STATISTICAL ANALYSIS PLAN

A Phase III, Prospective, Multicenter, Randomized, Controlled Clinical Trial to Demonstrate the Efficacy and Safety of Liposomal Cyclosporine A (L-CsA) Inhalation Solution Delivered via the [redacted] Device plus Standard of Care versus Standard of Care Alone in the Treatment of Chronic Lung Allograft Dysfunction / Bronchiolitis Obliterans Syndrome in Patients post Double Lung Transplantation

Protocol No: BT – L-CsA – 302 – DLT(BOSTON-2)

Statistical Analysis Plan Status: Final

Statistical Analysis Plan Version: 5.0

Statistical Analysis Plan Date: 27MARCH2024

SPONSOR

Zambon SpA

Via Lillo del Duca, 10

Bresso (MI) 20091, Italy

INVESTIGATIONAL MEDICINAL PRODUCT

Liposomal Cyclosporine A (L-CsA)

CCI [redacted]

Developed by PPD

1 SIGNATORIES

SPONSOR

Zambon SpA

Via Lillo del Duca, 10

Bresso (MI) 20091

Italy

PPD

PPD

Phone: PPD

Email: PPD

_____

Date Signature

PPD

PPD

Phone: PPD

Email: PPD

_____

Date Signature

PPD

PPD

Phone: PPD

Email: PPD

_____

Date Signature

2 TABLE OF CONTENTS

3 ABBREVIATIONS

AC	Adjudication Committee
AE	Adverse Event
AIC	Akaike Information Criterion
ALT	Alanine Aminotransferase
ANCOVA	Analysis of Covariance
AP	Alkaline Phosphatase
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Code
ATG	Antithymocyte Globulin
BLAD	Baseline Lung Allograft Dysfunction
BOS	Bronchiolitis Obliterans Syndrome
BUN	Blood Urea Nitrogen
CLAD	Chronic Lung Allograft Dysfunction
CMV	Cytomegalovirus
CNI	Calcineurin Inhibitor
COPD	Chronic Obstructive Pulmonary Disease
COVID-19	Coronavirus Disease of 2019
CSR	Clinical Study Report
CTC	Common Toxicity Criteria
DLT	Double Lung Transplant
DMC	Data Monitoring Committee
DRM	Data Review Meeting
EBV	Epstein Barr Virus
eCRF	Electronic Case Report Form
EoS	End of Study
EoT	End of Treatment
EQ-5D-5L	Euro QOL Health Questionnaire
FAS	Full Analysis Set
FDA	Food and Drug Administration
CCI	[redacted]
FEV_1	Forced Expiratory Volume in One Second
FVC	Forced Vital Capacity
γ-GT	Gamma Glutamyl Transferase
IA	Interim Analysis
IACP	Interim Analysis Communication Plan
IA-FAS	Interim Analysis Full Analysis Set
ICH	International Conference on Harmonization
IE	Intercurrent Event

IMP	Investigational Medicinal Product (study drug: L-CsA)
L	Liter
L-CsA	Liposomal Cyclosporine A Inhalation Solution
LMM	Linear Mixed Model
LSM	Least Square Means
MAR	Missing at Random
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MI-RD	MI from Retrieved Discontinued patients
mL	Milliliter
MMF	Mycophenolate Mofetil
MMRM	Mixed Model of Repeated Measures
MNAR	Missing Not at Random
mTOR	Mammalian Target of Rapamycin
PD	Protocol Deviation
PEF	Peak Expiratory Flow
PPS	Per Protocol Set
PT	Preferred Term
REML	Restricted Maximum Likelihood
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAF	Safety Analysis Set
[redacted]	[redacted]
SD	Standard Deviation
SDTM	Study Data Tabulation Model
SLT	Single Lung Transplant
SoC	Standard of Care
SOC	System Organ Class
TEAE	Treatment-Emergent Adverse Event
TFL	Table, figure and listing
USA	United States of America
VAS	Visual Analogue Scale
WHO-DD	World Health Organization-Drug Dictionary
WRO	Written Response Only

4 VERSION HISTORY

Date	Version	Brief Description of Changes
28 July 2018	1.0	Initial draft version
15 October 2021	2.0	Updates on the interim analysis, introducing the IA-FAS population, and updates from the protocol amendment 3, dated 19 January 2021. The aim was to allow for the interim analysis process to proceed, with a view to update the rest of the sections in the next version of the SAP. Added separated sections to describe IA objectives and analysis.
31 July 2023	3.0	Updates reflecting the protocol amendment 5, dated 7 April 2023 (V6). The aim was to specify the re-screening procedures as per current procedure since the beginning of the study, CCI [redacted], CCI [redacted], to clarify the need to identify only change in terms of a deterioration of BOS grade (for the time to BOS progression endpoint), CCI [redacted]. Additional explanations have been provided regarding the results of the interim analysis and the final decision on the appropriateness of the primary analysis through LMM. Further details on: - Handling of missing data; - Analysis of immunosuppressants; - Analysis of CCI [redacted] safety endpoints. The methods used for the primary analysis have been further explained and CCI code excerpts have been added.

Date	Version	Brief Description of Changes
		Additional sensitivity and supportive analyses have been included to evaluate the reliability of the assumptions made in the primary analysis. The previous sensitivity and supportive analyses have also been verified and expanded upon. Supportive analyses to assess the impact of COVID-19 pandemic on study results have been developed. It is anticipated that if any evidence emerges that the pandemic has affected the study endpoints, then it is necessary to consider how the initial study assumptions have changed and apply changes in the statistical criteria for final judgment of the study results. The present amendment also includes minor updates on section referencing and sentencing restructuring.
31 October 2023	4.0	Section 9.2.1 has been modified to clarify the definition of actual treatment in the SAF population. Additional explanations have been added in Section 12.1 to clarify the calculation of intervals between visits and in Section 14.1.1 (Table 5) to summarize the different approaches for handling intercurrent events. In Section 14.2.5 a ranked ANCOVA has been planned in case of violation of normality assumption even after log-transformation of data. Sections 14.3.3.1, 14.3.5, and 14.3.6 have been further elaborated to provide additional details. Sensitivity analyses have added in sections 14.4.1, 14.4.2, 14.5.2, and 14.5.3. Additional subgroup analyses (by use of azithromycin at baseline and by BLAD disease at baseline) have been added in Section 14.6. The BLAD definition has been added in section 11.3. In section 15.2 more instructions for the analysis of Extent of Exposure have been added, while in Section 15.3.1 it has

Date	Version	Brief Description of Changes
		been added an analysis of infection rates by year and overall.
27 March 2024	5.0	CCI [redacted], the following changes have been applied: • Section 14.1.1 has been amended to consider a treatment composite strategy instead of a treatment policy one. Therefore, death and re-transplantation IE are included in the endpoint through an imputation of FEV1 as a treatment failure (FEV1 equal to zero). This applies also to 14.1.3, 14.2.1, 14.2.2, 14.2.3, 14.2.5, 14.3.3.3, 14.3.4, 14.3.5, 14.3.6, 14.4.1, and Table 5. • SAP text has been revised to better separate the concept of missing data and intercurrent events. • Table 5 has been updated for clarifying how study discontinuation for lack of efficacy are treated in the trial. Clarifications in section 14.3.3.1, 14.3.3.2, and 14.4.2 have been added to specify analysis, comparisons, and figures. It is clarified in section 14.4.2 that BOS progressions will be considered up to study end, even after treatment completion/discontinuation. It is specified in section 14.5.1 that in case of not enough responses for using logistic regression, then exact methods will be used for CI and test. Section 15.5.1 has been updated with the full list of Hematology and Biochemistry exams. Minor errors in the CCI [redacted] have been fixed in section 17.

5 INTRODUCTION

This Statistical Analysis Plan (SAP) provides details of specific planned analyses to be performed on the data to be collected from clinical trial BT – L-CsA – 302 – DLT (BOSTON-2). This SAP is based on the Clinical trial protocol version 6.0 (amendment 5.0), dated 07 April 2023.

Modifications to this SAP will be subject to detailed documentation relating to timing and rationale, with a focus on identifying and providing a rationale for changes. Any deviations from the SAP will be fully described in the integrated clinical study report (CSR).

The main purpose of this SAP is to pre-specify statistical analyses that relate to identified clinical trial objectives. Any supportive or exploratory analyses presented in the integrated CSR which are not described in the SAP will be noted as such.

6 TRIAL OBJECTIVES

The objective of this trial is to assess the efficacy and safety of add-on aerosolized Liposomal Cyclosporine A Inhalation Solution (L-CsA) to Standard of Care (SoC) therapy as compared to SoC therapy alone in the treatment of bronchiolitis obliterans syndrome (BOS) in double lung transplant (DLT) recipients.

7 INVESTIGATIONAL PLAN

7.1 Trial Design

This is a Phase III, prospective, multicenter, randomized, controlled clinical trial of L-CsA for the treatment of BOS in adults with clinically defined BOS (CLAD – BOS phenotype) following DLT. Patients will receive either L-CsA (10 mg) via the CCI twice daily plus SoC treatment or SoC alone, for a period of 48 weeks. All patients will be eligible to continue in an open-label extension trial of L-CsA, BT – L-CsA – 303 – FU (BOSTON-3) following completion of BOSTON-2.

A total of 11 visits (Screening, V1 through V10) will be performed during the clinical trial. After informed consent has been obtained, a Screening Visit will be carried out in order to check general eligibility for participation. If a patient fails the screening process due to not meeting inclusion/exclusion criteria, a re-screening is allowed. At the Baseline Visit (V1, randomization visit), inclusion and exclusion criteria will be re-checked, and baseline serial spirometry will be performed. During the 48-week treatment period, visits are scheduled every 4-8 weeks (V2, …, V9). Visit 9 (End of Treatment, EoT) is scheduled to occur 48 weeks after Visit 1. If a patient has an event that meets one of the criteria for progression of BOS, progression of BOS must be confirmed by measurements that are taken with CCI spirometer at least 2 weeks apart.

Visit 10 (End of Study, EoS) is a safety follow-up visit performed 4 weeks after Visit 9/EoT only in patients not rolling over to the extension study (BOSTON-3) or if possible, in patients who decided to withdraw the consent during the study. For patients who will enroll to BOSTON-3, the EoT visit will be the EoS visit.

However, if one of the visits from Visit 2 to Visit 8 and discontinuation visits cannot be performed at site due to COVID-19, remote visits (e.g., by telephone) are possible.

7.2 Randomization and Stratification

At the Baseline visit (V1), after all inclusion and exclusion criteria have been fully evaluated, eligible patients will be randomized with equal probability to one of two treatment arms, a control arm (SoC alone) and an investigational arm (L-CsA plus SoC). The investigational medicinal product (IMP) is L-CsA. To assure balance between treatment arms with regard to key variables and potential confounders, stratification prior to randomization will be performed for the following variables:

- Screening FEV_1 ≥81% of personal best FEV_1 value post-transplant versus Screening FEV_1 between 80-51% of personal best FEV_1 value post-transplant;
- Age at the time of randomization: <55 years versus ≥55 years;
- Region: North America versus all other countries together.

Within each of the 8 strata, patients will be randomly assigned with equal probability (1:1) to either the investigational or control arm using a permuted blocks randomization. The randomization code, including specification of the block size, will be generated by a statistician not otherwise involved with the trial, using validated software. The unique randomization number and treatment assignment of each patient will be provided to the Investigator through the electronic Case Report Form (eCRF) after eligibility has been verified. The number assigned to a patient will be that patient's unique identifier throughout the clinical trial and cannot be re-assigned to another patient.

This is an open-label clinical trial. Clinical trial monitors, treating physicians, study nurses, study coordinators, and enrolled patients will not be blinded to treatment assignment.

