Supplementary materials

Supplementary methods

Efficacy endpoints (EU approach)

Primary and secondary endpoints for the EU approach were similar to the US approach, except they were measured over 24 weeks where appropriate. In addition, time to clinically important deterioration (CID), Transition Dyspnoea Index (TDI) focal score and change from baseline in Evaluating Respiratory Symptoms in COPD total score (RS-Total score) score were included as secondary endpoints. Non-inferiority analyses with BFF MDI 320/10 µg versus budesonide/formoterol DPI were performed for the EU approach where applicable (supplementary table S2).

Efficacy assessments

Patients were provided with an electronic diary (eDiary) at screening after being trained in its correct use. This was completed twice-daily to record time of study drug administration; daily symptoms using the Evaluating Respiratory Symptoms in COPD (E-RS™: COPD) measure and the use of any rescue medication (total number of puffs). Study drug compliance was checked at all visits, and any issues identified were documented in the appropriate study files.

Forced expiratory spirometry manoeuvres for forced expiratory volume at 1 second (FEV_1) were assessed with a spirometer that met or exceeded minimum performance recommendations of the American Thoracic Society. Spirometry was conducted 60 and 30 minutes prior to study drug administration on Day 1, and at each visit during treatment.

Patients completed the Baseline Dyspnoea Index (BDI) questionnaire prior to study drug administration on Day 1 and the TDI questionnaire at each post-randomisation visit including the treatment discontinuation/withdrawal visit. Patients completed the St. George's Respiratory Questionnaire (SGRQ) prior to study drug administration on Day 1 and at each post-randomisation visit including the treatment discontinuation/withdrawal visit.

A COPD exacerbation was defined as a change in usual symptoms beyond normal day-to-day variation that had an acute onset, lasted ≥ 2 days, and may have required a change in regular medication. The change in symptoms must have included ≥ 1 major symptom (dyspnoea or change in sputum volume or colour), and ≥ 1 minor symptom (cough, wheeze, sore throat, rhinorrhoea, nasal congestion or fever without another cause). If an event which did not meet all these criteria was defined as a COPD exacerbation, justification had to be provided by the investigator. A CID was defined as: a decrease from baseline in trough FEV_1 of ≥ 100 mL; an increase from baseline in SGRQ total score of ≥ 4; a TDI focal score of ≤ -1; or a treatment-emergent moderate/severe COPD exacerbation up to Week 24. The onset of action for budesonide (BD)/formoterol fumarate dihydrate (FF) metered dose inhaler (BFF MDI) was defined as the first time point where the difference from BD MDI for change from baseline in FEV_1 was statistically significant.

Safety evaluation

The following criteria were required for the diagnosis of pneumonia: clinical diagnosis by the investigator; compatible chest imaging obtained within 14 days of diagnosis; treatment with antibiotics and/or appropriate antiviral or antifungal agents; and ≥ 2 of increased cough, increased sputum purulence or production, adventitious breath sounds of auscultation, dyspnoea or tachypnoea, fever, elevated white blood cells counts or hypoxemia. MACE were defined as cardiovascular death, non-fatal myocardial infarction or non-fatal stroke.

Statistical analysis

The per-protocol (PP) population was a subset of the intent-to-treat population, defined as all patients with post-randomisation data obtained prior to any major protocol deviations. The PP estimand was the effect of treatment on patients with no major protocol deviations, including the use of randomised medication. Analysis was conducted using the PP population for the non-inferiority comparisons of BFF MDI 320/10 µg to budesonide/formoterol dry powder inhaler (DPI) 400/12 µg, with pre-specified non-inferiority margins of −50 mL for change from baseline in morning pre-dose trough FEV_1, −75 mL for FEV_1 area under the curve from 0–4 hours (AUC_{0-4}), −0.75 for TDI focal score, −1.5 for RS-Total score and 10% for SGRQ responders.

Analyses of peak FEV_1, TDI focal score, RS-Total Score, rescue medication (salbutamol sulfate) use and other secondary endpoint-related endpoints used similar repeated measures linear mixed models with their respective baseline measures as covariates. Time to onset of action was analysed using logistic regression, with baseline FEV_1, reversibility to salbutamol, and baseline eosinophil count as continuous covariates, and ICS use at screening as a categorical covariate. Time to CID was analysed using a Cox proportional hazards model, and time to first moderate-to-severe exacerbations was analysed using a negative binomial model. Responder analyses were performed for the SGRQ total score (improvement of ≥4 units) using a logistic regression model. Subgroups were analysed using the same model that was used for the overall analysis.

The Type I error control strategy in the EU approach was similar overall to the strategy in the US approach (supplementary figure S1). Type I error in the EU approach was strictly controlled for the analysis of the primary endpoints in each dose of BFF MDI. If these analyses were statistically significant, then the secondary analyses of BFF MDI versus the monocomponents and budesonide/formoterol DPI were conducted separately as three different families of hypotheses, with Type I error controlled within each family.

