Statistical Analysis Plan

Study ID: 208630

Study Official Title: Investigation of the Clinical, Radiological and Biological Factors Associated With Disease Progression, Phenotypes and Endotypes of COPD in China

NCT ID: NCT04853225

Date of Document (SAP Amendment 3): 9-Sept-2024

Date of Document (SAP Amendment 2): 10-Aug-2023

Information Type:	Statistical Analysis Plan (SAP)

TITLE PAGE

Protocol Title: Investigation of the Clinical, Radiological and Biological factors associated with disease progression, phenotypes and endotypes of COPD in China

Study Number: 208630208630

Compound Number: GSK2834425

Abbreviated Title: Investigation of disease progression, phenotypes and endotypes of COPD in China

Acronym: **COMPASS**

Sponsor Name: GlaxoSmithKline Research & Development Limited

Regulatory Agency Identifier Number(s)

Registry	ID
Clintrials.gov	NCT04853225

TABLE OF CONTENTS

PAGE

Version history

SAP Version	Approval Date	Protocol Version (Date) on which SAP is Based	Change	Rationale
RAP	10-Feb-2022	208630 / Amendment 02 (17-Jun-2021)	Not applicable	Original version
RAP amendment1	27-Oct-2022	208630 / Amendment 02 (17-Jun-2021)	Section 2.1: removed 'Per protocol populations were added to analyses populations. '	Correction
			Section 2.1: updated the content of 'treatment pattern, health resource utilization, HRCT, sputum'	Correction 6 months longitudinal data are available and baseline HRCT data are ready, sputum analysis had been done
			Section 3.1 table1: updated the content of Visit 1 to Visit 4	Clarification
			Section 4 table 2: Add HRCT analysis in sub-cohort population	Baseline HRCT data are ready
			Section 5.3.2: Added GOLD II patients in secondary comparison.	Clinical request
			Section 6.1 table 3: updated the content of Computed tomography (Lung HRCT)	Clinical request

SAP Version	Approval Date	Protocol Version (Date) on which SAP is Based	Change	Rationale
			Section 6.3 removed "Wilcoxon-rank sum test will be used to analyse mMRC dyspnea score among different type of subjects in Full population."	Clinical request
			Section 7.4.1: updated the sensitivity and specificity analysis of CAPTUR	Clinical request
			Section 9.1: updated the analysis strategy of treatment pattern	Clinical request
			Section 9.2: added Stratified analysis according to COPD and non-COPD patients	Clinical request
RAP amendment 2	10-Aug-2023	208630 / Amendment 02 (17-Jun-2021)	Section 7.4.1: added the logistic regression analysis for exacerbation	Clinical request
SAP amendment 3		208630 / Amendment 02 (17-Jun-2021)	Apply to SAP template, categorized the endpoints by clinical request.	Clinical request

1. INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the planned analyses and output of the planned analyses to be included in the CSR for Protocol: 208630.

This SAP is based on the following protocol and protocol amendments:

Revision Chronology:		
2017N345807_00	18-OCT-2018	Original
2017N345807_01	19-JUN-2019	Amendment 01
TMF-13816843	17-JUN-2021	Amendment 02

2. SUMMARY OF KEY PROTOCOL INFORMATION

2.1. Changes to the Protocol Defined Statistical Analysis Plan

There were changes to the original planned statistical analyses specified in the protocol amendment 02 (Dated:17JUN2021) of study 208630 [TMF-13816843].

The definition of population will follow the description of SAP, to better fit the study purpose. HRU population added due to data collection status.

Considering the COVID lockdown impact, analysis window for the 3 post-baseline visits (Month 6, Month 18, Month 30) will be changed to ±90 days.

Rate of change (decline) in CAT score analysis will be performed in further post-hoc analysis.

Phenotype analyses will be descriptive only, and further analyses of phenotype may be performed post hoc and based on separate documents.

2.2. Study Objective(s) and Endpoint(s)

The following table shows the objectives and endpoints. More details about the endpoints are provided after the table.

Research Questions & Objectives	Endpoints
1. Evaluate whether the predictors of Chronic Obstructive Pulmonary Disease (COPD) disease progression identified in Western cohorts are applicable in China.	**Main cohort** • Rate of decline in FEV_1 and change from baseline in forced expiratory volume in 1 second (FEV_1)

Research Questions & Objectives	Endpoints
• To evaluate disease progression in COPD and non-COPD participants over 2.5 years as assessed by lung function, exacerbation frequency, health status, lung high-resolution computed tomography (HRCT) scan, and changes in physical activity.	• Rate of decline in FVC and change from baseline in forced vital capacity (FVC) • Rate of moderate/severe exacerbations* • Change from baseline in COPD Assessment Test (CAT) score and rate of decline in CAT score • Frequency of Clinically Important Deterioration (CID) and its components • Death **Additional for sub-cohort** • Change in Evaluating Respiratory Symptoms in COPD (E-RS: COPD) scores • Characterisation of Exacerbation of COPD Tool (EXACT) events • Change from baseline in airway thickness by lung HRCT • Change in lung density by lung HRCT • Change in gas trapping • Change in physical activity measures (e.g. Daily steps, bouts of activity per day)
2. Characterise stable disease and exacerbation phenotypes in China through assessment of blood biomarkers, lung microbiome and radiological features of COPD	**Main cohort:** • Plasma fibrinogen and high-sensitivity C-Reactive Protein (hsCRP) • Differential blood cell count (e.g. eosinophils, neutrophils) and haemoglobin **Additional for sub-cohort** • Blood biomarkers including but not limited to Serum interferon- γ -inducible protein -10 (IP-10), soluble Receptor for Advanced Glycation End Products (sRAGE), Club cell protein (CC16) and HbA1c • Sputum microbiome as assessed by molecular methods

Research Questions & Objectives	Endpoints
	• Sputum total differential cell counts and percentage (e.g. total cell count, the counts and percentage of eosinophils and neutrophils) • CCI [redacted]
3. Characterise treatment pathways and healthcare resource utilization and costs in COPD	• Frequency of treatment with COPD medications; Treatment patterns at the time of study entry and during the study period • Count of outpatient visits, emergency visits, hospitalizations (overall and COPD-specific) • Direct medical costs associated with COPD medications and COPD disease management • Insurance coverage and patient out-of-pocket payment
4. Test the feasibility of integrating electronic health records, electronic clinical outcome assessments (eCOA) and physical activity data to support application of digital data analytics (DDA) platform in China studies (e.g. direct digital data capture and integration)	NA

*Moderate exacerbations are defined as COPD exacerbations that require either systemic corticosteroids (intramuscular (IM), intravenous, or oral) and/or antibiotics. Severe exacerbations are defined as COPD exacerbations requiring hospitalization (including intubation and admittance to an ICU) or result in death.

2.2.1. Additional information for endpoints of COPD progression

Exacerbation (Chapter 7)

Times of exacerbation 0-18m (baseline to Month 18, same for the followings) and 0-30m will be summarised. Rate of exacerbation (yearly) will be estimated based on negative binomial distribution at the end of study.

CID and its components are defined in 7.2. Percentage of participants who have any event(s) in 0-6m, 6-18m, 18-30m will be presented.

Clinical feature Analyses (Chapter 8)

Spirometry at each planned visit (baseline, Month 6, Month 18 and Month 30) and the post-baseline change from baseline will be summarised. FEV1 and FVC will be analysed using MMRM model. Analysis window for post-baseline visits of spirometry is defined as planned visit date ±90 days. Spirometry results out-of-window will be regarded as missing.

Rate of change (decline) in FEV1 and FVC (yearly, unit ml/year) will be analysed at the end of study. All results regardless of the visit window will be included in the analyses.

COPD GOLD grade change, as the result of change of spirometry. Spirometry results out-of-window (visit window is ±90 days) will be regarded as missing.

Patient Report Outcome (PRO) and Activity (Chapter 9)

The PROs and activities are not only collected by on site visit, and please refer to Protocol schedule of activities for further details.

CAT at each planned visits, including phone call visits, will be summarised.

EXACT and E-RS are collected for 2 period of 180 days after baseline and Month 18. Summary of these 2 questionnaires will be based on the mean score during baseline (defined as study D1-D7), Month 1-6 (defined as up to 180 days from study D8) and Month 18-24 (defined as up to 180 days from Month 18 visit).

Daily steps collected by wrist band is used to stands for physical activity measures. Daily steps are collected for 2 period of approximately 30 days after baseline and Month 18. Summary of daily steps will be based on the participants mean steps during baseline (defined as study D1-D7), Month 1 (defined as up to 30 days from study D8) and Month 18 (defined as up to 30 days from Month 18 visit).

HRCT data will be summarised at baseline and the end of study. In addition, for participants who had exacerbation visits (a pre-defined special unscheduled visit happened after exacerbation observed), the HRCT data of these visits will be summarised and compared with baseline and the end of study.

2.2.2. Additional information for other endpoints

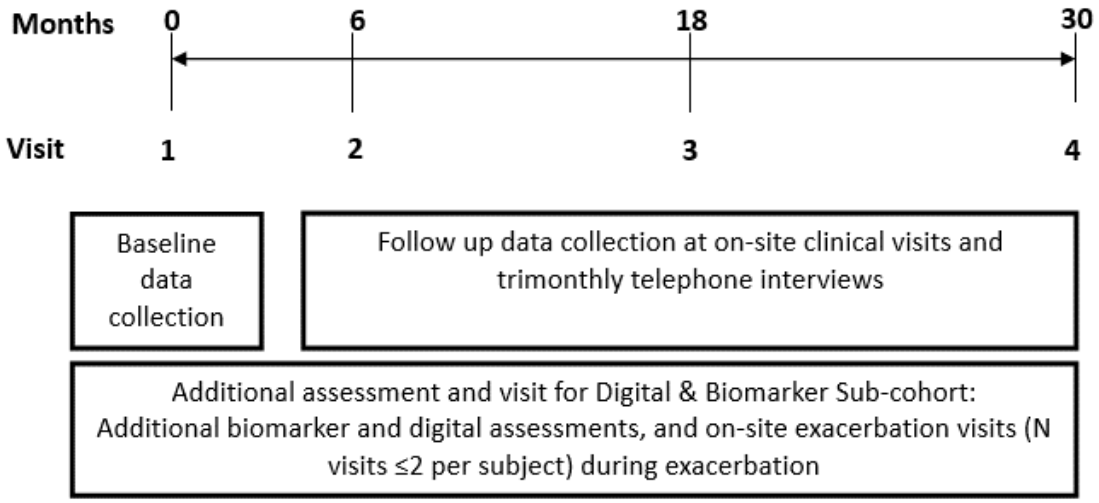
Treatment pattern will be analysed in baseline (up to 90 days prior baseline to baseline), 0-6m, 6-18m and 18-30m.

HRU data will be analysed in total and by season, as defined in 14.2 .

Unless otherwise specified, the other endpoints will be analysed by visits.

2.3. Study Design

Overview of Study Design and Key Features



Design Features	• This is a 2.5-year multi-center, prospective, longitudinal, non-drug interventional cohort study to investigate the pattern of COPD disease phenotypes, clinical, humanistic and healthcare utilization burden, disease progression, and management in China.
Dosing	• There is no investigational drug.
Treatment Assignment	• A total of 2,000 participants are planned to be enrolled, consisting of 1,700 participants with COPD, 180 with chronic bronchitis but without COPD, and 120 never smoker non-COPD healthy controls. • COPD participants will be recruited in the following manner: GOLD I: 700, GOLD II: 700, GOLD III: 200, GOLD IV: 100.
Interim Analysis	• Interim analyses will be conducted after the completion of each clinic visit by all subjects.

2.4. Statistical Hypotheses / Statistical Analyses

This study is mainly descriptive with no formal inference associated with the study objectives. Nominal P-value may be provided as reference. 2-sided 95% confidence intervals will accompany all effect estimates.

3. PLANNED ANALYSES

3.1. Interim Analyses

Interim analyses will be performed in the following sequential steps:

1. All participants have completed each clinical visit as defined in the protocol.
2. All required database cleaning activities in accordance with the Data Management Plan (DMP) have been completed and interim database release (DBR) and database freeze (DBF) has been declared by Data Management.

At the end of Visit 1, data from subjects in the full population will be used to address questions associated with subject phenotypes, characterisation of disease severity, and inter-individual variability of each endpoint. At the end of visit 2, 3 and 4, available clinical, radiological, biomarker, Patient-reported Outcomes (PRO), and health economics endpoints will be assessed based on the appropriate population.

Table 1. Overview of planned interim analyses

Time points	Analyses	Population	Data availability
Last Subject Complete Visit 1 (baseline)	Interim analyses 1	Full/sub-cohort	Study population analyses (baseline) Phenotypic analyses (baseline)
Last Subject Complete Visit 2 (month 6)	Interim analyses 2	Full/sub-cohort	Disease progression analyses Study population analysis (for Radiological examination) Treatment pattern and Health resource utilization/cost analyses
Last Subject Complete Visit 3 (month 18)	Interim analyses 3	Full/sub-cohort	Disease progression analyses (Exacerbation analysis)
Last Subject Complete Visit 4 (month 30)	Final analyses	Full/sub-cohort	Study population analyses Disease progression analyses Phenotypic analyses HRCT analyses Treatment pattern and Health resource utilization/cost Safety analyses

3.2. Final Analyses

The final planned primary analyses will be performed after the completion of the following sequential steps:

1. All subjects have completed/withdraw the study as defined in the protocol.
2. All required database cleaning activities have been completed and final database release and database freeze has been declared by Data Management.

4. ANALYSIS POPULATIONS

Table 2. Summary of Study Populations

Population	Definition / Criteria	Analyses Evaluated
All Screened	The All Screened Population will consist of all participants who sign the ICF to participate in the clinical trial. Participants in this population will be used for screen failure summary.	• Study Population
Full	The Full Population will consist of all subjects (except subject PPD) who are enrolled in the study and attend Visit 1, and not withdraw inform consent.	• Study Population • Disease Progression Analyses • Phenotypic Analyses • Treatment Pattern and Healthcare Resource Utilization/Cost • Safety Analyses
Sub-cohort	Sub-cohort is a subset of Full population. Subjects who are enrolled in the Biomarker & Digital sub-cohort and who provided a least one additional sub-cohort assessment.	• Study Population • Disease Progression Analyses • Phenotypic Analyses • Treatment Pattern and Healthcare Resource Utilization/Cost • HRCT analysis
HRU	HRU is a subset of Full population. The following sites are not feasible to captured data by HIS, thus the patients enrolled from the following sites will be excluded from HRU analysis: 238160, 238222, 237965, 237971, 238155.	• Treatment Pattern and Healthcare Resource Utilization/Cost

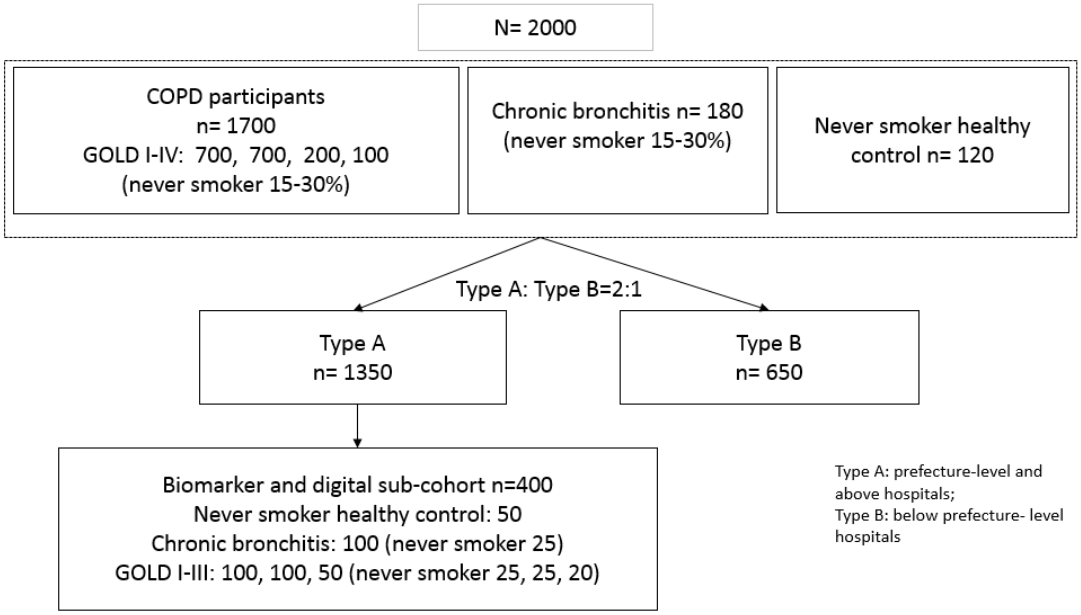
*: Subject PPD cannot be categorized into either of the study arms (COPD patients, Chronic Bronchitis patients or Healthy Control), thus excluded from the Full population.

5. CONSIDERATIONS FOR DATA ANALYSES AND DATA HANDLING CONVENTIONS

5.1. Number of Participants

A total of 2,000 participants are planned to be enrolled, consisting of 1,700 participants with COPD, 180 with chronic bronchitis but without fixed airflow limitation, and 120 never smoker healthy control. Specifically, COPD subjects will be classified using the GOLD grades I – IV (i.e. mild, moderate, severe and very severe). To better understand the disease progression of early stage COPD, i.e. GOLD I and II grades, COPD participants will be recruited in the following manner: GOLD I: 700, GOLD II: 700, GOLD III: 200, GOLD IV: 100 (Figure 1**Error! Reference source not found.**).

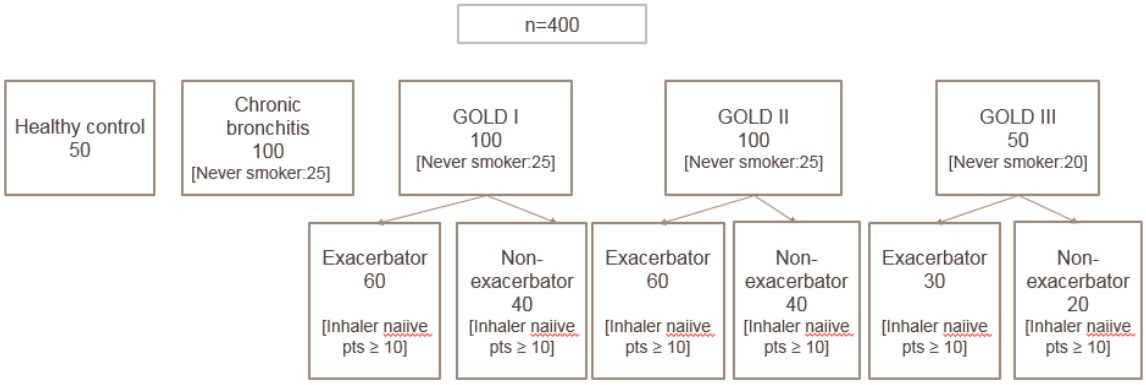
Figure 1 Study Population Schematic



More assessments will be conducted among the 400 Biomarker & Digital sub-cohort participants who will be all recruited from selected Type A hospitals. The sub-cohort consists of 50 never smoker healthy control, 100 chronic bronchitis (N never smoker: 25), 100 GOLD I (N never smoker: 25), 100 GOLD II (N never smoker: 25) and 50 GOLD III participants (N never smoker: 20) (Figure 2 Sub-cohort Population Schematic). In GOLD I-III grades, certain number of COPD participants with previous exacerbation history (i.e. exacerbators) will be recruited (target number of subjects are as follows - GOLD I: 60 exacerbators, GOLD II: 60 exacerbators, GOLD III: 30 exacerbators). Further, a proportion of those who are maintenance inhaler treatment naïve

will be recruited, aiming for: 10 participants who are maintenance inhaler treatment naive will be recruited among exacerbators and non-exacerbators in each GOLD I, II, III groups respectively (
Figure 2). All patient population sizes are approximate and allow +/- 10% for feasibility.

Figure 2 Sub-cohort Population Schematic



5.2. Arm Definition

This is a non-drug interventional study, and the "arm" in this SAP refers to the participants' health status at enrolment, i.e. Healthy Control, Chronic Bronchitis, COPD (GOLD grades I – IV).

5.3. Baseline Definitions

Unless otherwise specified , the baseline value will be the assessment with a non-missing value at Visit 1.

For EXACT/E-RS score and daily steps, the baseline is defined as the average value of D1-D7.

Unless otherwise stated, if baseline data is missing, no derivation will be performed, and baseline will be set to missing.

5.4. General Considerations in Statistical Analyses

Unless otherwise specified, continuous data will be summarized using descriptive statistics: n, mean, standard deviation (sd), median, minimum and maximum. Categorical data will be summarized as the number and percentage of participants in each category.

5.5. Examination of Covariates, Other Strata and Subgroups

5.5.1. Covariates and Other Strata

The potential covariates for statistical analyses may include type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control), gender, smoking status, baseline FEV_1, count of COPD exacerbations in the previous 12 months from baseline. For analyses where a baseline value (of the analysis variable) is available this will also be included as a covariate in model-based analyses, and baseline by visit interaction may be included depending on the model. CCI [redacted]

5.5.2. Examination of Subgroups

Subgroup definitions in main cohort were listed as follows,

- Age (<65 years, >65 at baseline)
- Site (Type A and Type B Hospital)
- Smoking status (Baseline ever/never smoker)
- History of exacerbation in the preceding years (Any exacerbation/No exacerbation)
- Inhaled treatment naïve (Yes/No)
- Childhood respiratory disease (Yes/No)
- Biomass fuel exposure (Yes/No)
- Presence of chronic bronchitis (Yes/No)

5.6. Other Considerations for Data Analyses and Data Handling Conventions

Other considerations for data analyses and data handling conventions are outlined in the appendices:

Section	Component
12.1	Appendix 1: Protocol Deviation Management and Definitions for Per Protocol Population
12.2	Appendix 2: Schedule of Activities
12.3	Appendix 3: Assessment Windows
12.4	Appendix 4: Data Display Standards & Handling Conventions
12.5	Appendix 5: Derived and Transformed Data
12.6	Appendix 6: Reporting Standards for Missing Data
12.7	Appendix 8: Abbreviations & Trademarks

6. STUDY POPULATION ANALYSES

Study population analyses will be based on the full population and sub-cohort population, unless otherwise specified.