7.3 Primary Efficacy Endpoint

- Mean change in FEV_1 (L) from baseline to Week 48

FEV_1 data collected from the on site CCI spirometer will be considered primary, while the data collected with the CCI home spirometer will be used for supportive analyses.

7.4 CCI

- CCI

- CCI

7.6 Safety Endpoints

- Adverse events (AEs) (including infections and malignancies);
- Acute tolerability of IMP (L-CsA);
- Clinical laboratory variables;
- Renal function;
- Vitals signs;
- Physical examinations.

8 STATISTICAL CONSIDERATIONS

8.1 General Considerations and Descriptive Statistics

All report outputs will be produced using [redacted] version 9.2 or later, in a secure and validated environment. All tables, figures and data listings (TFLs) to be included in the report will be independently checked for consistency and integrity.

The data recorded in the eCRF will be summarized as appropriate, per treatment group and for the total group, using the following descriptive measures:

- All continuous variables will be summarized using descriptive statistics (i.e., n, mean, standard deviation (SD), median, Q1, Q3, minimum, and maximum).
- All categorical variables will be summarized by using frequency counts and percentages.
- Discrete ordinal or higher-level data will be summarized as continuous data, but a tabulation of categories may also be included, depending on the number of categories.
- For time-to-event variables, cumulative event-free survival times will be evaluated using the Kaplan-Meier method. A frequency table showing the number of events and censored patients by treatment group, will be presented along with the event-free rates (and two-sided 95% confidence). Kaplan-Meier graphs will be presented along with the number of patient-at-risk at exact time points. If different rules are not specified, patients who are free from event will be censored at the study termination.

For all measures analyzed descriptively (i.e., measures to characterize the treatment groups and their baseline status, efficacy, and safety outcomes), tabulations comparing the treatment groups will be prepared. For outcomes measured more than once, the analyses will include tabulations of the measured values by visit, as well as change from baseline. Where provided, all confidence intervals will be two-sided 95% intervals.

8.1.1 Baseline Definition

CCI [redacted] baseline FEV_1: the FEV_1 value that is the mean of the best FEV_1 obtained with the CCI [redacted] study spirometer at the Screening Visit and the pre-randomization best FEV_1 obtained at the Baseline Visit (V1) is referred to as "baseline FEV_1". If the value at Screening visit is missing, then the baseline FEV_1 will be the pre-randomization value from the Baseline Visit (V1). If the pre-randomization value at Baseline Visit (V1) is missing, then the baseline FEV_1 will be the value from the Screening Visit.

Other CCI [redacted] baseline spirometry values: will be defined similarly as for Baseline FEV_1.

CCI [redacted] baseline FEV_1: will be defined similarly as for CCI [redacted] baseline FEV_1 but using the CCI [redacted] home spirometer.

Baseline laboratory values: the last measurement obtained prior to or at the Baseline Visit (V1).

Other baseline measurements: unless otherwise specified, baseline refers to measurement obtained at the Baseline Visit (V1).

8.1.2 Treatment Definition

Treatment refers to add-on L-CsA to SoC therapy, as well as SoC therapy alone. Treatments periods are defined in section 15.2.2.

8.2 Sample Size Considerations

The initial sample size was based on the following considerations. Assuming a treatment group difference for mean FEV_1 change from baseline of 200 mL (0.2 L) and a standard deviation of 350 mL (0.35 L) at the end of treatment Week 48, an independent samples t-test with a one-sided type I error level of α=0.025 will have at least 80% power to reject the null hypothesis when the sample size in each treatment group is at least 50 patients. To compensate for a loss of power due to premature withdrawals, a total of 110 patients will initially be randomized.

An interim analysis (IA) was performed to re-assess sample size and delivered to the Sponsor on December 1st, 2021. Results of the IA are presented in a separate IA report (section 19.1).

In accordance with the amended version of the protocol (version 6.1)[redacted]

It was anticipated that approximately 160-170 patients have been randomized in BOSTON-2 and approximately 60 patients in BOSTON-1. 169 patients were actually randomized in BOSTON-2 and 62 patients in BOSTON-1 .

8.3 Interim Analysis

An IA was to be conducted at the time when approximately 54 patients (50%) had at least completed Week 24 (for the rationale please refer to section 8.3.1), which is referred as IA Full Analysis Set (IA-FAS) population (refer to section 9.2.3 for the IA-FAS population definition), but only including patients who either had at least completed Week 24 assessments or had discontinued (at any timepoint) at the cutoff date, 30-Sep-2021.

Rejected spirometries were not to be analyzed and were excluded from the interim analysis.

The interim analysis was to be performed with the following objectives:

- Re-estimation of the SD of the primary endpoint, with the purpose of re-estimating the sample size.
- Evaluation of the actual amount of FEV_1 data collected at all timepoints using the CCI spirometer.
- Evaluation of the number of drop-outs and the number and pattern of the missing data.
- Appropriateness of the analysis based on the anticipated longitudinal model versus the analysis based on the analysis of covariance (ANCOVA) model in the light of the collected data.

All analyses were performed on the blinded data available at that time by an independent, blinded statistician.

For the study, a Blinding Charter describing the communication process and conduct of professional activities in order to ensure the proper blinding of the study personnel was in place.

In addition to the Blinding Charter, an Interim Analysis Communication Plan (IACP) had been produced specifically for the IA (please refer to Section 19.2). The IACP in particular aimed at describing the processes and procedures that were to be employed at the time of the interim analysis and the resulting controlled dissemination of interim analysis data. In order to control blinding, IACP clearly defined data access, roles, communication flows and procedures.

8.3.1 Interim Analysis for Sample Size Re-estimation

Given the uncertainty in the estimate of the SD of FEV_1 change from baseline at Week 48 due to differences in patient populations between this trial and that upon which the estimate was obtained, a blinded IA was to be conducted by an independent, blinded statistician to re-estimate the SD.

Kieser and Friede [1] have demonstrated that the type I error rate of the t-test is not affected if simple blind variance estimators are used for sample size recalculation. Hence, no adjustment to the type I error rate was to be made. If the sample size calculated on the basis of the new value for the SD is higher than the initial estimate, the sample size of the trial might be increased accordingly in order to preserve the pre-specified power (80%). If the re-calculated sample size is lower than the initial estimate, the clinical trial was to be continued with the initially planned sample size (110 patients).

According to the previous versions of the protocols, this IA was to be conducted approximately one or two months before the last patient was projected to be randomized in the clinical trial. However, at the time of SAP amendment version 2.0, enrolment of last patient was projected by March 2022. It was decided to prepare the IA a few months earlier CCI for the following reasons:

- The number of patients to be considered in the IA was CCI approximately 50% of the originally planned sample size, which is a common choice for conducting IA for sample size re-assessment
- Study enrolment was halted starting from CCI for all participating sites and it was re-activated site by site after the approval of the protocol version 3.0 (amendment 2.0) including the COVID-19 specific measures and contingency plan starting from CCI as a date of first site re-activated. Therefore, waiting a few more months would only marginally increase the number of patients with a long follow-up duration (for example, patients with 48 weeks of observation)
- If a sample size increase was needed, having this information earlier would render feasible to amend the study protocol, without pausing patient enrolment again.

For ongoing patients who were still at their early phases of the trial, it was not considered to be clinically meaningful to report a value that was heavily based on extrapolation. Therefore, the IA was conducted on the patients who had at least completed Week 24 (Visit 6), which was a halfway of their treatment. CCI

An independent statistician had been provided with the TFL outputs blinded to treatment assignment which included FEV_1 outcomes and the covariate data to re-estimate the SD and the results can be found in the interim analysis report (see section 19.1).

8.3.1.1 Data handling

For patients who were ongoing at or after Week 24 of the study (Visit 6) but still not reaching Week 48 at the time of the IA, their study status was to be randomly assigned to either the completer population or drop-outs population, in such a way that the drop-out pattern was as much as possible equal to the drop-out pattern observed at the time of the IA (see description below). The following proportions were to be calculated:

- Number of patients discontinued after Week 24 (Visit 6) at the time of IA;
- Number of patients who completed the study at the time of IA.

The proportions were then to be applied to all ongoing patients in the study and to assign a patient to either the completer population or drop-outs population, two random processes were to be implemented:

1. Patients completing Week 24 (Visit 6) but still not reaching Week 48 at the time of the IA or discontinued were to be randomly assigned either to the completer population or drop-outs population based on the proportion of the observed data at the time of the IA. This selection was performed using CCI. The dataset with complete study status is referred to as the complete data which was to be repeatedly generated for 100 times.

2. For patients who were to be assigned to the drop-outs population in each of the complete dataset, their last visit was to be randomly chosen from options of later visits among patients discontinued after Week 24 (Visit 6). If there were no later visits observed at the time of the IA, the patients were assumed to be discontinued at their last known visit.

8.3.1.2 Handling of Missing data

Following the random process, the additional visits were presented as missing rows which were handled based on the below rules:

- The same set of baseline covariates (e.g., age, region, baseline FEV_1) for the same patients were carried over to the missing rows;
- Missing data on temporal variables were placed by nominal time (see section 9.4).

Missing data on dependent variable, change in FEV_1 at Week 48 from baseline, were imputed based on multiple imputation model assuming missing at random (MAR) mechanism. The set of variables included in the multiple imputation model were region, age group, baseline FEV_1 and temporal (time) variables. The number of imputed datasets generated was 100. The number of datasets generated were 10000 with 100 imputations within each of the 100 complete datasets.

8.3.1.3 Linear Mixed Model

A linear mixed model (LMM) for repeated measurements, similar to the analysis of the primary endpoint (section 14.1) was performed to model FEV_1 measurements for each complete dataset. Since the IA was blinded, the treatment group, and the interactions of time splines by treatment were excluded from the model.

The model included change in FEV_1 between baseline and subsequent visits as the dependent variable, and contain the following covariates (except from the intercept):

- Time splines: $S_1=t$, $S_2=t^2$, $S_3=t^3$ and $S_4=\max(t-182,0)^3$ where t is the study day of the measurement and $S_4 = (t-182)^3$ if $t>182$ and 0 otherwise;
- base_FEV1: Baseline FEV_1;
- base_FEV1*S_1,…, base_FEV1*S_4: The interactions of time splines with the baseline FEV_1;
- region: North America versus all other countries together;
- age group: age (< 55 versus ≥ 55 years).

In case the model did not converge, or the covariance matrix was not positive definite, consideration was taken to drop from the model the highest order random effect.

The adjusted mean change from baseline at Week 48 derived from the LMM was adjusted at the mean baseline FEV_1 for all patients who were observed to complete the study or discontinued, not the mean of the covariates over all observations.

Results from each complete dataset was combined using Rubin's rule to provide the least square means (LSMs) for change at Week 48 from baseline as well as the standard error of the LSM. Let se_{all} denote the estimated standard error of the mean change from baseline of FEV_1 at Week 48.

The median of squared standard error over all complete datasets can be written as:

$$se^2_{all,median} = v^2/(n_{complete} + w\, n_{drop-out}),$$

where $n_{complete}$ and $n_{drop\text{-}out}$ denote the number of completer patients (i.e., having a 48 weeks measurement) and the number of patients who dropped-out, respectively, and w is the average weight of the drop-outs. The weight of the completer patients is by definition 1, and w (<1) quantifies the average contribution of the drop-outs to the estimation of the mean change from baseline of FEV_1 at Week 48. The variance v2 is the residual variance of the change from baseline of FEV_1 at Week 48 after adjustment for the covariates.