A sample size of 2420 patients (660 patients in the BFF MDI 320/10 µg, BFF MDI 160/10 µg and FF MDI 10 µg treatment groups, and 220 patients in the BD MDI 320 µg and budesonide/formoterol DPI 400/12 µg treatment groups) was estimated to provide power estimates of 90% or higher for all treatment comparisons for the two primary endpoints with both the US and EU approaches. All calculations assumed Type I error control at a 2-sided alpha level of 0.05 and 20% dropout rate. A standard deviation (SD) of 200 mL for the change from baseline for morning pre-dose trough FEV_1 and 220 mL for change from baseline in FEV_1 AUC_{0-4} at each visit was assumed. An effective SD for the change over 24 weeks of 157 mL (trough FEV_1) and 200 mL (FEV_1 AUC_{0-4}) was assumed.

Supplementary results

Primary endpoints (EU approach)

BFF MDI 320/10 µg statistically significantly improved least squares mean (LSM) change from baseline in morning pre-dose trough FEV_1 over 24 weeks compared with FF MDI (31 mL; p=0.0016), while the comparison of BFF MDI 160/10 µg versus FF MDI resulted in a numerical improvement (6 mL; p=0.5485; figure 2A; supplementary table S2). BFF MDI

320/10 µg and 160/10 µg statistically significantly improved FEV_1 AUC_{0-4} over 24 weeks compared with BD MDI (LSM difference: 181 mL and 165 mL respectively; both p<0.0001; figure 2B; supplementary table S2). BFF MDI 320/10 µg was non-inferior to budesonide/formoterol DPI for both primary endpoints over 24 weeks. The analyses of the two primary endpoints over 24 weeks were generally comparable between the attributable estimand and the efficacy estimand (supplementary table S2).

Secondary lung function endpoints (EU approach)

Both doses of BFF MDI nominally significantly improved morning pre-dose trough FEV_1 over 24 weeks versus BD MDI (figure 2A; supplementary table S2). BFF MDI 320/10 µg statistically significantly improved peak change from baseline in FEV_1 over 24 weeks versus BD MDI, with findings for BFF MDI 160/10 µg versus BD MDI reaching nominal significance (supplementary table S2).

Dyspnoea and health status endpoints (EU approach)

For the percentage of SGRQ responders over 24 weeks, comparisons for both doses of BFF MDI versus FF MDI and BD MDI were numerically higher. There was a lower risk of a CID event during treatment with both doses of BFF MDI versus FF MDI, with the comparison of BFF MDI 320/10 µg versus FF MDI reaching statistical significance, and the comparison of BFF MDI 160/10 µg versus FF MDI reaching nominal significance. Statistically significant improvements were observed for BFF MDI 320/10 µg versus BD MDI, and nominally significant improvements were observed for BFF MDI 160/10 µg versus both monocomponents for TDI focal score over 24 weeks, while improvements with BFF MDI 320/10 µg versus FF MDI were lower and did not reach significance. Nominally significant improvements were observed for BFF MDI 160/10 µg versus BD MDI for RS-Total Score over 24 weeks. However, the comparisons of BFF MDI 320/10 µg versus BD MDI and FF MDI, and BFF MDI 160/10 µg versus FF MDI did not reach significance.

BFF MDI 320/10 µg was non-inferior to budesonide/formoterol DPI for TDI focal score and RS-Total score. For the percentage of SGRQ responders, the 95% confidence interval (CI) for the treatment difference was not within the pre-specified non-inferiority margin of 10% (difference [95% CI]: −2.00 [−10.18, 6.19]%; supplementary table S2). However, the 95% CI

contained zero meaning that there was insufficient evidence to conclude that there were differences between BFF MDI 320/10 µg and budesonide/formoterol DPI.

Eosinophil subgroups (EU approach)

In the same manner as at Week 24, treatment differences in change from baseline in morning pre-dose trough FEV_1 over 24 weeks between both doses of BFF MDI and FF MDI were numerically slightly greater in the $\geq$150 cells/mm^3 subgroup than in the <150 cells/mm^3 subgroup (supplementary figure S4).