6.1. Participant Disposition

A summary of patient disposition, including screening status, reasons for screen failures and completers of each visit will be provided for the All Screened Population. In addition, the number of participants enrolled by site will be summarized by arm using the Full population.

A summary of the number of participants in each of the analysis set described will be provided.

A summary of study disposition including factors shown in study population schematic (Figure 1 and Figure 2, respectively) will be provided for Full and Sub-cohort Population, respectively.

Duration of study will be summarized by arm in Full Population and Sub-cohort Population in a same table, using mean(sd), median, Q1, Q3, min and max. For any participants, the duration of study is defined as the number of days from enrollment to study completion or study discontinuation, i.e.

$$Duration = Date\ of\ Final\ Visit - Date\ of\ Visit\ 1 + 1$$

The "Final Visit" is defined as the last scheduled visit (including exacerbation visit), regardless of Outpatient visits.

6.2. Protocol Deviations

Protocol deviations will be tracked by the study team throughout the conduct of the Study. Important protocol deviations (including deviations related to study inclusion/exclusion criteria, conduct of the trial, patient management or patient assessment) will be reviewed by study team before DBL and listed.

6.3. Demographics and Baseline Characteristics

The demographic characteristics, including age, gender, race, ethnicity, highest education level, household annual income (grade), occupation, basic insurance and commercial insurance will be summarized by arms.

Vital Signs including height, weight, BMI, blood pressure and heart rate will be summarized by arms.

Smoking status, passive smoker, and smoking related details will be summarized by arms.

Health and lifestyle information, including heating, cooking at home and cooking at work will be summarized by arms in Full Population.

6.4. Medical History

COPD disease history, including duration, current COPD maintenance therapy, first-degree relative diseases history, number of exacerbations and inhaled maintenance treatment naïve will be summarized in COPD patients by COPD I-IV arms. Number of exacerbation will be summarised into the following category: 0, 1, 2 and >2, and the median will also be provided.

Lower respiratory infections history will be summarized in chronic bronchitis patients by full population/sub-cohort population. Number of respiratory infections will be summarised into the following category: 0, 1, 2 and >2, and the median will also be provided.

Summary of medical history, current co-morbid conditions, childhood respiratory disease and other related questions will be provided by arms.

Concomitant medications including history and during study will be summarized by arm.

7. EXACERBATION

7.1. Exacerbation

A descriptive summary of exacerbation at 18 month and 30 month will be provided in COPD and Chronic Bronchitis patients for Full population and Sub-cohort population. Moderate, severe and moderate or severe exacerbation will be summarized by times of exacerbation [none, yes (1, 2, greater than 2)].

A generalized linear model assuming a negative binomial distribution will be used to estimate the exacerbation rate (95% CI) for COPD patients and chronic bronchitis patients in Full population, and then compare the exacerbation rate with GOLD I arm. Rate ratio (RR), 95% CI of RR and nominal P value will be provided for reference.

- **Covariates**: type of subjects (i.e. GOLD I-IV, chronic bronchitis), gender, smoking status, exacerbation history (moderate/severe), site (type of hospital), maintenance therapy (Y/N) and ICS containing (Y/N), Childhood chest disease (Y/N) and biomass fuel exposure (Y/N), age, and with offset logarithm of time on study.

7.2. Clinically Important Deterioration (CID)

Clinically Important Deterioration (CID) is a new composite outcome (developed by GSK) for the prediction of future risk of long-term adverse outcomes. CID components are defined as:

- Exacerbation: An occurrence of a moderate/severe exacerbation.
- FEV1: Decrease of ≥100 mL in postbronchodilator FEV1
- CAT: Increase of ≥2 units in the CAT

CID event, or CID-positive (CID+) is considered as to have ANY of the 3 components.

A summary of participants in Full population who had at least 1 CID event in Baseline to Month6, Month6 to Month18 and Month18 to Month30 will be provided, the occurrence of CID and the components will be summarised. Considering the Spirometry onsite test may delay due to COVID, an additional rule of analysis window for post-baseline FEV1 is defined as planned visit date ±90 days. If there's no available record, the result will be regarded as missing (no event).

8. CLINICAL FEATURE ANALYSES

8.1. Spirometry

Summary of pre- and post-bronchodilator spirometry (including FEV1, FVC, FEV1/FVC Ratio, FEV6, FEF25-75, FEV1 percent of predicted test, FEV1 percent of reversibility test, and predicted forced vital capacity) will be provided by arm.

Endpoint: FEV1 and FVC

Additional analysis of FEV1 and FVC will be provided in Full Population, described as followed.

Pre- and post-bronchodilator FEV1 at each visit, and their change from baseline will be summarized descriptively. N(missing), mean, sd, median, Q1, Q3, Minimum and Maximum will be provided. A same descriptive summary will be provided for FVC.

A line plot of pre- and post-bronchodilator FEV1 and FVC will be provided, showing the mean value of pre- and post-bronchodilator FEV1 and FVC at each visit on the same plot.

Analysis of Pre- and Post-bronchodilator FEV1 at each visit using mixed models with repeated measures (MMRM) will be provided. A same analysis will be provided for FVC. Details of MMRM is described in the following table.

Endpoint / Variables
Change of FEV_1 from baseline in milliliters (mL)

Change of FVC from baseline in milliliters (mL)
Model Specification

• The Mixed Models Repeated Measures (MMRM) analysis FEV1 or FVC at every on-site visit (6 months, 18 months and 30 months).
 	
• Terms in the model:
  • **Response**: FEV1 or FVC change from baseline at each visit.
  • **Population**: Full population
  • **Categorical**: visit, type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control), gender, smoking status, site (type of hospital), maintenance treatment (Y/N) and ICS containing (Y/N), Childhood chest disease (Y/N) and biomass fuel exposure (Y/N).
  • **Continuous**: age, baseline FEV_1, count of COPD exacerbations
  • **Interaction:** type of subjects*visit, baseline*visit
  • **Repeated:** visit
 	
• The model will be fitted with an unstructured variance-covariance matrix. If the model fails to converge, then alternative covariance structures may be investigated.
• The Kenward and Roger method (KR) for approximating the denominator degrees of freedom and correcting for bias in the estimated variance-covariance of the fixed effects will be used.
• Comparison with Health Control will be provided.
• Models will be fitted for pre- and post-bronchodilator FEV_1 and FVC, respectively.

Model Results Presentation
• Least-square (LS) means and LS mean change from baseline values for each arm at each visit will be presented with their associated standard errors as well as 95% CIs. • The LS mean differences and their associated standard errors (and associated 95% CIs) compared with Health Control for all visits will be presented in tables. Nominal P value will be presented just for reference.

Yearly rate of change (mL/year) in FEV1 and FVC will be summarized and analysed using random coefficients models respectively. The model is described in the following table.

Endpoint / Variables
• Rate of change in FEV_1 in mL/year • Rate of change in FVC in mL/year
Model Specification
• Random coefficients model • For the analyses of rate of decline, each subject's actual time on study will be used, rather than assigning visit as a categorical variable.

- While missing data are not explicitly imputed in the mixed effects model, there is an underlying assumption that the data are missing at random. Participants and time on study will be random effects.

- Terms in the model:
  - **Response**: FEV_1 or FVC at each visit.
  - **Population**: Full population
  - **Categorical**: type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control), gender, smoking status, site (type of hospital), maintenance treatment (Y/N) and ICS containing (Y/N), Childhood chest disease (Y/N) and biomass fuel exposure (Y/N)
  - **Continuous**: age, baseline FEV1, count of COPD exacerbations, time on study
  - **Interaction:** type of subjects * time on study

- Models will be fitted for pre- and post-bronchodilator FEV_1 or FVC, respectively.
- The model will be fitted with an unstructured variance-covariance matrix. If the model fails to converge, then alternative covariance structures may be investigated.
- A subject needs to have at last one post-baseline value to be included in the analysis.

Model Results Presentation

- Baseline FEV_1 or FVC will be provided with n (number of subjects included in the analysis) and mean(SD).
- The adjusted rate of change (decline), differences in different type of subjects compared with Health Control, nominal p-values and 95% confidence limits will be provided.

Subgroup analysis of rate of change in FEV1 will be analysed using the same random coefficients model. Number of subjects included in the model, baseline value, adjusted rate of change and difference compared with healthy control will be provided without p-value.

8.2. Smoking Status Change

Smoking status at each post-baseline visit in Full Population will be summarised by arm. The participants who do not complete the visit will be regarded as missing at the certain visit.

8.3. COPD Disease Progression

Disease status (healthy, chronic bronchitis, COPD [GOLD I-IV]) at each post-baseline visit in Full Population will be summarised by arm (reflecting baseline disease status). The participants who do not complete the visit will be regarded as missing at the certain visit. The rule for post-baseline COPD disease status is:

1. If post-bronchodilator FEV1(ml)/FVC(ml) ≥70%, the participants will be shown as "FEV1(ml)/FVC(ml) ≥70%", otherwise go to step 2;
2. If post-bronchodilator FEV1(ml)/FVC(ml)<70%, the participant will be regarded as COPD patient and the GOLD grade is defined as:
   GOLD 1: FEV1 ≥ 80% of predicted;
   GOLD 2: FEV1<80% but ≥ 50% of predicted;
   GOLD 3: FEV1<50% but ≥ 30% of predicted;
   GOLD 4: FEV1<30% of predicted.

9. PATIENT REPORT OUTCOME (PRO) AND ACTIVITY

For the details of the PROs, please refer to Patient-reported Outcome.

9.1. CAT score

A descriptive summary of CAT score at baseline and every planned test and change from baseline of post-baseline will be provided with N(missing), mean, sd, median, Q1, Q3, Minimum and Maximum. Besides the total score, the summary of the following 2 sub-domain will also be provided:

- Cough and sputum: Defined as the sum of items 1 and 2.
- Breathlessness and activity: Defined as the sum of items 4, 5 and 6.

9.2. Other PROs and Questionaries

Summary of SGRQ Total and subscales will be provided by arm.

Summary of mMRC Dyspnea score will be provided by arm.

Summary of CAPTURE score at baseline and V3 will be provided by arm. CAPTURE score category (defined as ≥2 and <2) will also be provided in the same table.

9.3. EXACT and E-RS

The EXACT and E-RS are questionnaire-based symptom scores. E-RS utilizes 11 respiratory symptom items from the existing and validated 14-item EXACT. The EXACT/E-RS are collected by daily digital diary, the measurements last for 180 days (6 months) from Visit 1 (baseline) and Visit 3 (month 18). Baseline of EXACT/E-RS is defined as the average of D1-D7. The follow up period start from D8, e.g. 1-6 month is defined as up to 180 days from D8, etc.

Frequency of EXACT events will be summarised by 0, 1, more than 2 for 1-6 month and 18-24 month in participants with at least 3 consecutive records, or 7 records. EXACT event is defined as an increase in EXACT score ≥9 points for 3 days or ≥12 points for 2 days, above baseline (average in 7-day period). For more details, please refer to Patient-reported Outcome.

9.4. Digital Activities

Physical activity will be evaluated by daily steps as assessed by a wrist band dispensed at baseline. Participants are required to keep wearing the wrist band for 1 month from Visit 1 (baseline) and Visit 3 (month 18). The data collected extended the 30-day period will be deleted and not included in analysis.

The value of daily step equal to 0 or >23,000 will be regarded as missing.

The summary of daily step is based on the average for each patient, and will be provided for the following 3 periods:

- Baseline: the average of first 7 days.
- Month 1: the average of up to 30 days from study D8.
- Month 18: the average of up to 30 days from Visit 3 (Month 18).

For the 2 post-baseline period (Visit 1 and Visit 3), participants with at least 3 consecutive days non-missing step data, or at least 7 days non-missing step data will be considered as evaluable. A table with number of evaluable patients, and descriptive summary for participants' average steps will be presented. A box plot will be provided.

10. BIOMAKER ANALYSIS

Summary of blood biomarkers (plasma fibrinogen, hsCRP, blood cell count, etc.) at each visit will be provided, geometrics mean and standard deviation will be presented in addition to the descriptive statistics (n, mean, standard deviation, median, minimum and maximum).

Summary of additional blood protein analyses (only collected in sub-cohort population) at each visit will be provided in sub-cohort population.

11. SPUTUM ANALYSIS

Summary of sputum collection at V1 (Day 1) and V4 (Month 30) will be provided by arms, including sample type, reason for not collected, and test results.

12. IMAGING ANALYSIS

Only sub-cohort participants are planned to have imaging tests; thus, the analysis defined in this chapter will only perform in Sub-cohort Population.

Imaging analysis is based on quantitative HRCT scanning. The CT scan data collected from the sub-cohort in this study provide information on lung structural damage, and the imaging measures are defined as:

Table 3 CT Scan Data and Imaging Endpoints

Imaging Analysis Category	Imaging Measures (Description)
Lung density/Emphysema	Perc15 (Corrected for Lung Volume) LAA-950 (Low Attenuation Area-950) DPM_Emphysema
Airway	Average wall thickness (mm) Airway Wall Area (%) Average of major inner diameter (mm) Average of minor inner diameter (mm) Total Airway Count (TAC) Pi10 (Including Pi10All and Pi10Le20)
Gas Trapping	E/I ratio Air Volume DPM_Airtrap (Gas Trapping)

Summary of imaging measures at baseline and Visit 4 (Month 30) will be provided by arm descriptively with 95% CI. Change from baseline will be provided in a same table.

Summary of imaging measures change with exacerbation will be provided in the COPD patients who had CT scan in exacerbation visits. The analysis will be based on events (exacerbations) instead of on patients, i.e., a same patient with 2 exacerbation imaging tests will be analysis as 2 records. Descriptive summary of baseline, exacerbation and the end of study (Visit 4 or early withdrawal) will be provided with 95% CI. Exacerbation visit change from baseline, end of study change compared with exacerbation visit will be provided in a same table with covariates of age, sex and height at baseline.

13. PHENOTYPIC ANALYSES

The data available at Visit 1 will assist in determination of subject phenotypes in disease manifestation.

Phenotypic analyses COPD patients as well as CB patients. Analysis for phenotypic in this SAP will be descriptive. Potential expletory analysis will be documented in separate documents.

13.1. Phenotypes

Summary of phenotypes in Full and Sub-cohort Population will be provided totally and by GOLD grades, including exacerbator, CMH, significant symptoms and ACO. The summary of phenotype will be performed in COPD patients.

- Exacerbator will be summarized with the following grades: Never, Occasional (1 in previous years), Yearly (1 each year in previous years) and Frequent (≥2 more in previous year).
- CMH (Chronic Mucus Hypersecretion) is identified using the SGRQ-C questionnaire, defined by answering cough for most days or several days a week (option 1 **OR** 2 for the "I cough" question) **AND** phlegm for most days or several days a week (option 1 **OR** 2 for the "I bring up phlegm (sputum)" question).
- Significant symptoms will be summarized by CAT score ≥10 and <10.
- ACO is defined as COPD patients that have physician diagnosed asthma ≤40 years of age. Status of ACO before the age of 18 and between age of 18 and 40 will be provided.

13.2. Endotypes

Summary of endotypes in Sub-cohort Population will be provided totally and by (baseline) GOLD grades, including Airflow limitation blood eosinophils and plasma fibrinogen.

- Airflow limitation is defined as GOLD grade increased from baseline, and will be summarized by Yes (GOLD grade increased) and No (GOLD grade remain stable or decreased)
- Blood Eosinophils will be summarized by > 300 /µl and ≤300 /µl
- Plasma Fibrinogen will be summarized by >400 mg/dl and ≤400 mg/dl

14. ANALYSES OF TREATMENT PATTERN AND HEALTHCARE RESOURCE UTILIZATION/COST

14.1. Treatment Pattern analyses

Treatment pattern analyses will be performed in Full Population. The treatment patterns and their grade are defined in Table 4.

Table 4 Treatment Pattern Definition

Treatment pattern	Treatment pattern Grade	Details in Treatment or the Combination
Naive	0	No any COPD treatment
Stop	0	Only COPD treatment records before 90 days from the first visit
LAMA	1	LAMA
LABA	1	LABA
ICS/LABA	2	ICS/LABA, or Combination 2, at the same time check: ”ICS”+”LABA”
LAMA/LABA	2	LABA/LAMA, or Combination 2, at the same time check: ”LABA”+”LAMA”
Triple(open)	3	When “Triple (open) =1”: Combination 1, at the same time check: “ICS/LABA”+”LAMA”, or Combination 2, at the same time check: “ICS/LABA”+”LABA/LAMA”, or Combination 3, at the same time check: “ICS”+”LABA/LAMA”, or Combination 4, at the same time check: ”ICS”+”LABA”+”LAMA”
Triple(closed)	3	ICS/LABA/LAMA

Descriptive summary of treatment pattern for each visit will be provided with frequency and percentage for the following subgroups: gender, age (<50, 50-59, 60-69, 70-80, >80), smoking status, employment status, Type of Hospital (Type A/B), GOLD grade, Previous Exacerbation status (Yes/No), Hospitalization in preceding year (Yes/No), Presence of chronic bronchitis (Yes/No). All the concomitate medication records collected on Visit 1 within 90 days before the will be considered as baseline and the post-baseline visits contains records collected from the date of last visit to the current visit, e.g. Visit 2 refers to the record from the date of Visit 1 +1 to date of Visit 2. The descriptive summary will be provided by the treatment pattern shown in Table 4.

A shift table will be provided to show the change (step up, stable, switch, step down or stop/missing, treatment naive) of treatment patterns for each visit compared with previous visit and visit 4 compared with baseline, with number of participants in the study, number and percentage of participant in the pre-defined category. The summary will be based on all the post-baseline concomitate medication collected from the date of last visit to the current visit, e.g. Visit 2 refers to the record from the date of Visit 1 +1 to date of Visit 2, and the changes of treatment pattern are defined as:

- Step up: the treatment pattern grade become larger than previous visit ;
- Switch: the treatment pattern changed within a same grade (except grade 0)
- Stable: the treatment pattern remain the same (except grade 0)
- Step down: participants receive treatment and the grade goes down compared with the previous visit;
- Stop/missing: No treatment record in the period

14.2. Health resource utilization (HRU)

The analysis of healthcare resource utilization/cost will be performed in HRU population based on data captured in HIS. Study scheduled visit information is not captured in the database and will not be derived. All the "visit" mentioned in this chapter refers to the information collected from database (i.e. the patient visits a hospital), instead of study scheduled visits (V1-V4).

Endpoint: Number of Visits

A descriptive summary of patient visits (Total, outpatient visits, emergency visits, and hospitalization) during the entire study period will be provided, both COPD related and non-COPD related visits will be presented.

A summary by season for COPD-related visits will also be provided, the definition of season is:

- Spring (Mar, Apr, May)
- Summer (Jun, Jul, Aug)
- Autumn (Sep, Oct, Nov)
- Winter (Dec, Jan, Feb)

The total follow-up period (year) and a summary of crude rate of COPD related and non-COPD related visits and the subcategory visits will be provided in a same table. The total follow-up and crude rate are defined as:

$$Total\ Followup\ Time = \sum(Last\ Visit\ Date - First\ Visit\ Date + 1)$$

$$Crude\ Rate = \frac{Sum\ of\ Number\ of\ Visits}{Total\ Followup\ Time\ (year)}$$

Endpoint: Direct Medical Costs associated with COPD

A descriptive summary of overall direct medical costs associated with COPD during the entire study period will be provided. Total cost and subcategory cost, as well as the insurance/out-of-pocket information will also be presented in the summary.

A summary by season for total cost will also be provided. The definition of season is the same with the definition in “Number of Visits” section.

15. SAFETY ANALYSES

Adverse Events (AE) will not be captured in the study as no study drug will be given in this study. Study procedure related serious adverse events (SAE) will be recorded.

15.1. Adverse Events

A listing will be provided for all study procedure related SAEs in Full Population.

Death and cause of death in Full Population will be summarized by arm. If no death case is reported, the table will be omitted.

16. REFERENCES

Atkinson AC. Plots, transformations and regression. Clarendon Press, Oxford. 1985.

GlaxoSmithKline document number 2017N345807_01: Investigation of the clinical, radiological and biological factors associated with disease progression, phenotypes and endotypes of COPD in China -Amendment 1.

Leidy NK, Murray LT. Patient-reported outcome (PRO) measures for clinical trials of COPD: the EXACT and E-RS. Journal of Chronic Obstructive Pulmonary Disease, 2013; 10:393-398.

Jones, P. St George 'S Respiratory Questionnaire for Copd Patients (Sgrq-C) Manual ,2016, Version No.1.3.

Kim, V., Dolliver, W.R., Nath, H.P. et al. Mucus plugging on computed tomography and chronic bronchitis in chronic obstructive pulmonary disease. Respir Res 22, 110 (2021). https://doi.org/10.1186/s12931-021-01712-0

17. APPENDICES

17.1. Appendix 1: Protocol Deviation Management and Definitions for Per Protocol Population

Protocol deviations will be tracked by the study team throughout the conduct of the study in accordance with the PDMP. Important protocol deviations as identified in the PDMP will be summarized and listed.

17.1.1. Exclusions from Per Protocol Population Analysis

This is an observational cohort study, any subjects with protocol deviations will not be removed from the study analyses. Per protocol population is not defined in this study.

