randomized to placebo, FF 200 AM, FF 200 PM, FF 400 AM, FF 400 PM, or FF 200 BD. Study FFA112202 was a randomized, double-blind, placebo- and active-controlled, cross-over trial to evaluate the frequency of dosing (with once-daily doses in the evening). Patients were randomized to 3 28-day treatment periods (with washout) of placebo, FF 200 OD, FF 100 BD, FP 200 OD, and/or FP 100 BD. Studies FFA109684, FFA109685, and

FFA109687 were 8-week, randomized, double-blind, double-dummy, parallel-group, placebo- and active-controlled trials to compare once-daily doses of FF. In Study FFA109684, patients with uncontrolled asthma despite treatment with medium-dose ICS were randomized to placebo, FF 200, FF 400, FF 600, FF 800, or FP 500 BD. In Study FFA109685, patients with uncontrolled asthma despite treatment with low-dose ICS were randomized to placebo, FF 100, FF 200, FF 300, FF 400, or FP 250 BD. In Study FFA109687, patients with uncontrolled asthma despite treatment with SABAs or non-ICS therapy were randomized to placebo, FF 25, FF 50, FF 100, FF 200, or FP 100 BD. Efficacy endpoints in these phase 2 studies typically included mean changes from baseline in FEV_1 and/or peak expiratory flow.

3.3 Statistical Methodologies

3.3.1 Primary and Secondary Analyses

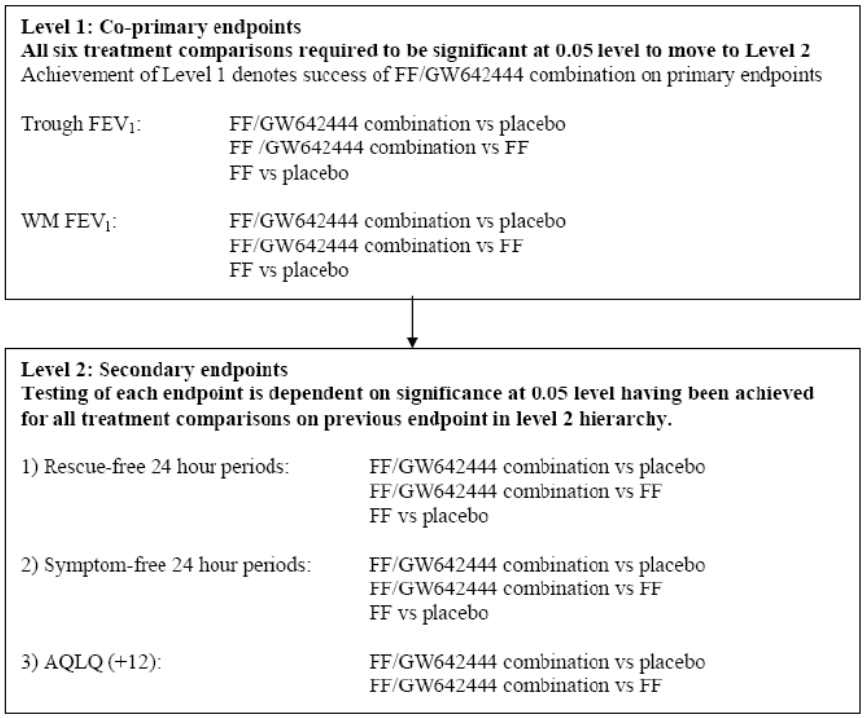
In Study 27, the primary analysis population consisted of all randomized subjects receiving at least one dose of study medication. The analyses of the co-primary endpoints trough FEV_1 and weighted mean FEV_1 were based on linear regression models (analyses of covariance [ANCOVA]) adjusting for baseline (pre-dose measurement on Day 0) FEV_1, region, sex, and age. In the primary analysis of trough FEV_1, last observation carried forward (LOCF) was used to impute missing measurements in patients who withdrew from the study. The primary analysis of weighted mean FEV_1 was restricted to patients who completed the study, with no imputation of missing data.

A mixed effects model for repeated measures (MMRM) based on the repeated measurements within patients over time was carried out as a supportive analysis (with visit as a categorical variable and a treatment-by-visit interaction, adjusting for the same baseline factors). Analyses of the percentage of rescue-free 24-hour periods, percentage of symptom-free 24-hour periods, AQLQ (+12) total score, and ACT score were based on analogous linear regression models to that of the primary analysis. There was no imputation of missing secondary endpoint data after patient dropout – the analyses evaluated differences between treatment groups only during the time period that patients remained in the study. At least two non-missing 24-hour period assessments were required to calculate the percentage of rescue- or symptom-free 24-hour periods within a patient, and missing 24-hour period assessments were not imputed. Percentages were calculated for each patient over the combined 12-week treatment and 2-week follow-up periods.

The applicant prespecified a sequential testing approach to control the one-sided 2.5% type I error rate across the multiple comparisons (Figure 1). The different pairwise comparisons of the three treatment arms with respect to the two co-primary endpoints were characterized as level 1 tests, and the pairwise comparisons with respect to three important secondary endpoints were

characterized as level 2 tests. The protocol indicated that all six comparisons in level 1 needed to be statistically significant in order for inference to be carried out on level 2 comparisons. Within the level 2 comparisons, if a given test was not statistically significant, then all tests below it in the hierarchy were considered descriptive. However, the protocol did not explicitly indicate how the type I error rate would be controlled across the six level 1 comparisons. This is relevant because this application relies on the two FF versus placebo comparisons, when the other four comparisons were not all statistically significant (see 5.1 for further discussion).

Figure 1: Applicant's Strategy to Control Type I Error Across Multiple Comparisons in Study 27 (Source: Applicant's Study Report)



In Study 59, the primary analysis was carried out in all randomized subjects receiving at least one dose of study medication, and was based on a a linear regression model (ANCOVA) adjusting for baseline FEV_1, region, sex, and age. LOCF was used to impute missing FEV_1 measurements in patients who withdrew from the study. An MMRM-based analysis was carried out as supportive. FF 100 was compared with placebo with respect to efficacy endpoints in the following sequential order to control the type I error rate across the multiple tests: (1) trough FEV_1; (2) rescue-free 24-hour periods; (3) trough PM PEF; (4) AM PEF; (5) symptom-free 24-hour periods; and (6): AQLQ (12+) score. Endpoints based on percentages or averages over 12- and 24-week periods were calculated based on the first 84 and 168 days after randomization, respectively, regardless of when follow-up visits occurred. Analyses of all secondary endpoints except AQLQ (12+) score were based on analogous linear regression models to that of the primary endpoint. With no imputation of missing secondary endpoint data, these analyses evaluated differences between

groups only during the time period that patients remained in the study. MMRM-based analyses were used to evaluate the treatment effects on AQLQ and ACT scores. The applicant conducted sensitivity analyses excluding data from one site (with 19 randomized patients) because of GCP issues identified during an audit.

In Study 29, the primary comparisons of interest were between FF 200 and the FF/VI 200/25 combination product (to evaluate the contribution of vilanterol). A comparison between FF 200 OD and FP 500 BD with respect to change from baseline in trough FEV_1 at Week 24 was prespecified but not included in the multiple testing framework to control the type I error rate. The applicant selected a non-inferiority bound of -125 mL for the comparison of FF with FP, but did not justify this margin. Analyses of primary and secondary endpoints were based on analogous statistical models to Studies 27 and 59. The applicant conducted sensitivity analyses excluding data from one site (with 48 randomized patients) because of GCP issues identified during an audit.

In Study 96, comparisons on FF 100 and 200 with respect to primary and secondary efficacy endpoints were based on similar analyses to those in the other phase 3 studies. The applicant prespecified subgroup analyses of interest according to baseline percent predicted FEV_1 (40–65% or 65–90%) and run-in ICS use (mid- or high-dose). The applicant excluded one site from all analyses because of GCP issues and conducted a sensitivity analysis excluding an additional site because of concerns regarding FEV_1 data quality.

All reported p-values are two-sided, and all reported intervals are 95% confidence intervals.

3.3.2 Evaluating the Potential Effect of Missing Data

The first recommendation in the National Research Council (NRC) report *The Prevention and Treatment of Missing Data in Clinical Trials* states:

> The trial protocol should explicitly define (a) the objective(s) of the trial; (b) the associated primary outcome or outcomes; (c) how, when, and on whom the outcome or outcomes will be measured; and (d) the measures of intervention effects, that is, the causal estimands of primary interest. These measures should be meaningful for all study participants, and estimable with minimal assumptions. Concerning the latter, the protocol should address the potential impact and treatment of missing data.

The protocols and statistical analysis plans for the key phase 3 studies of FF explicitly define (a), (b), and (c), but fail to identify (d). For example, consider the placebo-controlled Study 59. It is clear that the objective of the trial was to evaluate the safety and effectiveness of FF. The primary outcome was the change from baseline in trough FEV_1 at 24 weeks (assessed by spirometry), with the analysis aimed at comparing means in the intent-to-treat population. But the primary

causal estimand of interest was not clearly stated. Using terminology in the literature [1], it may be the *de facto* estimand, i.e., the difference in mean trough FEV_1 at 24 weeks, *regardless of adherence to the assigned treatment.* It may be the *de jure* estimand, i.e., the difference in mean FEV_1 at 24 weeks, *if everyone had adhered to the assigned treatment through 24 weeks.* Or it may be based on the last assessment during adherence, i.e., the difference in mean FEV_1 *until the last time point through 24 weeks at which patients adhere to the assigned treatment.* One could also consider utility estimands, in which patients who discontinue treatment early are assigned some "bad" score on the scale of trough FEV_1. In the absence of an explicitly prespecified, justified, and accepted primary estimand of interest, we must work backward from the primary analysis models. In the process, we consider whether each possible estimand is "meaningful for all study participants, and estimable with minimal assumptions," as recommended in the NRC report.

The primary analyses of trough FEV_1 for the key phase 3 clinical trials were based on linear regression models using the "last observation carried forward" for patients who discontinued assigned treatment early (and therefore also withdrew from the study, as dictated by the protocol). The "carried forward" part of the name implies that the interest might be in an actual value at 24 weeks. However, as noted in the NRC report, the last observed value carried forward is rarely, if ever, a reasonable approximation of the (de facto) value that would have been observed at the end of the study, had it been measured. In Study 59, for example, such an approach would assume that patients who demonstrated early FEV_1 improvement on FF but discontinued the treatment prior to 24 weeks would have maintained that improvement after stopping the therapy. Because inhaled corticosteroids are generally considered symptomatic and not disease-modifying therapies, and their effects on FEV_1 likely do not persist more than a few days after patients stop using them, this assumption is not plausible scientifically.

One might argue that LOCF provides a reasonable estimate of the (de jure) value that would have been observed if patients had hypothetically continued their assigned therapy. However, this is a strong assumption that cannot be verified, and we contend that this estimand is not relevant in the setting of a regulated, adequate and well-controlled, phase 3 trial. The difference in outcome improvement if everyone had adhered is a hypothetical rather than real-world measure of effectiveness, and the fact that everyone did not adhere in the phase 3 trials suggests that the de jure quantity cannot be achieved in real-world clinical practice. One final argument against the use of LOCF to impute a value at the end of the study is that, as a single-imputation approach, it does not take into account the uncertainty in the imputation. The bottom line is that LOCF does not reliably estimate any meaningful end-of-study quantity.

Alternatively, we choose to interpret the primary linear regression analysis with LOCF as an evaluation of the "last available observation" (LAO) estimand, that is, the difference in mean FEV_1 *until the last time point through 24 weeks at which patients adhere to the assigned*

treatment. For this estimand, primary analyses had very little missing data – values were missing only in those patients who did not have any post-baseline spirometry assessments. Therefore, the analysis estimates the LAO quantity with minimal statistical assumptions. However, although this estimand is likely a reasonable measure of drug activity, it may not provide a meaningful measure of effectiveness for all patients. For example, the LAO estimand will assign positive outcomes to patients who show an early FEV_1 improvement on FF but cannot tolerate or adhere to the therapy and therefore drop out prior to 24 weeks. In fact, FF was not effective in these patients because they need *long-term* treatments for their chronic asthma symptoms.

Therefore, an evaluation of the effectiveness of FF should not be based solely on the primary analysis of the LAO estimand. As a result, we first explored how much of the estimated treatment effect in the primary analysis was driven by an FEV_1 increase in dropouts. Second, we carried out supportive analyses aimed at evaluating additional estimands that are meaningful for all patients. The (de facto) difference in mean trough FEV_1 at 24 weeks in all randomized patients, regardless of adherence to the assigned treatment, is one important real-world measure of effectiveness. Because patients were not followed after treatment discontinuation, an evaluation of this estimand must be based on untestable assumptions about the unobserved values at 24 weeks. We found the most merit in the Jump to Reference multiple imputation approach carried out by the applicant – this analysis imputes missing data under the assumption that patients on both active and control arms tended to have outcomes at the end of the study similar to those observed in the completers on the control arm (in particular, the subset of control patients with similar baseline characteristics to the patient whose end-of-study value is being imputed). We note, however, that this analysis likely does not capture the effects of approved, effective ancillary therapies on the real-world end-of-study measure of pulmonary function. More information about this and other multiple imputation models used by the applicant can be found at www.missingdata.org.uk. We also used a simple tipping point analysis to determine how much worse end-of-study outcomes in patients who discontinued early on FF would have had to have been than end-of-study outcomes in dropouts on placebo such that the estimated de facto treatment effect would decrease to zero.

Finally, we considered a number of estimands aimed at the utility of the new treatment. We presented empirical distribution functions and results from responder analyses (using different threshold increases in FEV_1) in which patients who discontinued the assigned treatment prior to the end of the study (12 or 24 weeks) were considered treatment failures.