SUPPLEMENTARY TABLE S1 Important changes to protocol after trial commencement

Description of change	Rationale
Time to first moderate/severe COPD exacerbation moved from other efficacy endpoint/objective to become a secondary efficacy endpoint/objective	Exacerbations are an important endpoint and there is sufficient sample size to observe at least a numerical benefit in the form of a trend, even though the study is not optimised to demonstrate statistical significance. Analyses of severe COPD exacerbations will be conducted due to their clinical relevance
Time to first CID moved from other efficacy endpoint to be a secondary efficacy endpoint (EU approach)	Time to CID encompasses lung function, symptomatic benefit, and exacerbations into a single endpoint, and has become an endpoint of interest for COPD trials since this study commenced, as demonstrated by several publications [1-5]. Since CID is a composite of endpoints already being assessed in the study, its addition would not require the patients to complete any additional assessments
RS-Total Score moved from other endpoint to a secondary endpoint (EU approach) in the place of the EXACT Total Score, which is listed as an other endpoint	The E-RS: COPD [6] is an 11-item sub-set of the 14-item EXACT scale, which provides an indication of symptom severity and acute exacerbations, respectively [7]. Hence RS-Total Score is more aligned with the study population, i.e. all patients were symptomatic; baseline COPD Assessment Test score ≥10
Analysis of time to onset of action on Day 1 amended to be a formal comparison to BD MDI	Since BD MDI was not expected to produce bronchodilator effects immediately on Day 1, it was used as an active control in this analysis
Text has been updated with details of the estimands to be included in the efficacy analyses	The efficacy analysis section has been updated using wording around estimands in keeping with current statistical thinking
Type I error control strategy amended to account for new secondary endpoints and the analysis of the primary endpoints under the attributable estimand. The non-inferiority comparisons of BFF MDI 160/10 μg versus BUD/FORM DPI were removed.	New secondary endpoints and the need to include the analysis of the primary endpoint as a first secondary endpoint for the attributable estimand required a new Type I error control strategy.

BD: budesonide; BFF: budesonide/formoterol fumarate dihydrate;
BUD/FORM DPI: budesonide/formoterol dry powder inhaler; CID: clinically important deterioration;
COPD: chronic obstructive pulmonary disease; E-RS: COPD: RS-Total Score: Evaluating Respiratory Symptoms in COPD Total Score; EXACT: EXAcerbations of Chronic pulmonary disease Tool;
MDI: metered dose inhaler.

SUPPLEMENTARY TABLE S2 Primary, secondary and other efficacy endpoints (efficacy estimand, unless stated otherwise; mITT population)

	BFF MDI 320/10 µg (n=655)			BFF MDI 160/10 µg (n=637)	
	versus FF MDI 10 µg (n=644)	versus BD MD 320 µg (n=206)	versus BUD/FORM DPI 400/12 µg[a] (n=219)	versus FF MDI 10 µg (n=644)	versus BD MD 320 µg (n=206)
Primary endpoints					
Change from baseline in morning pre-dose trough FEV_1 (mL) over 24 weeks[b] (EU approach)					
LSM (95% CI)	31 (12, 50)	87[†] (59, 114)	–8 (–37, 21)	6 (–13, 25)	62[†] (34, 90)
p-value	0.0016*	<0.0001[#]	0.5734 NI	0.5485	<0.0001[#]
Change from baseline in FEV_1 AUC_{0-4} (mL) over 24 weeks[c] (EU approach)					
LSM (95% CI)	23 (5, 40)	181 (155, 206)	12 (–14, 38)	7 (–11, 25)	165 (140, 191)
p-value	0.0127[#]	<0.0001*	0.3755 NI	0.4328	<0.0001*
Secondary endpoints					
Change from baseline in morning pre-dose trough FEV_1 (mL) over 24 weeks (EU approach; attributable estimand)[c]					
LSM (95% CI)	32 (12, 51)	86 (58, 114)	See above	7 (–12, 26)	61 (33, 89)
p-value	0.0014*	<0.0001[#]		0.4819	<0.0001[#]
Change from baseline in FEV_1 AUC_{0-4} (mL) over 24 weeks (EU approach; attributable estimand)[c]					
LSM (95% CI)	25 (7, 42)	177 (152, 203)	See above	7 (–10, 25)	160 (134, 186)
p-value	0.0066[#]	<0.0001*		0.4107	<0.0001[#]
Peak change from baseline in FEV_1 (mL) over 24 weeks[c] (EU approach)					
LSM (95% CI)	19 (1, 38)	169 (142, 196)	12 (–15, 39)	5 (–14, 23)	154 (127, 181)
p-value	0.0393[#]	<0.0001*	0.3891[d]	0.6275	<0.0001[#]
Time to first moderate/severe COPD exacerbation (US and EU approach)					
Hazard ratio (95% CI)	0.675 (0.528, 0.863)	0.806 (0.560, 1.162)	1.163 (0.771, 1.755)	0.771 (0.608, 0.977)	0.921 (0.643, 1.319)
p-value[††]	0.0017*	0.2484	0.4719[d]	0.0310[#]	0.6535
Time to CID (EU approach)					
Hazard ratio (95% CI)	0.79 (0.69, 0.89)	0.70 (0.58, 0.83)	1.17 (0.96, 1.43)	0.85 (0.75, 0.96)	0.75 (0.63, 0.90)
p-value[††]	0.0002*	<0.0001[#]	0.1264[d]	0.0093[#]	0.0015[#]