Then we remove the rows of the drop-outs and calculate the standard error again. The median of squared standard error is now written as

$$se^2_{complete,median} = v^2/n_{complete}.$$

From the two equations, one can solve for v and w. The resulting v disregards the treatment group, so it is larger than the pooled within treatment group variance if the treatment effect is unequal to zero. Under the alternative hypothesis used in the original sample size calculation, H_1: δ= 200 mL (0.2 L), the pooled within groups residual variance estimate can be estimated as v^2 - $\delta^2/4$ (see Kieser and Friede [1]). This value can be used in the sample size formula for the independent samples t-test to calculate the effective sample size $n_{effective}$ needed for a certain power, or for calculating the power for a given $n_{effective}$. The corresponding number of randomized patients is $n_{randomized} = n_{effective}/(1-\pi+\pi w)$, where π is the probability of dropping out, estimated by a Kaplan-Meier analysis based on the observed data at the time of the IA.

The independent statistician had provided the Sponsor with the recommended sample size, defined as:

$$n_{recommended} = maximum\,[110,\ n_{updated}].$$

Here $n_{updated}$ is the total number of patients that has to be randomized in order to have 80% power against a treatment effect of 200 mL (0.2 L).

A sensitivity analysis was conducted for the sample size re-estimation by considering patients who have completed Week 48 or have discontinued when calculating the standard error of the change from baseline to Week 48 for FEV_1 measurements. The data were fitted using the same LMM for primary analysis for sample size re-estimation without any imputation. The LSM as well as its associated standard error overall were reported with all data under sensitivity analyses and data excluding drop-outs, respectively.

8.3.1.4 Analysis of Covariance

Since one of the objectives of the IA was to verify appropriateness of the analysis based on the anticipated longitudinal model vs the analysis based on the ANCOVA model, the sample size calculation was performed also assuming that the final model will be ANCOVA, so that if the decision were to conduct ANCOVA, the power information on this analysis was not missing.

The imputation method described for LMM was repeated and each of the complete datasets after imputation was then analyzed in an ANCOVA model adjusting for region, age group, and baseline FEV_1 as a continuous covariate. Time variables were not included as only Week 48 was analyzed for ANCOVA.

Results from each complete dataset were combined using Rubin's rule to provide the LSMs for change in FEV_1 at Week 48 from baseline as well as standard error of the LSM.

The adjusted mean change from baseline at Week 48 derived from the ANCOVA was adjusted at the mean of the other covariates at baseline for all patients who had completed the study at the time of the IA, not the mean of the covariates over all observations.

The median of squared standard error over all complete datasets is now written as:

$se^2_{complete,median} = v^2/n_{complete}$.

One could solve for *v*. The resulting *v* disregards the treatment group, and is larger than the pooled within treatment group variance if the treatment effect is unequal to zero. Under the alternative hypothesis used in the original sample size calculation, H_1: δ = 200 mL (0.2 L), the pooled within groups residual variance estimate could be estimated as $v^2 - \delta^2/4$ (see Kieser and Friede [1]). This value could be used in the sample size formula for the independent samples t-test to calculate the effective sample size $n_{effective}$ needed for a certain power, or for calculating the power for a given $n_{effective}$. The corresponding number of randomized patients is $n_{randomized} = n_{effective} / (1-\pi)$, where π is the probability of dropping out, estimated by a Kaplan-Meier analysis based on the observed data at the time of the IA.

The independent statistician was to provide to the Sponsor with the recommended sample size, defined as:

$n_{recommended} = maximum\,[\,110,\ n_{updated}]$.

Here $n_{updated}$ is the total number of patients that has to be randomized in order to have 80% power against a treatment effect of 200 mL (0.2 L).

8.3.2 Interim Analysis for missing values evaluation

In accordance with the previous versions of the protocol (version 4.0 and below), the other objectives of the IA were:

- Evaluation of the actual amount of FEV_1 data collected at all timepoints using the CCI spirometer;
- Evaluation of the number of drop-outs and the number and pattern of the missing data;
- Evaluation of the appropriateness of the analysis based on the anticipated longitudinal model vs the analysis based on the ANCOVA model in the light of the collected data.

These objectives were evaluated by means of descriptive statistics on the actual amount of FEV_1 data, number of discontinued patients and number of missing data at each study visit.

Pattern of the missing data were evaluated by number of missing values either due to drop-outs or data not collected at each timepoints (intercurrent missing values).

8.3.3 Communication of IA results to Sponsor and Investigators

Recommendations were communicated as described in the IACP. The IA was performed by the independent, blinded statistician and delivered on 01 December 2021. The recommendations were as reported in Table 1.

Table 1: IA recommendations.

	Section	Recommendation
1	Sample Size Re-estimation using LMM with multiple imputation - Primary Analysis	The recommendation is to continue with the initial sample size (110 patients) as the re-estimating sample size is lower than the initial estimate.
2	Sample Size Re-estimation using LMM - Sensitivity Analysis	LMM Sensitivity analysis supports the findings in the primary analysis. The analysis is only for information purpose due to low sample sizes.
3	Sample Size Re-estimation using ANCOVA with multiple imputation	The recommendation is to increase the sample size to a total number of 152 in order to preserve the pre-specified power.

	Section	Recommendation
4	OTHER OBJECTIVES. It may be concerning when using anticipated longitudinal model as there may be a few data collected at intermediate timepoints between Week 4 and Week 48	The completion rates are preserved at above 67% level and are fluctuated across the visits. There is no clear missing trend that can be observed from the data. In light of the collected data, the recommendation is to continue to the analysis based on the LMM as there is no early indication for trending of missing data.

The blinded statistician communicated these conclusions to the Sponsor. Confidentiality of results was ensured, and no IA results were provided to the Sponsor or the Investigators.

8.3.4 Final decision on appropriateness of primary analysis: LMM vs. ANCOVA.

The recommendations from the blinded statistician (December 1st, 2021) have been evaluated CCI [redacted]

- to continue to strive to achieve the originally planned total patient number of 220 for both clinical trials combined (BOSTON-1 and BOSTON-2) and revise the BOSTON-2 protocol and the new sample size accordingly.
- to implement efforts to minimize missing data in both studies. In general, all patients who maintain consent to be followed for additional outcome information, including patients who discontinue study treatment, should remain in the study for all important safety and efficacy assessments for the full duration of the trial.
- to clearly pre-specify a primary analysis (the meeting package indicates that the primary analysis may be based on ANCOVA, if the LMM model is not viable at the end of the study, in light of the ongoing COVID-19 pandemic). Modifications to the analysis of the primary CCI [redacted] endpoints should be reflected in an updated statistical analysis plan before locking the database and prior to viewing any comparative analysis results.

Based on the above considerations, the LMM is deemed applicable, and it will be used for the primary analysis, while ANCOVA will be used as sensitivity analysis, in addition to sensitivity analyses assessing robustness on missingness mechanism and LMM.

8.4 Type I Error Considerations

The overall type I error rate of the clinical trial will be $\alpha = 0.025$ one-sided. Strong control of the clinical trial-wise type I error level will be achieved by a-priori ordering of hypotheses:

1. The primary endpoint (mean change in FEV_1 from baseline to Week 48) will be tested first.
2. If, and only if, the null hypothesis associated with the primary endpoint can be rejected, a test CCI [redacted] will be performed.
3. If, and only if, the null hypothesis CCI [redacted] can be rejected, a test CCI [redacted] will be performed.

All tests will be performed using a local type I error level of $\alpha = 0.025$ (one-sided). If testing stops due to failure to reject a null hypothesis on a particular level of the hierarchy, all lower-level tests will still be performed but will be considered descriptive.

The blinded IA performed to re-assess the SD and to monitor the missing data does not have any impact on the overall type I error rate.

8.5 Data Monitoring Committee and Confidentiality of Results

An independent Data Monitoring Committee (DMC) will be established to monitor the safety of the IMP throughout the conduct of BOSTON-1 and BOSTON-2 studies. The DMC will monitor safety by evaluating the safety analysis generated by the unblinded independent statistician during the course of the clinical trial. Details of the safety analyses will be included in the DMC's Charter (from CCI [redacted]). The DMC will evaluate the treatment groups for possible trends in AEs, determine whether the basic clinical trial assumptions remain valid, evaluate whether the overall integrity, scientific merit and conduct of the clinical trials remain acceptable, and make recommendations to the Sponsor.

The DMC will perform a comparison between the rate of observed COVID-19 cases, rate of patients' withdrawal, and rate of missing data among trial participants in both treatment arms.

Members of the DMC will be unblinded to the treatment group label, i.e., the DMC will know which group is L-CsA + Standard of Care and which group is Standard of Care alone. The DMC will evaluate all safety and acute tolerability parameters as detailed in the DMC charter.

Safety and acute tolerability parameters include AEs, acute tolerability of L-CsA, infections, malignancies, clinical laboratory data (white blood cell count with differential, red blood cell count, platelet count, alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, gamma glutamyl transferase, blood urea nitrogen, serum bilirubin, serum creatinine), vital signs, physical examinations and FEV_1. The statistics provided will be merely descriptive.

Criteria for membership, responsibilities, meeting frequency, and identification of data listings, summaries and analyses to be provided to the DMC are detailed in the DMC charter.

Statistical analyses to be provided to DMC are planned in a specific DMC SAP.

8.6 Adjudication Committee

The AC – consisting of three independent physicians, with documented expertise in the clinical management of lung transplantation and diagnosis and treatment of BOS - will maintain the role of reviewing the BOS progression data in blind condition once each participating patient completes the study period, with the aim to ensure consistent and complete recording of data. If necessary, the AC may query the sites to ask for clarifications. In case there are discrepancies between the assessment of the AC and that of the Investigator regarding BOS progression events, the CRO will notify the Investigator. In all cases, the final decision as to whether or not specific event qualifies as BOS progression stands with the Investigator.

9 DEFINITIONS

9.1 Protocol Deviations

Major protocol deviations (PDs) are defined as those deviations from the clinical trial protocol for which there could be a significant influence on the assessment of the treatment efficacy. The impact of major protocol deviations on the efficacy results will be investigated by assessing the robustness of the clinical trial results and conclusions to the choice of analysis population, both including and excluding data potentially affected by major protocol deviations. Please note that other study documents may refer to the major PDs as important PDs, and minor PDs as non-important PDs. The SAP and TFL shells will be using the terminology major and minor PDs. For the study, a PD List has been pre-specified. The PD list contains the list of all study PDs, the categorization, the criticality [Major (important) or Minor (non-important)], the potential impact on analysis populations. All the processes around the collection and the review of PDs are described in detail in the study PD Management Plan.

9.2 Analysis Data Sets

The term analysis data set is used to define the set of patients to be included in a class of analyses. The analysis data set is described in terms of the specific criteria for patient inclusion.

Eligibility for the analysis data sets will be determined in a blinded Data Review Meeting (DRM) held after closing, but before locking the database and then producing the final analysis. The blinded DRM will be used to determine if patients should be excluded from the Per-Protocol Set (PPS) in cases of significant and major protocol deviations that would interfere with the assessment

of treatment efficacy. The protocol deviations and their impact are routinely collected and assessed during the conduct of the study.

9.2.1 Safety Analysis Set (SAF)

The SAF is defined as all randomized patients receiving SoC and/or at least one dose of L-CsA, independently of the treatment allocation at randomization. As per IE criteria, randomized patients should be already in SoC at enrollment, therefore it is expected that all randomized patients will enter the SAF population.

Independently of the treatment allocation at randomization, patients will be analyzed according to the treatment they actually received:

- L-CsA plus SoC group: for all randomized patients receiving at least one dose on IMP (L-CsA);
- SoC group: for all randomized patients not receiving any dose on IMP (L-CsA) but treated with SoC.

All data collected after baseline to the end of clinical trial participation will be included in the safety summaries.

All safety and tolerability data will be summarized and analyzed using the SAF.

