- **Protocol number:** Z7244L01
- **Document title:** A Phase III, multi-centre, randomized, rater- and patient blind, placebo- and active-controlled, parallel group clinical trial to compare the safety and efficacy of 1-week treatment with intravenous N-acetylcysteine (NAC) 600 mg twice daily (active test treatment), ambroxol hydrochloride 30 mg twice daily (active control treatment) and placebo as expectorant therapies in adult Chinese Patients with respiratory tract diseases and abnormal mucus secretions
- **NCT number**: NCT03843541
- **Version number:** 2.0
- **Date of the document:** 14 Apr 2021

PAREXEL International

Zambon S.p.A.

Z7244L01

A PHASE III, MULTI-CENTRE, RANDOMIZED, RATER- AND PATIENT-BLIND, PLACEBO- AND ACTIVE-CONTROLLED, PARALLEL GROUP CLINICAL TRIAL TO COMPARE THE EFFICACY AND SAFETY OF 1-WEEK TREATMENT WITH INTRAVENOUS N-ACETYLCYSTEINE (NAC) 600 MG TWICE DAILY (ACTIVE TEST TREATMENT), AMBROXOL HYDROCHLORIDE 30 MG TWICE DAILY (ACTIVE CONTROL TREATMENT) AND PLACEBO AS EXPECTORANT THERAPIES IN ADULT CHINESE PATIENTS WITHRESPIRATORY TRACT DISEASES AND ABNORMAL MUCUS SECRETIONS

Statistical Analysis Plan

PAREXEL Project Number: CCI

SPONSOR SIGNATURE PAGE

PPD

Approved by:

Project Manager
Zambon S.p.A

PAREXEL SIGNATURE PAGE

Signatures below confirm that the review process has been completed in accordance with CCI .

This document has been approved and signed electronically on the final page by the following:

	Signatory
Author	PPD Project Role: PPD

Modification History

Unique Identifier for this Version	Date of the Document Version	Author	Significant Changes from Previous Authorized Version
Draft 0.1	Aug. 23, 2018	PPD	Not Applicable – First Version
Final version 1	Jun. 03 2019	PPD	Finalized and add program algorithm into Appendix
Amendment Final version	Sep.09 2019	PPD	Add definition of prior and concomitant medication, concomitant procedure. Correct analysis population, to be consistent with protocol and SAP. Add new AE/SAE of relationship and severity tables, to be consistent with protocol. Newly added tables are 14-3.1.10/11 & table 1-3.1.17/18 Separate AE leading to discontinuation table to two different drug interrupted and drug withdrawn table.
Amendment Final version	Oct.09 2019	PPD	Add modified MW test dummy code in the appendix Update concomitant medication partial date imputation Add additional section of Changes to protocol-defined analyses Add Hochberg procedure into Appendix. Remove TOC of CS in the SAP.

V2.0	April 13 2021	PPD	Adjusted “change from baseline to 1-week treatment” to “change from baseline to day 7” for viscosity score, expectoration difficulty score sputum color score, cough severity score and sputum volume Modified the partial date conventions for prior and concomitant medication. Added summaries of Covid-19 to disposition, protocol deviations and adverse events sections. Added definition and explanation of minor protocol deviations for morning dose between 9am and 11am, and efficacy assessments occurring within 2 hours after morning dose of study medication. Added summary of major protocol deviations leading to exclusion from the Per-Protocol Population. Defined prior prohibited medications and prohibited concomitant medications. As sputum volume is a continuous measure, added new variable (sputum volume in quartiles) for the stratified Mann-Whitney U tests. Adjusted the replicate number of bootstraps from 5000 to 1000. Adjusted the code of two-way nonparametric ANOVA procedure in 4.14.3. Adjusted the code of EM algorithm in 4.14.4. Specified that assessments for the premature discontinuation visit are to be included for data imputation. [redacted] Clarified that missing baseline data will not be imputed. Clarified that the the decision to proceed or not proceed to the second gate for the non-inferiority comparison (for both the mITT and PP Population) only depends on the superiority results for the mITT Population). Explained the process for reporting Suspected, Unexpected Serious Adverse Reactions (SUSARs). Added description of the baseline definition to allow efficacy assessments after and within 2h of first IMP to be used. To further specify the number of multiple imputation of EM to 0 in SAS code as EM algorithms does not involve multiple imputation. Clarified that fiberoptic bronchoscopies are collected as concomitant procedures, regardless of relationship to AEs and SAEs.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AE	Adverse Event
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BUN	Blood Urea Nitrogen
CTR	Clinical Trial Report
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
EM	Expectation-Maximization
γGT	gamma glutamyl transferase
GSH	Glutathione
HOCl	Hypochorous acid
ICF	Informed Consent Form
IMP	Investigational Medicinal Product
ITT	Intention-to-treat
Kg	Kilograms
LOCF	Last Observation Carried Forward
MC	Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligram
mITT	Modified Intention to Treat
ml	Milliliter
MVTF	Missing Value Treatment Failure
NAC	N-acetylcysteine
PD	Protocol Deviation
PP	Per-Protocol Population
PT	Preferred Term
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SUSAR	Suspected Unexpected Serious Adverse Reaction
SOC	System Organ Class
TEAE	Treatment Emergent Adverse Event
TESAE	Treatment Emergent Serious Adverse Event
WHO-DD	World Health Organization-Drug Dictionary
WOCF	Worst Observation Carried Forward

1 INTRODUCTION

The study is intended to provide pivotal evidence of efficacy and safety for registration in China of slow intravenous infusion of N-acetylcysteine (NAC) in adult hospitalised patients with respiratory tract diseases and abnormal mucus secretions. The study is designed in rater and patient-blind fashion because of the distinguishable appearance and sulfuric smell of NAC.

This document presents the statistical analysis plan (SAP) for the study: A phase III, multi-centre, randomized, rater- and patient-blind, placebo- and active-controlled, parallel group clinical trial to compare the efficacy and safety of 1-week treatment of intravenous N-acetylcysteine (NAC) 600 mg twice daily (active test treatment), ambroxol hydrochloride 30 mg twice daily (active control treatment) and placebo as expectorant therapies in adult Chinese patients with respiratory tract diseases and abnormal mucus secretions. This SAP contains definitions of analysis populations, derived variables, and statistical analysis methods for the analysis of efficacy and safety parameters.

This SAP is based upon the following study documents:

- Study Protocol, Version 2.0 (Dec 19, 2018)
- Electronic Case Report Form (eCRF), Version 4.0 (May 19, 2020)
- Protocol Deviation Specification, Version 8.0 (Jan 15, 2021)

All report outputs will be produced using SAS® version 9.4 in a secure and validated environment. All report outputs will be provided to the Sponsor in a single Microsoft Word document.

2 STUDY OBJECTIVES

2.1 Primary objective

Demonstrate that slow (at least 5 minutes) intravenous infusion of NAC 600 mg twice daily is superior to placebo in change from baseline to the end of 1-week treatment of sputum viscosity score or expectoration difficulty score in adult Chinese hospitalized patients with respiratory tract diseases and abnormal mucus secretions

2.2 Secondary objectives

2.2.1 Efficacy objectives

Demonstrate in the same patient population that:

- slow (at least 5 minutes) intravenous infusion of NAC 600 mg twice daily is superior to placebo in change from baseline to day 3 and end of 1-week treatment of other indicators of mucolytic activity (individual sputum viscosity and expectoration difficulty scores, cough score, mucous color score and mucous volume);

- slow (at least 5 minutes) intravenous infusion of NAC 600 mg twice daily is non-inferior to slow intravenous infusion of ambroxol hydrochloride 30 mg twice daily in change from baseline to the end of 1-week treatment of sputum viscosity score or expectoration difficulty score;
- slow (at least 5 minutes) intravenous infusion of ambroxol hydrochloride 30 mg twice daily is superior to placebo in change from baseline to day 3 and to the end of 1-week treatment of sputum viscosity score or expectoration difficulty score and of other indicators of mucolytic activity: individual sputum viscosity and expectoration difficulty scores, cough score, mucous color score and mucous volume.

2.2.2 Safety objectives

- Confirm that 1-week treatment with iv NAC 600 mg twice daily is safe and well tolerated.