Change from baseline in average daily salbutamol use (puffs per day) over 24 weeks (US and EU approach)[e]					
LSM (95% CI)	−0.22 (−0.46, 0.01)	−0.70 (−1.04, −0.36)	−0.22 (−0.57, 0.12)	−0.17 (−0.41, 0.06)	−0.65 (−0.99, −0.31)
p-value	0.0610	<0.0001*	0.2010[d]	0.1535	0.0002[#]
TDI focal score over 24 weeks[f] (EU approach)					
LSM (95% CI)	0.15 (−0.06, 0.35)	0.53 (0.22, 0.83)	0.06 (−0.25, 0.36)	0.23 (0.02, 0.44)	0.61 (0.31, 0.91)
p-value	0.1676	0.0007*	0.7035 NI	0.0305[#]	<0.0001[#]
Change from baseline in RS-Total Score over 24 weeks[g] (EU approach)					
LSM (95% CI)	−0.17 (−0.59, 0.24)	−0.59 (−1.19, 0.01)	−0.18 (−0.79, 0.43)	−0.40 (−0.82, 0.01)	−0.82 (−1.42, −0.22)
p-value	0.4086	0.0524	0.5606 NI	0.0561	0.0072[#]
Percentage of patients achieving an MCID of ≥4 units in SGRQ total score over 24 weeks[h] (EU approach)					
Difference (95% CI)	2.55 (−3.04, 8.15)	3.64 (−4.39, 11.67)	−2.00 (−10.18, 6.19)	3.96 (−1.67, 9.59)	5.05 (−3.00, 13.10)
p-value	0.3712	0.3764	0.6321	0.1684	0.2223
Time to onset of action as assessed by FEV_1 (mL) on Day 1 (US and EU approach)[i,j]					
Time	NA	5 min	NA	NA	5 min
LSM (95% CI)	NA	132 (113, 151)	NA	NA	126 (108, 145)
p-value	NA	<0.0001*	NA	NA	<0.0001[#]
Other endpoints					
Rate of moderate/severe COPD exacerbations (US and EU approach)					
Rate ratio (95% CI)	0.63 (0.49, 0.82)	0.68 (0.47, 0.99)	1.32 (0.85, 2.06)	0.72 (0.56, 0.92)	0.77 (0.53, 1.11)
p-value	0.0005[#]	0.0433[#]	0.2158[d]	0.0094[#]	0.1634

Pre-specified primary and secondary efficacy comparisons are shaded in light grey. *=statistically significant; #=nominally significant (p<0.05 but not statistically significant after Type I error control or not included in Type I error control). [†]Pre-specified secondary endpoint comparison. [††]p values are "Cox-regression" p-values.

[a]Non-inferiority comparisons were conducted in the PP population; BFF MDI 320/10 μg (n=626); BFF MDI 160/10 μg (n=596); FF MDI 10 μg (n=591); BD MDI 320 μg (n=193); BUD/FORM DPI 400/12 μg (n=202). [b]BFF MDI 320/10 μg (n=627); BFF MDI 160/10 μg (n=615); FF MDI 10 μg (n=613); BD MDI 320 μg (n=190); BUD/FORM DPI 400/12 μg (n=211). [c]BFF MDI 320/10 μg (n=654); BFF MDI 160/10 μg (n=636); FF MDI 10 μg (n=643); BD MDI 320 μg (n=206); BUD/FORM DPI 400/12 μg (n=218). [d]Not evaluated for NI [e]BFF MDI 320/10 μg (n=654); BFF MDI 160/10 μg (n=636); FF MDI 10 μg (n=641); BD MDI 320 μg (n=206); BUD/FORM DPI 400/12 μg (n=218). [f]BFF MDI 320/10 μg (n=618); BFF MDI 160/10 μg (n=607); FF MDI 10 μg (n=605); BD MDI 320 μg (n=185); BUD/FORM DPI 400/12 μg (n=206). [g]BFF MDI 320/10 μg (n=655); BFF MDI 160/10 μg (n=637); FF MDI 10 μg (n=641); BD MDI 320 μg (n=206); BUD/FORM DPI 400/12 μg