9.2.2 Full Analysis Set (FAS)

The FAS is defined as all randomized patients. Patients will be analyzed according to the treatment group to which they were randomized. All primary, CCI [redacted] endpoints will be performed using the FAS, unless otherwise specified.

9.2.3 Interim Analysis Full Analysis Set (IA-FAS)

The IA-FAS is a subset of the FAS population, used for running the blinded IA foreseen by the study protocol; it included patients who had a baseline and had completed Week 24, or discontinued at the cutoff date, 30-Sep-2021.

No significant, major protocol deviations were identified for the IA-FAS.

Patients were analyzed according to the treatment group to which they were randomized.

The IA-FAS was only used for the IA completed on December 2021, and no further analyses or summaries were or will be produced for the IA-FAS.

9.2.4 Per Protocol Set (PPS)

The PPS is defined as all patients included in the FAS

- who complete randomized treatment as scheduled or who are withdrawn prematurely due to lack of efficacy or lack of tolerability of the clinical trial treatment, and
- for whom no major protocol deviations interfering with the assessment of treatment efficacy are observed.

For selected efficacy endpoints, including the primary endpoint, the PPS may be used as a supplementary analysis to assess the impact of drop-outs and major protocol deviations on the results.

Lack of efficacy is defined by the withdrawn reason "Lack of efficacy" as collected in the End Of Study CRF form. Lack of tolerability is defined by the withdrawn reason "Adverse Event (Including SAE/Death)" as collected in the End Of Study CRF form.

9.3 Study Day and Time from Randomization

Study Day 1 is defined to be the day of the randomization visit (visit 1, week 0). Time from randomization (in days) for subsequent visits is defined to be visit date – date of randomization visit +1.

9.4 Visit Windows

Following randomization, the visits are scheduled every 4-8 weeks, but the actual times of the visits may deviate from the scheduled times. With spirometry data, consideration needs to be taken to remove the rejected spirometry assessments for analysis before applying the visit windowing.

To allow for any presentation that summarizes values by visit, visit windows will be defined according to the table below. Measurements (both scheduled and unscheduled) will be allocated to visits (and analyzed) as reported in Table 2.

Table 2: Visit windows.

Nominal Visit	V1	V2	V3	V4	V5	V6	V7	V8	V9/EoT	V10/EoS
Nominal Week	W0	W4	W8	W12	W16	W24	W32	W40	W48	W52
Nominal Day	1	29	57	85	113	169	225	281	337	365
Time (days) from Randomization***	0	1-42	43-70	71-98	99-140	141-196	197-252	253-308	*309-364 **309-385	*365-385

**Patients who do not consent or participate in BT – L-CsA – 303 – FU (BOSTON-3).*
***Patients who consented and participated in BT – L-CsA – 303 – FU (BOSTON-3) and will not have a V10 / EoS*
**** Visit windows are based on [visit date – date of randomization visit] formula for time (days) from randomization.*

Data collected after 385 days or 364 days (as applicable, as per the table above) will not be used for analyses but will be included in the listings. If there are two or more measurements assigned to a nominal visit, only the one closest to the scheduled time of the visit will be used in the analysis, unless an additional measurement will be the last measurement during a patient's trial participation

(e.g., in case of an unscheduled visit before premature withdrawal), which will always be included into the analysis.

In the case a scheduled visit is performed in the same date of an unscheduled visit, the scheduled visit assessments will be used; in the specific case of spirometries, if the spirometry data is accepted in the scheduled visit, the scheduled visit data will be used, otherwise (if the spirometry in the scheduled visit is rejected) if the spirometry data is accepted in the unscheduled visit, the unscheduled visit data will be used.

Finally, if a nominal visit has more than one accepted spirometry data assigned, and some of them are performed after IMP inhalation, then only the spirometries done before inhaling the IMP will be used for the visit window mapping rules described above. Instead, if all the accepted spirometries of the nominal visit are done after inhaling the IMP, then all of them will be considered for the above mapping rules. Spirometries collected at 1h and 4h post dose will be used only for safety evaluations and will only be presented in data listings.

Windowing will be applied to the data prior to any missing data imputation.

9.5 Handling of Missing Data

9.5.1 Handling of Missing Efficacy Data

If a patient prematurely discontinues clinical trial participation for any reason apart death, re-transplantation or withdrawal of consent, he/she will be encouraged to complete the follow-up. All observed available data will be used for the primary CCI analyses. The assumption of the LMM (section 14.1) for missing data will be MAR. Since the definition of progression of BOS includes all relevant intercurrent events (IEs) as failures, drop-outs due to other reasons will be treated as non-informative censoring (section 14.4.2).

Several sensitivity analyses have been planned to assess robustness of primary results in case of deviation from MAR mechanism (sections 14.2).

9.5.2 Handling of Missing Safety Data

In general, missing clinical laboratory data, vital signs, and physical examination data will not be imputed. Unknown or partial medication and AE date imputations are given in the next section and to be used only for the assessment of prior/concomitant status for medications and treatment-emergent status for AEs.

9.5.3 Handling of Partial and Missing Dates

9.5.3.1 Partial and Missing Dates of Birth

Where the Date of Birth is missing the following convention will be used:

- Where the day is missing and month and year are available, the day will be completed as the 15th. For example, Date of Birth specified as --JAN1980 will be completed as 15JAN1980.
- If the day and month are missing and the year is available, the day and month will be completed as 02JUL (the 183rd day of the year). For example, Date of Birth specified as ----1980 will be completed as 02JUL1980.

9.5.3.2 Missing or Partial AE and Prior / Concomitant Medication Start and End Dates

Missing and/or incomplete dates for medications and AEs are imputed in a manner resulting in the earliest onset or the longest duration during the Treatment Period, whilst ensuring that the start date does not occur after the stop date.

For the derivation of the duration and cumulative of concomitant medication, if the medication is "Ongoing" then the medication end date will be imputed with the last visit date.

The stop date will not be imputed if the AE is "Ongoing". Technically, this will be done as follows:

- For a missing/incomplete start date/time, the earliest date/time of the following will be imputed:
  - The later date / time of: the earliest possible start date/time and the date/time of first dose of treatment (date/time of randomization).
  - The latest possible start date/time.
- For a missing/incomplete stop date/time, the later date/time of the following will be imputed:
  - The earlier date/time of: the latest possible stop date/time and the date/time of last dose of treatment (date/time of study completion or early termination).
  - The earliest possible stop date/time.
- Here, the earliest possible date/time is defined as:
  - The date/time itself if available.
  - The date/time of the first day of the month at 00:00hrs, if month and year are available but the day/time is missing.
  - The date/time of the first day of the year at 00:00hrs, if year is available but day/time and month are missing.
  - 00:00hrs on the day of informed consent, if the date/time is completely missing.
- The latest possible date/time is defined as:
  - The date / time itself if available.
  - The date / time of the last day of the month at 23:59hrs, if month and year are available but the day / time is missing.

- The date / time of the last day of the year at 23:59hrs, if year is available but day / time and month are missing.
- 23:59hrs on the date of last known date on the study for the patient plus one year, if the date / time is completely missing.

10 DISPOSITION

The number of patients screened, failed screening (patients who did not meet all eligibility criteria and was not randomized)**,** enrolled (patients with informed consent and all eligibility criteria met or randomized), treated but not randomized, randomized (treated, and not treated with clinical trial treatment), completing clinical trial treatment (and reasons for not completing), and completing the clinical trial (and reason for not completing) will be presented in frequency tables by treatment group.

The number of patients in each analysis population (SAF, FAS, and PPS) will be summarized by treatment group. A by-patient listing of disposition will be provided.

All major protocol deviations will be presented as a frequency table by treatment group and in a by-patient listing for all randomized patients, which will include the category of the deviation, the deviation text, and whether the patient was excluded from PPS.

In addition, the counts on COVID-19 related major PDs will be provided, together with a by-patient listing for all COVID-19 related PDs.

In addition, a listing of the patients affected by Coronavirus Disease of 2019 (COVID-19) and their COVID-19 AE details will be provided for the FAS.

A table for the summary of stratification factors by treatment group for FAS and a separate listing of patients with mis-stratifications will be displayed.

11 DEMOGRAPHICS AND BASELINE CHARACTERISTICS, AND DISEASE AND MEDICAL HISTORY

11.1 Demographic and Baseline Characteristics

All demographic, baseline characteristics, including azithromycin at randomization, and antibody-response data will be summarized by treatment group for SAF, FAS and PPS. For summary purposes, the most recent, non-missing result obtained on or prior to the baseline visit will be used.

Azithromycin at randomization will be derived by identifying azithromycin among medications collected in the study (generic name: azithromycin) using any of the following rules:

- azithromycin is ongoing at the date of randomization (regardless of start date);

- azithromycin start date is on the randomization date (regardless of end date).

By-patient listings will be provided for all demographic and baseline characteristics.

11.2 Disease and Medical History

Details of History of BOS diagnosis eCRF page [FVC (absolute), FVC - % of personal best post-transplant, FEV_1 (absolute), FEV_1 - % of personal best post-transplant, CCI (absolute), CCI - % of personal best post-transplant and FEV_1/FVC (ratio)], details of Historical Spirometry eCRF page [FVC (absolute), FEV_1 (absolute), CCI (absolute) and FEV_1/FVC (ratio)] as well as details related to each patient's lung transplant as recorded in the Lung Transplantation History eCRF page [type of underlying indication for lung transplant, time between lung transplantation and onset of BOS, donor gender, cytomegalovirus (CMV) status donor, CMV status recipient, Epstein Barr Virus (EBV)-status donor, EBV-status recipient, receive induction therapy, receive antiviral/antifungal/antibacterial therapy, cold ischemia duration double lung] will be summarized by treatment group for SAF, FAS and PPS.

Disease history, including the details from the Risk Factors eCRF will be listed by FAS.

Time (days) between lung transplantation and onset of BOS is defined as the duration between date of lung transplantation and onset of initial BOS + 1.

All other significant medical conditions (including gastro-esophageal reflux disease) and procedures (including past surgeries) occurring before the informed consent form is signed will be recorded as medical history.

Medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) [Version 23.1] and will be presented by Internationally Agreed order System Organ Class (SOC) and Preferred Term (PT) and total. The SOCs and PTs are to be sorted by SOC and descending PTs.

Medical history will be summarized by treatment group and overall for the SAF, FAS and PPS as follows:

- The number and percentage of patients with at least one medical history record will be presented.
- The number and percentage of patients with at least one medical history record within each primary SOC and PT will be presented. The summary will be sorted using the internationally agreed order for SOC and using descending order of overall numerical counts for PT. Where terms tie, these will be sorted alphabetically.

Medical history records will be listed by-patient and within-patient by medical history start date for the FAS.

11.3 Baseline Lung Allograft Dysfunction

Baseline Lung Allograft Dysfunction (BLAD) at baseline (Yes, No) will be reported in the demographic table. BLAD patients are defined as patients with baseline FEV_1/predicted FEV1*100 <80% [10]. The following formula for the calculation of predicted FEV1 will be used:

$$\text{Predicted } FEV_1 = b_0 + b_1 * age + b_2 * age^2 + b_3 * height^2$$

where the coefficients of the regression are reported in Table 3.