3.4 Evaluation of Efficacy

3.4.1 Dose Selection

The results of the phase 2 studies FFA20001, FFA106783, and FFA112202 suggested that either morning or evening dosing was reasonable, and that a once-daily dosing frequency was appropriate. The dose-ranging studies FFA109684, FFA109685, and FFA109687 suggested that FF 100 and 200 once-daily may lead to greater improvement in FEV_1 than lower doses. Higher doses did not show consistently greater treatment effects. Therefore, the 100 and 200 once-daily doses were reasonable choices to carry forward to the phase 3 clinical trials.

Two of the phase 2 studies included both the 100 and 200 OD doses of FF that are proposed for approval. In Study FFA109685, mean differences from placebo in Week 8 trough FEV_1 change were 0.21 L (95% CI: 0.10, 0.32) and 0.24 L (95% CI: 0.13, 0.35) for FF 100 and 200, respectively. In Study FFA109687, mean differences from placebo were 0.20 L (95% CI: 0.09, 0.32) and 0.23 L (95% CI: 0.11, 0.35) for FF 100 and 200, respectively. Study FFA109687 was different from the other phase 2 and 3 studies in that it consisted of asthma patients who had not been using inhaled corticosteroids. See 3.4.5 and 5.1 for more discussion on the comparison of the 100 and 200 doses of FF. More details on the results of the phase 2 studies are available in Dr. Tracy Kruzick's Medical Review.

3.4.2 Patient Disposition, Demographic, and Baseline Characteristics

Baseline characteristics were similar in the placebo-controlled Studies 27 and 59 (Tables 2 and 3). There were no large imbalances in baseline characteristics across the treatment arms. In the combined population from the two studies, 82% of patients were White, 58% were female, and the mean age was 40 years (with 13% of patients less than 18 years and 5% greater than 65 years of age). The average FEV_1 at baseline was approximately 2.3 L, and the mean duration of asthma was 14 years. Patient characteristics were largely similar in the two additional key phase 3 studies (Appendix: Tables 16 and 17).

In Study 27, there were 610 subjects enrolled at 64 sites in Germany, Japan, Poland, Romania, Ukraine, and the United States. There were 196 (32%) patients from U.S. sites. In Study 59, there were 343 subjects enrolled at 56 sites in Germany, Poland, Romania, Belgium, and the United States (197 patients; 57%). Study 29 consisted of 586 patients from sites in the Russian Federation, Romania, Germany, Poland, Japan, and the United States (143 patients; 24%). In Study 96, there were 238 subjects enrolled at 27 sites in Argentina, Chile, the Russian Federation, Mexico, France, and the United States. In the intent-to-treat population (219

patients), 55 patients were from U.S. sites (16%). Study 96 had a much greater proportion of Hispanic/Latino patients (59%) than the other studies.

As described previously, the design of the phase 3 studies was such that subjects who stopped treatment early would be also be withdrawn from the study. There were many prespecified reasons for withdrawal, such as adverse event, lack of efficacy, and protocol deviation. As a result, there was substantial patient dropout. The proportions of patients withdrawing over time in Studies 27 and 59 are displayed by treatment group in Figures 2 and 3. In Studies 27 and 59, 15% and 26% failed to complete the double-blind follow-up period, respectively (Tables 4 and 5). Dropout rates were greater on placebo than FF, with the differences primarily attributable to greater placebo dropout because of lack of efficacy. Similar disposition patterns were observed in the subset of sites in Study 27 that measured 0–24 hour weighted mean FEV_1, with 23 (19%) of the 121 patients receiving placebo, and 8 (7%) of the 119 patients receiving FF, withdrawing from the study early.

Overall dropout rates, and the distribution of reasons for withdrawal, were similar on FF and FP in Studies 59 and 29, and between patients receiving FF 100 and FF 200 in Study 96 (Appendix: Figures 13 and 14; Tables 18 and 19).

Table 2: Baseline Characteristics in Study 27

	Placebo	FF 100	FF/VI 100/25	Overall
N	203	205	201	609
Female	111 (55%)	126 (61%)	116 (58%)	353 (58%)
Age (years)	38.1 (16.5)	40.4 (16.8)	40.7 (16.4)	39.7 (16.6)
Age Group (years)				
< 18	33 (16%)	28 (14%)	21 (10%)	82 (13%)
18-65	160 (79%)	161 (79%)	169 (84%)	490 (80%)
≥ 65	10 (5%)	16 (8%)	11 (5%)	37 (6%)
Race				
White	169 (83%)	170 (83%)	172 (86%)	511 (84%)
Black	14 (7%)	16 (8%)	13 (6%)	43 (7%)
Asian	19 (9%)	16 (8%)	16 (8%)	51 (8%)
Other	1 (0%)	3 (1%)	0 (0%)	4 (1%)
Hispanic/Latino	12 (6%)	16 (8%)	9 (4%)	37 (6%)
Weight (kg)	75.3 (18.6)	75.8 (17.9)	75.6 (17.4)	75.6 (17.9)
Height (cm)	167.6 (9.0)	167.5 (9.4)	168.5 (9.1)	167.9 (9.2)
FEV_1	2.3 (0.6)	2.3 (0.6)	2.3 (0.6)	2.3 (0.6)
FEV_1 % Predicted	70.2 (10.1)	70.5 (11.0)	70.6 (11.9)	70.4 (11.0)
Morning PEF	355.5 (112.3)	366.3 (111.9)	361.5 (120.4)	361.1 (114.8)
Evening PEF	367.8 (110.5)	375.2 (113.1)	370.2 (122.7)	371.1 (115.4)
% Rescue-free Days	20.5 (32.0)	21.1 (31.4)	18.9 (30.0)	20.2 (31.1)
% Rescue-free Nights	45.8 (39.5)	55.0 (38.6)	63.6 (37.6)	54.8 (39.2)
% 24-hour Rescue-free Periods	14.5 (29.9)	15.3 (29.3)	13.4 (27.4)	14.4 (28.8)
Daily Rescue Use	3.0 (2.6)	3.0 (2.1)	3.0 (2.2)	3.0 (2.3)
% Symptom-free Days	6.8 (17.5)	9.9 (21.9)	9.3 (20.3)	8.7 (20.0)
% Symptom-free Nights	21.1 (34.5)	18.8 (31.5)	17.3 (30.0)	19.1 (32.0)
% 24-hour Symptom-free Periods	3.5 (12.8)	5.8 (16.5)	5.0 (15.2)	4.8 (14.9)
24-hour Symptom Score	2.7 (1.2)	2.8 (1.4)	2.7 (1.3)	2.7 (1.3)
Duration of Asthma (years)	11.3 (10.2)	13.2 (11.7)	11.8 (12.0)	12.1 (11.4)
At USA site	65 (32%)	71 (35%)	60 (30%)	196 (32%)

Source: Reviewer

Cell contents are mean (standard deviation) or frequency (percent)

Table 3: Baseline Characteristics in Study 59

	Placebo	FF 100	FP 250	Overall
N	115	114	114	343
Female	68 (59%)	63 (55%)	72 (63%)	203 (59%)
Age (years)	40.3 (17.7)	40.1 (16.2)	41.4 (15.6)	40.6 (16.5)
Age Group (years)				
< 18	18 (16%)	17 (15%)	11 (10%)	46 (13%)
18-65	92 (80%)	93 (82%)	98 (86%)	283 (83%)
≥ 65	5 (4%)	4 (4%)	5 (4%)	14 (4%)
Race				
White	88 (77%)	90 (80%)	92 (81%)	270 (79%)
Black	23 (20%)	22 (19%)	19 (17%)	64 (19%)
Asian	2 (2%)	1 (1%)	2 (2%)	5 (1%)
Other	2 (2%)	0 (0%)	0 (0%)	2 (1%)
Hispanic/Latino	4 (3%)	7 (6%)	5 (4%)	16 (5%)
Weight (kg)	80.0 (24.3)	80.0 (24.6)	80.0 (20.8)	80.0 (23.2)
Height (cm)	167.9 (9.2)	168.1 (10.2)	167.9 (10.6)	168.0 (10.0)
FEV_1	2.3 (0.6)	2.4 (0.6)	2.4 (0.7)	2.4 (0.7)
FEV_1 % Predicted	72.3 (10.9)	72.2 (10.4)	73.0 (11.9)	72.5 (11.1)
Morning PEF	347.2 (112.1)	351.6 (113.2)	347.2 (111.8)	348.7 (112.0)
Evening PEF	358.8 (115.4)	370.4 (113.7)	355.3 (109.6)	361.5 (112.8)
% Rescue-free Days	25.6 (34.8)	19.8 (28.1)	23.6 (34.5)	23.0 (32.6)
% Rescue-free Nights	34.8 (37.9)	48.4 (40.6)	53.0 (40.2)	45.4 (40.2)
% 24-hour Rescue-free Periods	18.2 (29.2)	13.5 (24.6)	17.1 (30.5)	16.3 (28.2)
Daily Rescue Use	3.5 (7.3)	3.2 (2.7)	2.9 (2.3)	3.2 (4.7)
% Symptom-free Days	10.8 (22.2)	12.7 (24.4)	10.9 (24.7)	11.5 (23.7)
% Symptom-free Nights	22.0 (31.9)	24.9 (36.6)	19.9 (30.6)	22.3 (33.1)
% 24-hour Symptom-free Periods	4.0 (10.6)	7.8 (20.5)	7.0 (21.0)	6.3 (18.0)
24-hour Symptom Score	2.6 (1.3)	2.3 (1.2)	2.7 (1.4)	2.5 (1.3)
Duration of Asthma (years)	17.8 (14.1)	18.3 (13.0)	18.7 (15.0)	18.3 (14.0)
At USA site	65 (57%)	66 (58%)	66 (58%)	197 (57%)

Source: Reviewer

Cell contents are mean (standard deviation) or frequency (percent)

Figure 2: Patient Withdrawal over Time in Study 27 (Source: Reviewer)

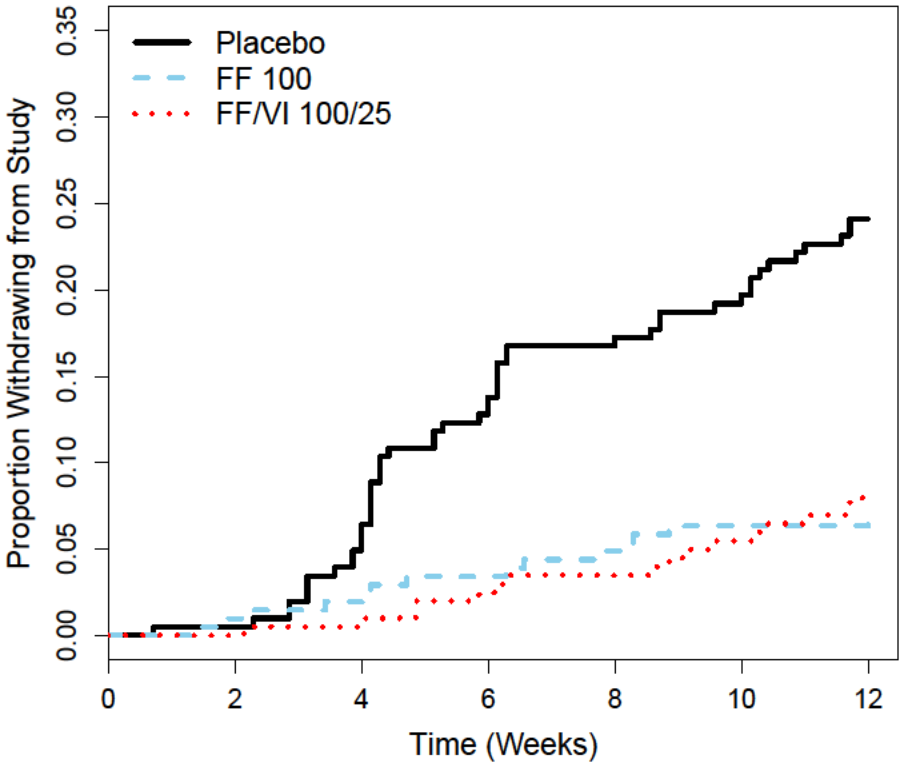


Figure 3: Patient Withdrawal over Time in Study 59 (Source: Reviewer)

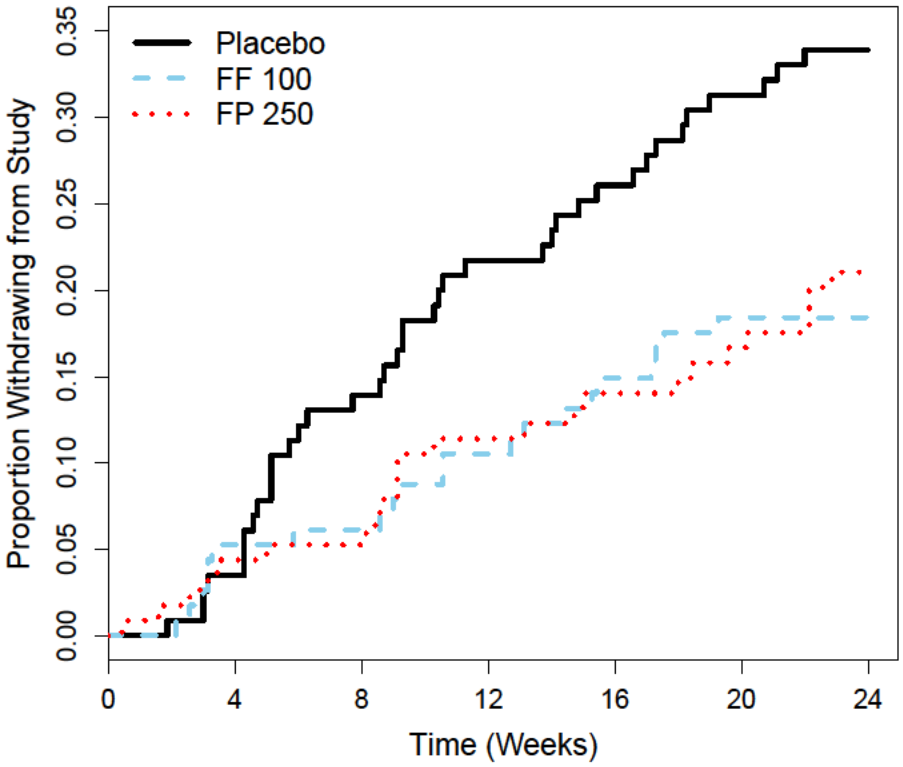


Table 4: Patient Dropout, by Reason for Withdrawal, in Study 27

	Placebo	FF 100	FF/VI 100/25	Overall
Completed Study	151 (74%)	185 (90%)	179 (89%)	515 (85%)
Withdrew from Study	52 (26%)	20 (10%)	22 (11%)	94 (15%)
Adverse event	1 (0%)	0 (0%)	2 (1%)	3 (0%)
Investigator discretion	6 (3%)	7 (3%)	6 (3%)	19 (3%)
Lack of efficacy	32 (16%)	6 (3%)	7 (3%)	45 (7%)
Lost to follow-up	0 (0%)	1 (0%)	2 (1%)	3 (0%)
Protocol deviation	7 (3%)	0 (0%)	2 (1%)	9 (1%)
Withdrew consent	6 (3%)	6 (3%)	3 (1%)	15 (2%)