17.2. Appendix 2: Schedule of Activities

17.2.1. Protocol Defined Schedule of Events

Protocol Activity		Screen/Baseline	Study visit				
visit		**V1**	**V2**	**V3**	**V4**	**EW**	**Exacerbation visit (≤2)[1]**
Study month		**M0**	**M6**	**M18**	**M30**		
window		**+3 days**	**±28 days**	**±28 days**	**±28 days**		
Procedures							
Informed consent[2]		X	-	-	-	-	-
Demography		X	-	-	-	-	-
Medical history		X	-	-	-	-	-
Health and lifestyle information/Exposures		X	-	-	-	-	-
Inclusion/Exclusion Criteria		X	-	-	-	-	-
COPD and Exacerbation History[3]		X	-	-	-	-	-
Concurrent medications		X	X	X	X	X	-
Smoking status		X	X	X	X	X	-
Spirometry[7]		X	X	X	X	X	X
Reversibility Testing		X	-	-	-	-	-
Patient reported	mMRC	X	-	-	-	-	
	SGRQ-C	X	-	-	-	-	
	CAPTURE	X		X			
Questionnaires	CAT (on-site visit)	X	X	X	X	-	X

Protocol Activity		Screen/Baseline	Study visit				
	CAT (Phone Call)	Trimonthly Phone Call[4]					
Healthcare resource utilization and cost		-	X	X	X	X	X
Assessments							
COPD exacerbation assessment (on site visit)[5]		-	X	X	X	X	X
COPD exacerbation assessment (Phone Call)[5]		Trimonthly Phone Call[4]					
COPD medication and treatment (on site visit)[5]		X	X	X	X	X	X
COPD medication and treatment (Phone Call)[5]		Trimonthly Phone Call[4]					
Death		-	X				
Serious Adverse Event assessment[12]			X	X	X	X	X
Lab							
Blood Draw for e.g. fibrinogen/HsCRP/Cell count		X	-	-	X	X	X
Sub-cohort only							
Lung HRCT scan[6]		X	-	-	X	X	X
Daily steps		X[8]	-	X[8]	-	-	-
Evening diary (EXACT plus additional questions)		X[9]	-	X[9]	-	-	-
Monthly CAT		X[10]	-	X[10]	-	-	-
Blood biomarkers (inclusive of serum sRAGE, CC16, IP-10, HbA1c)		X	-	-	X	X	X
Sputum cytology[11]		X	-	-	X	X	X
Sputum microbiome[11]		X	-	-	X	X	X

CAT=COPD assessment test, COPD=Chronic Obstructive Pulmonary Disease, HRCT=High-resolution computed tomography, EW=Early Withdrawal, EXACT=EXAcerbation of COPD Tool; SGRQ-C=St.George's Respiratory Questionnaire for COPD.

1. Exacerbation visits only apply to subjects included in the Biomarker & Digital sub-cohort. Moderate exacerbations are defined as COPD exacerbations that require either systemic corticosteroids (intramuscular (IM), intravenous, or oral) and/or antibiotics. Severe exacerbations are defined as COPD exacerbations requiring hospitalization or result in death.
2. Informed consent must be conducted at the screen visit prior to performing any study procedures.
3. For COPD participants only;
4. When a trimonthly telephone interview overlaps a clinical visit, the COPD exacerbation assessment, COPD medication and treatment questionnaires will be completed at the clinic visit.
5. Data collection of COPD exacerbation assessment, COPD medication and treatment is not required for the healthy control participants.
6. CT scan should only be performed at withdrawal visit if the last study CT performed was more than a year ago. Inspiratory and expiratory HRCT scans are required at Visit 1 and Visit 4, while only inspiratory scan will be conducted on exacerbation visit.
7. Pre-and post- bronchodilator values of spirometry test will be collected at each site visit.

8. Physical activity will be evaluated by daily steps as assessed by a wrist band dispensed at baseline. Participants are required to keep wearing the wrist band for 1 month from Visit 1 (baseline) and Visit 3
9. Daily digital EXACT measurement for 6 months from Visit 1 (baseline) and Visit 3
10. Monthly digital measurement from baseline for 6 months, and monthly digital measurement from V3 for 6 months
11. Sputum can be collected spontaneously or can be induced, as per investigator judgement. Sputum sampling should only be done if, in the opinion of the investigator, it is safe for the subject.
12. Only SAEs relating to the study procedures are collected.

17.3. Appendix 3: Assessment Windows

17.3.1. Visit Based Assessments

Nominal visits (scheduled visit) will be used for reporting and analysis.

Assessment windows for the scheduled study visits are as follows:

Full population:

- Baseline/Screen (+ 3 days)
- Visit 2 (± 28 days)
- Visit 3 (± 28 days)
- Visit 4 (± 28 days)

In addition, COPD and chronic bronchitis participants will be contacted by phone on trimonthly basis within ± 7 days of the planned date to assess the CAT and to collect information on exacerbation rates and COPD medications; healthy controls will be contacted by phone on trimonthly basis within ± 7 days of the planned date to assess the CAT only.

17.4. Appendix 4: Data Display Standards & Handling Conventions

17.4.1. Reporting Process

Software
• The version 9.4 or later of SAS software will be used.
Analysis Datasets
• Analysis datasets will be created according to CDISC standards

• For creation of ADaM datasets (ADCM/ADAE), the same version of dictionary datasets will be implemented for conversion from SI to SDTM.
Generation of RTF Files
• Rich Text Format (RTF) files will be generated for the final reporting effort for use in writing the CSR.

17.4.2. Reporting Standards

Reporting Standards	
General	
• The current GSK Integrated Data Standards Library (IDSL) will be applied for reporting, unless otherwise stated (IDSL Standards Location: https://spope.gsk.com/sites/IDSLLibrary/SitePages/Home.aspx): • 4.03 to 4.23: General Principles • 5.01 to 5.08: Principles Related to Data Listings • 6.01 to 6.11: Principles Related to Summary Tables • 7.01 to 7.13: Principles Related to Graphics	
Formats	
• GSK IDSL Statistical Principles (5.03 & 6.06.3) for decimal places (DP's) will be adopted for reporting of data based on the raw data collected, unless otherwise stated. • Numeric data will be reported at the precision collected on the eCRF. • The reported precision from non eCRF sources will follow the IDSL statistical principles but may be adjusted to a clinically interpretable number of DP's.	
Planned and Actual Time	

• Reporting for tables, figures and formal statistical analyses :
  • Planned time will be used in figures, summaries, statistical analyses and calculation of any derived parameters, unless otherwise stated.
  • The impact of any major deviation from the planned assessment times and/or scheduled visit days on the analyses and interpretation of the results will be assessed as appropriate.
• Reporting for Data Listings:
  • Planned and actual time will be shown in listings.
  • Unscheduled or unplanned readings will be presented within the subject's listings.
  • Visits outside the protocol defined time-windows (i.e. recorded as protocol deviations) will be included in statistical analyses.	

Unscheduled Visits	
• Unscheduled visits will not be included in summary tables. • Unscheduled visits will not be included in figures. • All unscheduled visits will be included in listings.	
Descriptive Summary Statistics	
Continuous Data	Refer to IDSL Statistical Principle 6.06.1

Reporting Standards	
Categorical Data	N, n, frequency, %
Graphical Displays	
• Refer to IDSL Statistical Principals 7.01 to 7.13.	

17.5. Appendix 5: Derived and Transformed Data

17.5.1. General

Multiple Measurements at One Analysis Time Point
• Mean of the measurements will be calculated and used in any derivation of summary statistics but if listed, all data will be presented. • If there are two values within a time window, the value closest to the target day for that window will be used. If values are the same distance from the target, then the mean will be taken.
Study Day

• Calculated as the number of days from visit 1:
  • Ref Date = Missing → Study Day = Missing
  • Ref Date < Visit 1 → Study Day = Ref Date – Visit 1
  • Ref Data ≥ Visit 1 → Study Day = Ref Date – Visit 1 + 1

17.5.2. Study Population

Demographics
Age

• Age will be calculated based on the Visit 1.
• Birth date will be imputed as follows:
  • Any subject with a missing day will have this imputed as day ’15’.
  • Any subject with a missing date and month will have this imputed as ’30JUN’.
• Birth date will be presented in listings as ‘YYYY’.

Age categories are based on at screening and are defined as: • <65 years • ≥65 years

Subject disposition
Discontinuation from Study Treatment
Not Applicable.

Subject disposition
Discontinuation from Study Treatment
Discontinuation from Study
Censoring will be performed as follows: • For subject that withdraws consent for disclosure of future information, any data collected prior to withdrawal will be retained and reported. • For subject that withdraws consent for disclosure, the site may use any publicly information to determine vital status at the end of the study and this information would be included in the final analyses. • If a subject withdraws from the study, he/she may request destruction of any samples taken and not tested, such request will be documented in study records. • Subject that is unreachable and fails to come back for both Visit 3 and Visit 4 will be considered lost to follow-up. Data collected prior to missed visits will be retained and reported.

Smoking status
• Smoking status: Never smokers are defined as those with a lifetime exposure of <1 package/year in their life (excluding passive smoker). Ever smokers (current or former) are defined as those with a lifetime exposure of ≥10 package/year. Former smokers are those who have stopped smoking for at least 6 months. Current smokers are those who are currently smoking. Passive smokers are defined as those with the presence of smokers living in the same household, or co-workers smoking nearby while indoors • Smoking status will be reassessed at each scheduled clinic visit evaluated by participants' self-report (e.g. never smoker, current smoker, former smoker). • Smoking history including Pack-years of smoking: Number of pack years (number of cigarettes per day / 20) x number of years smoked, e.g., 10 packyears means 20 cigarettes per day for 10 years, or 10 cigarettes per day for 20 years)

COPD Exacerbation History
• Individual exacerbations during the 12/24 months prior to Screening and during the study were to be collected on the eCRF by the investigator • COPD exacerbations with onset date on or after the date of Screening will not be counted in the counts of number of exacerbations in the past year; all other exacerbations recorded on the eCRF will be included in the counts, even if onset/resolution dates were missing or earlier than the start of the exact 12 month period prior to Screening • Number of COPD exacerbations reported in the past year prior to Screening will be summarised according to three categories: moderate COPD exacerbation, severe COPD exacerbation and moderate/severe COPD exacerbation. • Moderate COPD exacerbations are defined as exacerbation that required treatment with oral/systemic corticosteroids and/or antibiotics (not involving hospitalisation). • Severe COPD exacerbations are defined as exacerbations that required in-patient hospitalisation.

COPD Exacerbation History
• Total number of moderate/severe COPD exacerbations are defined as total numbers of moderate and severe COPD exacerbation for each subject.

Baseline Lung Function
Reversibility
Reversibility will be calculated as follows: (Highest post-bronchodilator FEV_1 – Highest pre-bronchodilator FEV_1)*100/(Highest post-bronchodilator)
GOLD Grade 1-4 at Screening
Subjects will be classified into Global Initiative on Obstructive Lung Disease (GOLD) Grades 1-4 using the post-salbutamol percent predicted FEV1 assessment at Screening: • GOLD Grade 1 (Mild): percent predicted FEV1 ≥80% • GOLD Grade 2 (Moderate): 50%≤percent predicted FEV1 < 80% • GOLD Grade 3 (Severe): 30%≤percent predicted FEV1 < 50% • GOLD Grade 4 (Very Severe): percent predicted FEV1 < 30%

Death
Details of any patient deaths, including date of death, will be recorded in the CRF. Information will be obtained from the telephone interview and from medical records in HIS. Cause of death will be recorded if available. Data from publicly available source will be used if participant is not contactable at the end of the study.

Blood Biomarker
• Blood samples will be taken at Visit 1 and throughout the study as defined in the time and events table. (Appendix 2: Schedule of Activities). • Plasma fibrinogen, hsCRP, total/differential blood cell count measurements and haemoglobin will be conducted at central lab for all participants • Additional protein analysis will be conducted at central lab for the sub-cohort using validated assays. Analytes chosen will include serum sRAGE, CC16 and IP-10， HbA1c. • The blood biomarker assessments will be conducted according to the study reference manual.

Sputum Microbiome and Sputum Cytology
• Sputum samples for microbiome and sputum cytology will be collected at Vist1 and Visit 4 in sub-cohort only as defined in the time and events table (Appendix 2: Schedule of Activities). • Sputum can be collected spontaneously or can be induced, as per investigator judgement, induced sputum is preferred. Sputum sampling should only be done if, in the opinion of the investigator, it is safe for the subject. The sputum microbiome and sputum cytology assessments will be conducted according to the sputum assessment procedure (Refer to study reference manual).

Sputum Microbiome and Sputum Cytology
• Additional analyses may include the measurement of viruses and fungi.

CCI

[redacted]

COPD-related Healthcare Resource Utilization

Within site hospital COPD-related healthcare resource utilization and cost will be captured on the study questionnaire form where possible using the HIS (based on site feasibility). Information collected include:
• Unscheduled COPD outpatient visits and emergency visits
  ○ Number of visits, date of visit, diagnosis of visit, cost
• COPD hospitalization (including intensive care)
  ○ Number of hospitalizations, date and length of stay, diagnosis of visit, cost
• Prescribed COPD medications
  ○ Medication name, date of prescription, dosage, administration, cost, whether the prescription is picked up at hospital, etc.
• Other COPD treatment, e.g. oxygen therapy, nebulizers
  ○ Treatment name, date, cost, etc.
• Lab tests, image tests, surgeries, and other tests and procedures (e.g., spirometry, Chest X-Ray, CT)
  ○ Test or procedure, date of operation, cost, etc.
• Healthcare cost if feasible, e.g. general service usage, consultation fee
  ○ Item, date of charge
• Insurance coverage and patient out-of-pocket payment if feasible
 	
Additionally, healthcare resource utilization could occur in other places outside of the site hospitals including: (1) hospitals other than the site hospital where the subject is recruited; (2) retail pharmacy stores; (3) subjects' home.
 	
Outside of site hospital COPD-related healthcare resource utilization and cost will be collected using standard questionnaires at trimonthly subjects' phone interviews. Information collected include COPD-related diagnosis and treatment in the following sources:
• Outpatient visit, emergency visit, and hospitalization

COPD-related Healthcare Resource Utilization
○ Date, reason, length of stay (for hospitalization) • Prescribed medications • Over-the-counter (OTC) medications • Other therapies, e.g. oxygen therapy, nebulizers • Associated costs if feasible • Lab tests, image tests, surgeries, and other tests and procedures (e.g., spirometry, Chest X-Ray, CT)

Adverse Events
Adverse Events (AE) will not be captured in the study as no study drug will be given in this study. • If an AE (serious or non-serious) or complaint with any specifically named GSK product is reported spontaneously by a subject during the course of the study, the site staff will complete the AE reporting form and transmit the report to the GSK China within 24 hours of being made aware of the AE. Study procedure related serious adverse events (SAE) will be recorded. • All SAEs will be collected from the signing of the ICF until the follow-up visit at the time points specified in the Schedule of Activities (Appendix 2: Schedule of Activities) • The investigator is responsible for the detection and documentation of events meeting the criteria and definition of an SAE, provided in this protocol. During the study when there is a safety evaluation, the investigator or site staff will be responsible for detecting, documenting, and reporting SAEs Serious Adverse Event (SAE) is defined as any untoward medical occurrence that: • Results in death, • Is life-threatening, • Requires hospitalization or prolongation of existing hospitalization, • Results in disability/incapacity, • Is a congenital anomaly/birth defect, • Other: medical or scientific judgement should be exercised in deciding whether reporting is appropriate in other situations, such as important medical events that may not be immediately life-threatening or result in death or hospitalization but may jeopardize the subject or may require medical or surgical intervention to prevent one of the other outcomes listed in the above definition. These should also be considered serious. • Possible drug-induced liver injury See study Pharmacovigilance Plan (sPVP) for details. sPVP will include the following elements to ensure a comprehensive approach to safety event collection and reporting: • Supplier pharmacovigilance training • Investigator and site staff pharmacovigilance training • Safety-specific roles

Adverse Events
• SAEs collection and reporting processes • SAEs collection forms • Frequency of data review • Reporting process and timelines • Interim reports • Study-specific PVP monitoring process

17.5.3. Patient-reported Outcome

CAT
• The CAT is an 8-item questionnaire suitable for completion by all patients diagnosed with COPD. When completing the questionnaire, participants rate their experience on a 6-point scale, ranging from 0 (no impairment) to 5 (maximum impairment) with a scoring range of 0-40. Higher scores indicate greater disease impact. • For all participants, CAT will be completed at baseline, each follow up clinic visit, and tri-monthly phone interviews. Interviewer-administered approach (face-to-face or over the phone) will be used. • Additionally, for monitoring purpose, monthly CAT will be interviewer-administered over phone call in the sub-cohort for 6 months after baseline visit and 6 months after Visit 3. CAT will also be assessed in the sub-cohort at exacerbation visits.

CAPTURE
• The CATPURE is a short, five-item questionnaire that can be easily completed by patients and are used to identify individuals who may have undiagnosed, clinically significant COPD. • The algorithm is a simple summation of patient responses to each of the five items. (For Question1-4, Score 0 for No, Score 1 for Yes, Question 5, Score 0,1,2. So the total score is ranging from 0 to 6). Higher score means higher risk of COPD (0-1 is at low risk, 2-6 is at high risk).

St. George's Respiratory Questionnaire for COPD Patients (SGRQ-C)
• The SGRQ-C contains 14 questions with a total of 40 items grouped into three domains (Symptoms, Activity and Impacts). • The details for how to score the SGRQ-C are outlined in the SGRQ-C manual (Jones, 2016). This includes details on how to handle missing data. • SGRQ-C domain and total score will be converted to SGRQ scores as described in the manual. • The converted SGRQ scores will be used for all summaries/analyses.

mMRC Dyspnea Scale

- The modified Medical Research Council (mMRC) Dyspnea Scale quantifies disability attributable to breathlessness and is useful for characterizing baseline dyspnea in patients with respiratory diseases.
- mMRC is a simple one-item questionnaire that is completed by patients to grade their degree of baseline functional disability due to dyspnea. The 5 options in the question item translate to a 0-4 score range, with higher score indicating higher severity.

EXACT and E-RS:COPD

- The EXACT is a 14-item daily diary designed to provide a measure of patient-reported symptoms of COPD exacerbation. An EXACT Total Score, ranging from 0 to 100, where higher scores indicate a more severe condition, will be derived for each day of diary collection according to the instructions in the EXACT User Manual (Version 4.0, Evidera, 2011).
- Definition of EXACT events: magnitude of responder for symptomatic events defined as acute, sustained symptomatic worsening of COPD, defined as an increase in EXACT score ≥9 points for 3 days or ≥12 points for 2 days, above baseline (average in 7-day period)
- The EXACT daily diary will be completed each evening electronically for a month following Visit 1 and Visit 3.
- The E-RS™ involves the use of 11 of the 14 EXACT-PRO™ items and a different scoring algorithm to assess the severity of respiratory symptoms in patients with COPD. The E-RS™ score ranges from 0 to 40 with higher scores indicating more severe symptoms.
- Three subscales of the E-RS: COPD are used to describe different symptoms: breathlessness, cough and sputum, and chest symptoms.
- EXACT and E-RS score's scoring algorithms are from the EXACT-PRO user manual (Evidera, 2011).

Scale	Item-level Construct	Item Number	Score Range
RS-Breathlessness			
	Breathless today	7	0-4
	Breathless with activity	8	0-3
	Short of Breath – personal care	9	0-4
	Short of Breath – indoor activity	10	0-3
	Short of Breath – outdoor activity	11	0-3
RS-Cough and Sputum			
	Cough frequency	2	0-4
	Mucus quantity	3	0-3
	Difficulty with mucus	4	0-4
RS-Chest Symptoms			
	Congestion	1	0-4
	Discomfort	5	0-4
	Tightness	6	0-4
Respiratory Symptoms	E-RS Total	11 items	0-40

EXACT and E-RS:COPD			
Additional Attributes			
Used in the EXACT total score for characterizing exacerbations, specifically	Tired or weak	12	0-4
	Sleep disturbance	13	0-4
	Scared or worried	14	0-3
EXACT – Exacerbation	EXACT Raw Total Score[1]	14 items	0-55

[1]EXACT Total Score will be transformed to a 0-100 scale based on the raw total score

Exact Event
Definition of EXACT events: magnitude of responder for symptomatic events defined as acute, sustained symptomatic worsening of COPD, defined as an increase in EXACT score ≥9 points for 3 days or ≥12 points for 2 days, above baseline (average in 7-day period).

Clinically important deterioration (CID)
Clinically important deterioration (CID) is a new composite outcome (developed by GSK) for the prediction of future risk of long-term adverse outcomes. The definition of composite CID used in this study is CAT-CID: • Occurrence of a moderate/severe exacerbation and/or • ≥100 mL deterioration from baseline in postbronchodilator FEV1 and/or • ≥2 units deterioration in the CAT.

COPD Exacerbation
• COPD subjects will be asked for a history of AECOPD using components of moderate/severe exacerbation definition at each clinic visit including exacerbation visits. • Additionally, COPD subjects will be contacted by phone every 3 months by the study staff and asked about their exacerbation details for the previous 3 months. • Study staff will check Hospital Information System (HIS) during the interview to confirm the event of exacerbation wherever possible.

Digital Physical Activity
• Daily physical activity will be collected in sub-cohort only. Participants are required to keep wearing the wrist band for 1 month from visit 1 and 3 as outlined in the time and event table (Appendix 2: Schedule of Activities). • Daily activity as measured by number of daily steps will be captured using digital mobility devices (wrist bands) worn by participants, dispensed at baseline. • Daily steps walked, sedentary minutes, active minutes, very active minutes, sedentary minutes and total sleeping minutes will be derived and summarized. • Definition of intensity data

Digital Physical Activity

- SEDENTARY - Little to no activity monitored. This could be due to minimal movement, sitting, resting, or sleeping.
- ACTIVE - Some activity monitored. A brisk walk could achieve this intensity.
- HIGHLY_ACTIVE - High activity monitored. Running or speed walking could achieve this intensity.