Table 3: Coefficients for FEV_1 Prediction

Sex	Ethnicity	Intercept (b_0)	Age (b_1)	Age^2 (b_2)	Ht (cm)^2 (b_3)
Male	Caucasian < 20 yr of age	−0.7453	−0.04106	0.004477	0.00014098
Male	Caucasian ⩾ 20 yr of age	0.5536	−0.01303	−0.000172	0.00014098
Male	African-American < 20 yr of age	−0.7048	−0.05711	0.004316	0.00013194
Male	African-American ⩾ 20 yr of age	0.3411	−0.02309	0	0.00013194
Male	Mexican-American < 20 yr of age	−0.8218	−0.04248	0.004291	0.00015104
Male	Mexican-American ⩾ 20 yr of age	0.6306	−0.02928	0	0.00015104
Female	Caucasian < 18 yr of age	−0.8710	0.06537	0	0.00011496
Female	Caucasian ⩾ 18 yr of age	0.4333	−0.00361	−0.000194	0.00011496
Female	African-American < 18 yr of age	−0.9630	0.05799	0	0.00010846
Female	African-American ⩾ 18 yr of age	0.3433	−0.01283	−0.000097	0.00010846
Female	Mexican-American < 18 yr of age	−0.9641	0.06490	0	0.00012154
Female	Mexican-American ⩾ 18 yr of age	0.4529	−0.01178	−0.000113	0.00012154

Due to eCRF study design, ethnicity has not been recorded as required for Table 3, and as a result, the race field as recorded in the eCRF will be used for mapping as detailed below in Table 4.

Table 4: eCRF Race Classification

eCRF Term	Remapping
White	Caucasian
Black or African American	African-American
Asian	0.88% reduction will be used versus the Caucasian
Other	check "other, specify" to see additional race info. - If patients are defined as “white” or “black/African”, please use the corresponding classification. - If patients are defined as "latinos" , then use “Mexican-American”. - If patient is defined as “Arabian” then treat as the Asian category with the 0.88% reduction used versus the Caucasian. Otherwise consider them as "unknown"

For Asian or Other field with Arabian ethnicity, the calculation will be performed using the Caucasian coefficients of Table 3 (considering the corresponding sex, age, and height values) and then applying a correction factor of 0.88 [12]. For example, in case of an Asian male patient older than 20 years, the calculation will be:

Predicted $FEV_1 = 0.88 * (0.5536 - 0.01303 * age - 0.000172 * age^2 + 0.00014098 * height^2)$.

In case of other races (American Indian or Alaskan, Native, Native Hawaiian or Other, Pacific Islander), unknown, or unspecified ethnicities (including "multiracial", "multicultural", "mixed", etc.), parameters calculated as average of the Caucasian, African-American, and Mexican-American parameters will be used, according to Table 4:

Table 4: Coefficients for FEV_1 Prediction for other/unspecified ethnicities.

Sex	Age class	Intercept (b_0)	Age (b_1)	Age^2 (b_2)	Ht (cm)^2 (b_3)
Male	< 20 yr of age	-0,7573	-0,04688	0,004361333	0,000141
Male	⩾ 20 yr of age	0,508433	-0,0218	-0,0000573	0,000141
Female	< 18 yr of age	-0,9327	0,062753	0	0,000115
Female	⩾ 18 yr of age	0,409833	-0,00941	-0,000134667	0,000115

12 Immunosuppressants

Immunosuppressants are collected in the Concomitant Medications CRF form.

As per CRF completion guidelines, immunosuppressants are to be collected using the same measurement unit i.e., mg. In those cases where it is not done, considerations will be taken on ongoing basis (at DRM at the latest) to convert units for each immunosuppressive agent to allow cumulative dose to be calculated. Immunosuppressant medications will be coded using the World Health Organization (WHO-DD) classification system and grouped by drug class and PT.

In order to identify immunosuppressants among concomitant medications, any of the following two rules will be applied.

1. Only the following routes are to be considered: oral, intramuscular, intravenous, and subcutaneous;
   the following codes are to be considered:
   - L04A Immunosuppressants:
     1. L04AA Selective immunosuppressants,
     2. L04AB Tumor necrosis factor alpha (TNF-α) inhibitors,
     3. L04AC Interleukin inhibitors,
     4. L04AD Calcineurin inhibitors,
     5. L04AX Other immunosuppressants;
   - H02AB Glucocorticoids (all);

2. L04AD02 Tacrolimus, and route = sublingual.

CCI

12.1 Basic Immunosuppressants

As per protocol, regardless of treatment assignment, all clinical trial participants must be on a maintenance regimen of immunosuppressive agents including tacrolimus, a second agent such as but not limited to mycophenolate motfetil (MMF) or azathioprine, and a systemic corticosteroid such as prednisone as third agent. At each visit, the average maintenance doses of immunosuppressive agents will be recorded by the Investigator on the Concomitant Medication eCRF.

Basic immunosuppressants refer to maintenance regimen of immunosuppressive agents and include the following:

- Calcineurin inhibitors (Tacrolimus, Cyclosporine A);
- Antimetabolicte (Antimetabolite) agents (MMF, Azathioprine);
- mTOR inhibitor (Everolimus, Sirolimus);
- Selective costimulation blocker (Belatacept);
- Corticosteroids.

The total basic immunosuppression administered from the day of randomization will be evaluated.

Basic immunosuppressives will be coded by Anatomical Therapeutic Code (ATC) codes, and will be summarized by treatment group for SAF, FAS and PPS.

A p-value will be obtained from the Fisher's Exact test for the comparison of patients with any basic immunosuppressants from L-CsA plus SoC versus patients with any basic immunosuppressants from SoC alone, for each class of basic immunosuppressants listed above.

Descriptive statistics on cumulative doses for basic immunosuppressants will be provided:

- for the whole treatment period (from day of randomization to end of study),
- for each period between study visits (V1 to V2, V2 to V3, etc.),
- cumulatively for each period between study visits (V1 to V2, V1 to V3, V1 to V4, etc.),

where periods between study visits V_i to V_{i+1} (with i=2,3,4, etc) is defined as the interval starting from the first day post visit V_i and ending the last day of visit V_{i+1} (included). For V1 to V2 the date of V1 must be considered too.

Descriptive statistics will be calculated for each class of basic immunosuppressants listed above and overall. If a patient took more than one basic immunosuppressant related to the same class, then the total dose will be taken into consideration in the calculation.

P-values will be obtained from the Fisher's Exact test for the comparison of patients with any basic immunosuppressants, including each of basic immunosuppressant listed above from L-CsA plus SoC versus patients with any basic immunosuppressants from SoC alone at each time interval (V1 to V2, V1 to V3, …. ,V1 to V9).

Exposure to immunosuppressant medication details will be summarized similarly for each category of basic immunosuppressant, and overall for all.

This includes the calculation of the duration of intake (date of last basic immunosuppressant intake – date of first basic immunosuppressant intake + 1). The duration will be calculated by summing the duration of the associated treatments. The overlapping days will be counted only once.

The ratio between the cumulative dose and intake duration will be used to produce the dose per day. The overall dose per day per patient will then be displayed for each treatment group for each time interval along with the 95% confidence interval.

By-patient listings will be provided for all basic immunosuppressants for FAS.

12.2 Other Immunosuppressants

Other immunosuppressants refer to all the immunosuppressive therapies that are not classified as maintenance regimen as described in section 12.1 (but meet the conditions reported in section 12). Other immunosuppressive therapies will be coded by Anatomical Therapeutic Code (ATC) codes, and will be summarized by treatment group for SAF, FAS and PPS.

All other immunosuppressants administered from the day of randomization will be evaluated. The number (%) of patients taking any other immunosuppressants will be provided overall and for each ATC code.

The same summary statistics for cumulative dose and for exposure as described in section 12.1 will be provided.

13 Prior and concomitant medications

Prior medications are those medications taken prior to randomization with a stop date and time prior to the randomization date/time (standard definition of prior/concomitant medications considers the first dose of study treatment but, as in this study SOC group won't have any first dose date, randomization date will be considered instead).

Concomitant medications are those medications with a start date and time on or after the randomization date/time, or those with a start date and time before the randomization date/time

and either a stop date and time on or after the study randomization or are ongoing at the end of the study.

Prior and concomitant medications will be summarized by ATC codes and listed for FAS. Prior/concomitant medications will be coded using the WHO-DD classification system and grouped by drug class and PT.

14 EFFICACY

For the efficacy analysis, the adequacy of the model assumptions will be checked, and if it is found that these are inadequate then alternative approaches may be performed.

For all variables assumed to be normally distributed, the conditional studentized residuals from this model will be tested using the Shapiro-Wilk test. If the assumption of normality of these residuals is violated, then a log-transformed model will be fitted. The need for a possible log transformation of these variables will be investigated prior to unblinding by assessing the model residuals without a treatment term. Where data are transformed for analysis, back transformation will be applied for the presentation of results. Presentation of the tables will be amended as needed for this, with the addition of an explanatory footnote.
Classification and imputation of missing data will be completed prior to any transformation of the data for analysis.

As a general remark for all efficacy endpoints based on spirometry data, values collected at 1h and 4h post dose will not be used for any efficacy analyses as they are collected only for safety evaluations.

14.1 PRIMARY EFFICACY ANALYSIS

14.1.1 Analysis Population and Estimand

The primary endpoints will be analyzed using a "composite strategy" (International Conference on Harmonization (ICH) guideline E9 addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials). The estimand is defined by the following:

- Population: adult recipients of a double pulmonary allograft with clinically defined BOS, matching the inclusion and exclusion protocol criteria.
- Variable: FEV_1 data at Week 48 (visit 9) collected from the onsite CCI spirometer (data collected with the CCI home spirometer will be used for supplementary analyses). Only accepted over-reader best FEV_1 data for the session from CCI will be used for analysis; data will not be re-derived.

Intercurrent events: After IEs such as stopping treatment with L-CsA, augmentation of SoC therapy, etc., FEV_1 measurements will still be collected and used in the analysis. In case of IEs preventing the use of FEV_1 measurements, i.e. death and re-transplantation, the FEV_1 will be imputed as treatment failure: FEV_1 equal to zero for all visits subsequent the event.

Rejected spirometries are those that did not meet the American Thoracic Society criteria; if rejected, then they would be re-done but there may be cases where they are rejected twice and resulting in missing data. Rejected spirometry is not an indicator of a better/worse outcome and can be assumed to be MAR.

- Population-level summary: difference in mean change from baseline in the FEV_1 at week 48 between treatment groups.

Table 5 summarizes the statistical approaches for handling IEs and missing data.

Table 5: Summary of strategies for Intercurrent Events and Missing data.

Type of IEs	**Statistical analysis**				
	Primary analysis *(section 14.1.3)*	**Missing data not at random (MNAR)** *(section 14.2.1)*	**"While on treatment" strategy (for COVID-19)** *(section 14.3.3.3)*	**"While on treatment" strategy** *(section 14.3.5)*	**Trimmed Means** *(section 14.3.6)*
Death	The FEV_1 value will be imputed equal to zero.	The FEV_1 value will be imputed equal to zero.	Not applicable.	Not applicable.	The FEV_1 value will be imputed equal to zero at W48.
Re-transplantation			The FEV_1 value will be imputed equal to zero.	The FEV_1 value will be imputed equal to zero.	
Treatment discontinuation Study discontinuation for lack of efficacy*	FEV_1 measurements after IE will still be collected and included in the analysis.	FEV_1 measurements after IE will still be collected and included in the analysis.	FEV_1 measurements after IE will still be collected and included in the analysis.	Data after IE will not be considered**.	FEV_1 measurements after IE will still be collected and included in the analysis. Last observation will be used for determining the ordering of patients to be trimmed.
Initiation of additional immunosuppressive					
COVID-19 infection			Data after IE will not be considered.	FEV_1 measurements after IE will still be collected and included in the analysis.	
In case of: - residual missing data, or - study discontinuation for consent withdrawal or lost to follow-up:	Missing data handled under MAR	Missing data imputed using fully conditional specification based on data from retrieved discontinued patients ***.	Missing data handled under MAR	Missing data handled under MAR	Last observation will be used for determining the ordering of patients to be trimmed. Missing data handled under MAR.