Source: Reviewer

Table 5: Patient Dropout, by Reason for Withdrawal, in Study 59

	Placebo	FF 100	FP 250	Overall
Completed Study	75 (65%)	92 (81%)	88 (77%)	255 (74%)
Withdrew from Study	40 (35%)	22 (19%)	26 (23%)	88 (26%)
Adverse event	2 (2%)	2 (2%)	3 (3%)	7 (2%)
Investigator discretion	0 (0%)	0 (0%)	3 (3%)	3 (1%)
Lack of efficacy	23 (20%)	15 (13%)	14 (12%)	52 (15%)
Lost to follow-up	4 (3%)	0 (0%)	0 (0%)	4 (1%)
Protocol deviation	1 (1%)	2 (2%)	3 (3%)	6 (2%)
Withdrew consent	10 (9%)	3 (3%)	3 (3%)	16 (5%)

Source: Reviewer

3.4.3 Results in Studies 27 and 59

Results from the primary analyses of Studies 27 and 59 are described in Tables 6, 7, and 8. In Study 27, treatment with fluticasone furoate 100 mcg resulted in greater estimated changes from baseline than placebo in the co-primary endpoints 12-week mean trough FEV_1 and 12-week mean 0–24 hour postdose weighted mean FEV_1, with differences of 0.14 L (95% CI: 0.05, 0.22; p=0.002) and 0.19 L (95% CI: 0.06, 0.31; p=0.003), respectively. The determination of the adequacy of the statistical evidence to support a treatment effect in Study 27 is complicated by the lack of a strategy to control the type I error rate across the evaluations of FF and FF/VI (see 5.1 for further discussion). In Study 59, there was a statistically significant 0.15 L (95% CI: 0.04, 0.26; p=0.01) greater improvement on FF than placebo in the primary endpoint 24-week mean trough FEV_1 change. The estimated treatment effects on trough FEV_1 should be interpreted as differences in the mean change from baseline to the last available visit prior to 12 (Study 27) or 24 (Study 59) weeks (see 5.1 for further discussion).

Observed effects of FF on trough FEV_1 were evident as early as Week 2 and then remained relatively stable over the 12-week and 24-week treatment periods in Studies 27 and 59, respectively (Figures 4 and 5). Empirical distribution plots, in which dropouts were treated as the worst potential outcomes, suggested benefits of FF with respect to summary measures of the FEV_1 distribution besides the mean, such as the median or the proportions achieving 0.1 or 0.2 L improvements from baseline (Figures 6 and 7).

FF also showed benefit, or trends toward benefit, for additional endpoints of interest, including the percent of rescue-free 24-hour periods, percent of symptom-free 24-hour periods, morning peak expiratory flow, evening peak expiratory flow, AQLQ (+12) total score, and ACT score (Table 9). For example, treatment with FF, relative to placebo, led to estimated mean increases in the percent of symptom-free periods of 6% (95% CI: 0%, 12%) and 9% (95% CI: 1%, 17%) in Studies 27 and 59, respectively. Estimated mean improvements in the Asthma Control Test score on FF were 1.3 (95% CI: 0.6, 2.0) and 1.4 (95% CI: 0.4, 2.5), respectively. In Study 27, none of the treatment effects on the prespecified secondary endpoints of interest were statistically significant because of the multiple testing strategy – secondary endpoint tests were only to be evaluated for support of additional claims if all primary endpoint comparisons were successful, and the FF versus FF/VI comparisons (with respect to both trough and weighted mean FEV_1) were not statistically significant (p-values of 0.41 and 0.06, respectively). In Study 59, there was statistical evidence of a treatment effect on the percent of rescue-free 24-hour periods. There was not evidence of effects on any additional efficacy endpoints because the next test in the sequential multiple testing hierarchy evaluated differences in evening PEF, and this test was not statistically significant. As with the primary endpoints, the interpretation of the evaluation of these secondary endpoints is clouded by the substantial missing data (see 5.1 for additional

discussion).

Data are also available from Study 59 for comparisons of FF 100 OD against an active control, the approved inhaled corticosteroid FP 250 BD. The estimated effects of FF and FP on trough FEV_1 were nearly identical (Table 8) – the estimated difference in mean changes from baseline between the two ICS products was 0.0 L (95% CI: -0.11, 0.11). The effects of FF and FP were also similar for additional endpoints of interest (Table 9).

There were few severe asthma exacerbations in these placebo-controlled studies. In Study 27, 9 patients (4%) on placebo and 4 patients (2%) on FF had a severe asthma exacerbation while receiving treatment. In Study 59, 8 patients (7%) on placebo, 3 patients (3%) on FF, and 2 patients (2%) on FP experienced an on-treatment exacerbation. None of the exacerbations in these studies resulted in hospitalization or emergency room visits.

The ANCOVA and MMRM-based analyses used to evaluate data from Studies 27 and 59 have important assumptions. Both analyses assume constant variance, and the mixed effects model also assumes normally distributed errors and normally distributed random intercepts. Residual plots suggested some departures from constant variance and normality. Therefore, we also fit simple linear regression models (using only baseline and last available visit data) to estimate treatment effects, with adjustment for baseline FEV_1, region, sex, and age, and the use of robust Huber-White standard errors. These analyses, which do not rely on assumptions of normality or constant variance, produced nearly identical estimates and similar confidence intervals (results not shown) to the primary and secondary analyses.

Table 6: Analyses of the Co-Primary Endpoint Trough FEV_1 in Study 27: Differences in Mean Changes from Baseline to the Last Available Visit up to Week 12

		Placebo (N=203)	FF 100 (N=205)
Last Available Follow-up Visit, N	None	10	2
	Week 2	17	4
	Week 4	14	7
	Week 8	8	4
	Week 12	154	188
Mean (SD) Trough FEV_1, L	Baseline, All Patients	2.33 (0.63)	2.29 (0.62)
	Change, Completers	0.28 (0.45)	0.33 (0.45)
	Change, Dropouts	-0.03 (0.46)	0.23 (0.49)
	Change, All Patients	0.22 (0.47)	0.32 (0.45)
Estimated Difference from Placebo[1]			0.14
(95% CI)			(0.05, 0.22)
p-value			0.002

Source: Reviewer

Abbreviations: SD = standard deviation; CI = confidence interval

[1] Based on linear regression model of last available observation adjusting for baseline FEV_1, region, sex, and age

Table 7: Analyses of the Co-Primary Endpoint 0–24 Hour Postdose Weighted Mean FEV_1 in Study 27: Differences in Mean Changes from Baseline to Week 12

		Placebo (N^1=121)	FF 100 (N^1=119)
N (%) with 12-Week Measurements		96 (79%)	108 (91%)
Mean (SD) Weighted Mean FEV_1, L	Baseline, All Patients	2.33 (0.63)	2.29 (0.62)
	Change, Completers	0.25 (0.48)	0.38 (0.50)
Estimated Difference from Placebo[2]			0.19
(95% CI)			(0.06, 0.31)
p-value			0.003

Source: Reviewer

Abbreviations: SD = standard deviation; CI = confidence interval

[1] Only a subset of sites were selected to perform serial FEV_1 measurements

[2] Based on linear regression model of observed Week 12 data adjusting for baseline FEV_1, region, sex, and age

Table 8: Analyses of the Primary Endpoint Trough FEV_1 in Study 59: Differences in Mean Changes from Baseline to the Last Available Visit up to Week 24

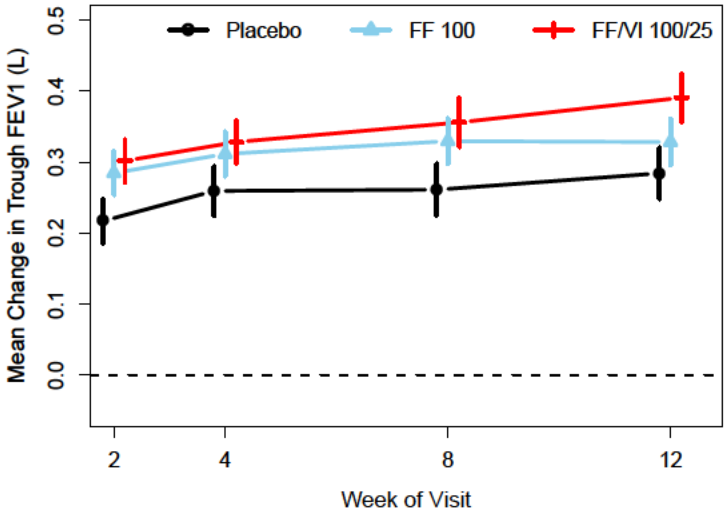
		Placebo (N=115)	FF 100 (N=114)	FP 250 (N=114)
Last Available Follow-up Visit, N	None	2	2	7
	Week 2	8	5	0
	Week 4	9	2	2
	Week 8	7	4	9
	Week 12	5	4	1
	Week 16	5	4	4
	Week 20	4	1	4
	Week 24	75	92	87
Mean (SD) Trough FEV_1, L	Baseline, All Patients	2.33 (0.65)	2.37 (0.63)	2.36 (0.73)
	Change, Completers	0.13 (0.45)	0.19 (0.40)	0.17 (0.42)
	Change, Dropouts	-0.20 (0.42)	0.02 (0.65)	0.05 (0.32)
	Change, All Patients	0.02 (0.47)	0.17 (0.45)	0.15 (0.40)
Estimated Difference from Placebo[1]			0.15	0.14
(95% CI)			(0.04, 0.26)	(0.03, 0.26)
p-value			0.01	0.01
Estimated Difference from FP 250[1]			0.00	
(95% CI)			(-0.11, 0.11)	
p-value			0.98	

Source: Reviewer

Abbreviations: SD = standard deviation; CI = confidence interval

[1] Based on linear regression model of last available observation adjusting for baseline FEV_1, region, sex, and age

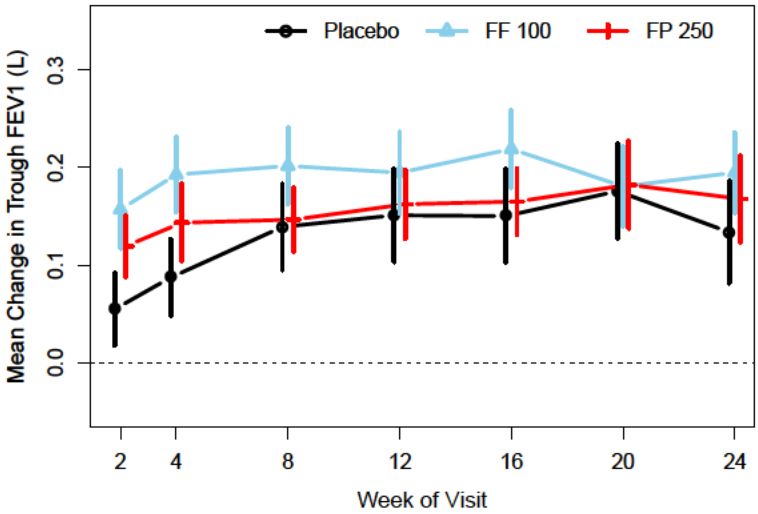
Figure 4: Mean Change from Baseline in Trough FEV_1 over Time in Study 27 Based on Observed Data. Error Bars Represent Plus or Minus One Standard Error. (Source: Reviewer)



Number of Patients with Data

Placebo	192	176	161	154
FF 100	203	199	192	188
FF/VI 100/25	200	192	189	184

Figure 5: Mean Change from Baseline in Trough FEV_1 over Time in Study 59 Based on Observed Data. Error Bars Represent Plus or Minus One Standard Error. (Source: Reviewer)



Number of Patients with Data

Placebo	112	105	96	88	84	79	75
FF 100	110	106	105	100	97	93	92
FP 250	105	107	104	95	94	90	87

Figure 6: Empirical Distribution Function for Change from Baseline in Trough FEV_1 at 12 Weeks in Study 27 (Source: Reviewer)

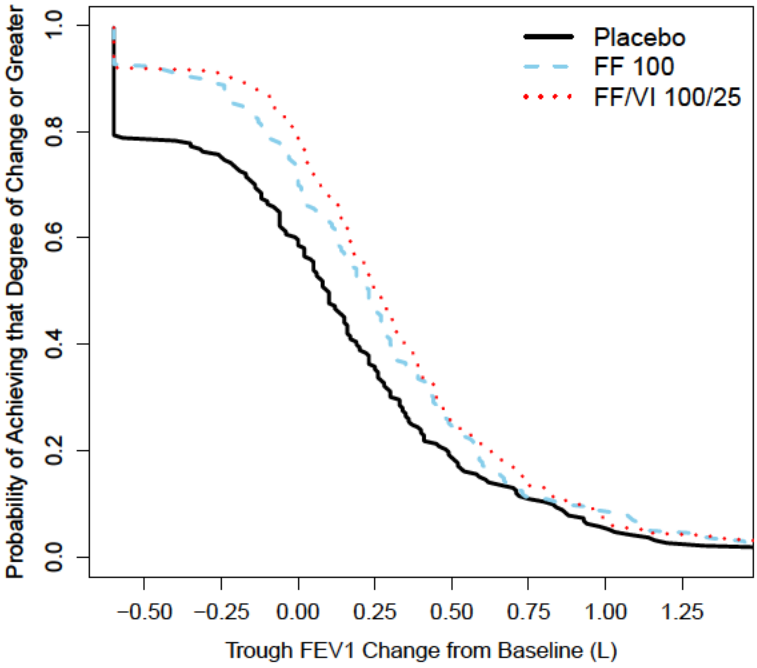


Figure 7: Empirical Distribution Function for Change from Baseline in Trough FEV_1 at 24 Weeks in Study 59 (Source: Reviewer)