- Below is a template of how activity information is summarized for site personnel in Rave EDC

CRF Component	Value Format	Calculation	Example Values
Date	Text	Date of log line record	26 MAR 2015
Sync Date/Time	Text	Date and time data was synced to Garmin	26 MAR 2015 14:00
Daily Steps Walked	Numeric	Total # of steps in routine data objects for a subjects day (timezone is subject's)	2240
Sedentary Minutes	Numeric	Total # of minutes in routine data objects for a subject's day (timezone is subject's) with intensity parameter = sedentary	438
% of time sedentary	Numeric	Total # of minutes in routine data objects for a subject's day (timezone is subject's) with intensity parameter = sedentary divided by 1440 minutes (total # of minutes per day)	85.32
Active Minutes	Numeric	Total # of minutes in routine data objects for a subject's day (timezone is subject's) with intensity parameter = active	74
% of time active	Numeric	Total # of minutes in routine data objects for a subject's day (timezone is subject's) with intensity parameter = active divided by 1440 minutes (total # of minutes per day)	14.47
Very Active Minutes	Numeric	Total # of minutes in routine data objects for a subject's day (timezone is subject's) with intensity parameter = very active	1
% of time very active	Numeric	Total # of minutes in routine data objects for a subject's day (timezone is subject's) with intensity parameter = very active divided by 1440 minutes (total # of minutes per day)	0.20

Digital Physical Activity			
Total Daily Minutes Sleep	Numeric	Total # of minutes in sleep data objects for a subject's day (timezone is subject's)	362

Phenotype
Exacerbator

• Exacerbator is categorized based on the number of exacerbations patients have in the past:
  ○ Never
  ○ Occasional (1 in the last 2 years)
  ○ Yearly (1 each year in the last 2 years)
• Frequent (≥2 in the previous year)

Chronic mucus hypersecretion (CMH)
• Chronic mucus hypersecretion (CMH) is defined as: At least 3 months of cough & phlegm in past year. This will be identified by asking a question as "How many months in the past 12 months have you had bronchitis or chronic coughing with phlegm or sputum from the chest?" • SGRQ-C will be used for identification of CMH in this study. This would be identified by cough and phlegm related questions in SGRQ-C or CAT. • SGRQ-CMH is defined by answering cough for most days or several days a week (option 1 or 2 for the "I cough" question) and phlegm for most says or several days a week (option 1 or 2 for the "I bring up phlegm [sputum]" question) (Kim et al. 2021) • Significant Symptoms is defined as a CAT score of 10 or higher.
Significant Symptoms
• Significant Symptoms is defined as soring 10 or higher in the CAT questionnaire.
Asthma-COPD Overlap (ACO)
• Asthma-COPD overlap (ACO) has been used to identify patients with airway disease who have features of both asthma and COPD. • ACO is defined as COPD patients that have physician diagnosed asthma ≤40 years of age

Endotype
Airflow Limitation
• FEV1 (GOLD grade)
PRISm
• Normal FEV1/FVC ratio, but FEV1 <80% pred
'Pure' Airways Disease
• Airway wall thickening on CT • No emphysema on CT
Presence of emphysema
• CT scan – standard definition now available

Higher Blood Eosinophils
• Eosinophils > 300 µl
Fibrinogen
• Plasma fibrinogen > 400 mg/dL
Mibrobiome

17.6. Appendix 6: Reporting Standards for Missing Data

17.6.1. Premature Withdrawals

Element	Reporting Detail
General	• Subject study completion (i.e. as specified in the protocol) was defined as completed all scheduled visit and completed exit visit. • Withdrawn subjects were not replaced in the study. • All available data from participants who were withdrawn from the study will be listed and all available planned data will be included in summary tables and figures, unless otherwise specified. • Withdrawal visits will be slotted as per Appendix 3: Assessment Windows or will be summarised as withdrawal visits.

17.6.2. Handling of Missing Data

Element	Reporting Detail

General	• Missing data occurs when any requested data is not provided, leading to blank fields on the collection instrument:
	  ○ These data will be indicated by the use of a "blank" in subject listing displays. Unless all data for a specific visit are missing in which case the data is excluded from the table.
	  ○ Answers such as "Not applicable" and "Not evaluable" are not considered to be missing data and should be displayed as such.

MAR	• The data are said to be missing at random (MAR) when drop-out probability is conditionally independent of missing data given the observed outcomes and covariates • There is an underlying assumption that the data are missing at random. Missing data are not explicitly imputed in the Mixed Model Repeated Measures model. • If data are MAR, the source of the missing data may be ignored and inference is valid from a mixed-effects model. In other words, it is not necessary to include in the data analysis information about the source of the missing data. • For sensitivity analyses, Multiple imputation will be utilized under the assumption of MAR for MMRM model for participants who withdrawal from the study as well as lost to follow up the study, post- withdrawal counts were imputed conditional upon the subject's own observed measurements prior to withdrawal.

17.7. Appendix 8: Abbreviations & Trademarks

17.7.1. Abbreviations

Abbreviation	Description
CAPTURE	COPD Assessment in Primary Care to Identify Undiagnosed Respiratory Disease and Exacerbation Risk
CAT	COPD Assessment Test
CDISC	Clinical Data Interchange Standards Consortium
CID	Clinically Important Deterioration
COPD	Chronic Obstructive Pulmonary Disease
CSR	Clinical Study Report
CTR	Clinical Trial Register
DDA	Digital Data Analytics
DOB	Date of Birth
EXACT	EXAcerbation of COPD Tool
FVC	forced vital capacity
GOLD	Global Initiative for Chronic Obstructive Lung Disease
GSK	GlaxoSmithKline
HRCT	High-resolution computed tomography
ICH	International Conference on Harmonization
IDSL	Integrated Data Standards Library
MMRM	Mixed Model Repeated Measures
PDMP	Protocol Deviation Management Plan
PRO	Patient-reported Outcomes
RAP	reporting and analysis plan
SAC	Statistical Analysis Complete
SGRQ-C	St. George's Respiratory Questionnaire for COPD
SOP	Standard Operation Procedure
TFL	Tables, Figures & Listings
TLC	Total Lung Capacity

17.7.2. Trademarks

Trademarks of the GlaxoSmithKline Group of Companies
COPD Assessment Test (CAT)

Trademarks not owned by the GlaxoSmithKline Group of Companies
EXACT-PRO
E-RS:COPD
SAS
NONMEM
SGRQ

Division	:	Worldwide Development
Information Type	:	Reporting and Analysis Plan (RAP)
Title	:	Reporting and Analysis Plan for 208630: Investigation of the Clinical, Radiological and Biological factors associated with disease progression, phenotypes and endotypes of COPD in China
Compound Number	:	GSK2834425
Effective Date	:	10 Aug 2023

Description:

- The purpose of this RAP (Core RAP) is to describe the planned analyses and output to be included in the Clinical Study Report (CSR) for Protocol 208630.
- This RAP is intended to describe the planned study population analyses, disease progression analyses, phenotypic analyses, treatment pattern and healthcare resource utilization analyses. This RAP will be provided to the study team members to convey the content of the Statistical Analysis Complete (SAC) deliverable.

RAP Author(s):

Author's Name and Functional Area:	
PPD	07-Aug-2023
Statistician, Biostatistics, Clinical Development	

RAP Team Review Confirmations:

Reviewer	Date
PPD GF Respiratory CMO COPD, GSK	07-Aug-2023
PPD Medical Affairs Lead, Medical Affair, China	07-Aug-2023
PPD Principle Clinical Data Scientist, Data Management, GSK R&D China	07-Aug-2023
PPD Programming, GSK R&D China	07-Aug-2023
PPD Senior Study Manager, Study Management, GSK R&D China	07-Aug-2023

Clinical Statistics and Clinical Programming Line Approvals:

Approver	Date
PPD Principle Statistician, Biostatistics, Clinical development, GSK R&D China	10 Aug 2023

TABLE OF CONTENTS

PAGE

Version history

RAP Version	Approval Date	Protocol Version (Date) on which RAP is Based	Change	Rationale
RAP	10-Feb-2022	208630 / Amendment 02 (17-Jun-2021)	Not applicable	Original version
RAP amendment1	27-Oct-2022	208630 / Amendment 02 (17-Jun-2021)	Section 2.1: removed 'Per protocol populations were added to analyses populations. '	Correction
			Section 2.1: updated the content of 'treatment pattern, health resource utilization, HRCT, sputum'	Correction 6 months longitudinal data are available and baseline HRCT data are ready, sputum analysis had been done
			Section 3.1 table1: updated the content of Visit 1 to Visit 4	Clarification
			Section 4 table 2: Add HRCT analysis in sub-cohort population	Baseline HRCT data are ready
			Section 5.3.2: Added GOLD II patients in secondary comparison.	Clinical request
			Section 6.1 table 3: updated the content of Computed tomography (Lung HRCT)	Clinical request

RAP Version	Approval Date	Protocol Version (Date) on which RAP is Based	Change	Rationale
			Section 6.3 removed "Wilcoxon-rank sum test will be used to analyse mMRC dyspnea score among different type of subjects in Full population."	Clinical request
			Section 7.4.1: updated the sensitivity and specificity analysis of CAPTUR	Clinical request
			Section 9.1: updated the analysis strategy of treatment pattern	Clinical request
			Section 9.2: added Stratified analysis according to COPD and non-COPD patients	Clinical request
RAP amendment2	10 Aug 2023	208630 / Amendment 02 (17-Jun-2021)	Section 7.4.1: added the logistic regression analysis for exacerbation	Clinical request

1. INTRODUCTION

The purpose of this reporting and analysis plan (RAP) is to describe the planned analyses and output of the planned analyses to be included in the CSR for Protocol: 208630.

This RAP is based on the following protocol and protocol amendments:

Revision Chronology:		
2017N345807_00	18-OCT-2018	Original
2017N345807_01	19-JUN-2019	Amendment 01
TMF-13816843	17-JUN-2021	Amendment 02

2. SUMMARY OF KEY PROTOCOL INFORMATION

2.1. Changes to the Protocol Defined Statistical Analysis Plan

There were changes to the original planned statistical analyses specified in the protocol amendment 02 (Dated:17JUN2021) of study 208630 [TMF-13816843].

One of the endpoints in disease progression changed from 'Assess COPD Assessment in Primary Care to Identify Undiagnosed Respiratory Disease and Exacerbation Risk (CAPTURE) to identify underdiagnosed COPD patients in Chinese population' into 'Test the sensitivity and specificity of CAPTURE to identify possible COPD patients in a Chinese population'.

Add rate of decline in CAT score for additional endpoint analyses.

2.2. Study Objective(s) and Endpoint(s)

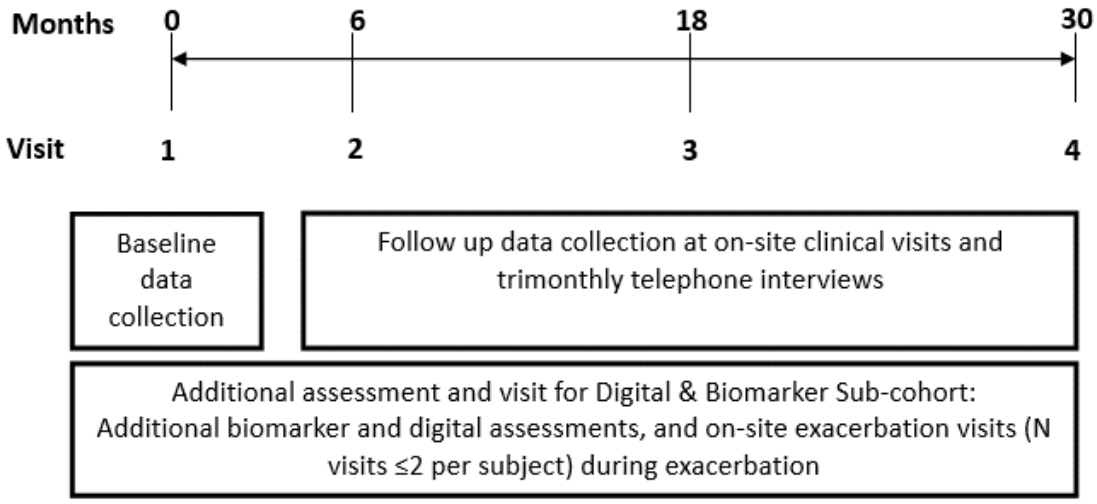
Research Questions & Objectives	Endpoints
1. Evaluate whether the predictors of Chronic Obstructive Pulmonary Disease (COPD) disease progression identified in Western cohorts are applicable in China. • To evaluate disease progression in COPD and non-COPD participants over 2.5 years as assessed by lung function, exacerbation frequency, health status, lung high-	**Main cohort** • Rate of decline in FEV_1 and change from baseline in forced expiratory volume in 1 second (FEV_1) • Rate of decline in FVC and change from baseline in forced vital capacity (FVC) • Rate of moderate/severe exacerbations* • Change from baseline in COPD Assessment Test (CAT) score and rate of decline in CAT score

Research Questions & Objectives	Endpoints
resolution computed tomography (HRCT) scan, and changes in physical activity.	• Test the sensitivity and specificity of CAPTURE to identify possible COPD patients in a Chinese population • Frequency of Clinically Important Deterioration (CID) and its components • Death **Additional for sub-cohort** • Change in Evaluating Respiratory Symptoms in COPD (E-RS: COPD) scores • Characterisation of Exacerbation of COPD Tool (EXACT) events • Change from baseline in airway thickness by lung HRCT • Change in lung density by lung HRCT • Change in gas trapping • Change in physical activity measures (e.g. Daily steps, bouts of activity per day)
2. Characterise stable disease and exacerbation phenotypes in China through assessment of blood biomarkers, lung microbiome and radiological features of COPD	**Main cohort:** • Plasma fibrinogen and high-sensitivity C-Reactive Protein (hsCRP) • Differential blood cell count (e.g. eosinophils, neutrophils) and haemoglobin **Additional for sub-cohort** • Blood biomarkers including but not limited to Serum interferon- γ -inducible protein -10 (IP-10), soluble Receptor for Advanced Glycation End Products (sRAGE), Club cell protein (CC16) and HbA1c • Sputum microbiome as assessed by molecular methods • Sputum total differential cell counts and percentage (e.g. total cell count, the counts and percentage of eosinophils and neutrophils)

Research Questions & Objectives	Endpoints
	• CCI [redacted]
3. Characterise treatment pathways and healthcare resource utilization and costs in COPD	• Frequency of treatment with COPD medications; Treatment patterns at the time of study entry and during the study period • Count of outpatient visits, emergency visits, hospitalizations (overall and COPD-specific) • Direct medical costs associated with COPD medications and COPD disease management • Insurance coverage and patient out-of-pocket payment
4. Test the feasibility of integrating electronic health records, electronic clinical outcome assessments (eCOA) and physical activity data to support application of digital data analytics (DDA) platform in China studies (e.g. direct digital data capture and integration)	NA

2.3. Study Design

*Moderate exacerbations are defined as COPD exacerbations that require either systemic corticosteroids (intramuscular (IM), intravenous, or oral) and/or antibiotics. Severe exacerbations are defined as COPD exacerbations requiring hospitalization (including intubation and admittance to an ICU) or result in death.

Overview of Study Design and Key Features	
	
Design Features	• This is a 2.5-year multi-center, prospective, longitudinal, non-drug interventional cohort study to investigate the pattern of COPD disease phenotypes, clinical, humanistic and healthcare utilization burden, disease progression, and management in China.
Dosing	• There is no investigational drug.
Treatment Assignment	• A total of 2,000 participants are planned to be enrolled, consisting of 1,700 participants with COPD, 180 with chronic bronchitis but without COPD, and 120 never smoker non-COPD healthy controls. • COPD participants will be recruited in the following manner: GOLD I: 700, GOLD II: 700, GOLD III: 200, GOLD IV: 100.
Interim Analysis	• Interim analyses will be conducted after the completion of each clinic visit by all subjects.

2.4. Statistical Hypotheses / Statistical Analyses

This study is mainly descriptive with no formal inference associated with the study objectives. Confidence intervals will accompany all effect estimates.

3. PLANNED ANALYSES

3.1. Interim Analyses

Interim analyses will be performed in the following sequential steps:

1. All participants have completed each clinical visit as defined in the protocol.
2. All required database cleaning activities in accordance with the Data Management Plan (DMP) have been completed and interim database release (DBR) and database freeze (DBF) has been declared by Data Management.

At the end of Visit 1, data from subjects in the Full Population will be used to address questions associated with subject phenotypes, characterisation of disease severity, and inter-individual variability of each endpoint. At the end of visit 2, 3 and 4, available clinical, radiological, biomarker, Patient-reported Outcomes (PRO), and health economics endpoints will be assessed based on the appropriate population.

Table 1. Overview of planned interim analyses

Time points	Analyses	Population	Data availability
Visit 1 (baseline)	Interim analyses 1	Full/sub cohort	• Study population analysis: Characterisation of disease severity (including spirometry, blood and digital biomarkers, PROs) • Phenotypic analyses: Phenotypes and endotypes
Visit 2 (month 6)	Interim analyses 2	Full/sub cohort	• Disease progression analyses: Disease progression in spirometry and PROs • Study population analysis: Radiological examination • Treatment pattern and Health resource utilization/cost: Treatment pattern and Health economics

			• Disease progression analyses: CAPTURE sensitivity examination
Visit 3 (month 18)	Interim analyses 3	Full/sub cohort	• Disease progression analyses: Disease progression in spirometry and PROs • Treatment pattern and Health resource utilization/cost: Treatment pattern and Health economics
Visit 4 (month 30)	Final analyses	Full/sub cohort completer	• Disease progression analyses: Disease progression (including spirometry, blood, sputum and digital biomarkers, PROs, radiological) • Treatment pattern and Health resource utilization/cost

3.2. Final Analyses

The final planned primary analyses will be performed after the completion of the following sequential steps:

1. All subjects have completed the study as defined in the protocol.
2. All required database cleaning activities have been completed and final database release and database freeze has been declared by Data Management.

4. ANALYSIS POPULATIONS

Table 2. Summary of Study Populations

Population	Definition / Criteria	Analyses Evaluated
Full population	The Full Population will consist of all subjects who are enrolled in the study and attend Visit 1.	• Disease Progression Analyses • Phenotypic Analyses • Treatment Pattern and Healthcare Resource Utilization/Cost • Safety Analyses

Population	Definition / Criteria	Analyses Evaluated
Sub-cohort population	Subjects who are enrolled in the Biomarker & Digital sub-cohort and who provided a least one additional sub-cohort assessment.	• Disease Progression Analyses • Phenotypic Analyses • Treatment Pattern and Healthcare Resource Utilization/Cost • HRCT analysis
Completer Population	The Completer Population will consist of all subjects in the Full Population who complete all scheduled clinic visits over 2.5 years.	• Disease Progression Analyses • Phenotypic Analyses • Treatment Pattern and Healthcare Resource Utilization/Cost

4.1. Protocol Deviations

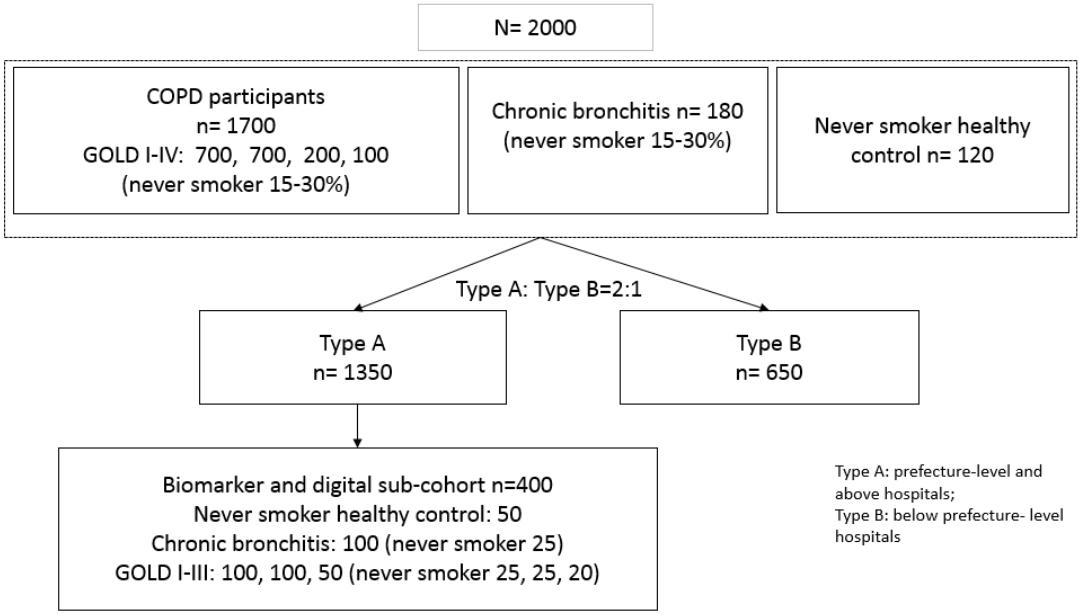
- Important protocol deviations (including deviations related to study inclusion/exclusion criteria, conduct of the trial, patient management or patient assessment) will be summarized and listed.
- Protocol deviations will be tracked by the study team throughout the conduct of the study in accordance with the Protocol Deviation Management Plan.
  - Data will be reviewed prior to freezing the database to ensure all important deviations are captured and categorized on the protocol deviations dataset.
  - This dataset will be the basis for the summaries and listings of protocol deviations.
- A separate summary and listing of all inclusion/exclusion criteria deviations will also be provided. This summary will be based on data as recorded on the inclusion/exclusion page of the eCRF.