(n=218). [h]BFF MDI 320/10 µg (n=649); BFF MDI 160/10 µg (n=635); FF MDI 10 µg (n=640); BD MDI 320 µg (n=204); BUD/FORM DPI 400/12 µg (n=217). [i]The onset of action for BFF MDI was defined as the first timepoint where the difference from BD MDI for change from baseline in FEV_1 was statistically significant. [j]BFF MDI 320/10 µg (n=535); BFF MDI 160/10 µg (n=536); BD MDI 320 µg (n=171).
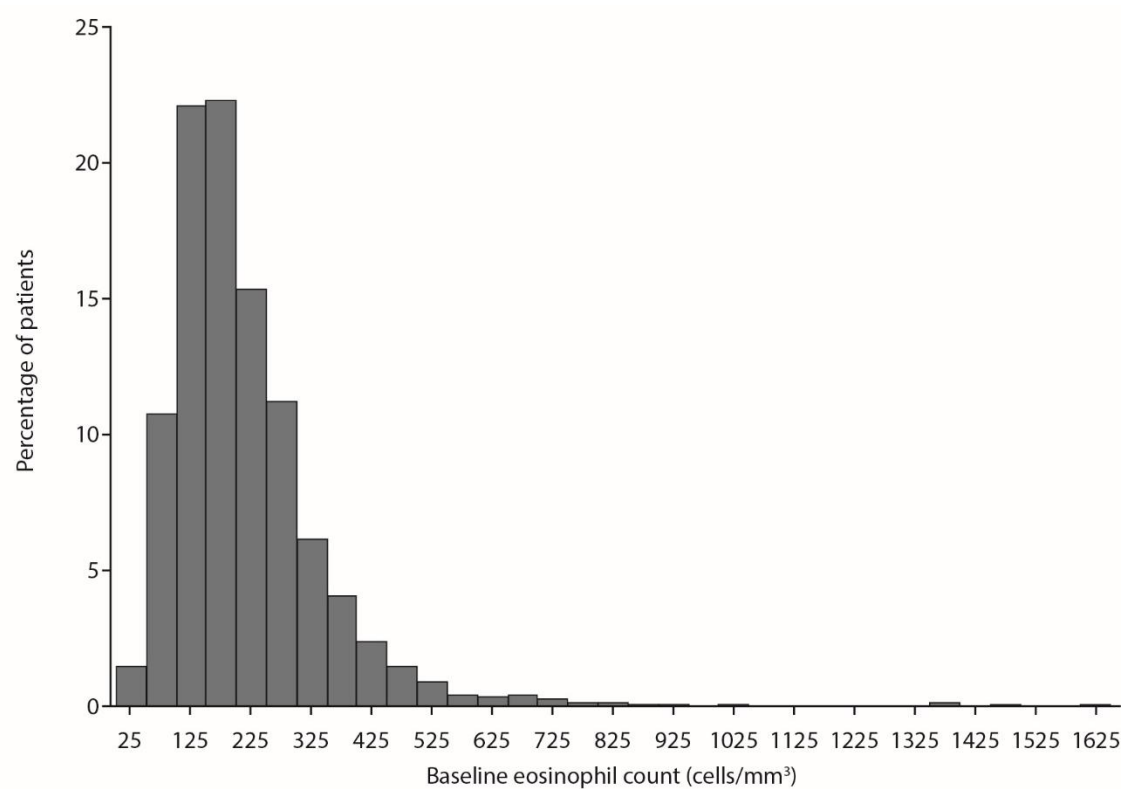
$AUC_{0–4}$: area under curve from 0 to 4 hours; BD: budesonide; BFF: budesonide/formoterol fumarate dihydrate; BUD/FORM DPI: budesonide/formoterol dry powder inhaler; CI: confidence interval; CID: clinically important deterioration; COPD: chronic obstructive pulmonary disease; FEV_1: forced expiratory volume in 1 second; FF: formoterol fumarate dihydrate; LSM: least squares mean; MCID: minimal clinically important difference; MDI: metered dose inhaler; mITT: modified intent-to-treat; NA: not applicable; NI: non-inferior; RS-Total Score: Evaluating Respiratory Symptoms in COPD Total Score; SGRQ: St George's Respiratory Questionnaire; TDI: Transition Dyspnoea Index.