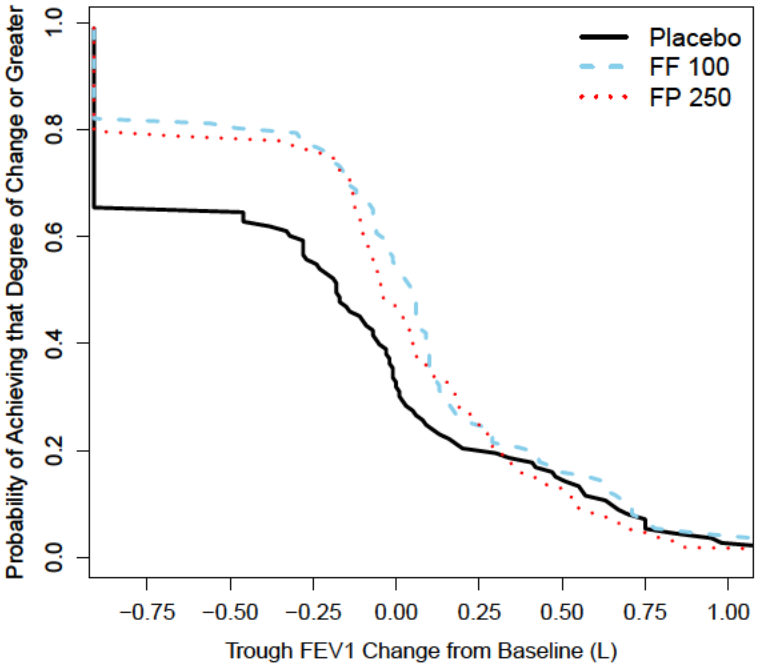


Table 9: Analyses of Additional Supportive Endpoints in Studies 27 and 59: Differences in Mean Changes from Baseline During Adherence to Treatment

	FF 100 vs Placebo (95% CI)		FF 100 vs FP 250 (95% CI)
Endpoint	Study 27 (Week 12)	Study 59 (Week 24)	Study 59 (Week 24)
% Rescue-free 24-hour periods	9 (2, 15)	15 (7, 23)	-3 (-11, 5)
% Symptom-free 24-hour periods	6 (0, 12)	9 (1, 17)	0 (-8, 8)
Morning PEF (L/min)	19 (12, 25)	9 (-1, 19)	4 (-6, 14)
Evening PEF (L/min)	16 (9, 23)	3 (-7, 12)	-3 (-12, 7)
AQLQ (+12) total score	0.15 (-0.01, 0.31)	0.33 (0.09, 0.57)	0.17 (-0.07, 0.40)
ACT score	1.3 (0.6, 2.0)	1.4 (0.4, 2.5)	0.3 (-0.6, 1.3)

Source: Reviewer

Abbreviations: CI = confidence interval

[1] All analyses based on linear regression models adjusting for baseline FEV_1, region, sex, and age, except for analyses of AQLQ (+12) and ACT scores in Study 59, which were based on analogous mixed effects models

[2] Statistical evidence of a difference between FF 100 and placebo was not demonstrated for any of these endpoints in Study 27, according to prespecified multiple testing strategy

[3] Statistical evidence of a difference between FF 100 and placebo was demonstrated only for the % of rescue-free periods in Study 59, according to prespecified multiple testing strategy

3.4.4 Potential Effect of Missing Data

As described in detail in 3.4.2, there were substantial missing data in the placebo-controlled Studies 27 and 59. Overall dropout rates were 15% and 26%, respectively. We used a number of approaches to investigate the potential effect of missing data on the reliability of efficacy results. First, we explored whether patients who dropped out were similar to patients who completed the studies. Patients who would go on to withdraw early tended to have a slightly greater disease burden at baseline than patients who would go on to complete the studies (Appendix: Tables 20 and 21). For example, there was a noticeable difference in the percent of nights that were rescue-free, with an average of 59% in completers, as compared to 30% in dropouts, in Study 27. The averages were 49% and 34%, respectively, in Study 59.

We also examined trends in trough FEV_1 before dropout within each treatment arm. Tables 6 and 8 show the mean changes from baseline to the last available visit by treatment arm in the subset of patients who dropped out early from Studies 27 and 59, respectively. Figures 8 and 9 display average pulmonary function over time in these patients. There is substantial variability in the estimated means because of the small numbers of patients (particularly at later visits), but an important pattern was evident: patients on FF tended to have better pulmonary function than placebo patients before study withdrawal. On the one hand, it therefore seems unlikely that patients treated with FF who withdrew from the study early went on to have substantially worse lung function at the end of the study than patients on placebo who dropped out. This is reassuring, especially in combination with the observation of greater dropout on placebo than FF because of lack of efficacy. On the other hand, this pattern highlights a deficiency of the primary evaluation of the last available observation estimand – those patients on FF who showed early FEV_1 improvements but then dropped out will be assigned positive outcomes in the primary analysis despite the fact that they will not receive long-term benefit from the treatment.

Therefore, we also evaluated alternative estimands that may be more meaningful for all patients. We considered the Jump to Reference multiple imputation approach performed by the applicant as a potential evaluation of the de facto estimand, i.e., the difference in mean changes from baseline in trough FEV_1 at the end of study in all patients, regardless of adherence. Under the Jump to Reference approach, statistical significance was maintained for the comparison of FF with placebo in Study 27, but not in Study 59 (Table 10). Estimated magnitudes of treatment effect were approximately 20-30% smaller than those based on the primary analyses in the two studies. For example, in Study 59 (with the most missing end-of-study data), the estimated mean improvement in FEV_1 on FF, relative to placebo, was 0.11 L (95% CI: -0.01, 0.23), as compared to 0.15 L (95% CI: 0.04, 0.26) in the primary analysis. Although the scientific justification of the Jump to Reference assumptions seems reasonable, this and all other potential analyses to evaluate the de facto estimand rely on untestable assumptions about unobserved data.

In addition, none of the sensitivity analyses conducted by the applicant allow for the possibility that dropouts on FF could have experienced worse outcomes after discontinuation than dropouts on control. That being said, the observed trend toward greater FEV_1 on FF than placebo before dropout (Figures 8 and 9) somewhat mitigates this concern, at least with respect to pulmonary function. We also conducted a simple tipping point analysis in which different mean changes were imputed for the subsets of placebo and FF patients who withdrew early from the two studies. In Study 27, it requires the assumption of approximately a 1 L decrease in trough FEV_1 at 12 weeks in dropouts on FF, as compared to no change in dropouts on placebo, for the estimated difference to change from favoring FF to favoring placebo (Table 11). In Study 59, where there were more missing data, the estimated difference would change signs with around a 0.4 L decrease in FF dropouts, as compared to no change in placebo dropouts (Table 12).

We also evaluated estimands aimed at the utility of the new treatment, in which patients who discontinued were assigned a bad outcome. The empirical distribution functions discussed earlier (Figures 6 and 7) show some separation between the placebo and FF FEV_1 distributions. Table 13 also presents the results of analyses to compare the proportion of patients achieving certain threshold improvements in trough FEV_1 (with dropouts considered non-responders). There were trends toward greater probabilities of 100–500 mL increases on FF than placebo in both studies, with larger estimated differences in Study 27 than Study 59. The most consistent evidence was for a 100 mL increase, with estimated 15% (95% CI: 5%, 25%) and 14% (95% CI: 1%, 27%) greater absolute probabilities on FF than placebo in Studies 27 and 59, respectively. These analyses aimed at evaluating alternative utility estimands largely provide support for the results of the primary analyses.

Figure 8: Trough FEV_1 over Time in Dropouts Prior to Withdrawal in Study 27 (Source: Reviewer)

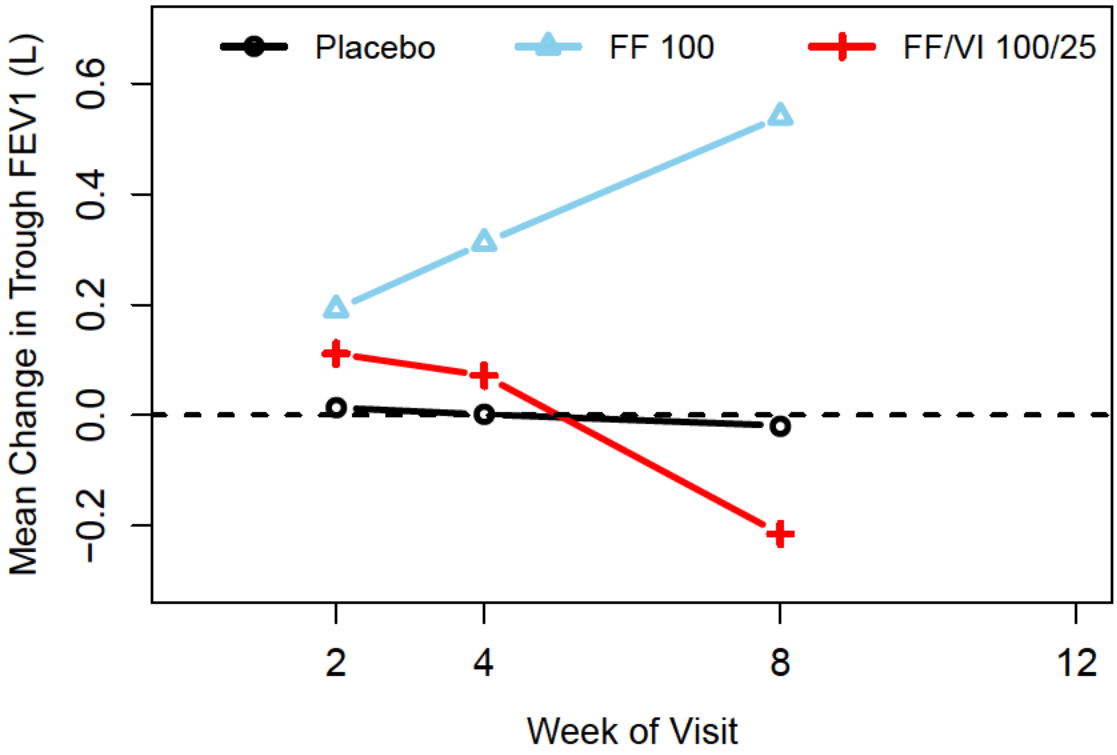


Table 10: Exploring the Potential Effect of Missing Data: Results for the Primary Endpoint Trough FEV_1 with the Primary Analysis as Compared to a Multiple Imputation Supportive Analysis in Studies 27 and 59

	FF 100 versus Placebo Mean Difference in Trough FEV_1 Change from Baseline, L (95% CI)	
	Primary Analysis[1]	Supportive Analysis[2]
Study 27 (12 Weeks)	0.14 (0.05, 0.22)	0.12 (0.03, 0.21)
Study 59 (24 Weeks)	0.15 (0.04, 0.26)	0.11 (-0.01, 0.23)

Source: Reviewer

Abbreviations: CI = confidence interval

[1] Based on linear regression model adjusting for age, sex, region, and baseline FEV_1, with last observation carried forward (i.e., last available observation)

[2] Based on Jump to Reference Multiple Imputation Approach

Figure 9: Trough FEV_1 over Time in Dropouts Prior to Withdrawal in Study 59 (Source: Reviewer)

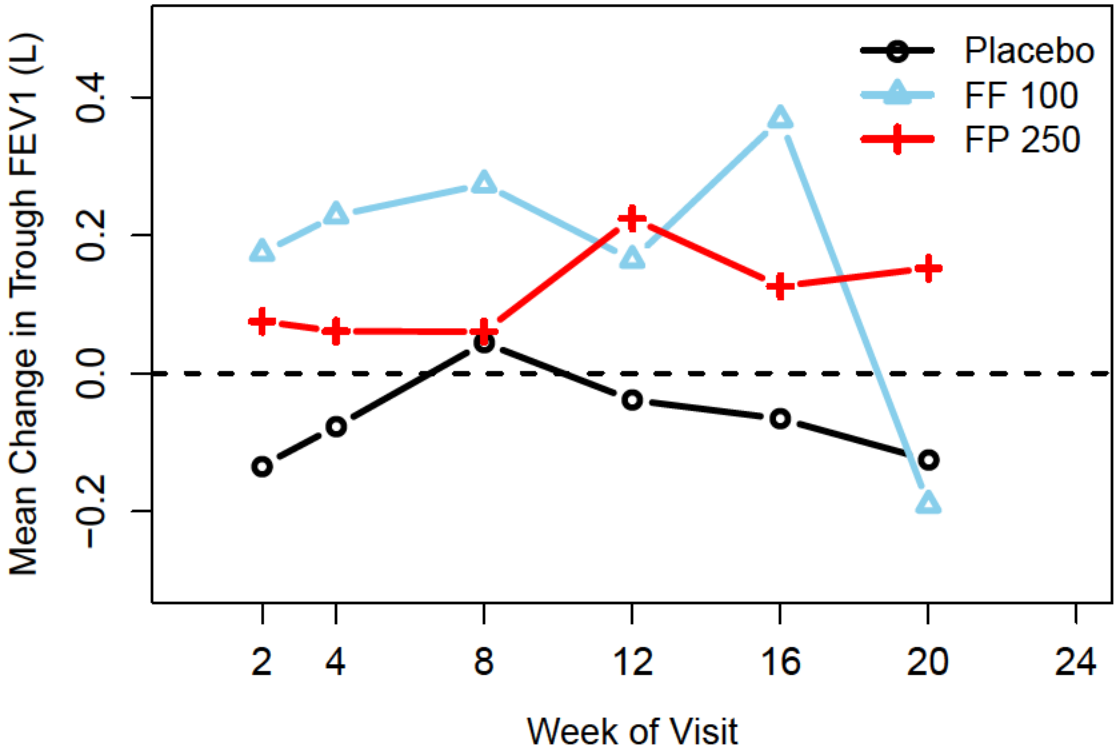


Table 11: Tipping Point Analysis in Study 27: Estimated Differences Between FF and Placebo in the Mean Change from Baseline to 12 Weeks in Trough FEV_1 (L) under Different Imputed Values for the Missing Mean Changes in Patients who Withdrew from the Study Early on Each Treatment Arm