5. CONSIDERATIONS FOR DATA ANALYSES AND DATA HANDLING CONVENTIONS

5.1. Number of Participants and sub-groups

A total of 2,000 participants are planned to be enrolled, consisting of 1,700 participants with COPD, 180 with chronic bronchitis but without fixed airflow limitation, and 120 never smoker healthy control. Specifically, COPD subjects will be classified using the GOLD grades I – IV (i.e. mild, moderate, severe and very severe). To better understand the disease progression of early stage COPD, i.e. GOLD I and II grades, COPD participants will be recruited in the following manner: GOLD I: 700, GOLD II: 700, GOLD III: 200, GOLD IV: 100 (Figure 1**Error! Reference source not found.**).

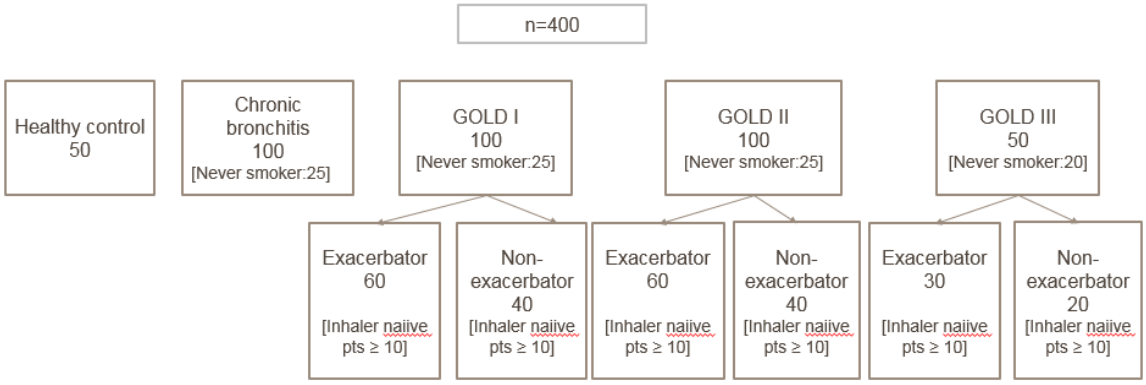
Figure 1 Study Population Schematic



More assessments will be conducted among the 400 Biomarker & Digital sub-cohort participants who will be all recruited from selected Type A hospitals. The sub-cohort consists of 50 never smoker healthy control, 100 chronic bronchitis (N never smoker: 25), 100 GOLD I (N never smoker: 25), 100 GOLD II (N never smoker: 25) and 50 GOLD III participants (N never smoker: 20) (Figure 2 Sub-cohort Population Schematic). In GOLD I-III grades, certain number of COPD participants with previous exacerbation history (i.e. exacerbators) will be recruited (target number of subjects are as follows - GOLD I: 60 exacerbators, GOLD II: 60 exacerbators, GOLD III: 30 exacerbators). Further, a proportion of those who are maintenance inhaler treatment naïve will be recruited, aiming for: 10 participants who are maintenance inhaler treatment naive

will be recruited among exacerbators and non-exacerbators in each GOLD I, II, III groups respectively (
Figure 2). All patient population sizes are approximate and allow +/- 10% for feasibility.

Figure 2 Sub-cohort Population Schematic



5.2. Baseline Definitions

For all endpoints, the baseline value will be the assessment with a non-missing value at Visit 1.

Unless otherwise stated, if baseline data is missing no derivation will be performed and baseline will be set to missing.

5.3. Examination of Covariates, Other Strata and Subgroups

5.3.1. Covariates and Other Strata

The potential covariates for statistical analyses include type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control), COPD phenotypes (e.g. Exacerbation frequency, % emphysema, % airways disease)/endotypes (e.g. blood biomarker levels such as high eosinophils or low sRAGE), gender, smoking status, baseline FEV_1, count of COPD exacerbations in the previous 12 months and 24 months. For analyses where a baseline value (of the analysis variable) is available this will also be included as a covariate in model-based analyses. CCI [redacted]

5.3.2. Examination of Subgroups

Statistical comparisons will be performed in the following subgroups, unless other specified

- The primary comparison is between COPD patients and Chronic Bronchitis patients/ never smoker healthy control, p value as well as 95%CI will be provided.

- The secondary comparison is between GOLD I (or GOLD II) and Chronic Bronchitis patients/ never smoker healthy control, p value as well as 95%CI will be provided.

Baseline covariates considered for inclusion will be including type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control), age, gender, smoking status, baseline FEV_1, count of COPD exacerbations, site, time on study and type of subjects by time interaction.

.

Subgroup definitions in main cohort were listed as follows,

Age (<65 years, >65)

Site (Type A and Type B Hospital)

Smoking status (ever/never smoker)

History of exacerbation in the preceding years (Yes/No)

Inhaled treatment naïve (Yes/No)

Childhood respiratory disease (Yes/No)

Biomass fuel exposure (Yes/No)

5.4. Other Considerations for Data Analyses and Data Handling Conventions

Other considerations for data analyses and data handling conventions are outlined in the appendices:

Section	Component
12.1	Appendix 1: Protocol Deviation Management and Definitions for Per Protocol Population
12.2	Appendix 2: Schedule of Activities
12.3	Appendix 3: Assessment Windows
12.4	Appendix 4: Data Display Standards & Handling Conventions
12.5	Appendix 5: Derived and Transformed Data

Section	Component
12.6	Appendix 6: Reporting Standards for Missing Data
12.7	Appendix 7: Model Checking and Diagnostics for Statistical Analyses
12.8	Appendix 8: Abbreviations & Trademarks

6. STUDY POPULATION ANALYSES

6.1. Overview of Planned Study Population Analyses

The study population analyses will be based on the Full population, unless otherwise specified.

Table 3 provides an overview of the planned study population displays.

Table 3 Overview of Planned Study Population Analyses

Endpoint / Parameter / Display Type	Data Displays Generated		
	Table	Figure	Listing
Subject Disposition			
Study populations and Reasons for Screen Failure	Y[1,2]		Y[1]
Study Dispositions	Y[1,2]		
Enrollment Site	Y[1,2]		
Protocol Deviations			
Important Protocol Deviations	Y[1,2]		Y[1]
Subjects with Inclusion/Exclusion Criteria Deviations	Y[1]		Y[1]
Populations Analysed			
Study Populations and Exclusions	Y[1]		
Subjects Excluded from Completer Population	Y[1]		Y[1]
Demography[1]			
Demographic Characteristics	Y[1,2]		
Race	Y[1,2]		
Social-demographic Characteristics	Y[1,2]		
Health and Lifestyle Information			
Vital Signs	Y[1,2]		
Smoking Status and History	Y[1,2]		
Biomass Exposure	Y[1,2]		
Occupational Exposures	Y[1,2]		
COPD disease History			
Disease Duration	Y[1,2,3]		Y[1]
Initial COPD Maintenance Treatment	Y[1,2,3]		Y[1]
Current COPD Therapy	Y[1,2,3]		Y[1]
History of COPD Exacerbation	Y[1,2,3]		Y[1]
Inhaled Maintenance Treatment History	Y[1,2,3]		Y[1]
Lower Respiratory Infections History			
Number of Infections	Y[1,2,4]		Y[1]
Number of Infections Treated with Oral Corticosteroids and/or Antibiotics and/or Required Hospital Admission	Y[1,2,4]		Y[1]
Medical History			

Endpoint / Parameter / Display Type	Data Displays Generated		
	Table	Figure	Listing
Previous Asthma History	Y[1,2]		Y[1]
Current Co-morbid Conditions	Y[1,2]		Y[1]
Childhood Respiratory Diseases	Y[1,2]		Y[1]
Family History of COPD or Other Respiratory Diseases	Y[1,2,3]		Y[1]
Concomitant Medications			
Concomitant Medications History	Y[1,2]		Y[1]
Spirometry[5]			
FEV_1	Y[1,2]		
FVC and FEV1/FVC Ratio	Y[1,2]		
FEV_6	Y[1,2]		
Forced Expiratory Flow Between 25% and 75% of Vital Capacity (FEF 25-75)	Y[1,2]		
FEV1 Percent of Predicted and Reversibility Test	Y[1,2]		
Biomarkers			
Blood Biomarker	Y[1,2]		
Additional Blood Protein Analyses	Y[2]		
Sputum Microbiome	Y[2]		
Sputum Collection			
Sputum Cytology	Y[2]		
Patient Reported Outcome Measures			
CAT	Y[1,2]		
SGRQ	Y[1]		
mMRC Dyspnea Score	Y[1]		
CAPTURE	Y[1,2]		
EXACT events and E-RS scores	Y[2]		
Computed tomography (Lung HRCT)			
Volumetric measure	Y[2]		
Emphysema	Y[2]		
Airway wall	Y[2]		
DPMs (disease probability measure)	Y[2]		
Death			
Death	Y[1,2]		Y[1,2]
Digital physical activity			
Daily average of digital steps	Y[2]	Y[2]	
Daily average of sedentary minutes/ percentage	Y[2]		
Daily average of active minutes/ percentage	Y[2]		
Daily average of very active minutes/ percentage	Y[2]		
Daily average of total minutes sleep	Y[2]		

NOTES : Y = Yes display generated.
[1] For Full Population

Endpoint / Parameter / Display Type	Data Displays Generated		
	Table	Figure	Listing

[2] For Sub-cohort Population
[3] COPD patients
[4] Chronic Bronchitis Patients
[5] Pre- and post- bronchodilator

6.2. Population of Interest

The disease progression analyses will be based on the full population and sub-cohort population, unless otherwise specified.

6.3. Statistical Analyses / Methods

Unless otherwise specified, endpoints / variables defined in Section 6.1 will be summarised using descriptive statistics and listed.

For the endpoints listed below, summary comparison will be conducted between different type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control).

- modified Medical Research Council (mMRC) Dyspnea Scale (mMRC)
- St. George's Respiratory Questionnaire for COPD

SGRQ score will be summarized using descriptive statistics in Full population.

7. DISEASE PROGRESSION ANALYSES

7.1. Overview of Planned Disease Progression Analyses

Table 4 provides an overview of the planned disease progression analysis displays.

Table 4 Overview of Planned Disease Progression Analyses

Endpoint / Display Type	Absolute							Change from Baseline						
	Stats Analysis			Summary		Individual		Stats Analysis			Summary		Individual	
	T	F	L	T	F	F	L	T	F	L	T	F	F	L
Main Cohort Endpoints														
Post-bronchodilator FEV1 rate of decline	Y[1]													
Post- bronchodilator FEV1 (MMRM analysis)	Y[1]			Y[1]			Y	Y[1,2,3,4,5]	Y[1,2]		Y			
Pre-bronchodilator FEV1 rate of decline	Y[1]													
Pre- bronchodilator FEV1 (MMRM analysis)	Y[1]			Y[1]			Y	Y[1,2,3,4,5]	Y[1,2]		Y			

Endpoint / Display Type	Absolute							Change from Baseline						
	Stats Analysis			Summary		Individual		Stats Analysis			Summary		Individual	
	T	F	L	T	F	F	L	T	F	L	T	F	F	L
Post- bronchodilator FVC rate of decline	Y[1]													
Post- bronchodilator FVC (MMRM analysis)	Y[1]			Y[1]			Y	Y[1,2,3,4,5]	Y[1,2]		Y			
Pre- bronchodilator FVC rate of decline	Y[1]													
Pre- bronchodilator FVC (MMRM analysis)	Y[1]			Y[1]			Y	Y[1,2,3,4,5]	Y[1,2]		Y			
Rate of Moderate/severe exacerbations	Y[1,3,4]			Y[1]			Y							
CAT score(MMRM analysis)	Y[1]			Y[1]			Y	Y[1,2,3,4,5]	Y[1,2]		Y			
Assess CAPTURE to identify under diagnosed COPD patients	Y			Y			Y							
CID composite outcome and components	Y	Y		Y			Y							
death				Y[3,4]			Y[3,4]							
Sub Cohort Endpoints														
Characterization of EXACT events	Y	Y		Y			Y							
E-RS: COPD scores				Y			Y	Y	Y		Y			
Airway thickness				Y[3.4]			Y				Y[3.4]			
Lung density by Lung HRCT				Y[3.4]			Y				Y[3.4]			
Gas trapping as a measure of small airway disease				Y[3.4]			Y				Y[3.4]			
Daily steps				Y			Y				Y			

NOTES :

- T = Table, F = Figure, L = Listing, Y = Yes display generated.
- Stats Analysis = Represents TFL related to any formal statistical analyses (i.e. modelling) conducted.
- Summary = Represents TFL related to any summaries (i.e. descriptive statistics) of the observed raw data.
- Individual = Represents FL related to any displays of individual subject observed raw data.
- [1] Subcohort analysis additionally
- [2] Sensitivity analysis
- [3] Separate tables for early withdrawal subjects from full population
- [4] Completer Populations
- [5] Separate tables for Exacerbation population

7.2. Summary Measure

CCI [redacted] The analyses will be conducted in each interim analyses as well as the final analyses.

Mean change in FEV1, FVC, CAT and Daily steps subgroups of interest based on baseline information will be compared using least square mean of change from baseline.

For rate of exacerbation, adjusted exacerbation rate will be provided. In addition, subgroups of interest based on baseline information will be compared using ratio of events for main cohort.

Change in E-RS scores, airway thickness by HRCT ,lung density by HRCT and gas trapping by HRCT will be compared using least square mean of change from baseline in sub-cohort.

Frequency of EXACT events will be compared using the ratio of events rate for subgroups in sub-cohort.

7.3. Population of Interest

The disease progression analyses will be based on the full population and sub-cohort population, unless otherwise specified.

7.4. Statistical Analyses / Methods

Unless otherwise specified, endpoints / variables defined in Section 7.1**Error! Reference source not found.** will be summarised using descriptive statistics, graphically presented (where appropriate) and listed.

7.4.1. Statistical Methodology Specification

Endpoint / Variables
• Rate of decline in FEV1 in mL/year • Rate of decline in FVC in mL/year
Model Specification
• Random coefficients model (a particular form of a mixed effects model) • For the analyses of rate of decline, each subject's actual time on study will be used, rather than assigning visit as a categorical variable. • While missing data are not explicitly imputed in the mixed effects model, there is an underlying assumption that the data are missing at random. Participants and time on study will be random effects.

- Terms in the model:
  - **Response**: FEV1/FVC at each visit.
  - **Categorical**: type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control), smoking status, site (type of hospital).
  - **Continuous**: age, baseline FEV1, count of COPD exacerbations, time on study
    Interaction: type of subjects*time on study
- Models will be fitted for pre- and post-bronchodilator FEV_1, respectively.
- The model will be fitted with an unstructured variance-covariance matrix. If the model fails to converge, then alternative covariance structures may be investigated.
- Only FEV1 values measured after baseline are used as the response. So in this model a subject needs to have at last one post-baseline value to be included in the analysis.
- To investigate the effects of particular covariates
  - Base model: type of patients, time on study, type of patients* time on study (interaction term) terms for the covariate and covariate-by-time interaction were added to the base model described above, these covariates including gender, smoking status, age, count of COPD exacerbations, maintenance treatment and ICS containing, Childhood chest disease (Y/N) and biomass fuel exposure (Y/N)
  - The slopes were estimated for each type of subjects ((i.e. GOLD I-IV, chronic bronchitis, healthy control)

Model Results Presentation

- The estimate of the slope is the parameter estimate associated with the type of subjects (Overall COPD subjects, GOLD I-IV, chronic bronchitis, healthy control)/ gender/ smoking status/ age/ count of COPD exacerbationsmaintenance treatment and ICS containing/ Childhood chest disease /biomass fuel exposure by time interaction in the model. The adjusted means, pairwise differences in different type of subjects, p-values and 95% confidence limits for the differences of interests will be summarized overall and also estimated for each visit.

Model Checking & Diagnostics

Refer to Appendix 7: Model Checking and Diagnostics for Statistical Analyses.

Main Cohort Statistical Analyses

Endpoint / Variables

Change from baseline FEV_1 in millilitres (mL)
Change from baseline FVC in millilitres (mL)

Model Specification

- Mixed Models Repeated Measures (MMRM) model

- While missing data are not explicitly imputed in the MMRM model, there is an underlying assumption that the data are missing at random.

- The MMRM analysis for FEV1/FVC at on-site visit (6 months, 18 months and 30 months). The MMRM analysis for FEV1 will include FEV1/FVC at each visit.

- Terms in the model:
  - **Response**: FEV1/FVC change from baseline at each visit.
  - **Categorical**: visit, type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control), gender, smoking status, site (type of hospital), maintenance treatment (Y/N) and ICS containing (Y/N), Childhood chest disease (Y/N) and biomass fuel exposure (Y/N).
  - **Continuous**: age, baseline FEV_1count of COPD exacerbations
  - **Interaction:** type of subjects*visit, baseline*visit
  - **Repeated:** visit

- The model will be fitted with an unstructured variance-covariance matrix. If the model fails to converge, then alternative covariance structures may be investigated.
- The Kenward and Roger method (KR) for approximating the denominator degrees of freedom and correcting for bias in the estimated variance-covariance of the fixed effects will be used.
- Baseline is defined in Section 5.2.
- Two models will be fitted; one with a response variable of change from baseline and one with the response variable as the raw value.
-
- Models will be fitted for pre- and post-bronchodilator FEV_1/FVC, respectively.

Model Results Presentation

- Least-square (LS) means and LS mean change from baseline values for each type of subject will be presented with their associated standard errors as well as 95% CIs. The estimated difference of interest (Overall COPD subjects, GOLD I-IV, chronic bronchitis, healthy control) defined in section 7.2 along with corresponding standard error, 95% CI for each visit will be presented.
- The LS mean differences (and associated 95% CIs) for all visits will be presented in tables.
- The LS mean change from baseline values (and 95% CIs) for all visits will be presented for each subgroup.

Model Checking & Diagnostics

Refer to Appendix 7: Model Checking and Diagnostics for Statistical Analyses.

Subgroup Analyses

- Subgroup analysis will be performed as specified in section 5.3.2. This will include analysis by age (<65 years, ≥65), gender (Female, Male), GOLD I-IV baseline reversibility, number of exacerbations in previous year (2, 3, ≥4) and smoking status (ever/never smoker).
- Statistical analyses for sub-cohort participants should be conducted for scheduled visits and exacerbation visits, respectively.

Sensitivity and Supportive Analyses

- A sensitivity analysis will also be performed on the subjects in completer Population.
- Multiple imputation will be utilized under the assumption of MAR for MMRM model for participants who withdrawal from the study as well as lost to follow up the study,

post- withdrawal counts were imputed conditional upon the subject's own observed measurements prior to withdrawal.

Endpoint / Variables

- Rate of moderate/severe exacerbations

Moderate and severe exacerbations occurring from baseline will be included in the analysis.

For all summaries and analyses, exacerbations occurring within 7 days from the end of the first exacerbation of each other within a given subject will be handled as a continuation of the same exacerbation, assigned to the greatest severity. All individual exacerbations will be listed.

Model Specification

- Generalized linear model assuming a negative binomial distribution
- Terms in the model:
  - **Response**: number of moderate/severe exacerbations experienced per subject.
  - **Categorical:** visit, type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control), gender, smoking status, exacerbation history (moderate/severe), site (type of hospital), maintenance therapy (Y/N) and ICS containing (Y/N), Childhood chest disease (Y/N) and biomass fuel exposure (Y/N).
  - **Continuous:** age, baseline FEV_1, time on study
  - **Offset:** logarithm of time on study

Model Results Presentation

- The model estimated mean rate of exacerbations per year will be calculated using the observed marginal distributions of the study population covariates by inclusion of the OM (obsmargins) option in the LSMEANS statement of the GENMOD procedure. The estimated mean rate of moderate/severe exacerbations per year for each type of subjects (Overall COPD subjects, GOLD I-IV, chronic bronchitis, healthy control) and the rate ratio for each type of subjects with 95% confidence intervals and associated p-values will be presented in a table and associated figure.

Endpoint(s)

- Change from baseline in CAT score

Model Specification

- Mixed Models Repeated Measures (MMRM) model.
- While missing data are not explicitly imputed in the MMRM analyses, there is an underlying assumption that the data are missing at random.
- The MMRM analysis for CAT score measurements at on-site visit (month 6, months 18 and month 30) and trimonthly telephone interview (month 3, month 9, month 12, month 15, month 21, month 24, month 27). Additional visits will be added for sub-cohort (month

2, month 4, month 5, month 19, month 20, month 22 and month 23). The MMRM analysis for CAT score will include CAT scores at each visit.

- Terms in the model:
  - **Response**: Change from baseline in CAT score at each visit.
  - **Categorical**: visit, type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control), gender, smoking status, site (type of hospital), maintenance therapy (Y/N) and ICS containing (Y/N).
  - **Continuous**: age, baseline FEV_1, baseline CAT score, count of COPD exacerbations,
  - **Interaction:** baseline CAT score*visit, type of subjects*visit
  - **Repeated**: visit

- The model will be fitted with an unstructured variance-covariance matrix.
- The Kenward and Roger method (KR) for approximating the denominator degrees of freedom and correcting for bias in the estimated variance-covariance of the fixed effects will be used.
- Baseline is defined in Section 5.2.
- Two models will be fitted; one with a response variable of change from baseline and one with the response variable as the raw value.

Model Results Presentation

- Least-square (LS) means and LS mean change from baseline values for each type of subject (Overall COPD subjects, GOLD I-IV, chronic bronchitis, healthy control) will be presented with their associated standard errors as well as 95% CIs. The estimated difference along with corresponding standard error, 95% CI for each visit will be presented.
- The LS mean differences (and associated 95% CIs) for all visits will also be presented in tables.