3 INVESTIGATIONAL PLAN

3.1 Overall Study Design and Plan

This is a phase 3, multicenter, randomized, rater- and patient-blind, placebo- and active-controlled, 3-arm parallel group clinical trial.
The trial will be conducted in approximately 25 sites in China, and a total of 333 patients with respiratory tract diseases and abnormal mucus secretions will be enrolled.
The study will consist of a total of 4 core hospital visits (screening, baseline, treatment day 3, treatment day 7) and of one follow-up phone call (or follow-up hospital visit, at the patient's convenience) 2 weeks after discontinuation of treatment; if the patient is still hospitalized or if more convenient for the patient, the follow-up phone call can be replaced by a follow-up hospital visit.
At screening visit (Visit 1), patients will be asked to provide Informed Consent prior to any trial related procedure and will be checked against inclusion and exclusion criteria. Their medical history will be recorded. Furthermore, patients will undergo a 12-lead electrocardiogram (ECG) and laboratory assessments. Results must be available before Visit 2.
At randomization visit (Visit 2), to occur within 7 days after Visit 1, female patients will undergo a pregnancy test; all patients will be checked again against the full set of inclusion and exclusion criteria and eligible subjects will be randomized with a 1:1:1 ratio to receive NAC or ambroxol hydrochloride or placebo.
Procedures performed at each core hospital visit will include documentation of concomitant medications and treatment emergent adverse events (TEAEs), 24-hour sputum collection and assessment on a 4-point ordinal scale of expectoration difficulty, sputum viscosity, sputum color and cough.
In case a patient experiences a deterioration of expectoration capacity and related symptoms, and/or of an underlying condition, he/she should contact a study investigator immediately and undergo an unscheduled deterioration visit. If the patient and/or the

investigator believe that it is in the patient's best interest to receive a rescue mucolytic, the patient will be discontinued from the study and rescue mucolytic will be administered, with any additional treatment the investigator deems appropriate.

3.2 Efficacy and Safety Variables

3.2.1 Efficacy Variables

- Sputum viscosity, expectoration difficulty, sputum color and cough are assessed by means of ordinal categorical 4-point scales with 0 = best and 3= worst.

For each patient the assessment is to be carried out by the blinded rater on treatment days 1, 3 and 7 in the morning before 9 am before administration of study drug. Should a patient discontinue prematurely from the trial (including discharge from the hospital before day 7) an assessment should be conducted on the day of discharge whenever possible.

Scoring criteria are as follows:

	0	1	2	3
Sputum viscosity	Liquid (normal viscosity)	Fluid (mildly increased viscosity)	Viscous (moderately increased viscosity)	Sticky (severely increased viscosity)
Expectoration difficulty	No difficulty	Mild difficulty	Moderate difficulty	Marked difficulty
Sputum Color	Mostly white	Mostly pale yellow	Mostly dark yellow	Very dark yellow /green
Cough	No cough	Sporadic and mild cough	Moderate cough	Severe Cough

- Sputum volume
  At Randomization Visit (Visit 2), Visit 3 and Visit 4 the patients will be given a graduated container and instructed to use it to collect all sputum generated in the following 24 hours (i.e. same time of the following day). The day the container will be collected, and the 24-hour sputum volume measured in mL/24h.

3.2.2 Safety Variables

- Adverse events assessed for severity and seriousness.
- 12-lead electrocardiogram
- Laboratory measurements including hematology, liver enzymes, serum creatinine.

- Vital signs

4 STATISTICAL METHODS

4.1 Data Quality Assurance

All tables, figures and data listings to be included in the report will be independently checked for consistency, integrity and in accordance with standard PAREXEL procedures.

4.2 General Presentation Considerations

Continuous data will be summarized in terms of the mean, standard deviation (SD), median, minimum, maximum and number of observations, unless otherwise stated. The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean, median, lower quartile and upper quartile will be reported to one more decimal place than the raw data recorded in the database. The SD will be reported to two more decimal places than the raw data recorded in the database. In general, the maximum number of decimal places reported shall be four for any summary statistic.

Categorical data will be summarized in terms of the number of subjects providing data at the relevant time point (n), frequency counts and percentages. Any planned collapsing of categories will be detailed in the SAP text and the data displays.

Percentages will be presented to one decimal place. Percentages will not be presented for zero counts. Percentages will be calculated using n as the denominator. If sample sizes are small, the data displays will show the percentages, but any textual report will describe frequencies only.

Changes from baseline in categorical data will be summarized using shift tables where appropriate.

P-values greater than or equal to 0.001, in general, will be presented to three decimal places. p-values less than 0.001 will be presented as "<0.001".

Confidence intervals will be presented to one more decimal place than the raw data.

The data from all sites will be pooled. Unless stated otherwise, all available data from withdrawn subjects will be included in the analysis up to the time of withdrawal.

All tests will be one-sided and performed at the significance nominal level of $\alpha = 0.025$.

Baseline for Efficacy and safety

Generally, baseline measurements are the last available data obtained prior to the first administration of investigational medicinal product (IMP) at Visit 2.

For sputum viscosity, expectoration difficulty, sputum color and cough, the last available data at Visit 2 (Day 1) within 2 hours after the first dose of IMP can be used for baseline if no data was obtained prior to the first dose of IMP.
For 24-hour sputum collection, the baseline measurement will be the last available data captured on Visit 2 (Day 1) or the next day (Day 2). This is because the collection of sputum is to begin on Day 1 and lasts for 24 hours up until the same time the next day (Day 2). Therefore, the baseline sputum volume measurement may be after the first dose of IMP.

Scheduled and Unscheduled Assessments
The by-visit tables will only include the scheduled assessments, unless specifically stated otherwise. If more than one value is available for a given visit, the first valid observation will be used in summary tables, unless specifically stated otherwise, and all observations will be presented in listings.

Unscheduled assessments will be included in listings, but not in summary tables.

Detection of Outliers
Any outliers that are detected during the review of the data before database lock will be investigated. If necessary, queries will be issued to Investigators, to either correct or confirm the outlier.

4.3 Study Subjects

4.3.1 Disposition of Subjects

A clear accounting of the disposition of all subjects who enter the study will be provided, from screening to study completion.

The following summaries will be provided:

- A summary of the number and percentage of subjects who were screened, randomized, and subjects who are excluded prior to randomization by major reason and overall, using the Enrolled population (Section 4.4.1).
- A summary of the number and percentage of patients treated (with at least one dose or partial dose of IMP) and discontinued from treatment, and the number and percentage of major reason of patient discontinuation from treatment (including due to Covid-19) by treatment group and overall, using the Intention-To-Treat (ITT) population (Section 4.4.2).
- A summary of the number and percentage of patients who completed the study and discontinued from the study, and the number and percentage of major reasons for discontinuation from the study (including withdrawn due to Covid-19), by treatment group and overall, using the ITT population (Section 4.4.2).

A by-subject listing of disposition data will be provided, using the Enrolled Population (Section 4.4.1).

4.3.2 Protocol Deviations

Protocol Deviations (PDs) will be classified as "major" or "minor", and any action to be taken regarding the exclusion of subjects or affected data from specific analyses will be defined in the project-specific Protocol Deviation Specification. Exclusion of subjects from the PP analyses will be decided jointly by the CRO and Sponsor's Medical Monitor, Clinical Trial Manager and Biostatistician prior to database lock. The subjects or observations to be excluded, and the reasons for their exclusion will be documented and approved by the above-mentioned persons prior to database release. The documentation will be filed together with the remaining trial documentation.

Major PDs are defined as those deviations from the protocol likely to have an impact on the subject's rights, safety, well-being, and/or on the validity of the data for analysis. Major PDs that have impact on the efficacy assessment will lead to the exclusion of a subject from the PP population. PDs will be assessed prior to final database lock.

Note that per the Protocol, morning study drug administration is meant to occur prior to 9am. In addition, efficacy assessments are meant to occur before the morning study drug administration for sputum viscosity, expectoration difficulty, sputum color and cough. Following medical discussion, it was decided that morning study drug administration between 9am and 11am would be considered minor protocol deviations. In addition, efficacy assessments occurring within 2 hours after the morning dose of study medication would also be classified as minor deviations. Efficacy assessments occurring after 2 hours of the morning dose of study medication will be classified as a major deviation.

The PD summaries will include:

- the number and percentage of subjects with a major PD by treatment group and overall, and by type of major deviation, based on the ITT population (Section 4.4.2).
- the number and percentage of subjects with a major PD leading to exclusion from the Per- Protocol Population by treatment group and overall, and by type of major deviation leading to exclusion, based on the ITT Population (Section 4.4.2).

A by-subject listing of PDs should be provided based on the ITT population (Section 4.4.2).

4.4 Analysis Populations

4.4.1 Enrolled Population

Enrolled population is defined as all patients who provided informed consent for this study.

4.4.2 Intention-To-Treat (ITT) Population

The Intention-To-Treat Population will include all subjects who provided informed consent and received a patient number (randomization number) whether or not they receive IMP. Following the ITT principle, subjects will be analyzed according to the treatment they have been assigned to at randomization.

4.4.3 Modified Intention-To-Treat (mITT) Population

The Modified ITT (mITT) Population will comprise all subjects in the ITT population who received at least 1 dose or partial dose of the IMP.

4.4.4 Safety Population

The Safety Population will comprise all subjects who provide Informed Consent and received at least 1 dose or partial dose of IMP. Subjects will be analyzed according to the treatment they actually received.

4.4.5 Per-protocol Population

The Per-protocol (PP) Population will include all mITT subjects who were compliant with study drug administration (i.e. had a compliance of at least 80%) and who had no major protocol deviations that were considered as potentially impacting the efficacy results. Major protocol deviations might include, but are not be limited to, subjects taking a not-permitted concomitant medication, the IMP not being administered during the trial as defined in the protocol, subjects receiving a treatment different than the one assigned by randomization, major PDs related to Covid-19. With the exclusion of non-inferiority comparisons, results of the primary and secondary efficacy endpoints analyses conducted in the PP will be considered as supportive.