		Imputed Mean Change in FF Dropouts, L				
		-1.00	-0.75	-0.50	-0.25	0.00
Imputed Mean Change in Placebo Dropouts, L	-1.00	0.24	0.26	0.29	0.31	0.33
	-0.75	0.18	0.20	0.23	0.25	0.27
	-0.50	0.12	0.14	0.16	0.19	0.21
	-0.25	0.06	0.08	0.10	0.13	0.15
	0.00	0.00	0.02	0.04	0.06	0.09

Source: Reviewer

The estimated mean difference in the primary analysis of last available observation was 0.14 L

Table 12: Tipping Point Analysis in Study 59: Estimated Differences Between FF and Placebo in the Mean Change from Baseline to 24 Weeks in Trough FEV_1 (L) under Different Imputed Values for the Missing Mean Changes in Patients who Withdrew from the Study Early on Each Treatment Arm

		Imputed Mean Change in FF Dropouts, L				
		-0.4	-0.3	-0.2	-0.1	0.0
Imputed Mean Change in Placebo Dropouts, L	-0.4	0.13	0.15	0.17	0.19	0.21
	-0.3	0.10	0.12	0.14	0.16	0.17
	-0.2	0.06	0.08	0.10	0.12	0.14
	-0.1	0.03	0.05	0.07	0.09	0.10
	0.0	-0.01	0.01	0.03	0.05	0.07

Source: Reviewer

The estimated mean difference in the primary analysis of last available observation was 0.15 L

Table 13: Differences Between FF and Placebo in the Probability of Achieving Certain Threshold Changes in Trough FEV_1 in Studies 27 and 59

	Study 27 (12 Weeks)			Study 59 (24 Weeks)		
Threshold	Placebo	FF	Difference (95% CI)[1]	Placebo	FF	Difference (95% CI)[1]
100 mL	0.48	0.63	0.15 (0.05, 0.25)	0.24	0.39	0.14 (0.01, 0.27)
200 mL	0.38	0.52	0.14 (0.04, 0.24)	0.21	0.26	0.05 (-0.06, 0.17)
300 mL	0.30	0.41	0.11 (0.01, 0.21)	0.20	0.21	0.01 (-0.10, 0.12)
400 mL	0.23	0.33	0.10 (0.01, 0.19)	0.18	0.20	0.02 (-0.09, 0.13)
500 mL	0.18	0.24	0.07 (-0.02, 0.15)	0.15	0.16	0.01 (-0.09, 0.11)

Source: Reviewer

Abbreviations: CI = confidence interval

[1] Based on unadjusted difference in proportions with patients who withdrew from the study early considered non-responders, and confidence intervals based on the normal approximation

3.4.5 Results in Additional Studies

Study 29 allows for a comparison between the higher 200 mcg once-daily dose of FF and the approved 500 mcg twice-daily dose of FP, a higher dose than the 250 mcg dose used as the comparator in Study 59 (Table 14). There was no statistical evidence of a difference between the treatments in the mean change from baseline in trough FEV_1 (estimated 0.02 L greater change on FF 200; 95% CI: -0.07, 0.10 L). There was perhaps a slight trend toward greater improvement on FF than FP in a plot of the mean change in trough FEV_1 over time, as well as an empirical distribution plot of the change at Week 24 (Appendix: Figures 15 and 17). The 95% CI ruled out the sponsor's prespecified non-inferiority margin of -0.125 L, although this margin was not adequately justified. Changes from baseline in key secondary endpoints were similar on FF and FP. This study demonstrated assay sensitivity (the ability to detect a difference, if one exists), as patients on FF/VI 200/25 had a statistically significantly greater improvement in trough FEV_1 than patients on FF 200 ($p<0.001$).

Study 96 compares the 100 and 200 mcg once-daily doses of FF (Table 15). There was no statistical evidence of a difference between the treatments in the mean change from baseline in trough FEV_1, although there was a trend toward greater improvement on FF 200 than FF 100 (estimate: 0.08 L, 95% CI: -0.04, 0.19 L). Slight trends were also evident in plots of FEV_1 over time and in comparing the empirical distribution functions (Appendix: Figures 16 and 18). Changes from baseline in key secondary endpoints were largely similar on FF 200 and 100.

One of the factors cited by the applicant to support the approval of the 200 mcg dose of FF (in addition to the 100 mcg dose) is the observed greater difference between FF 200 and 100 in patients who were using a high-dose ICS during the run-in period. In patients who had been using a mid-dose ICS, the estimated difference between FF 200 and 100 in mean trough FEV_1 change was 0.06 L, as compared to 0.13 L in the high-dose ICS subgroup. However, a test for interaction suggested that the observed difference between the treatment effects in the two subgroups (estimate=0.07 L; 95% CI: -0.36, 0.22) could have been due to random chance ($p=0.63$). In addition, a plot of differences between FF 200 and 100 by selected important subgroups (Figure 10) showed similar variability across subgroup effects for several other patient characteristics.

Table 14: Analyses of Trough FEV_1 and Additional Endpoints in Study 29: Differences in Mean Changes from Baseline During Adherence to Treatment

		FP 500 BD (N=195)	FF 200 OD (N=194)
Last Available Follow-up Visit, N	None	5	7
	Week 3	8	9
	Week 4	7	7
	Week 8	7	7
	Week 12	3	8
	Week 16	1	2
	Week 20	6	7
	Week 24	158	147
Mean (SD) Trough FEV_1, L	Baseline, All Patients	2.14 (0.67)	2.19 (0.68)
	Change, Completers	0.21 (0.38)	0.29 (0.47)
	Change, Dropouts	0.01 (0.42)	-0.05 (0.51)
	Change, All Patients	0.17 (0.39)	0.22 (0.50)
	Mean Difference from FP 500 (95% CI)		0.02 (-0.07, 0.10)
0–24 Hour Weighted Mean FEV_1, L Difference from FP 500 (95% CI)			0.07 (-.07, 0.21)
% Rescue-free 24-hour periods Difference from FP 500 (95% CI)			-5 (-12, 1)
% Symptom-free 24-hour periods Difference from FP 500 (95% CI)			-4 (-10, 3)
AM PEF (L/min) Difference from FP 500 (95% CI)			-1 (-8, 8)
PM PEF (L/min) Difference from FP 500 (95% CI)			-5 (-13, 4)
ACT score Difference from FP 500 (95% CI)			0.4 (-0.4, 1.3)
AQLQ score Difference from FP 500 (95% CI)			-0.03 (-0.22, 0.17)

Source: Reviewer

Abbreviations: SD = standard deviation; CI = confidence interval

[1] Analyses based on linear regression models adjusting for baseline FEV_1, region, sex, and age, or analogous mixed effects models (for AQLQ (+12) and ACT scores)

Table 15: Analyses of Trough FEV_1 and Additional Endpoints in Study 96: Differences in Mean Changes from Baseline During Adherence to Treatment

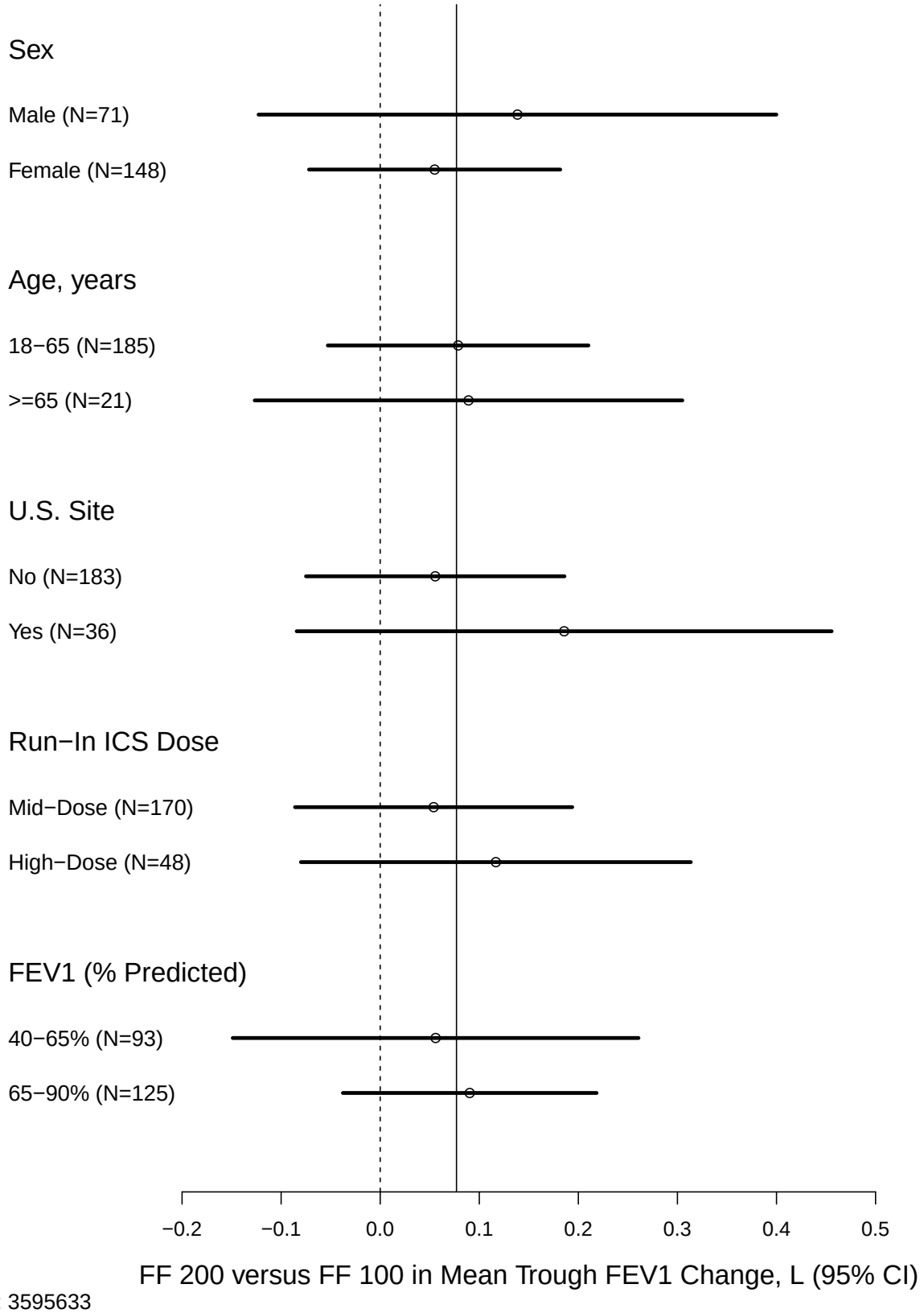
		FF 100 (N=108)	FF 200 (N=111)
Last Available Follow-up Visit, N	None	2	2
	Week 2	2	1
	Week 4	0	2
	Week 8	1	3
	Week 12	3	1
	Week 18	5	3
	Week 24	95	99
Mean (SD) Trough FEV_1, L	Baseline, All Patients	2.04 (0.67)	2.08 (0.65)
	Change, Completers	0.22 (0.42)	0.30 (0.48)
	Change, Dropouts	0.03 (0.21)	0.15 (0.42)
	Change, All Patients	0.20 (0.41)	0.29 (0.48)
	Mean Difference from FF 100 (95% CI)		0.08 (-0.04, 0.19)
% Rescue-free 24-hour periods Difference from FF 100 (95% CI)			2 (-7, 10)
% Symptom-free 24-hour periods Difference from FF 100 (95% CI)			2 (-6, 10)
AM PEF (L/min) Difference from FF 100 (95% CI)			0 (-9, 9)
PM PEF (L/min) Difference from FF 100 (95% CI)			1 (-8, 10)
ACT score Difference from FF 100 (95% CI)			0.2 (-0.7, 1.2)

Source: Reviewer

Abbreviations: SD = standard deviation; CI = confidence interval

[1] Analyses based on linear regression models adjusting for baseline FEV_1, region, sex, and age, or an analogous mixed effects model (for ACT score)

Figure 10: Estimated Difference Between FF 200 and FF 100, Stratified by Selected Subgroups, in Study 96. Solid Vertical Line Represents Estimated Treatment Effect in Overall Population, and Dashed Vertical Line Represents No Difference. (Source: Reviewer)



3.5 Evaluation of Safety

The reader is referred to the Medical Review by Dr. Tracy Kruzick for an evaluation of the safety of fluticasone furoate.

4 Findings in Special/Subgroup Populations

Figures 11 and 12 present the results of subgroup analyses by sex, race (White, Black, or Asian), age ($\leq$18, 18–65, $\geq$65), and geographic region (non-U.S. versus U.S.) in Studies 27 and 59, respectively. Estimated differences in mean trough FEV_1 comparing FF with placebo were largely consistent across the subgroups. There was a trend toward a smaller observed treatment effect in older patients in both studies, although tests for interaction between treatment and age (as a continuous variable) suggested that these observed differences may have been due to random chance (p-values of 0.63 and 0.25 in Studies 27 and 59, respectively). The limited numbers of Black and Asian patients led to large variability in the estimated treatment effects in these subgroups, and the number of Asians in Study 59 was too small to get a sufficiently reliable estimated treatment effect to report.

