Respiratory Research

https://doi.org/10.1186/s12931-026-03799-9

Article in Press

Uncovering the heterogeneous effect of inhaled corticosteroids on COPD exacerbations with causal machine learning

Received: 3 November 2025

Accepted: 23 June 2026

Published online: 22 July 2026

Cite this article as: Huts H., Verstraete K., Strypsteen T. *et al.* Uncovering the heterogeneous effect of inhaled corticosteroids on COPD exacerbations with causal machine learning. *Respir Res* (2026). https://doi.org/10.1186/s12931-026-03799-9

Helene Huts, Kenneth Verstraete, Thomas Strypsteen, Iwein Gyselinck, Michaël Staes, Amber Beersaerts, Jeroen Berrevoets, Wim Janssens & Maarten De Vos

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Uncovering the heterogeneous effect of inhaled corticosteroids on COPD exacerbations with causal machine learning

Helene Huts[1, 2], Kenneth Verstraete[2], Thomas Strypsteen[1], Iwein Gyselinck[3], Michaël Staes[2,3], Amber Beersaerts[2], Jeroen Berrevoets[4], Wim Janssens[2, 3, *], Maarten De Vos[1, 5, *]

[1] STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Department of Electrical Engineering (ESAT), KU Leuven, Leuven, Belgium

[2] Laboratory of Respiratory Diseases and Thoracic Surgery (BREATHE), Department of Chronic Diseases and Metabolism (CHROMETA), KU Leuven, Leuven, Belgium

[3] Department of Respiratory Diseases, University Hospitals Leuven, Leuven, Belgium

[4] Department of Informatics, King's College London, Londen, United Kingdom

[5] Woman and Child, Department of Development and Regeneration, KU Leuven, Leuven, Belgium

ARTICLE IN PRESS

* Co-senior authors

Corresponding Author

Wim Janssens, MD, PhD.

ORCID iD: http://orcid.org/0000-0003-1830-2982

Department of Chronic Diseases and Metabolism, ON1bis Herestraat 49 - box 706, 3000 Leuven, Belgium

E-mail: wim.janssens@uzleuven.be

ARTICLE IN PRESS

ORDIC iDs

Helene Huts: http://orcid.org/0000-0002-0617-1316

Kenneth Verstraete: http://orcid.org/0000-0003-3790-417X

Thomas Strypsteen: http://orcid.org/0000-0001-7485-0321

Iwein Gyselinck: https://orcid.org/0000-0002-4068-7228

Michaël Staes: https://orcid.org/0009-0008-4560-9078

Amber Beersaerts: https://orcid.org/0009-0005-8586-2033

Jeroen Berrevoets: https://orcid.org/0000-0002-0396-5655

Wim Janssens: https://orcid.org/0000-0003-1830-2982

Maarten De Vos: https://orcid.org/0000-0002-3482-5145

ARTICLE IN PRESS

Word count

Abstract: 325 words (350 words max)

Main text: 3327 words

ARTICLE IN PRESS

Abstract

Background: High blood eosinophil counts can predict an enhanced response to inhaled corticosteroids (ICS) in patients with chronic obstructive pulmonary disease (COPD), but additional insight in treatment effect heterogeneity is needed to optimize clinical disease management. We investigated if causal machine learning models can detect heterogeneity in the effect of two ICS-containing therapies on both time to first exacerbation and exacerbation rate, and if this approach can identify additional predictors of ICS response.

Methods: Baseline characteristics from patients in the fluticasone furoate/umeclidinium/vilanterol (FF/U/V) and U/V arms of the InforMing the PAthway of COPD Treatment (IMPACT) trial (ClinicalTrials.gov identifier NCT02164513) were used to train a causal survival forest to estimate the effect of fluticasone furoate (an ICS) on time to first exacerbation for each patient. Similarly, a causal forest was trained for exacerbation rate. Results were averaged over 100 train and validation sets in a Monte-Carlo cross-validation approach. The analysis was repeated for a comparison of the FF/V and U/V arms.

Results: This analysis included 4048 FF/U/V, 4034 FF/V and 2025 U/V patients. Significant p-values for AUTOC (area under the targeting operator characteristic) curve indicated that all models ranked patients well according to individualized treatment and that significant heterogeneity was detected in the effect of ICS on both exacerbation outcomes. Respiratory symptoms, reversibility of airflow limitation and lung function, in addition to eosinophils, were identified as important predictors by all models. The contribution of

ARTICLE IN PRESS

additional predictors in determining the effect of ICS was most apparent for patients at the lower end of the eosinophil spectrum.

Conclusions: This hypothesis-generating secondary analysis of IMPACT detected significant heterogeneity in the effect of ICS and indicated that, in addition to eosinophils, other clinical patient characteristics also contribute to the prediction of ICS response on both time to first exacerbation and exacerbation rate. The causal machine learning approach in this analysis can identify treatment effect predictors and patterns of response, to enhance insight in COPD disease phenotypes and heterogeneous treatment effects.

ARTICLE IN PRESS

Key words

COPD Exacerbations

Bronchodilators

Corticosteroids

Eosinophils

Machine learning

Predictive Learning Models

Treatment effect heterogeneity

ARTICLE IN PRESS

Background

Chronic obstructive pulmonary disease (COPD) is a complex disease that is characterized by a notable heterogeneity in clinical features, in underlying biological mechanisms and in the effect of treatments [1- 4]. Optimal clinical management of the disease requires a precision medicine approach, where treatment decisions are made based on the biological and clinical characteristics of each individual patient [1, 2, 4].

Inhaled corticosteroids (ICS) are effective to prevent COPD exacerbations [5-8]. Predefined [6-8] and post-hoc [9-11] analyses have shown that high blood eosinophil counts predict an enhanced effect of ICS treatment. Yet, the commonly proposed cutoffs – 100 and 300 eosinophils per µL, depending on the clinical context [12] – fall well within the commonly observed physiological range for patients with COPD [13], making them broadly inclusive. Given the concern over ICS-associated adverse events [14], additional traits or biomarkers may strengthen confidence in initiating or withholding ICS, especially for exacerbating patients in the lower end of this eosinophil spectrum.

ARTICLE IN PRESS

Causal machine learning (ML) models aim to estimate the effect of a treatment for each patient, based on their individual baseline characteristics [15]. In this context, individualized treatment effect (ITE) is defined as the difference between a patient's outcome with and without a certain treatment [16]. Forest-based causal ML models are flexible models that allow for non-linear interactions between baseline characteristics and a data-driven selection of the most important predictors [15]. Compared to the regression models that were used to investigate exacerbations in several other analyses

ARTICLE IN PRESS

[9-11, 17-21], forest-based methods are fitted to the data in a different way, which creates an interesting opportunity to generate new insights and hypotheses.

The InforMing the PAthway of COPD Treatment (IMPACT; ClinicalTrials.gov: NCT02164513) trial investigated the benefits of a triple therapy of fluticasone furoate (FF, an ICS), umeclidinium (U, a long-acting muscarinic antagonist (LAMA)) and vilanterol (V, a long-acting β_2-agonist (LABA)) compared to dual therapies of FF/V and U/V [6]. In previous analyses of this trial, heterogeneity was detected in the effect of ICS on exacerbation rate with a subgroup analysis for blood eosinophil counts above 150 cells/µL [6], by modelling continuous blood eosinophil counts with fractional polynomials [10] and with a causal forest model that used a wide range of baseline characteristics as input [22]. In addition to exacerbation rate, time to first exacerbation is an outcome that could enhance insights in the effect of ICS treatment [23]. Recent advancements have extended causal ML strategies to allow for the estimation of ITE for right-censored survival outcomes [24].