A summary of analysis populations will be provided using the ITT population.

A by-subject listing of analysis populations will be provided using the ITT population.

If not otherwise stated, the mITT population will be used for the primary analyses of each of the efficacy endpoints whilst results from supplemental analyses using the Per Protocol Population will be compared to those based on the mITT population to assess the effects of dropouts, missing data and exclusions for major protocol violations.

4.5 Demographic and Other Baseline Characteristics

At Visit 1 (Screening) the subjects' demographic data will be documented, including age, ethnicity, gender, height, weight as well as smoking history and alcohol use.

The following summaries will be provided:

- A summary of the demographic data, baseline height (cm), and baseline weight (kg) will be provided using the mITT population (Section 4.4.3).
- A summary of the demographic data, baseline height (cm), and baseline weight (kg) will be provided using the PP population (Section 4.4.5).

Age will be calculated as the number of complete years between a patient's date of birth and the date of informed consent. Age (in years) will be calculated as:

(Date of informed consent – date of birth + 1)/365.25

If date of birth is a partial date missing month only, month will be assumed to be June. If date of birth is a partial date missing day only, day will be assumed to be 15th of the month. If date of birth is a partial date missing both month and day, it will be assumed to be Jun 30 of the year.

A by-subject listing of the demographic data and the baseline characteristics will be provided using the ITT population (Section 4.4.2).

4.6 Medical and Surgical History

Medical and surgical history will be summarized separately using the ITT population (Section 4.4.2). Frequency table will be provided by treatment group and total group, system organ class (SOC) and preferred term (PT), where SOC and PT are coded by Medical Dictionary for Regulatory Activities (MedDRA) Version 23.1.

The following listings will be provided using the ITT population (Section 4.4.2):

- A by-subject listing of medical history, including the reported term, the coded term, the start date, and the end date.
- A by-subject listing of surgical history, including the reported term, the coded term, the start date, and the end date.

4.7 Prior and Concomitant Medication

Medication start and stop dates will be compared to the date of first dose of study medication to allow medications to be classified as either Prior, or Concomitant.

Medications that start and stop prior to the date of first dose of study medication will be classified as Prior medication. If a medication stops on or after the date of first dose of study medication up to the last dose of study medication, then the medication will be classified as a Concomitant medication.

If medication start and/or stop dates are missing or partial, the dates will be compared as far as possible with the date of first dose of study medication. Medications will be assumed to be Concomitant, unless there is clear evidence (through comparison of partial dates) to suggest that the medication stopped prior to the first dose of study medication. If there is clear evidence to suggest that the medication stopped prior to the first dose of study medication, the medication will be assumed to be Prior.
Imputation of partial dates of prior and concomitant medications is described in Appendix 4.14 (section 4.14.7).

Prior and concomitant medication will be summarized using the ITT population (Section 4.4.2) separately. Frequency table will be provided by treatment group, anatomical therapeutic chemical (ATC) drug class and preferred term (PT), where ATC and PT will be coded using the World Health Organization-Drug Dictionary Enhanced (WHO-DDE) September 2020 B3.

A by-subject listing of prior and concomitant medications will be provided using the ITT population.

In addition, the following treatments are not permitted from Visit 1 to Visit 4 (end of IMP treatment) or unscheduled Deterioration Visit:

- Expectorants
- Antitussive agents
- Sedatives
- Traditional Chinese Medicine (TCM) treatments.

Prior prohibited medications are defined as those prohibited medications with a stop date on or after the Visit 1 date, and prior to the first dose of study medication date. Prohibited concomitant medications are defined above as for concomitant medications for those treatments not permitted.
Prior prohibited medications and prohibited concomitant medications will be summarized for the ITT population. Frequency tables will be provided following the same structure as for prior and concomitant medications.

A by-subject listing of prohibited medications will be provided using the ITT population.

4.8 Concomitant Procedures

Concomitant procedures are defined as any procedures starting after the first dose of study drug up to the last dose date of study drug.

Concomitant procedures will be collected only if they are related to AEs and SAEs, with the exception of fiberoptic bronchoscopies, which will be captured regardless of relationship to AEs and SAEs. Concomitant procedures will be summarized using the Safety population (Section 4.4.4). Frequency table will be provided by treatment group, system organ class (SOC) and preferred term (PT), where SOC and PT are coded by Medical Dictionary for Regulatory Activities (MedDRA) Version 23.1.

A by-subject listing of concomitant procedures will be provided using the Safety population.

4.9 Efficacy Evaluation

The co-primary efficacy endpoints are: “the change from baseline to the 1-week treatment in sputum viscosity score” and “the change from baseline to the 1-week treatment in expectoration difficulty score” and will be analyzed using the mITT and PP populations.
The primary study objective will be assessed by testing superiority of NAC compared to placebo and will be achieved if at least one of the two co-primary end-points will be found statistically significant.
The key secondary objective will be assessed through the analysis of non-inferiority of NAC compared to ambroxol and, similarly, will be reached if at least one of the two co-

primary end-points will be found statistically significant for both the mITT and PP populations. Details will be found in section 4.9.1.4.
Conversely, the analyses of the remaining secondary efficacy endpoints will be treated as non-key secondary objectives since they will be performed only to support primary endpoints findings.

A by-subject listing of all efficacy variables (score of sputum viscosity, expectoration difficulty, sputum color and cough, and sputum volume) will be provided using the mITT population (Section 4.4.3).

4.9.1 Analysis and Data Conventions

This study is designed to test for the superiority of NAC vs. placebo and the non-inferiority of NAC vs. ambroxol hydrochloride in change from baseline to day 7 of sputum viscosity score as well as change from baseline to day 7 of expectoration difficulty score.

4.9.1.1 Multi-center Studies

A test of site-by-treatment interaction will be carried-out using a two-way nonparametric ANOVA. The interaction will be considered statistically significant for a two-tailed α=0.1. Also, in order to graphically assess the possible heterogeneity of the treatment effect across centres, forest plots will be generated to display the results at each center. CCI [redacted] In the case of non-convergence or SAS errors/warnings in the log (due to low numbers of patients at some sites), sites with 8 or fewer subjects randomized will be grouped together.

4.9.1.2 Adjustments for Covariates

Adjustments for covariates will be performed for the following two parts.

First, for the two-way non-parametric ANOVA, modelling adjustments for covariates will be performed by including treatment, site, and site-by-treatment interaction terms in the model.
Secondly, adjustments of covariate for stratified Mann-Whitney U tests will also be performed by considering baseline score as a block/stratification variable in order to adjust imbalance. Block/stratification for baseline score refers to baseline results of sputum viscosity score, expectoration difficulty score, sputum color score, cough severity score and sputum volume. A new variable will be created dividing sputum volume into quartiles and this variable will be used as the block/stratification variable for baseline sputum score (as sputum score is a continuous measure unlike the other efficacy variables). Details of two-way non-parametric ANOVA can be seen in Appendix (Section 4.14.3).

4.9.1.3 Handling of Dropouts or Missing Data

Missing data on the primary and secondary efficacy endpoints will be imputed as treatment failures (MVTF) by replacing missing values using worst observation carried forward

(WOCF) with the worst scores recorded at baseline or post baseline (follow up) (including measurements from the premature discontinuation visit) imputed to missing visits. MVTF is the primary imputation method. The Expectation-Maximization (EM) algorithm and last observation carried forward (LOCF) will be used as sensitivity analyses.
Data missing at baseline will not be imputed.

4.9.1.4 Multiple Comparisons/Multiplicity

The overall type I family-wise error rate will be preserved at the one-tailed 0.025 nominal level (i.e. $\alpha = 0.025$) by controlling multiplicity over the two co-primary endpoints and over the two study targets (superiority and non-inferiority) using a multiple-sequence gatekeeping procedure (Keays et al., 1991). To achieve this, the analyses on primary endpoints will be grouped into two family of comparisons (gatekeepers) and processed as follows:

1. The first gatekeeper includes the superiority contrasts "NAC vs placebo" with the adjustment for multiple comparisons across the two co-primary endpoints performed according to Bonferroni. This implies that a difference will be considered statistically significant for a one-tailed p-value ≤ 0.0125 (i.e. $\alpha/2 = 0.0125$).
   The transition to the second gatekeeper requires that at least one of the two co-primary endpoints is statistically significant in the mITT population.
2. The second gatekeeper includes the family of non-inferiority contrasts "NAC vs ambroxol" conducted for both the mITT population and PP population.
   The analyses in the second gatekeeper will be performed depending upon the results obtained in the first gatekeeper as shown below:
   a. If both superiority contrasts are found statistically significant in the mITT population, non-inferiority will be tested on both co-primary endpoints for both the mITT and PP populations with an overall probability α set to 0.025 (one-tailed) and with the adjustment for multiple comparisons across the two co-primary endpoints performed according to Hochberg. Details of the Hochberg procedure can be found in Appendix (Section 4.14.1).
   b. If only one of the two superiority contrasts are found statistically significant in the mITT population, non-inferiority will be tested for that co-primary endpoint only for both the mITT and PP populations and will be considered achieved if the associated one-tailed p-values are ≤ 0.0125.