** If a patient discontinues the study for lack of efficacy, he/she will be encouraged to complete the follow-up visits planned in the protocol. All patients who prematurely discontinue the study should complete at least the evaluation at EoT prior to discontinuation whenever possible.*

*** Two analyses will be performed: first considering only treatment discontinuation as IEs, then considering the initiation of additional immunosuppressive too.*

**** If there are not enough retrieved patients for the convergence of MI-RD regression model, a reference-based MI will be adopted.*

14.1.2 Working Hypothesis

The alternative hypothesis (H_1) for this trial is that L-CsA plus SoC is superior to SoC alone in preserving FEV_1, i.e., under H_1 it is expected that patients treated with L-CsA plus SoC will exhibit a smaller mean FEV_1 decline between baseline and the end of Week 48 than those who receive SoC alone.

14.1.3 Analysis Model

The primary efficacy analysis will be carried out using a LMM for repeated measurements, using all observed available FEV_1 measurements. In case of death or re-transplantation events, FEV_1 will be imputed as zero at each nominal day post event.

The dependent variables are the absolute change from baseline of FEV_1 measurements at visit 2 (i.e., the first post-baseline visit) through visit 9 (Week 48). For the definition of baseline FEV_1 please refer to section 8.1.1.

A flexible and robust model will be chosen for the random part of the LMM according to a pre-specified stepwise procedure (see next section). The fixed part of the LMM is specified as follows: the model will contain (except from the intercept) time cubic splines, treatment, the interactions of time splines by treatment, baseline FEV_1, the interactions of time splines with the baseline FEV_1, region (North America versus all other countries together), age (< 55 versus ≥ 55 years), and use of azithromycin at randomization (i.e. Baseline, Visit 1).

The primary efficacy parameter will be estimated from the model by the adjusted difference between treatment groups in mean FEV_1 change from baseline after 48 weeks. The p-value for the null hypothesis of no difference and a two-sided 95% confidence interval for the difference will be provided.

Superiority of L-CsA will be demonstrated by a statistically significant result (defined as $p<0.025$, one-sided) in favor of the L-CsA plus SoC treatment compared to SoC alone.

The above comparison between treatment groups at all the other visits (2-8) will also be presented.

A figure of observed mean change from baseline and standard errors at each visit, by treatment group will be presented. A figure with adjusted mean changes from baseline at each visit by treatment group derived from the LMM will be provided.

14.1.4 Specification of the Model

For valid inference on the primary efficacy parameter, the evolution of FEV_1 in both treatment groups should be modeled as a flexible function in time, without unnecessarily strong assumptions.

Therefore, time will be modeled assuming that the time evolution of FEV_1 is a smooth function of time. Cubic splines will be used with a large enough number of parameters to minimize the risk of misspecification. We will choose one knot at 182 days, i.e., halfway between the first (day 28) and

the last scheduled measurement (337 days). So, it is assumed that the mean of FEV_1 is a third-degree polynomial at each of the two intervals, smoothly connected, equal at day 182 with equal first and second derivative. Thus, 4 parameters per group are used to model time. In total, as many as 12 parameters in the model will involve time, ensuring robust estimation of the time evolution of FEV_1. Denoting the covariates associated with the splines by $S_1=t$, $S_2=t^2$, $S_3=t^3$ and $S_4=\max(t-182,0)^3$, the treatment group by T (0=SoC; 1=L-CsA plus SoC), the day of the measurement by t and the baseline FEV_1 by X, the fixed part of the model will contain the following covariates:

Intercept, S_1, …, S_4, T, $T*S_1$,…, $T*S_4$, X, $X*S_1$,…, $X*S_4$, other covariates

The Kenward-Roger adjustment will be used for the denominator degrees of freedom needed for hypothesis tests and confidence intervals.

The other covariates are:

- Age (< 55 versus ≥ 55 years);
- Region (North America versus all other countries together);
- Use of azithromycin at randomization.

In the unlikely case of collinearity, covariates will be dropped in the above order, starting with the last.

The random part of the LMM will be selected by comparing the Akaike Information Criterion (AIC) of three models:

- model with random intercept and random time;
- model with random intercept, random time and random squared time;
- model with random intercept, random time, random squared time, and random cubic time.

The model with the lowest AIC will be adopted for the primary analysis.

The core excerpt of CCI code for the primary analysis is reported in section 17.1.

14.1.5 Specification of the Covariance Model

The model for the mean uses the actual times of the visits as a continuous variable instead of the scheduled visit times. Therefore, the covariance model will also use the actual visit times, using random effects. For valid inference, the model for the covariance should be correctly specified. Thus, the preference will be for models with relatively many parameters in order to ensure robust estimation of the covariance, though this may lead to a loss of efficiency.

Therefore a model will be specified with four per patient random effects: intercept, t, t^2, t^3, with an unstructured covariance matrix. This covariance model contains 11 parameters, ensuring sufficient robustness. In case the model does not converge, or the covariance matrix is not positive definite, we will drop terms from the highest order random effect.

14.2 Sensitivity Analyses for the Primary Endpoint

14.2.1 Missing data not at random (MNAR)

To investigate the impact of departures from MAR assumption, a MNAR approach will be evaluated using the LMM described for the primary analysis. After remapping of the visits as per section 9.4, FEV_1 at specific visit will be imputed depending on the IE experienced by the patient.

In case of IEs as death or re-transplantation, FEV1 will be imputed as zero at each post-event nominal day as per section 9.4. Other IEs (see Table 5) will be addressed per treatment policy strategy, and FEV1 measurements collected after IE will be used in the analysis.

Residual missing data (or missing data due to discontinuation for consent withdrawal or patient lost to follow-up) will be imputed using fully conditional specification, a type of multiple imputation using chained equations, implemented with CCI [redacted] under the FCS option. The imputation model will include baseline FEV_1, age (as a continuous variable), region, use of azithromycin at randomization, and intermediate FEV_1 up to week 48 as covariates. Two hundred data sets will be generated, and the random seed number will be 52082. In case of imputation, the associated nominal day of assessment will be considered as per section 9.4.

Only observed values from retrieved discontinued patients will be used to inform the MI regression model (this model will be defined as MI-RD). Retrieved discontinued patients are defined as patients who discontinue study treatment and decide to remain in the study by following the schedule of assessments and continuing to adhere to protocol requirements. If there are not enough retrieved patients for the convergence of MI-RD regression model, a reference-based MI will be adopted: imputation of values in the investigational arm will be done using the non-missing values from the control group (this approach will be referred as "Copy Reference"). This approach does not assume benefits for the investigational arm in case of discontinuation and limits a post-discontinuation clinical effect to that of SoC. The final decision on the use of the MI-RD vs Copy-Reference will be done at the time of the analysis and reported in the CSR.

Each of the 200 imputed datasets will be analyzed using the LMM (see primary analysis) including all observed and imputed data. Rubin's rule [2] will be used for combining results to draw inference.

The robustness of primary results will be evaluated by means of the adjusted difference between treatment groups in mean FEV_1 change from baseline after 48 weeks. The p-value for the null hypothesis of no difference and a two-sided 95% confidence interval for the difference will be provided. The same comparison between treatment groups at all the other visits (2-8) will be presented.

The core excerpt of CCI [redacted] code for this sensitivity analysis is reported in section 17.2

The 200 imputed datasets will be used also for sensitivity analysis reported in section 14.2.4.

14.2.2 Delta-Based Controlled Imputation (Tipping point)

A delta-based controlled imputation approach using the tipping point analysis will be performed as a sensitivity analysis of treatment superiority, if shown. Tipping point analysis will assess how departures from MAR assumption must be to overturn conclusions from the primary superiority analysis. Tipping point will be based on iterative application of MI.

In case of death or re-transplantation events, FEV_1 will be imputed as zero at each nominal day post event.

In the first iteration, the missing FEV_1 outcomes will be imputed separately for each treatment group, using the fully conditional specification. The imputation model will include baseline FEV_1, age, region, use of azithromycin at randomization, and intermediate FEV_1 up to week 48 as covariates. Two hundred data sets will be generated, and the random seed number will be 99823. Non-missing values from all patients will be used.

In successive iterations, to represent departures from MAR, values $\Delta 1$ and $\Delta 2$ will be added and subtracted to the imputed values for L-CsA plus SoC treatment and SOC treatment groups respectively. $\Delta 1$ and $\Delta 2$ will range from 0 to 100 mL (0.1 L) by 10 mL (0.01 L).

For each scenario (i.e., each combination of $\Delta 1$ and $\Delta 2$), the 200 delta-adjusted imputed datasets will be analyzed using the primary LMM including all observed and imputed data. Rubin's rule will be used for combining results to draw inference. The adjusted difference between treatment groups in mean FEV_1 change from baseline after 48 weeks will be provided along with its p-value for the null hypothesis of no difference.

These results will show how extreme the departures from MAR must be to overturn the results of the primary analysis. Controlled imputation is always feasible, independent from the incidence of dropouts.

The core excerpt of CCI code for this sensitivity analysis is reported in section 17.3.

14.2.3 Mixed model repeated measures (MMRM)

To investigate robustness against deviations from the model, a MMRM model will be carried out using all observed available FEV_1 measurements (under MAR assumption). The dependent variables are the absolute change from baseline of FEV_1 measurements at visit 2 (i.e., the first post-baseline visit) through visit 9 (Week 48). In case of death or re-transplantation events, FEV_1 will be imputed as zero at each nominal day post event.

As fixed effects, the model will contain the following covariates (except from the intercept): categorical (scheduled) visit, treatment, the interactions of treatment by categorical visit, baseline FEV_1, the interactions of categorical visit with the baseline FEV_1, region (North America versus all other countries together), age (< 55 versus ≥ 55 years), and use of azithromycin at

randomization (i.e. Baseline, Visit 1). Random effects per patient for this model will include categorical visit.

An unstructured covariance pattern will be used to estimate the variance–covariance of the within-patient repeated measures. If the model fails to converge, a heterogeneous Toeplitz structure will be used. Parameters will be estimated using REML with the Newton–Raphson algorithm and using the Kenward–Roger method for calculating the denominator degrees of freedom.

Treatment group comparisons at each visit will be estimated by differences between least squares (LS) means from the visit by treatment interaction, with accompanying p-values and 95% CIs. The robustness of primary results will be evaluated by means of the difference of LS means at week 48.

The core excerpt of CCI code for the sensitivity analysis is reported in section 17.4.

14.2.4 MMRM under MNAR

Each of the 200 imputed datasets created under MNAR condition as described in section 14.2.1, will be analyzed using the MMRM described in previous section 14.2.3 including all observed and imputed data. Rubin's rule will be used for combining results to draw inference.

Treatment group comparisons at each visit will be estimated by differences between least squares (LS) means from the visit by treatment interaction, with accompanying p-values and 95% CIs. The robustness of primary results will be evaluated by means of the difference of LS means at week 48.

14.2.5 ANCOVA

The population-level summary for the variable will be altered. An ANCOVA model will be performed on the changes from baseline in FEV_1 to Week 48, using the same stratification factors and the same covariates described for the LMM for the primary efficacy analysis (except for time factors). In case of death or re-transplantation events, FEV_1 will be imputed as zero at each nominal day post event.

In case of violation of the normality assumption even after log transformation (section 14), then rank transformation will be applied to normalize the data. The same ANCOVA model will be applied to the rank-transformed data.

14.2.6 Selection Modeling

A selection model fits simultaneously a model for the repeated FEV_1 measurements and a model for the probabilities of dropping out at the different visits. These models are referred to as the measurement model and the drop-out model.

Applying Diggle and Kenward's Selection Model on incomplete longitudinal data, the response is modeled through a repeated measures model, while a logistic regression model is used to describe the dependency of the missing data indicators on the longitudinal responses.

