GlaxoSmithKline	**Statistical Analysis Plan**
Detailed Title:	A prospective, epidemiological, multi-country, cohort study to assess the occurrence of potential bacterial and viral pathogens in stable chronic obstructive pulmonary disease (COPD) and during acute exacerbations of COPD (AECOPD), in moderate to very severe COPD patients, in Asia Pacific.
eTrack study number and Abbreviated Title	201112 (EPI-NTHI-001 BOD APA)
Scope:	All data pertaining to the above study.
Date of Statistical Analysis Plan	Final: 28 July 2017 Amendment 1 Final:14 September 2018 Amendment 2 Final: 05 August 2020
Co-ordinating author:	PPD, Biostatistician in Epidemiology PPD, Biostatistician in Epidemiology
Reviewed by:	PPD, Lead Epidemiologist PPD, Lead Statistician in Epidemiology PPD, Lead Statistician PPD, Lead Statistical Analyst PPD, Statistical Analyst PPD, Statistical Analyst PPD, Statistical Analyst PPD, Statistical Analyst PPD, Scientific writer PPD, Public Disclosure Representative
Approved by:	PPD, Clinical and Epidemiology Project Lead delegating to PPD, Lead Epidemiologist PPD, Lead Statistician PPD, delegating to PPD, Scientific Writer PPD, Lead Statistical Analyst

APP 9000058193 Statistical Analysis Plan Template (Effective date: 14 April 2017)

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

LIST OF ABBREVIATIONS

A. baumannii	*Acinetobacter baumannii*
AE	Adverse event
AECOPD	Acute exacerbation of COPD
CAT	COPD assessment test
CI	Confidence Interval
COPD	Chronic Obstructive Pulmonary Disease
CRF	Case Report Form
EXACT-PRO	Exacerbations of Chronic Pulmonary Disease Tool - Patient Reported Outcome
FAS	Full Analysis Set
FEV_1	Forced expiratory volume in 1 second
GOLD	Global Initiative for Chronic Obstructive Lung Disease
GSK	GlaxoSmithKline
H. haemolyticus	*Haemophilus haemolyticus*
H. influenzae	*Haemophilus influenzae*
HRQOL	Health-related quality of life
HRV	Human rhinoviruses
K. pneumoniae	*Klebsiella pneumoniae*
LL	Lower Limit of the confidence interval
M. catarrhalis	*Moraxella catarrhalis*
MA	Main Analyses
MedDRA	Medical Dictionary for Regulatory Activities
NTHi	Non-Typeable *Haemophilus influenzae*
P. aeruginosa	*Pseudomonas aeruginosa*
PCR	Polymerase chain reaction

PEF	Peak Expiratory Flow
RSV	Respiratoir Syncytiaal Virus
RT-PCR	Real-time polymerase chain reaction
SAE	Serious adverse event
S. aureus	*Staphylococcus aureus*
SAP	Statistical Analysis Plan
SD	Standard Deviation
SGRQ-C	St. George's Respiratory Questionnaire for COPD patients
S. pneumoniae	*Streptococcus pneumoniae*
SR	Study Report
TFL	Tables Figures and Listing template annexed to SAP
UL	Upper Limit of the confidence interval

1. DOCUMENT HISTORY

Date	Description	Protocol Version
28-JUL-2017	Version 1	Protocol Amendment 1 13-FEB-2017
14-SEPT-2018	SAP Amendment 1 Updates done before Interim Analysis: - analysis population was adapted to Protocol Amendment 2 - elimination codes were aligned to PDMP v.4 - analyses added to Interim Analysis timepoint - sputum quality - alignment with the protocol/CRF - more details about laboratory methods were added - several minor corrections	Protocol Amendment 2 19-OCT-2017
05-AUG-2020	SAP Amendment 2 Updated made before final analysis: - elimination codes were aligned to PDMP v.9 - Statistical modelling details added or updated based on preliminary analyses results - Addition, removal or update of a few new analyses related to secondary (incidence rate) and tertiary objectives as per team's request - some general minor edits were implemented	Protocol Amendment 2 19-OCT-2017

2. STUDY DESIGN

Type of design: epidemiological, prospective, interventional, multi-country, cohort study. The study targets enrolling approximately 200 stable moderate to very severe COPD patients, males and females, aged 40 years or older at the time of enrolment, with at least 1 documented moderate or severe AECOPD in the year before enrolment.

This study will be conducted in several sites among several countries in Asia Pacific [Hong Kong, Philippines, Korea, Taiwan].

Type of study: self-contained.

Duration of the study: approximately 1 year for each patient.

Study visits/contacts:

- Screening visit
- Three scheduled visits occurring at 6 months intervals.
- For each AECOPD: AECOPD visit (within 96 hours of the onset of the symptoms) and follow-up phone call(s) (at least every 2 weeks until the AECOPD has resolved). Follow-up phone contacts will define end of AECOPD.

The following group names will be used for the statistical analyses:

Group order in tables	Group label in tables	Group definition for footnote
1	COPD	Patients with moderate to very severe COPD

3. OBJECTIVES

3.1. Primary objective

To estimate the proportion of potential bacterial and viral pathogens (overall and by species) detected in sputum of stable COPD patients and during AECOPD, using respectively bacteriological methods and viral PCR, over the course of 1 year.

3.2. Secondary objectives

- To evaluate the concordance between PCR and bacteriological methods data in sputum for potential bacterial pathogens.
- To estimate the proportion of potential bacterial and viral pathogens (overall and by species) detected in sputum of stable COPD patients, by GOLD grade.
- To estimate the proportion of potential bacterial and viral pathogens (overall and by species) detected in sputum during AECOPD, by severity of AECOPD.
- To estimate the incidence rate of all-cause AECOPD, overall and by GOLD grade.
- To describe the severity and duration of all-cause AECOPD, overall and by GOLD grade.
- To assess the impact of all-cause AECOPD on HRQOL.
- To assess the impact of all-cause AECOPD on lung function.
- To assess the impact of all-cause AECOPD on healthcare utilisation.

3.3. Tertiary objectives

- To assess the use of EXACT-PRO for determining the end date of AECOPD.
- To evaluate the load of bacterial and viral pathogens in stable COPD and during AECOPD.
- To collect biological specimens for future respiratory disease-related testing:
  - Aliquots of sputum samples for assay development and microbiome analysis at stable visits and during exacerbation.
  - Blood sampling for potential biomarkers for identification/quantification of biomarkers at Visit 1 and Visit 3.

4. ENDPOINTS

4.1. Primary endpoint

- Occurrence of potential bacterial and viral pathogens in sputum of stable COPD patients and during AECOPD, over the course of 1 year:
  - Bacterial pathogens, as identified by bacteriological methods (culture), including (but not necessarily limited to) *H. influenzae*, *M. catarrhalis*, *S. pneumoniae*, *S. aureus*, *P. aeruginosa*, *K. pneumoniae* and *A. baumannii*.
  - Viral pathogens, as identified by PCR, including (but not necessarily limited to) RSV, parainfluenza virus, enterovirus/ HRV, metapneumovirus, influenza virus, adenovirus, bocavirus and coronavirus.

4.2. Secondary endpoints

- Occurrence of potential bacterial pathogens in sputum of stable COPD patients and during AECOPD, as measured by real-time qualitative PCR/ quantitative PCR and compared to data from bacteriological methods, over the course of 1 year:
  - Including (but not necessarily limited to) *H. influenzae, M. catarrhalis, S. pneumoniae, S. aureus* and *P. aeruginosa.*
- Occurrence of potential bacterial and viral pathogens (overall and by species) in sputum of stable COPD patients by GOLD grade, over the course of 1 year.
- Occurrence of potential bacterial and viral pathogens (overall and by species) in sputum during AECOPD by severity of AECOPD, over the course of 1 year.
- Incidence rate (per patient per year) of AECOPD, overall and by GOLD grade, over the course of 1 year.
- Severity of AECOPD, overall and by GOLD grade, over the course of 1 year.
- Duration of AECOPD, overall and by AECOPD severity, over the course of 1 year.
- CAT score in stable COPD patients and during AECOPD, over the course of 1 year.
- SGRQ-C score in stable COPD patients, over the course of 1 year.
- FEV_1% of predicted normal value in stable COPD patients, at Pre-Month 0 and Month 12.
- Healthcare utilisation, over the course of 1 year.

4.3. Tertiary endpoints

- EXACT-PRO scores in stable COPD patients and during AECOPD, over the course of 1 year.
- Bacterial load measured by both culture and PCR at stable and exacerbation visits.
- HRV load measured by PCR at stable and exacerbation visits.

5. ANALYSIS SETS

The following data sets will be defined.

5.1. All screened set

The all screened set will include all patients that were evaluated for enrolment and that signed the informed consent. Patients with invalid informed consent or fraud data will be exluded from the all screened set.

5.2. All enrolled set

The all enrolled set will include patients from the all screened set minus those that were screening failures.

5.3. Full Analysis Set

The Full Analysis Set (FAS) will include patients from the all enrolled set except for those who discontinued the study at Visit 1.

Study objectives will be assessed on the FAS. The population set for each analysis will change according to the subjects evaluable for the specific endpoint.

5.3.1. Change in analysis population from protocol amendment 2

The analysis population is aligned with protocol amendment 2.

5.4. Criteria for eliminating data from the Analysis Sets

Details about elimination codes used to identify patients to be eliminated from analysis are provided below for each set.

5.4.1. Elimination from All Screened Set

Code 800 (fraudulent data) and code 900 (invalid informed consent) will be used for identifying patients eliminated from All Screened Set.

5.4.2. Elimination from All Enrolled Set

Code 800 (fraudulent data), code 900 (invalid informed consent) and code 2010 (Protocol violation - inclusion/exclusion criteria) will be used for identifying patients eliminated from All Enrolled Set.

5.4.3. Elimination from the Full Analysis Sets (FAS)

Code 800 (fraudulent data), code 900 (invalid informed consent) and code 2010 (Protocol violation - inclusion/exclusion criteria) will be used for identifying patients eliminated from FAS.

Of note, the only difference between the All Enrolled Set and the FAS comes from patients that withdraw from the study at visit 1. These patients will be in the All Enrolled Set but will not be in the FAS. However, this is not a protocol deviation and is not mentioned in this section.

5.4.4. Elimination codes details

A list of reasons for elimination from the analyses (interim analysis and final analysis), is reported below:

Code	Condition under which the code is used
800	Fraudolent data
900	1. Informed consent deviations 1.a. Signed informed consent not available on site 1.b. Wrong informed consent version signed 1.c. Informed consent not signed and/or dated by subject 1.d. Informed consent not signed and/or dated by appropriated site staff 1.e. Informed consent not signed prior to any study procedure 1.f. Other informed consent deviations 2. Eligibility deviations 2.a. Biological sample specimen procedures 2.a.a. PII recorded in label of sample
2010	1. Eligibility criteria not met
2040	1. Excluded medication, vaccine, or device 1.a. Prohibited medication as per protocol (e.g., antibiotics administration will impact the primary endpoint)

Code	Condition under which the code is used
2050	1. Not withdrawn after developing withdrawal criteria 1.a. Not withdrawn after developing withdrawal criteria other than prohibited vaccination/medication
2090	1. Assessment or time point completion 1.a. Out of window visit (assessment) for efficacy 1.a.a. Sputum sample collection outside the 96 hour-interval since exacerbation onset
2100	1. Study procedures 1.a. Biological sample specimen procedures 1.a.a. Low volume (not sufficient to perform testing) 1.a.b. Mislabeling (sample not tested)
2110	1. Study procedures 1.a. Biological sample specimen procedures 1.a.a. Incorrect or expired sample tube used 1.a.b. Incorrect spinning/processing of sample
2120	1. Assessment or time point completion 1.a. Other assessment not properly performed 1.a.a. Sputum culture not completed within six hours, 1.a.b. Culture: error in testing algorithm for identification (vs AP, eg.no optochin for Strep.) 1.b. Assessmet not properly performed 1.b.a. Sputum processing priorities not respected. 1.b.b. Frozen sputum DTT before culture

Code	Condition under which the code is used
	1.b.c. Sputum DTT aliquots not frozen <6hrs 1.b.d. Sputum not diluted after glass smear preparation 1.b.e. Sputum diluted with concentrated DTT 1.b.f. Gram stain not read on the same day as preparation (not an issue if still done < 10 days) 1.b.g. Sputum glass smear for quality not prepared or assessment not performed 2. Study procedures 2.a. Biological sample specimen procedures 2.a.a. Central/internal/external lab deviations, 2.a.b. Mislabelling (sample tested), 2.a.c. Wrong packaging that impacts quality of the sample (i.e. room temperature packaging for shipping frozen ELISA samples). 2.a.d. Temperature log not maintained as required or not tracked 2.a.e. "Blood" or Sputum temperature deviations from range defined in protocol and/or SPM/lab manual - room temperature and freezer -70°C 2.a.e.a. Sputum storage (only if not tested on same day), then storage cooled (+2-8°C): no impact if < 24hrs. If > 24hrs but still <72hrs, potential impact, to evaluate case by case. If >72hrs, elimination 2.a.e.b. Sputum DTT (-70°C): range +/-20°C. Lab is looking for additional stability data. Normally, ok at -20°C for 1 month (for PCR) 2.a.f. Sputum collection procedure not following process outlined in the protocol protocol (e.g. sputum collected at home for stable visit)

Code	Condition under which the code is used
2120a	1. Study procedures 1.a. Biological sample specimen procedures 1.a.a. "Blood" or Sputum temperature deviations from range defined in protocol and/or SPM/lab manual - room temperature and freezer -70°C 1.a.a.a. Sputum DTT - Only culture samples affected
2120b	1. Study procedures 1.a. Biological sample specimen procedures 1.a.a. "Blood" or Sputum temperature deviations from range defined in protocol and/or SPM/lab manual - room temperature and freezer -70°C 1.a.a.a. Sputum DTT - Only PCR samples affected
2120c	1. Study procedures 1.a. Biological sample specimen procedures 1.a.a. "Blood" or Sputum temperature deviations from range defined in protocol and/or SPM/lab manual - room temperature and freezer -70°C 1.a.a.a. Sputum DTT - Only Microbiome samples affected
2120d	1. Study procedures 1.a. Biological sample specimen procedures 1.a.a. "Blood" or Sputum temperature deviations from range defined in protocol and/or SPM/lab manual - room temperature and freezer -70°C 1.a.a.a. Sputum DTT - Only Fibrinogen samples affected

Code	Condition under which the code is used
2120e	1. Study procedures 1.a. Biological sample specimen procedures 1.a.a. "Blood" or Sputum temperature deviations from range defined in protocol and/or SPM/lab manual - room temperature and freezer -70°C 1.a.a.a. Sputum DTT - Only hsCRP samples affected
2120f	1. Study procedures 1.a. Biological sample specimen procedures 1.a.a. "Blood" or Sputum temperature deviations from range defined in protocol and/or SPM/lab manual - room temperature and freezer -70°C 1.a.a.a. Sputum DTT - Only CXCL10 samples affected
2120g	1. Study procedures 1.a. Biological sample specimen procedures 1.a.a. "Blood" or Sputum temperature deviations from range defined in protocol and/or SPM/lab manual - room temperature and freezer -70°C 1.a.a.a. Blood - Only "other biomarker" samples affected
2120h	1. Study procedures 1.a. Biological sample specimen procedures 1.a.a. "Blood" or Sputum temperature deviations from range defined in protocol and/or SPM/lab manual - room temperature and freezer -70°C 1.a.a.a. Blood - Only samples for hematology profile affected

Code	Condition under which the code is used
2130	1. Study procedures 1.a. Biological sample specimen procedures 1.a.a.Testing performed on samples not aligned with ICF, 1.a.b. Subjects bled but not supposed to be bled
2500	1. Assessment or time point completion 1.a. Incorrect assessment performed 1.a.a. Other
2600	1. Other 1.a. Other 1.a.a. Any other GCP non-compliance

Code 2040 and codes from 2090 to 2600 are protocol deviations not leading to elimination from the population sets above mentioned but that may lead, judged case by case, to sample elimination from a specific statistical analysis.

6. STATISTICAL ANALYSES

The total number of patients in the following data sets overall and among the study country will be tabulated:

- All screened set and screening failures.
- All enrolled set and patients that withdrew from the study at visit 1.
- Full analysis set.
- Patients that withdraw from the study at each stable visit and patients that completed the study.

In addition, screening failures will be summarized according to the reason for failure.

Withdrawal status will be summarised according to the reason for withdrawal. The number of withdrawn patients will be tabulated by study visit and overall.

6.1. Analysis of demographics/baseline characteristics

The analysis of demographic/baseline characteristics will be performed on the FAS and on the all enrolled set only if the percentage of enrolled patients excluded from the FAS is 5% or more.

The following characteristics at enrolment will be summarized using descriptive statistics:

- Demographics such as age at enrolment, sex and race.
- Present and past medical history.
- History of AECOPD within the previous year (overall and by AECOPD severity).
- Pre- and post-bronchodilator spirometry results.
- COPD status using the GOLD grade.
- Smoking/biomass exposure history status.
- Selected results and several risk factors from the ATS-DLD-78-A questionnaire.
- Physical examination/vital signs results such as height, weight, BMI, heart rate, respiratory rate, systolic and diastolic blood pressure.
- Chest X-rays results.
- Intercurrent comorbidities (reported at visit 1)
- History of pneumococcal and influenza vaccination (reported at visit 1)

The following descriptive statistics will be adopted:

- The frequency and the percentage of each category will be reported for categorical variable.
- Number of non-missing observations, mean, median, standard deviation, minimum and maximum will be provided for quantitative variable.

All tables will be also generated by country.

Demographics and baseline characteristics will be also summarized for study sample for interim analysis.

6.2. Analysis of primary objective

The analysis of the primary objective will be performed on the FAS.

Average proportion of patients with sputum samples positive for bacterial or viral pathogens overall and for specific species, and its exact 95% confidence interval (CIs)will be computed at any stable scheduled visit and at any exacerbation visit for the entire study period. In addition, the average proportion of sputum samples positive for bacterial and viral pathogens overall and for specific species, and its 95% confidence interval will also be computed at any stable scheduled visit and at any exacerbation visit for the entire study period. Confidence intervals will be estimated using a Generalized Estimating Equations (GEE) model assuming a binomial distribution for the response variable with logit as link function and a compound symmetry correlation matrix (exchangeable structure) to account for the within-patient correlations (Liang,. 1986) (more details in section 6.2.1). If the GEE model will not run due to only one level of the outcome variable, the exact 95% confidence interval will be computedProportion of patients with sputum samples positive for bacterial/viral pathogens overall and for specific species, will be computed, with exact 95% CIs, at each stable* scheduled visit (visit 1, visit 2, visit3) and at each exacerbation visit confirming acute exacerbation**,***.

** A confirmed stable visit will be defined as a scheduled study visit for which the investigator confirms in the eCRF that the patient is stable/has recovered from a previous exacerbation.*

*** If an AECOPD occurs at the time when one of the scheduled [stable] study visits is planned, it should be handled and recorded as an AECOPD visit. If possible, the stable study visit should be re-scheduled to a later date, within the ± 14 days' time window from the scheduled appointment, when the patient is stable again.*

**** The definition of acute exacerbation of COPD (AECOPD) event is detailed in section 12.*

In details the following pathogens will be evaluated:

- Bacteria including, but not necessarily limited to, *H. influenzae**, *M. catarrhalis*, *S. pneumoniae*, *S. aureus*, *P. aeruginosa*, *K. pneumoniae* and *A. baumannii* by bacteriological methods (culture).

  **H. influenzae* positive cultures will be further characterized to differentiate *H. haemolyticus,* typeable *H. influenzae* and non typeable *H. influenzae* by PCR assay.

- Viruses including, but not necessarily limited to, RSV, parainfluenza virus, enterovirus/ HRV, metapneumovirus, influenza virus, adenovirus, bocavirus and coronavirus using qualitative PCR.

Patients providing a sputum sample as per protocol are all evaluable patients (*i.e.* those meeting all eligibility criteria and complying with procedures described in the protocol*,**). According to protocol, the interval of ± 14 days for any scheduled visit will not be a criterion for exclusion from the analyses. The number and percentage of patients with a stable visit within and without this time window will be presented.

** Patients attending a study visit and providing sputum sample.*

*** Patients using antibiotics on a continual basis (defined as more than 1 month in total) will be allowed to continue study participation, but may be eliminated from the analyses.*

If more than 10% of sputum samples have bad quality (as defined in section 11.2.3), the analyses for primary objective will be repeated in the subset of sputum samples of good and moderate quality.

An interim analysis will be performed when at least 40 AECOPD sputum samples are collected. This interim analysis will compute the rate of sputum sample positive any bacterium and by species (including Hi, NTHi and Mcat).

6.2.1. Statistical methods: analysis of correlated Data Using Generalized Estimating Equations

Average proportion of sputum samples positive for bacteria/viruses and its 95% CI will be computed using a GEE model. GEE approach uses weighted combinations of observations to extract the appropriate amount of information from correlated data such as the response measurements across multiple stable visits on the same patient (Burton, 1998).

For the purpose of the analyses a binomial distribution for the response variable (presence or absence of bacterial/viral pathogens in the sputum) with logit as link function will be assumed and an exchangeable correlation matrix will be chosen to account for the within-patient correlations.

The exchangeable correlation matrix assumes that the correlation between any two responses of the i^{th} individual is the same.

Other commonly used within-patient correlation matrices are as follows:

- Independence, repeated observations are uncorrelated
- Unspecified (unstructured), correlations within any two responses are unknown and need to be estimated
- Autoregressive of first order [AR(1)], assuming the interval length is the same between any two observations. Measurements taken further apart are less correlated than those taken closer together.

The model with the exchangeable correlation matrix will be adopted in the main analysis. Models with the other matrices could be run as sensitivity analyses. QIC and the related QICu statistic will be used to compare GEE models with different correlation matrices.

SAS codes

```
ods output ParameterEstimates=out_parm ClassLevels=Class;
proc genmod data=<dataset> descending;
 class pid;
 model y =  / dist = bin link = logit lrci ;
 Repeated subject=pid/sorted corr=cs PRINTMLE;
 by bacteria visit;
run;

data result;
 set out_parm(where=(parameter='Intercept') drop=ChiSq ProbChiSq DF);
 format percent LL UL percent8.1;
 Percent=exp(estimate)/(1+exp(estimate));
 LL=exp(LowerLRCL)/(1+exp(LowerLRCL));
 UL=exp(UpperLRCL)/(1+exp(UpperLRCL));
 unit=_N_;
run;
```

where pid = personal identifier, y = positivity to pathogen (0, 1), bacteria = type of bacteria, visit = type of visit (stable or exacerbation).

The REPEATED statement invokes the GEE method, specifies the correlation structure, and controls the displayed output from the GEE model. The option SUBJECT = PID identifies *pid* as the clustering variable, and the CORR = CS option specifies an exchangeable working correlation structure. Other matrices will be specified as CORR =: IND [independence], UN [unstructured], AR(1) [autoregressive (1)].

6.3. Analysis of secondary objectives

The analyses of the secondary objectives will be performed on the full analysis set.