Figure 11: Estimated Treatment Effect of FF, Stratified by Selected Subgroups, in Study 27. Solid Vertical Line Represents Estimated Treatment Effect in Overall Population, and Dashed Vertical Line Represents No Difference. (Source: Reviewer)

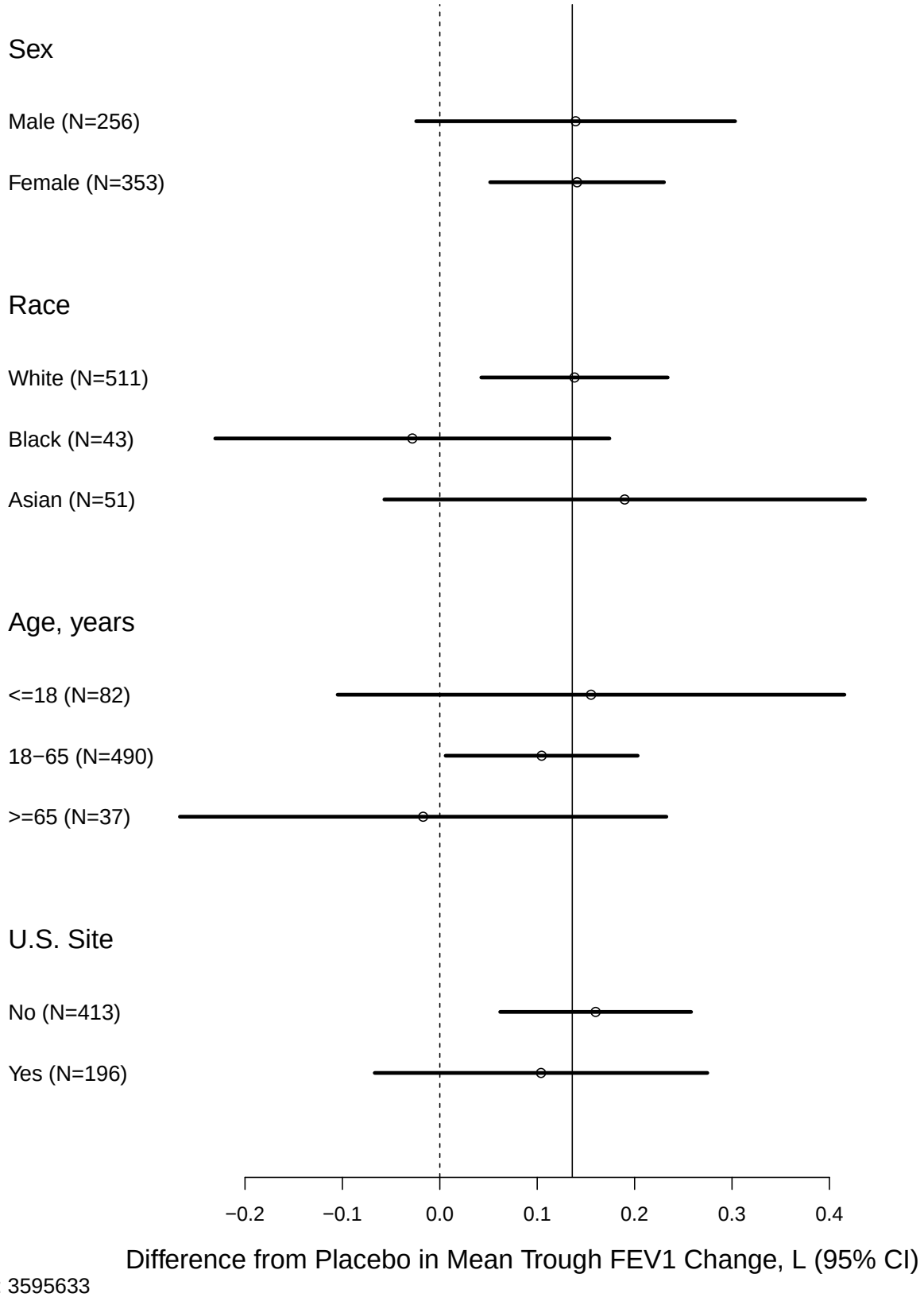
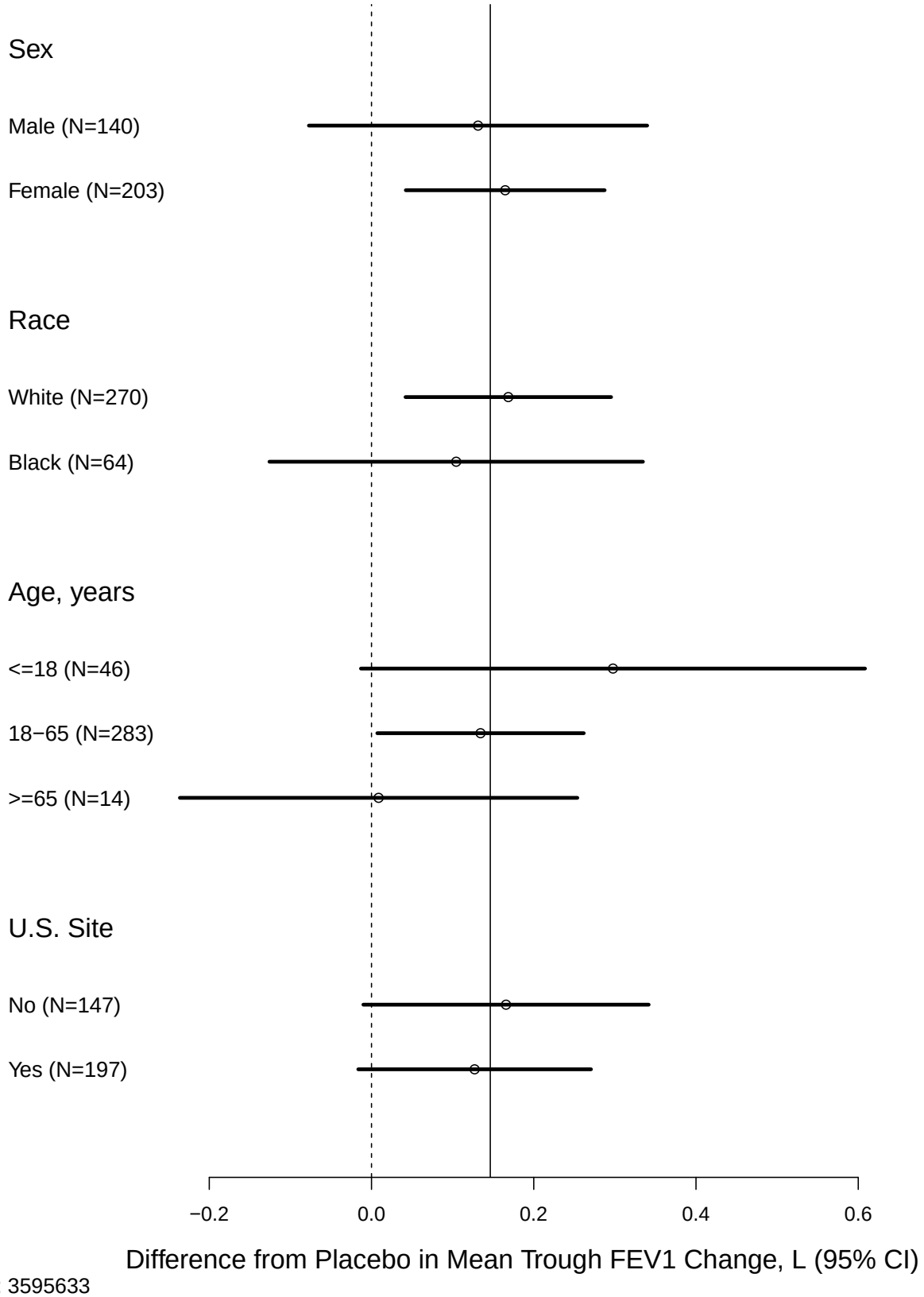


Figure 12: Estimated Treatment Effect of FF, Stratified by Selected Subgroups, in Study 59. Solid Vertical Line Represents Estimated Treatment Effect in Overall Population, and Dashed Vertical Line Represents No Difference. (Source: Reviewer)



5 Summary and Conclusions

5.1 Statistical Issues

During this statistical review, we identified the following important issues:

- Potential effect of missing data on the reliability of efficacy results

  This issue was discussed in detail in 3.3.2 and 3.4.4. There were substantial missing data at the end of the study in the placebo-controlled Studies 27 and 59, with overall dropout rates of 15% and 26%, respectively. We interpret the primary linear regression analyses using last observation carried forward as evaluations of the last available observation estimand, i.e., the difference in mean trough FEV_1 until the last time point prior to the end of the study at which patients adhere to the assigned treatment. There were little missing data with respect to this estimand, and results demonstrated benefit for FF over placebo in both studies. However, the LAO estimand may not be meaningful for all patients because it assigns positive outcomes to patients who showed an early FEV_1 improvement but could not tolerate or adhere to the therapy. Therefore, we gave importance to supportive analyses evaluating alternative estimands.

  We considered a multiple imputation approach performed by the applicant aimed at evaluating the de facto estimand, i.e., the difference in mean trough FEV_1 at the end of the study in all randomized patients, regardless of adherence. Statistical significance was maintained for the FF versus placebo comparison in Study 27, but not in Study 59, and estimated treatment effects were approximately 20–30% smaller than those based on the primary analyses in the two studies. This multiple imputation-based analysis relies on untestable assumptions about the missing data and does not allow for the possibility that dropouts on FF could have experienced worse outcomes after discontinuation than dropouts on placebo. However, a simple tipping point analysis suggested that dropouts on FF would have had to experience far worse future outcomes than dropouts on placebo for the estimated difference between the groups to decline to zero. Such an assumption is likely not plausible, especially given the fact that patients on FF tended to show greater FEV_1 prior to withdrawal than patients on placebo.

  The results of the primary analyses were also supported by evaluations of utility estimands, in which patients who dropped out were assigned a bad outcome (e.g., considered a non-responder). Empirical distribution plots showed separation between the placebo and FF FEV_1 distributions, and there were trends toward greater probabilities of certain threshold increases in trough FEV_1 on FF than placebo in both studies.

- Use of the surrogate marker FEV_1 as the primary efficacy endpoint

  The primary endpoint in the phase 3 efficacy studies was the mean change from baseline in trough (or weighted mean) FEV_1 at 12 or 24 weeks. We consider FEV_1 to be a surrogate endpoint, because it does not directly measure how a patient functions or feels in daily life, or how long a patient survives [2]. Spirometric assessments like FEV_1 provide standardized, easy to perform, and reproducible assessments of pulmonary function and are commonly used and accepted by FDA as primary efficacy endpoints in asthma clinical trials. However, because they do not directly measure the asthma symptoms (e.g., wheezing, chest tightness, shortness of breath, coughing, and exacerbation) that are important to patients, the claim of effectiveness based on the primary analyses relies on the conclusion that the treatment effect on FEV_1 will reliably predict effects on a clinically meaningful endpoint. Therefore, we also considered the analyses of several secondary endpoints to be important in the overall evaluation of effectiveness.

  The following additional endpoints ascertained in the phase 3 studies might be considered to provide some direct measure of how patients function or feel in daily life: percent rescue-free 24-hour periods, percent symptom-free 24-hour periods, AQLQ (+12) score, and ACT score. In both studies, FF showed benefit, or trends toward benefit, for these additional endpoints of interest. For example, treatment with FF led to estimated improvements over placebo in the mean percent of symptom-free periods of 6% (95% CI: 0%, 12%) and 9% (95% CI: 1%, 17%) in Studies 27 and 59, respectively. Estimated mean improvements in the Asthma Control Test score on FF were 1.3 (95% CI: 0.6, 2.0) and 1.4 (95% CI: 0.4, 2.5), respectively. Therefore, results for these secondary assessments provide additional support for the effectiveness of FF in asthma, increasing confidence that the treatment effect on the surrogate marker FEV_1 will reliably predict clinical benefit.

- Statistical evidence in Study 27 in the presence of multiple comparisons

  In Study 27, the estimated mean differences between FF and placebo with respect to changes from baseline in the co-primary endpoints trough FEV_1 and 0–24 hour weighted mean FEV_1 were 0.14 L (95% CI: 0.05, 0.22; p=0.002) and 0.19 L (95% CI: 0.06, 0.31; p=0.003). Despite the low p-values, the statistical significance of these evaluations might be questioned because of the multiple comparisons in this study. The applicant prespecified six primary analyses of interest, including comparisons of FF with placebo, FF/VI with placebo, and FF/VI with FF, with respect to the co-primary endpoints. However, the applicant did not clearly indicate how the family-wise type I error rate would be controlled across these six comparisons. The co-primary endpoint comparisons of FF/VI with FF did not show evidence of differences (p-values of 0.41 and 0.06). Therefore, if a sequential gatekeeping strategy was carried out in the order the comparisons were listed in the

protocol (see Figure 1), the FF versus placebo comparisons would not be statistically significant because of the failed FF/VI versus FF tests.

That being said, we consider the statistical evidence against the null hypothesis of no FF treatment effect in Study 27 to be strong for the following reasons. First, if a conservative Bonferroni adjustment were applied to the six tests (resulting in a threshold for significance of $0.05/6 = 0.0083$), the p-values of 0.002 and 0.003 would provide evidence of statistically significant improvements on FF over placebo with respect to the co-primary endpoints. Second, the concern that the observed FF versus placebo difference might represent a false positive is somewhat mitigated by the highly statistically significant difference between FF/VI and placebo ($p<0.001$ for both co-primary endpoints), which appears to have been primarily driven by the effect of FF. The consistent effects of FF across secondary endpoints (including non-spirometric endpoints not strongly correlated to the primary endpoints) also alleviates this concern. Third, the failed tests that motivated this discussion evaluated a different product (FF/VI) than is being considered in this review and therefore do not directly lend doubt to the interpretation of the observed FF versus placebo differences. Finally, although the multiplicity approach carried out by the applicant was prespecified, we did not convey its limitations to the applicant prior to the unblinding of study results. Therefore, it is difficult to require that the results hold up to a worst-case testing strategy (a gatekeeping hierarchy) when statistical significance was evident using an approach (Bonferroni) known to provide conservative control of the family-wise error rate.