Statistically significant results will be needed in both the mITT and PP populations to accept the hypothesis of non-inferiority of NAC versus ambroxol.

Consistent with the hypothesis testing process performed on the co-primary endpoints, the strategy described above will also be applied to control the coverage of confidence intervals at the two-tailed 95% nominal level (i.e. $1 - 2\alpha = 0.95$).

The analyses on secondary efficacy endpoints (not key) will be performed only to support primary endpoints findings and/or for explorative reasons, therefore no adjustment of significance level will be made to account for multiple comparisons.

4.9.1.5 Interim Analyses

No interim analysis will be performed.

4.9.1.6 Examination of Subgroups

Upon completion of the study, the study statistician in consultation with Zambon's Clinical Team will determine the proportion of concomitant medications between the treatment groups and will decide whether to perform subgroup efficacy analyses if deemed appropriate. The analyses described for the primary and secondary efficacy endpoints will be performed separately for each of the identified stratums.

4.9.2 Primary Efficacy Analysis

The co-primary efficacy endpoints are: "the change from baseline to day 7 in sputum viscosity score" and "the change from baseline to day 7 in expectoration difficulty score" and they will be analyzed on the mITT and PP populations by means of stratified Mann-Whitney U Statistics for testing the hypotheses of superiority of NAC versus placebo (primary objective) and by means of a modified Mann-Whitney U Statistics (Rothmann, et al., 2012) for testing the hypotheses of non-inferiority of NAC versus ambroxol (key secondary objective) assuming a margin (i.e. δ') of non-inferiority equal to 0.30.
Statistically significant results will be needed on both mITT and PP populations to accept the hypothesis of non-inferiority of NAC versus ambroxol. Non-inferiority will be declared if the one-sided confidence interval will be within the interval [1/2-δ'', 1) so, equivalently, the lower limit of the confidence interval will be greater than 1/2-δ'', where δ'' is the non-inferiority margin converted on a probability scale from a point scale δ'. Details of the converting margin can be found in Appendix 4.14.6. Note that the one-sided confidence interval will be conducted depending on the outcome of the first gatekeeper procedure for the superiority contrasts.

In addition, for all superiority test in efficacy analysis, Hodges Lehmann point estimate of the median treatment differences together with associated two-sided 95% confidence intervals will be computed using nonparametric bootstrap; For all non-inferiority tests in the efficacy analysis, point estimates of treatment differences in probability scale together with associated one-side confidence interval will be provided. Nonparametric bootstrap point estimates of treatment differences in probability scale together with two-sided 95% confidence intervals will also be computed. Details of bootstrap can be found in Appendix (Section 4.14.2).

The following summaries of testing the hypotheses of superiority of NAC versus placebo and non-inferiority of NAC versus ambroxol will be provided:

- A summary of the change from baseline to day 7 in sputum viscosity score using the mITT population, in which the missing data will be handled using the MVTF
- A summary of the change from baseline to day 7 in expectoration difficulty score using the mITT population, in which the missing data will be handled using the MVTF

- A summary of the change from baseline to day 7 in sputum viscosity score using the PP population, in which the missing data will be handled using the MVTF
- A summary of the change from baseline to day 7 in expectoration difficulty score using the PP population, in which the missing data will be handled using the MVTF

The following sensitivity analyses will be provided:

- A summary of p-value of the superiority and non-inferiority tests for the change from baseline to day 7 in sputum viscosity score using the mITT population and PP population, in which the missing data will be handled using the EM algorithm
- A summary of p-value of the superiority and non-inferiority tests for the change from baseline to day 7 in expectoration difficulty score using the mITT population and PP population, in which the missing data will be handled using the EM algorithm
- A summary of p-value of the superiority and non-inferiority tests for the change from baseline to day 7 in sputum viscosity score using the mITT population and PP population, in which the missing data will be handled using the LOCF
- A summary of p-value of the superiority and non-inferiority tests for the change from baseline to day 7 in expectoration difficulty score using the mITT population and PP population, in which the missing data will be handled using the LOCF

The test statistics for the superiority and non-inferiority tests can be found in the Appendix (Section 4.14.5 and 4.14.6).

4.9.3 Secondary Efficacy Analysis

The secondary efficacy variables will be analyzed using the mITT, and the missing data will be handled using the MVTF, EM and LOCF methods:

- Change from baseline to day 3 in sputum viscosity score: the hypothesis of superiority of NAC compared to placebo will be assessed using the same nonparametric test (stratified Mann-Whitney U Statistics blocking for baseline score), whilst Hodges Lehmann point estimate of the median treatment differences together with associated two-sided 95% confidence intervals will be computed using nonparametric bootstrap.
- Change from baseline to day 3 in expectoration difficulty score: the hypothesis of superiority of NAC compared to placebo will be assessed using the same nonparametric test (stratified Mann-Whitney U Statistics blocking the baseline score), whilst Hodges Lehmann point estimate of the median treatment differences together with associated two-sided 95% confidence intervals will be computed using nonparametric bootstrap.
- Change from baseline to day 3 and to day 7 period in sputum color score: the hypothesis of superiority of NAC compared to placebo will be assessed using the same nonparametric test (stratified Mann-Whitney U Statistics blocking for baseline score), whilst Hodges Lehmann point estimate of the median treatment differences together with associated two-sided 95% confidence intervals will be computed using nonparametric bootstrap.

- Change from baseline to day 3 and to day 7 period in cough severity score: the hypothesis of superiority of NAC compared to placebo will be assessed using the same nonparametric test (stratified Mann-Whitney U Statistics blocking for baseline score), whilst Hodges Lehmann point estimates of the median treatment differences together with associated two-sided 95% confidence intervals will be computed using nonparametric bootstrap.
- Change from baseline to day 3 and to day 7 in sputum volume: the hypothesis of superiority of NAC compared to placebo will be assessed using the same nonparametric test (stratified Mann-Whitney U Statistics blocking for baseline score), whilst Hodges Lehmann point estimate of the median treatment differences together with associated two-sided 95% confidence intervals will be computed using nonparametric bootstrap.

The same statistical analysis (stratified Mann-Whitney U Statistics blocking for baseline score) will be carried out for testing superiority of ambroxol vs. placebo on the same efficacy endpoints. Results will be used for the purpose of verifying study's assay sensitivity.

The following summaries will be provided using the mITT population (Section 4.4.3):

- A summary of comparing ambroxol and placebo for the change from baseline to day 3 and to day 7 in sputum viscosity score, in which the missing data will be handled using the MVTF, expectation-maximization (EM) algorithm and last observation carried forward (LOCF) as sensitivity analyses.
- A summary of comparing ambroxol and placebo for the change from baseline to day 3 and to day 7 in expectoration difficulty score, in which the missing data will be handled using the MVTF, expectation-maximization (EM) algorithm and last observation carried forward (LOCF) as sensitivity analyses.
- A summary of comparing ambroxol and placebo for the change from baseline to day 3 and to day 7 in sputum color score, in which the missing data will be handled using the MVTF, expectation-maximization (EM) algorithm and last observation carried forward (LOCF) as sensitivity analyses.
- A summary of comparing ambroxol and placebo for the change from baseline to day 3 and to day 7 in cough severity score, in which the missing data will be handled using the MVTF, expectation-maximization (EM) algorithm and last observation carried forward (LOCF) as sensitivity analyses.
- A summary of comparing ambroxol and placebo for the change from baseline to day 3 and to day 7 in sputum volume, in which the missing data will be handled using the MVTF, expectation-maximization (EM) algorithm and last observation carried forward (LOCF) as sensitivity analyses.

The stratified Mann-Whitney U test statistic can be seen in the Appendix (Section 4.14.5).

4.9.4 Multi-center Analysis

[redacted] In the first instance, since the aim of the analysis is to examine sites, the analysis will be performed using individual

sites (i.e. no pooling of sites). If the situation occurs where there are convergence issues or SAS errors/warnings due to some sites with small numbers of subjects, sites with 8 or fewer randomized subjects will be grouped for the multi-center analysis.

The following summaries will be provided for each co-primary efficacy endpoint:

- A summary of the change from baseline to day 7 in sputum viscosity score using the mITT population, in which the missing data will be handled using the MVTF. Two-way non-parametric ANOVA will be performed by including treatment, site, and site-by-treatment interaction terms in the model.
- A summary of the change from baseline to day 7 in expectoration difficulty score using the mITT population, in which the missing data will be handled using the MVTF. Two-way non-parametric ANOVA will be performed by including treatment, site, and site-by-treatment interaction terms in the model.

Forest plots will be generated to display the results at each center. Note that the Forest plots utilize a GLM fit to the ranked datasets used by the two-way non-parametric ANOVA. This is because it is not possible to include an interaction term (site-by-treatment) into the stratified Mann-Whitney U Statistic test utilized for analysis of the co-primary endpoints. Caution must be taken in interpreting the results of the multi-center analysis owing to difference in statistical methodology compared to the primary and secondary endpoints analysis as well as a lack of appropriate stratification for the randomization.