6.3.1. Concordance between molecular (PCR) and bacteriological methods

The frequency distribution of sputum samples positive or negative for specific bacterial species (including but not necessarily limited to H. influenzae, M. catarrhalis, S. pneumoniae, S. aureus and P. aeruginosa) obtained from PCR will be tabulated versus results obtained from culture at any visit (stable scheduled visits or AECOPD visits).

The concordance between culture and PCR, at any visit, will be evaluated using the Cohen's kappa (k) statistic and 95% CI. Concordance will be interpreted as good (>0.79), substantial (0.60–0.79), moderate (0.40–0.59), fair (0.21–0.39) or no concordance (<0.21) (McHugh, 2012).

Proportion of patients with sputum samples positive for bacterial pathogens (including but not necessarily limited to H. influenzae , M. catarrhalis, S. pneumoniae, S. aureus, P. aeruginosa, K. pneumoniae and A. baumannii) by PCR will be computed, with 95% CIs, at each and any stable scheduled visit and at each and any AECOPD visit confirming acute exacerbation.

6.3.2. Proportion of potential bacterial/viral pathogens by country, GOLD grade and by AECOPD severity

Average proportion of patients with sputum samples positive for bacterial or viral pathogens overall and for specific species, will be computed, with 95% CIs, at any stable scheduled visit and at any AECOPD visit confirming acute exacerbation [mild, moderate, severe] by GOLD clasification [GOLD 2, 3, 4].

Average proportion of patients with sputum samples positive for bacterial or viral pathogens overall and for specific species, will be computed, with 95% CIs, at any stable scheduled visit and at any AECOPD visit confirming acute exacerbation by country.

6.3.3. Sputum sample collection and quality

6.3.3.1. Sputum sample collection

The percentage of patients at each scheduled [stable] and exacerbation visit for whom a sputum sample is obtained will be computed overall and by the method the sample is obtained [i.e. spontaneous (at study visit or at patient's home), induced using 0.9% saline or induced using 3% saline]. The percentage of patients with induced sputum will be tabulated by GOLD grade at enrolment and by severity of the AECOPD. The proportion of sputum samples obtained at each confirmed stable or AECOPD visit [any, mild, moderate and severe] with or without induction and positive for specific bacterial\viral pathogens by bacteriological culture and bacterial\viral PCR will be computed.

The distribution of the number of sputum samples will be presented according to time (days) since the start of exacerbation. The distribution of the proportion of sputum sample positive for bacteria/virus given the number of days between the start of exacerbation and sample collection will be presented.

The impact of previous antibiotic administration on bacterial results presence will be evaluated. The proportion of sputum samples obtained at each confirmed stable or AECOPD visit [any, mild, moderate and severe] with previous administration of antibiotics or not and positive for specific bacterial pathogens by bacteriological culture and PCR, respectively will be computed.

The impact of previous vaccination with a pneumococcal vaccine or an influenza vaccine (regardless of time) on *S. pneumoniae* and influenza virus will be evaluated. The proportion of sputum samples obtained at each confirmed stable or AECOPD visit with previous vaccination or not and positive for *S. pneumoniae,* by culture and PCR or influenza virus, by PCR, will be computed.

6.3.3.2. Sputum sample quality

Sputum sample quality will be summarized (squamous cell count, neutrophils cells count and bacteria direct smear) at any stable visit and at any mild, moderate or severe AECOPD visit. Sputum sample characteristics will be also summarized by bacterial and /or viral presence in any stable visit and in any exacerbation visit.

Sputum quality, derived as detailed in section 11.2.3, will be tabulated given the type of visit.

6.3.4. Incidence rates of AECOPD

This study will provide an estimate of average number of exacerbations per person per year, in a COPD population at increased risk of exacerbation (i.e. moderate to very severe COPD patients with a documented history of at least 1 moderate or severe AECOPD in the year prior to enrolment).

To compute the incidence rate all the exacerbations will be taken into account. The observation period is from visit 1 to visit 3 or from visit1 to the last visit perfomed for patients lost to follow up.

6.3.4.1. Overall incidence

The following incidence rates (per patient per year) will be computed, with 95% CIs [using negative binomial regression detailed in section 6.3.4.4]:.

- Confirmed exacerbations (exacerbation events confirmed by the investigator plus missed exacerbation visits with medical records: 5 and 6 dash boxes for the former and 2 and 7 for the latter from Figure 1).
- Confirmed exacerbations plus potential exacerbations with alert confirmed by phone call/or at the study site (potential exacerbations correspond to 3 dash box in Figure 1).
- Exacerbation events with sputum sample positive for selected pathogens

The incidence rate of confirmed exacerbations and of confirmed exacerbations plus potential exacerbations confirmed by phone call/or at the study site, with 95% CIs, will be estimated according to COPD severity at enrolment based on GOLD classification [GOLD 2, 3, 4].

The incidence rates of confirmed exacerbations, with 95% CIs, will be also computed by exacerbation severity.

In addition, the incidence rates of confirmed exacerbations, with 95% CIs, will be estimated according to covariates such as sex, country, age at enrolment (below median age or equal to/above median age), smoking status at enrolment (current and former smokers) and number of exacerbations reported in the 12 months prior to enrolment (one, two and equal or above three).

6.3.4.2. Incidence bacteria/viruses associated AECOPD

The following incidence rates (per patient per year) will be computed, with 95% CIs:

- AECOPD events with sputum sample containing bacterial pathogens found by culture or by PCR (overall and by, but not limited to, the following bacterial species: *H. influenzae* (and *H.influenzae* after differentiation form *H. Haemolyticus* by PCR assay), NTHi (after differentiation from typeable *H. influenzae* by PCR assay), *M. catarrhalis*, *S. pneumoniae*, *S. aureus*, and *P. aeruginosa*).
- AECOPD events with sputum sample containing viral pathogens found by PCR (overall and by, but not limited to, the following viral species: RSV, parainfluenza virus, enterovirus/ HRV, metapneumovirus, influenza virus, adenovirus, bocavirus and coronavirus using qualitative PCR and HRV also by quantitative RT-PCR).

These incidence rates, with 95% CIs, will be estimated also according to covariates such as COPD severity, sex, country, age at enrolment, smoking status at enrolment and number of exacerbations reported in the 12 months prior to enrolment.

In addition, incidence rate of mild, moderate and severe exacerbation having sputum containing any bacterial/viral pathogen will be computed overall and by species.

Incidence rate of AECOPD overall and associated to bacteria (detected by culture and by PCR) will be computed also at interim analysis.

6.3.4.3. Analyses of exacerbations in relation to morning e-diaries

Among morning e-diary alerts the frequency distribution of the following groups will be tabulated (refer to Figure 1 dashed boxes):

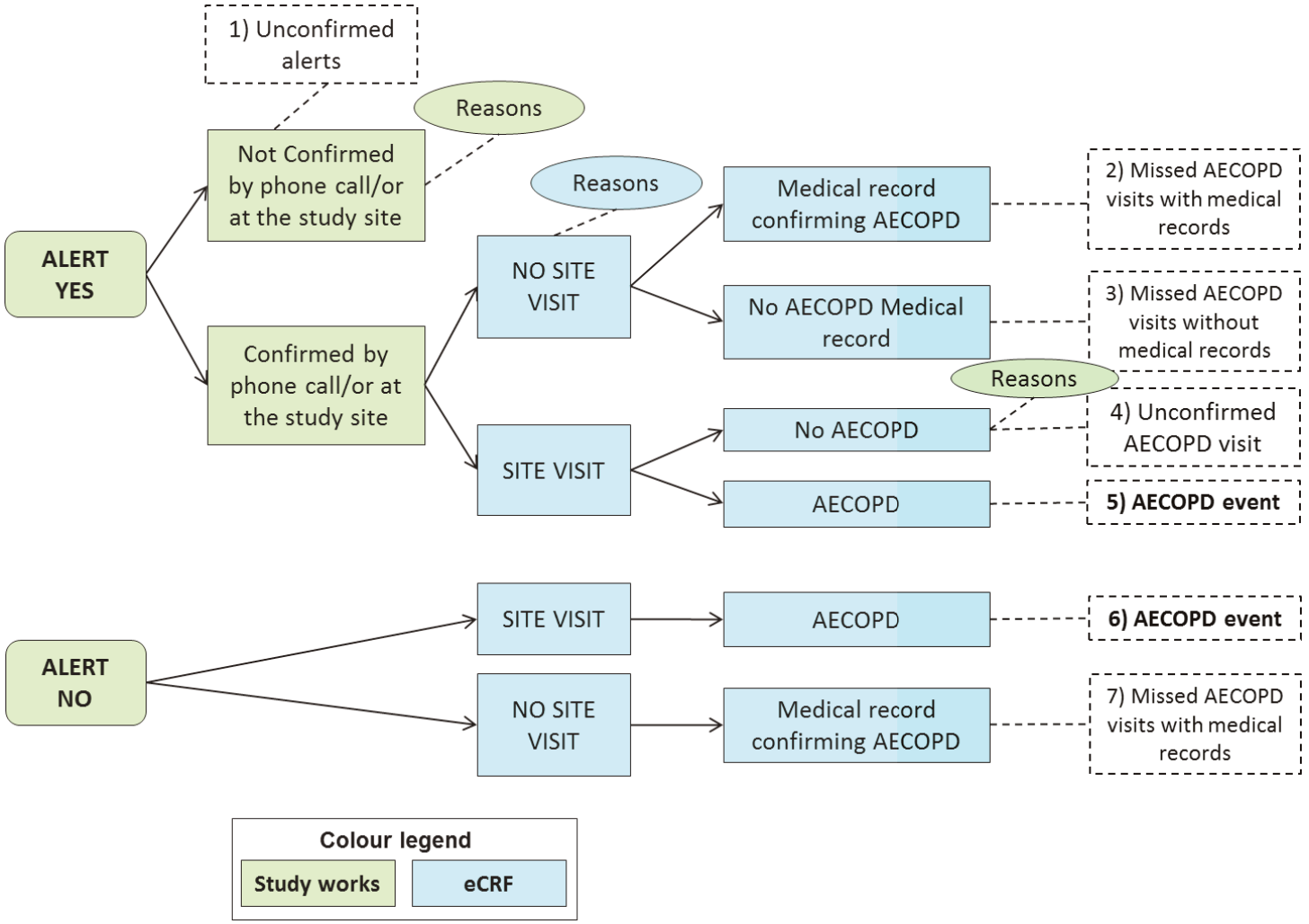
1. Potential event not confirmed by phone call/or at the study site (unconfirmed alerts, box 1),
2. Potential event confirmed by phone call/or at the study site, without site visit, with medical records confirming exacerbation (missed AECOPD visits with medical records, box 2),

3. Potential event confirmed by phone call/or at the study site, without site visit and without medical records (missed AECOPD visits without medical records, box 3).
4. Potential event confirmed by phone call/or at the study site, with site visit but not confirmed by study investigator (unconfirmed AECOPD visit, box 4),
5. Potential event confirmed by phone call/or at the study site, with site visit and confirmed by study investigator (AECOPD event, box 5),

Reasons for inconsistencies between morning e-diary alert and no confirmation by phone call, reasons for missing study visit after phone call confirmation, and reasons for inconsistencies between phone call confirmation and study visit will be tabulated.

In absence of morning diary alert the percentage of spontaneous visits confirming exacerbation or missed AECOPD visit with medical record confirming AECOPD will be tabulated (Figure 1, boxes 6 and 7).

Figure 1 Workflow of study assessment for AECOPD



6.3.4.4. Statistical methods: analysis of count data with overdispersion

The incidence rate and its 95% CI will be estimated using the Generalised Linear Model (GLM) assuming a negative binomial distribution for the response variable with logarithm as link function and the logarithm of follow up time as an offset variable. The fit of the model will be examined using Q-Q plots of standardised residuals. It is possible that the negative binomial model will not converge due to underdispersion of data; in such cases, to overcome the problem of underdispersion, a Poisson model will be used to obtain the incidence rate and its 95% CI.

SAS codes:

```
proc genmod data = <dataset>;
  class gold age smoking exacHist;
  model n_exac = GOLD age smoking exacHist country
      / link=log dist=negbin offset=logfuinyears ;
  ods output parameterestimates=out_parm nobs=nobs;
  ods exclude modelinfo modelfit parameterestimates nobs;
  run;

  data parm_est(keep= val rate rate_ll rate_ul);
    set out_parm(where=(parameter='intercept'));
    format rate rate_ll rate_ul 8.2;
    val= "mean number per year from negative binomial model without
covariate";
    rate=exp(estimate);
    rate_ll=exp(lowerwaldcl);
    rate_ul=exp(upperwaldcl);
  run;
```

where GOLD = COPD grade (2, 3 or 4), age = age at enrolment (40-49, 50-59, 60-69, ≥70 years), smoking = smoking status at enrolment (current and former smokers), exacHist = number of exacerbations reported in the 12 months prior to enrolment (one, two and equal or above three), country = countries in the study (Hong Kong, Philippines, Korea, Taiwan), n_exac = number of exacerbations during the study and logfuinyears = logarithm of the follow-up period in years.

6.3.5. Severity and duration of AECOPD

Total number of exacerbations recorded during the study, exacerbation rate (number of exacerbation per patient-year) and the proportion of COPD patients experiencing a certain number of exacerbations will be presented overall and per bacterial/viral species. The exacerbation rate per year will be reported with and without correction for early withdrawal (see Section 11.3).

Total number of exacerbations recorded during the study, the mean exacerbation rate per year and the proportion of COPD patients experiencing a certain number of exacerbations will be presented overall and divided by AECOPD severity.

The number and proportion of COPD patients experiencing a certain number of exacerbations will be presented overall and by GOLD grade at enrolment.

The number of patients that report at least 1 AECOPD will be tabulated and descriptive statistics (N, mean, median, SD, minimum, maximum, first and third quartiles) on the average length of AECOPD episodes, as estimated by the investigator, will be presented, for any, mild, moderate and severe AECOPD.

Number of days between 2 consecutive exacerbations, as estimated by the investigator, will be summarized using the same descriptive statistics.

Descriptive statistics of the number of patients according to their percentage of days within an exacerbation period out of the total observation period as estimated by the investigator will be presented. For this analysis, the exacerbations for which the end date was unknown were not excluded and the length was estimated by the investigator as the average length per severity grade.

Results from the chest X-rays during exacerbation visit will be tabulated.

The percentage of scheduled visit in which the status of the patient is stable/recovered, or not recovered will be tabulated.

6.3.5.1. Seasonal distribution

Seasonal distribution of the exacerbations (total number of exacerbations, AECOPD events and of AECOPD events containing NTHi, Mcat, Strep, Entero/HRV, QPCR-Rhino and Influenza A and B) regardless of intensity will be reported by month.

6.3.5.2. Time to exacerbation

To investigate the effect of several baseline risk factors upon the time to first exacerbation since enrolment/first visit, Cox regression models will be fitted including several risk factors such as smoking status at enrolment (active smoker vs. former-smoker), number of exacerbation reported in the 12 months prior to enrolment (one, two and equal or above three), COPD grade at enrolment, age (40-49, 50-59, 60-69, ≥70 years), sex. Time to the first recurrence overall and by covariates will be depicted in Kaplan-Meier survival curves. Log-log survival curves will be displayed as well. If the proportionality of hazards assumption is met the log–log survival curves will be parallel for covariate strata. If there is an intersection between curves the covariate will be added as a stratification factor in the cox regression model.

To take into account that exacerbations may occur more than once over the follow-up time for a given patient, a Cox regression model for recurrent events using the counting process approach (Andersen, 1993) will be adopted (more details in section 6.3.5.2.1). The model will be fitted including smoking status at enrolment, history of exacerbation in the 12 months prior to enrolment, COPD severity at enrolment, age and sex as covariates.

Additional stratification/risk factor may be defined during the analyses.

6.3.5.2.1. Statistical methods: survival analysis

Kaplan-Meier curves

Kaplan-Meier survival curves will be produced to represent the cumulative survival probability over time.

The code follows.

```
Proc lifetest data = <dataset> method = KM plot = (S, LLS);
  time survt*Status(0);
  strata smoking;
run;
```

where survt = survival time in days (days to first exacerbation, for censored data the date of last contact will be used), status = censoring variable (0 = censored, 1 = event), and smoking = smoking status at enrolment (current and former smokers). The PLOTS=(S,LLS) option produces log–log curves as well as survival curves.

Cox Proportional Hazards regression model

To investigate the effect of several baseline risk factors upon the time to first exacerbation, Cox Proportional Hazards (PH) regression models will be fitted.

SAS codes

```
Proc phreg data = <dataset>;
  class GOLD age exacHist sex
  model survt*Status(0) = gold age exacHist / ties = exact rl;
  strata smoking;
run;
```

where GOLD = GOLD grade (2,3 or 4), age = age at enrolment (40-49, 50-59, 60-69, ≥70 years), smoking = smoking status at enrolment (current and former smokers), exacHist = number of exacerbations reported in the 12 months prior to enrolment (one, two and equal or above three) and sex (male and female). For the scope of the example, the variable *smoking* is considered to be a stratification factor.

Recurrent event survival analysis

SAS code

```
Proc phreg data = <dataset> covs(aggregate);
  model (start,stop)*event(0) = exacerbationHistory smoker;
  id pid;
run;
```

The code (START,STOP)∗EVENT(0) in the MODEL statement indicates that the time intervals for each observation are defined by the variables START and STOP and that EVENT = 0 denotes a censored observation.

6.3.6. Impact of AECOPD on HRQOL

Descriptive statistics (N, mean, median, SD, minimum, maximum, first and third quartiles) on the CAT and SGRQ-C scores (total score and symptoms, activity and impacts component scores) will be tabulated at each stable scheduled visit, overall and by COPD GOLD grade at enrolment.

CAT and SGRQ-C (total score and symptoms, activity and impacts component scores) changes from first visit to final scheduled visit will be displayed according to the number of exacerbation occurred in the observation period. Correlation will be evaluated using Spearman's rank-order correlation coefficient.

Descriptive statistics (N, mean, median, SD, minimum, maximum, first and third quartiles) on the CAT score will be tabulated for any stable and any mild, moderate or severe exacerbation visit.

6.3.7. Impact of AECOPD on lung function

Descriptive statistics (N, mean, median, standard deviation, minimum and maximum) on post-bronchodilator FEV_1% of predicted normal value and PEF will be tabulated at enrolment and final visit. Change from baseline will be also summarized overall and represented according to the number of exacerbations.

6.3.8. Impact of AECOPD on healthcare utilisation

Descriptive statistics of the daily number of healthcare utilisation (number of HCU in the relevant period divided by the number of days) within the stable period and within the exacerbation period will be presented overall and divided by the following unscheduled visits:

1. Number of physician office consultations,
2. Number of visits to urgent care,
3. Number of visits to emergency department,
4. Number of hospitalizations.

Healthcare use for each patient will be obtained through review of the patient's medical record (aided by patient daily self-reporting in the eDiary).

Healthcare utilisation will be summarized according to country and COPD grade at enrolment. In addition, healthcare utilisation will be summarized by AECOPD severity during exacerbations.

Overall HCU [total number and divided by type of HCU] per patient will be correlated to the number of exacerbations experienced during the observation time.

Current medication for COPD and additional COPD treatments prescribed by primary and secondary care physicians within the stable period and within the exacerbation period will be described by drug category (chronic use for COPD and chronic use for other

disorders) overall, by country and by COPD grade at enrolment. In addition, percentage of patients with or without inhaled corticosteroids at enrolment and incidence rate of confirmed exacerbation by use of corticosteroids will be tabulated.

6.4. Analysis of tertiary objective

The analyses of the tertiary objectives will be performed on the full analysis set.

6.4.1. End date of AECOPD using EXACT-PRO

Descriptive statistics (N, mean, median, standard deviation, minimum and maximum, first and third quartiles) on the average length of AECOPD episodes, as estimated through the EXACT-PRO (section 12), will be presented, for any, mild, moderate and severe AECOPD.

The frequency distribution of exacerbation confirmed at site visit or through medical records (with or without morning symptoms diary alert) will be tabulated versus EXACT-PRO exacerbations.

Descriptive statistics (N, mean, median, SD, minimum, maximum, first and third quartiles) for the EXACT daily scores will be tabulated at baseline, during the day before exacerbation, at exacerbation onset and at recovery according to investigator judgement. These statistics will be reported overall and for any mild, moderate or severe exacerbation visit.

Graphs detailing the EXACT-PRO scores before, during and after exacerbation onset as defined clinically will be also presented.

Difference from baseline, with 95% CI, of EXACT score measured at stable scheduled visit will be plotted.

6.4.2. Mixed infections

The simultaneous bacterial [culture/PCR] and viral presence in sputum at enrolment, any stable and any exacerbation visit [mild, moderate, severe] will be computed overall and by pathogens.

6.4.3. Load of bacteria and virus

Total number of patients providing a sputum sample, the number of sputum samples positive to specific bacteria (by culture) and the frequencies of semi-quantitative bacteriological load (few scattered, +, ++, +++) will be tabulated at each scheduled [stable] visit and at each exacerbation visit.

Total number of patients providing a sputum sample, the number of sputum sample positive to specific bacteria (by PCR)/HRV and average load will be tabulated at each scheduled [stable] visit and at each exacerbation visit.

6.4.4. Impact of occurrence of bacterial/viral pathogens in sputum on AECOPD

The following proportions and 95% CIs will be computed:

- Proportion of sputum samples which are positive for specific pathogens [NTHi, Mcat, S. pneumoniaeP. aeruginosa, K. pneumoniae, HRV] at an AECOPD visit but not at the previous stable visit and proportion of sputum samples which are positive for specific pathogens [NTHi, Mcat, S. pneumoniae, P. aeruginosa, K. pneumoniae, HRV] at an AECOPD visit but not at the previous visit (either stable or AECOPD visit) (apparition).

Proportion of sputum samples which are positive for specific pathogens [NTHi, Mcat, S. pneumoniae, P. aeruginosa, K. pneumoniae, HRV] which appeared for the first time, during the study, at an AECOPD visit (acquisition).For both acquisition and apparition, the CIs will be estimated using a GEE model assuming a binomial distribution for the response variable with logit as link function and a compound symmetry correlation matrix (exchangeable structure) to account for the within-patient correlation.

In addition to these proportions the following odds ratios (ORs) will be calculated:

- Odds of experiencing an AECOPD (vs. being in a stable state) for patients with sputum positive for selected pathogens (vs. those with negative sputum)
- Odds of experiencing an AECOPD (vs. being in a stable state) for patients with sputum positive for selected pathogens for the first time during the study (acquisition) vs. those without acquisition
- Odds of experiencing an AECOPD (vs. being in a stable state) for patients with sputum positive for selected pathogens which was negative at the previous visit (apparition) vs. those without apparition

To obtain the odds ratios conditional logistic regression models, stratified by patient, will be fitted. Separate analyses considering bacteria detected by culture and PCR as well as viruses detected by PCR will be conducted. For each analysis, three models will be run: the first will include the pathogen, season and the interaction term between the pathogen and season, the second will include the pathogen and will be adjusted for season (without the interaction term) and the third will only include the pathogen. Season will be treated as a dichotomous variable with high season (October to March) and low season (April to September)..