ARTICLE IN PRESS

In a secondary analysis of the IMPACT trial, we aimed to investigate (1) if a causal ML model can detect heterogeneity in the effect of ICS on time to first exacerbation and (2) if a combination of causal ML models trained for related outcomes –time to first exacerbation and exacerbation rate – can identify a clear profile of strong ICS responders. We first compared the FF/U/V and U/V arms and extended the analysis to evaluate (3) if a pattern of heterogeneous response, similar to that in the FF/U/V versus U/V comparison, can also be observed when comparing the FF/V and U/V arms.

ARTICLE IN PRESS

Methods

Study Participants, Outcomes and Baseline Characteristics

IMPACT was a phase III, 52-week, randomized, double-blind, 3-arm parallel group, global multicentre study. Details of the study design, the study protocol, and main results have been published previously [6]. Briefly, the trial was conducted in 37 countries from June 2014 to July 2017 and included 10,355 patients with symptomatic COPD and a history of exacerbations. Moderate exacerbations were defined per protocol as exacerbations that required treatment with antibiotics or systemic glucocorticoids. Severe exacerbations were defined as exacerbations that resulted in hospitalization or death. Patients who received triple therapy of FF/U/V had significantly lower annual rates of moderate or severe exacerbations (primary outcome) and significantly longer times until first exacerbation (a secondary outcome) compared to patients who received dual therapies of U/V or FF/V.

ARTICLE IN PRESS

The secondary analysis of data from IMPACT in this work was approved by the KU Leuven Institutional Privacy and Ethics review committee (SMEC/PRET, ID: G-2022-5402) and followed guidelines from the Predictive Approaches to Treatment effect Heterogeneity (PATH) [25] and Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD+AI) [26] statements.

A set of 17 patient baseline characteristics was used as input features for the models. The selection was made based on patient characteristics that were previously associated with an enhanced effect of ICS [6-11] or increased exacerbation risk [17-21, 27]. Completeness

ARTICLE IN PRESS

of baseline characteristics was very high. Of the 10,355 patients in the intention-to-treat population, 10,107 patients who had no missing values for any of the selected baseline characteristics were included in the analysis. More details on selected characteristics, missing data and preprocessing steps are provided in Supplementary File S1.

Model Development

To optimally use all available data, patients in the FF/U/V and U/V arms were split 100 times in a 70% training and 30% internal validation set to perform Monte-Carlo cross-validation [28]. Using patient baseline characteristics, a causal survival forest [24] learned to estimate the effect of FF on top of U/V for time to first COPD exacerbation for each patient. This effect was expressed as a difference in restricted mean survival time. Similarly, a causal forest [29] learned to estimate the individualized effect of FF on top of U/V for COPD exacerbation rate. A limit of the causal forest is that it cannot take differences in follow-up time into account when modelling exacerbation rate. To reduce the influence of unrealistically high apparent annual rates caused by very short follow-up times, apparent annual rates higher than 15 were set to 15 for 125 patients in the analysis. Both exacerbation outcomes were estimated with a time horizon of 52 weeks, in line with the follow-up time of the IMPACT trial.

Both forest-based causal machine learning models split patients in the train set based on their baseline characteristics to create a decision-tree-like structure that groups patients with similar treatment effects together. The final model is referred to as a “forest” because it is an ensemble of many such trees. To reduce bias and overfitting, each tree is

ARTICLE IN PRESS

built using a random subset of patients and characteristics, and different patients are used for building the tree structure and for determining the treatment effect in each leaf (a principle called “honesty”). An estimate of individualized treatment effect for an individual patient in the validation set is made by aggregating estimates from all trees. Estimates of treatment effect for each outcome were averaged over 100 internal validation sets to get a final estimate for each patient. R version 4.4.2 (R Foundation for Statistical Computing, Vienna, Austria) and the R *grf* package were used for both models. A detailed description of model development is provided in Supplementary File S1. The same methodology was repeated for FF/V versus U/V.

Model Performance Evaluation

Model performance was evaluated by calculating the area under the targeting operator characteristic (AUTOC) curve [30]. This is a metric that quantifies observed treatment effect gain relative to the observed average treatment effect. Patient are ranked from highest to lowest estimated effect and a curve is constructed by calculating the average treatment effect for increasing top fractions of ranked patients. The overall average treatment effect is subtracted to ensure that the curve ends at zero when the last top fraction includes all patients. A higher area under the curve indicates an overall better ranking of patients from highest to lowest treatment effect (better discrimination). An AUTOC significantly different from zero indicates that significant heterogeneity in treatment effect was detected. To visualize heterogeneity and to evaluate overlap for both modelled outcomes, patients were ranked from highest to lowest estimated effect and split into 5 groups (quintiles). Model calibration was evaluated with calibration curves

ARTICLE IN PRESS

and by comparing the average estimated and observed effects per quintile. A detailed description of evaluation methods is provided in Supplementary File S1.

Results

Study Participants, Outcomes and Baseline Characteristics

This analysis included 4048 FF/U/V patients, 2025 U/V patients and 4034 FF/V patients who had no missing data for any of the baseline characteristics of interest. For patients included in the analysis, average observed outcome values in the FF/U/V versus U/V comparison were significantly in favour of FF/U/V (rate ratio for exacerbation rate: 0.75, $p<0.0001$, hazard ratio for time to first exacerbation: 0.83, $p<0.0001$). The second treatment comparison in this analysis – FF/V versus U/V – was not part of the original trial analysis [6], but may also show the heterogeneous effect of ICS. On average, observed exacerbation rates were significantly in favour of FF/V (rate ratio 0.87, $p=0.0005$), while no effect was observed for time to first exacerbation (hazard ratio 0.98, $p=0.6556$).

ARTICLE IN PRESS

Model Development

The relative importance of patient baseline characteristics was extracted from the models and is shown in Fig. 1. The importance of patient baseline characteristics was determined based on how many times a characteristic was chosen to partition the data in the first splits of the forest models. Blood eosinophil counts were the most important characteristic for all models. Importance was distributed more equally among other characteristics. Some other characteristics important for all models were St. George's

ARTICLE IN PRESS

Respiratory Questionnaire (SGRQ) symptoms score, forced expiratory volume in 1 second (FEV1) reversibility, and FEV1 percent predicted.

Model Performance Evaluation

Table 1 provides an overview of performance metrics for the final ranking of patients and estimates of treatment effect that were obtained by averaging over 100 internal validation sets. Significant p-values for AUTOC indicated good discrimination performance: all models ranked patients well according to their estimated individualized treatment effect and significant treatment effect heterogeneity was detected. An evaluation of model calibration based on a comparison of average observed and estimated treatment effects per quintile (Fig. 2) showed that average estimated effects for the time-to-first-exacerbation model were close to average observed effects per quintile for FF/U/V versus U/V, while for FF/V versus U/V the model underestimated the highest and lowest effects. The exacerbation-rate model tended to overestimate the magnitude of treatment effect for both treatment comparisons.