4.10 Safety Evaluation

All safety summaries and listings will be based upon the Safety Population (Section 4.4.4).

4.10.1 Extent of Exposure

Exposure will be summarized both in terms of the treatment duration (in days) and the total dose administered up to the end of the last administration of IMP, as recorded on the eCRF. The duration of treatment will be calculated from the first administration date to the last administration date at the end of the last administration of treatment arm (duration = the last administration date - the first administration date +1). The total dose of the treatment arm will be the summation of actual administration dose at every administration.

Investigator will assess adherence with IMP dosing regimen on an ongoing basis by determining the amount of IMP dispensed, used (i.e. returned open ampoules/vials) and the amount of IMP returned (i.e. returned, unopened ampoules/vials) at Visit 2 (first dose), Visit 3 and Visit 4 (end of treatment) or unscheduled Deterioration Visit/premature discontinuation visit.

The evaluation of compliance will be done using the following formula:

$$\%\ of\ administered\ drug = 100 \times \frac{Total\ number\ of\ administered\ doses}{Total\ number\ of\ scheduled\ doses}$$

The compliance calculation will be based on visit basis. For example, if subject takes IMP only in the morning but withdraws in the afternoon without taking IMP as planned, then the total number of scheduled doses at this visit will count two doses in the denominator.

A subject would be considered as compliant if % of administered drug is at least 80%.
The following summaries will be provided:

- A summary of the duration and the actual total dose by treatment group
- A summary of number and percentage of subjects with treatment compliance by treatment group and visit.

The following listings will be provided:

- A by-subject listing of administration of IMP
- A by-subject listing of treatment compliance

4.10.2 Adverse Events

An AE is "any untoward medical occurrence in a patient or clinical trial subject administered a medicinal product and which does not necessarily have a causal relationship with this treatment".

Adverse events will be coded using Medical Dictionary for Regulatory Activities (MedDRA) Version 23.1.

Any AE occurring onset date on or after the date of the first IMP administration, i.e. Visit 2, until the end of the trial, i.e. follow-up phone call, will be considered as a treatment emergent AE (TEAE). AEs which are already present before the first IP and increase in severity after the first IP will be considered as TEAEs. Pre-existing AEs before the first IP with no increase in severity after the first IP will not be considered as TEAEs. A Pre-treatment AE will be defined as any AE with an onset date before the date of first administration of IMP.

Imputation of partial dates of AE record is described in Appendix 4.14 (section 4.14.7).

Following summaries will be provided:

- An overview of all TEAEs including any TEAEs with relationship to IMP, any TEAEs leading to drug withdrawn, any TEAEs leading to drug interrupted, any TEAEs leading to fatal, any TEAEs of Covid-19, by treatment group.
- A summary of Pre-treatment Adverse Events by SOC, PT and CTCAE Grade (Safety Set)
- A summary of the number and percentage of subjects reporting an TEAE and the number of TEAEs, by treatment group, SOC, and PT
- A summary of the number and percentage of subjects reporting a TEAE with Incidence $\geqslant$ 5% of Patients by treatment group, SOC, and PT
- A summary of the number and percentage of subjects reporting Other adverse events with Incidence $\geqslant$ 5% of Patients by treatment group, SOC, and PT
- A summary of the number and percentage of subjects reporting a TEAE Related to IMP by treatment group, SOC, and PT

- A summary of the number and percentage of subjects reporting an TEAE and the number of TEAEs, by treatment group, relationship, SOC, and PT
- A summary of the number and percentage of subjects reporting an TEAE and the number of TEAEs, by treatment group, severity, SOC, and PT
- A summary of the number and percentage of subjects reporting a TEAE Leading to Study Discontinuation by treatment group, SOC, and PT
- A summary of the number and percentage of subjects reporting a TEAE Leading to Drug withdrawal of IMP by treatment group, SOC, and PT

For the summarization of number of subjects, a subject will be counted only once for each SOC and each PT, even if the subject reported one or more event under each subcategory; for the summarization of number of events, the number of TEAEs will count all TEAEs of each subject under each SOC and each PT.

For each TEAE, the most extreme grade recorded in eCRF will be attributed and used in the by-severity summaries. If the severity is missing, it will be imputed as the most extreme case, i.e., Severe. If relationship to study treatment is missing, the relationship will be imputed as Related.

Unless specified otherwise, the table of a summary will be ordered in terms of decreasing number of subjects for system organ class and then preferred term within system organ class in the overall group, and then alphabetically for SOC and PT within SOC if the frequency is tied.

A by-subject listing of all AEs will be provided.

4.10.3 Deaths and Serious Adverse Events

Any SAE occurring from the day of the first IMP administration, i.e. Visit 2, until the end of the trial, i.e. follow-up phone call, will be considered as a treatment-emergent SAE.

Following summaries will be provided:

- An overview of all treatment-emergent SAEs including any treatment-emergent SAEs with relationship to IMP, any treatment-emergent SAEs leading to drug withdrawn, any treatment-emergent SAEs leading to drug interrupted, any treatment-emergent SAEs leading to death, any treatment-emergent SAEs of Covid-19, by treatment arm.
- A Summary of Death
- A summary of the TEAE Leading to Death by System Organ Class, Preferred Term
- A summary of the number and percentage of subjects reporting a SAE and the number of SAEs, by treatment group, SOC and PT
- A summary of the number and percentage of subjects reporting a treatment-emergent SAE and the number of treatment-emergent SAEs, by treatment group, SOC and PT
- A summary of the number and percentage of subjects reporting a treatment-emergent SAE Related to IMP, by treatment group, SOC, and PT

- A summary of the number and percentage of subjects reporting a treatment-emergent SAE and the number of treatment-emergent SAEs, by treatment group, severity (mild, moderate, severe), SOC, and PT
- A summary of the number and percentage of subjects reporting a treatment-emergent SAE and the number of treatment-emergent SAEs, by treatment group, relationship, SOC, and PT
- A summary of the number and percentage of subjects reporting a treatment-emergent SAE leading to drug withdrawal, by treatment group, SOC and PT
- A summary of the number and percentage of subjects reporting a treatment-emergent SAE leading to Study discontinuation, by treatment group, SOC and PT

Tables should follow the similar format described for TEAEs if appropriate.

A by-subject listing of all SAEs will be provided.

4.10.4 Laboratory Evaluations

Laboratory evaluations will include Hematology, Clinical Chemistry (biochemistry), and Pregnancy test. Lab value will be converted to standard unit.

Quantitative laboratory measurements reported as '< X', i.e. below the lower limit of quantification (BLQ), or '> X', i.e. above the upper limit of quantification (ULQ), will be converted to X for the purpose of quantitative summaries, but will be presented as recorded, i.e. as '< X' or '> X' in the listings.

The evaluations of hematology and clinical chemistry (biochemistry) will be performed at Screening Visit (Visit 1) and Visit 4 and in case of premature discontinuation.

A dipstick urine pregnancy test for women of childbearing potential will be performed at Randomization Visit (Visit 2) and at Visit 4 and/or at unscheduled deterioration visit/premature discontinuation visit.

The parameters of hematology, clinical chemistry(biochemistry) and pregnancy test are presented in the following table.

Hematology:			
Hematocrit	Hemoglobin	Platelet count	Neutrophils
Lymphocytes			
Eosinophils	Basophils		
Clinical Chemistry(biochemistry):			
BUN	Chloride	AST	
Creatinine	Uric acid	ALT	
Sodium	Amylase	γGT	
Potassium	ALP	Bilirubin (direct and total)	
Calcium	UREA		
Pregnancy Test:			
Urine β-hCG (dipstick)			

The following summaries will be provided:

- A summary of hematology observed values at Visit 1 and Visit 4, and the changes from Visit 1 for each laboratory parameter by treatment group

  For the continuous parameters, it will be summarized using the descriptive statistics. For the categorical parameters, a shift table of the test results at Visit 4 compared to the test results at Visit 1 will be provided
- A summary of clinical chemistry (biochemistry) observed values at Visit 1 and Visit 4, and the changes from Visit 1 for each laboratory parameter by treatment group
- A summary of urine Pregnancy Test observed values at Visit 1 and Visit 4
- A shift table of the pregnancy test results at Visit 4 compared to the pregnancy test results at Visit 2
- Summary of subjects with values outside central laboratory normal range by visit for each treatment group (Safety Population)

The following listings will be provided:

- A by-subject listing of all laboratory data, in which all values appearing outside the laboratory normal range will be highlighted
- A by-subject list of pregnancy test results

4.10.5 Vital Signs and Physical Examination

Vital Signs

Vital Signs will be assessed at Screening Visit (Visit 1), Randomization Visit (Visit 2), Visit 3 and Visit 4 and/or at the unscheduled deterioration visit/premature discontinuation.

Vital signs include:

- heart rate (beats/minute),

- systolic and diastolic blood pressure, measured after at least 5 minutes in the supine position (mmHg)
- respiration rate (breaths/min)
- body temperature, measured under the arm pit (°C)

A summary of the observed values and changes from Visit 1 for each vital sign by treatment group and overall at each visit will be provided. Both raw values and changes from Visit 1 will be summarized using descriptive statistics.