The conditional logistic model allows estimating the impact of independent variables from observations from each individual patient. In other words, by stratifying by patient, there is no need to adjust for time-invariant covariates such as age and sex.

6.4.5. Microbiome analysis

Further characterization of *H. influenzae, M. catarrhalis* and HRV strains will be performed. Details of the statistical analyses will be released in a separate additional analysis request.

6.4.6. Biomarkers

Descriptive statistics (N, mean, median, SD, minimum, maximum, first and third quartiles) for the level of each quantitative biomarker/haematology parameter will be tabulated for visit 1 and visit 3. The difference in sample mean will be reported together with a confidence interval.

Number and percentage of patients with fibrinogen, human serum C reactive protein (hsPCT), C-X-C Motif Chemokine Ligand 10 (CXCL10) values outside the laboratory normal ranges will be tabulated for visit 1 and visit 3.

Descriptive statistics will be presented overall and by GOLD grade at enrolment.

Changes in biomarkers will be evaluated according to the number of exacerbations occurred in the observation period using Pearson's correlation coefficient (or Spearman's rank-order correlation coefficient if the normal distribution assumption is not met).

The mean number of exacerbations per year overall, by severity of exacerbation and by presence of Hi and Mcat in the sputum collected at exacerbation visits will be calculated for patients with blood eosinophil levels ≥2% and for those with blood eosinophil levels <2% (the level measured for each patient at visit 1 will be used for this categorization).

6.5. Analysis of safety

The analysis of serious adverse events will be performed on the full analysis set.

Adverse events (AEs) and serious adverse events (SAEs) considered possibly related to the study participation and patients withdrawal due to AEs and SAEs during the entire study period will be described in detail.

7. ANALYSIS INTERPRETATION

All analyses are descriptive. The use of these descriptive analyses wil be limited to supportive analysis or hypothesis generation.

8. CONDUCT OF ANALYSES

8.1. Sequence of analyses

The analyses will be performed in two steps:

- An interim analysis will be performed when at least 40 AECOPD evaluable sputum samples are available. It will include all data available at the time of analysis. The rate of positive samples will be computed for bacterial pathogen (including Hi, NTHi and Mcat) together and per pathogen. The incidence rate of AECOPD overall and associated to bacterial pathogen will be computed as well. The analysis will be done based on as clean as possible data. No study report will be written at this stage.

- A final analysis of all objectives will be performed when data obtained up to planned last patient last visit are available. A final study report will be written at this stage.

Description	Analysis ID	Disclosure Purpose (CTRS=public posting, SR=study report, internal)	Dry run review needed (Y/N)	Study Headline Summary (SHS)requiring expedited communication to upper management (Yes/No)	Reference for TFL
Final analysis	E1_01	Study report	YES	NO	See TFL TOC
Interim analysis	E1_02	Internal	NO	NO	See TFL TOC

8.2. Statistical considerations for interim analyses

All analyses are descriptive and therefore no statistical adjustment for the interim analysis is required. No study report will be written for interim analysis.

9. CHANGES FROM PLANNED ANALYSES

Not applicable

10. LIST OF FINAL REPORT TABLES, LISTINGS AND FIGURES

The TFL TOC provides the list of tables/listings and figures needed for the study report. It also identifies the tables eligible for each analyses and their role (synopsis, in-text, post-text, SHS, CTRS,…). Note that all TFL aimed to be included as post-text are noted as post-text even if these are tabulation of individual data such as listing of SAE. The post-text material contain all source material for the study report and accordingly a post-text table may be redundant with an in-text table.

11. ANNEX 1 DATA DERIVATION RULE AND STATISTICAL METHODS

11.1. Statistical Method References

Andersen P.K., Borgan O., Gill R.D., and Keiding N. Statistical Models Based on Counting Processes. New York: Springer-Verlag 1993.

Atkinson AC. Plots, Transformations and Regression. Clarendon Press. 1985

Burton P, Gurrin L, Sly P. Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level mixed modelling. *Stat Med*. 1998;17:1261–91.

Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of binomial. *Biometrika*. 1934;26:404-413

Liang KY, and Zeger SL. Longitudinal Data Analysis Using Generalized Linear Models." Biometrika. 1986; 73:13–22.

McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med*. 2012;22(3):276-82.

11.2. Data derivation

11.2.1. COPD severity stage

The baseline COPD severity stage will be derived by the investigator from the spirometry results at the screening visit. The spirometric classification of airflow limitation in COPD patients is based on post-bronchodilator FEV_1 and can be divided into four GOLD grades (GOLD 1 - mild, GOLD 2 - moderate, GOLD 3 - severe and GOLD 4 - very severe) [Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of Chronic Obstructive Pulmonary Disease, updated 2013. Available from: http://www.goldcopd.org/uploads/users/files/GOLD_Report_2013_Feb20.pdf]. For the purpose of this study only patients with COPD graded from moderate to very severe will be enrolled.

Table 1 COPD severity stages included in the study

GOLD grade	**In patients with $FEV_1/FVC<0.70$**
GOLD 2: Moderate	$50\% \leq FEV_1 < 80\%$ predicted
GOLD 3: Severe	$30\% \leq FEV_1 < 50\%$ predicted
GOLD 4: Very severe	$FEV_1 < 30\%$ predicted

The resulting COPD severity variable will be used in all models in which COPD at baseline will be taken into account.

A four-grade refined assessment of COPD severity stages will be computed according to the newest GOLD criteria [GOLD, 2017 Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of Chronic Obstructive Pulmonary Disease, updated 2017. Available from: http://goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/]. In the assessment scheme, patients should undergo spirometry to determine the severity of airflow limitation (i.e., spirometric grade). They should then undergo assessment of either dyspnea using mMRC or symptoms using CAT. Finally, their history of exacerbations (including prior hospitalizations) should be recorded.

For the purpose of this study we will only adopt CAT rather than adopting mMRC dyspnea scale. This choice is because the two tools are not interchangeable [Kim S, Oh J, Kim YI, Ban HJ, Kwon YS, Oh IJ, Kim KS, Kim YC, Lim SC. Differences in classification of COPD group using COPD assessment test (CAT) or modified Medical Research Council (mMRC) dyspnea scores: a cross-sectional analyses. *BMC Pulm Med*. 2013;13:35] and it is now recognized that COPD impacts patients beyond just dyspnea [Jones PW. Health status measurement in chronic obstructive pulmonary disease. *Thorax.* 2001; 56(11):880-7].

The four grades are:

- Patient Group A – Low Risk, Less Symptoms Typically GOLD 1 or GOLD 2 (Mild or Moderate airflow limitation); and 0-1 exacerbation per year and no hospitalization for exacerbation; and CAT score < 10;
- Patient Group B – Low Risk, More Symptoms Typically GOLD 1 or GOLD 2 (Mild or Moderate airflow limitation); and 0-1 exacerbation per year and no hospitalization for exacerbation; and CAT score ≥ 10;
- Patient Group C – High Risk, Less Symptoms Typically GOLD 3 or GOLD 4 (Severe or Very Severe airflow limitation); and ≥ 2 exacerbations per year or ≥ 1 with hospitalization for exacerbation; and CAT score < 10;
- Patient Group D – High Risk, More Symptoms Typically GOLD 3 or GOLD 4 (Severe or Very Severe airflow limitation); and ≥ 2 exacerbations per year or ≥ 1 with hospitalization for exacerbation; and CAT score ≥ 10.

Spirometry results for GOLD classifications, number of exacerbations per year, and number of exacerbation requiring hospitalization will be taken from the screening visit (visit 0). CAT score will be derived from visit 1.

11.2.2. HRQOL

11.2.2.1. CAT

The CAT index will be derived as the sum of the ratings recorded for each of the eight individual items. Each of these items has 6 possible scores (0, 1, 2, 3, 4 or 5), leading to a range of 0 to 40 for CAT score.

No missing items are expected because the e-diary system prompts the patient to compete an item of CAT to proceed to the following one. Partial questionnaires will be excluded.

11.2.2.2. SGRQ-C

The SGRQ-C total score will be derived as the weighted sum of the forty individual items leading to a range of 0 to 100 as detailed in the reference manual [St George's Respiratory Questionnaire for COPD patients (SGRQ-C), version 1.3, 2016].

The SGRQ-C symptoms, activity and impacts component scores will be derived as the weighted sum of subset of items as detailed in the reference manual.

No missing missing items are expected because the e-diary system prompts the patient to compete an item of SGRQ-C to proceed to the following one. Partial questionnaires will be excluded.

11.2.3. Sputum Quality

The quality of sputum is defined as follow:

- < 10 squamous epithelial cells/field: sample of good quality
- 10-25 squamous epithelial cells/field with significant numbers of cells derived from the lower respiratory tractus (ciliary epithelial cells, bronchial cells, round cells): sample of moderate quality
- > 25 squamous epithelial cells/field: sample of bad quality

11.2.4. Bacterial and viral PCR assays

- Multiplex real-time PCR (MPB1) assay will target *Haemophilus influenzae, Moraxella catarrhalis, Streptococcus pneumoniae* with quantitative results (copies/mL of sample).

For qualitative analyses, quantitative results must be converted in POS/NEG results. Only samples above the cut off will be considered as POSITIVE.

The positivity cut-off for each quantitative PCR assay will be defined at time of data release.

For quantitative analyses, only samples with load above the cut off will contribute to mean and SD computation.

- Multiplex real-time PCR (MPB2) assay will target Streptococcus pyogenes, Staphylococcus aureus, Pseudomonas aeruginosa with qualitative results only (POS/NEG).
- Multiplex reverse transcription PCR assay will target different viral pathogens with qualitative results only (POS/NEG).
- HRV qPCR assay will target HRV with quantitative results (copies/mL).

For qualitative analyses, qualitative and quantitative results will be combined with the following algorithm:

Sample	Allplex HRV result	Allplex HEV result	HRV RT-qPCR	Final HRV result (variable to be derived)
1	POS	NEG	≥ LOD*	POS
2	POS	NEG	< LOD	POS
3	NEG	POS	≥ LOD	POS
4	NEG	POS	< LOD	NEG
5	NEG	NEG	Not tested	NEG
6	POS	NEG	0	POS
7	POS	POS	≥ LOD	POS

*The positivity cut-off for each quantitative PCR assay will be defined at time of data release.For quantitative analyses, only positive samples will contribute to mean and SD computation.

11.3. Handling missing data

Missing or non-evaluable measurements will not be replaced. The only exception is detailed below.

The number of exacerbations per year will be extrapolated for patients withdrawing from the study to provide an estimate of the number of exacerbations over the one year observation period. The number of exacerbations in a year will be calculated by multiplying the number of exacerbations experienced by the patient by 13 and dividing by the number of 4-week periods the patient was followed up [Stockley R.A., Chopra N, Rice L. Addition of salmeterol to existing treatment in patients with COPD: a 12 month study. *Thorax.* 2006;61:122-128].

$$\text{Number of exacerbations per year} = \frac{\text{Number of exacerbations} * 13}{\text{Number of 4-week observation period intervals}}$$

The calculation of exacerbation rate will be based on follow-up period intervals of four weeks to avoid obtaining high imputed rates if a patient withdrew very early from the study after experiencing an exacerbation. Four-week intervals will be adopted since treatment courses for moderate/severe exacerbations are <=2 - 4 weeks when appropriate.

11.3.1. Date derivation

Partial dates will not be considered as missing or not evaluable date. In case day is missing, 15th of month is used. In case day and month are missing, 30th of June is used.

Partial or missing initial and end dates for exacerbation will not be considered to measure the length of an exacerbation.

11.3.2. Missing data for electronic Diary Card alert

- Days with missing data will not be ignored when defining start of an e-diary signal: e.g., if a patient has two qualifying symptoms followed by a missing day, if the next day with data has two qualifying symptoms this will not be defined as an exacerbation. Hence, the start of an e-diary signal of an exacerbation will require 2 consecutive non-missing days of worsening. If there is partially recorded symptom data at a visit, however, only the observed values will be considered but the day will not be discarded.
- Days with missing data will be ignored when defining end of an e-diary signal of an exacerbation: e.g., if a patient has two days missing in the baseline period and one day missing in the 3- day moving average, both baseline and 3-day moving average values will be calculated using observed data only. Similar rules will apply when there is partially recorded symptom data at a visit (and daily average values will be calculated among the observed values).
- All available data will be used for each patient.
- As a general rule, the baseline symptom count score, and 3 day moving averages are calculated as the average count score of the observed days in the baseline period. For example, if an e-diary signal occurs from day 10 onwards, the baseline period will be the average between day1 and day2 (day1 being the first study day). If day 13 is missing, the 3 day moving average at day 14 will only take into account day 14 and day 15.
- If there are more than 2 consecutive days of missing data after an e-diary signal not resolved, the e-diary signal will not be considered resolved until the next day with observed data if the corresponding moving average is below the baseline value. In order words, if the 3-day average daily score is missing, the day will be counted in the duration of the episode.

11.4. Number of decimals

The following decimal description will be used for the analyses.

Parameters	Number of decimal digits
% of count, including LL & UL of CI	1
p-value	3
Mean, median	Number of decimals in the raw data + 1
SD	Number of decimals in the raw data + 1
Minimum, maximum, range	Number of decimals in the raw data

LL = Lower Limit UL = Upper Limit CI = Confidence Interval
SD = Standard deviation

11.5. Methodology for computing CI

- All CI computed will be two-sided at 95%significant level.
- The exact 95% CIs for a proportion within a group will be calculated using Clopper Pearson method [Clopper 1934].

12. ANNEX 2: AECOPD DEFINITION AND CALCULATION RULE

12.1. Definition, onset, recovery, duration of an AECOPD event

AECOPD events are confirmed according to investigator judgment after an AECOPD visit which is aimed to exclude worsening in symptoms not related to an AECOPD event.

AECOPD visits are scheduled after an electronic Diary Card alert confirmed by the investigator (by phone call or at the study site) or after spontaneous site visits due to worsening symptoms without any alert.

Electronic Diary Card alerts refer to daily symptoms recorded the morning after (also called morning symptoms). Alerts are based on the Anthonisen criteria:

- Worsening of two or more of the following major symptoms for at least two consecutive days*,**: dyspnea, sputum volume, sputum purulence (colour)

Or

- Worsening of any major symptom together with any of the following minor symptoms for at least two consecutive days*,**: sore throat, colds (nasal discharge and/or nasal congestion), fever (oral temperature $\geq 37.5°C$) without other cause, increased cough, increased wheeze.

** The same two symptoms do not have to be present on both days as long as at least one major symptom is present on both days.*

AECOPD onset will be defined as the first day of the two consecutive days of worsening symptoms.

AECOPD recovery will be determined/confirmed by the investigator/delegate during (a) follow-up phone call(s) which will take place every 2 weeks until the AECOPD has resolved. The end date will be based on when the investigator/delegate determines that the AECOPD symptoms have resolved. In determining this end date, consideration will be given to symptoms recorded in the electronic Diary Card and patient assessment during the phone calls.

AECOPD duration will be defined as the number of days from AECOPD onset (included) and AECOPD recovery (not included).

Duration = date2 – date1 + 1,

with date2: AECOPD recovery date,

date1: AECOPD onset date

12.2. Recovery and duration of an exacerbation according to EXACT-PRO:

An EXACT-PRO AECOPD event is defined as a 12 point increase above baseline for 2 consecutive days, or a nine point increase for 3 consecutive days.

The baseline EXACT-PRO score is calculated as the average daily EXACT-PRO score over days 14 to 8 preceding AECOPD symptomatic onset.

AECOPD recovery according to EXACT-PRO corresponds to the day at which the EXACT-PRO score return to its baseline value (or below) [Mackay AJ, Donaldson GC, Patel AR, Singh R, Kowlessar B, Wedzicha JA. Detection and severity grading of COPD exacerbations using the exacerbations of chronic pulmonary disease tool (EXACT). *Eur Respir J*. 2014;43(3):735-44.].

The AECOPD duration according to EXACT-PRO will be defined as the number of days from AECOPD onset (included) and AECOPD recovery according to EXACT-PRO (not included).

12.2.1. EXACT-PRO baseline score at enrolment

An EXACT-PRO baseline score at enrolment will be computed as the mean within-patient score over 7 days, with data present for a minimum of 4 of the 7 days. If fewer than 4 days of data are available, the EXACT baseline score cannot be calculated.

12.3. Time between 2 *e-diary* signals for exacerbations/evaluation of baseline symptom score:

The minimum duration between 2 e-diary signals with distinct baseline value is 15 days, which is the minimal period for which a new full exacerbation-free baseline can be calculated for the next e-diary episode.

If a new worsening of symptoms occurs between 6 and 14 days after the end of a previous exacerbation, the baseline period used will be the same baseline as the one of the original exacerbation. Therefore, it will last until the daily symptom count score has gone below the baseline symptom score of the previous event.

12.4. Definition of Missed AECOPD visits

Missed AECOPD visits with medical records (dashed boxes 2 and 7 in Figure 1) will capture:

- an alert from the electronic Diary Card without a site visit but that cannot be discounted as a possible exacerbation signal after reconciliation (detailed in the Study Procedure Manual),

and/or

- an exacerbation reported retrospectively during a scheduled visit clinically documented for which there has not been a corresponding exacerbation visit.

Onset date, recovery date and severity will be retrieved by medical records.

For the analyses not involving the collection of any biological specimen, exacerbations will refer to AECOPD events plus missed AECOPD visits with medical records.

Missed AECOPD visits without medical records, also referred to as Potential AECOPDs, (dashed box 3 in Figure 1) will capture:

- an alert from the electronic Diary Card, assessed AECOPD during the phone call, without a site visit and without AECOPD medical records.

12.5. Unconfirmed AECOPD event with morning e-diary signal alert notification.

An alert from the electronic Diary Card not confirmed to be an AECOPD after contact by phone call or at the study site.

12.6. Grading of severity of an exacerbation

Severity of exacerbations is defined as per protocol:

- Mild: Worsening symptoms of COPD that are self-managed by the patient.
- Moderate: Worsening symptoms of COPD that require treatment with oral corticosteroids and/or antibiotics.
- Severe: Worsening symptoms of COPD that require treatment with in-patient hospitalisation or home care intervention.

Severity of the exacerbation will not be derived but they will be taken directly from the information entered in the CRF (i.e. Severity of exacerbation will be taken from the investigator conclusion of the exacerbation visit).

13. ANNEX 3: STUDY SPECIFIC MOCK TFL

Study specific mock TFL is provided in a separate document.

GlaxoSmithKline	**Statistical Analysis Plan**
Detailed Title:	A prospective, epidemiological, multi-country, cohort study to assess the occurrence of potential bacterial and viral pathogens in stable chronic obstructive pulmonary disease (COPD) and during acute exacerbations of COPD (AECOPD), in moderate to very severe COPD patients, in Asia Pacific.
eTrack study number and Abbreviated Title	201112 (EPI-NTHI-001 BOD APA)
Scope:	All data pertaining to the above study.
Date of Statistical Analysis Plan	Amendment 1:14 September 2018
Co-ordinating author:	PPD, Biostatistician in Epidemiology PPD, Biostatistician in Epidemiology
Reviewed by:	PPD, Lead Epidemiologist PPD, Lead Statistician in Epidemiology PPD, Lead Statistical Analyst PPD, Statistical Analyst PPD, Scientific writer PPD, Regulatory Affair PPD, SERM physician PPD, Public Disclosure Representative
Approved by:	PPD, Clinical and Epidemiology Project Lead delegating to PPD, Lead Epidemiologist PPD, Lead Statistician in Epidemiology PPD, delegating to Binutha Pereira, Scientific Writer PPD, Lead Statistical Analyst

APP 9000058193 Statistical Analysis Plan Template (Effective date: 14 April 2017)

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

LIST OF ABBREVIATIONS

A. baumannii	*Acinetobacter baumannii*
AE	Adverse event
AECOPD	Acute exacerbation of COPD
CAT	COPD assessment test
CI	Confidence Interval
COPD	Chronic Obstructive Pulmonary Disease
CRF	Case Report Form
EXACT-PRO	Exacerbations of Chronic Pulmonary Disease Tool - Patient Reported Outcome
FAS	Full Analysis Set
FEV_1	Forced expiratory volume in 1 second
GOLD	Global Initiative for Chronic Obstructive Lung Disease
GSK	GlaxoSmithKline
H. haemolyticus	*Haemophilus haemolyticus*
H. influenzae	*Haemophilus influenzae*
HRQOL	Health-related quality of life
HRV	Human rhinoviruses
K. pneumoniae	*Klebsiella pneumoniae*
LL	Lower Limit of the confidence interval
M. catarrhalis	*Moraxella catarrhalis*
MA	Main Analyses
MedDRA	Medical Dictionary for Regulatory Activities
NTHi	Non-Typeable *Haemophilus influenzae*
P. aeruginosa	*Pseudomonas aeruginosa*
PCR	Polymerase chain reaction

PEF	Peak Expiratory Flow
RSV	Respiratoir Syncytiaal Virus
RT-PCR	Real-time polymerase chain reaction
SAE	Serious adverse event
S. aureus	*Staphylococcus aureus*
SAP	Statistical Analysis Plan
SD	Standard Deviation
SGRQ-C	St. George's Respiratory Questionnaire for COPD patients
S. pneumoniae	*Streptococcus pneumoniae*
SR	Study Report
TFL	Tables Figures and Listing template annexed to SAP
UL	Upper Limit of the confidence interval

1. DOCUMENT HISTORY

Date	Description	Protocol Version
14-SEP-2018	SAP Amendment 1 Updates done before Interim Analysis: - analysis population was adapted to Protocol Amendment 2 - elimination codes were aligned to PDMP v.4 - analyses added to Interim Analysis timepoint - sputum quality - alignment with the protocol/CRF - more details about laboratory methods were added - several minor corrections	Protocol Amendment 2 19-OCT-2017
28-JUL-2017	Version 1	Protocol Amendment 1 13-FEB-2017

2. STUDY DESIGN

Type of design: epidemiological, prospective, interventional, multi-country, cohort study. The study targets enrolling approximately 200 stable moderate to very severe COPD patients, males and females, aged 40 years or older at the time of enrolment, with at least 1 documented moderate or severe AECOPD in the year before enrolment.

This study will be conducted in several sites among several countries in Asia Pacific [China (Hong Kong), Philippines, Korea, Taiwan].

Type of study: self-contained.

Duration of the study: approximately 1 year for each patient.

Study visits/contacts:

- Screening visit
- Three scheduled visits occurring at 6 months intervals.
- For each AECOPD: AECOPD visit (within 96 hours of the onset of the symptoms) and follow-up phone call(s) (at least every 2 weeks until the AECOPD has resolved). Follow-up phone contacts will define end of AECOPD.

The following group names will be used for the statistical analyses:

Group order in tables	Group label in tables	Group definition for footnote
1	COPD	Patients with moderate to very severe COPD

3. OBJECTIVES

3.1. Primary objective

To estimate the proportion of potential bacterial and viral pathogens (overall and by species) detected in sputum of stable COPD patients and during AECOPD, using respectively bacteriological methods and viral PCR, over the course of 1 year.