Under this approach, the probability of dropout is considered a function of the observed longitudinal response measured prior to dropout at time j and the unobserved response at the time of dropout. For individuals with complete data, dropout depends on the response prior to the last measurement ($y_{i,j-1}$) and the final observed response (y_{ij}). In either case, dropout cannot depend on future responses. Specifically, the logit of the probability of dropout depends on the observation prior to dropout ($y_{i,j-1}$) (assuming the individual was observed prior to dropout) and the unobserved response at dropout (y_{ij}):

$$\text{Logit}[P(Di=j|Di>=j] = \text{Intercept} + \alpha_1 y_{i,j-1} + \alpha_2 y_{ij}$$

where the coefficients α_1 and α_2 are the effects of $y_{i,j-1}$ and y_{ij} on the logit, respectively. Under this model, if both α_1 and α_2 are equal to zero, then the missingness is independent of the previous and missing responses. In such cases, the missing data are Missing Completely at Random, and consequently, the missing data process is ignorable. If α_2 but not α_1 is equal to zero, then the missingness is related to the response prior to dropout. In this case, the missing data are MAR, and again, the missing data process is ignorable. If α_2 is not equal to zero, then the missingness depends on the missing data at dropout. Consequently, the missing data process cannot be ignored.

These scenarios will describe how large the departure from the MAR assumption should be in order to substantially affect the results of the primary analysis. If the number of drop-outs turns out to be small, this sensitivity analysis may not be feasible because of insufficient information to fit the logistic drop-out model, and the CSR will mention this as a caution. This model can be fitted using the selection modelling macros from the Drug Information Association Scientific Working Group on Estimands and Missing Data [8].

14.2.7 Joint Model

This analysis jointly models the repeated FEV_1 measurements and the time to dropout. The longitudinal measurement process model follows a standard random-coefficient mixed effect model, while the dropout mechanism model uses a complementary log-log link.

This model will be fitted using the shared parameters macros from the Drug Information Association Scientific Working Group on Estimands and Missing Data [8].

14.3 Supplementary Analyses for the Primary Endpoint

Supplementary analyses can provide additional insights into the understanding of a treatment effect.

14.3.1 PPS

The population for the primary estimand will be altered from the FAS to the PPS. The analysis will be conducted as described for the primary endpoint.

14.3.2 CCI data

The endpoint for the estimand will be spirometry data stored on the CCI. The analysis will be conducted as described for the primary endpoint.

14.3.3 The impact of COVID-19

The COVID-19 pandemic has impacted the conduct of the study and might have impacted its results for different aspects including infections, quarantines, social distances [5], and other considerations linked to the different ways that each site has faced the pandemic. Date of COVID-19 outbreak was set at March 11, 2020 and different analyses will be performed to assess the impact of the COVID-19 pandemic to the study pre and post pandemic. In case evidence of an impact of pandemic on study endpoints should emerge, considerations on how the initial study assumptions have been altered must be done, and changes in the statistical criteria for final judgment should be applied.

14.3.3.1 Patients impacted by COVID-19 pandemic

The analysis of primary endpoint will be performed using the LMM described in the primary analysis with the inclusion of additional categorical covariates:

- A binary covariate representing if patients have been affected by COVID-19 during the treatment period:
  - Affected, if an AE of COVID-19 is reported in the eCRF AE page between randomization and end of treatment date;
  - Not affected, if no AEs of COVID-19 are reported in the eCRF AE page between randomization and end of treatment date.
- A binary covariate representing if patients have been randomized before or after the COVID-19 pandemic outbreak:
  - Before, if date of randomization is before 11 March 2020;
  - After, if date of randomization is equal or after 11 March 2020.

FEV_1 data will be descriptively summarized by impact of COVID-19 (COVID-19 affected vs Not COVID-19 affected, and randomized before COVID-19 Pandemic Outbreak vs randomized after COVID-19 outbreak) at each visit, by treatment and overall. The treatment comparison for difference in mean change from baseline in the FEV_1 at week 48 will be carried out using t-tests as described in section 14.3.4. This will also be performed for each of the subpopulations defined

by the binary covariates above i.e. patients affected/not affected by COVID-19 during the treatment period and patients randomized before/after the COVID-19 pandemic outbreak.

Furthermore, it is of interest to categorize each visit for each patient to assign these before or after the COVID-19 pandemic outbreak. A summary table will be produced to descriptively display the effects of FEV1 over time by taking each visit with respect to the COVID-19 outbreak pandemic date into consideration. The treatment comparison for difference in mean change from baseline in the FEV_1 at week 48 will be analyzed similarly using the t-test as described in section 14.3 for both groups, visits before COVID-19 pandemic outbreak, visits after COVID-19 pandemic outbreak.

The observed mean change from baseline and standard errors of FEV1 at each timepoint will be graphically reported by treatment and by impact of COVID-19 categorization. This will also be displayed by COVID-19 pandemic outbreak at each visit. Furthermore, a figure with adjusted mean changes from baseline at each visit by treatment group derived from the LMM with impact of COVID-19 will be provided.

14.3.3.2 Piecewise regression

The analysis of primary endpoint will be carried out using a modified LMM for repeated measurements that will allow to assess if FEV1 trajectory slopes are different before and after the COVID pandemic outbreak. All observed available FEV1 measurements will be used and the model will contain (except from the intercept) time (from randomization), time after COVID pandemic outbreak (defined as days after March 11, 2020, with March 11, 2020 considered as equal to zero; for assessments of FEV1 done before 11 March 2020 this time is set at zero), treatment, the interactions of time (from randomization) by treatment, the interactions of time after COVID pandemic outbreak by treatment, baseline FEV1, the interactions of time (from randomization) with the baseline FEV1, the interactions of time after COVID pandemic outbreak with the baseline FEV1, region (North America versus all other countries together), age (< 55 versus 55 years), and use of azithromycin at randomization (i.e. Baseline, Visit 1). The model will include random intercept, time and time after COVID pandemic outbreak.

Denoting the time from randomization and the time after COVID pandemic outbreak by S_1 and S_2 respectively, treatment group by T (0=SoC; 1=L-CsA plus SoC), the day of the measurement by t and the baseline FEV_1 by X, the fixed part of the model will contain the following covariates:

Intercept, S_1, S_2, T, $T*S_1$, $T*S_2$, X, $X*S_1$, $X*S_2$, other covariates.

The Kenward-Roger adjustment will be used for the denominator degrees of freedom needed for hypothesis tests and confidence intervals.

The primary efficacy parameter will be estimated from the model by the adjusted difference between treatment groups in mean FEV_1 change from baseline after 48 weeks. The p-value for the null hypothesis of no difference and a two-sided 95% confidence interval for the difference will

be provided. Superiority of L-CsA will be confirmed by a statistically significant result in favor of the L-CsA plus SoC treatment compared to SoC alone.

The slopes of FEV_1 before and after March 11, 2020 will be assessed and compared.

The core excerpt of CCI code for the primary analysis is reported in section 17.5.

14.3.3.3 "While on treatment" strategy (for COVID-19)

A further supplementary analysis on the impact of COVID-19 will be performed. The IE of COVID-19 infection (per eCRF page, see definition in section 14.3.3.1) will be considered and will be handled by the "while on treatment" strategy: the analysis will be conducted as the primary analysis with the exception that, in case of a COVID-19 IEs, only the FEV_1 measurements up to the date the COVID-19 AE started will be used.

In case re-transplantation events, FEV_1 will be imputed as zero at each nominal day post event up to the COVID-19 infection date. According Table 5, in case of death, no imputation will be done.

14.3.4 T-test

The population-level summary for the variable will be altered. The mean treatment difference in change from baseline to Week 48 in FEV_1, standard deviation and 95% confidence interval will be presented. A two-sample t-test will be performed by treatment group using change from baseline in FEV_1 to Week 48.

In case of death or re-transplantation events, FEV_1 will be imputed as zero at each nominal day post event.

14.3.5 "While on treatment" strategy"

The IE discontinuation of randomized treatment will be considered and treated following a while on treatment strategy. In case of such IE, all FEV_1 measurements up to the date of the IE will be included, while the observation post IE will not be considered. The analysis will be conducted as described for the primary endpoint (section 14.1).

As an additional analysis, the initiation of additional immunosuppressive therapy (see section 14.5.1 for details on additional immunosuppressive therapies) will be included in the analysis: IEs of either discontinuation of randomized treatment or initiation of additional immunosuppressive therapy will be analyzed according to the above while on treatment strategy.

In both analyses, the summary statistics of time to treatment discontinuation (weeks) and time to initiation of additional immunosuppressive therapy (weeks) (for the second analysis only) will be provided for a better interpretation of results.

In case of re-transplantation events, FEV_1 will be imputed as zero at each nominal day post event. According Table 5, in case of death, no imputation will be done.

14.3.6 Trimmed Means

The trimmed means analysis uses a different estimand and follows the approach of Permutt and Li (2017) [3]. This method provides an exact test using all randomized patients that is very robust and does not require any modeling assumptions. The estimand of the primary analysis is the difference between treatment groups in mean FEV_1 change from baseline at week 48 in the population of all randomized patients, including the patients who die or are re-transplanted and have no FEV_1 measurement at week 48. The primary analysis implicitly models the FEV_1 at week 48 for patients who died or were re-transplanted, which is somewhat unnatural. In the approach of Permutt and Li, all dropouts are considered to be failures and these patients are given a FEV_1 value of zero at week 48. The exact imputed value does not matter, as long as it is worse than all values observed in patients with FEV_1 measurements at week 48. The imputed value will be zero. The estimand is the between group difference in mean change from baseline in the percentage p subpopulation with the best FEV_1 values at week 48. Typical choices for p are in the range of 50 to 90%, depending on the anticipated percentage of patients with an imputed low value. For this sensitivity analysis p = 70% will be used. The method is based on a permutation test for the null hypothesis that the outcome distributions at week 48 are equal. The variant of the method described in Section 4.3 of Permutt and Li (2017) that exploits the intermediate FEV_1 values and covariates in the same way as the primary analysis will be used. The only requirement for this method is that it must be reasonable to consider the outcome of dropouts as worse than the lowest outcome of the completers. This is reasonable for dropout due to death or re-transplantation but may not always be reasonable for dropout due to other reasons. Therefore, imputation will be applied only in case of death, re-transplantation, or lack of efficacy IEs.

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

CCI

[redacted]

CCI	[redacted]

[redacted]

CCI [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

- [redacted]
  - [redacted]
  - [redacted]
  - [redacted]
  - [redacted]

[redacted]

CCI

CCI [redacted]

CCI [redacted]

14.6 SUBGROUP ANALYSES

Analyses of primary CCI [redacted] endpoints will be performed on the following subgroups of patients:

- Age class (<55 years, ≥55 years),
- Gender (female, male, undifferentiated),
- Region (North America, all other countries together),
- Severity of disease, defined by:
  - % of baseline FEV_1 over personal best FEV_1 post-transplant (baseline FEV_1 >80%, >65–80%, >50–65%, and ≤50% of personal best FEV_1 post-transplant),
  - baseline FEV_1 % of predicted (<70%, ≥ 70%),
  - baseline FEV_1/FVC ratio (< 0.7, ≥ 0.7),
  - BLAD at baseline (Yes, No);
- Race (American Indian or Alaskan Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Other),
- Use of azithromycin at randomization (Yes/No).

Subgroup levels may be pooled to have an adequate number (≥ 10) of patients per strata to perform the analysis. Decision on pooling will be taken during the blinded DRM, before data base lock.

Within each subgroup level, descriptive analyses (section 8.1) will be performed on:

- FEV_1,
- FEV_1/FVC,
- Time to Progression of BOS.