Model Checking & Diagnostics

- Refer to Appendix 7: Model Checking and Diagnostics for Statistical Analyses.

Subgroup Analyses

Subgroup analysis will be performed as specified in section 5.3.2. This will include analysis by age (<65 years, ≥65), gender (Female, Male), GOLD I-IV baseline reversibility, number of exacerbations in previous year (2, 3, ≥4), maintenance treatment (Y/N) and ICS containing (Y/N) and smoking status (ever/never smoker).

For CAT score, analyses will be conducted for main cohort and sub-cohort, respectively. Statistical analyses for sub-cohort participants should be conducted for scheduled visits and exacerbation visits, respectively.

Sensitivity and Supportive Analyses

- A mixed effects model whereas time is entered as a random effect will be analysed as a supportive analysis to estimate the longitudinal changes if necessary.

Endpoint / Variables

- Rate of decline in CAT score

Model Specification

• Random coefficients model (a particular form of a mixed effects model)
• For the analyses of rate of decline, each subject's actual time on study will be used, rather than assigning visit as a categorical variable.
• While missing data are not explicitly imputed in the mixed effects model, there is an underlying assumption that the data are missing at random. Participants and time on study will be random effects.
 	
• Terms in the model:
  • **Response**: CAT score at each visit.
  • **Categorical**: type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control), smoking status, site (type of hospital).
  • **Continuous**: age, baseline CAT score, count of COPD exacerbations, time on study **Interaction:** type of subjects*time on study
• The model will be fitted with an unstructured variance-covariance matrix. If the model fails to converge, then alternative covariance structures may be investigated.
• Only CAT score measured after baseline are used as the response. So in this model a subject needs to have at least one post-baseline value to be included in the analysis.
• To investigate the effects of a particular covariates
  ○ Base model: type of patients, time on study, type of patients* time on study (interaction term) terms for the covariate and covariate-by-time interaction were added to the base model described above, these covariates including gender, smoking status, age, count of COPD exacerbations, maintenance therapy and ICS containing treatment
  ○ The slopes were estimated for each type of subjects ((i.e. GOLD I-IV, chronic bronchitis, healthy control)

Model Results Presentation
• The estimate of the slope is the parameter estimate associated with the type of subjects (Overall COPD subjects, GOLD I-IV, chronic bronchitis, healthy control) by time interaction in the model. The adjusted means, pairwise differences in different type of subjects, p-values and 95% confidence limits for the differences of interests will be summarized overall and also estimated for each visit.
Model Checking & Diagnostics
Refer to Appendix 7: Model Checking and Diagnostics for Statistical Analyses.

Endpoint(s)
Test the sensitivity and specificity of CAPTURE to identify possible COPD patients in a Chinese population
Model Specification
• CAPTURE (COPD Assessment in Primary Care to Identify Undiagnosed Respiratory Disease and Exacerbation Risk) is a simple, five-item, patient-completed questionnaire. • The algorithm is a simple summation of patient responses to each of the five items. (For Question1-4, Score 0 for No, Score 1 for Yes, Question 5, Score 0,1,2. So the total score is

ranging from 0 to 6). Higher score means higher risk of COPD (0-1 is at low risk, 2-6 is at high risk).

- Cut-off value (2 or 3 points) are used to test the sensitivity and specificity of CAPTURE questionnaire.
- Reference standard: COPD patients (GOLDI and GOLDII) and non-COPD patients (Chronic Bronchitis and never smoker healthy control)

Cut Point:	COPD patients	Non-COPD patients
Capture score (≥Cut-off)	TP	FP
Capture score (<Cut-off)	FN	TN

TP = number of true positive events
FP = number of false positive events
TN = number of true negative events
FN = number of false negative events
Positive predictive value (PPV)=TP/(TP+FP)
Negative predictive value (NPV)=TN/(TN+FN)
Estimated sensitivity = 100% * TP/(TP+FN)
Estimated specificity = 100% * TN/(FP+TN)

Sensitivity and Supportive Analyses

Sensitivity and specificity of CAPTURE score should be (score 3-6) vs (score 0-2) and (score 4-6) vs (score 0-3), respectively.

Model Results Presentation

- Estimated sensitivity and specificity for Capture Score

Endpoint(s)

Frequency of Clinically Important Deterioration (CID) composite outcome and its components

Model Specification

- The CID will be derived from three key clinical assessments, which are 1) moderate/severe exacerbations 2) worsening of FEV_1 and 3) worsening of health status using the CAT score.
- Definition of the composite CID event*: moderate/severe exacerbation and/or ⩾2-unit deterioration from baseline in CAT score and/or ⩾100 mL deterioration from baseline in FEV1 (CAT-containing CID).
- Two subgroups of participants were created based on CID status at Month 6, Month 18 and Month 30.
- CID status will be evaluated at visit 2/3/4.
-

Model Results Presentation

- Number and percentage of CID-positive and CID negative for each subgroup (GOLD I-IV, chronic bronchitis, healthy control) at each visit.

- Number and percentage of positive and negative response of each CID components (Exacerbation CID, FEV1 CID, CAT CID) at each visit for each subgroup

*There are two definitions of CID: one is SGRQ-containing definitions; the other is CAT containing definitions, here we use CAT containing definitions, because we only have the baseline data for SGRQ, but we have both baseline and follow up data for CAT score.

Endpoint(s)

- Time to the first CID composite event

Model Specification

- The CID will be derived from three key clinical assessments, which are 1) moderate/severe exacerbations 2) worsening of FEV_1 and 3) worsening of health status using the CAT score.
- Definition of the composite CID event*: moderate/severe exacerbation and/or ⩾2-unit deterioration from baseline in CAT score and/or ⩾100 mL deterioration from baseline in FEV1 (CAT-containing CID).
- Two subgroups of participants will be created based on CID status at Month 6, Month 18 and Month 30.
- CID status will be evaluated at visit 2/3/4.
- Cox proportional hazards model
- Terms in the model:
  - **Response**: Time to the first CID composite event.
    - **Categorical**: type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control), gender, smoking status, site (type of hospital), maintenance therapy and ICS containing treatment
    - **Continuous**: age, baseline FEV_1, baseline CAT score, count of COPD exacerbations
- The actual time to the first deterioration, of each individual component and the composite endpoint, will be analyzed using a Cox's proportional hazards model, with the 'exact' method for handling ties. (If the analysis will not run using the 'exact' method, then the 'Efron' method for handling ties will be used instead.)
- The time to the first deterioration based on the composite endpoint will be taken as the time to the first deterioration of any component event making up the composite.

Model Results Presentation

- A Kaplan-Meier plot with estimates of the proportion of subjects with CID event over time will be produced for each type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control) will be plotted on the same figure.
- The ratio of the estimated hazard rates for the comparison between different type of subject will be presented with 95% confidence intervals.
- Estimated probabilities of the occurrence of CID event at each visit will be shown with 95% confidence intervals for each type of subject.

Model Checking & Diagnostics

- Refer to Appendix 7: Model Checking and Diagnostics for Statistical Analyses.

Endpoints

Rate of EXACT events

Model Specification

- Generalized linear mixed model
- In sub-cohort population.
- Terms in the model:
  - **Response:** definition of EXACT events: magnitude of responder for symptomatic events defined as acute, sustained symptomatic worsening of COPD, defined as an increase in EXACT score ≥9 points for 3 days or ≥12 points for 2 days, above baseline (average in 7-day period)
  - **Dependent variable:** Events response (Yes/No)
  - **Categorical**: month, type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control), gender, smoking status, site (type of hospital), maintenance therapy (Y/N) and ICS containing (Y/N).
  - **Continuous**: age, baseline EXACT total score
  - **Interaction:** baseline*month
  - **Repeated:** month
- The model will be fit with an unstructured variance-covariance matrix
- For sub cohort population, visit 1 (month1-6) and visit 3 (month 18-23)

Model Results Presentation

- Number and percentage of events and non-events for each subgroup (GOLD I-IV, chronic bronchitis, healthy control) at each month

Endpoint(s)

- Change in Evaluating Respiratory Symptoms in COPD (E-RS: COPD) scores

Model Specification

- Mixed Models Repeated Measures (MMRM) model.
- While missing data are not explicitly imputed in the MMRM analyses, there is an underlying assumption that the data are missing at random.
- The E-RS utilizes 11 respiratory symptom items from the existing and validated 14-item EXACT, which measures the symptoms of exacerbation with 3 subscales assess breathlessness, cough and sputum, and chest symptoms. Our analyses will include both the total score and the 3 subscale scores.
- The MMRM analysis for E-RS scores measurements were available for sub-cohort at (months 1-6 and months 18-23).
- Terms in the model:
  - **Response**: Change in E-RS score at each visit.
  - **Categorical**: visit, type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control), gender, smoking status, site (type of hospital), maintenance therapy (Y/N) and ICS containing (Y/N).
  - **Continuous**: age, baseline E-RS score
  - **Interaction:** baseline E-RS score*visit, type of subjects*visit
  - **Repeated**: visit

- The model will be fitted with an unstructured variance-covariance matrix.
- The Kenward and Roger method (KR) for approximating the denominator degrees of freedom and correcting for bias in the estimated variance-covariance of the fixed effects will be used.
- Baseline is defined in Section 5.2.

Two models will be fitted; one with a response variable of change from baseline and one with the response variable as the raw value.

Model Results Presentation

- Least-square (LS) means and LS mean change from baseline values for each type of subject (Overall COPD subjects, GOLD I-IV, chronic bronchitis, healthy control) will be presented with their associated standard errors as well as 95% CIs. The estimated difference along with corresponding standard error, 95% CI for each visit will be presented.
- The LS mean differences (and associated 95% CIs) for all visits will also be presented graphically.

Model Checking & Diagnostics

- Refer to Appendix 7: Model Checking and Diagnostics for Statistical Analyses.

Sensitivity and Supportive Analyses

A mixed effects model whereas time is entered as a random effect will be analysed as a supportive analysis to estimate the longitudinal changes if necessary.

Endpoints

- Change from baseline in airway thickness by lung HRCT
- Change in lung density by lung HRCT
- Small airways disease measured using estimates of gas trapping

Model Specification

- Similar to analyses of change in CAT score with MMRM
- In sub-cohort population.

Model Checking & Diagnostics

- Refer to Appendix 7: Model Checking and Diagnostics for Statistical Analyses.

Endpoint(s)

- Change in physical activity measures (daily steps)

Model Specification

- Mixed Models Repeated Measures (MMRM) model.
- While missing data are not explicitly imputed in the MMRM analyses, there is an underlying assumption that the data are missing at random.
- The MMRM analysis for daily steps were available for sub-cohort for one month period at visit 1 and visit 3 .
- Terms in the model:
  - **Response**: Change in daily steps.

- **Categorical**: week, type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control), gender, smoking status, site (type of hospital). maintenance therapy (Y/N) and ICS containing (Y/N).
- **Continuous**: age, baseline daily steps
- **Interaction:** baseline daily steps*week, type of subjects*week
- **Repeated**: week

- The model will be fitted with an unstructured variance-covariance matrix.
- The Kenward and Roger method (KR) for approximating the denominator degrees of freedom and correcting for bias in the estimated variance-covariance of the fixed effects will be used.
- Baseline is defined in Section 5.2.

Model Results Presentation

- Least-square (LS) means and LS mean change from baseline values for each type of subject (Overall COPD subjects, GOLD I-IV, chronic bronchitis, healthy control) will be presented with their associated standard errors as well as 95% CIs. The estimated difference along with corresponding standard error, 95% CI for each weeks will be presented.
- The LS mean differences (and associated 95% CIs) for all weeks will also be presented graphically.

Model Checking & Diagnostics

- Refer to Appendix 7: Model Checking and Diagnostics for Statistical Analyses.

Subgroup Analyses

This will include analysis by age (<65 years, ≥65), gender (Female, Male), GOLD I-IV baseline reversibility, number of exacerbations in previous year (2, 3, ≥4) and smoking status (ever/never smoker).

Sensitivity and Supportive Analyses

A mixed effects model whereas time is entered as a random effect will be analysed as a supportive analysis to estimate the longitudinal changes if necessary.

Endpoint(s)

- Exacerbation in the 18 months in the full population

Model specification

- Logistic regression with binary endpoint
- Response variable:
- moderate exacerbation in the 18 months period vs exacerbation free for 18 months in the full population
  - severe exacerbation vs exacerbation free or only moderate exacerbation for 18 months in the full population
  - moderate or severe exacerbation vs exacerbation free for 18 months in the full population
- Predictors/ covariates

- **Continuous variables**: age, Number of pack-years, Total SGRQ score, FEV1 % pred (post), FVC % pred (post), mMRC, CAT, Hemoglobin and remove (some of) aforementioned covariate(s) or add other covariate(s) if needed.
- **Categorical variables**: Gender, Smoking status (former vs never, current vs never), Education level (high school vs less high, above school vs less high), Type of hospital, Household income (low income <=80000 vs high income >80000), GOLD grade (II vs I, III vs I, IV vs I), Neutrophil/lymphocyte ratio (low <5, high ≥5), Eosinophils (low <300, high >=300), CRP (low <=3, high >30), Fibrinogen (low <400 high >=400), Asthma diagnosed before age 40 (either in Medical History or Childhood Respiratory Asthma is diagnosed as yes, then labelled as yes), Moderate exacerbations in one year (yes or no), Severe exacerbations in one year (yes or no), history of CB (Has the subject had bronchitis or chronic coughing with phlegm or sputum from the chest, for at least past 2 years, and at least 3 months in past year), Diabetes (yes-no), Gastroesophageal Reflux and remove (some of) aforementioned covariate(s) or add other covariate(s) if need.

- The data set will be randomly participated into training dataset (approximate two thirds of the full sample size, adjusted if necessary) and testing dataset (approximate one third of the full sample size, adjusted if necessary).
- The training dataset is used to fit the model using the predictors significant in the univariate model (P<0.05, the criteria can be adjusted if necessary).
- The testing dataset is used to evaluate the performance of the fitted model where the C-index will be calculated, the predictors of the testing dataset will be same as those of training dataset.

Model Results Presentation

- For multivariate logistic regression model, fitted Statistics (-2 Log Likelihood, AIC, BIC, r square), regression coefficient, 95% confidence intervals of the regression coefficient, P value, C-index (Concordance index), will be presented based on the training dataset.
- C-index (Concordance index) based on the testing dataset will be reported separately.
- Other Statistics will be reported if applicable.

Sensitivity Analyses

- Two multivariate logistic regression models will be built based on AIC stepwise, BIC stepwise criteria to be compared.
- Use the same training dataset and testing dataset as the previous multivariate logistic regression model.
- The training dataset is used to fit the model while the predictors include all continuous variables and categorical variables mentioned in model specification
- The reported statistics will be same as model results presentation

Demo examples for multivariate logistic regression of the full population

For the model using the predictors significant in the univariate model

```
ods graphics on;

 ods output
```

```
  CLparmWald=CLparmWald

  FitStatistics=FitStatistics

  ParameterEstimates=ParameterEstimates

  RSquare=RSquare

  ROCAssociation=ROCAssociation

  ROCCurve=ROCCurve;

 proc logistic data=mulregdata;

/*the data set consists of response variable and significant
predictors in the univariate logistic model*/

    model response variable(event='1')= x1-xn/rsquare cl;

/*significant predictors in the univariate logistic model, named as
Model 1 for easy reference*/

   score data= mytestingdata out=valpred outroc=vroc;

    roc;

run;
```

For the model using AIC and BIC criteria

```
ods graphics on;

 ods output FitStatistics=FitStatistics

            ParameterEstimates=ParameterEstimates

            Association=Association;

proc hplogistic data=pooldata;

/*the data set consists of training dataset, grouped as 1, and testing
dataset, grouped as 2*/

   model response variable(event='1') = all the listed
predictors/association ctable=Roc lackfit;

   partition rolevar=group(train='1' test='2');

   selection method=stepwise (SELECT=AIC CHOOSE=AIC STOP=AIC);

/*For BIC, change to (SELECT=BIC CHOOSE=BIC STOP=BIC)*/

run;
```


Endpoint(s)
Exacerbation in the 18 months in the sub cohort
Model specification

• Logistic regression with binary endpoint
• Response variable: Refer to the full population case
• Predictors/ covariates
  • **Continuous variables**: all continuous variables in full population case and LAA950, pi10, Steps walked within first week, DPM functional small airway disease, DPM tissue destruction, RV/TLC ratio and remove (some of) aforementioned sub cohort covariate(s) or add other sub cohort continuous variables if necessary.
  • **Categorical variables**: all categorical variables in full population case and other sub cohort categorical variables if necessary
• All the sub cohort data will be used for model fitting, i.e., the training dataset is sub cohort data where the predictors are those are significant in the univariate model (P<0.05, relaxed to 0.1 if necessary).
• Create testing data by bootstrap method, i.e., resampling method with replacement, where the sample size of testing dataset is same as the sub cohort training dataset. The boot strap time is set to be 100, adjusted if necessary. For each bootstrap procedure, seed is required to set. For each testing dataset, the first type of C-index will be calculated based on the (same) model trained using the sub cohort data. The second type of C-index will be calculated based on the (same) model trained using the training dataset of the full population of the previous full population analysis section.
• The final two types of C-indexes derived based on the testing datasets will be present as the means and 95%CIs of the C-indexes based on the 100 bootstrap testing datasets.

Model Results Presentation
• For multivariate logistic regression model, fitted Statistics (-2 Log Likelihood, AIC, BIC, r square), regression coefficient, 95% confidence intervals of the regression coefficient, P value, C-index (Concordance index), will be presented based on the training dataset. • The means and 95%CIs of the two types of C-indexes based on the 100 bootstrap testing datasets will be reported separately. • Other Statistics will be reported if applicable.
Sensitivity Analyses
• Two multivariate logistic regression models will be built based on AIC stepwise, BIC stepwise criteria to be compared.

- The training dataset is sub cohort data and it is used to fit the model while the predictors include all continuous variables and categorical variables mentioned in model specification
- Create testing data by bootstrap method, i.e., resampling method with replacement, where the sample size of testing dataset is same as the sub cohort training dataset. The boot strap time is set to be 100, adjusted if necessary. For each bootstrap procedure, seed is required to set. For each testing dataset, the first type of C-index will be calculated based on the (same) model trained using the sub cohort data. The second type of C-index will be calculated based on the (same) model trained using the training dataset of the full population of the previous full population analysis section.
- The final two types of C-indexes derived based on the testing datasets will be present as the means and 95%CIs of the C-indexes based on the 100 bootstrap testing datasets.

Demo examples for multivariate logistic regression of the sub cohort population

Train model using the predictors significant in the univariate model

```
ods graphics on;

   ods output

       CLparmWald=CLparmWald

        FitStatistics=FitStatistics

        ParameterEstimates=ParameterEstimates

        RSquare=RSquare

       ROCAssociation=ROCAssociation;

 proc logistic data=subcohortmulregdata;

/*the data set consists of response variable and significant
predictors in the univariate logistic model of sub cohort data*/

    model response variable (event='1')= x1-xm/rsquare cl;

/*significant predictors in the univariate logistic model, named as
Model 2 for easy reference */

run;
```

Train model using AIC and BIC criteria

```
ods graphics on;

   ods output FitStatistics=FitStatistics

              ParameterEstimates=ParameterEstimates
```

```
                Association=Association;

proc hplogistic data=subcohortdata;

/*the data set is sub cohort dataset*/

   model response variable(event='1') = all the listed
predictors/rsquare cl;

   selection method=stepwise (SELECT=AIC CHOOSE=AIC STOP=AIC);

/*For BIC, change to (SELECT=BIC CHOOSE=BIC STOP=BIC)*/

run;
```

Calculate C-index of the bootstrap testing dataset when the training model uses the predictors significant in the univariate model

```
%macro repeat();

 %do j=1 %to number of bootstrap time;
      %let seednum=%eval(100+&j);

      %put &seednum;

     /* resampling the j-th bootstrap testing dataset, i.e.,
subcohortBOOTdata_&j._ is the j-th bootstrap resampling testing
dataset in loop.*/

    proc surveyselect data= subcohortdata out= subcohortBOOTdata_&j._
NOPRINT

      seed=&seednum

      method=urs sampsize=sample size of subcohortdata

      outhits;

   run;

/*The first type of C-index will be calculated based on the (same)
model trained using the sub cohort data. */
  ods graphics on;

 ods output ROCCurve=ROCCurve_repeatSUB_&j._;

 proc logistic data= subcohortmulregdata;

 model response variable(event='1')= x1-xm; /* this is the same model
as the model in the above step of training  model using the predictors
significant in the univariate model of the sub cohort data, i.e., same
model as Model 2 */
```

```
 score data= subcohortBOOTdata_&j._ out=valpred outroc=vroc;

 roc;

run;

/* The second type of C-index will be calculated based on the (same)
model trained using the training dataset of the full population */
  ods graphics on;

 ods output ROCCurve=ROCCurve_repeatFULL_&j._;

 proc logistic data=mulregdata;

 model response variable(event='1')= x1-xn; /* this is the model of
using the predictors significant in the univariate model of the full
population in the previous full population analysis section, i.e.,
same model as Model 1*/

 score data= subcohortBOOTdata_&j._ out=valpred outroc=vroc;

 roc;

run;

%end;

%mend repeat;

%repeat;
```