ARTICLE IN PRESS

Table 1 Model performance

Treatment comparison	**Modelled outcome**	**AUTOC**	
		Mean (95% CI)	**p-value***
FF/U/V-U/V	Time to first exacerbation	11.88 (3.54 to 20.22)	0.0026
	Exacerbation rate	0.39 (0.27 to 0.50)	<0.0001
FF/V-U/V	Time to first exacerbation	22.06 (13.86 to 30.26)	<0.0001

ARTICLE IN PRESS

	Exacerbation rate	0.27 (0.14 to 0.41)	<0.0001

The unit of the area under the targeting operator characteristic (AUTOC) curve depends on the modelled outcome. For time to first exacerbation this is a gain in treatment effect in days of restricted mean survival time. For exacerbation rate this is a gain in treatment effect in exacerbation rate. A significant p-value indicates that the AUTOC was significantly different from zero and that significant treatment effect heterogeneity was detected. FF, Fluticasone Furoate; U, Umeclidinium; V, Vilanterol; CI, Confidence interval. * p-value from a one-sided test

Patients in the FF/U/V versus U/V comparison who were estimated by the models to be in quintile 1 (highest estimated benefit from FF/U/V) showed the largest observed treatment effects in favour of FF/U/V (Fig. 2). For the time-to-first-exacerbation model, this was a mean increase of 43.69 days until first exacerbation if patients received FF/U/V compared to if they would have received U/V therapy. For the exacerbation rate model, this was a mean reduction of 0.80 exacerbations per year. For both outcomes, the observed effects gradually decreased towards quintile 5 (patients estimated to benefit least from FF/U/V). Patient rankings for both outcomes overlapped most in quintiles 1 and 5 (Fig. 3). Patients in quintile 1 or 2 for both outcomes had high blood eosinophil counts, higher SGRQ symptoms scores (more severe respiratory symptoms), higher FEV1 reversibility and lower FEV1 percent predicted values (worse lung function) compared to patients in quintile 4 or 5 for both outcomes (Fig 3 and Table 2).

ARTICLE IN PRESS

Table 2 Characteristics of patients selected to benefit most or least for both modelled outcomes

Characteristic	Level	FF/U/V versus U/V		FF/V versus U/V	
		Q1 and/or Q2 Benefits most from FF/U/V (n = 1906)	Q4 and/or Q5 Benefits least from FF/U/V (n = 1902)	Q1 and/or Q2 Benefits most from FF/V (n = 2002)	Q4 and/or Q5 Benefits least from FF/V (n = 1889)
Subject Level Analysis					
Age (years)		66.04 (7.89)	64.32 (8.61)	64.91 (8.31)	65.41 (8.08)
Sex	Male	1330 (69.78)	1301 (68.40)	1360 (67.93)	1395 (73.85)
	Female	576 (30.22)	601 (31.60)	642 (32.07)	494 (26.15)
Height (cm)		167.74 (9.94)	167.87 (7.97)	168.03 (10.02)	168.79 (7.51)
Weight (kg)		74.20 (20.25)	76.48 (18.01)	77.52 (19.80)	73.11 (18.01)
GOLD grade	1	5 (0.26)	7 (0.37)	4 (0.20)	2 (0.11)
	2	619 (32.48)	991 (52.10)	794 (39.66)	571 (30.23)
	3	960 (50.37)	780 (41.01)	956 (47.75)	949 (50.24)
	4	322 (16.89)	124 (6.52)	248 (12.39)	367 (19.43)
Previous exacerbation count		1.79 (1.01)	1.72 (0.72)	1.84 (0.95)	1.58 (0.77)
Smoking status	Former	1264 (66.32)	1240 (65.19)	1309 (65.38)	1251 (66.23)
	Current	642 (33.68)	662 (34.81)	693 (34.62)	638 (33.77)
Substance Use Analysis					
Number of pack years		47.69 (26.76)	44.02 (26.08)	45.68 (26.74)	48.89 (26.73)
Laboratory Analysis					
Eosinophils (10^9/L)		0.35 [0.25,0.50]	0.10 [0.06,0.15]	0.34 [0.26,0.48]	0.11 [0.06,0.15]
Neutrophils (10^9/L)		4.91 [3.85,6.09]	4.74 [3.76,6.16]	4.92 [3.90,6.14]	4.83 [3.81,6.22]
Pulmonary Function Test Analysis					
FEV1 (L)		1.12 (0.49)	1.38 (0.43)	1.22 (0.49)	1.18 (0.49)
FEV1 reversibility (mL)		132.54 (151.42)	83.18 (133.94)	140.02 (154.61)	69.58 (94.40)
FVC (L)		2.49 (0.83)	2.81 (0.72)	2.59 (0.83)	2.63 (0.79)
FEV1pp (%)		39.64 (14.23)	48.99 (14.25)	42.66 (14.21)	40.88 (15.29)
SGRQ Analysis					
SGRQ Activity score		67.47 (19.28)	59.21 (18.89)	67.62 (19.55)	58.76 (19.89)
SGRQ Impacts score		39.71 (18.80)	30.84 (17.88)	40.98 (19.57)	29.64 (17.53)
SGRQ Symptoms score		69.91 (17.93)	57.70 (17.97)	71.03 (18.02)	57.64 (17.46)

Data are presented as mean (standard deviation), median [interquartile range], and number (%). Q1 and/or Q2, group of patients estimated to be in quintile 1 and/or quintile 2 for both modelled outcomes; Q4 and/or Q5, group of patients estimated to be in quintile 4 and/or quintile 5 for both modelled outcomes; FF, Fluticasone Furoate; U, Umeclidinium; V, Vilanterol; GOLD, Global Initiative for Chronic Obstructive Lung Disease;

ARTICLE IN PRESS

FEV1, Forced Expiratory Volume in 1 second; FEV1pp, Forced Expiratory Volume in 1 second percent predicted; FVC, Forced Vital Capacity; SGRQ, St. George's Respiratory Questionnaire; n, number of patients.

In the FF/V versus U/V comparison, largest treatment effects in favour of FF/V were observed in quintiles 1 and 2 (patients estimated to benefit most from FF/V, Fig. 2). No clear benefit of FF/V over U/V was observed in other quintiles. For the time-to-first-exacerbation model, the observed benefit in quintile 4 and 5 (patients estimated to benefit least from FF/V) is even significantly in favour of U/V. Overlap in patient rankings was strongest in quintile 1 and 5 (Fig. 4). Patients in quintile 1 or 2 for both outcomes had high blood eosinophil counts, higher SGRQ symptoms scores, higher FEV1 reversibility and a reduced lung function, while patients in quintile 4 or 5 for both outcomes had low blood eosinophil counts, lower FEV1 reversibility, lower SGRQ symptoms scores and a reduced lung function (Fig. 4 and Table 2). More detailed performance evaluations and patient characteristics per quintile are provided in Supplementary File S1.

ARTICLE IN PRESS

Discussion

Previous analyses have shown that high blood eosinophil counts can be used as a biomarker to predict an enhanced effect of ICS treatment [6-11], but additional insight in treatment effect heterogeneity is needed to optimize clinical disease management. This analysis used forest-based causal ML models to investigate heterogeneity in the effect of ICS on COPD exacerbations and to search for new insights and hypotheses. The models in

ARTICLE IN PRESS

this internal validation study in the IMPACT trial ranked patients well according to their estimated individualized treatment effect and detected significant heterogeneity in the effect of ICS on both time to first COPD exacerbation and COPD exacerbation rate.

The analysis first focused on investigating heterogeneity in the effect of ICS on top of LABA/LAMA by modelling the FF/U/V arm versus the U/V arm. For both exacerbation outcomes, blood eosinophil counts were selected as most important predictor of treatment effect. Highest eosinophil counts were observed in quintile 1 (patients estimated to benefit most from FF/U/V) and gradually decreased towards quintile 5 (patients estimated to benefit least from FF/U/V). This was in line with the continuous relationship between blood eosinophil counts and ICS effect observed in previous literature [9-10] and the existing hypothesis that patients with eosinophilic inflammation (identified by high blood eosinophil counts) have the highest benefit of ICS treatment [1, 2, 4].