SUPPLEMENTARY FIGURE S1 Type I error control for the US approach
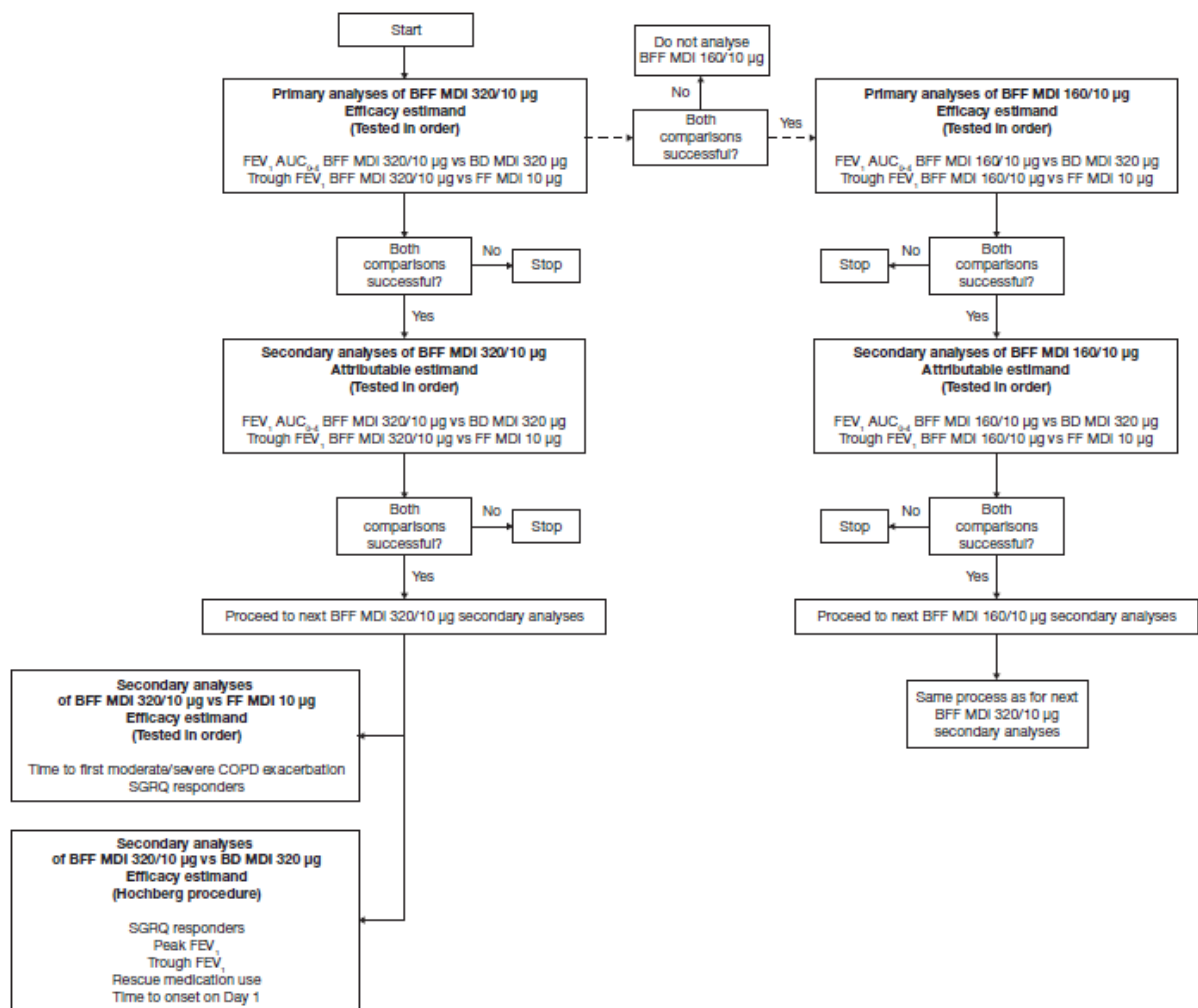
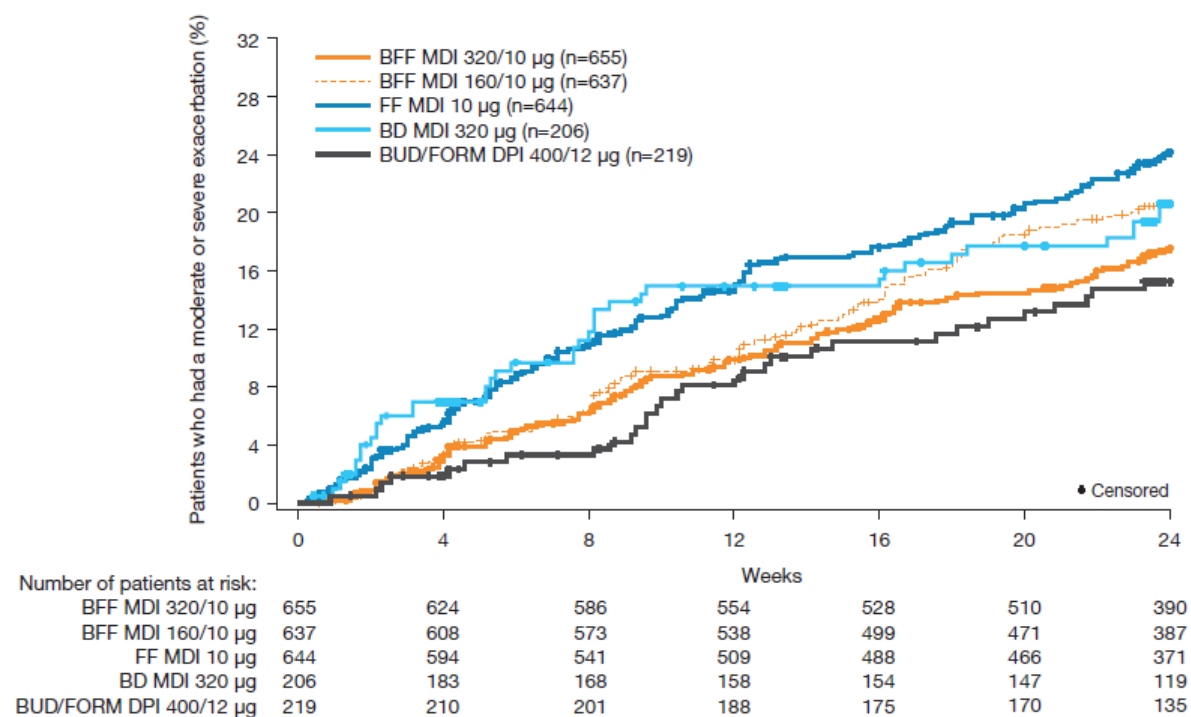AUC$_{0-4}$: area under curve from 0 to 4 hours; BD: budesonide; BFF: budesonide/formoterol fumarate dihydrate; COPD: chronic obstructive pulmonary disease; FEV$_1$: forced expiratory volume in 1 second; FF: formoterol fumarate dihydrate; MDI: metered dose inhaler; SGRQ: St George's Respiratory Questionnaire.