Calculate C-index of the bootstrap testing dataset when the training model uses AIC and BIC criteria

```
%macro repeat();

 %do j=1 %to number of bootstrap time;
      %let seednum=%eval(100+&j);

      %put &seednum;

     /* resampling the j-th bootstrap testing dataset, for each loop,
it is the same dataset as subcohortBOOTdata_&j._ in the previous loop
example */

    proc surveyselect data= subcohortdata out= subcohortBOOTdata_&j._
NOPRINT

      seed=&seednum

      method=urs sampsize=sample size of subcohortdata
```

```
      outhits;

   run;

/*The first type of C-index will be calculated based on the (same)
model trained using the sub cohort data. */
     ods graphics on;

     ods output Association=Association_repeatSUB_&j._;

      proc hplogistic data= subcohortBOOt_all_&j._;

/* For j-th loop, subcohortBOOt_all_&j._ consists of the sub cohort
data, grouped as 1, and a bootstrap testing subcohortBOOTdata_&j._,
grouped as 2, which is obtained in the resampling process*/

      model response variable (event='1') = all the predictors
listed/association          ctable=Roc lackfit;

     partition rolevar=group(train='1' test='2');

      selection method=stepwise (SELECT=AIC CHOOSE=AIC STOP=AIC);

     /*For BIC, change to (SELECT=BIC CHOOSE=BIC STOP=BIC)*/

run;

/* The second type of C-index will be calculated based on the (same)
model trained using the training dataset of the full population */
     ods graphics on;

     ods output Association=Association_repeatFULL_&j._;

      proc hplogistic data= FULLsubcohortBOOt_all_&j._;

/* For j-th loop, FULLsubcohortBOOt_all_&j._ consists of the (same)
training dataset from the full population of the previous full
population analysis section, grouped as 1, and a bootstrap testing
subcohortBOOTdata_&j._, grouped as 2, which is obtained in the
resampling process*/

      model response variable (event='1')= all the predictors
listed/association          ctable=Roc lackfit;

     partition rolevar=group(train='1' test='2');

      selection method=stepwise (SELECT=AIC CHOOSE=AIC STOP=AIC);

     /*For BIC, change to (SELECT=BIC CHOOSE=BIC STOP=BIC)*/

run;

%end;
```

```
%mend repeat;

%repeat;
```

Endpoint(s)

- Moderate or severe exacerbation in the 18 months in the full population

Model specification

- Logistic regression with binary endpoint
- Response variable:
  - moderate or severe exacerbation vs exacerbation free for 18 months in the full population
- Predictors/ covariates
  - **Set 1:**
  - **Continuous variables**: CAT
  - **Categorical variables**: GOLD grade (II vs I, III vs I, IV vs I), Moderate exacerbations in one year (yes or no), Severe exacerbations in one year (yes or no), history of CB (Has the subject had bronchitis or chronic coughing with phlegm or sputum from the chest, for at least past 2 years, and at least 3 months in past year), Gastroesophageal Reflux and remove (some of) aforementioned covariate(s) or add other covariate(s) if need.
  - **Set 2:**
  - **Continuous variables**: mMRC
  - **Categorical variables**: GOLD grade (II vs I, III vs I, IV vs I), Moderate exacerbations in one year (yes or no), Severe exacerbations in one year (yes or no), history of CB (Has the subject had bronchitis or chronic coughing with phlegm or sputum from the chest, for at least past 2 years, and at least 3 months in past year), Gastroesophageal Reflux and remove (some of) aforementioned covariate(s) or add other covariate(s) if need.
  - **Set 3:**
  - **Continuous variables**: CAT, FEV1 % pred (post)
  - **Categorical variables**: Moderate exacerbations in one year (yes or no), Severe exacerbations in one year (yes or no), history of CB (Has the subject had bronchitis or chronic coughing with phlegm or sputum from the chest, for at least past 2 years, and at least 3 months in past year), Gastroesophageal Reflux and remove (some of) aforementioned covariate(s) or add other covariate(s) if need.
  - **Set 4:**
  - **Continuous variables**: mMRC, FEV1 % pred (post)
  - **Categorical variables**: Moderate exacerbations in one year (yes or no), Severe exacerbations in one year (yes or no), history of CB (Has the subject had bronchitis or chronic coughing with phlegm or sputum from the chest, for at least past 2 years, and at least 3 months in past year), Gastroesophageal Reflux and remove (some of) aforementioned covariate(s) or add other covariate(s) if need.
- The whole data set used to fit the model. If the predictor is significant in the univariate model ($P<0.05$, the criteria can be adjusted if necessary), it will be incorporated in

multivariate logistic regression model. Four multivariate logistic regression models will be fitted, the covariates set will used the for four sets separately.

Model Results Presentation

- For multivariate logistic regression model, fitted Statistics (-2 Log Likelihood, AIC, BIC, r square), regression coefficient, 95% confidence intervals of the regression coefficient, P value, C-index (Concordance index), will be presented.
- Other Statistics will be reported if applicable.

Sensitivity Analyses

- Two multivariate logistic regression models based on each covariates set will be built based on AIC stepwise, BIC stepwise criteria to be compared.
- The whole data set is used to fit the model while the predictors include all continuous variables and categorical variables mentioned in model specification.
- The reported statistics will be same as model results presentation

Demo examples for multivariate logistic regression of the full population

For the model using the predictors significant in the univariate model

```
ods graphics on;

 ods output

  CLparmWald=CLparmWald

  FitStatistics=FitStatistics

  ParameterEstimates=ParameterEstimates

  RSquare=RSquare

  ROCAssociation=ROCAssociation

proc logistic data=mulregdata;

/*the data set consists of response variable and significant
predictors in the univariate logistic model of all GOLD III and GOLD
IV */

    model response variable(event='1')= x1-xn/rsquare cl;

/*significant predictors in the univariate logistic model of each
covariates set, i.e., set 1-4 will be fitted separately*/

run;
```

For the model using AIC and BIC criteria

```
ods graphics on;

 ods output FitStatistics=FitStatistics

            ParameterEstimates=ParameterEstimates

            Association=Association;

proc hplogistic data= fullpopulationdata;

/*the dataset consists of all subjects of full population, each
covariates set, set 1-4 will be fitted separately */

    model response variable(event='1') = all the listed
predictors/association ctable=Roc lackfit;

selection method=stepwise (SELECT=AIC CHOOSE=AIC STOP=AIC);

/*For BIC, change to (SELECT=BIC CHOOSE=BIC STOP=BIC)*/

run;
```

8. PHENOTYPIC ANALYSES

The data available at Visit 1 will assist in determination of subject phenotypes in disease manifestation. Phenotypes associated with progression of disease will be based on longitudinal data. Phenotype identification will also refer to the findings of previous Western cohorts.

Unless otherwise specified, endpoints / variables defined in Section 8.1 will be summarised using descriptive statistics and listed.

8.1. Endpoint / Variables

Table 5. Overview of Planned Phenotypic Analyses

Endpoint / Parameter / Display Type	Data Displays Generated		
	Table	Figure	Listing
Phenotypes			
Exacerbator	Y[1,2,3]		Y[1,2]
CMH	Y[1,2,3]		Y[1,2]
Significant Symptoms	Y[1,2,3]		Y[1,2]

Endpoint / Parameter / Display Type	Data Displays Generated		
	Table	Figure	Listing
ACO	Y[1,2,]		Y[1,2]
Endotypes			
Airflow Limitation	Y[1,2,]		Y[1,2,]
Higher Blood Eosinophils	Y[1,2,]		Y[1,2]
Fibrinogen	Y[1,2,]		Y[1,2]
Pure airways disease	Y[2,3]		Y[2,3]
Sputum microbiome	Y[2,3]		Y[2,3]
PRISm (preserved ratio impaired spirometry)	Y[2,3]		Y[2,3]
Presence of emphysema	Y[2,3]		Y[2,3]

Note:
1. Main cohort population
2. Sub-cohort population
3. Exacerbation population

8.2. Population of Interest

The phenotypic analyses will be based on the full population and sub-cohort population. frequency and percentage for categorical variables and mean and standard deviation for continues variables.

9. ANALYSES OF TREATMENT PATTERN AND HEALTHCARE RESOURCE UTILIZATION/COST

9.1. Treatment Pattern analyses

Endpoint 1 Each type of treatment for COPD patients

Descriptive statistics will be calculated and tabulated using frequency and percentage for categorical variables and mean and standard deviation for continues variables during baseline period. The overall Year 1 treatment (LAMA, LABA, ICS/LABA, LAMA/LABA, open triple, closed triple) will be presented, overall, by specific characteristics at enrolment (Gender, Age, Smoking Status, enrolment status, Type of hospital, GOLD grade, COPD group with CAT score, Exacerbation status, Hospitalization in preceding year, Presence of chronic bronchitis, Prescribing costs changed for triple during COMPASS and season of prescription). The entire treatment data will be presented in the individual data listings appended to the Clinical Study Report.

Endpoint(s)

Change between treatment pattern during study period

Model Specification

- Initiation of maintenance treatment

Treatment pattern	Combination
LAMA	LAMA
LABA	LABA
ICS/LABA	ICS/LABA, or Combination 2, at the same time check: ”ICS”+”LABA”
LAMA/LABA	LABA/LAMA, or Combination 2, at the same time check: ”LABA”+”LAMA”
Triple(open)	When “Triple (open) =1”: Combination 1, at the same time check: “ICS/LABA”+”LAMA”, or Combination 2, at the same time check: “ICS/LABA”+”LABA/LAMA”, or Combination 3, at the same time check: “ICS”+”LABA/LAMA”, or Combination 4, at the same time check: ”ICS”+”LABA”+”LAMA”
Triple(closed)	ICS/LABA/LAMA

- Switches to another maintenance treatment
- Addition of other maintenance treatment
- Stepping up to triple therapy
- Discontinuation of treatment

All the data were captured by the HIS (hospital information system).

9.2. Health resource utilization

Table 6. Overview of the Health Resource Utilization

Endpoint / Parameter / Display Type	Data Displays Generated		
	Table	Figure	Listing
Health Resource Utilization			
Number of outpatient visits	Y		Y
Number of emergency visits	Y		Y
Number of hospitalizations, length of stay	Y		Y
Treatment of exacerbation	Y		Y
Cost			
Total medical cost	Y		Y
Pharmacy cost	Y		Y
Other cost	Y		Y
Outpatient cost	Y		Y
Emergency cost	Y		Y

Endpoint / Parameter / Display Type	Data Displays Generated		
	Table	Figure	Listing
Hospitalization cost	Y		Y
Exacerbation cost	Y		Y

Summaries of healthcare resource utilization/cost will be provided at visit 2, 3, 4 for main cohort population and sub-cohort population. Subgroup analyses will include GOLD I-IV patients, chronic bronchitis patients and healthy control participants. Descriptive analyses will be calculated and tabulated by the abovementioned subgroups, using frequency and percentage for categorical variables and mean and standard deviation for continues variables, for healthcare resource utilization during the study. Stratified analysis will be performed according to COPD and non-COPD patients.

10. SAFETY ANALYSES

The safety analyses will be based on the full population, unless otherwise specified.

10.1. Adverse Events Analyses

Number and percentage of subjects with study procedure-related SAEs will be summarized.

10.2. COVID-19 Assessment and COVID-19 AEs

Number of participants with suspected, probable or confirmed for COVID-19 infection will be summarized according to the “COVID-19 Coronavirus Infection Diagnosis” and “COVID-19 Coronavirus Infection Assessment” eCRF page.

All COVID-19 AEs will be identified prior the DBF via a medical review of all AE preferred terms. Number of participants with COVID-19 as reported as an AE are summarized. All COVID-19 AEs will be flagged in the AE listings.

10.3. Length of Time in Study

Length of time in study will be calculated as the number of days from enrollment to study completion or study discontinuation (i.e., the span of time between Visit 1 and final visit: Final visit date – Visit 1 date + 1). This will be summarized as a discrete variable based on ranges of days. The number of subjects that attend each visit will also be tabulated.

11. REFERENCES

Atkinson AC. Plots, transformations and regression. Clarendon Press, Oxford. 1985.

GlaxoSmithKline document number 2017N345807_01: Investigation of the clinical, radiological and biological factors associated with disease progression, phenotypes and endotypes of COPD in China -Amendment 1.

Leidy NK, Murray LT. Patient-reported outcome (PRO) measures for clinical trials of COPD: the EXACT and E-RS. Journal of Chronic Obstructive Pulmonary Disease, 2013; 10:393-398.

Jones, P. St George 'S Respiratory Questionnaire for Copd Patients (Sgrq-C) Manual ,2016, Version No.1.3.

Kim, V., Dolliver, W.R., Nath, H.P. et al. Mucus plugging on computed tomography and chronic bronchitis in chronic obstructive pulmonary disease. Respir Res 22, 110 (2021). https://doi.org/10.1186/s12931-021-01712-0

12. APPENDICES

12.1. Appendix 1: Protocol Deviation Management and Definitions for Per Protocol Population

Protocol deviations will be tracked by the study team throughout the conduct of the study in accordance with the PDMP. Important protocol deviations as identified in the PDMP will be summarized and listed.

12.1.1. Exclusions from Per Protocol Population Analysis

This is an observational cohort study, any subjects with protocol deviations will not be removed from the study analyses. Per protocol population is not defined in this study.

12.2. Appendix 2: Schedule of Activities

12.2.1. Protocol Defined Schedule of Events

Protocol Activity		Screen/Baseline	Study visit				
visit		**V1**	**V2**	**V3**	**V4**	**EW**	**Exacerbation visit (≤2)[1]**
Study month		**M0**	**M6**	**M18**	**M30**		
window		**+3 days**	**±28 days**	**±28 days**	**±28 days**		
Procedures							
Informed consent[2]		X	-	-	-	-	-
Demography		X	-	-	-	-	-
Medical history		X	-	-	-	-	-
Health and lifestyle information/Exposures		X	-	-	-	-	-
Inclusion/Exclusion Criteria		X	-	-	-	-	-
COPD and Exacerbation History[3]		X	-	-	-	-	-
Concurrent medications		X	X	X	X	X	-
Smoking status		X	X	X	X	X	-
Spirometry[7]		X	X	X	X	X	X
Reversibility Testing		X	-	-	-	-	-
Patient reported	mMRC	X	-	-	-	-	
	SGRQ-C	X	-	-	-	-	
	CAPTURE	X		X			
Questionnaires	CAT (on-site visit)	X	X	X	X	-	X
	CAT (Phone Call)	Trimonthly Phone Call[4]					
Healthcare resource utilization and cost		-	X	X	X	X	X
Assessments							
COPD exacerbation assessment (on site visit)[5]		-	X	X	X	X	X
COPD exacerbation assessment (Phone Call)[5]		Trimonthly Phone Call[4]					
COPD medication and treatment (on site visit)[5]		X	X	X	X	X	X
COPD medication and treatment (Phone Call)[5]		Trimonthly Phone Call[4]					
Death		-	X				
Serious Adverse Event assessment[12]			X	X	X	X	X
Lab							
Blood Draw for e.g. fibrinogen/HsCRP/Cell count		X	-	-	X	X	X

Protocol Activity	Screen/Baseline	Study visit				
Sub-cohort only						
Lung HRCT scan[6]	X	-	-	X	X	X
Daily steps	X[8]	-	X[8]	-	-	-
Evening diary (EXACT plus additional questions)	X[9]	-	X[9]	-	-	-
Monthly CAT	X[10]	-	X[10]	-	-	-
Blood biomarkers (inclusive of serum sRAGE, CC16, IP-10, HbA1c)	X	-	-	X	X	X
Sputum cytology[11]	X	-	-	X	X	X
Sputum microbiome[11]	X	-	-	X	X	X

CAT=COPD assessment test, COPD=Chronic Obstructive Pulmonary Disease, HRCT=High-resolution computed tomography, EW=Early Withdrawal, EXACT=EXAcerbation of COPD Tool; SGRQ-C=St.George's Respiratory Questionnaire for COPD.

1. Exacerbation visits only apply to subjects included in the Biomarker & Digital sub-cohort. Moderate exacerbations are defined as COPD exacerbations that require either systemic corticosteroids (intramuscular (IM), intravenous, or oral) and/or antibiotics. Severe exacerbations are defined as COPD exacerbations requiring hospitalization or result in death.
2. Informed consent must be conducted at the screen visit prior to performing any study procedures.
3. For COPD participants only;
4. When a trimonthly telephone interview overlaps a clinical visit, the COPD exacerbation assessment, COPD medication and treatment questionnaires will be completed at the clinic visit.
5. Data collection of COPD exacerbation assessment, COPD medication and treatment is not required for the healthy control participants.
6. CT scan should only be performed at withdrawal visit if the last study CT performed was more than a year ago. Inspiratory and expiratory HRCT scans are required at Visit 1 and Visit 4, while only inspiratory scan will be conducted on exacerbation visit.
7. Pre-and post- bronchodilator values of spirometry test will be collected at each site visit.
8. Physical activity will be evaluated by daily steps as assessed by a wrist band dispensed at baseline. Participants are required to keep wearing the wrist band for 1 month from Visit 1 (baseline) and Visit 3
9. Daily digital EXACT measurement for 6 months from Visit 1 (baseline) and Visit 3
10. Monthly digital measurement from baseline for 6 months, and monthly digital measurement from V3 for 6 months
11. Sputum can be collected spontaneously or can be induced, as per investigator judgement. Sputum sampling should only be done if, in the opinion of the investigator, it is safe for the subject.
12. Only SAEs relating to the study procedures are collected.

12.3. Appendix 3: Assessment Windows

12.3.1. Visit Based Assessments

Nominal visits (scheduled visit) will be used for reporting and analysis.

Assessment windows for the scheduled study visits are as follows:

Full population:

- Baseline/Screen (+ 3 days)
- Visit 2 (± 28 days)
- Visit 3 (± 28 days)
- Visit 4 (± 28 days)

In addition, COPD and chronic bronchitis participants will be contacted by phone on trimonthly basis within ± 7 days of the planned date to assess the CAT and to collect information on exacerbation rates and COPD medications; healthy controls will be contacted by phone on trimonthly basis within ± 7 days of the planned date to assess the CAT only.

12.4. Appendix 4: Data Display Standards & Handling Conventions

12.4.1. Reporting Process

Software
• The version 9.4 or later of SAS software will be used.
Analysis Datasets
• Analysis datasets will be created according to CDISC standards • For creation of ADaM datasets (ADCM/ADAE), the same version of dictionary datasets will be implemented for conversion from SI to SDTM.
Generation of RTF Files
• Rich Text Format (RTF) files will be generated for the final reporting effort for use in writing the CSR.

12.4.2. Reporting Standards

Reporting Standards
General
• The current GSK Integrated Data Standards Library (IDSL) will be applied for reporting, unless otherwise stated (IDSL Standards Location: https://spope.gsk.com/sites/IDSLLibrary/SitePages/Home.aspx): • 4.03 to 4.23: General Principles • 5.01 to 5.08: Principles Related to Data Listings • 6.01 to 6.11: Principles Related to Summary Tables • 7.01 to 7.13: Principles Related to Graphics
Formats
• GSK IDSL Statistical Principles (5.03 & 6.06.3) for decimal places (DP's) will be adopted for reporting of data based on the raw data collected, unless otherwise stated. • Numeric data will be reported at the precision collected on the eCRF. • The reported precision from non eCRF sources will follow the IDSL statistical principles but may be adjusted to a clinically interpretable number of DP's.
Planned and Actual Time

• Reporting for tables, figures and formal statistical analyses :
  • Planned time will be used in figures, summaries, statistical analyses and calculation of any derived parameters, unless otherwise stated.
  • The impact of any major deviation from the planned assessment times and/or scheduled visit days on the analyses and interpretation of the results will be assessed as appropriate.
• Reporting for Data Listings:
  • Planned and actual time will be shown in listings.
  • Unscheduled or unplanned readings will be presented within the subject's listings.

Reporting Standards	
• Visits outside the protocol defined time-windows (i.e. recorded as protocol deviations) will be included in statistical analyses.	
Unscheduled Visits	
• Unscheduled visits will not be included in summary tables. • Unscheduled visits will not be included in figures. • All unscheduled visits will be included in listings.	
Descriptive Summary Statistics	
Continuous Data	Refer to IDSL Statistical Principle 6.06.1
Categorical Data	N, n, frequency, %
Graphical Displays	
• Refer to IDSL Statistical Principals 7.01 to 7.13.	

12.5. Appendix 5: Derived and Transformed Data

12.5.1. General

Multiple Measurements at One Analysis Time Point
• Mean of the measurements will be calculated and used in any derivation of summary statistics but if listed, all data will be presented. • If there are two values within a time window, the value closest to the target day for that window will be used. If values are the same distance from the target, then the mean will be taken.
Study Day

• Calculated as the number of days from visit 1:
  • Ref Date = Missing → Study Day = Missing
  • Ref Date < Visit 1 → Study Day = Ref Date – Visit 1
  • Ref Data ≥ Visit 1 → Study Day = Ref Date – Visit 1 + 1

12.5.2. Study Population

Demographics
Age

• Age will be calculated based on the Visit 1.
• Birth date will be imputed as follows:
  • Any subject with a missing day will have this imputed as day '15'.
  • Any subject with a missing date and month will have this imputed as '30JUN'.
• Birth date will be presented in listings as 'YYYY'.

Age categories are based on at screening and are defined as: • <65 years • ≥65 years

Subject disposition
Discontinuation from Study Treatment
Not Applicable.
Discontinuation from Study
Censoring will be performed as follows: • For subject that withdraws consent for disclosure of future information, any data collected prior to withdrawal will be retained and reported. • For subject that withdraws consent for disclosure, the site may use any publicly information to determine vital status at the end of the study and this information would be included in the final analyses. • If a subject withdraws from the study, he/she may request destruction of any samples taken and not tested, such request will be documented in study records.

Subject disposition
Discontinuation from Study Treatment
• Subject that is unreachable and fails to come back for both Visit 3 and Visit 4 will be considered lost to follow-up. Data collected prior to missed visits will be retained and reported.