A by-subject listing of vital signs will be provided.

Physical Examination

At Visit 1, Visit 2, Visit 3 and Visit 4, and/or at the unscheduled deterioration Visit/premature discontinuation, a physical examination of general body systems will be performed according to current medical standards and site practice. The general body systems include:

- Respiratory System
- Cardiovascular System
- Abdomen
- General Appearance
- Head and Neck
- Lymph Nodes
- Musculoskeletal System
- Nervous System
- Skin
- Endocrine System
- Ears, nose and throat
- Hematopoietic System
- Reproductive System
- Urinary System

The examination is to include a chest auscultation.

A summary of the number and percentage of subjects for each body system by category (normal, abnormal not clinically significant, and abnormal clinically significant), treatment group and overall at each visit will be provided.

A by-subject listing of physical examination will be provided.

4.10.6 12-Lead Electrocardiogram

12-lead electrocardiogram (ECG) recordings will be obtained at each study center using site machines at the Screening Visit (Visit 1) to verify the eligibility of the subject, and at Visit 4 or unscheduled deterioration visit.

ECG will be performed after obtaining vital signs. Prior to recording, the subject should be at rest for at least 5 mins.

ECG will be evaluated by the investigator/local medical staff and the recording will be reported in the eCRF as "normal", "abnormal clinically significant" or "abnormal not clinically significant".

A shift table of the ECG assessments at Visit 4 compared to the ECG assessments at Visit 1 by treatment group will be provided. When there are multiple assessments of ECG taken at a visit, the worst ECG finding will be used.

A by-subject listing of 12-lead ECG will be provided.

4.10.7 Suspected Unexpected Serious Adverse Reactions

A Suspected, Unexpected Serious Adverse Reaction (SUSAR) is a subset of SAEs that, although foreseeable, are potentially related to the IMP and not identified in the reference safety information and will be reported to the appropriate regulatory authorities and investigators following local and global guidelines and requirements. The identification and reporting of SUSARs is the responsibility of Zambon's Pharmacovigilance department. Following database lock, a listing of SUSARs will be requested from Zambon. If any occurred during the study, they will be incorporated in the TLFs.

The following summaries will be provided:

- A summary of number and proportion of subjects reporting at least one SUSAR. The corresponding two-sided 95% CI of the proportion for each treatment group using the Clopper-Pearson method
- A summary of number and percentage of subjects reporting at least one SUSAR, and the number of SUSARs presented by System Organ Class, and Preferred Term.

A by-subject listing of SUSARs will be provided.

4.11 Determination of Sample Size

This study has been powered to reach statistical significance for the superiority comparisons of N-acetyl cysteine (NAC) vs. placebo (primary study objective), and for the non-inferiority comparisons of NAC vs. ambroxol hydrochloride (key secondary study objective) in at least one of the two co-primary endpoints at the end of the 1-week treatment. This implies that sample size computation must account for the type I error inflation caused by the multiplicity over the two co-primary endpoints and by the multiplicity over the study targets (superiority and non-inferiority assessments) as well as the logical relationships among the two family of hypotheses with the non-inferiority testable only provided that the superiority contrast, assessed on the same endpoint, is statistically significant. To accomplish this, a multiple-sequence gatekeeping procedure [21] was implemented in a Monte Carlo (MC) study with 1000 runs and with the Bonferroni correction applied to the set of superiority comparisons (first gatekeeper) and

the Hochberg correction applied to the set of non-inferiority comparisons (second gatekeeper) in order to control the overall type I family-wise error rate at the one-tailed 0.025 significance level. Details on the adopted gatekeeping strategy are reported further on.

For the superiority tests, and with group sample sizes of 100 patients in placebo group and 100 patients in NAC group, the Monte Carlo study estimated an overall power ("disjunctive" power) close to 93% to reach statistical significance in at least one co-primary end-point. These calculations are based on a one-sided Mann-Whitney test and assuming a minimal clinically relevant difference of 0.35 points between placebo and NAC on a 0-3-point scale, standard deviation equal to 0.79 and 0.77 for sputum viscosity and expectoration difficulty scores (change from baseline to the 1-week treatment) respectively and a correlation between the two end-points equal to 0.31.

For the non-inferiority tests, a sample of 100 patients per each active treatment group ensures sufficient disjunctive power (Monte Carlo disjunctive power = 82%) to detect non-inferiority between NAC and ambroxol in at least one co-primary end-point. These calculations are based on a one-sided non-inferiority t-test assuming a margin of non-inferiority equal to 0.30, standard deviation equal to 0.79 and 0.77 for sputum viscosity and expectoration difficulty scores (change from baseline to the 1-week treatment) respectively and a correlation between the two endpoints equal to 0.31.

Assuming a drop-out rate of about 10% over one week of entry, a total of 333 patients will be randomized (i.e. 111 patients in the placebo group, 111 patients in the NAC group and 111 patients in the ambroxol group.

4.12 Changes to protocol-defined analyses

The following changes from protocol defined analyses were made:

- In protocol, a test of site-by-treatment interaction was planned to be carried-out using a two-way nonparametric ANOVA performed by resorting to the "Aligned Rank Transform" (ART) procedure. However, this ART procedure will not be performed in final analysis. The main reason is due to the program limitation for implementing in the SAS 9.4 and no existing and validated procedure to perform this ART analysis. From the program maintenance and quality control perspective, the study team decides to use canonical two-way nonparametric ANOVA procedure instead.

4.13 Reference

1. Wobbrock, J. O., Findlater L., Gergle D., Higgins J. J. (2011). The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only ANOVA Procedures. Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '11). Vancouver, British Columbia (May 7-12, 2011). New York: ACM Press, pp. 143-146.

2. Keays, et al. (1991). Intravenous acetylcysteine in paracetamol induced fulminant hepatic failure: a prospective controlled trial. BMJ; 303: 1026-9.
3. Rothmann, M. D., Wiens, B. L., and Chan, I. S. F. (2012). Design and Analysis of Non-Inferiority Trials. Chapman & Hall/CRC Biostatistics Series, 350-351.
4. Qu Y., Zhao Y. D., and Rahardia D. (2008). Wilcoxon-Mann-Whitney test: stratify or not? Journal of Biopharmaceutical Statistics; 18: 1103-1111.
5. Munzel U. and Hauschke D. (2003). A nonparametric test for proving noninferiority in clinical trials with ordered categorical data. Pharmaceutical Statistics; 2:31-37.

4.14 Appendix

4.14.1 Hochberg procedure

In the case the co-primary endpoints for the superiority of NAC versus placebo are statistically significant (first gatekeeper), in the second gatekeeper for the family of non-inferiority tests the Hochberg procedure for adjustment of multiple comparisons across the two co-primary endpoints will be performed. The Hochberg procedure assumes p-values are independent and uniformly distributed under respective null hypotheses and control the familywise error rate calculated in step-up fashion.

Suppose test [redacted] null hypotheses and denote the ordered p-values as $p_{(1)} \leq \cdots \leq p_{(m)}$, the Hochberg-adjusted p-values $\tilde{p}_{(1)} \leq \cdots \leq \tilde{p}_{(m)}$ are obtained from:

$$\tilde{p}_{(i)} = \begin{cases} p_{(m)} & for\ i = m \\ \min\left(\tilde{p}_{(i+1)}, (m-i+1)p_{(i)}\right) & for\ \ i = m-1, \ldots, 1 \end{cases}$$

The SAS syntax of Hochberg procedure is specified as follows:

```
PROC MULTTEST INPVALUES=<input_raw_pvalue>  OUT= <output_adjust_pvalue>
HOCHBERG;
RUN;
```

Where input_raw_pvalue= “the raw p-value calculated”.
output_adjust_pvalue= “the adjusted p-value output”.

The procedure of MULTTEST will compute the adjusted p-value and can be tested on both co-primary endpoints with an overall probability α set to 0.025 (one-tailed).

4.14.2 Bootstrap Sampling Algorithm

The algorithm to generate bootstrap samples is by using the procedure of PROC SURVEYSELECT to perform the random sampling with replacement stratified by treatment group. And the following statements will generate 1000 bootstrap samples by repeatedly drawing 333 random observations from the original data.

The SAS syntax of Bootstrap Sampling Algorithm is specified as follows:

```
%let rep = 1000; /* number of bootstrap resamples */
PROC SURVEYSELECT DATA= <input_dataset> OUT=<output_dataset>
(rename=(Replicate=SampleID)) SEED =<seed number>  OUTHITS
METHOD = urs  /* resample with replacement */
 SAMPRATE = 1 /* sample rate */
REP = &rep; /* generate NumSamples bootstrap resamples */
```

```
STRATA trt01pn ; /*stratified by treatment group*/
RUN;
```

After resampling 1000 times from the data to form 333 bootstrap samples, the efficacy analysis based on both mITT and PP populations will be performed by using either the stratified WMW test or modified WMW test on each bootstrap sample. The stratified and modified WMW syntax can be referenced in 0 and 4.14.6.