3.2. Secondary objectives

- To evaluate the concordance between PCR and bacteriological methods data in sputum for potential bacterial pathogens.
- To estimate the proportion of potential bacterial and viral pathogens (overall and by species) detected in sputum of stable COPD patients, by GOLD grade.
- To estimate the proportion of potential bacterial and viral pathogens (overall and by species) detected in sputum during AECOPD, by severity of AECOPD.
- To estimate the incidence rate of all-cause AECOPD, overall and by GOLD grade.
- To describe the severity and duration of all-cause AECOPD, overall and by GOLD grade.
- To assess the impact of all-cause AECOPD on HRQOL.
- To assess the impact of all-cause AECOPD on lung function.
- To assess the impact of all-cause AECOPD on healthcare utilisation.

3.3. Tertiary objectives

- To assess the use of EXACT-PRO for determining the end date of AECOPD.
- To evaluate the load of bacterial and viral pathogens in stable COPD and during AECOPD.
- To collect biological specimens for future respiratory disease-related testing:
  - Aliquots of sputum samples for assay development and microbiome analysis at stable visits and during exacerbation.
  - Blood sampling for potential biomarkers for identification/quantification of biomarkers at Visit 1 and Visit 3.

4. ENDPOINTS

4.1. Primary endpoint

- Occurrence of potential bacterial and viral pathogens in sputum of stable COPD patients and during AECOPD, over the course of 1 year:
  - Bacterial pathogens, as identified by bacteriological methods, including (but not necessarily limited to) *H. influenzae*, *M. catarrhalis*, *S. pneumoniae*, *S. aureus*, *P. aeruginosa*, *K. pneumoniae* and *A. baumannii*.
  - Viral pathogens, as identified by PCR, including (but not necessarily limited to) RSV, parainfluenza virus, enterovirus/ HRV, metapneumovirus, influenza virus, adenovirus, bocavirus and coronavirus and by HRV quantitative RT-PCR.

4.2. Secondary endpoints

- Occurrence of potential bacterial pathogens in sputum of stable COPD patients and during AECOPD, as measured by real-time qualitative PCR/ quantitative PCR and compared to data from bacteriological methods, over the course of 1 year:
  - Including (but not necessarily limited to) *H. influenzae, M. catarrhalis, S. pneumoniae, S. aureus* and *P. aeruginosa.*
- Occurrence of potential bacterial and viral pathogens (overall and by species) in sputum of stable COPD patients by GOLD grade, over the course of 1 year.
- Occurrence of potential bacterial and viral pathogens (overall and by species) in sputum during AECOPD by severity of AECOPD, over the course of 1 year.
- Incidence rate (per patient per year) of AECOPD, overall and by GOLD grade, over the course of 1 year.
- Severity of AECOPD, overall and by GOLD grade, over the course of 1 year.
- Duration of AECOPD, overall and by AECOPD severity, over the course of 1 year.
- CAT score in stable COPD patients and during AECOPD, over the course of 1 year.
- SGRQ-C score in stable COPD patients, over the course of 1 year.
- FEV_1% of predicted normal value in stable COPD patients, at Pre-Month 0 and Month 12.
- Healthcare utilisation, over the course of 1 year.

4.3. Tertiary endpoints

- EXACT-PRO scores in stable COPD patients and during AECOPD, over the course of 1 year.
- Bacterial load measured by both culture and PCR in COPD and during AECOPD.
- HRV load measured by PCR in COPD and during AECOPD.

5. ANALYSIS SETS

The following data sets will be defined.

5.1. All screened set

The all screened set will include all screened patients.

5.2. All enrolled set

The all enrolled set will include all screened patients except for screening failures.

5.3. Full Analysis Set

The Full Analysis Set (FAS) will include all enrolled patients except for those who discontinued the study at Visit 1.

Study objectives will be assessed on the FAS. The population set for each analysis will change according to the subjects evaluable for the specific endpoint.

5.3.1. Change in analysis population from protocol amendment 2

The analysis population is aligned with protocol amendment 2.

5.4. Criteria for eliminating data from the Analysis Sets

Elimination codes are used to identify patients to be eliminated from analysis. Detail is provided below for each set.

5.4.1. Elimination from All Screened Set

Code 900 (invalid informed consent or fraud data) will be used for identifying patients eliminated from All Screened Set.

5.4.2. Elimination from All Enrolled Set

Code 900 (invalid informed consent or fraud data) and code 2010 (Protocol violation (inclusion/exclusion criteria) will be used for identifying patients eliminated from All Enrolled Set.

5.4.3. Elimination from the Full Analysis Sets (FAS)

Code 900 (invalid informed consent or fraud data) and code 2010 (Protocol violation (inclusion/exclusion criteria) will be used for identifying patients eliminated from FAS.

5.4.4. Elimination codes details

A list of reasons for elimination from the analyses (interim analysis and final analysis), is reported below:

Code	Condition under which the code is used
800	Fraudolent data
900	1. Informed consent deviations 1.a. Signed informed consent not available on site 1.b. Wrong informed consent version signed 1.c. Informed consent not signed and/or dated by subject 1.d. Informed consent not signed and/or dated by appropriated site staff 1.e. Informed consent not signed prior to any study procedure 1.f. Other informed consent deviations 2. Eligibility deviations 2.a. Biological sample specimen procedures 2.a.a. PII recorded in label of sample
2010	1. Eligibility criteria not met
2040	1. Excluded medication, vaccine, or device 1.a. Prohibited medication as per protocol (e.g., antibiotics administration will impact the primary endpoint)
2050	1. Not withdrawn after developing withdrawal criteria 1.a. Not withdrawn after developing withdrawal criteria other than prohibited vaccination/medication

Code	Condition under which the code is used
2090	1. Assessment or time point completion 1.a. Out of window visit (assessment) for efficacy 1.a.a. Sputum sample collection outside the 96 hour-interval since exacerbation onset
2100	1. Study procedures 1.a. Biological sample specimen procedures 1.a.a. Low volume (not sufficient to perform testing) 1.a.b. Mislabeling (sample not tested)1.a.c. Incorrect or expired sample tube used 1.a.d. Incorrect spinning/processing of sample
2110	1. Study procedures 1.a. Biological sample specimen procedures 1.a.a. Incorrect or expired sample tube used 1.a.b. Incorrect spinning/processing of sample
2120	1. Assessment or time point completion 1.a. Other assessment not properly performed 1.a.a. Sputum culture not completed within six hours, 1.a.b. Culture: error in testing algorithm for identification (vs AP, eg.no optochin for Strep.) 1.b. Assessmet not properly performed 1.b.a. Sputum processing priorities not respected.

Code	Condition under which the code is used
	1.b.b. Frozen sputum DTT before culture 1.b.c. Sputum DTT aliquots not frozen <6hrs 1.b.d. Sputum not diluted after glass smear preparation 1.b.e. Sputum diluted with concentrated DTT 1.b.f. Gram stain not read on the same day as preparation (not an issue if still done < 10 days) 1.b.g. Sputum glass smear for quality not prepared or assessment not performed 2. Study procedures 2.a. Biological sample specimen procedures 2.a.a. Central/internal/external lab deviations, 2.a.b. Mislabelling (sample tested), 2.a.c. Wrong packaging that impacts quality of the sample (i.e. room temperature packaging for shipping frozen ELISA samples). 2.a.d. Temperature log not maintained as required or not tracked 2.a.e. "Blood" or Sputum temperature deviations from range defined in protocol and/or SPM/lab manual - room temperature and freezer -70°C 2.a.e.a. Sputum storage (only if not tested on same day), then storage cooled (+2-8°C): no impact if < 24hrs. If > 24hrs but still <72hrs, potential impact, to evaluate case by case. If >72hrs, elimination 2.a.e.b. Sputum DTT (-70°C): range +/-20°C. Lab is looking for additional stability data. Normally, ok at -20°C for 1 month (for PCR)

Code	Condition under which the code is used
2130	1. Study procedures 1.a. Biological sample specimen procedures 1.a.a.Testing performed on samples not aligned with ICF, 1.a.b. Subjects bled but not supposed to be bled

Codes from 2040 to 2130 are protocol deviations not leading to elimination from the population sets above mentioned but that may lead, judged case by case, to sample elimination from a specific statistical analysis.

6. STATISTICAL ANALYSES

The total number and the distribution of the following data sets overall and among the study country will be tabulated:

- All screened set and screening failures.
- All enrolled set.
- The full analysis set.
- Patients completing the study and withdrawn patients.

In addition, screening failures will be summarized according to the reason for failure.

Withdrawal status will be summarised according to the reason for withdrawal. The number of withdrawn patients will be tabulated by study visit and overall.

6.1. Analysis of demographics/baseline characteristics

The analysis of demographic/baseline characteristics will be performed on the full analysis set and on the all enrolled set only if the percentage of enrolled patients excluded from the FAS is 5% or more.

The following characteristics at enrolment will be summarized using descriptive statistics:

- Demographics such as age at enrolment, sex and race.
- Present and past medical history.
- History of AECOPD within the previous year (overall and by AECOPD severity).
- Pre- and post-bronchodilator spirometry results.
- COPD status using the GOLD grade.
- Smoking/biomass exposure history status.
- Selected results and several risk factors from the ATS-DLD-78-A questionnaire.
- Physical examination/vital signs results such as height, weight, BMI, heart rate, respiratory rate, systolic and diastolic blood pressure.
- Chest X-rays results.
- Intercurrent comorbidities (reported at visit 1)
- History of pneumococcal and influenza vaccination (reported at visit 1)

The following descriptive statistics will be adopted:

- The frequency and the percentage of each category will be reported for categorical variable.
- Number of non-missing observations, mean, median, standard deviation, minimum and maximum will be provided for quantitative variable.

All tables will be also generated by country.

Demographics and baseline characteristics will be also summarized for study sample for interim analysis.

6.2. Analysis of primary objective

The analysis of the primary objective will be performed on the FAS.

Average proportion of patients with sputum samples positive for bacterial or viral pathogens overall and for specific species, and 95% CIs will be computed at any stable scheduled visit and at any exacerbation visit of the entire study period. Confidence intervals will be estimated using a Generalized Estimating Equations (GEE) model assuming a binomial distribution for the response variable with logit as link function and a compound symmetry correlation matrix (exchangeable structure) to account for the within-patient correlations (Liang,. 1986) (more details in section 6.2.1). The crude proportions (i.e. number of positive sputum samples by number of visits with a sputum sample) and 95% CI will be reported as well.

Proportion of patients with sputum samples positive for bacterial/viral pathogens overall and for specific species, will be computed, with exact 95% confidence intervals (CIs), at each stable* scheduled visit (enrolment, visit 1, visit 2, visit3) and at each exacerbation visit confirming acute exacerbation**,***.

** A confirmed stable visit will be defined as a scheduled study visit for which the investigator confirms in the eCRF that the patient is stable/has recovered from a previous exacerbation.*

*** If an AECOPD occurs at the time when one of the scheduled [stable] study visits is planned, it should be handled and recorded as an AECOPD visit. If possible, the stable study visit should be re-scheduled to a later date, within the ± 14 days' time window from the scheduled appointment, when the patient is stable again.*

**** The definition of acute exacerbation of COPD (AECOPD) event is detailed in section 12.*

In details the following pathogens will be evaluated:

- Bacteria including, but not necessarily limited to, *H. influenzae**, *M. catarrhalis*, *S. pneumoniae*, *S. aureus*, *P. aeruginosa*, *K. pneumoniae* and *A. baumannii*) by bacteriological methods.

  **H. influenzae* positive cultures will be further characterized to differentiate *H. haemolyticus,* typeable *H. influenzae* and non typeable *H. influenzae* by PCR assay.

- Viruses including, but not necessarily limited to, RSV, parainfluenza virus, enterovirus/ HRV, metapneumovirus, influenza virus, adenovirus, bocavirus and coronavirus using qualitative PCR and HRV also by quantitative RT-PCR.

- Patients providing a sputum sample as per protocol are all evaluable patients (*i.e.* those meeting all eligibility criteria and complying with procedures described in the protocol*,**). According to protocol, the interval of ± 14 days for any scheduled visit will not be a criterion for exclusion from the analyses. The number and percentage of patients with a stable visit within and without this time window will be presented.

** Patients attending a study visit and providing sputum sample.*

*** Patients using antibiotics on a continual basis (defined as more than 1 month in total) will be allowed to continue study participation, but may be eliminated from the analyses.*

If more than 10% of sputum samples have bad quality (as defined in section 11.2.3 in section 11), the analyses for primary objective will be repeated including only sputum samples with good and moderate quality.

An interim analysis will be performed when at least 40 AECOPD sputum samples are collected. This analysis will compute the rate of sputum sample positive any bacterium and by species (including Hi, NTHi and Mcat).

6.2.1. Statistical methods: analysis of correlated Data Using Generalized Estimating Equations

Average proportion of patients with sputum positive for bacteria/viruses and 95% CI will be computed using a GEE model. GEE approach uses weighted combinations of observations to extract the appropriate amount of information from correlated data such as the response measurements across multiple stable visits on the same patient (Burton, 1998).

For the purpose of the analyses a binomial distribution for the response variable with logit as link function will be assumed and an exchangeable correlation matrix will be chosen to account for the within-patient correlations.

The exchangeable correlation matrix assumes that the correlation between any two responses of the i^{th} individual is the same.

Other commonly used within-patient correlation matrices are as follows:

- Independence, repeated observations are uncorrelated
- Unspecified (unstructured), correlations within any two responses are unknown and need to be estimated
- Autoregressive of first order [AR(1)], assuming the interval length is the same between any two observations. Measurements taken further apart are less correlated than those taken closer together.

The model with the exchangeable correlation matrix will be adopted in the main analysis. Models with the other matrices could be run as sensitivity analyses. QIC and the related QICu statistic will be used to compare GEE models with different correlation matrices.

SAS codes

```
ods output ParameterEstimates=out_parm ClassLevels=Class;
proc genmod data=<dataset> descending;
 class pid;
 model y =  / dist = bin link = logit lrci ;
 Repeated subject=pid/sorted corr=cs PRINTMLE;
 by bacteria visit;
run;

data result;
 set out_parm(where=(parameter='Intercept') drop=ChiSq ProbChiSq DF);
 format percent LL UL percent8.1;
 Percent=exp(estimate)/(1+exp(estimate));
 LL=exp(LowerLRCL)/(1+exp(LowerLRCL));
 UL=exp(UpperLRCL)/(1+exp(UpperLRCL));
 unit=_N_;
run;
```

where pid = personal identifier, y = positivity to pathogen (0, 1), bacteria = type of bacteria, visit = type of visit (stable or exacerbation).

The REPEATED statement invokes the GEE method, specifies the correlation structure, and controls the displayed output from the GEE model. The option SUBJECT = PID identifies *pid* as the clustering variable, and the CORR = CS option specifies an exchangeable working correlation structure. Other matrices will be specified as CORR =: IND [independence], UN [unstructured], AR(1) [autoregressive (1)].

6.3. Analysis of secondary objectives

The analyses of the secondary objectives will be performed on the full analysis set.

6.3.1. Concordance between molecular and bacteriological methods

The frequency distribution of sputum samples positive or negative for specific bacterial species (including but not necessarily limited to *H. influenzae*, *M. catarrhalis*, *S. pneumoniae*, *S. aureus* and *P. aeruginosa*) obtained from PCR will be tabulated versus results obtained from culture at any visit (stable scheduled visits or AECOPD visits).

The concordance between culture and PCR at any visit will be evaluated using the Cohen's kappa (k) statistic and 95% CI. Concordance will be interpreted as good (>0.79), substantial (0.60–0.79), moderate (0.40–0.59), fair (0.21–0.39) or no concordance (<0.21) (McHugh, 2012).

The load of *H. influentiae, M. catarrhalis* and S. *pneumoniae* as identified in sputum culture versus RT-PCR, at any visit (stable scheduled visits or AECOPD visits), will be plotted.

The concordance between culture and PCR at any visit will be evaluated using Spearman's rank correlation coefficient. Concordance will be interpreted as good (>0.79), substantial (0.60–0.79), moderate (0.40–0.59), or fair (0.21–0.39) or no concordance (<0.21).

Proportion of patients with sputum samples positive for bacterial pathogens (including but not necessarily limited to *H. influenzae* , *M. catarrhalis*, *S. pneumoniae*, *S. aureus*, *P. aeruginosa*, *K. pneumoniae* and *A. baumannii*) by PCR will be computed, with 95% confidence intervals (CIs), at each and any stable scheduled visit and at each and any AECOPD visit confirming acute exacerbation.

6.3.2. Proportion of potential bacterial/viral pathogens by country, GOLD grade and by AECOPD severity

Average proportion of patients with sputum samples positive for bacterial or viral pathogens overall and for specific species, will be computed, with 95% confidence intervals (CIs), at any stable scheduled visit and at any AECOPD visit confirming acute exacerbation [mild, moderate, severe] by GOLD grade at enrolment [moderate, severe, very severe].

Average proportion of patients with sputum samples positive for bacterial or viral pathogens overall and for specific species, will be computed, with 95% confidence intervals (CIs), at any stable scheduled visit and at any AECOPD visit confirming acute exacerbation by country.

6.3.3. Sputum sample collection and quality

6.3.3.1. Sputum sample collection

The percentage of patients at each scheduled [stable] and exacerbation visit for whom a sputum sample is obtained will be computed overall and by the method the sample is obtained [i.e. spontaneous (at study visit or at patient's home), induced using 0.9% saline or induced using 3% saline]. The percentage of patients with induced sputum will be tabulated by GOLD grade at enrolment and by severity of the AECOPD. The proportion of sputum samples obtained at each confirmed stable or AECOPD visit [any, mild, moderate and severe] with or without induction and positive for specific bacterial\viral pathogens by bacteriological culture and bacterial\viral PCR will be computed.

The distribution of the number of sputum samples will presented according to time (days) since the start of exacerbation. The distribution of the proportion of sputum sample positive for bacteria/virus given the number of days between the start of exacerbation and sample collection will be presented.

The distribution of the number of sputum samples collected on the same day, after 1, 2 or more days since the start of antibiotics administration will be tabulated. The impact of previous antibiotic administration on bacterial results will be evaluated. The proportion of sputum samples obtained at each confirmed stable or AECOPD visit [any, mild, moderate and severe] with previous administration of antibiotics or not and positive for specific bacterial pathogens by bacteriological culture and PCR, respectively will be computed.

The impact of previous vaccination with a pneumococcal vaccine or an influenza vaccine (regardless of time) on *S. pneumoniae* and influenza virus will be evaluated. The proportion of sputum samples obtained at each confirmed stable or AECOPD visit with previous vaccination or not and positive for *S. pneumoniae* by culture and PCR or influenza virus by PCR will be computed.

6.3.3.2. Sputum sample quality

Sputum sample quality will be summarized (squamous cell count, neutrophils cells count and bacteria direct smear) at any stable visit and at any mild, moderate or severe AECOPD visit. Sputum sample characteristics will be also summarized by bacterial/viral presence in any stable visit and in any exacerbation visit.

Sputum quality, derived as detailed in section 11.2.3 in section 11, will be tabulated given the type of visit.

6.3.4. Incidence rates of AECOPD

This study will provide an estimate of average number of exacerbations per person per year, in a COPD population at increased risk of exacerbation (i.e. moderate to very severe COPD patients with a documented history of at least 1 moderate or severe AECOPD in the year prior to enrolment).

To compute the incidence rate all the exacerbations will be taken into account. The observation period is from visit 1 to visit 3 or from visit1 to the last visit perfomed for patients lost to follow up.

6.3.4.1. Overall incidence

The following incidence rates (per patient per year) will be computed, with 95% CIs [using negative binomial regression detailed in section 6.3.4.4]:

- AECOPD events clinically confirmed by the investigator (5 and 6 dash boxes in Figure 1).
- Confirmed exacerbations (AECOPD events plus missed AECOPD visits with medical records: 2, 5, 6 and 7 dash boxes in Figure 1).
- Potential AECOPD with alert confirmed by phone call/or at the study site (2, 3, 5, 6 and 7 dash boxes).

The incidence rate of AECOPD events, AECOPD events plus missed AECOPD visits with medical records and potential AECOPD confirmed by phone call/or at the study site, with 95% CIs, will be estimated according to COPD severity at enrolment based on GOLD grade [moderate, severe, very severe].

The incidence rates of AECOPD events and AECOPD events plus missed AECOPD visits with medical records, with 95% CIs, will be also computed for mild AECOPD, moderate AECOPD and for severe AECOPD.

In addition, the incidence rates of confirmed exacerbations (events and AECOPD events plus missed AECOPD visits with medical records), with 95% CIs, will be estimated according to covariates such as sex, country, age at enrolment (40-49, 50-59, 60-69, ≥70 years), smoking status at enrolment (current and former smokers) and number of exacerbations reported in the 12 months prior to enrolment (one, two and equal or above three).

An attempt to control simultaneously for all the above mentioned variables/factors will be done with multivariable negative binomial regression.

6.3.4.2. Incidence bacteria/viruses associated AECOPD

The following incidence rates (per patient per year) will be computed, with 95% CIs:

- AECOPD events with sputum sample containing bacterial pathogens found by culture or by PCR (overall and by, but not limited to, the following bacterial species: *H. influenzae* (and *H.influenzae* after differentiation form *H. Haemolyticus* by PCR assay), NTHi (after differentiation from typeable *H. influenzae* by PCR assay), *M. catarrhalis*, *S. pneumoniae*, *S. aureus*, and *P. aeruginosa*).
- AECOPD events with sputum sample containing viral pathogens found by PCR (overall and by, but not limited to, the following viral species: RSV, parainfluenza virus, enterovirus/ HRV, metapneumovirus, influenza virus, adenovirus, bocavirus and coronavirus using qualitative PCR and HRV also by quantitative RT-PCR).

These incidence rates, with 95% CIs, will be estimated also according to covariates such as COPD severity, sex, country, age at enrolment, smoking status at enrolment and number of exacerbations reported in the 12 months prior to enrolment.

In addition, incidence rate of mild, moderate and severe exacerbation having sputum containing any bacterial/viral pathogen will be computed overall and by species.

Incidence rate of AECOPD overall and associated to bacteria (by culture/PCR) will be computed also at interim analysis.