Smoking status
• Smoking status: Never smokers are defined as those with a lifetime exposure of <1 package/year in their life (excluding passive smoker). Ever smokers (current or former) are defined as those with a lifetime exposure of ≥10 package/year. Former smokers are those who have stopped smoking for at least 6 months. Current smokers are those who are currently smoking. Passive smokers are defined as those with the presence of smokers living in the same household, or co-workers smoking nearby while indoors • Smoking status will be reassessed at each scheduled clinic visit evaluated by participants' self-report (e.g. never smoker, current smoker, former smoker). • Smoking history including Pack-years of smoking: Number of pack years (number of cigarettes per day / 20) x number of years smoked, e.g., 10 packyears means 20 cigarettes per day for 10 years, or 10 cigarettes per day for 20 years)

COPD Exacerbation History
• Individual exacerbations during the 12/24 months prior to Screening and during the study were to be collected on the eCRF by the investigator • COPD exacerbations with onset date on or after the date of Screening will not be counted in the counts of number of exacerbations in the past year; all other exacerbations recorded on the eCRF will be included in the counts, even if onset/resolution dates were missing or earlier than the start of the exact 12 month period prior to Screening • Number of COPD exacerbations reported in the past year prior to Screening will be summarised according to three categories: moderate COPD exacerbation, severe COPD exacerbation and moderate/severe COPD exacerbation. • Moderate COPD exacerbations are defined as exacerbation that required treatment with oral/systemic corticosteroids and/or antibiotics (not involving hospitalisation). • Severe COPD exacerbations are defined as exacerbations that required in-patient hospitalisation. • Total number of moderate/severe COPD exacerbations are defined as total numbers of moderate and severe COPD exacerbation for each subject.

Baseline Lung Function
Reversibility
Reversibility will be calculated as follows: (Highest post-bronchodilator FEV_1 – Highest pre-bronchodilator FEV_1)*100/(Highest post-bronchodilator)
GOLD Grade 1-4 at Screening

Baseline Lung Function
Reversibility
Reversibility will be calculated as follows: (Highest post-bronchodilator FEV_1 – Highest pre-bronchodilator FEV_1)*100/(Highest post-bronchodilator)
Subjects will be classified into Global Initiative on Obstructive Lung Disease (GOLD) Grades 1-4 using the post-salbutamol percent predicted FEV1 assessment at Screening: • GOLD Grade 1 (Mild): percent predicted FEV1 ≥80% • GOLD Grade 2 (Moderate): 50%□□percent predicted FEV1 < 80% • GOLD Grade 3 (Severe): 30%□□percent predicted FEV1 < 50% • GOLD Grade 4 (Very Severe): percent predicted FEV1 < 30%

Death
Details of any patient deaths, including date of death, will be recorded in the CRF. Information will be obtained from the telephone interview and from medical records in HIS. Cause of death will be recorded if available. Data from publicly available source will be used if participant is not contactable at the end of the study.

Blood Biomarker
• Blood samples will be taken at Visit 1 and throughout the study as defined in the time and events table. (Appendix 2: Schedule of Activities). • Plasma fibrinogen, hsCRP, total/differential blood cell count measurements and haemoglobin will be conducted at central lab for all participants • Additional protein analysis will be conducted at central lab for the sub-cohort using validated assays. Analytes chosen will include serum sRAGE, CC16 and IP-10， HbA1c. • The blood biomarker assessments will be conducted according to the study reference manual.

Sputum Microbiome and Sputum Cytology
• Sputum samples for microbiome and sputum cytology will be collected at Vist1 and Visit 4 in sub-cohort only as defined in the time and events table (Appendix 2: Schedule of Activities). • Sputum can be collected spontaneously or can be induced, as per investigator judgement, induced sputum is preferred. Sputum sampling should only be done if, in the opinion of the investigator, it is safe for the subject. The sputum microbiome and sputum cytology assessments will be conducted according to the sputum assessment procedure (Refer to study reference manual). • Additional analyses may include the measurement of viruses and fungi.

CCI

CCI [redacted]

COPD-related Healthcare Resource Utilization

Within site hospital COPD-related healthcare resource utilization and cost will be captured on the study questionnaire form where possible using the HIS (based on site feasibility). Information collected include:
• Unscheduled COPD outpatient visits and emergency visits
  ○ Number of visits, date of visit, diagnosis of visit, cost
• COPD hospitalization (including intensive care)
  ○ Number of hospitalizations, date and length of stay, diagnosis of visit, cost
• Prescribed COPD medications
  ○ Medication name, date of prescription, dosage, administration, cost, whether the prescription is picked up at hospital, etc.
• Other COPD treatment, e.g. oxygen therapy, nebulizers
  ○ Treatment name, date, cost, etc.
• Lab tests, image tests, surgeries, and other tests and procedures (e.g., spirometry, Chest X-Ray, CT)
  ○ Test or procedure, date of operation, cost, etc.
• Healthcare cost if feasible, e.g. general service usage, consultation fee
  ○ Item, date of charge
• Insurance coverage and patient out-of-pocket payment if feasible
 	
Additionally, healthcare resource utilization could occur in other places outside of the site hospitals including: (1) hospitals other than the site hospital where the subject is recruited; (2) retail pharmacy stores; (3) subjects' home.
 	
Outside of site hospital COPD-related healthcare resource utilization and cost will be collected using standard questionnaires at trimonthly subjects' phone interviews. Information collected include COPD-related diagnosis and treatment in the following sources:
• Outpatient visit, emergency visit, and hospitalization
  ○ Date, reason, length of stay (for hospitalization)
• Prescribed medications
• Over-the-counter (OTC) medications
• Other therapies, e.g. oxygen therapy, nebulizers
• Associated costs if feasible
• Lab tests, image tests, surgeries, and other tests and procedures (e.g., spirometry, Chest X-Ray, CT)

Adverse Events

Adverse Events (AE) will not be captured in the study as no study drug will be given in this study.

- If an AE (serious or non-serious) or complaint with any specifically named GSK product is reported spontaneously by a subject during the course of the study, the site staff will complete the AE reporting form and transmit the report to the GSK China within 24 hours of being made aware of the AE.

Study procedure related serious adverse events (SAE) will be recorded.

- All SAEs will be collected from the signing of the ICF until the follow-up visit at the time points specified in the Schedule of Activities (Appendix 2: Schedule of Activities)
- The investigator is responsible for the detection and documentation of events meeting the criteria and definition of an SAE, provided in this protocol. During the study when there is a safety evaluation, the investigator or site staff will be responsible for detecting, documenting, and reporting SAEs

Serious Adverse Event (SAE) is defined as any untoward medical occurrence that:

- Results in death,
- Is life-threatening,
- Requires hospitalization or prolongation of existing hospitalization,
- Results in disability/incapacity,
- Is a congenital anomaly/birth defect,
- Other: medical or scientific judgement should be exercised in deciding whether reporting is appropriate in other situations, such as important medical events that may not be immediately life-threatening or result in death or hospitalization but may jeopardize the subject or may require medical or surgical intervention to prevent one of the other outcomes listed in the above definition. These should also be considered serious.
- Possible drug-induced liver injury

See study Pharmacovigilance Plan (sPVP) for details. sPVP will include the following elements to ensure a comprehensive approach to safety event collection and reporting:

- Supplier pharmacovigilance training
- Investigator and site staff pharmacovigilance training
- Safety-specific roles
- SAEs collection and reporting processes
- SAEs collection forms
- Frequency of data review
- Reporting process and timelines
- Interim reports
- Study-specific PVP monitoring process

12.5.3. Patient-reported Outcome

CAT
• The CAT is an 8-item questionnaire suitable for completion by all patients diagnosed with COPD. When completing the questionnaire, participants rate their experience on a 6-point scale, ranging from 0 (no impairment) to 5 (maximum impairment) with a scoring range of 0-40. Higher scores indicate greater disease impact. • For all participants, CAT will be completed at baseline, each follow up clinic visit, and tri-monthly phone interviews. Interviewer-administered approach (face-to-face or over the phone) will be used. • Additionally, for monitoring purpose, monthly CAT will be interviewer-administered over phone call in the sub-cohort for 6 months after baseline visit and 6 months after Visit 3. CAT will also be assessed in the sub-cohort at exacerbation visits.

CAPTURE
• The CATPURE is a short, five-item questionnaire that can be easily completed by patients and are used to identify individuals who may have undiagnosed, clinically significant COPD. • The algorithm is a simple summation of patient responses to each of the five items. (For Question1-4, Score 0 for No, Score 1 for Yes, Question 5, Score 0,1,2. So the total score is ranging from 0 to 6). Higher score means higher risk of COPD (0-1 is at low risk, 2-6 is at high risk).

St. George's Respiratory Questionnaire for COPD Patients (SGRQ-C)
• The SGRQ-C contains 14 questions with a total of 40 items grouped into three domains (Symptoms, Activity and Impacts). • The details for how to score the SGRQ-C are outlined in the SGRQ-C manual (Jones, 2016). This includes details on how to handle missing data. • SGRQ-C domain and total score will be converted to SGRQ scores as described in the manual. • The converted SGRQ scores will be used for all summaries/analyses.

mMRC Dyspnea Scale
• The modified Medical Research Council (mMRC) Dyspnea Scale quantifies disability attributable to breathlessness and is useful for characterizing baseline dyspnea in patients with respiratory diseases. • mMRC is a simple one-item questionnaire that is completed by patients to grade their degree of baseline functional disability due to dyspnea. The 5 options in the question item translate to a 0-4 score range, with higher score indicating higher severity.

EXACT and E-RS:COPD

- The EXACT is a 14-item daily diary designed to provide a measure of patient-reported symptoms of COPD exacerbation. An EXACT Total Score, ranging from 0 to 100, where higher scores indicate a more severe condition, will be derived for each day of diary collection according to the instructions in the EXACT User Manual (Version 4.0, Evidera, 2011).
- Definition of EXACT events: magnitude of responder for symptomatic events defined as acute, sustained symptomatic worsening of COPD, defined as an increase in EXACT score ≥9 points for 3 days or ≥12 points for 2 days, above baseline (average in 7-day period)
- The EXACT daily diary will be completed each evening electronically for a month following Visit 1 and Visit 3.
- The E-RS™ involves the use of 11 of the 14 EXACT-PRO™ items and a different scoring algorithm to assess the severity of respiratory symptoms in patients with COPD. The E-RS™ score ranges from 0 to 40 with higher scores indicating more severe symptoms.
- Three subscales of the E-RS: COPD are used to describe different symptoms: breathlessness, cough and sputum, and chest symptoms.
- EXACT and E-RS score's scoring algorithms are from the EXACT-PRO user manual (Evidera, 2011).

Scale	Item-level Construct	Item Number	Score Range
RS-Breathlessness			
	Breathless today	7	0-4
	Breathless with activity	8	0-3
	Short of Breath – personal care	9	0-4
	Short of Breath – indoor activity	10	0-3
	Short of Breath – outdoor activity	11	0-3
RS-Cough and Sputum			
	Cough frequency	2	0-4
	Mucus quantity	3	0-3
	Difficulty with mucus	4	0-4
RS-Chest Symptoms			
	Congestion	1	0-4
	Discomfort	5	0-4
	Tightness	6	0-4
Respiratory Symptoms	E-RS Total	11 items	0-40
Additional Attributes			
Used in the EXACT total score for characterizing exacerbations, specifically	Tired or weak	12	0-4
	Sleep disturbance	13	0-4
	Scared or worried	14	0-3
EXACT – Exacerbation	EXACT Raw Total Score[1]	14 items	0-55

[1]EXACT Total Score will be transformed to a 0-100 scale based on the raw total score

Exact Event
Definition of EXACT events: magnitude of responder for symptomatic events defined as acute, sustained symptomatic worsening of COPD, defined as an increase in EXACT score ≥9 points for 3 days or ≥12 points for 2 days, above baseline (average in 7-day period).

Clinically important deterioration (CID)
Clinically important deterioration (CID) is a new composite outcome (developed by GSK) for the prediction of future risk of long-term adverse outcomes. The definition of composite CID used in this study is CAT-CID: • Occurrence of a moderate/severe exacerbation and/or • ≥100 mL deterioration from baseline in postbronchodilator FEV1 and/or • ≥2 units deterioration in the CAT.

COPD Exacerbation
• COPD subjects will be asked for a history of AECOPD using components of moderate/severe exacerbation definition at each clinic visit including exacerbation visits. • Additionally, COPD subjects will be contacted by phone every 3 months by the study staff and asked about their exacerbation details for the previous 3 months. • Study staff will check Hospital Information System (HIS) during the interview to confirm the event of exacerbation wherever possible.

Digital Physical Activity

- Daily physical activity will be collected in sub-cohort only. Participants are required to keep wearing the wrist band for 1 month from visit 1 and 3 as outlined in the time and event table (Appendix 2: Schedule of Activities).
- Daily activity as measured by number of daily steps will be captured using digital mobility devices (wrist bands) worn by participants, dispensed at baseline.
- Daily steps walked, sedentary minutes, active minutes, very active minutes, sedentary minutes and total sleeping minutes will be derived and summarized.
- Definition of intensity data
    - SEDENTARY - Little to no activity monitored. This could be due to minimal movement, sitting, resting, or sleeping.
    - ACTIVE - Some activity monitored. A brisk walk could achieve this intensity.
    - HIGHLY_ACTIVE - High activity monitored. Running or speed walking could achieve this intensity.
- Below is a template of how activity information is summarized for site personnel in Rave EDC

CRF Component	Value Format	Calculation	Example Values
Date	Text	Date of log line record	26 MAR 2015

Digital Physical Activity			
Sync Date/Time	Text	Date and time data was synced to Garmin	26 MAR 2015 14:00
Daily Steps Walked	Numeric	Total # of steps in routine data objects for a subjects day (timezone is subject's)	2240
Sedentary Minutes	Numeric	Total # of minutes in routine data objects for a subject's day (timezone is subject's) with intensity parameter = sedentary	438
% of time sedentary	Numeric	Total # of minutes in routine data objects for a subject's day (timezone is subject's) with intensity parameter = sedentary divided by 1440 minutes (total # of minutes per day)	85.32
Active Minutes	Numeric	Total # of minutes in routine data objects for a subject's day (timezone is subject's) with intensity parameter = active	74
% of time active	Numeric	Total # of minutes in routine data objects for a subject's day (timezone is subject's) with intensity parameter = active divided by 1440 minutes (total # of minutes per day)	14.47
Very Active Minutes	Numeric	Total # of minutes in routine data objects for a subject's day (timezone is subject's) with intensity parameter = very active	1
% of time very active	Numeric	Total # of minutes in routine data objects for a subject's day (timezone is subject's) with intensity parameter = very active divided by 1440 minutes (total # of minutes per day)	0.20
Total Daily Minutes Sleep	Numeric	Total # of minutes in sleep data objects for a subject's day (timezone is subject's)	362

Phenotype
Exacerbator

• Exacerbator is categorized based on the number of exacerbations patients have in the past:
  ○ Never
  ○ Occasional (1 in the last 2 years)
  ○ Yearly (1 each year in the last 2 years)
• Frequent (≥2 in the previous year)

Chronic mucus hypersecretion (CMH)
• Chronic mucus hypersecretion (CMH) is defined as: At least 3 months of cough & phlegm in past year. This will be identified by asking a question as "How many months in the past 12 months have you had bronchitis or chronic coughing with phlegm or sputum from the chest?" • SGRQ-C will be used for identification of CMH in this study. This would be identified by cough and phlegm related questions in SGRQ-C or CAT. • SGRQ-CMH is defined by answering cough for most days or several days a week (option 1 or 2 for the "I cough" question) and phlegm for most says or several days a week (option 1 or 2 for the "I bring up phlegm [sputum]" question) (Kim et al. 2021) • Significant Symptoms is defined as a CAT score of 10 or higher.
Significant Symptoms
• Significant Symptoms is defined as soring 10 or higher in the CAT questionnaire.
Asthma-COPD Overlap (ACO)
• Asthma-COPD overlap (ACO) has been used to identify patients with airway disease who have features of both asthma and COPD. • ACO is defined as COPD patients that have physician diagnosed asthma ≤40 years of age

Endotype
Airflow Limitation
• FEV1 (GOLD grade)
PRISm
• Normal FEV1/FVC ratio, but FEV1 <80% pred
'Pure' Airways Disease
• Airway wall thickening on CT • No emphysema on CT
Presence of emphysema
• CT scan – standard definition now available
Higher Blood Eosinophils
• Eosinophils > 300 µl
Fibrinogen
• Plasma fibrinogen > 400 mg/dL
Mibrobiome

12.6. Appendix 6: Reporting Standards for Missing Data

12.6.1. Premature Withdrawals

Element	Reporting Detail
General	• Subject study completion (i.e. as specified in the protocol) was defined as completed all scheduled visit and completed exit visit. • Withdrawn subjects were not replaced in the study. • All available data from participants who were withdrawn from the study will be listed and all available planned data will be included in summary tables and figures, unless otherwise specified. • Withdrawal visits will be slotted as per Appendix 3: Assessment Windows or will be summarised as withdrawal visits.

12.6.2. Handling of Missing Data

Element	Reporting Detail

General	• Missing data occurs when any requested data is not provided, leading to blank fields on the collection instrument:
	  ○ These data will be indicated by the use of a "blank" in subject listing displays. Unless all data for a specific visit are missing in which case the data is excluded from the table.
	  ○ Answers such as "Not applicable" and "Not evaluable" are not considered to be missing data and should be displayed as such.

MAR	• The data are said to be missing at random (MAR) when drop-out probability is conditionally independent of missing data given the observed outcomes and covariates • There is an underlying assumption that the data are missing at random. Missing data are not explicitly imputed in the Mixed Model Repeated Measures model. • If data are MAR, the source of the missing data may be ignored and inference is valid from a mixed-effects model. In other words, it is not necessary to include in the data analysis information about the source of the missing data. • For sensitivity analyses, Multiple imputation will be utilized under the assumption of MAR for MMRM model for participants who withdrawal from the study as well as lost to follow up the study, post- withdrawal counts were imputed conditional upon the subject's own observed measurements prior to withdrawal.

12.7. Appendix 7: Model Checking and Diagnostics for Statistical Analyses

12.7.1. Statistical Analysis Assumptions

Endpoint(s)	• Change from baseline in FEV_1 • Change from baseline in FVC

	• Change from baseline in CAT score • Change in Evaluating Respiratory Symptoms in COPD (E-RS: COPD) scores • Change from baseline in airway thickness by lung HRCT • Change in lung density by lung HRCT
Analysis	Mixed Model Repeated Measures

- The Kenward-Roger method for approximating the denominator degrees of freedom and correcting for bias in the estimated variance-covariance of the fixed effects will be used. An unstructured covariance structure for the R matrix will be used by specifying 'type=UN' on the REPEATED line.
    - In the event that this model fails to converge, alternative correlation structures may be considered.
- Distributional assumptions underlying the model used for the analysis will be examined by obtaining a normal probability plot of the residuals and a plot of the residuals versus fitted values (i.e. checking the normality assumption and constant variance assumption of the model respectively) to gain confidence that the model assumptions are reasonable. These plots will be generated using the option plots=RESIDUALPANEL.

Endpoint(s)	• Rate of decline in FEV_1 • Rate of decline in FVC
Analysis	Random Coefficient Models

Distributional assumptions underlying the model used for the analysis will be examined by obtaining a normal probability plot of the residuals and a plot of the residuals versus fitted values (i.e. checking the normality assumption and constant variance assumption of the model respectively) to gain confidence that the model assumptions are reasonable. These plots will be generated using the option plots=RESIDUALPANEL.

Endpoint(s)	• Rate of moderate/severe exacerbation • Clinically important deterioration (CID) composite outcome and its components • Characterization of EXACT events
Analysis	Negative binomial distribution

The fit of the regression models will be examined using "Q-Q" plots of the standardized residuals. Interpretation of these plots will be aided by the addition of simulated envelopes as proposed by Atkinson (1).

Endpoint(s)	Time to the first CID composite event
Analysis	Cox proportional hazard method
The proportional hazards (PH) assumption for this method of analysis will be checked by comparing log-log survival curves, i.e. ln (-ln$\hat{S}$). The log-log survival curves will be estimated based on Kaplan-Meier (KM) method. The predictors include type of subjects (i.e. GOLD I-IV, chronic bronchitis, healthy control), gender, smoking status, site (type of hospital), age, baseline FEV1, count of COPD exacerbations. The variables will be checked one at a time. The curves will be approximately parallel if the PH assumption is met. If the curves intersect or are not parallel in some other way, the PH assumption is considered not satisfied for this variable.	

12.8. Appendix 8: Abbreviations & Trademarks

12.8.1. Abbreviations

Abbreviation	Description
CAPTURE	COPD Assessment in Primary Care to Identify Undiagnosed Respiratory Disease and Exacerbation Risk
CAT	COPD Assessment Test
CDISC	Clinical Data Interchange Standards Consortium
CID	Clinically Important Deterioration
COPD	Chronic Obstructive Pulmonary Disease
CSR	Clinical Study Report
CTR	Clinical Trial Register
DDA	Digital Data Analytics
DOB	Date of Birth
EXACT	EXAcerbation of COPD Tool
FVC	forced vital capacity
GOLD	Global Initiative for Chronic Obstructive Lung Disease
GSK	GlaxoSmithKline
HRCT	High-resolution computed tomography
ICH	International Conference on Harmonization
IDSL	Integrated Data Standards Library
MMRM	Mixed Model Repeated Measures
PDMP	Protocol Deviation Management Plan
PRO	Patient-reported Outcomes
RAP	reporting and analysis plan
SAC	Statistical Analysis Complete
SGRQ-C	St. George's Respiratory Questionnaire for COPD
SOP	Standard Operation Procedure
TFL	Tables, Figures & Listings
TLC	Total Lung Capacity

12.8.2. Trademarks

Trademarks of the GlaxoSmithKline Group of Companies
COPD Assessment Test (CAT)

Trademarks not owned by the GlaxoSmithKline Group of Companies
EXACT-PRO
E-RS:COPD
SAS
NONMEM
SGRQ