SUPPLEMENTARY FIGURE S2 Baseline distribution of blood eosinophil count (mITT population)


mITT: modified intent-to-treat.

SUPPLEMENTARY FIGURE S3 Time to first moderate/severe COPD exacerbation (mITT population; efficacy estimand)

Number of patients at risk:							
BFF MDI 320/10 µg	655	624	586	554	528	510	390
BFF MDI 160/10 µg	637	608	573	538	499	471	387
FF MDI 10 µg	644	594	541	509	488	466	371
BD MDI 320 µg	206	183	168	158	154	147	119
BUD/FORM DPI 400/12 µg	219	210	201	188	175	170	135

BD: budesonide; BFF: budesonide/formoterol fumarate dihydrate; BUD/FORM DPI: budesonide/formoterol dry powder inhaler; COPD: chronic obstructive pulmonary disease; FF: formoterol fumarate dihydrate; MDI: metered dose inhaler; mITT: modified intent-to-treat.

SUPPLEMENTARY FIGURE S4 Forest plots of primary lung function endpoints by baseline eosinophil count (efficacy estimand; mITT population; EU approach). (A) Change from baseline in morning pre-dose trough FEV_1 over 24 weeks (B) Change from baseline in FEV_1 AUC_{0-4} over 24 weeks
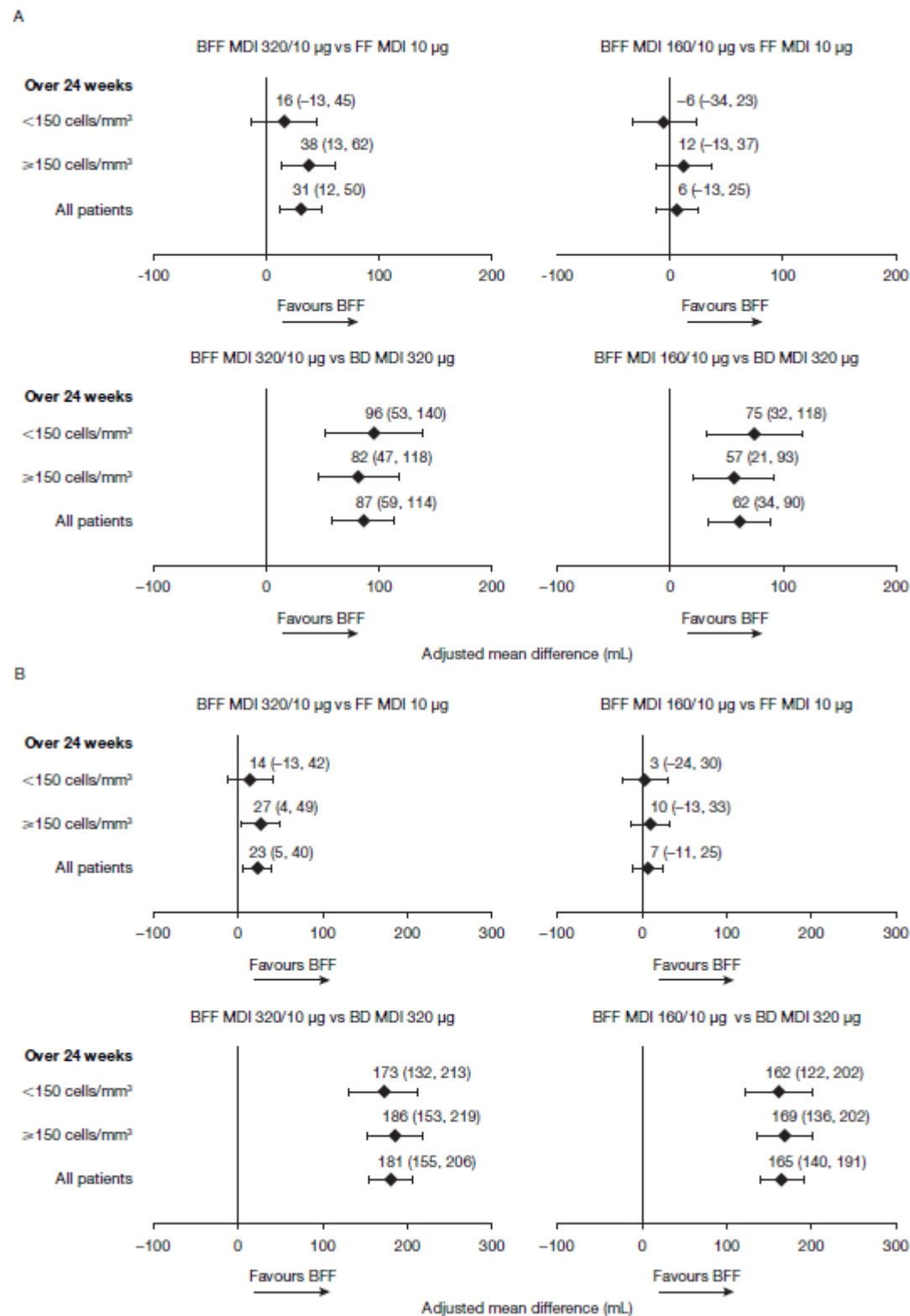
Data represent least squares mean. Error bars represent 95% confidence intervals.