6.3.4.3. Analyses of exacerbations in relation to morning e-diaries

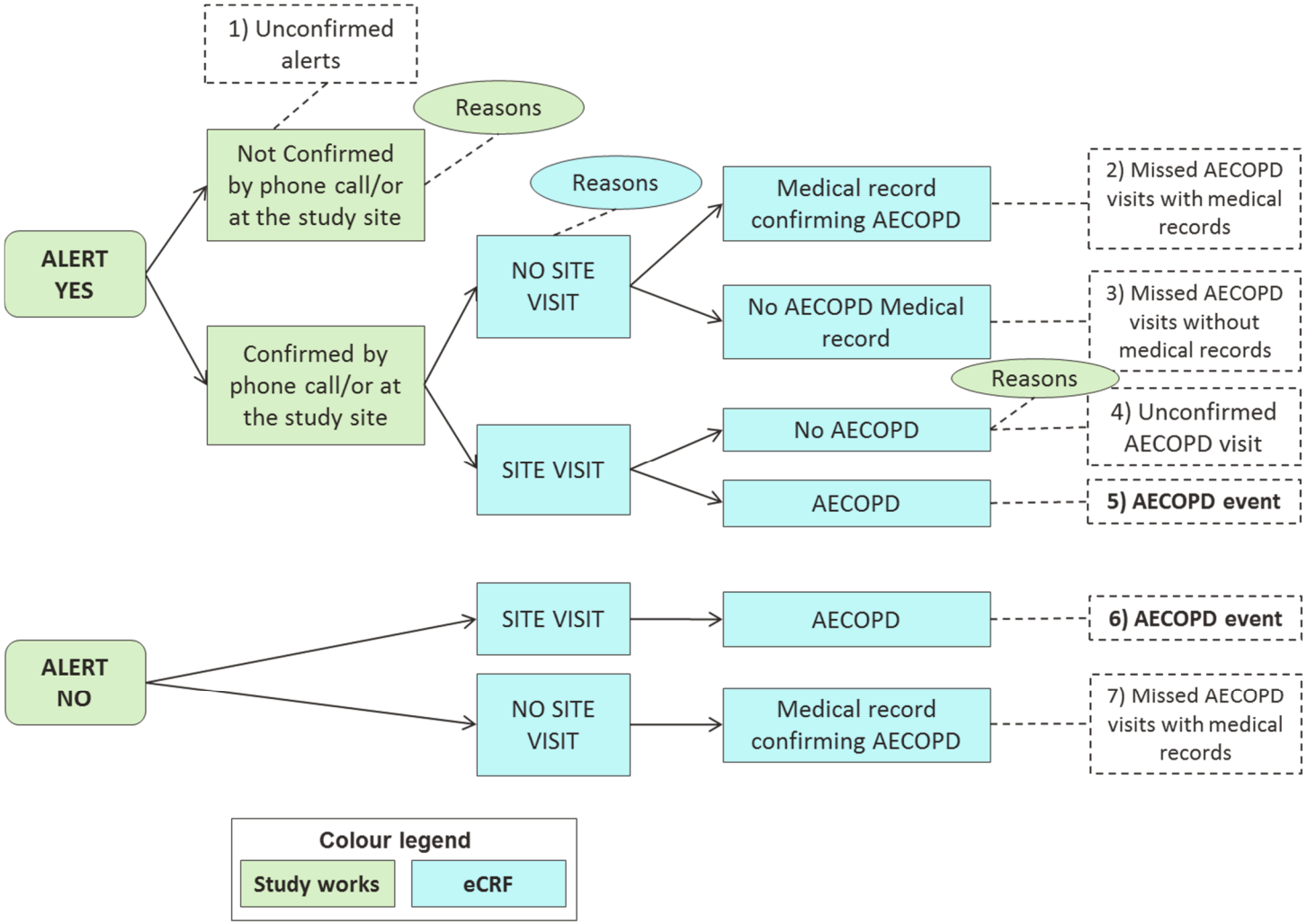
Among morning e-dairy alerts the frequency distribution of the following groups will be tabulated (refer to Figure 1 dashed boxes):

1. Potential event not confirmed by phone call/or at the study site (unconfirmed alerts, box 1),
2. Potential event confirmed by phone call/or at the study site, without site visit, with medical records confirming exacerbation (missed AECOPD visits with medical records, box 2),
3. Potential event confirmed by phone call/or at the study site, without site visit and without medical records (missed AECOPD visits without medical records, box 3).
4. Potential event confirmed by phone call/or at the study site, with site visit but not confirmed by study investigator (unconfirmed AECOPD visit, box 4),
5. Potential event confirmed by phone call/or at the study site, with site visit and confirmed by study investigator (AECOPD event, box 5),

Reasons for inconsistencies between morning e-diary alert and no confirmation by phone call, reasons for missing study visit after phone call confirmation, and reasons for inconsistencies between phone call confirmation and study visit will be tabulated.

In absence of morning diary alert the percentage of spontaneous visits confirming exacerbation or missed AECOPD visit with medical record confirming AECOPD will be tabulated (Figure 1, boxes 6 and 7).

Figure 1 Workflow of study assessment for AECOPD



6.3.4.4. Statistical methods: analysis of count data with overdispersion

The 95% CIs of the incidence rates will be estimated using the Generalised Linear Model (GLM) assuming a negative binomial distribution for the response variable with logarithm as link function and the logarithm of time for follow-up as an offset variable. The fit of the model will be examined using Q-Q plots of standardised residuals.

To overcome convergence issues, the Poisson regression model with adjustment for overdispersion will also be used to estimate the 95% CI of the incidence rate. The fit of the model will be examined using Q-Q plots of standardised residuals. Simulated envelopes will be produced.

SAS codes:

```
proc genmod data = <dataset>;
  class gold age smoking exacHist;
  model n_exac = GOLD age smoking exacHist country
      / link=log dist=negbin offset=logfuinyears ;
  ods output parameterestimates=out_parm nobs=nobs;
  ods exclude modelinfo modelfit parameterestimates nobs;
  run;

  data parm_est(keep= val rate rate_ll rate_ul);
    set out_parm(where=(parameter='intercept'));
    format rate rate_ll rate_ul 8.2;
    val= "mean number per year from negative binomial model without
covariate";
    rate=exp(estimate);
    rate_ll=exp(lowerwaldcl);
    rate_ul=exp(upperwaldcl);
  run;
```

where GOLD = COPD severity, age = age at enrolment (40-49, 50-59, 60-69, ≥70 years), smoking = smoking status at enrolment (current and former smokers), exacHist = number of exacerbations reported in the 12 months prior to enrolment (one, two and equal or above three), country = countries in the study (China [Hong Kong], Philippines, Korea, Taiwan), n_exac = number of exacerbations during the study and logfuinyears = logarithm of the follow-up period in years.

6.3.5. Severity and duration of AECOPD

Total number of exacerbations recorded during the study, exacerbation rate (number of exacerbation per patient-year) and the proportion of COPD patients experiencing a certain number of exacerbations will be presented overall and per bacterial/viral species. The exacerbation rate per year will be reported with and without correction for early withdrawal (see Section 11).

Total number of exacerbations recorded during the study, the mean exacerbation rate per year and the proportion of COPD patients experiencing a certain number of exacerbations will be presented overall and divided by AECOPD severity.

The number and proportion of COPD patients experiencing a certain number of exacerbations will be presented overall and by GOLD grade at enrolment.

The number of patients that report at least 1 AECOPD will be tabulated and descriptive statistics (N, mean, median, SD, minimum, maximum, first and third quartiles) on the average length of AECOPD episodes, as estimated by the investigator, will be presented, for any, mild, moderate and severe AECOPD.

Number of days between 2 consecutive exacerbations, as estimated by the investigator, will be summarized using the same descriptive statistics.

Descriptive statistics of the number of patients according to their percentage of days within an exacerbation period out of the total observation period as estimated by the investigator will be presented. For this analysis, the exacerbations for which the end date was unknown were not excluded and the length was estimated by the investigator as the average length per severity grade.

Results from the chest X-rays during exacerbation visit will be tabulated.

The percentage of scheduled visit in which the status of the patient is stable/recovered, or not recovered will be tabulated.

6.3.5.1. Seasonal distribution

Seasonal distribution of the exacerbations (total number of exacerbations, AECOPD events and of AECOPD events containing NTHi, Mcat, Strep, Entero/HRV, QPCR-Rhino and Influenza A and B) regardless of intensity will be reported by month.

6.3.5.2. Time to exacerbation

To investigate the effect of several baseline risk factors upon the time to first exacerbation since enrolment/first visit, Cox regression models will be fitted including several risk factors such as smoking status at enrolment (active smoker vs. former-smoker), number of exacerbation reported in the 12 months prior to enrolment (one, two and equal or above three), COPD grade at enrolment, age (40-49, 50-59, 60-69, ≥70 years), sex. Time to the first recurrence overall and by covariates will be depicted in Kaplan-Meier survival curves. Log-log survival curves will be displayed as well. If the proportionality of hazards assumption is met the log–log survival curves will be parallel for covariate strata. If there is an intersection between curves the covariate will be added as a stratification factor in the cox regression model.

To take into account that exacerbations may occur more than once over the follow-up time for a given patient, a Cox regression model for recurrent events using the counting process approach (Andersen, 1993) will be adopted (more details in section 6.3.5.2.1). The model will be fitted including smoking status at enrolment, history of exacerbation in the 12 months prior to enrolment, COPD severity at enrolment, age and sex as covariates.

Additional stratification/risk factor may be defined during the analyses.

6.3.5.2.1. Statistical methods: survival analysis

Kaplan-Meier curves

Kaplan-Meier survival curves will be produced to represent the cumulative survival probability over time.

The code follows.

Proc lifetest data = <dataset> method = KM plot = (S, LLS);
time survt*Status(0);
strata smoking;
run;

where survt = survival time in days (days to first exacerbation, for censored data the date of last contact will be used), status = censoring variable (0 = censored, 1 = event), and smoking = smoking status at enrolment (current and former smokers). The PLOTS=(S,LLS) option produces log–log curves as well as survival curves.

Cox Proportional Hazards regression model

To investigate the effect of several baseline risk factors upon the time to first exacerbation, Cox Proportional Hazards (PH) regression models will be fitted.

SAS codes

Proc phreg data = <dataset>;
class GOLD age exacHist sex
model survt*Status(0) = gold age exacHist / ties = exact rl;
strata smoking;
run;

where GOLD = COPD severity, age = age at enrolment (40-49, 50-59, 60-69, ≥70 years), smoking = smoking status at enrolment (current and former smokers), exacHist = number of exacerbations reported in the 12 months prior to enrolment (one, two and equal or above three) and sex (male and female). For the scope of the example, the variable *smoking* is considered to be a stratification factor.

Recurrent event survival analysis

SAS code

Proc phreg data = <dataset> covs(aggregate);
model (start,stop)*event(0) = exacerbationHistory smoker;
id pid;
run;

The code (START,STOP)∗EVENT(0) in the MODEL statement indicates that the time intervals for each observation are defined by the variables START and STOP and that EVENT = 0 denotes a censored observation.

6.3.6. Impact of AECOPD on HRQOL

Descriptive statistics (N, mean, median, SD, minimum, maximum, first and third quartiles) on the CAT and SGRQ-C scores (total score and symptoms, activity and impacts component scores) will be tabulated at each stable scheduled visit, overall and by COPD GOLD grade at enrolment.

CAT and SGRQ-C (total score and symptoms, activity and impacts component scores) changes from first visit to final scheduled visit will be displayed according to the number of exacerbation occurred in the observation period. Correlation will be evaluated using Spearman's rank-order correlation coefficient.

Descriptive statistics (N, mean, median, SD, minimum, maximum, first and third quartiles) on the CAT score will be tabulated for any stable and any mild, moderate or severe exacerbation visit.

6.3.7. Impact of AECOPD on lung function

Descriptive statistics (N, mean, median, standard deviation, minimum and maximum) on post-bronchodilator FEV_1% of predicted normal value and PEF will be tabulated at enrolment and final visit. Change from baseline will be also summarized overall and represented according to the number of exacerbations. Changes in FEV_1% will be evaluated according to the number of exacerbations occurred in the observation period using Spearman's rank-order correlation coefficient.

6.3.8. Impact of AECOPD on healthcare utilisation

Descriptive statistics of the daily number of healthcare utilisation (number of HCU in the relevant period divided by the number of days) within the stable period and within the exacerbation period will be presented overall and divided by the following unscheduled visits:

1. Number of physician office consultations,
2. Number of visits to urgent care,
3. Number of visits to emergency department,
4. Number of hospitalizations.

Healthcare use for each patient will be obtained through review of the patient's medical record (aided by patient daily self-reporting in the eDiary).

Healthcare utilisation will be summarized according to COPD grade at enrolment, smoking status, sex and age at enrolment. In addition, healthcare utilisation will be summarized by AECOPD severity during exacerbations.

Overall HCU [total number and divided by type of HCU] per patient will be correlated to the number of exacerbation experienced during the observation time.

Current medication for COPD and additional COPD treatments prescribed by primary and secondary care physicians within the stable period and within the exacerbation period will be described by drug category overall and by COPD grade at enrolment, smoking status, sex and age at enrolment. In addition, current medication for COPD and additional COPD treatments prescribed will be summarized by AECOPD severity during exacerbations.

6.4. Analysis of tertiary objective

The analyses of the tertiary objectives will be performed on the full analysis set.

6.4.1. End date of AECOPD using EXACT-PRO

Descriptive statistics (N, mean, median, standard deviation, minimum and maximum, first and third quartiles) on the average length of AECOPD episodes, as estimated with the EXACT-PRO (section 12), will be presented, for any, mild, moderate and severe AECOPD.

Descriptive statistics of the number of patients according to their percentage of days within exacerbation periods out of the total observation period as estimated with the EXACT-PRO will be presented.

The frequency distribution of exacerbation confirmed at site visit or through medical records (with or without morning symptoms diary alert) will be tabulated versus EXACT-PRO exacerbations.

Descriptive statistics (N, mean, median, SD, minimum, maximum, first and third quartiles) for the EXACT daily scores will be tabulated at baseline, during the day before exacerbation, at exacerbation onset and at recovery according to investigator judgement. These statistics will be reported overall and for any mild, moderate or severe exacerbation visit.

Graphs detailing the EXACT-PRO scores before during and after exacerbation onset as defined clinically will be also presented.

Difference from baseline, with 95% CI, of EXACT score measured at stable scheduled visit will be plotted.

6.4.2. Mixed infections

The simultaneous bacterial [culture/PCR] and viral presence in sputum at enrolment, any stable and any exacerbation visit [mild, moderate, severe] will be computed overall and by pathogens.

6.4.3. Load of bacteria and virus

Total number of patients providing a sputum sample, the number of sputum samples positive to specific bacteria (by culture) and the frequencies of semi-quantitative bacteriological load (few scattered, +, ++, +++) will be tabulated at each scheduled [stable] visit and at each exacerbation visit.

Total number of patients providing a sputum sample, the number of sputum sample positive to specific bacteria (by PCR)/HRV and average load will be tabulated at each scheduled [stable] visit and at each exacerbation visit.

6.4.4. Impact of occurrence of bacterial/viral pathogens in sputum on AECOPD

The following proportions will be computed:

- Proportion of patients with sputum samples which are positive for specific pathogens [NTHi, Mcat, S. pneumoniae, HRV] during an AECOPD visit but not in the previous stable visit/in the previous visit
- Proportion of patients with sputum samples which are negative for specific pathogens [NTHi, Mcat, *S. pneumoniae*, HRV] during an AECOPD visit but not in the previous stable visit/in the previous visit
- Proportion of patients with sputum samples which are positive for specific pathogens [NTHi, Mcat, *S. pneumoniae*, HRV] during an AECOPD visit and also in the previous stable visit/in the previous visit
- Proportion of patients with sputum samples which are negative for specific pathogens [NTHi, Mcat, *S. pneumoniae*, HRV] during an AECOPD visit and also in the previous stable visit/in the previous visit

Conditional logistic regression models, stratified by patient, will be fitted to estimate the effect of the presence of bacterial/viral pathogens in sputum on the odds of experiencing an acute exacerbation rather than remain in a stable COPD state. Separate models considering bacteria detected by bacteriological culture and by PCR will be fitted. Both the impact of presence and new occurrence of each specific pathogen will be considered. New occurrence of a pathogen will be defined as the detection of the specific pathogen after negative sputum sample at previous [any/stable] visit. Seasonality will be taken into account as a dichotomous variable: high season (October to March) and low season (April to September). Study country will be also taken in consideration. Backward selection of variables will be adopted until all remaining variables have individual P values smaller than 0.10.

Complete models with interactions between NTHi, Mcat, *S. pneumoniae*, HRV and season will be also fitted.

The conditional logistic model allows estimating the impact of independent variables from observations from each individual patient. In other words, by stratifying by patient, there is no need to adjust for time-invariant covariates such as age and sex.

6.4.5. Microbiome analysis

Further characterization of *H. influenzae, M. catarrhalis* and HRV strains will be performed. Details of the statistical analyses will be released in a separate additional analysis request.

6.4.6. Biomarkers

Descriptive statistics (N, mean, median, SD, minimum, maximum, first and third quartiles) for the level of each quantitative biomarker/haematology parameter will be tabulated for visit 1 and visit 3. The difference in sample mean will be reported together with a confidence interval.

Number and percentage of patients with fibrinogen, human serum C reactive protein (hsPCT), C-X-C Motif Chemokine Ligand 10 (CXCL10) values outside the laboratory normal ranges will be tabulated for visit 1 and visit 3.

Descriptive statistics will be presented overall and by GOLD grade at enrolment.

Changes in biomarkers will be evaluated according to the number of exacerbations occurred in the observation period using Pearson’s correlation coefficient (or Spearman's rank-order correlation coefficient if the normal distribution assumption is not met).

6.5. Analysis of safety

The analysis of serious adverse events will be performed on the full analysis set.

Adverse events (AEs) and serious adverse events (SAEs) considered possibly related to the study participation and withdrawal due to AEs and SAEs during the entire study period will be described in detail.

7. ANALYSIS INTERPRETATION

All analyses are descriptive. The use of these descriptive analyses should be limited to supportive analysis of confirmatory analyses or hypothesis generation.

8. CONDUCT OF ANALYSES

8.1. Sequence of analyses

The analyses will be performed in two steps:

- An interim analysis will be performed when at least 40 AECOPD evaluable sputum samples are available. It will include all data available at the time of analysis. The rate of positive samples will be computed for bacterial pathogen (including Hi, NTHi and Mcat) together and per pathogen. The incidence rate of AECOPD overall and associated to bacterial pathogen will be computed as well. The analysis will be done based on as clean as possible data. No study report will be written at this stage.
- A final analysis of all objectives will be performed when data obtained up to planned last patient last visit are available. A final study report will be written at this stage.

Description	Analysis ID	Disclosure Purpose (CTRS=public posting, SR=study report, internal)	Dry run review needed (Y/N)	*Study Headline Summary (SHS)*requiring expedited communication to upper management (Yes/No)	Reference for TFL
Final analysis	E1_01	Study report	YES	NO	See TFL TOC
Interim analysis	E1_02	Internal	NO	NO	See TFL TOC

8.2. Statistical considerations for interim analyses

All analyses are descriptive and therefore no statistical adjustment for the interim analysis is required. No study report will be written for interim analysis.

9. CHANGES FROM PLANNED ANALYSES

Not applicable

10. LIST OF FINAL REPORT TABLES, LISTINGS AND FIGURES

The TFL TOC provides the list of tables/listings and figures needed for the study report. It also identifies the tables eligible for each analyses and their role (synopsis, in-text, post-text, SHS, CTRS,…). Note that all TFL aimed to be included as post-text are noted as post-text even if these are tabulation of individual data such as listing of SAE. The post-text material contain all source material for the study report and accordingly a post-text table may be redundant with an in-text table.

11. ANNEX 1 DATA DERIVATION RULE AND STATISTICAL METHODS

11.1. Statistical Method References

Andersen P.K., Borgan O., Gill R.D., and Keiding N. Statistical Models Based on Counting Processes. New York: Springer-Verlag 1993.

Atkinson AC. Plots, Transformations and Regression. Clarendon Press. 1985

Burton P, Gurrin L, Sly P. Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level mixed modelling. *Stat Med*. 1998;17:1261–91.

Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of binomial. *Biometrika*. 1934;26:404-413

Liang KY, and Zeger SL. Longitudinal Data Analysis Using Generalized Linear Models." Biometrika. 1986; 73:13–22.

McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med*. 2012;22(3):276-82.

11.2. Data derivation

11.2.1. COPD severity stage

The baseline COPD severity stage will be derived by the investigator from the spirometry results at the screening visit. The spirometric classification of airflow limitation in COPD patients is based on post-bronchodilator FEV_1 and can be divided into four GOLD grades (mild, moderate, severe and very severe) [Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of Chronic Obstructive Pulmonary Disease, updated 2013. Available from: http://www.goldcopd.org/uploads/users/files/GOLD_Report_2013_Feb20.pdf]. For the purpose of this study only patients with COPD graded from moderate to very severe will be enrolled.

Table 1 COPD severity stages included in the study

GOLD grade	In patients with $FEV_1/FVC<0.70$
Moderate	$50\% \leq FEV_1 < 80\%$ predicted
Severe	$30\% \leq FEV_1 < 50\%$ predicted
Very severe	$FEV_1 < 30\%$ predicted

The resulting COPD severity variable will be used in all models in which COPD at baseline will be taken into account.

A four-grade refined assessment of COPD severity stages will be computed according to the newest GOLD criteria [GOLD, 2017 Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of Chronic Obstructive Pulmonary Disease, updated 2017. Available from: http://goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/]. In the assessment scheme, patients should undergo spirometry to determine the severity of airflow limitation (i.e., spirometric grade). They should then undergo assessment of either dyspnea using mMRC or symptoms using CAT. Finally, their history of exacerbations (including prior hospitalizations) should be recorded.

For the purpose of this study we will only adopt CAT rather than adopting mMRC dyspnea scale. This choice is because the two tools are not interchangeable [Kim S, Oh J, Kim YI, Ban HJ, Kwon YS, Oh IJ, Kim KS, Kim YC, Lim SC. Differences in classification of COPD group using COPD assessment test (CAT) or modified Medical Research Council (mMRC) dyspnea scores: a cross-sectional analyses. *BMC Pulm Med*. 2013;13:35] and it is now recognized that COPD impacts patients beyond just dyspnea [Jones PW. Health status measurement in chronic obstructive pulmonary disease. *Thorax.* 2001; 56(11):880-7].

The four grades are:

- Patient Group A – Low Risk, Less Symptoms Typically GOLD 1 or GOLD 2 (Mild or Moderate airflow limitation); and 0-1 exacerbation per year and no hospitalization for exacerbation; and CAT score < 10;
- Patient Group B – Low Risk, More Symptoms Typically GOLD 1 or GOLD 2 (Mild or Moderate airflow limitation); and 0-1 exacerbation per year and no hospitalization for exacerbation; and CAT score ≥ 10;
- Patient Group C – High Risk, Less Symptoms Typically GOLD 3 or GOLD 4 (Severe or Very Severe airflow limitation); and ≥ 2 exacerbations per year or ≥ 1 with hospitalization for exacerbation; and CAT score < 10;
- Patient Group D – High Risk, More Symptoms Typically GOLD 3 or GOLD 4 (Severe or Very Severe airflow limitation); and ≥ 2 exacerbations per year or ≥ 1 with hospitalization for exacerbation; and CAT score ≥ 10.

Spirometry results for GOLD classifications, number of exacerbations per year, and number of exacerbation requiring hospitalization will be taken from the screening visit (visit 0). CAT score will be derived from visit 1.