Then bootstrap distribution will be used to obtain bootstrap estimates including the point estimates of treatment differences together with associated two-sided 95% confidence intervals. Here nonparametric confidence interval can be estimated by percentiles of the bootstrap distribution. The 95% bootstrap confidence interval of SAS syntax can be referred as follow:

```
PROC UNIVARIATE DATA= <input_stats>;
  VAR stats;
  OUTPUT OUT=Pctl  PCTLPRE =CI95_
   MEAN=BootMean  STD=BootStdErr /* compute bootstrap mean and standard error */
  PCTLPTS =2.5  97.5       /* compute 95% bootstrap confidence interval */
  PCTLNAME=Lower Upper;
RUN;
```

where input_dataset = “individual efficacy data of changing from baseline”
input_stat= “Bootstrap distribution of stratified WMW statistics or modified WMW statistics”.
stats= “Computed point estimates of stratified WMW statistics or modified WMW statistics”.
trt01pn = “treatment group, 1 stands for test group; 2 stands for control group”.

4.14.3 Two-way nonparametric ANOVA procedure

A test of site-by-treatment interaction will be carried-out using a two-way nonparametric ANOVA and modelling adjustment for covariates will be performed and tested at two-tailed α=0.1 by including treatment, site, and site-by-treatment interaction terms in the model. SAS syntax can be referred as follow:

```
PROC RANK DATA=<input_dataset>  OUT= <ranked_dataset>;
  BY SITE;
  VAR response;
 RANKS ranked_response;
RUN;

 PROC ANOVA DATA=<ranked_dataset>;
   CLASS site trt01pn;
        MODEL ranked_response = site trt01pn site* trt01pn;
```

```
RUN;
```

To obtain the treatment difference by site for the forest plot, SAS syntax can be referred as below:

```
PROC GLM DATA=<ranked_dataset>;
   CLASS site trt01pn ;
        MODEL ranked_response = site trt01pn site* trt01pn;
        LSMEANS site*trt01pn /cl PDIFF STDERR;
RUN;
```

Where trt01pn = "treatment group, 1 stands for test group; 2 stands for control group".
response = "change from baseline to day 7 in Sputum viscosity" or "change from baseline to day 7 in Expectoration difficulty"

4.14.4 EM Algorithm

The expectation-maximization (EM) algorithm is a technique for maximum likelihood estimation in parametric models for incomplete data. The EM algorithm computes the MLE for the means and covariance matrix, of a multivariate normal distribution from the input data set with missing values.

The SAS syntax of EM algorithm is specified as follows:

```
PROC MI DATA=<input_dataset> SEED= <Seed number> NIMPUTE=0;
        EM OUTEM=out_data;
        VAR spv epd spc cough sv;
RUN;
```

Please note that the imputed values will round to the nearest integer. For value range of sputum viscosity, expectoration difficulty, sputum color and cough severity score will be integer from 0 to 3. The range of Sputum volume imputed value can be non-negative number.

Where input_data= "Individual efficacy data"

Seed number= "Random Seed"
spv="Sputum viscosity"
epd="Expectoration difficulty"
spc="Sputum Color"
cough = "Cough severity score"
sv="sputum volume"

4.14.5 The Stratified Wilcoxon-Mann-Whitney Test Statistic

Let $x_k = (x_{k,1}, \dots, x_{k,m_k})$ and $y_k = (y_{k,1}, \dots, y_{k,n_k})$, $k = 1,..,d$, be the d pairs (strata) of independent random samples from distribution F_k and G_k, respectively, where d is the number of blocks of baseline score. The x-values are from the active test group, and the y-values are from the placebo group.

The null hypotheses of superiority test can be expressed as

$$H_0: \pi = 0.5$$

where $\pi = \sum_{k=1}^{d} c_k Pr(Y_k > X_k)$, X_k and Y_k are random variables with distribution F_k and G_k, respectively, and $c_k = m_k n_k (m_k + n_k + 1)^{-1}$.

The stratified WMW test statistic (van Elteren statistic) is computed as

$$T = \frac{\hat{\pi} - 0.5}{\sqrt{v_0}}$$

where

$$\hat{\pi} = \sum_{k=1}^{d} c_k W_k$$

$$W_k = (m_k n_k)^{-1} \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} I(y_{k,j} > x_{k,i})$$

and

$$v_0 = \sum_{k=1}^{d} c_k^2 \, (12 m_k n_k)^{-1} (m_k + n_k + 1).$$

The detail can be seen in Qu et al. (2008).

The SAS syntax of stratified WMW test (Van Elteren test) is specified as follows:

```
PROC NPAR1WAY DATA = <input_dataset> ;
CLASS trt01pn ;
STRATA  baseline_score;
VAR Change_from_baseline;
RUN;
```

where input_data = “Individual efficacy data of changing from baseline”

trt01pn = “treatment group, 1 stands for test group; 2 stands for control group”
baseline_score = “The following Baseline score will be implemented separately, including sputum viscosity score, expectoration difficulty score, sputum color score, cough severity score and sputum volume”

Change_from_baseline= "The response will consider : (1) Change from baseline to day 3 (2) Change from baseline to day7. For each condition, the following change from baseline score will be considered: sputum viscosity score, expectoration difficulty score, sputum color score, cough severity score, sputum volume."

4.14.6 The Modified Wilcoxon-Mann-Whitney Test Statistic

Let $x = (x_1, \ldots, x_{n_T})$ and $y = (y_1, \ldots, y_{n_C})$ denote the independent random samples from distribution F_C and F_T, respectively. The x-values are from the active test group, and the y-values are from the active control group.

The null hypotheses of non-inferiority test are expressed as

$$H_0: p \leq 0.5 - \delta$$

where $p = Pr(X < Y) + 0.5 \times Pr(X = Y)$, and the value for δ is the non-inferiority margin.

The unbiased estimator of p is given by

$$\hat{p} = {n_T}^{-1}(\bar{R}_C - (n_C + 1)/2)$$

where $\bar{R}_C$ is the arithmetic average of the rank of the observation in the active control group among all observations. That is $\bar{R}_C = \sum_{j=1}^{n_C} R(y_j)/n_C$, where $R(y_j)$ is the rank of y_j in ordering of the combined sample. Define also $\bar{R}_T = \sum_{j=1}^{n_T} R(x_j)/n_T$, where $R(x_j)$ is the rank of x_j in the ordering of the combined sample. When ties occur, $R(y_j)$ is the midrank, as well as $R(x_j)$. Let $R^{(C)}(y_j)$ denote the rank of y_j among $y_1, \ldots, y_{n_C}$, and let $R^{(T)}(x_j)$ denote the rank of x_j among $x_1, \ldots, x_{n_T}$. Define

$$J_T^2 = \sum_{j=1}^{n_T} \left(R(x_j) - R^{(T)}(x_j) - \bar{R}_T + (n_T + 1)/2\right)^2/(n_T - 1),$$

and

$$J_C^2 = \sum_{j=1}^{n_C} \left(R(y_j) - R^{(C)}(y_j) - \bar{R}_C + (n_C + 1)/2\right)^2/(n_C - 1).$$

Furthermore, for $i = T, C$, define $\hat{u}_i^2 = J_i^2/(N - n_i)^2$, where $N = n_T + n_C$. The test statistics is given by

$$Q = \frac{\hat{p} - (0.5 - \delta)}{\sqrt{\frac{\hat{u}_C^2}{n_C} + \frac{\hat{u}_T^2}{n_T}}}.$$

The detail can be seen in (Munzel and Hauschke, 2003) as well.

Since for using Munzel and Hauschke approach, a non-inferiority margin expressed on a probability scale is needed, the non-inferiority margin (as used in the sample size estimation) needs to be converted from a point scale (since it is expressed in the same scale of data) to a probability scale.
The converted non-inferiority margin can be derived as below.

To derive δ''(probability scale margin) from δ'(point scale margin) we apply the following formula of Cohen D effect size by assuming normality between two effect size:

$$\delta'' = \delta' \times \frac{SD(\delta'')}{SD(\delta')}$$

Where $\delta' = 0.30$; $SD(\delta') = \sqrt{\sigma_c^2 + \sigma_T^2}$; $SD(\delta'') = \sqrt{\mu_c^2 + \mu_T^2}$

- SD is the term used here to indicate the standard deviation associated to statistic based on δ' and δ''
- σ_c and σ_T are the standard deviation of the changes from baseline of the original scores for control group and test group respectively
- μ_c and μ_T are estimated according to the formula as shown above.