- Evidence to support the 200 mcg dose of FF

The applicant is seeking the approval of both the 100 and 200 mcg doses of FF for treatment of asthma. There was statistical evidence of efficacy for FF 100 over placebo in two phase 3 clinical trials. Direct comparisons between FF 100 and 200 were available in one phase 3 trial, in addition to two phase 2 dose-ranging studies. In the phase 3 trial Study 96, there was no statistical evidence of a difference between the treatments in trough FEV_1, although there was a trend toward greater improvement on FF 200 than FF 100 (estimate: 0.08 L, 95% CI: -0.04, 0.19 L). The applicant also noted that the estimated difference between FF 200 and 100 in mean trough FEV_1 change was 0.13 L in patients who had been using a high-dose ICS, as compared to 0.06 L in the mid-dose ICS subgroup. However, this degree of variability across subgroup effects would not be unusual by random chance if there were no truly effect modification (p-value for interaction=0.63).

Trends toward slightly greater FEV_1 improvement on FF 200 than 100 in the full study population were also observed in the phase 2 studies. In both Studies FFA109685 and FFA109687, the mean difference between FF 200 and 100 in Week 8 trough FEV_1 change

was 0.03 L. The phase 3 Study 29 also demonstrated similar effects between FF 200 and the approved high-dose ICS comparator FP 500 BD (estimated 0.02 L greater change on FF 200; 95% CI: -0.07, 0.10 L).

In summary, although we do not find the applicant's argument regarding a possible interaction between FF dose and past ICS dose to be convincing, there was a slight trend toward greater trough FEV_1 improvement on FF 200 than 100. There was not statistical evidence in any single study to support a greater treatment effect for FF 200, but estimated differences between the effects of the doses were relatively similar across the three phase 2 and 3 studies, with estimates ranging from 0.03 to 0.08 L.

5.2 Collective Evidence

The collective evidence supports the effectiveness of once-daily fluticasone furoate 100 mcg for maintenance treatment of asthma as prophylactic therapy in patients 12 years of age and older. In Studies 27 and 59, treatment with FF 100 resulted in statistically significantly greater mean changes from baseline than placebo in the primary endpoint trough FEV_1 (in addition to the co-primary endpoint 0–24 hour weighted mean FEV_1 in Study 27). The estimated treatment effects on trough FEV_1, which should be interpreted as differences in mean changes to the last visit during adherence, were 0.14 L (95% CI: 0.05, 0.22) and 0.15 L (95% CI: 0.04, 0.26) in Studies 27 and 59, respectively. There were trends toward slightly greater FEV_1 improvement on FF 200 than 100 mcg.

Analyses to evaluate the potential impact of missing data generally supported the effectiveness of FF. However, estimates of the treatment effect on the mean change in trough FEV_1 at the end of the study, regardless of adherence to assigned therapy, were approximately 20–30% smaller than estimates from the primary analyses. The effectiveness of FF was also supported by trends toward benefit with respect to several additional endpoints, including the proportion of rescue-free days, the proportion of symptom-free days, and the patient-reported outcomes AQLQ (+12) total score and ACT score. These trends toward benefit increase confidence that the treatment effect on the surrogate endpoint trough FEV_1 is likely to predict clinical benefit, i.e., improvements in how asthma patients function or feel in daily life, or survive.

5.3 Labeling Recommendations

We have made a number of recommended edits to the labeling proposed by the applicant. In particular, we recommended that a statistically valid approach be used to integrate data from multiple studies in tables of adverse event rates. The originally proposed approach simply

(b) (4)

This approach is subject to confounding by study (Simpson's Paradox). We also recommended that the description of primary results from Studies 27, 59, and 96 reflect the (last available observation) approach that was taken to address missing data. For example, we suggested the following language to describe the findings from Study 27: "At Week 12 or the last available on-treatment visit prior to Week 12, the mean change from baseline in trough FEV1 was greater among patients receiving ARNUITY ELLIPTA 100 mcg once daily than among those receiving placebo (treatment difference from placebo 0.14L and 95% confidence interval [0.05, 0.22])."

Appendix

Table 16: Baseline Characteristics in Study 29

	FF 200	FF/VI 200/25	FP 500	Overall
N	194	197	195	586
Female	113 (58%)	116 (59%)	116 (59%)	345 (59%)
Age (years)	44.6 (14.3)	46.6 (15.1)	47.3 (14.1)	46.2 (14.5)
Age Group (years)				
< 18	7 (4%)	8 (4%)	8 (4%)	23 (4%)
18-65	173 (89%)	167 (85%)	171 (88%)	511 (87%)
≥ 65	14 (7%)	22 (11%)	16 (8%)	52 (9%)
Race				
White	165 (85%)	165 (84%)	162 (83%)	492 (84%)
Black	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%)
Hispanic/Latino	6 (3%)	2 (1%)	3 (2%)	11 (2%)
Weight (kg)	81.1 (18.2)	79.1 (18.2)	79.6 (19.4)	79.9 (18.6)
Height (cm)	168.3 (9.7)	168.1 (9.3)	167.6 (9.4)	168.0 (9.4)
FEV_1	2.2 (0.7)	2.1 (0.7)	2.1 (0.7)	2.2 (0.7)
FEV_1 % Predicted	66.7 (12.4)	66.6 (12.6)	67.6 (12.2)	66.9 (12.4)
Morning PEF	333.5 (123.6)	327.4 (113.3)	330.2 (114.1)	330.3 (116.9)
Evening PEF	348.5 (120.0)	342.6 (112.4)	344.3 (116.1)	345.1 (116.0)
% Rescue-free Days	11.5 (24.5)	12.1 (24.2)	10.2 (22.0)	11.3 (23.6)
% Rescue-free Nights	46.0 (39.3)	57.7 (40.6)	50.9 (40.3)	51.6 (40.3)
% 24-hr Rescue-free Periods	7.8 (20.7)	7.6 (19.2)	6.3 (18.0)	7.2 (19.3)
Daily Rescue Use	4.0 (2.8)	4.1 (3.0)	4.2 (2.7)	4.1 (2.8)
% Symptom-free Days	6.3 (18.4)	8.1 (20.1)	5.2 (14.7)	6.6 (17.9)
% Symptom-free Nights	17.4 (32.0)	19.3 (32.1)	16.5 (30.8)	17.8 (31.6)
% 24-hr Symptom-free Periods	4.7 (16.1)	5.1 (15.2)	2.7 (9.8)	4.1 (14.0)
24-hour Symptom Score	3.4 (1.7)	3.3 (1.7)	3.5 (1.7)	3.4 (1.7)
Duration of Asthma (years)	14.7 (11.9)	17.0 (13.2)	14.9 (12.5)	15.5 (12.6)
At USA site	48 (25%)	48 (24%)	47 (24%)	143 (24%)

Source: Reviewer

Cell contents are mean (standard deviation) or frequency (percent)

Abbreviations: hr = hour

Table 17: Baseline Characteristics in Study 96

	FF 100	FF 200	Overall
N	108	111	219
Female	75 (69%)	73 (66%)	148 (68%)
Age (years)	47.3 (15.5)	45.5 (15.4)	46.4 (15.4)
Age Group (years)			
< 18	7 (6%)	6 (5%)	13 (6%)
18-65	89 (82%)	96 (86%)	185 (84%)
≥ 65	12 (11%)	9 (8%)	21 (10%)
Race			
White	94 (87%)	96 (86%)	190 (87%)
Black	2 (2%)	1 (1%)	3 (1%)
Asian	0 (0%)	2 (2%)	2 (1%)
Other	12 (11%)	12 (11%)	24 (11%)
Hispanic/Latino	67 (62%)	63 (57%)	130 (59%)
Weight (kg)	76.4 (16.1)	77.1 (18.6)	76.7 (17.4)
Height (cm)	163.0 (9.6)	163.8 (9.5)	163.4 (9.6)
FEV_1	2.0 (0.7)	2.1 (0.7)	2.1 (0.7)
FEV_1 % Predicted	68.6 (13.9)	68.0 (13.4)	68.3 (13.7)
Morning PEF	329.3 (111.2)	325.4 (109.3)	327.3 (110.0)
Evening PEF	340.5 (113.6)	332.7 (105.6)	336.6 (109.5)
% Rescue-free Days	19.6 (33.2)	16.6 (30.2)	18.1 (31.7)
% Rescue-free Nights	45.4 (37.6)	46.2 (37.8)	45.8 (37.6)
% 24-hour Rescue-free Periods	14.3 (28.9)	11.6 (25.5)	12.9 (27.2)
Daily Rescue Use	3.7 (2.6)	3.9 (2.6)	3.8 (2.6)
% Symptom-free Days	9.5 (22.7)	8.7 (20.8)	9.1 (21.7)
% Symptom-free Nights	17.1 (31.1)	12.9 (24.2)	15.0 (27.9)
% 24-hour Symptom-free Periods	6.1 (17.7)	4.9 (14.3)	5.5 (16.1)
24-hour Symptom Score	3.1 (1.6)	3.0 (1.3)	3.0 (1.4)
Duration of Asthma (years)	20.0 (16.3)	20.9 (14.8)	20.5 (15.5)
At USA site	16 (15%)	20 (18%)	36 (16%)

Source: Reviewer

Cell contents are mean (standard deviation) or frequency (percent)

Figure 13: Patient Withdrawal over Time in Study 29 (Source: Reviewer)

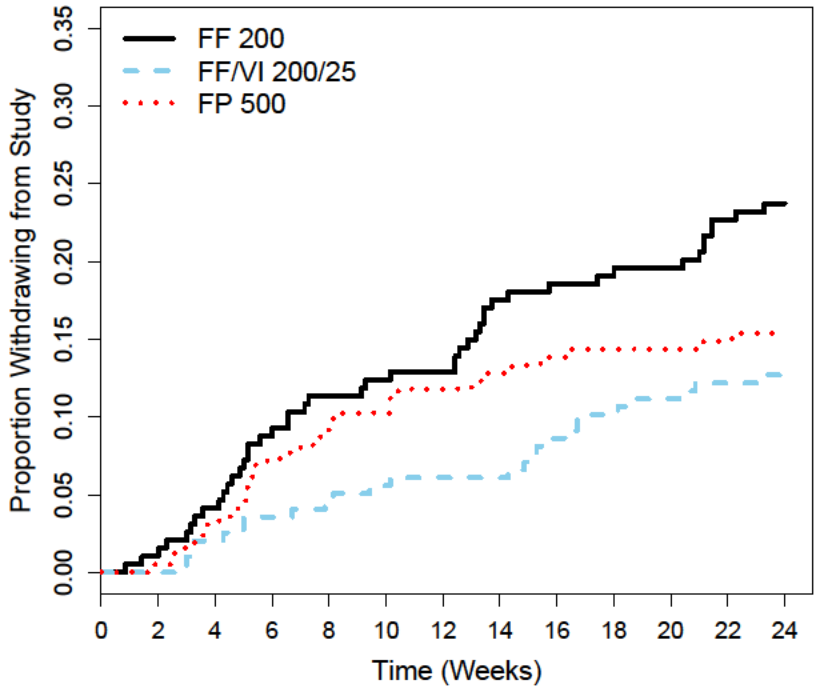


Figure 14: Patient Withdrawal over Time in Study 96 (Source: Reviewer)

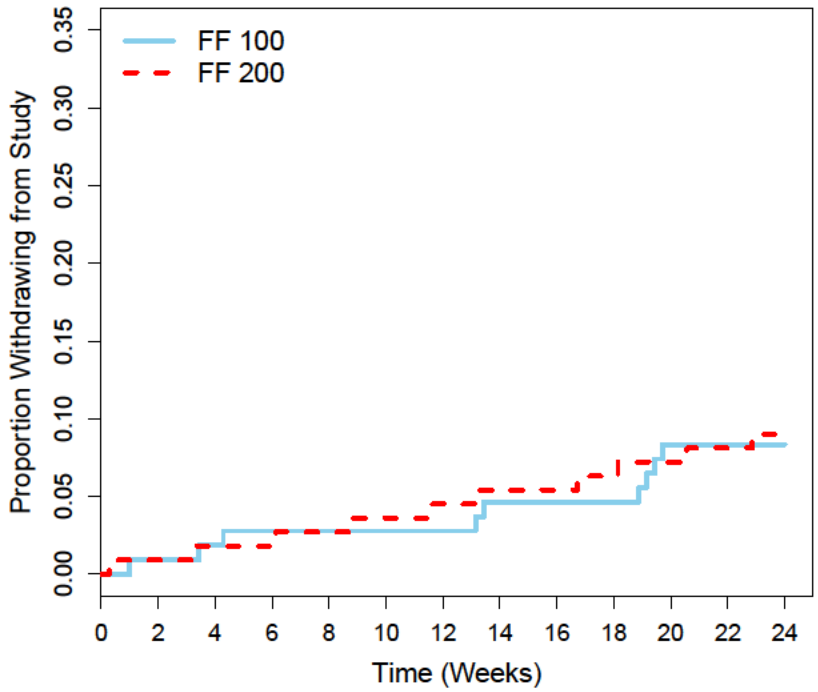


Table 18: Patient Dropout, by Reason for Withdrawal, in Study 29

	FF 200	FF/VI 200/25	FP 500	Overall
Completed Study	146 (75%)	169 (86%)	161 (83%)	476 (81%)
Withdrew from Study	48 (25%)	28 (14%)	34 (17%)	110 (19%)
Adverse event	3 (2%)	7 (4%)	2 (1%)	12 (2%)
Investigator discretion	4 (2%)	8 (4%)	1 (1%)	13 (2%)
Lack of efficacy	21 (11%)	6 (3%)	18 (9%)	45 (8%)
Lost to follow-up	2 (1%)	0 (0%)	1 (1%)	3 (1%)
Protocol deviation	5 (3%)	3 (2%)	5 (3%)	13 (2%)
Withdrew consent	13 (7%)	4 (2%)	7 (4%)	24 (4%)