11.2.2. HRQOL

11.2.2.1. CAT

The CAT index will be derived as the sum of the ratings recorded for each of the eight individual items. Each of these items has 6 possible scores (0, 1, 2, 3, 4 or 5), leading to a range of 0 to 40 for CAT score.

No missing items are expected because the e-diary system prompts the patient to compete an item of CAT to proceed to the following one. Partial questionnaires will be excluded.

11.2.2.2. SGRQ-C

The SGRQ-C total score will be derived as the weighted sum of the forty individual items leading to a range of 0 to 100 as detailed in the reference manual [St George's Respiratory Questionnaire for COPD patients (SGRQ-C), version 1.3, 2016].

The SGRQ-C symptoms, activity and impacts component scores will be derived as the weighted sum of subset of items as detailed in the reference manual.

No missing missing items are expected because the e-diary system prompts the patient to compete an item of SGRQ-C to proceed to the following one. Partial questionnaires will be excluded.

11.2.3. Sputum Quality

The quality of sputum is defined as follow:

- < 10 squamous epithelial cells/field: sample of good quality
- 10-25 squamous epithelial cells/field with significant numbers of cells derived from the lower respiratory tractus (ciliary epithelial cells, bronchial cells, round cells): sample of moderate quality
- > 25 squamous epithelial cells/field: sample of bad quality

11.2.4. Bacterial and viral PCR assays

- Multiplex real-time PCR (MPB1) assay will target *Haemophilus influenzae, Moraxella catarrhalis, Streptococcus pneumoniae* with quantitative results (copies/mL of sample).

For qualitative analyses, quantitative results must be converted in POS/NEG results. Only samples above the cut off will be considered as POSITIVE.

The positivity cut-off for each quantitative PCR assay will be defined at time of data release.

For quantitative analyses, only samples with load above the cut off will contribute to mean and SD computation.

- Multiplex real-*time* PCR (MPB2) assay will target *Streptococcus pyogenes, Staphylococcus aureus, Pseudomonas aeruginosa* with qualitative results only (POS/NEG).

Multiplex reverse transcription PCR assay will target different viral pathogens with qualitative results only (POS/NEG).

HRV qPCR assay will target HRV with quantitative results (copies/mL).

For qualitative analyses, qualitative and quantitative results will be combined with the following algorithm:

Sample	Allplex HRV result	Allplex HEV result	HRV RT-qPCR	Final HRV result (variable to be derived)
1	POS	NEG	≥ LOD*	POS
2	POS	NEG	< LOD	POS
3	NEG	POS	≥ LOD	POS
4	NEG	POS	< LOD	NEG
5	NEG	NEG	Not tested	NEG
6	POS	NEG	0	POS
7	POS	POS	≥ LOD	POS

*The positivity cut-off for each quantitative PCR assay will be defined at time of data release.For quantitative analyses, only positive samples will contribute to mean and SD computation.

11.3. Handling missing data

Missing or non-evaluable measurements will not be replaced. The only exception is detailed below.

The number of exacerbations per year will be extrapolated for patients withdrawing from the study to provide an estimate of the number of exacerbations over the one year observation period. The number of exacerbations in a year will be calculated by multiplying the number of exacerbations experienced by the patient by 13 and dividing by the number of 4-week periods the patient was followed up [Stockley R.A., Chopra N, Rice L. Addition of salmeterol to existing treatment in patients with COPD: a 12 month study. *Thorax.* 2006;61:122-128].

$$\text{Number of exacerbations per year} = \frac{\text{Number of exacerbations} * 13}{\text{Number of 4-week observation period intervals}}$$

The calculation of exacerbation rate will be based on follow-up period intervals of four weeks to avoid obtaining high imputed rates if a patient withdrew very early from the study after experiencing an exacerbation. Four-week intervals will be adopted since treatment courses for moderate/severe exacerbations are <=2 - 4 weeks when appropriate.

11.3.1. Date derivation

Partial dates will not be considered as missing or not evaluable date. In case day is missing, 15th of month is used. In case day and month are missing, 30th of June is used.

Partial or missing initial and end dates for exacerbation will not be considered to measure the length of an exacerbation.

11.3.2. Missing data for electronic Diary Card alert

- Days with missing data will not be ignored when defining start of an e-diary signal: e.g., if a patient has two qualifying symptoms followed by a missing day, if the next day with data has two qualifying symptoms this will not be defined as an exacerbation. Hence, the start of an e-diary signal of an exacerbation will require 2 consecutive non-missing days of worsening. If there is partially recorded symptom data at a visit, however, only the observed values will be considered but the day will not be discarded.
- Days with missing data will be ignored when defining end of an e-diary signal of an exacerbation: e.g., if a patient has two days missing in the baseline period and one day missing in the 3- day moving average, both baseline and 3-day moving average values will be calculated using observed data only. Similar rules will apply when there is partially recorded symptom data at a visit (and daily average values will be calculated among the observed values).
- All available data will be used for each patient.
- As a general rule, the baseline symptom count score, and 3 day moving averages are calculated as the average count score of the observed days in the baseline period. For example, if an e-diary signal occurs from day 10 onwards, the baseline period will be the average between day1 and day2 (day1 being the first study day). If day 13 is missing, the 3 day moving average at day 14 will only take into account day 14 and day 15.
- If there are more than 2 consecutive days of missing data after an e-diary signal not resolved, the e-diary signal will not be considered resolved until the next day with observed data if the corresponding moving average is below the baseline value. In order words, if the 3-day average daily score is missing, the day will be counted in the duration of the episode.

11.4. Number of decimals

The following decimal description will be used for the analyses.

Parameters	Number of decimal digits
% of count, including LL & UL of CI	1
p-value	3
Mean, median	Number of decimals in the raw data + 1
SD	Number of decimals in the raw data + 2
Minimum, maximum, range	Number of decimals in the raw data
First and third quartiles	Number of decimals in the raw data + 1

LL = Lower Limit UL = Upper Limit CI = Confidence Interval
SD = Standard deviation

11.5. Methodology for computing CI

- All CI computed will be two-sided 95% CI.
- The exact 95% CIs for a proportion within a group will be calculated [Clopper 1934].

12. ANNEX 2: AECOPD DEFINITION AND CALCULATION RULE

12.1. Definition, onset, recovery, duration of an AECOPD event

AECOPD events are confirmed according to investigator judgment after an AECOPD visit which is aimed to exclude worsening in symptoms not related to an AECOPD event.

AECOPD visits are scheduled after an electronic Diary Card alert confirmed by the investigator (by phone call or at the study site) or after spontaneous site visits due to worsening symptoms without any alert.

Electronic Diary Card alerts refer to daily symptoms recorded the morning after (also called morning symptoms). Alerts are based on the Anthonisen criteria:

- Worsening of two or more of the following major symptoms for at least two consecutive days*,**: dyspnea, sputum volume, sputum purulence (colour)

Or

- Worsening of any major symptom together with any of the following minor symptoms for at least two consecutive days*,**: sore throat, colds (nasal discharge and/or nasal congestion), fever (oral temperature $\geq 37.5°C$) without other cause, increased cough, increased wheeze.

** The same two symptoms do not have to be present on both days as long as at least one major symptom is present on both days.*

AECOPD onset will be defined as the first day of the two consecutive days of worsening symptoms.

AECOPD recovery will be determined/confirmed by the investigator/delegate during (a) follow-up phone call(s) which will take place every 2 weeks until the AECOPD has resolved. The end date will be based on when the investigator/delegate determines that the AECOPD symptoms have resolved. In determining this end date, consideration will be given to symptoms recorded in the electronic Diary Card and patient assessment during the phone calls.

AECOPD duration will be defined as the number of days from AECOPD onset (included) and AECOPD recovery (not included).

Duration = date2 – date1 + 1,

with date2: AECOPD recovery date,

date1: AECOPD onset date

12.2. Recovery and duration of an exacerbation according to EXACT-PRO:

An EXACT-PRO AECOPD event is defined as a 12 point increase above baseline for 2 consecutive days, or a nine point increase for 3 consecutive days.

The baseline EXACT-PRO score is calculated as the average daily EXACT-PRO score over days 14 to 8 preceding AECOPD symptomatic onset.

AECOPD recovery according to EXACT-PRO corresponds to the day at which the EXACT-PRO score return to its baseline value (or below) [Mackay AJ, Donaldson GC, Patel AR, Singh R, Kowlessar B, Wedzicha JA. Detection and severity grading of COPD exacerbations using the exacerbations of chronic pulmonary disease tool (EXACT). *Eur Respir J*. 2014;43(3):735-44.].

The AECOPD duration according to EXACT-PRO will be defined as the number of days from AECOPD onset (included) and AECOPD recovery according to EXACT-PRO (not included).

12.2.1. EXACT-PRO baseline score at enrolment

An EXACT-PRO baseline score at enrolment will be computed as the mean within-patient score over 7 days, with data present for a minimum of 4 of the 7 days. If fewer than 4 days of data are available, the EXACT baseline score cannot be calculated.

12.3. Time between 2 *e-diary* signals for exacerbations/evaluation of baseline symptom score:

The minimum duration between 2 e-diary signals with distinct baseline value is 15 days, which is the minimal period for which a new full exacerbation-free baseline can be calculated for the next e-diary episode.

If a new worsening of symptoms occurs between 6 and 14 days after the end of a previous exacerbation, the baseline period used will be the same baseline as the one of the original exacerbation. Therefore, it will last until the daily symptom count score has gone below the baseline symptom score of the previous event.

12.4. Definition of Missed AECOPD visits

Missed AECOPD visits with medical records (dashed boxes 2 and 7 in Figure 1) will capture:

- an alert from the electronic Diary Card without a site visit but that cannot be discounted as a possible exacerbation signal after reconciliation (detailed in the Study Procedure Manual),

and/or

- an exacerbation reported retrospectively during a scheduled visit clinically documented for which there has not been a corresponding exacerbation visit.

Onset date, recovery date and severity will be retrieved by medical records.

For the analyses not involving the collection of any biological specimen, exacerbations will refer to AECOPD events plus missed AECOPD visits with medical records.

Missed AECOPD visits without medical records (dashed box 3 in Figure 1) will capture:

- an alert from the electronic Diary Card, assessed AECOPD during the phone call, without a site visit and without AECOPD medical records.

12.5. Unconfirmed AECOPD event with morning e-diary signal alert notification.

An alert from the electronic Diary Card not confirmed to be an AECOPD after contact by phone call or at the study site.

12.6. Grading of severity of an exacerbation

Severity of exacerbations is defined as per protocol:

- Mild: Worsening symptoms of COPD that are self-managed by the patient.
- Moderate: Worsening symptoms of COPD that require treatment with oral corticosteroids and/or antibiotics.
- Severe: Worsening symptoms of COPD that require treatment with in-patient hospitalisation or home care intervention.

Severity of the exacerbation will not be derived but taken directly from the information entered in the CRF (i.e. Severity of exacerbation will be taken from the conclusion of the exacerbation visit).

13. ANNEX 3: STUDY SPECIFIC MOCK TFL

Study specific mock TFL is provided in a separate document.

GlaxoSmithKline	**Statistical Analysis Plan**
Detailed Title:	A prospective, epidemiological, multi-country, cohort study to assess the occurrence of potential bacterial and viral pathogens in stable chronic obstructive pulmonary disease (COPD) and during acute exacerbations of COPD (AECOPD), in moderate to very severe COPD patients, in Asia Pacific.
eTrack study number and Abbreviated Title	201112 (EPI-NTHI-001 BOD APA)
Scope:	All data pertaining to the above study.
Date of Statistical Analysis Plan	Final version: 28-Jul-2017
Co-ordinating author:	PPD , Biostatistician in Epidemiology
Reviewed by:	PPD , Lead Epidemiologist PPD , Lead Statistician in Epidemiology PPD , Lead Statistical Analyst PPD , Statistical Analyst PPD Scientific writer PPD , Regulatory Affair PPD , SERM physician PPD , Public Disclosure Representative
Approved by:	PPD , Clinical and Epidemiology Project Lead delegating to PPD , Lead Epidemiologist PPD , Lead Statistician in Epidemiology PPD , delegating to PPD , Scientific Writer PPD , Lead Statistical Analyst

APP 9000058193 Statistical Analysis Plan Template (Effective date: 14 April 2017)

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

LIST OF ABBREVIATIONS

A. baumannii	*Acinetobacter baumannii*
AE	Adverse event
AECOPD	Acute exacerbation of COPD
CAT	COPD assessment test
CI	Confidence Interval
COPD	Chronic Obstructive Pulmonary Disease
CRF	Case Report Form
EXACT-PRO	Exacerbations of Chronic Pulmonary Disease Tool - Patient Reported Outcome
FAS	Full Analysis Set
FEV_1	Forced expiratory volume in 1 second
GOLD	Global Initiative for Chronic Obstructive Lung Disease
GSK	GlaxoSmithKline
H. haemolyticus	*Haemophilus haemolyticus*
H. influenzae	*Haemophilus influenzae*
HRQOL	Health-related quality of life
HRV	Human rhinoviruses
I1	Interim analysis 1
K. pneumoniae	*Klebsiella pneumoniae*
LL	Lower Limit of the confidence interval
M. catarrhalis	*Moraxella catarrhalis*
MA	Main Analyses
MedDRA	Medical Dictionary for Regulatory Activities
NTHi	Non-Typeable *Haemophilus influenzae*
P. aeruginosa	*Pseudomonas aeruginosa*

PCR	Polymerase chain reaction
PEF	Peak Expiratory Flow
RSV	Respiratoir Syncytiaal Virus
RT-PCR	Real-time polymerase chain reaction
SAE	Serious adverse event
S. aureus	*Staphylococcus aureus*
SAP	Statistical Analysis Plan
SD	Standard Deviation
SGRQ-C	St. George's Respiratory Questionnaire for COPD patients
S. pneumoniae	*Streptococcus pneumoniae*
SR	Study Report
TFL	Tables Figures and Listing template annexed to SAP
UL	Upper Limit of the confidence interval

1. DOCUMENT HISTORY

Date	Description	Protocol Version
28-JUL-2017	Version 1	Protocol Amendment 1 13-FEB-2017

2. STUDY DESIGN

Type of design: epidemiological, prospective, interventional, multi-country, cohort study. The study targets enrolling approximately 200 stable moderate to very severe COPD patients, males and females, aged 40 years or older at the time of enrolment, with at least 1 documented moderate or severe AECOPD in the year before enrolment.

This study will be conducted in several sites among several countries in Asia Pacific [China (Hong Kong), Philippines, Korea, Taiwan].

Type of study: self-contained.

Duration of the study: approximately 1 year for each patient.

Study visits/contacts:

- Screening visit
- Three scheduled visits occurring at 6 months intervals.
- For each AECOPD: AECOPD visit (within 96 hours of the onset of the symptoms) and follow-up phone call(s) (at least every 2 weeks until the AECOPD has resolved). Follow-up phone contacts will define end of AECOPD.

The following group names will be used for the statistical analyses:

Group order in tables	Group label in tables	Group definition for footnote
1	COPD	Patients with moderate to very severe COPD

3. OBJECTIVES

3.1. Primary objective

To estimate the proportion of potential bacterial and viral pathogens (overall and by species) detected in sputum of stable COPD patients and during AECOPD, using respectively bacteriological methods and viral PCR, over the course of 1 year.

3.2. Secondary objectives

- To evaluate the concordance between PCR and bacteriological methods data in sputum for potential bacterial pathogens.
- To estimate the proportion of potential bacterial and viral pathogens (overall and by species) detected in sputum of stable COPD patients, by GOLD grade.
- To estimate the proportion of potential bacterial and viral pathogens (overall and by species) detected in sputum during AECOPD, by severity of AECOPD.
- To estimate the incidence rate of all-cause AECOPD, overall and by GOLD grade.
- To describe the severity and duration of all-cause AECOPD, overall and by GOLD grade.
- To assess the impact of all-cause AECOPD on HRQOL.
- To assess the impact of all-cause AECOPD on lung function.
- To assess the impact of all-cause AECOPD on healthcare utilisation.

3.3. Tertiary objectives

- To assess the use of EXACT-PRO for determining the end date of AECOPD.
- To evaluate the load of bacterial and viral pathogens in stable COPD and during AECOPD.
- To collect biological specimens for future respiratory disease-related testing:
  - Aliquots of sputum samples for assay development and microbiome analysis at stable visits and during exacerbation.
  - Blood sampling for potential biomarkers for identification/quantification of biomarkers at Visit 1 and Visit 3.

4. ENDPOINTS

4.1. Primary endpoint

- Occurrence of potential bacterial and viral pathogens in sputum of stable COPD patients and during AECOPD, over the course of 1 year:
  - Bacterial pathogens, as identified by bacteriological methods, including (but not necessarily limited to) *H. influenzae*, *M. catarrhalis*, *S. pneumoniae*, *S. aureus*, *P. aeruginosa*, *K. pneumoniae* and *A. baumannii*.
  - Viral pathogens, as identified by PCR, including (but not necessarily limited to) RSV, parainfluenza virus, enterovirus/ rhinovirus, metapneumovirus, influenza virus, adenovirus, bocavirus and coronavirus and by rhinovirus quantitative RT-PCR.

4.2. Secondary endpoints

- Occurrence of potential bacterial pathogens in sputum of stable COPD patients and during AECOPD, as measured by real-time qualitative PCR/ quantitative PCR and compared to data from bacteriological methods, over the course of 1 year:
  - Including (but not necessarily limited to) *H. influenzae, M. catarrhalis, S. pneumoniae, S. aureus* and *P. aeruginosa.*
- Occurrence of potential bacterial and viral pathogens (overall and by species) in sputum of stable COPD patients by GOLD grade, over the course of 1 year.
- Occurrence of potential bacterial and viral pathogens (overall and by species) in sputum during AECOPD by severity of AECOPD, over the course of 1 year.
- Incidence rate (per patient per year) of AECOPD, overall and by GOLD grade, over the course of 1 year.
- Severity of AECOPD, overall and by GOLD grade, over the course of 1 year.
- Duration of AECOPD, overall and by AECOPD severity, over the course of 1 year.
- CAT score in stable COPD patients and during AECOPD, over the course of 1 year.
- SGRQ-C score in stable COPD patients, over the course of 1 year.
- FEV_1% of predicted normal value in stable COPD patients, at Pre-Month 0 and Month 12.
- Healthcare utilisation, over the course of 1 year.

4.3. Tertiary endpoints

- EXACT-PRO scores in stable COPD patients and during AECOPD, over the course of 1 year.
- Bacterial load measured by both culture and PCR in COPD and during AECOPD.

- Viral load measured by PCR in COPD and during AECOPD.

5. ANALYSIS SETS

The following data sets will be defined.

5.1. All screened set

The all screened set will include all screened patients.

5.2. All enrolled set

The all enrolled set will include all screened patients except for screening failures.

5.3. Full Analysis Set

Since screening visit is conducted before the first scheduled stable visit (ideally it will occur not more than 6 weeks before visit 1), a full analysis set will be defined including all enrolled patients except for those who discontinued the study at visit 1 (according to protocol: for serious adverse event, protocol violation, consent withdrawal, patients not willing to participate, moved from the study area, lost to follow-up, sponsor study termination, other).

5.3.1. Change in analysis population from protocol amendment 1

A Full Analysis Set will be defined as the primary analysis population instead of the Per Protocol cohort as mentioned in the Protocol.

The Full Analysis Set will include all enrolled patients that fulfil eligible criteria at visit 1 (regardless if they provide a sputum sample) and do not withdraw their consent between screening visit and visit 1.

This choice is to follow ICH E9 recommendations. Furthermore, in observational studies describing the burden of disease, given the high number of objectives and the absence of adherence to treatment; it is not possible to define a single or few per protocol sets.

5.4. Criteria for eliminating data from the Analysis Sets

Elimination codes are used to identify patients to be eliminated from analysis. Detail is provided below for each set.

5.4.1. Elimination from All Screened Set

Code 900 (invalid informed consent or fraud data) will be used for identifying patients eliminated from All Screened Set.

5.4.2. Elimination from All Enrolled Set

Code 900 (invalid informed consent or fraud data) and code 2010 (Protocol violation (inclusion/exclusion criteria) will be used for identifying patients eliminated from All Enrolled Set.

5.4.3. Elimination from the Full Analysis Sets (FAS)

Code 900 (invalid informed consent or fraud data) and code 2010 (Protocol violation (inclusion/exclusion criteria) will be used for identifying patients eliminated from FAS.

5.4.4. Elimination from the analyses

A list of reasons for elimination from the analyses (interim analysis and final analysis), is reported below:

Code	Condition under which the code is used
900	Invalid informed consent or fraud data
2010	Protocol violation (inclusion/exclusion criteria)
2040	Antibiotics taken on a continual basis, defined as more than 1 month in total)
2050	Not withdrawn after developing withdrawal criteria (e.g. pregnancy)
2110	Missed or incomplete spirometry, missed blood collection, sputum collection or sputum culture
2120	Obvious incoherence or abnormality or error in data (e.g. spirometry not properly performed)
2140	Missing entries in CAT or SGRQ-C
3000	Missed scheduled visit, AECOPD visit or initial phone contact after potential AECOPD alert
3010	Out of window for scheduled visit or AECOPD visit

Codes from 2040 to 3010 are protocol deviations not leading to elimination from the population sets above mentioned.

6. STATISTICAL ANALYSES

The total number and the distribution of the following data sets overall and among the study country will be tabulated:

- All screened set and screening failures.
- All enrolled set.
- The full analysis set.
- Patients completing the study and withdrawn patients.

In addition, screening failures will be summarized according to the reason for failure.

Withdrawal status will be summarised according to the reason for withdrawal. The number of withdrawn patients will be tabulated by study visit and overall.

6.1. Analysis of demographics/baseline characteristics

The analysis of demographic/baseline characteristics will be performed on the full analysis set and on the all enrolled set only if the percentage of enrolled patients excluded from the FAS is 5% or more.

The following characteristics at enrolment will be summarized using descriptive statistics:

- Demographics such as age at enrolment, sex and race.
- Present and past medical history.
- History of AECOPD within the previous year (overall and by AECOPD severity).
- Pre- and post-bronchodilator spirometry results.
- COPD status using the GOLD grade.
- Smoking/biomass exposure history status.
- Selected results and several risk factors from the ATS-DLD-78-A questionnaire.
- Physical examination/vital signs results such as height, weight, BMI, heart rate, respiratory rate, systolic and diastolic blood pressure.
- Chest X-rays results.
- Intercurrent comorbidities (reported at visit 1)
- History of pneumococcal and influenza vaccination (reported at visit 1)

The following descriptive statistics will be adopted:

- The frequency and the percentage of each category will be reported for categorical variable.
- Number of non-missing observations, mean, median, standard deviation, minimum and maximum will be provided for quantitative variable.

All tables will be also generated by country.

Demographics and baseline characteristics will be also summarized for study sample for interim analysis.

6.2. Analysis of primary objective

The analysis of the primary objective will be performed on the FAS.