mITT population: <150 cells/mm^3 n=810; ≥150 cells/mm^3 n=1551; all patients, N=2361.

AUC$_{0-4}$: area under the curve from 0 to 4 hours; BD: budesonide; BFF: budesonide/formoterol fumarate dihydrate; FEV$_1$: forced expiratory volume in 1 second; FF: formoterol fumarate dihydrate; MDI: metered dose inhaler; mITT modified intent-to-treat.

References

1. Rabe KF, Martinez FJ, Rodriguez-Roisin R, Fabbri LM, Ferguson GT, Orevillo C, Darken P, Maes A, Martin UJ, Reisner C. LAMA/LABA glycopyrrolate/formoterol fixed-dose combination, delivered using a novel MDI Co-Suspension™ delivery technology reduces risk of clinically important deteriorations in COPD versus placebo and monocomponent MDIs. *Am J Respir Crit Care Med* 2017: 195: A3594.

2. Anzueto AR, Vogelmeier CF, Kostikas K, Mezzi K, Fucile S, Bader G, Shen S, Banerji D, Fogel R. The effect of indacaterol/glycopyrronium versus tiotropium or salmeterol/fluticasone on the prevention of clinically important deterioration in COPD. *Int J Chron Obstruct Pulmon Dis* 2017: 12: 1325-1337.

3. Singh D, D'Urzo AD, Chuecos F, Muñoz A, Garcia Gil E. Reduction in clinically important deterioration in chronic obstructive pulmonary disease with aclidinium/formoterol. *Respir Res* 2017: 18: 106.

4. Singh D, Maleki-Yazdi MR, Tombs L, Iqbal A, Fahy WA, Naya I. Prevention of clinically important deteriorations in COPD with umeclidinium/vilanterol. *Int J Chron Obstruct Pulmon Dis* 2016: 11: 1413-1424.

5. Buhl R, McGarvey L, Korn S, Ferguson GT, Gronke L, Hallmann C, Vob F, Rabe KF, Maltais F. Benefits of tiotropium + olodaterol over tiotropium at delaying clinically significant events in patients with COPD classified as GOLD B. *Am J Respir Crit Care Med* 2016: 193: A6779.

6. Leidy NK, Murray LT. Patient-reported outcome (PRO) measures for clinical trials of COPD: the EXACT and E-RS. *COPD* 2013: 10: 393-398.

7. Leidy NK, Murray LT, Monz BU, Nelsen L, Goldman M, Jones PW, Dansie EJ, Sethi S. Measuring respiratory symptoms of COPD: performance of the EXACT- Respiratory Symptoms Tool (E-RS) in three clinical trials. *Respir Res* 2014: 15: 124.