ARTICLE IN PRESS

The models also identified other patient characteristics – SGRQ symptoms score, FEV1 reversibility and FEV1 percent predicted – as important features that contribute to the detection of treatment effect heterogeneity. The influence of these characteristics was strongest for patients with lower values of eosinophils. Overall, patients estimated to benefit most from ICS had higher SGRQ symptoms scores (more severe respiratory symptoms), higher FEV1 reversibility, and reduced lung function values. FEV1 reversibility has been linked to an enhanced ICS effect in a previous subgroup analysis [8], but another

ARTICLE IN PRESS

subgroup analysis did not confirm this finding [7]. Higher SGRQ symptoms scores and worsening lung function have been linked to increased risk of exacerbations [17, 18, 21].

Previous analyses have associated higher exacerbation counts in the previous year with higher exacerbation risk [17, 18, 20]. Model-based feature importance (Fig. 1) did not indicate the count of exacerbations in the previous year as an important predictor of treatment effect, but higher counts were observed for those patients who were estimated to respond best to ICS (Table 2). IMPACT only included patients who had at least one exacerbation in the previous year. Possibly, in patients with a history of exacerbations and for this specific forest-based model structure, the count of previous exacerbations was less informative compared to other characteristics.

The analysis was repeated for a comparison of the FF/V and U/V arms. In this case, the trade-off between the ICS/LABA in the FF/V arm and the LAMA/LABA in the U/V arm emphasizes traits that predict response to adding ICS versus maximizing bronchodilation. Overall, patients estimated to benefit most from FF/V therapy (quintile 1) showed high blood eosinophil counts, higher SGRQ symptoms scores and higher FEV1 reversibility. These values fit with a patient profile of eosinophilic inflammation as responder to ICS that was also present in the FF/U/V versus U/V treatment comparison. Patients estimated to benefit least from FF/V therapy (quintile 5) showed low blood eosinophil counts and lower FEV1 reversibility, higher SGRQ symptoms scores, a reduced lung function and higher Global Initiative for Chronic Obstructive Lung Disease (GOLD) grades.

ARTICLE IN PRESS

Interestingly, the model results in Figs. 2 and 4 suggest that in this modelled trade-off between ICS/LABA and LAMA/LABA, patients in quintile 5 do not benefit from ICS/LABA but clearly benefit from LAMA/LABA therapy. In patients who present with reduced lung function but no eosinophilic inflammation, the effect of the LAMA on top of LABA therapy may be more beneficial for preventing future exacerbations than adding ICS. For these patients, the current GOLD guidelines [12], that recommend an initial treatment with LAMA/LABA are correct. On the other hand, in patients in GOLD group E for whom LABA/LAMA is the initial standard, this analysis confirms eosinophils as the main predictor for ICS response but also suggests that eosinophils are not the only predictor. Additional characteristics (like symptoms score, lung function and reversibility of airflow) contribute to ICS response prediction at the lower end of the eosinophil spectrum. Including other characteristics for these patients may therefore improve the allocation of ICS.

The models in this analysis aimed to investigate the heterogeneous effect of ICS, but several factors may have influenced the observed results. Firstly, the findings in this internal validation analysis may be very specific to the inclusion criteria of the IMPACT trial, which only included patients with symptomatic COPD and a history of exacerbations. Patients who were taking ICS before the start of the trial also had to abruptly quit this treatment if they were randomized to the LAMA/LABA arm. Part of the observed effect may therefore be due to ICS withdrawal.

While both models succeeded well in ranking patients, model calibration was slightly less accurate in the FF/V versus U/V comparison, possibly because estimating the trade-off

ARTICLE IN PRESS

between ICS/LABA and LAMA/LABA is more complex than estimating the add-on effect of ICS in the FF/U/V versus U/V comparison. Calibration was also overestimated for exacerbation rate, likely because causal forests cannot take differences in follow-up time into account. Capping apparent rates at 15 reduces the bias caused by some very short follow-up times, but an extension of the causal forest that corrects for different follow-up times could improve results. When focusing on a single model, the causal survival forest may be preferred, since it was developed specifically for right-censored survival outcomes and does not encounter the same issue. Future analyses could investigate if overlap between the two outcomes can be improved or if slight differences in models are inherently present due to differences in pathophysiology and underlying treatment mechanisms.

This analysis used internal validation and all generated insights are therefore hypothesis generating. The Monte-Carlo cross-validation approach with 100 train and internal validation sets, and the combination of results from two related exacerbation outcomes and two treatment comparisons increases confidence in the generated hypotheses, but further evidence from external validation studies and prospective trials should be gathered before additions to treatment allocation guidelines are made.

In the long run, a well-validated causal ML model could be turned into a tool that supports treatment decisions by estimating a clear and nuanced treatment effect for each individual patient. Future analyses could expand the number of included features and investigate if treatment effect is also influenced by characteristics that are less known

ARTICLE IN PRESS

from previous literature. In addition to time to first exacerbation and exacerbation rate, a model for ICS treatment of patient with COPD could also include other outcomes, like trough FEV1 and changes in SGRQ, and adverse events like pneumonia, which is a notable concern for ICS treatment [14]. The development of modelling strategies that combine data from randomized controlled trials with observational data also holds potential to further improve models and enhance the generalizability of conclusions [31].

Conclusions

This analysis on data from the large IMPACT trial demonstrated a forest-based causal machine learning approach that used a selection of baseline characteristics to estimate the effect of ICS treatment on COPD exacerbations for each individual patient. We combined a novel causal survival forest to model individualized effect on time to first exacerbation with a causal forest for exacerbation rate and investigated both exacerbation outcomes in a comparison of two different ICS-containing therapies versus a therapy of only bronchodilators. In this internal validation study, models ranked patients well according to their treatment effect and detected significant heterogeneity in the effect of ICS. All models identified blood eosinophil counts as the most important predictor of treatment effect. Other characteristics that consistently ranked high in importance were respiratory symptoms, reversibility of airflow limitation and lung function. The contribution of these other characteristics in determining the effect of ICS was most apparent at the lower end of the eosinophil spectrum. This analysis illustrates how causal machine learning models can generate hypotheses to identify new predictors

ARTICLE IN PRESS

ARTICLE IN PRESS

and patterns of response, and to enhance insight in complex disease phenotypes and heterogeneous treatment effects.

List of abbreviations

AUTOC	Area under the targeting operator characteristic
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
FEV1	Forced expiratory volume in 1 second
FEV1pp	Forced expiratory volume in 1 second percent predicted
FF	Fluticasone furoate
FVC	Forced vital capacity
GOLD	Global Initiative for Chronic Obstructive Lung Disease
HR	Hazard ratio
ICS	Inhaled corticosteroids
IMPACT	InforMing the PAthway of COPD Treatment
ITE	Individualized treatment effect
LABA	Long-acting β_2-agonist
LAMA	Long-acting muscarinic antagonist
ML	Machine learning
Q	Quintile
RMST	Restricted mean survival time
RR	Rate ratio
SGRQ	St. George's Respiratory Questionnaire
U	Umeclidinium
V	Vilanterol

ARTICLE IN PRESS

ARTICLE IN PRESS

Declarations

Ethics approval and consent to participate

This study was approved by the KU Leuven Institutional Privacy and Ethics review committee (SMEC/PRET, ID: G-2022-5402). The requirement for informed consent was waived due to the retrospective nature of the study design.