Average proportion of patients with sputum samples positive for bacterial or viral pathogens overall and for specific species, and 95% CIs will be computed at any stable scheduled visit and at any exacerbation visit of the entire study period. Confidence intervals will be estimated using a Generalized Estimating Equations (GEE) model assuming a binomial distribution for the response variable with logit as link function and a compound symmetry correlation matrix (exchangeable structure) to account for the within-patient correlations (Liang,. 1986) (more details in section 6.2.1). The crude proportions (i.e. number of positive sputum samples by number of visits with a sputum sample) and 95% CI will be reported as well.

Proportion of patients with sputum samples positive for bacterial/viral pathogens overall and for specific species, will be computed, with exact 95% confidence intervals (CIs), at each stable* scheduled visit (enrolment, visit 1, visit 2, visit3) and at each exacerbation visit confirming acute exacerbation**,***.

** A confirmed stable visit will be defined as a scheduled study visit for which the investigator confirms in the eCRF that the patient is stable/has recovered from a previous exacerbation.*

*** If an AECOPD occurs at the time when one of the scheduled [stable] study visits is planned, it should be handled and recorded as an AECOPD visit. If possible, the stable study visit should be re-scheduled to a later date, within the ± 14 days' time window from the scheduled appointment, when the patient is stable again.*

**** The definition of acute exacerbation of COPD (AECOPD) event is detailed in section 12.*

In details the following pathogens will be evaluated:

- Bacteria including, but not necessarily limited to, *H. influenzae**, *M. catarrhalis*, *S. pneumoniae*, *S. aureus*, *P. aeruginosa*, *K. pneumoniae* and *A. baumannii*) by bacteriological methods.

  **H. influenzae* will be also differentiated form *H. haemolyticus* by PCR assay. When possible, further differentiation (i.e., *Hi/NTHi*) will be reported.

- Viruses including, but not necessarily limited to, RSV, parainfluenza virus, enterovirus/ rhinovirus, metapneumovirus, influenza virus, adenovirus, bocavirus and coronavirus using qualitative PCR and rhinovirus by quantitative RT-PCR.

Patients providing a sputum sample as per protocol are all evaluable patients (*i.e.* those meeting all eligibility criteria and complying with procedures described in the protocol*,**). According to protocol, the interval of ± 14 days for any scheduled visit will not be a criterion for exclusion from the analyses. The number and percentage of patients with a stable visit within and without this time window will be presented.

** Patients attending a study visit and providing sputum sample.*

*** Patients using antibiotics on a continual basis (defined as more than 1 month in total) will be allowed to continue study participation, but may be eliminated from the analyses.*

If more than 10% of sputum samples have bad quality (as defined in section 11.2.3 in section 11), the analyses for primary objective will be repeated including only sputum samples with good and moderate quality.

An interim analysis will be performed when at least 40 AECOPD sputum samples are collected. This analysis will be limited to compute the rate of sputum sample positive for Hi, NTHi and Mcat together and per pathogen.

6.2.1. Statistical methods: analysis of correlated Data Using Generalized Estimating Equations

Average proportion of patients with sputum positive for bacteria/viruses and 95% CI will be computed using a GEE model. GEE approach uses weighted combinations of observations to extract the appropriate amount of information from correlated data such as the response measurements across multiple stable visits on the same patient (Burton, 1998).

For the purpose of the analyses a binomial distribution for the response variable with logit as link function will be assumed and an exchangeable correlation matrix will be chosen to account for the within-patient correlations.

The exchangeable correlation matrix assumes that the correlation between any two responses of the i^{th} individual is the same.

Other commonly used within-patient correlation matrices are as follows:

- Independence, repeated observations are uncorrelated
- Unspecified (unstructured), correlations within any two responses are unknown and need to be estimated
- Autoregressive of first order [AR(1)], assuming the interval length is the same between any two observations. Measurements taken further apart are less correlated than those taken closer together.

The model with the exchangeable correlation matrix will be adopted in the main analysis. Models with the other matrices could be run as sensitivity analyses. QIC and the related QICu statistic will be used to compare GEE models with different correlation matrices.

SAS codes

```
ods output ParameterEstimates=out_parm ClassLevels=Class;
proc genmod data=<dataset> descending;
 class pid;
 model y =  / dist = bin link = logit lrci ;
 Repeated subject=pid/sorted corr=cs PRINTMLE;
 by bacteria visit;
run;

data result;
 set out_parm(where=(parameter='Intercept') drop=ChiSq ProbChiSq DF);
 format percent LL UL percent8.1;
 Percent=exp(estimate)/(1+exp(estimate));
 LL=exp(LowerLRCL)/(1+exp(LowerLRCL));
 UL=exp(UpperLRCL)/(1+exp(UpperLRCL));
 unit=_N_;
run;
```

where pid = personal identifier, y = positivity to pathogen (0, 1), bacteria = type of bacteria, visit = type of visit (stable or exacerbation).

The REPEATED statement invokes the GEE method, specifies the correlation structure, and controls the displayed output from the GEE model. The option SUBJECT = PID identifies *pid* as the clustering variable, and the CORR = CS option specifies an exchangeable working correlation structure. Other matrices will be specified as CORR =: IND [independence], UN [unstructured], AR(1) [autoregressive (1)].

6.3. Analysis of secondary objectives

The analyses of the secondary objectives will be performed on the full analysis set.

6.3.1. Concordance between molecular and bacteriological methods

The frequency distribution of sputum samples positive or negative for specific bacterial species (including but not necessarily limited to *H. influenzae*, *M. catarrhalis*, *S. pneumoniae*, *S. aureus* and *P. aeruginosa*) obtained from PCR will be tabulated versus results obtained from culture at any visit (stable scheduled visits or AECOPD visits).

The concordance between culture and PCR at any visit will be evaluated using the Cohen's kappa (k) statistic and 95% CI. Concordance will be interpreted as good (>0.79), substantial (0.60–0.79), moderate (0.40–0.59), fair (0.21–0.39) or no concordance (<0.21) (McHugh, 2012).

The load of *H. influentiae, M. catarrhalis* and S. *pneumoniae* as identified in sputum culture versus RT-PCR, at any visit (stable scheduled visits or AECOPD visits), will be plotted.

The concordance between culture and PCR at any visit will be evaluated using Spearman's rank correlation coefficient. Concordance will be interpreted as good (>0.79), substantial (0.60–0.79), moderate (0.40–0.59), or fair (0.21–0.39) or no concordance (<0.21).

Proportion of patients with sputum samples positive for bacterial pathogens (including but not necessarily limited to *H. influenzae* (*Hi*/*H. haemolyticus*, *Hi/NTHi*), *M. catarrhalis*, *S. pneumoniae*, *S. aureus*, *P. aeruginosa*, *K. pneumoniae* and *A. baumannii*) by PCR will be computed, with 95% confidence intervals (CIs), at each and any stable scheduled visit and at each and any AECOPD visit confirming acute exacerbation.

6.3.2. Proportion of potential bacterial/viral pathogens by country, GOLD grade and by AECOPD severity

Average proportion of patients with sputum samples positive for bacterial or viral pathogens overall and for specific species, will be computed, with 95% confidence intervals (CIs), at any stable scheduled visit and at any AECOPD visit confirming acute exacerbation [mild, moderate, severe] by GOLD grade at enrolment [moderate, severe, very severe].

Average proportion of patients with sputum samples positive for bacterial or viral pathogens overall and for specific species, will be computed, with 95% confidence intervals (CIs), at any stable scheduled visit and at any AECOPD visit confirming acute exacerbation by country.

6.3.3. Sputum sample collection and quality

6.3.3.1. Sputum sample collection

The percentage of patients at each scheduled [stable] and exacerbation visit for whom a sputum sample is obtained will be computed overall and by the method the sample is obtained [i.e. spontaneous (at study visit or at patient's home), induced using 0.9% saline or induced using 3% saline]. The percentage of patients with induced sputum will be tabulated by GOLD grade at enrolment and by severity of the AECOPD. The proportion of sputum samples obtained at each confirmed stable or AECOPD visit [any, mild, moderate and severe] with or without induction and positive for specific bacterial\viral pathogens by bacteriological culture and bacterial\viral PCR will be computed.

The distribution of the number of sputum samples will presented according to time (days) since the start of exacerbation. The distribution of the proportion of sputum sample positive for bacteria/virus given the number of days between the start of exacerbation and sample collection will be presented.

The distribution of the number of sputum samples collected on the same day, after 1, 2 or more days since the start of antibiotics administration will be tabulated. The impact of previous antibiotic administration on bacterial results will be evaluated. The proportion of sputum samples obtained at each confirmed stable or AECOPD visit [any, mild, moderate and severe] with previous administration of antibiotics or not and positive for specific bacterial pathogens by bacteriological culture and PCR, respectively will be computed.

The impact of previous vaccination with a pneumococcal vaccine or an influenza vaccine (regardless of time) on *S. pneumoniae* and influenza virus will be evaluated. The proportion of sputum samples obtained at each confirmed stable or AECOPD visit with previous vaccination or not and positive for *S. pneumoniae* by culture and PCR or influenza virus by PCR will be computed.

6.3.3.2. Sputum sample quality

Sputum sample characteristics will be summarized (sputum sample weight, sputum colour, sputum appearance) at any stable visit and at any mild, moderate or severe AECOPD visit. Sputum sample characteristics will be also summarized by bacterial/viral presence in any stable visit and in any exacerbation visit.

Summary statistic (N, mean, median, SD, minimum, maximum) of neutrophils, eosinophils, macrophages, lymphocytes and bronchial epithelial cells will be tabulated given the type of visit and given induction of sputum

Sputum quality, derived as detailed in section 11.2.3 in section 11, will be tabulated given the type of visit.

6.3.4. Incidence rates of AECOPD

This study will provide an estimate of average number of exacerbations per person per year, in a COPD population at increased risk of exacerbation (i.e. moderate to very severe COPD patients with a documented history of at least 1 moderate or severe AECOPD in the year prior to enrolment).

To compute the incidence rate all the exacerbations will be taken into account. The observation period is from visit 1 to visit 3 or from visit1 to the last visit perfomed for patients lost to follow up.

6.3.4.1. Overall incidence

The following incidence rates (per patient per year) will be computed, with 95% CIs [using negative binomial regression detailed in section 6.3.4.4]:

- AECOPD events clinically confirmed by the investigator (5 and 6 dash boxes in Figure 1).
- Confirmed exacerbations (AECOPD events plus missed AECOPD visits with medical records: 2, 5, 6 and 7 dash boxes in Figure 1).
- Potential AECOPD with alert confirmed by phone call/or at the study site (2, 3, 4, 5, 6 and 7 dash boxes).

The incidence rate of AECOPD events, AECOPD events plus missed AECOPD visits with medical records and potential AECOPD confirmed by phone call/or at the study site, with 95% CIs, will be estimated according to COPD severity at enrolment based on GOLD grade [moderate, severe, very severe].

The incidence rates of AECOPD events and AECOPD events plus missed AECOPD visits with medical records, with 95% CIs, will be also computed for mild AECOPD, moderate AECOPD and for severe AECOPD.

In addition, the incidence rates of confirmed exacerbations (events and AECOPD events plus missed AECOPD visits with medical records), with 95% CIs, will be estimated according to covariates such as sex, country, age at enrolment (40-49, 50-59, 60-69, ≥70 years), smoking status at enrolment (current and former smokers) and number of exacerbations reported in the 12 months prior to enrolment (one, two and equal or above three).

An attempt to control simultaneously for all the above mentioned variables/factors will be done with multivariable negative binomial regression.

6.3.4.2. Incidence bacteria/viruses associated AECOPD

The following incidence rates (per patient per year) will be computed, with 95% CIs:

- AECOPD events with sputum sample containing bacterial pathogens found by culture or by PCR (overall and by, but not limited to, the following bacterial species: *H. influenzae* (and *H.influenzae* after differentiation form *H. Haemolyticus* by PCR assay), NTHi (after differentiation from *H. influenzae* by PCR assay), *M. catarrhalis*, *S. pneumoniae*, *S. aureus*, and *P. aeruginosa*).
- AECOPD events with sputum sample containing viral pathogens found by PCR (overall and by, but not limited to, the following viral species: RSV, parainfluenza virus, enterovirus/ rhinovirus, metapneumovirus, influenza virus, adenovirus, bocavirus and coronavirus using qualitative PCR and rhinovirus by quantitative RT-PCR).

These incidence rates, with 95% CIs, will be estimated also according to covariates such as COPD severity, sex, country, age at enrolment, smoking status at enrolment and number of exacerbations reported in the 12 months prior to enrolment.

In addition, incidence rate of mild, moderate and severe exacerbation having sputum containing any bacterial/viral pathogen will be computed overall and by species.

6.3.4.3. Analyses of exacerbations in relation to morning e-diaries

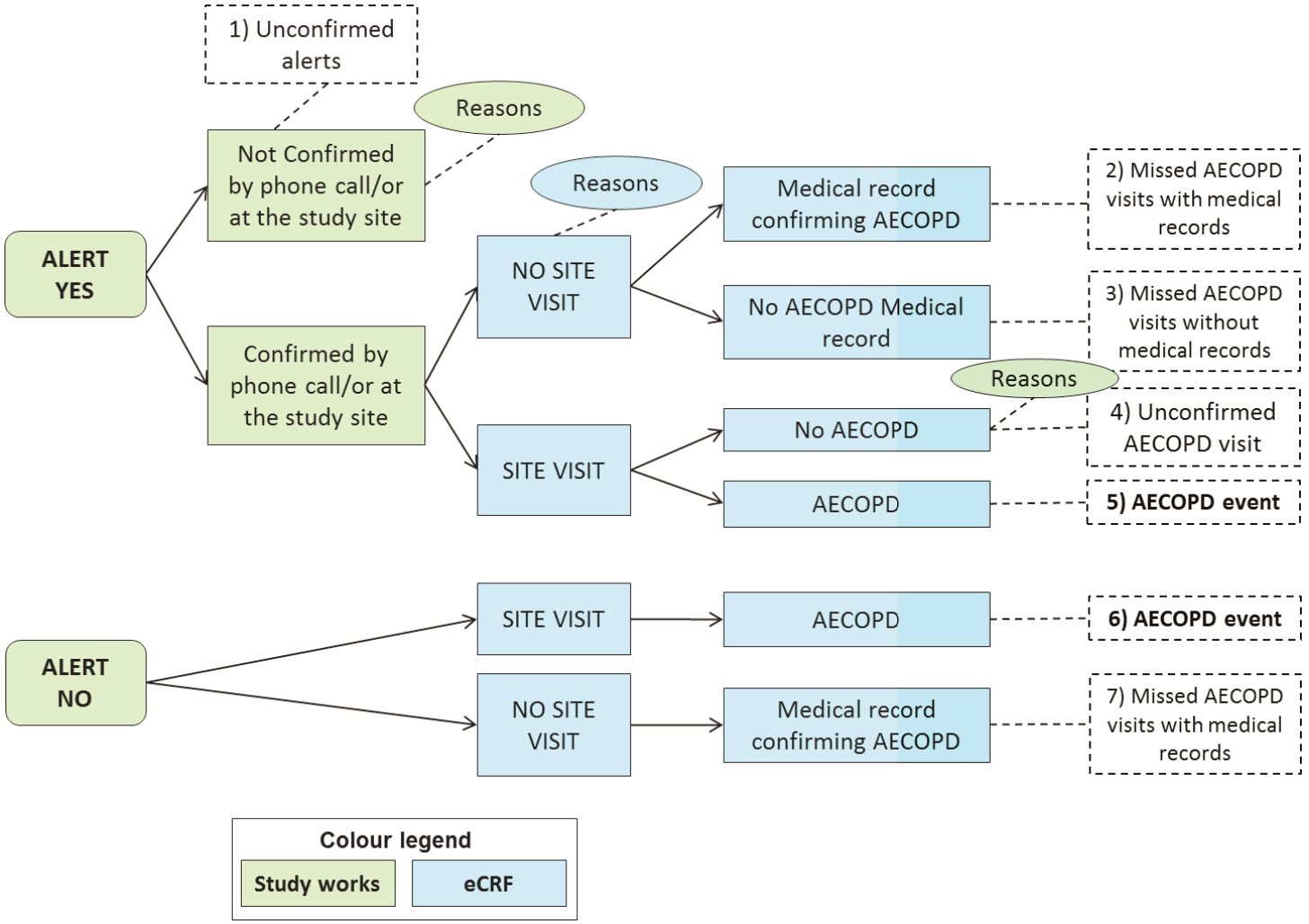
Among morning e-dairy alerts the frequency distribution of the following groups will be tabulated (refer to Figure 1 dashed boxes):

1. Potential event not confirmed by phone call/or at the study site (unconfirmed alerts, box 1),
2. Potential event confirmed by phone call/or at the study site, without site visit, with medical records confirming exacerbation (missed AECOPD visits with medical records, box 2),
3. Potential event confirmed by phone call/or at the study site, without site visit and without medical records (missed AECOPD visits without medical records, box 3).
4. Potential event confirmed by phone call/or at the study site, with site visit but not confirmed by study investigator (unconfirmed AECOPD visit, box 4),
5. Potential event confirmed by phone call/or at the study site, with site visit and confirmed by study investigator (AECOPD event, box 5),

Reasons for inconsistencies between morning e-diary alert and no confirmation by phone call, reasons for missing study visit after phone call confirmation, and reasons for inconsistencies between phone call confirmation and study visit will be tabulated.

In absence of morning diary alert the percentage of spontaneous visits confirming exacerbation or missed AECOPD visit with medical record confirming AECOPD will be tabulated (Figure 1, boxes 6 and 7).

Figure 1 Workflow of study assessment for AECOPD



6.3.4.4. Statistical methods: analysis of count data with overdispersion

The 95% CIs of the incidence rates will be estimated using the Generalised Linear Model (GLM) assuming a negative binomial distribution for the response variable with logarithm as link function and the logarithm of time for follow-up as an offset variable. The fit of the model will be examined using Q-Q plots of standardised residuals.

To overcome convergence issues, the Poisson regression model with adjustment for overdispersion will also be used to estimate the 95% CI of the incidence rate. The fit of the model will be examined using Q-Q plots of standardised residuals. Simulated envelopes will be produced.

SAS codes:

```
proc genmod data = <dataset>;
  class gold age smoking exacHist;
  model n_exac = GOLD age smoking exacHist country
      / link=log dist=negbin offset=logfuinyears ;
  ods output parameterestimates=out_parm nobs=nobs;
  ods exclude modelinfo modelfit parameterestimates nobs;
  run;

  data parm_est(keep= val rate rate_ll rate_ul);
    set out_parm(where=(parameter='intercept'));
    format rate rate_ll rate_ul 8.2;
    val= "mean number per year from negative binomial model without
covariate";
    rate=exp(estimate);
    rate_ll=exp(lowerwaldcl);
    rate_ul=exp(upperwaldcl);
  run;
```

where GOLD = COPD severity, age = age at enrolment (40-49, 50-59, 60-69, ≥70 years), smoking = smoking status at enrolment (current and former smokers), exacHist = number of exacerbations reported in the 12 months prior to enrolment (one, two and equal or above three), country = countries in the study (China [Hong Kong], Philippines, Korea, Taiwan), n_exac = number of exacerbations during the study and logfuinyears = logarithm of the follow-up period in years.

6.3.5. Severity and duration of AECOPD

Total number of exacerbations recorded during the study, exacerbation rate (number of exacerbation per patient-year) and the proportion of COPD patients experiencing a certain number of exacerbations will be presented overall and per bacterial/viral species. The exacerbation rate per year will be reported with and without correction for early withdrawal (see Section 11).

Total number of exacerbations recorded during the study, the mean exacerbation rate per year and the proportion of COPD patients experiencing a certain number of exacerbations will be presented overall and divided by AECOPD severity.

The number and proportion of COPD patients experiencing a certain number of exacerbations will be presented overall and by GOLD grade at enrolment.

The number of patients that report at least 1 AECOPD will be tabulated and descriptive statistics (N, mean, median, SD, minimum, maximum, first and third quartiles) on the average length of AECOPD episodes, as estimated by the investigator, will be presented, for any, mild, moderate and severe AECOPD.

Number of days between 2 consecutive exacerbations, as estimated by the investigator, will be summarized using the same descriptive statistics.

Descriptive statistics of the number of patients according to their percentage of days within an exacerbation period out of the total observation period as estimated by the investigator will be presented. For this analysis, the exacerbations for which the end date was unknown were not excluded and the length was estimated by the investigator as the average length per severity grade.

Results from the chest X-rays during exacerbation visit will be tabulated.

The percentage of scheduled visit in which the status of the patient is stable/recovered, or not recovered will be tabulated.

6.3.5.1. Seasonal distribution

Seasonal distribution of the exacerbations (total number of exacerbations, AECOPD events and of AECOPD events containing NTHi, Mcat, Strep, Entero/Rhinovirus, QPCR-Rhino and Influenza A and B) regardless of intensity will be reported by month.

6.3.5.2. Time to exacerbation

To investigate the effect of several baseline risk factors upon the time to first exacerbation since enrolment/first visit, Cox regression models will be fitted including several risk factors such as smoking status at enrolment (active smoker vs. former-smoker), number of exacerbation reported in the 12 months prior to enrolment (one, two and equal or above three), COPD grade at enrolment, age (40-49, 50-59, 60-69, $\geq$70 years), sex. Time to the first recurrence overall and by covariates will be depicted in Kaplan-Meier survival curves. Log-log survival curves will be displayed as well. If the proportionality of hazards assumption is met the log–log survival curves will be parallel for covariate strata. If there is an intersection between curves the covariate will be added as a stratification factor in the cox regression model.

To take into account that exacerbations may occur more than once over the follow-up time for a given patient, a Cox regression model for recurrent events using the counting process approach (Andersen, 1993) will be adopted (more details in section 6.3.5.2.1). The model will be fitted including smoking status at enrolment, history of exacerbation in the 12 months prior to enrolment, COPD severity at enrolment, age and sex as covariates.

Additional stratification/risk factor may be defined during the analyses.

6.3.5.2.1. Statistical methods: survival analysis

Kaplan-Meier curves

Kaplan-Meier survival curves will be produced to represent the cumulative survival probability over time.

The code follows.

```
Proc lifetest data = <dataset> method = KM plot = (S, LLS);
  time survt*Status(0);
  strata smoking;
run;
```

where survt = survival time in days (days to first exacerbation, for censored data the date of last contact will be used), status = censoring variable (0 = censored, 1 = event), and smoking = smoking status at enrolment (current and former smokers). The PLOTS=(S,LLS) option produces log–log curves as well as survival curves.

Cox Proportional Hazards regression model

To investigate the effect of several baseline risk factors upon the time to first exacerbation, Cox Proportional Hazards (PH) regression models will be fitted.

SAS codes

```
Proc phreg data = <dataset>;
  class GOLD age exacHist sex
  model survt*Status(0) = gold age exacHist / ties = exact rl;
  strata smoking;
run;
```

where GOLD = COPD severity, age = age at enrolment (40-49, 50-59, 60-69, ≥70 years), smoking = smoking status at enrolment (current and former smokers), exacHist = number of exacerbations reported in the 12 months prior to enrolment (one, two and equal or above three) and sex (male and female). For the scope of the example, the variable *smoking* is considered to be a stratification factor.