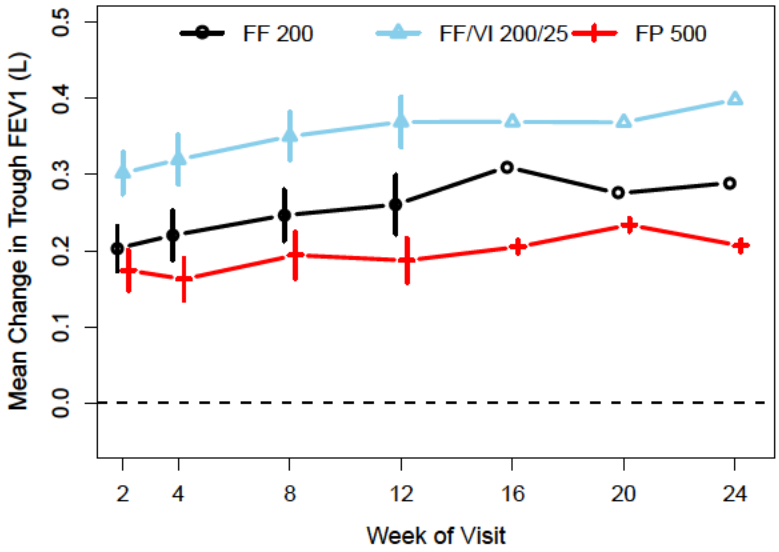
Source: Reviewer

Table 19: Patient Dropout, by Reason for Withdrawal, in Study 96

	FF 100	FF 200	Overall
Completed Study	96 (89%)	100 (90%)	196 (89%)
Withdrew from Study	12 (11%)	11 (10%)	23 (11%)
Adverse event	2 (2%)	2 (2%)	4 (2%)
Investigator discretion	2 (2%)	1 (1%)	3 (1%)
Lack of efficacy	2 (2%)	1 (1%)	3 (1%)
Lost to follow-up	0 (0%)	1 (1%)	1 (0%)
Protocol deviation	2 (2%)	3 (3%)	5 (2%)
Withdrew consent	4 (4%)	3 (3%)	7 (3%)

Source: Reviewer

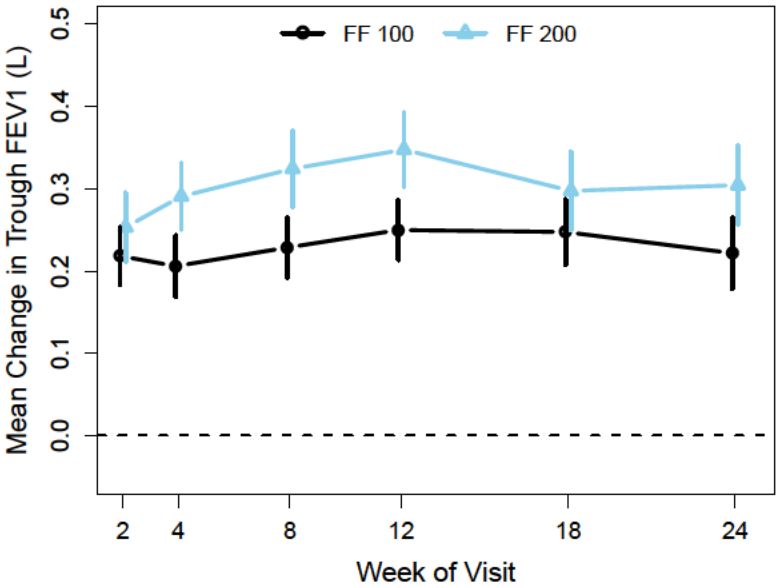
Figure 15: Mean Change from Baseline in Trough FEV_1 over Time in Study 29 Based on Observed Data. Error Bars Represent Plus or Minus One Standard Error. (Source: Reviewer)



Number of Patients with Data

FF 200	185	177	170	163	156	154	147
FF/VI 200/25	188	185	180	177	172	168	163
FP 500	186	181	173	168	165	163	158

Figure 16: Mean Change from Baseline in Trough FEV_1 over Time in Study 96 Based on Observed Data. Error Bars Represent Plus or Minus One Standard Error. (Source: Reviewer)



Number of Patients with Data

FF 100	106	104	104	103	100	95
FF 200	109	108	106	103	102	99

Figure 17: Empirical Distribution Function for Change from Baseline in Trough FEV_1 at 24 Weeks in Study 29 (Source: Reviewer)

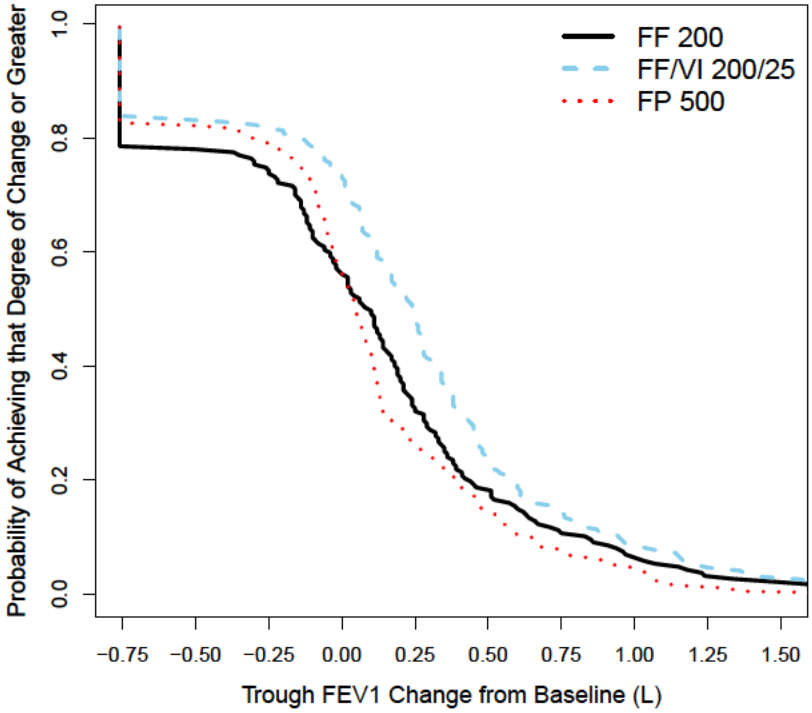


Figure 18: Empirical Distribution Function for Change from Baseline in Trough FEV_1 at 24 Weeks in Study 96 (Source: Reviewer)

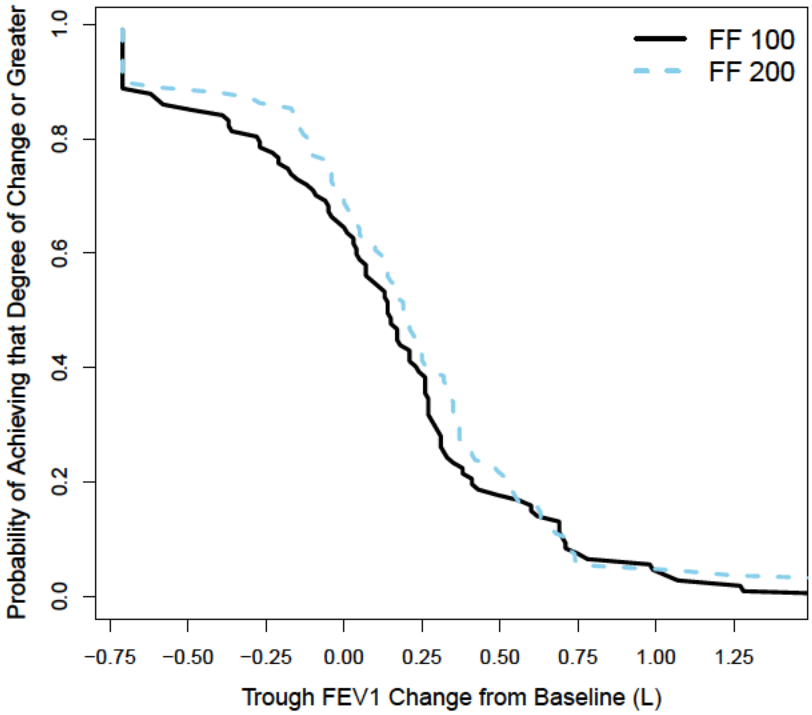


Table 20: Baseline Characteristics in Study 27, Stratified According to Patient Withdrawal

	Completed Study	Withdrew from Study	Overall
N	515	94	609
Female	292 (57%)	61 (65%)	353 (58%)
Age (years)	39.8 (16.5)	39.4 (17.1)	39.7 (16.6)
Age Group (years)			
<18	66 (13%)	16 (17%)	82 (13%)
18-65	419 (81%)	71 (76%)	490 (80%)
≥ 65	30 (6%)	7 (7%)	37 (6%)
Race			
White	438 (85%)	73 (78%)	511 (84%)
Black	38 (7%)	5 (5%)	43 (7%)
Asian	36 (7%)	15 (16%)	51 (8%)
Other	3 (1%)	1 (1%)	4 (1%)
Hispanic/Latino	19 (4%)	18 (19%)	37 (6%)
Weight (kg)	75.6 (17.9)	75.2 (18.3)	75.6 (17.9)
Height (cm)	168.2 (9.3)	166.1 (8.2)	167.9 (9.2)
FEV_1	2.3 (0.6)	2.2 (0.6)	2.3 (0.6)
FEV_1 % Predicted	70.5 (10.9)	70.3 (11.6)	70.4 (11.0)
Morning PEF	367.9 (117.8)	323.2 (88.3)	361.1 (114.8)
Evening PEF	377.2 (118.2)	336.8 (91.3)	371.1 (115.4)
% Rescue-free Days	20.0 (30.4)	21.0 (34.8)	20.2 (31.1)
% Rescue-free Nights	59.3 (37.7)	29.8 (38.0)	54.8 (39.2)
% 24-hour Rescue-free Periods	14.1 (28.2)	16.2 (32.3)	14.4 (28.8)
Daily Rescue Use	2.8 (2.0)	4.2 (3.2)	3.0 (2.3)
% Symptom-free Days	8.7 (19.7)	8.6 (21.3)	8.7 (20.0)
% Symptom-free Nights	19.1 (31.6)	19.1 (34.5)	19.1 (32.0)
% 24-hour Symptom-free Periods	4.7 (14.8)	5.3 (15.5)	4.8 (14.9)
24-hour Symptom Score	2.7 (1.2)	3.1 (1.6)	2.7 (1.3)
Duration of Asthma (years)	11.9 (11.6)	13.5 (9.7)	12.1 (11.4)
At USA site	161 (31%)	35 (37%)	196 (32%)

Source: Reviewer

Cell contents are mean (standard deviation) or frequency (percent)

Table 21: Baseline Characteristics in Study 59, Stratified According to Patient Withdrawal

	Completed Study	Withdrew from Study	Overall
N	255	88	343
Female	148 (58%)	55 (62%)	203 (59%)
Age (years)	40.0 (16.3)	42.3 (16.9)	40.6 (16.5)
Age Group (years)			
<18	39 (15%)	7 (8%)	46 (13%)
18-65	205 (80%)	78 (89%)	283 (83%)
≥65	11 (4%)	3 (3%)	14 (4%)
Race			
White	204 (80%)	66 (76%)	270 (79%)
Black	45 (18%)	19 (22%)	64 (19%)
Asian	5 (2%)	0 (0%)	5 (1%)
Other	0 (0%)	2 (2%)	2 (1%)
Hispanic/Latino	10 (4%)	6 (7%)	16 (5%)
Weight (kg)	80.3 (23.9)	79.4 (21.1)	80.0 (23.2)
Height (cm)	168.1 (10.3)	167.7 (9.0)	168.0 (10.0)
FEV_1	2.4 (0.7)	2.2 (0.7)	2.4 (0.7)
FEV_1 % Predicted	73.5 (10.0)	69.7 (13.3)	72.5 (11.1)
Morning PEF	353.3 (114.8)	334.7 (102.7)	348.7 (112.0)
Evening PEF	368.0 (115.2)	342.0 (103.6)	361.5 (112.8)
% Rescue-free Days	22.1 (31.8)	25.8 (35.0)	23.0 (32.6)
% Rescue-free Nights	49.2 (40.9)	33.7 (36.1)	45.4 (40.2)
% 24-hour Rescue-free Periods	15.7 (27.8)	17.9 (29.6)	16.3 (28.2)
Daily Rescue Use	2.9 (2.3)	4.1 (8.5)	3.2 (4.7)
% Symptom-free Days	11.6 (24.0)	11.3 (23.1)	11.5 (23.7)
% Symptom-free Nights	24.2 (34.4)	16.5 (28.2)	22.3 (33.1)
% 24-hour Symptom-free Periods	7.0 (19.4)	4.1 (13.1)	6.3 (18.0)
24-hour Symptom Score	2.5 (1.2)	2.8 (1.3)	2.5 (1.3)
Duration of Asthma (years)	18.2 (13.7)	18.6 (15.0)	18.3 (14.0)
At USA site	145 (57%)	52 (59%)	197 (57%)

Source: Reviewer

Cell contents are mean (standard deviation) or frequency (percent)

References

[1] Michael G Kenward. The handling of missing data in clinical trials. *Clinical Investigation*, 3:241–250, 2013.

[2] Thomas R Fleming and John H Powers. Biomarkers and surrogate endpoints in clinical trials. *Statistics in Medicine*, 31:2973–2984, 2012.

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

/s/

GREGORY P LEVIN
07/18/2014

DAVID M PETULLO
07/18/2014