Consent for publication

Not applicable

Availability of data and materials

The anonymized individual participant data of IMPACT (Sponsor ID CTT116855) used in this study was provided upon request by www.ClinicalStudyDataRequest.com with the approval of GlaxoSmithKline Research & Development Ltd. The data are available from www.ClinicalStudyDataRequest.com upon request.

Competing interests

HH received research funding from Chiesi, IG received payment for a presentation from Chiesi and a travel grant from Astra Zeneca. WJ received grants or contracts from GSK and Chiesi, consulting fees from GSK, Sanofi, Roche and Astra Zeneca, payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events

ARTICLE IN PRESS

from GSK, Sanofi, Chiesi and Astra Zeneca. MDV received research funding from Chiesi. The other authors have nothing to disclose.

Funding

This research work was funded by a Strategic Basic Research Fellowship from the Research Foundation – Flanders, Belgium (FWO) for HH (FWO project number 1S30225N); by a research grant from the Research Foundation – Flanders, Belgium (FWO) for MDV and WJ (Research Project: "Artificial Intelligence (AI) for data-driven personalised medicine", FWO project number G0C9623N); by the Horizon Europe 2022 Research and Innovation Program, HORIZON-HLTH-2022-IND-13: "Privacy compliant health data as a service for AI development (PHASE IV AI)", funded by the European Union, under Grant Agreement #101095384; by the KU Leuven Special Research Fund (BOF): "Digital technologies and machine learning models for diagnosis and treatment of chronic respiratory diseases" (BOF project number FKO/24/001); and received funding from the Flemish Government (AI Research Program) and the KU Leuven Chiesi Chair on Phenotyping and individualized therapy in COPD. MDV is affiliated to Leuven.AI - KU Leuven institute for AI, B-3000, Leuven, Belgium.

ARTICLE IN PRESS

Authors' contributions

HH, KV, JB, WJ and MDV contributed to the conceptualization and design of this work. HH, KV, TS, IG, MS, AB, WJ and MDV acquired, analysed and/or interpreted the data. HH drafted the manuscript. All authors critically revised the manuscript for important intellectual content and approved the final version.

ARTICLE IN PRESS

Acknowledgements

Not applicable

Authors' information

None

Prior abstract publication/presentation

This work was previously presented as two abstracts at the ERS Congress 2025.

- Huts H, Verstraete K, Strypsteen S, et al. Investigating individual effects of ICS on COPD exacerbations with causal machine learning. Eur Respir J. 2025;66: Suppl. 69, PA5694. Presented at the European Respiratory Society (ERS) Congress 2025, Amsterdam, 27 Sep 2025 – 01 Oct 2025. doi: 10.1183/13993003.congress-2025.PA5694
- Huts H, Verstraete K, Strypsteen S, et al. Causal machine learning to investigate heterogeneity in time to first COPD exacerbation for ICS/LABA versus LAMA/LABA. Eur Respir J. 2025;66: Suppl. 69, PA5691. Presented at the European Respiratory Society (ERS) Congress 2025, Amsterdam, 27 Sep 2025 – 01 Oct 2025. doi: 10.1183/13993003.congress-2025.PA5691

ARTICLE IN PRESS

ARTICLE IN PRESS

References

1. Agustí A, Bafadhel M, Beasley R, et al. Precision medicine in airway diseases: moving to clinical practice. Eur Respir J. 2017;50(4):1701655. doi: 10.1183/13993003.01655-2017.
2. Agusti A, Bel E, Thomas M, et al. Treatable traits: toward precision medicine of chronic airway diseases. Eur Respir J. 2016;47(2):410–419. doi: 10.1183/13993003.01359-2015.
3. Barnes PJ. Inflammatory endotypes in COPD, Allergy. 2019;74(7):1249–1256. doi: 10.1111/all.13760.
4. David B, Bafadhel M, Koenderman L, De Soyza A. Eosinophilic inflammation in COPD: from an inflammatory marker to a treatable trait. Thorax. 2021;76(2):188–195. doi: 10.1136/thoraxjnl-2020-215167.
5. Ferguson GR, Rabe KF, Martinez FJ, et al. Triple therapy with budesonide/glycopyrrolate/formoterol fumarate with co-suspension delivery technology versus dual therapies in chronic obstructive pulmonary disease (KRONOS): a double-blind, parallel-group, multicentre, phase 3 randomised controlled trial. Lancet Respir Med. 2018;6(10):747–758. doi: 10.1016/S2213-2600(18)30327-8.
6. Lipson DA, Barnhart F, Brealey N, et al. Once-daily single-inhaler triple versus dual therapy in patients with COPD. N Engl J Med. 2018;378(18):1671–1680. doi: 10.1056/NEJMoa1713901.
7. Papi A, Vestbo J, Fabbri L, et al. Extrafine inhaled triple therapy versus dual bronchodilator therapy in chronic obstructive pulmonary disease (TRIBUTE): a

ARTICLE IN PRESS

double-blind, parallel group, randomised controlled trial. Lancet. 2018;391(10125):1076–1084. doi: 10.1016/S0140-6736(18)30206-X.

8. Rabe KF, Martinez FJ, Ferguson GT, et al. Triple Inhaled Therapy at Two Glucocorticoid Doses in Moderate-to-Very-Severe COPD. N Engl J Med. 2020;383(1):35–48. doi: 10.1056/NEJMoa1916046.
9. Bafadhel M, Peterson S, De Blas MA, et al. Predictors of exacerbation risk and response to budesonide in patients with chronic obstructive pulmonary disease: a post-hoc analysis of three randomised trials. Lancet Respir Med. 2018;6(2):117–126. doi: 10.1016/S2213-2600(18)30006-7.
10. Pascoe S, Barnes N, Brusselle G, et al. Blood eosinophils and treatment response with triple and dual combination therapy in chronic obstructive pulmonary disease: analysis of the IMPACT trial. Lancet Respir Med. 2019;7(9):745–756. doi: 10.1016/S2213-2600(19)30190-0.
11. Pascoe S, Locantore N, Dransfield MT, Barnes NC, Pavord ID. Blood eosinophil counts, exacerbations, and response to the addition of inhaled fluticasone furoate to vilanterol in patients with chronic obstructive pulmonary disease: a secondary analysis of data from two parallel randomised controlled trials. Lancet Respir Med. 2015;3(6):435–442. doi: 10.1016/S2213-2600(15)00106-X.
12. GOLD: 2025 Global Strategy for Prevention, Diagnosis and Management of COPD. www.goldcopd.org. Accessed June 22 2025.
13. Kolsum U, Southworth T, Jackson N, Singh D. Blood eosinophil counts in COPD patients compared to controls. Eur Respir J. 2019;54(4):1900633. doi: 10.1183/13993003.00633-2019.

ARTICLE IN PRESS

14. Heerfordt CK, Eklöf J, Sivapalan P, et al. Inhaled Corticosteroids in Patients with Chronic Obstructive Pulmonary Disease and Risk of Acquiring Streptococcus Pneumoniae Infection. A Multiregional Epidemiological Study. Int J Chron Obstruct Pulmon Dis. 2023;18:373–384. doi: 10.2147/COPD.S386518.
15. Feuerriegel S, Frauen D, Melnychuk V, et al. Causal machine learning for predicting treatment outcomes. Nat Med. 2024;30(4):958–968. doi: 10.1038/s41591-024-02902-1.
16. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol. 1974;66(5):688-701. doi: 10.1037/h0037350.
17. Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. N Engl J Med. 2010;363(12):1128–1138. doi: 10.1056/NEJMoa0909883.
18. Jenkins CR, Celli B, Anderson JA, et al. Seasonality and determinants of moderate and severe COPD exacerbations in the TORCH study. Eur Respir J. 2012;39(1):38-45. doi: 10.1183/09031936.00194610.
19. Lonergan M, Dicker AJ, Crichton ML, et al. Blood neutrophil counts are associated with exacerbation frequency and mortality in COPD. Respir Res. 2020;21(1):166. doi: 10.1186/s12931-020-01436-7.
20. Make BJ, Eriksson G, Calverley PM, et al. A score to predict short-term risk of COPD exacerbations (SCOPEX). Int J Chron Obstruct Pulmon Dis. 2015;10:201–209. doi: 10.2147/COPD.S69589.