The SAS syntax of obtaining the probability margin converting from point scales and modified Wilcoxon-Mann Whitney test are specified as follows.

```
/*R(x j ) and R(y j )*/
proc rank data= input_data  out=out1 ties=mean;
   var chg;
   ranks r_chg;
run;

data out1t;
      set out1;
      where trt01pn=1;
run;

data out1c;
      set out1;
      where trt01pn=2;
run;

/*R^((T) ) (x_j )*/
proc rank data=input_data (where=(trt01pn=1)) out=out_rt ties=mean;
   var chg;
   ranks r_chg_rt;
```

```
run;

/*R^((C) ) (y j )*/
proc rank data=input_data (where=(trt01pn=2)) out=out_rc ties=mean;
   var chg;
   ranks r_chg_rc;
run;

/*R mean T and  n_T*/
proc sql;
      select mean(r chg), count(r chg) into: avg_rt, :n_t
      from out1(where=(trt01pn=1));
quit;

%put &avg_rt &n_t;

/* R_mean C and  n_C*/
proc sql;
      select mean(r_chg), count(r_chg) into: avg_rc, :n_c
      from out1(where=(trt01pn=2));
quit;

%put &avg_rc &n_c;

/*JT2*/
proc sql noprint;
 select sum(((r_chg - r_chg_rt - &avg_RT + (&n_T+1)/2)**2)/(&n_T-1))
into :JT2
 from out1t as a
 left join out rt as b
 on a.usubjid = b.usubjid;
quit;
%put &JT2;

/*JC2*/
proc sql noprint;
 select sum(((r_chg - r_chg_rc - &avg_RC + (&n_C+1)/2)**2)/(&n_C-1))
into :JC2
 from out1c as a
 left join out_rc as b
 on a.usubjid = b.usubjid;
quit;
%put &JC2;

/*UT2 and UC2*/
%let UT2=%sysevalf(&JT2/(&n_C**2));

%let UC2=%sysevalf(&JC2/(&n_T**2));

%put &UT2; %put &UC2;

/*Std_RC and Std_RT*/

proc sql;
      select std(chg) into : std_rc
      from out_rc;
quit;
```

```
%put &std_rc;

proc sql;
      select std(chg)  into : std_rt
      from out_rt;
quit;

%put &std_rt;
```

All above calculated variables will be used to calculate the δ″to implement the following modified Wilcoxon-Mann-Whitney tests.

The SAS syntax of modified Wilcoxon-Mann-Whitney test to compute test statistics and one side confidence intervals are specified as follows:

```
proc freq data= input_data;  /*somers d statistics*/
      table trt01pn*chg /measures;
      test smdcr;
      output smdcr out=smd_out;
run;

data outf;
      set smd_out;
      p 1=(1+ smdcr )/2; /* unbiased estimator of p */
      e_p1=e_smdcr/2;
      keep _smdcr_ e_smdcr p_1 e_p1;
run;

data test;
      set outF;
      delta1=0.3;
      delta2= delta1*sqrt(&UT2+&UC2)/sqrt(&Std_RT**2+&Std_RC**2) ;
      p 10=0.5- delta2 ; /* defined margin*/
      Z p1=(p_1-p_10)/e_p1; /* modified Wilcoxon-Mann-Whitney test
 statistics*/

      p_p1=1-probnorm(Z_p1); /*p value for P_1*/

      lower limit 1=p 1-1.96* e p1; /* Depends on results of primary
 endpoint, here is the sample code of one side 97.5% lower CI of
 modified Wilcoxon-Mann-Whitney test statistics */
      upper limit_1=1;  /* Upper CI of modified Wilcoxon-Mann-Whitney
test statistics*/

       lower limit 2=p 1-2.24* e p1; /* Depends on results of primary
 endpoint, here is the sample code of one side 98.75% lower CI of
 modified Wilcoxon-Mann-Whitney test statistics  */
      upper limit_2=1;  /* Upper CI of modified Wilcoxon-Mann-Whitney
test statistics*/

 run;
```

where input_data = “individual efficacy data of changing from baseline”

trt01pn = “treatment group, 1 stands for test group; 2 stands for control group”
chg = “change from baseline score of ordinal categorical 4-point scales starting from 0(normal) to 3(worst) of sputum viscosity score, expectoration difficulty score, sputum color score, and cough severity score”

4.14.7 Partial Date Conventions

Imputed dates will not be presented in the listings.

However, in general, when calculating relative days, partial dates with missing day only will be assumed to be 15th of the month, and partial dates with both missing day and month will be assumed to be June 30.

Otherwise, the following rules in the given table will be applied for each case.

4.14.7.1 Algorithm for Prior/Concomitant Medications

Start Date	Stop Date	Action
Known	Known	If start and stop dates are prior to the date of first dose of study medication, considered as prior; If stop date is on or after the date of first dose of study medication, considered as concomitant.
	Partial	The last day of the month will be used if only the day of stop date is missing; Date of last dose of study medication will be used if day and month of stop date is missing; December 31st will be used if day and month of stop date is missing and the year of the year of last dose of study medication is greater than the year of stop date. If start date and the imputed stop date is prior to the date of first dose of study medication, considered as prior; if the imputed stop date is on or after the date of first dose of study medication, considered as concomitant.
	Missing	Stop date will not be imputed. Considered as concomitant.

Start Date	Stop Date	Action
Partial	Known	January 1st will be used if the start day and month is missing and the year of first dose of study medication is not the same as the year of the start date; the date of first dose of study mediation will be used if the start day and month is missing and the year of first dose of study medication is the same as the year of the start date; the date of first dose of study mediation will be used if only the day of start date is missing, and the month and year of start date is the same as the first dose of study medication; The first day of the month will be used if only the date of start date is missing, and the month and year of start date is not the same as the first dose of study medication. If the imputed start date and stop date is prior to the date of first dose of study medication, considered as prior; if the stop date is on or after the date of first dose of study medication, considered as concomitant.
	Partial	The same imputation rule of partial start date when start date is partial, stop date is known will be applied; the same imputation rule of partial stop date when start date is known, stop date is partial will be applied. If the imputed start date and stop date is prior to the date of first dose of study medication, considered as prior; if the imputed stop date is on or after the date of first dose of study medication and is up to the last dose of study medication, considered as concomitant.
	Missing	The same imputation rule of partial start date when start date is partial, stop date is known will be applied; The stop date will not be imputed. Considered as concomitant.

Start Date	Stop Date	Action
Missing	Known	The date of first dose will be used if stop date is not prior to the date of first dose of study medication. Start date will not be imputed if stop date is prior to the date of first dose of study medication. If stop date is prior to the date of first dose of study medication, considered as prior; if stop date is on or after the date of first dose of study medication, considered as concomitant.
	Partial	Start date will not be imputed; the same imputation rule of partial stop date when start date is known, stop date is partial will be applied. If the imputed stop date is prior to the date of first dose of study medication, considered as prior; if the imputed stop date is on or after the date of first dose of study medication, considered as concomitant.
	Missing	The date of first dose will be used to impute start date; the stop date will not be imputed. Considered as concomitant.

Note: If the recorded start date or imputed start date of the medication is after the date of last dose of study drug, it will not be classified to prior or concomitant medications, just set the category to missing.

4.14.7.2 Algorithm for Adverse Events (include TEAE and pre-treatment AE)

Start/ Increase Severity Date	Stop Date	Action
Known	Known	Considered as a TEAE if start date is on or after the date of the first dose of study drug.
	Partial	Considered as a TEAE if start date on or after the date of the first dose of study drug. The last day of the month and the last month (i.e. December) will be used if the stop day/month is missing.
	Missing	Considered as a TEAE if start date on or after the date of the first dose of study drug.
Partial, but known components show that it cannot be on or after first study drug taken date	Known	Not a TEAE. The first day of the month and January will be used if the start day/month is missing.
	Partial	Not a TEAE. The first day of the month and January will be used if the start day/month is missing. The last day of the month and the last month (i.e. December) will be used if the stop day/month is missing.
	Missing	Not a TEAE. The first day of the month and January will be used if the start day/month is missing.
Partial, could be on or after first study drug taken date	Known	Considered as TEAE, if stop date is after first study drug taken date. The first study drug taken date will be used if start date is in the same month/year with first study drug taken date, or the first day of the month and January will be used if the start day/month is after first of study drug taken date Considered as not TEAE, if stop date is prior to first of study drug

Start/ Increase Severity Date	Stop Date	Action
		taken date. The first day of the month and January will be used if the start day/month is missing.
	Partial	Considered as TEAE. The first of study drug taken date will be used if start date is in the same month/year with first study drug taken date, or the first day of the month and January will be used if the start day/month is after first study drug taken date. The last day of the month and the last month (i.e. December) will be used if the stop day/month is missing.
	Missing	Considered as TEAE. The first study drug taken date will be used if start date is in the same month/year with first study drug taken date, or the first day of the month and January will be used if the start day/month is after first study drug taken date.
Missing	Known	Considered as TEAE if stop date is on or after the date of the first dose of study drug.
	Partial	The last day of the month and the last month (i.e. December) will be used if the stop day/month is missing. If the imputed stop date is on or after the first dose of study drug considered as a TEAE; if the year is missing, considered as a TEAE.
	Missing	Considered as a TEAE

PAREXEL International Electronic Signature Page

This page is the manifestation of the electronic signature(s) used in compliance with PAREXEL International's electronic signature policies and procedures and in compliance with applicable regulations.

UserName: PPD
Title: PPD
Date: Thursday, 15 April 2021, 07:20 AM GMT Standard Time
Meaning: Document contents approved.
==========