Recurrent event survival analysis

SAS code

```
Proc phreg data = <dataset> covs(aggregate);
  model (start,stop)*event(0) = exacerbationHistory smoker;
  id pid;
run;
```

The code (START,STOP)∗EVENT(0) in the MODEL statement indicates that the time intervals for each observation are defined by the variables START and STOP and that EVENT = 0 denotes a censored observation.

6.3.6. Impact of AECOPD on HRQOL

Descriptive statistics (N, mean, median, SD, minimum, maximum, first and third quartiles) on the CAT and SGRQ-C scores (total score and symptoms, activity and impacts component scores) will be tabulated at each stable scheduled visit, overall and by COPD GOLD grade at enrolment.

CAT and SGRQ-C (total score and symptoms, activity and impacts component scores) changes from first visit to final scheduled visit will be displayed according to the number of exacerbation occurred in the observation period. Correlation will be evaluated using Spearman's rank-order correlation coefficient.

Descriptive statistics (N, mean, median, SD, minimum, maximum, first and third quartiles) on the CAT score will be tabulated for any stable and any mild, moderate or severe exacerbation visit.

6.3.7. Impact of AECOPD on lung function

Descriptive statistics (N, mean, median, standard deviation, minimum and maximum) on post-bronchodilator FEV_1% of predicted normal value and PEF will be tabulated at enrolment and final visit. Change from baseline will be also summarized overall and represented according to the number of exacerbations. Changes in FEV_1% will be evaluated according to the number of exacerbations occurred in the observation period using Spearman's rank-order correlation coefficient.

6.3.8. Impact of AECOPD on healthcare utilisation

Descriptive statistics of the daily number of healthcare utilisation (number of HCU in the relevant period divided by the number of days) within the stable period and within the exacerbation period will be presented overall and divided by the following unscheduled visits:

1. Number of physician office consultations,
2. Number of visits to urgent care,
3. Number of visits to emergency department,
4. Number of hospitalizations.

Healthcare use for each patient will be obtained through review of the patient's medical record (aided by patient daily self-reporting in the eDiary).

Healthcare utilisation will be summarized according to COPD grade at enrolment, smoking status, sex and age at enrolment. In addition, healthcare utilisation will be summarized by AECOPD severity during exacerbations.

Overall HCU [total number and divided by type of HCU] per patient will be correlated to the number of exacerbation experienced during the observation time.

Current medication for COPD and additional COPD treatments prescribed by primary and secondary care physicians within the stable period and within the exacerbation period will be described by drug category overall and by COPD grade at enrolment, smoking status, sex and age at enrolment. In addition, current medication for COPD and additional COPD treatments prescribed will be summarized by AECOPD severity during exacerbations.

6.4. Analysis of tertiary objective

The analyses of the tertiary objectives will be performed on the full analysis set.

6.4.1. End date of AECOPD using EXACT-PRO

Descriptive statistics (N, mean, median, standard deviation, minimum and maximum, first and third quartiles) on the average length of AECOPD episodes, as estimated with the EXACT-PRO (section 12), will be presented, for any, mild, moderate and severe AECOPD.

Descriptive statistics of the number of patients according to their percentage of days within exacerbation periods out of the total observation period as estimated with the EXACT-PRO will be presented.

The frequency distribution of exacerbation confirmed at site visit or through medical records (with or without morning symptoms diary alert) will be tabulated versus EXACT-PRO exacerbations. The concordance between investigator judgement and EXACT-PRO exacerbations will be evaluated using the Cohen's kappa (k) test, with 95% CI. Concordance will be interpreted as good (>0.79), substantial (0.60–0.79), moderate (0.40–0.59), or fair (0.21–0.39) or no concordance (<0.21).

Descriptive statistics (N, mean, median, SD, minimum, maximum, first and third quartiles) for the EXACT daily scores will be tabulated at baseline, during the day before exacerbation, at exacerbation onset and at recovery according to investigator judgement. These statistics will be reported overall and for any mild, moderate or severe exacerbation visit.

Graphs detailing the EXACT-PRO scores before during and after exacerbation onset as defined clinically will be also presented.

Difference from baseline, with 95% CI, of EXACT score measured at stable scheduled visit will be plotted.

6.4.2. Mixed infections

The simultaneous bacterial [culture/PCR] and viral presence in sputum at enrolment, any stable and any exacerbation visit [mild, moderate, severe] will be computed overall and by pathogens.

6.4.3. Load of bacteria and virus

Total number of patients providing a sputum sample, the number of sputum samples positive to specific bacteria (by culture) and the frequencies of semi-quantitative bacteriological load (few scattered, +, ++, +++) will be tabulated at each scheduled [stable] visit and at each exacerbation visit.

Total number of patients providing a sputum sample, the number of sputum sample positive to specific bacteria (by PCR)/virus and average load will be tabulated at each scheduled [stable] visit and at each exacerbation visit.

6.4.4. Impact of occurrence of bacterial/viral pathogens in sputum on AECOPD

The following proportions will be computed:

- Proportion of patients with sputum samples which are positive for specific pathogens [NTHi, Mcat, S. pneumoniae, HRV] during an AECOPD visit but not in the previous stable visit/in the previous visit
- Proportion of patients with sputum samples which are negative for specific pathogens [NTHi, Mcat, *S. pneumoniae*, HRV] during an AECOPD visit but not in the previous stable visit/in the previous visit
- Proportion of patients with sputum samples which are positive for specific pathogens [NTHi, Mcat, *S. pneumoniae*, HRV] during an AECOPD visit and also in the previous stable visit/in the previous visit
- Proportion of patients with sputum samples which are negative for specific pathogens [NTHi, Mcat, *S. pneumoniae*, HRV] during an AECOPD visit and also in the previous stable visit/in the previous visit

Conditional logistic regression models, stratified by patient, will be fitted to estimate the effect of the presence of bacterial/viral pathogens in sputum on the odds of experiencing an acute exacerbation rather than remain in a stable COPD state. Separate models considering bacteria detected by bacteriological culture and by PCR will be fitted. Both the impact of presence and new occurrence of each specific pathogen will be considered. New occurrence of a pathogen will be defined as the detection of the specific pathogen after negative sputum sample at previous [any/stable] visit. Seasonality will be taken into account as a dichotomous variable: high season (October to March) and low season (April to September). Study country will be also taken in consideration. Backward selection of variables will be adopted until all remaining variables have individual P values smaller than 0.10.

Complete models with interactions between NTHi, Mcat, *S. pneumoniae*, Rhinovirus and season will be also fitted.

The conditional logistic model allows estimating the impact of independent variables from observations from each individual patient. In other words, by stratifying by patient, there is no need to adjust for time-invariant covariates such as age and sex.

6.4.5. Microbiome analysis

Because *H. influenzae* species could not be differentiated from *H. haemolyticus* species based on culture data, these isolates will be further processed by PCR to confirm the presence of *H. influenzae*. Distribution of Hi versus Hh in positive culture samples will be reported. An attempt to differentiate NTHi from Hi will be also made.

Sputum samples positive for enterovirus will be further tested for quantifying human rhinovirus. Results for Entero/Rhinovirus and QPCR rhinovirus will be reported.

6.4.6. Biomarkers

Descriptive statistics (N, mean, median, SD, minimum, maximum, first and third quartiles) for the level of each quantitative biomarker/haematology parameter will be tabulated for visit 1 and visit 3. The difference in sample mean will be reported together with a confidence interval.

Number and percentage of patients with fibrinogen, human serum C reactive protein (hsPCT), C-X-C Motif Chemokine Ligand 10 (CXCL10) values outside the laboratory normal ranges will be tabulated for visit 1 and visit 3.

Descriptive statistics will be presented overall and by GOLD grade at enrolment.

Changes in biomarkers will be evaluated according to the number of exacerbations occurred in the observation period using Pearson’s correlation coefficient (or Spearman's rank-order correlation coefficient if the normal distribution assumption is not met).

6.5. Analysis of safety

The analysis of serious adverse events will be performed on the full analysis set.

Adverse events (AEs) and serious adverse events (SAEs) considered possibly related to the study participation and withdrawal due to AEs and SAEs during the entire study period will be described in detail.

7. ANALYSIS INTERPRETATION

All analyses are descriptive. The use of these descriptive analyses should be limited to supportive analysis of confirmatory analyses or hypothesis generation.

8. CONDUCT OF ANALYSES

8.1. Sequence of analyses

The analyses will be performed in two steps:

- An interim analysis will be performed when at least 40 AECOPD sputum samples are available. It will include all data available at the time of analysis. The rate of positive samples will be computed for Hi, NTHi and Mcat together and per pathogen. The analysis will be done based on as clean as possible data. No study report will be written at this stage.
- A final analysis of all objectives will be performed when data obtained up to planned last patient last visit are available. A final study report will be written at this stage.

Description	**Analysis ID**	**Disclosure Purpose (CTRS=public posting, SR=study report, internal)**	**Dry run review needed (Y/N)**	***Study Headline Summary (SHS)*requiring expedited communication to upper management (Yes/No)**	**Reference for TFL**
Final analysis	E1_01	Study report	YES	NO	See TFL TOC
Interim analysis	E1_02	Internal	NO	NO	See TFL TOC

8.2. Statistical considerations for interim analyses

All analyses are descriptive and therefore no statistical adjustment for the interim analysis is required. No study report will be written for interim analysis.

9. CHANGES FROM PLANNED ANALYSES

Not applicable

10. LIST OF FINAL REPORT TABLES, LISTINGS AND FIGURES

The TFL TOC provides the list of tables/listings and figures needed for the study report. It also identifies the tables eligible for each analyses and their role (synopsis, in-text, post-text, SHS, CTRS,…). Note that all TFL aimed to be included as post-text are noted as post-text even if these are tabulation of individual data such as listing of SAE. The post-text material contain all source material for the study report and accordingly a post-text table may be redundant with an in-text table.

11. ANNEX 1 DATA DERIVATION RULE AND STATISTICAL METHODS

11.1. Statistical Method References

Andersen P.K., Borgan O., Gill R.D., and Keiding N. Statistical Models Based on Counting Processes. New York: Springer-Verlag 1993.

Atkinson AC. Plots, Transformations and Regression. Clarendon Press. 1985

Burton P, Gurrin L, Sly P. Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level mixed modelling. *Stat Med*. 1998;17:1261–91.

Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of binomial. *Biometrika*. 1934;26:404-413

Liang KY, and Zeger SL. Longitudinal Data Analysis Using Generalized Linear Models." Biometrika. 1986; 73:13–22.

McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med*. 2012;22(3):276-82.

11.2. Data derivation

11.2.1. COPD severity stage

The baseline COPD severity stage will be derived by the investigator from the spirometry results at the screening visit. The spirometric classification of airflow limitation in COPD patients is based on post-bronchodilator FEV_1 and can be divided into four GOLD grades (mild, moderate, severe and very severe) [Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of Chronic Obstructive Pulmonary Disease, updated 2013. Available from: http://www.goldcopd.org/uploads/users/files/GOLD_Report_2013_Feb20.pdf]. For the purpose of this study only patients with COPD graded from moderate to very severe will be enrolled.

Table 1 COPD severity stages included in the study

GOLD grade	In patients with $FEV_1/FVC<0.70$
Moderate	$50\% \leq FEV_1 < 80\%$ predicted
Severe	$30\% \leq FEV_1 < 50\%$ predicted
Very severe	$FEV_1 < 30\%$ predicted

The resulting COPD severity variable will be used in all models in which COPD at baseline will be taken into account.

A four-grade refined assessment of COPD severity stages will be computed according to the newest GOLD criteria [GOLD, 2017 Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of Chronic Obstructive Pulmonary Disease, updated 2017. Available from: http://goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/]. In the assessment scheme, patients should undergo spirometry to determine the severity of airflow limitation (i.e., spirometric grade). They should then undergo assessment of either dyspnea using mMRC or symptoms using CAT. Finally, their history of exacerbations (including prior hospitalizations) should be recorded.

For the purpose of this study we will only adopt CAT rather than adopting mMRC dyspnea scale. This choice is because the two tools are not interchangeable [Kim S, Oh J, Kim YI, Ban HJ, Kwon YS, Oh IJ, Kim KS, Kim YC, Lim SC. Differences in classification of COPD group using COPD assessment test (CAT) or modified Medical Research Council (mMRC) dyspnea scores: a cross-sectional analyses. *BMC Pulm Med*. 2013;13:35] and it is now recognized that COPD impacts patients beyond just dyspnea [Jones PW. Health status measurement in chronic obstructive pulmonary disease. *Thorax.* 2001; 56(11):880-7].

The four grades are:

- Patient Group A – Low Risk, Less Symptoms Typically GOLD 1 or GOLD 2 (Mild or Moderate airflow limitation); and 0-1 exacerbation per year and no hospitalization for exacerbation; and CAT score < 10;
- Patient Group B – Low Risk, More Symptoms Typically GOLD 1 or GOLD 2 (Mild or Moderate airflow limitation); and 0-1 exacerbation per year and no hospitalization for exacerbation; and CAT score ≥ 10;
- Patient Group C – High Risk, Less Symptoms Typically GOLD 3 or GOLD 4 (Severe or Very Severe airflow limitation); and ≥ 2 exacerbations per year or ≥ 1 with hospitalization for exacerbation; and CAT score < 10;
- Patient Group D – High Risk, More Symptoms Typically GOLD 3 or GOLD 4 (Severe or Very Severe airflow limitation); and ≥ 2 exacerbations per year or ≥ 1 with hospitalization for exacerbation; and CAT score ≥ 10.

Spirometry results for GOLD classifications, number of exacerbations per year, and number of exacerbation requiring hospitalization will be taken from the screening visit (visit 0). CAT score will be derived from visit 1.

11.2.2. HRQOL

11.2.2.1. CAT

The CAT index will be derived as the sum of the ratings recorded for each of the eight individual items. Each of these items has 6 possible scores (0, 1, 2, 3, 4 or 5), leading to a range of 0 to 40 for CAT score.

No missing items are expected because the e-diary system prompts the patient to compete an item of CAT to proceed to the following one. Partial questionnaires will be excluded.

11.2.2.2. SGRQ-C

The SGRQ-C total score will be derived as the weighted sum of the forty individual items leading to a range of 0 to 100 as detailed in the reference manual [St George's Respiratory Questionnaire for COPD patients (SGRQ-C), version 1.3, 2016].

The SGRQ-C symptoms, activity and impacts component scores will be derived as the weighted sum of subset of items as detailed in the reference manual.

No missing missing items are expected because the e-diary system prompts the patient to compete an item of SGRQ-C to proceed to the following one. Partial questionnaires will be excluded.

11.2.3. Sputum Quality

The quality of sputum is defined as follow:

- < 10 squamous epithelial cells/field: sample of good quality
- 10-25 squamous epithelial cells/field with significant numbers of cells derived from the lower respiratory tractus (ciliary epithelial cells, bronchial cells, round cells): sample of moderate quality
- > 25 squamous epithelial cells/field: sample of bad quality

11.3. Handling missing data

Missing or non-evaluable measurements will not be replaced. The only exception is detailed below.

The number of exacerbations per year will be extrapolated for patients withdrawing from the study to provide an estimate of the number of exacerbations over the one year observation period. The number of exacerbations in a year will be calculated by multiplying the number of exacerbations experienced by the patient by 13 and dividing by the number of 4-week periods the patient was followed up [Stockley R.A., Chopra N, Rice L. Addition of salmeterol to existing treatment in patients with COPD: a 12 month study. *Thorax.* 2006;61:122-128].

$$\text{Number of exacerbations per year} = \frac{\text{Number of exacerbations} * 13}{\text{Number of 4-week observation period intervals}}$$

The calculation of exacerbation rate will be based on follow-up period intervals of four weeks to avoid obtaining high imputed rates if a patient withdrew very early from the study after experiencing an exacerbation. Four-week intervals will be adopted since treatment courses for moderate/severe exacerbations are <=2 - 4 weeks when appropriate.

11.3.1. Date derivation

Partial dates will not be considered as missing or not evaluable date. In case day is missing, 15th of month is used. In case day and month are missing, 30th of June is used.

Partial or missing initial and end dates for exacerbation will not be considered to measure the length of an exacerbation.

11.3.2. Missing data for electronic Diary Card alert

- Days with missing data will not be ignored when defining start of an e-diary signal: e.g., if a patient has two qualifying symptoms followed by a missing day, if the next day with data has two qualifying symptoms this will not be defined as an exacerbation. Hence, the start of an e-diary signal of an exacerbation will require 2 consecutive non-missing days of worsening. If there is partially recorded symptom data at a visit, however, only the observed values will be considered but the day will not be discarded.
- Days with missing data will be ignored when defining end of an e-diary signal of an exacerbation: e.g., if a patient has two days missing in the baseline period and one day missing in the 3- day moving average, both baseline and 3-day moving average values will be calculated using observed data only. Similar rules will apply when there is partially recorded symptom data at a visit (and daily average values will be calculated among the observed values).
- All available data will be used for each patient.
- As a general rule, the baseline symptom count score, and 3 day moving averages are calculated as the average count score of the observed days in the baseline period. For example, if an e-diary signal occurs from day 10 onwards, the baseline period will be the average between day1 and day2 (day1 being the first study day). If day 13 is missing, the 3 day moving average at day 14 will only take into account day 14 and day 15.
- If there are more than 2 consecutive days of missing data after an e-diary signal not resolved, the e-diary signal will not be considered resolved until the next day with observed data if the corresponding moving average is below the baseline value. In order words, if the 3-day average daily score is missing, the day will be counted in the duration of the episode.

11.4. Number of decimals

The following decimal description will be used for the analyses.

Parameters	Number of decimal digits
% of count, including LL & UL of CI	1
p-value	3
Mean, median	Number of decimals in the raw data + 1
SD	Number of decimals in the raw data + 2
Minimum, maximum, range	Number of decimals in the raw data
First and third quartiles	Number of decimals in the raw data + 1

LL = Lower Limit UL = Upper Limit CI = Confidence Interval
SD = Standard deviation

11.5. Methodology for computing CI

- All CI computed will be two-sided 95% CI.
- The exact 95% CIs for a proportion within a group will be calculated [Clopper 1934].

12. ANNEX 2: AECOPD DEFINITION AND CALCULATION RULE

12.1. Definition, onset, recovery, duration of an AECOPD event

AECOPD events are confirmed according to investigator judgment after an AECOPD visit which is aimed to exclude worsening in symptoms not related to an AECOPD event.

AECOPD visits are scheduled after an electronic Diary Card alert confirmed by the investigator (by phone call or at the study site) or after spontaneous site visits due to worsening symptoms without any alert.

Electronic Diary Card alerts refer to daily symptoms recorded the morning after (also called morning symptoms). Alerts are based on the Anthonisen criteria:

- Worsening of two or more of the following major symptoms for at least two consecutive days*,**: dyspnea, sputum volume, sputum purulence (colour)

Or

- Worsening of any major symptom together with any of the following minor symptoms for at least two consecutive days*,**: sore throat, colds (nasal discharge and/or nasal congestion), fever (oral temperature $\geq 37.5°C$) without other cause, increased cough, increased wheeze.

** The same two symptoms do not have to be present on both days as long as at least one major symptom is present on both days.*

AECOPD onset will be defined as the first day of the two consecutive days of worsening symptoms.

AECOPD recovery will be determined/confirmed by the investigator/delegate during (a) follow-up phone call(s) which will take place every 2 weeks until the AECOPD has resolved. The end date will be based on when the investigator/delegate determines that the AECOPD symptoms have resolved. In determining this end date, consideration will be given to symptoms recorded in the electronic Diary Card and patient assessment during the phone calls.

AECOPD duration will be defined as the number of days from AECOPD onset (included) and AECOPD recovery (not included).

Duration = date2 – date1 + 1,

with date2: AECOPD recovery date,

date1: AECOPD onset date

12.2. Recovery and duration of an exacerbation according to EXACT-PRO:

An EXACT-PRO AECOPD event is defined as a 12 point increase above baseline for 2 consecutive days, or a nine point increase for 3 consecutive days.

The baseline EXACT-PRO score is calculated as the average daily EXACT-PRO score over days 14 to 8 preceding AECOPD symptomatic onset.

AECOPD recovery according to EXACT-PRO corresponds to the day at which the EXACT-PRO score return to its baseline value (or below) [Mackay AJ, Donaldson GC, Patel AR, Singh R, Kowlessar B, Wedzicha JA. Detection and severity grading of COPD exacerbations using the exacerbations of chronic pulmonary disease tool (EXACT). *Eur Respir J*. 2014;43(3):735-44.].

The AECOPD duration according to EXACT-PRO will be defined as the number of days from AECOPD onset (included) and AECOPD recovery according to EXACT-PRO (not included).

12.2.1. EXACT-PRO baseline score at enrolment

An EXACT-PRO baseline score at enrolment will be computed as the mean within-patient score over 7 days, with data present for a minimum of 4 of the 7 days. If fewer than 4 days of data are available, the EXACT baseline score cannot be calculated.

12.3. Time between 2 *e-diary* signals for exacerbations/evaluation of baseline symptom score:

The minimum duration between 2 e-diary signals with distinct baseline value is 15 days, which is the minimal period for which a new full exacerbation-free baseline can be calculated for the next e-diary episode.

If a new worsening of symptoms occurs between 6 and 14 days after the end of a previous exacerbation, the baseline period used will be the same baseline as the one of the original exacerbation. Therefore, it will last until the daily symptom count score has gone below the baseline symptom score of the previous event.

12.4. Definition of Missed AECOPD visits

Missed AECOPD visits with medical records (dashed boxes 2 and 7 in Figure 1) will capture:

- an alert from the electronic Diary Card without a site visit but that cannot be discounted as a possible exacerbation signal after reconciliation (detailed in the Study Procedure Manual),

and/or

- an exacerbation reported retrospectively during a scheduled visit clinically documented for which there has not been a corresponding exacerbation visit.

Onset date, recovery date and severity will be retrieved by medical records.

For the analyses not involving the collection of any biological specimen, exacerbations will refer to AECOPD events plus missed AECOPD visits with medical records.

Missed AECOPD visits without medical records (dashed box 3 in Figure 1) will capture:

- an alert from the electronic Diary Card, assessed AECOPD during the phone call, without a site visit and without AECOPD medical records.

12.5. Unconfirmed AECOPD event with morning e-diary signal alert notification.

An alert from the electronic Diary Card not confirmed to be an AECOPD after contact by phone call or at the study site.

12.6. Grading of severity of an exacerbation

Severity of exacerbations is defined as per protocol:

- Mild: Worsening symptoms of COPD that are self-managed by the patient.
- Moderate: Worsening symptoms of COPD that require treatment with oral corticosteroids and/or antibiotics.
- Severe: Worsening symptoms of COPD that require treatment with in-patient hospitalisation or home care intervention.

Severity of the exacerbation will not be derived but taken directly from the information entered in the CRF (i.e. Severity of exacerbation will be taken from the conclusion of the exacerbation visit).

13. ANNEX 3: STUDY SPECIFIC MOCK TFL

Study specific mock TFL is provided in a separate document.