ARTICLE IN PRESS

ARTICLE IN PRESS

21. Seemungal TA, Donaldson GC, Paul EA, Bestall JC, Jeffries DJ, Wedzicha JA. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 1998;157(5 Pt 1):1418–1422. doi: 10.1164/ajrccm.157.5.9709032.

22. Verstraete K, Gyselinck I, Huts H, et al. Estimating individual treatment effects on COPD exacerbations by causal machine learning on randomised controlled trials. Thorax. 2023;78(10):983–989. doi: 10.1136/thorax-2022-219382.

23. Gyselinck I, Huts H, De Vos M, Janssens W. Time T-to understand treatment response. Lancet Respir Med. 2025;13(6):482-484. doi: 10.1016/S2213-2600(25)00118-3.

24. Cui Y, Kosorok MR, Sverdrup E, Wager S, Zhu R. Estimating heterogeneous treatment effects with right-censored data via causal survival forests. Journal of the Royal Statistical Society Series B: Statistical Methodology. 2023;85(2):179–211. doi: 10.1093/jrsssb/qkac001.

25. Kent DM, Paulus JK, van Klaveren D, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement. Ann Intern med. 2020;172(1):35–45. doi: 10.7326/M18-3667.

26. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. doi: 10.1136/bmj-2023-078378.

27. Bhowmik A, Seemungal TA, Sapsford RJ, Wedzicha JA. Relation of sputum inflammatory markers to symptoms and lung function changes in COPD exacerbations. Thorax. 2000;55(2):114–120. doi: 10.1136/thorax.55.2.114.

ARTICLE IN PRESS

28. Xu Q-S, Liang Y-Z. Monte Carlo cross validation. Chemometrics and Intelligent Laboratory Systems. 2001;56(1):1–11. doi: 10.1016/S0169-7439(00)00122-2.

29. Athey S, Tibshirani J, Wager S. Generalized random forests. The Annals of Statistics. 2019;47(2):1148–1178. doi: 10.1214/18-AOS1709.

30. Yadlowsky S, Fleming S, Shah N, Brunskill E, Wager S. Evaluating Treatment Prioritization Rules via Rank-Weighted Average Treatment Effects. J Am Stat Assoc. 2024;120(549):38–51. doi: 10.1080/01621459.2024.2393466.

31. Colnet B, Mayer I, Chen G, et al. Causal Inference Methods for Combining Randomized Trials and Observational Studies: A Review. Statistical science. 2024;39(1):165–191. doi: 10.1214/23-STS889.

ARTICLE IN PRESS

Supplementary Information

File name: Supplementary File S1

File format: Portable document format (pdf) .pdf

Title of the data: Supplementary File S1

Description of the data: A structured pdf file with additional details on the methodology used in this study, and tables and figures with additional results.

ARTICLE IN PRESS

Figure legends

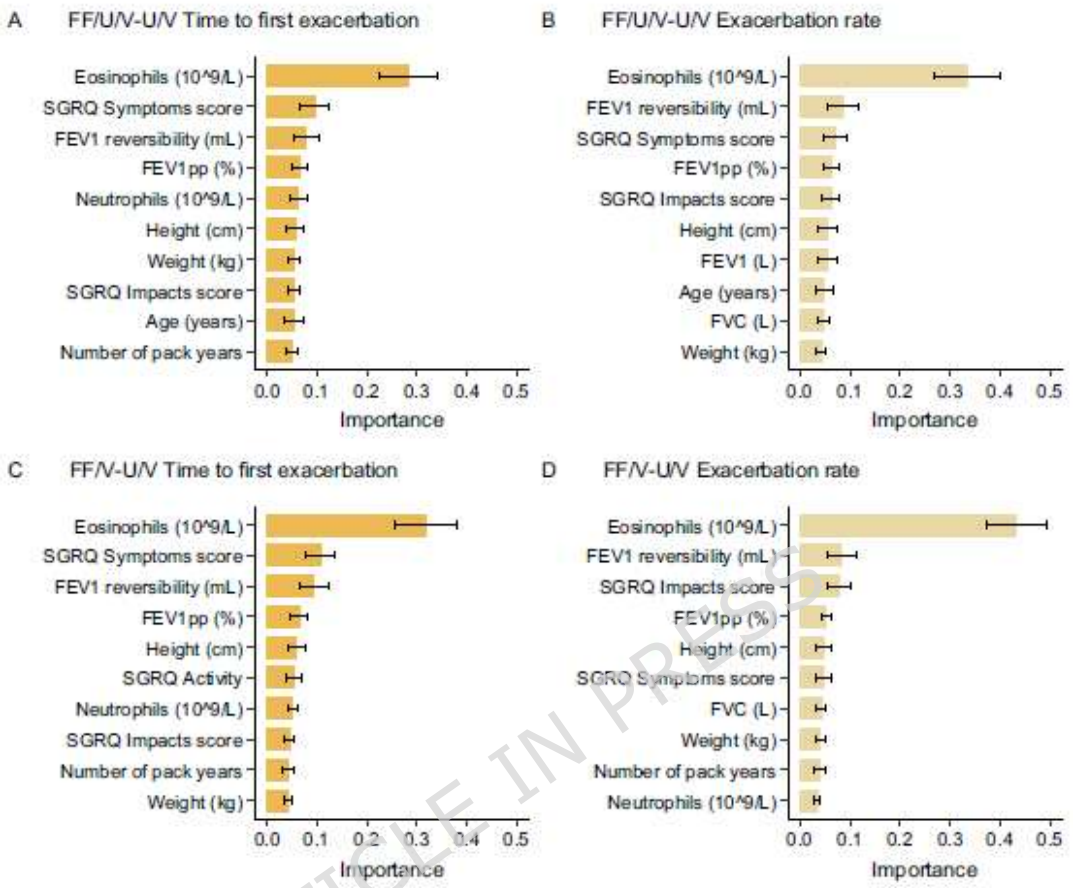


Fig. 1 Model-based feature importance

Importance is plotted on a relative scale: for each model, the sum of all importances is equal to 1. In **A** and **B**, a triple therapy of fluticasone furoate (FF), umeclidinium (U) and vilanterol (V) was compared against a dual therapy of U/V. **C** and **D** compared dual therapy of FF/V against U/V. For **A** and **C**, models estimated individualized treatment effect for time to first exacerbation. For **B** and **D**, models estimated individualized treatment effect for exacerbation rate. The 10 most important features are shown. Error bars indicate the standard deviation over 100 training sets. SGRQ, St. George's Respiratory Questionnaire; FEV1, forced expiratory volume in 1 second; FEV1pp, forced expiratory

ARTICLE IN PRESS

volume in 1 second percent predicted; FVC, forced vital capacity; GOLD, Global Initiative for Chronic Obstructive Lung Disease

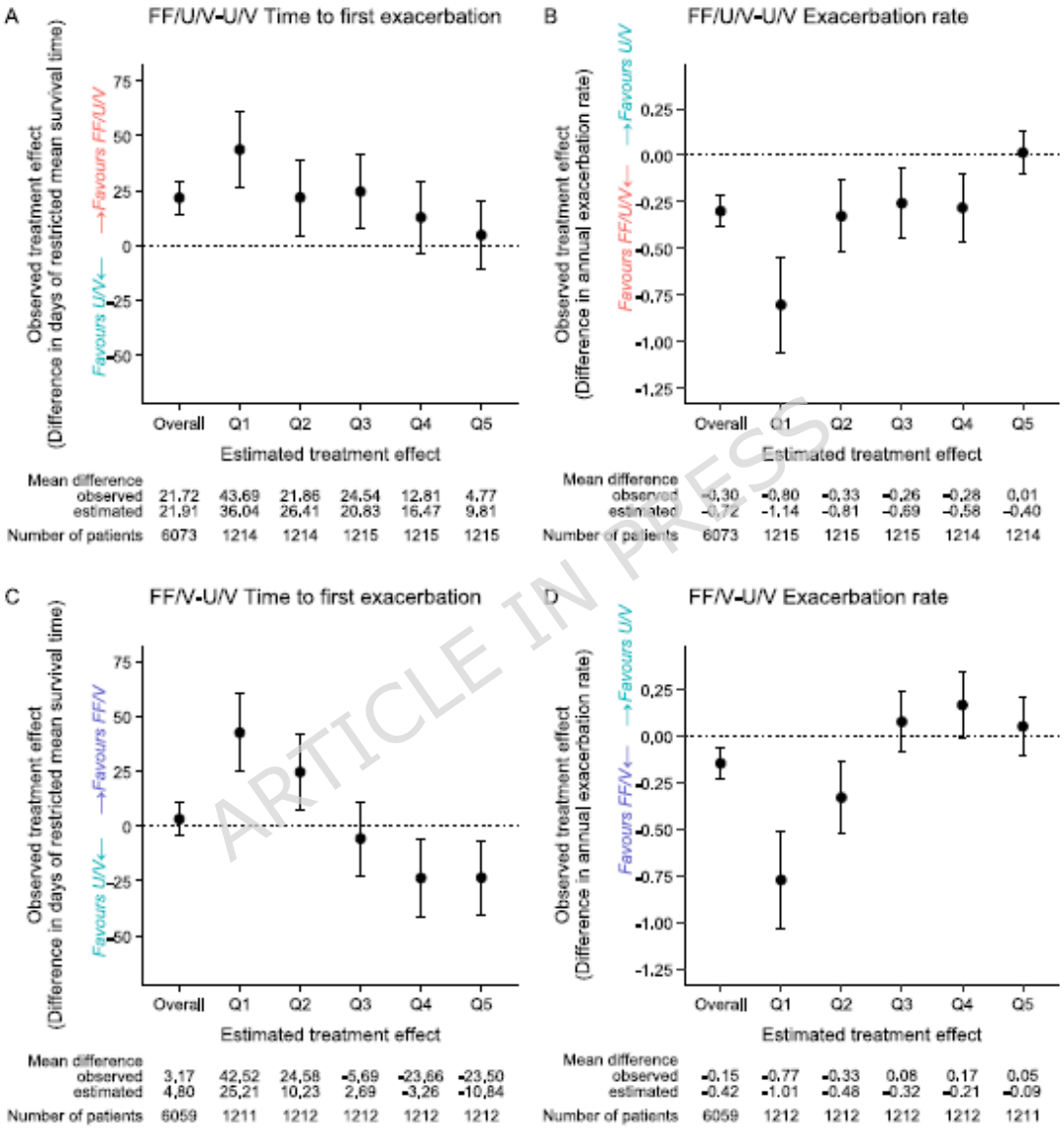


Fig. 2 Observed treatment effects per quintile of estimated treatment effect

In **A** and **B**, a triple therapy of fluticasone furoate (FF), umeclidinium (U) and vilanterol (V) was compared against dual therapy of U/V. **C** and **D** compared dual therapy of FF/V against U/V. Patients were ranked according to their estimated individualized treatment effect (ITE) and split into five quintiles. Quintile 1 (Q1) contains patients estimated to

ARTICLE IN PRESS

benefit most from an inhaled corticosteroid-containing therapy (either FF/U/V or FF/V), while patients in quintile 5 (Q5) were estimated to benefit least from this therapy. For **A** and **C**, models estimated ITE for time to first exacerbation. Observed treatment effects are expressed as a difference in restricted mean survival time. For **B** and **D**, models estimated ITE for exacerbation rate. Observed treatment effects are expressed as a difference in mean annual exacerbation rate between the inhaled corticosteroid-containing therapy and U/V therapy. Error bars indicate 95% confidence intervals.

ARTICLE IN PRESS

ARTICLE IN PRESS

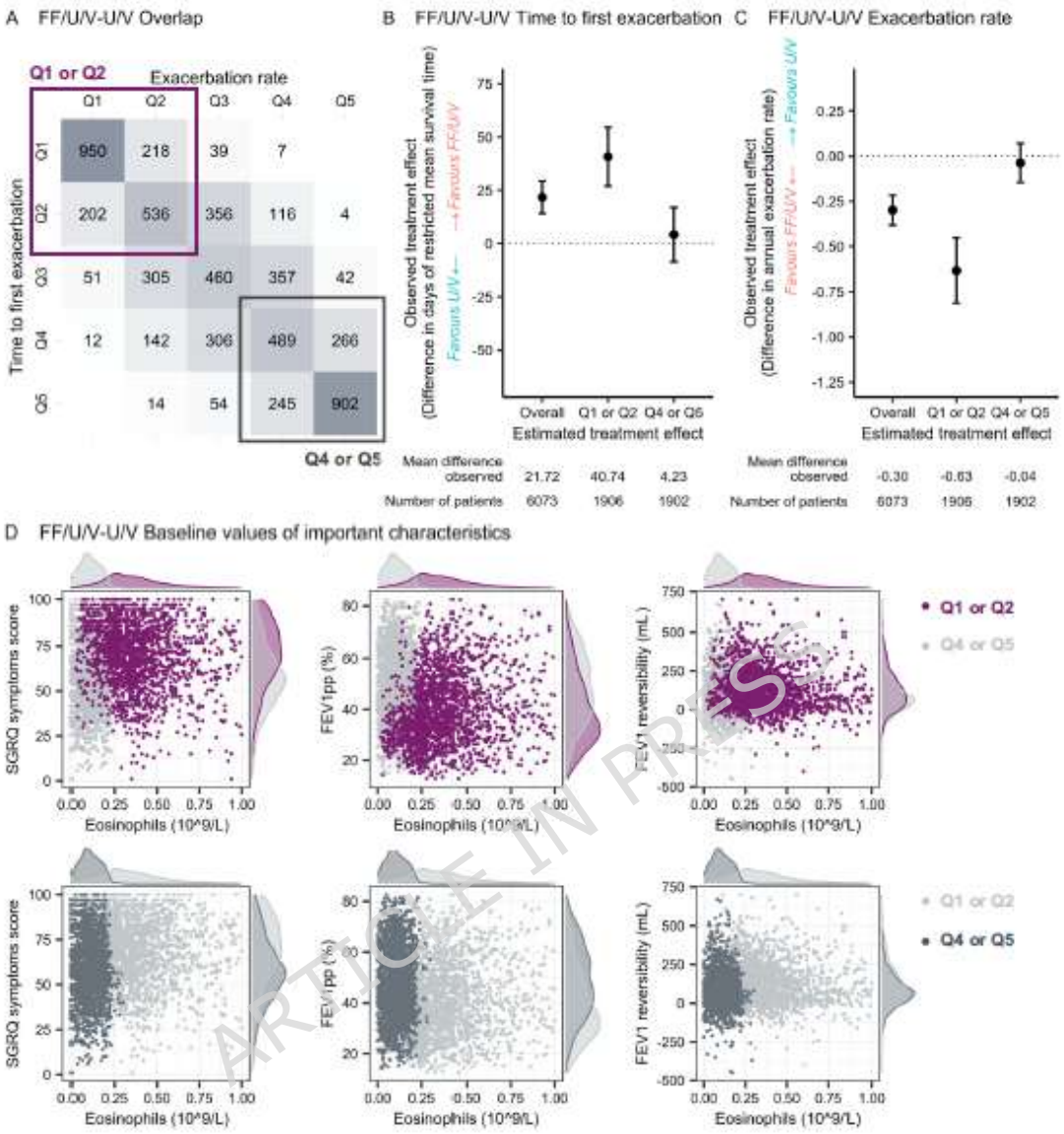


Fig. 3 Overlap in patient rankings for exacerbation rate and time to first exacerbation for FF/U/V versus U/V

A triple therapy of fluticasone furoate (FF), umeclidinium (U) and vilanterol (V) was compared against dual therapy of U/V. Patients were ranked according to their estimated individualized treatment effect and split into five quintiles. **A** illustrates overlap in patient rankings for both exacerbation outcomes with a confusion matrix. **B** and **C** visualize observed treatment effect for patients estimated to benefit most from inhaled

ARTICLE IN PRESS

corticosteroid-containing therapy (FF/U/V) according to both rankings (Q1 or Q2), and for patients estimated to benefit least from this therapy according to both rankings (Q4 or Q5). Observed treatment effects are expressed as the difference in restricted mean survival time and the difference in mean annual exacerbation rate between FF/U/V and U/V therapy. Error bars indicate 95% confidence intervals. **D** visualizes baseline values of the most important characteristics for patients estimated to benefit most and least. FEV1, Forced Expiratory Volume in 1 second; FEV1pp, Forced Expiratory Volume in 1 second percent predicted; SGRQ, St. George's Respiratory Questionnaire.

ARTICLE IN PRESS

ARTICLE IN PRESS

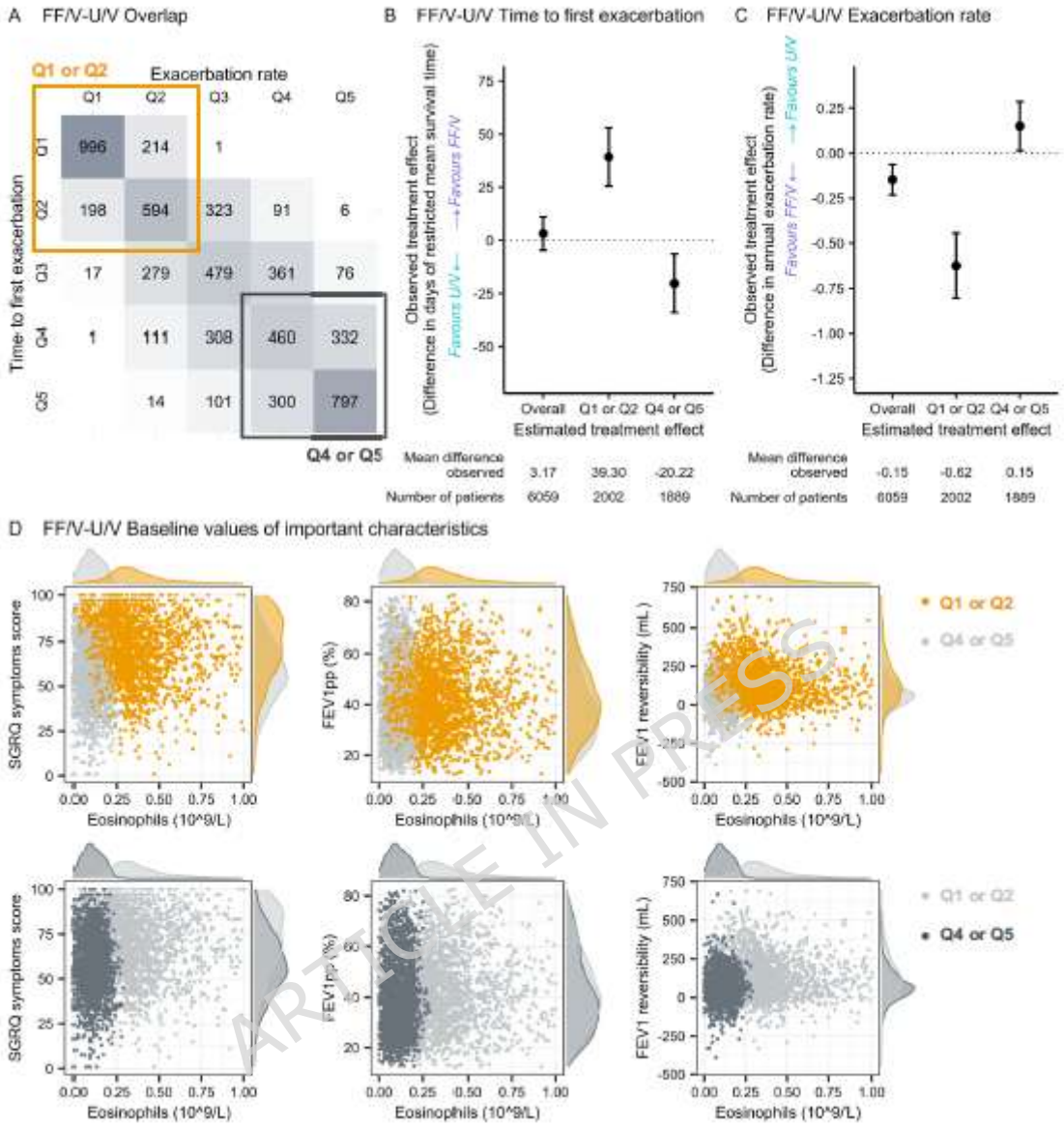


Fig. 4 Overlap in patient rankings for exacerbation rate and time to first exacerbation for FF/V versus U/V

A dual therapy of fluticasone furoate and vilanterol (FF/V) was compared against dual therapy of umeclidinium and vilanterol (U/V). Patients were ranked according to their estimated individualized treatment effect and split into five quintiles. **A** illustrates overlap in patient rankings for both exacerbation outcomes with a confusion matrix. **B** and **C** visualize observed treatment effect for patients estimated to benefit most from inhaled corticosteroid-containing therapy (FF/V) according to both rankings (Q1 or Q2), and for

ARTICLE IN PRESS

patients estimated to benefit least from this therapy according to both rankings (Q4 or Q5). Observed treatment effects are expressed as the difference in restricted mean survival time and the difference in mean annual exacerbation rate between FF/V and U/V therapy. Error bars indicate 95% confidence intervals. **D** visualizes baseline values of the most important characteristics for patients estimated to benefit most and least. FEV1, Forced Expiratory Volume in 1 second; FEV1pp, Forced Expiratory Volume in 1 second percent predicted; SGRQ, St. George's Respiratory Questionnaire.

ARTICLE IN PRESS