FEV_1 and FEV_1/FVC change from baseline will be compared using a two-sample t-test (assuming unequal variances) or, if the required assumptions are not met, a two-sample Mann–Whitney U test.

Comparison of time to progression of BOS curves among treatments will be performed with the log-rank test.

15 SAFETY DATA

15.1 General Considerations

Summaries of safety data (e.g., AEs, laboratory data, vital signs) will be provided for the SAF. No inferential analyses are planned for safety endpoints.

15.2 Extent of Exposure

15.2.1 Exposure to L-CsA

Exposure to the investigational medicinal product (IMP), L-CsA will be confirmed based on the report of IMP vials returned and dosing recorded in the eCRF. IMP vials that were missing will not be included in the counts of returned (i.e. used) vials; this will also apply where vials are missing but the patients were considered to have taken the vials. A summary table will be prepared including the following details about administration of L-CsA:

- Descriptive statistics for the total vials planned. This is calculated as ([date of last dose – date of randomization + 1] x 2 vials);
- Descriptive statistics for the total actual dose (mg). This is calculated as total empty (used) vials multiplied by dose;
- Descriptive statistics for the duration of treatment in days and weeks. For each patient, the treatment duration will be calculated as the difference between the dates of last and first administration of L-CsA plus 1;
- Descriptive statistics for the duration of treatment in days and weeks, excluding the inhalation temporarily discontinued. For each patient, the treatment duration will be calculated as the difference between the dates of last and first administration of L-CsA plus

1, excluding the days of any inhalation temporarily discontinued. Derivation: (date of last dose – date of first dose + 1) - (stop date of inhalation temporarily discontinued– start date of inhalation temporarily discontinued +1).

Compliance of L-CsA using the drug accountability data is calculated as the total empty (used) vials returned divided by the total vials planned and will be summarized on the FAS population. For the evaluation of compliance, all the planned vials up to the actual end of treatment for each patient will be considered in the calculation.

By-patient listings will present all IMP administration data.

15.2.2 Exposure to Study treatment (SoC or SoC+L-CsA)

SoC therapy period starts at randomization date and stops at date of study completion or early termination. L-CsA treatment period starts at date of first dose of L-CsA and stops at the date of last dose administered.

Descriptive statistics for the duration of any study treatment (SoC or SoC+L-CsA) in days and weeks will be provided. Since the SoC will cover the entire study period post-randomization, the study treatment duration (SoC or SoC+L-CsA) will be calculated as the difference between the date of study completion/early termination and randomization plus 1.

15.3 Adverse Events

All AEs occurring during the clinical trial, whether or not attributable to study treatment, will be documented in the eCRF. All AEs will be coded to SOCs and PTs using the MedDRA dictionary (version 23.1).

Treatment-emergent adverse events (TEAEs) are defined as AEs with onset on or after the date of randomization.

The Investigator will assess the relationship (unrelated, possible, definite) of each AE to the clinical trial treatment assignment (L-CsA plus SoC or SoC alone). Treatment-related AEs are defined as events the Investigator considers to be possibly or definitely related to clinical trial treatment assignment.

All AEs reported in the eCRF will be presented in by-patient listings for the SAF.

The following summaries will be presented for AEs by treatment group and overall:

- Overview table including the number and percent of patients with:
  - AEs,
  - TEAEs,
  - TEAEs leading to discontinuation of study treatment,
  - TEAEs leading study discontinuation,

  - TEAEs leading to death.
  - TEAEs by severity (Mild, Moderate, Severe),
  - Study treatment related TEAEs,
  - Serious TEAEs,
  - Serious study treatment related TEAEs;
- TEAEs, grouped within SOC by decreasing frequency of PT;
- TEAEs by decreasing frequency of PT;
- TEAEs leading to discontinuation of study treatment, grouped within SOC by decreasing frequency of PT;
- TEAEs leading to study discontinuation, grouped within SOC by decreasing frequency of PT;
- TEAEs leading to death, grouped within SOC by decreasing frequency of PT;
- TEAEs by maximum severity, grouped within SOC by decreasing frequency of PT;
- TEAEs by relationship, grouped within SOC by decreasing frequency of PT;
- Serious TEAEs, grouped within SOC by decreasing frequency of PT;
- Related serious TEAEs, grouped within SOC by decreasing frequency of PT.

Patients who have multiple events coded to the same PT or within the same SOC will be counted only once in the patient counts for the corresponding summaries, but all events will be considered and listed. Table presentations will contain counts of patients, percentages of patients, and the number of events.

Any AE with missing relationship will be assumed related and summarized in the possible related category to clinical trial treatment assignment. Patients who experience the same event multiple times will be included in the most related category for summarization.

In summaries including severity, the following severity categories will be summarized: ‘Mild’, ‘Moderate’, ‘Severe’. Any AEs with missing severity will be assumed as ‘Severe’. Patients who experience the same event multiple times will be included in the most severe category.

AE summaries will be ordered in terms of decreasing frequency for SOC, and PT within SOC, in the L-CsA plus SoC column.

The following by-patient listings will be provided:

- AEs;
- TEAEs;
- TEAEs leading to discontinuation of study treatment;
- TEAEs leading to study discontinuation;
- TEAEs leading to death;
- Study treatment related TEAEs;
- Serious TEAEs.

These listings will include at least: center, patient identifier, treatment group, AEs (SOC, PT, and verbatim term), AE start date/time, AE end date/time, duration of AE, severity, seriousness, action taken, outcome and relationship.

15.3.1 Infections

All infections will be reported as AEs on the AE eCRF. The summaries described in this section will be provided for treatment-emergent infections (defined as for TEAEs). Treatment-emergent infections will be summarized on the basis of information reported on the AE eCRF.

For any infection, the following summaries will be presented by treatment group:

- Number (percent) of patients with at least one infection,
- Number and % of infections per patient (0, 1, 2, …),
- Descriptive statistics for number of infections per patient.

The number of infections per patient and per patient-years (cumulative follow-up of each patient) will be provided by year (e.g., 2019, 2020, 2021, etc.) and overall.

A by-patient listing summarizing information on infections will be provided on SAF set.

15.3.2 Malignancies

All malignancies must be confirmed by histopathology. Malignancies are documented as AEs, and will be coded by SOC and PT. The summaries described in this section will be provided for treatment-emergent malignancies (defined as for TEAEs). Malignancies will be identified through a blinded medical review of collected AEs. The result of the medical review will be a flag for the identification of malignancies added in an excel sheet, that will be imported for analysis.

A summary table by SOC and PT and by-patient listing summarizing information collection will be provided on SAF.

15.4 Acute Tolerability of IMP

At Visit 1, acute tolerability of IMP (L-CsA) during initial dosing will be determined by measuring:

- Spirometry (FEV_1, CCI , FVC) measured before and CCI
- Spontaneous report of cough or dyspnea, recorded as AE in the appropriate eCRF.

Descriptive statistics of pre-dose (V1), post-dose CCI and change from pre-dose will be presented for each spirometry parameter on the SAF population.

In addition, the number and percent of patients experiencing a decline of CCI in FEV_1 will be presented.

The number and percent of patients experiencing new onset or worsening of TEAEs having the following PTs on the date of the first visit at which L-CsA is administered will be summarized:

- Cough;
- Dyspnea.

15.5 Clinical Laboratory Data

15.5.1 Hematology and Biochemistry

Clinical laboratory assessments will be performed at each visit by a certified local facility following the routine protocols at each site. As described in the clinical trial protocol (Section 13.1.5), the following clinical laboratory parameters will be reported in the eCRF:

Hematology

- White Blood Cell (WBC) count with differential,
- Red Blood Cell (RBC) count,
- Platelet count,
- Neutrophils,
- Lymphocytes,
- Monocytes,
- Eosinophils,
- Basophils,
- Neutrophils Absolute Count,
- Lymphocytes absolute counts,
- Eosinophils absolute count,
- Basophils absolute count,
- Monocyte absolute count.

Biochemistry

- Alkaline phosphatase (AP),
- Alanine aminotransferase (ALT),
- Aspartate aminotransferase (AST),
- Gamma Glutamyl Transferase (γ-GT),
- Blood urea nitrogen (BUN),
- Serum bilirubin,
- Serum creatinine,
- Sodium,
- Potassium,
- Chloride,

- Bicarbonate,
- Calcium,
- Magnesium.

Descriptive statistics of observed values and changes from baseline will be presented by visit and treatment group. The summaries described in this section will be presented relative to baseline, defined as the most recent, non-missing result on or before the date of randomization. No imputation for missing data will be performed.

All clinical laboratory data reported in the eCRF will be presented in by-patient listings.

15.5.2 Blood Sampling CNI/mTOR Inhibitors

A whole blood sample for the determination of blood sampling CNI/mTOR inhibitors will be collected at each visit; this includes cyclosporine A and/or tacrolimus trough level. In patients allocated to L-CsA plus SoC, the sample should be collected 60 min before L-CsA inhalation. Blood samples will be evaluated by the local laboratories associated with each site. The summaries described in this section will be presented relative to baseline, defined as the most recent, non-missing result on or before the date of randomization. No imputation for missing data will be performed.

All CNI/mTOR data reported in the eCRF will be presented in by-patient listings. Descriptive statistics of observed values and changes from baseline will be presented by visit and treatment group.

15.5.3 Donor Specific Antibody Tests

Serum samples for donor specific antibody testing will be collected at Screening, and at Visits 4, 6, and 9. A by-patient listing of all results will be provided.

15.5.4 Pregnancy Tests

Serum samples from women of childbearing potential will be collected for pregnancy. A by-patient listing of all results will be provided.

15.6 Renal Function

In addition to the summarization of serum creatinine as described in previous section, renal dysfunction will be further listed and summarized by treatment group based upon the classification of AEs with the following PTs (see protocol section 13.1.5.4, Table 7):

- Blood Creatinine Increased;
- Renal Impairment;
- Renal Failure.

15.7 Vital Signs

Vital signs will be assessed at each visit. The following vital signs will be recorded:

- Systolic and diastolic blood pressure (sitting, after 5 minutes of rest);
- Heart rate (sitting, after 5 minutes of rest);
- Body temperature (the same method should be used for each patient and throughout the trial);
- Respiratory rate;
- Weight.

The summaries described in this section will be presented relative to baseline, defined as the most recent, non-missing result on or before the date of randomization. No imputation for missing data will be performed.

Descriptive statistics of observed values and for changes from baseline will be presented by visit and treatment group on SAF. Vital signs data reported in the eCRF will be presented in by-patient listings.

15.8 Physical Examination

A physical examination will be carried out at Screening and every subsequent visit.

The following body systems will be examined: body weight, cardiovascular, respiratory, nervous, gastrointestinal, hepatic, renal, dermatological, musculoskeletal, extremities, eyes, ears, nose, throat, and lymphatic. Patient height will be measured at Screening Visit only.

All physical examination data, including abnormal findings reported in the eCRF will only be presented in by-patient listing.

16 Changes from Planned Analyses

The changes indicated below are being made, none of these changes are documented as formal protocol amendments.

- The SAF population has been further detailed to better define the actual treatment received by patients.
- CCI [redacted]
- The protocol amendment version 6.0 Table 1 displaying the 2019 CLAD staging has been amended in this SAP. Table 7 shows the correct descriptions for the 2019 CLAD Staging as reported in Verleden et al., 2019 [4].

- CCI [redacted]
- In section 9.4 further details compared to protocol have been added for handling visits that have more than one accepted spirometry assigned, with some of them performed after IMP inhalation.
- CCI [redacted]
- BOS Progressions will be evaluated up to last assessment, even after treatment discontinuation or completion.
- CCI [redacted]

17 CCI [redacted]

Please refer to Table 8 for details on names and content of the variables used in the CCI [redacted] of this section.

Table 8: Variable names and descriptions.

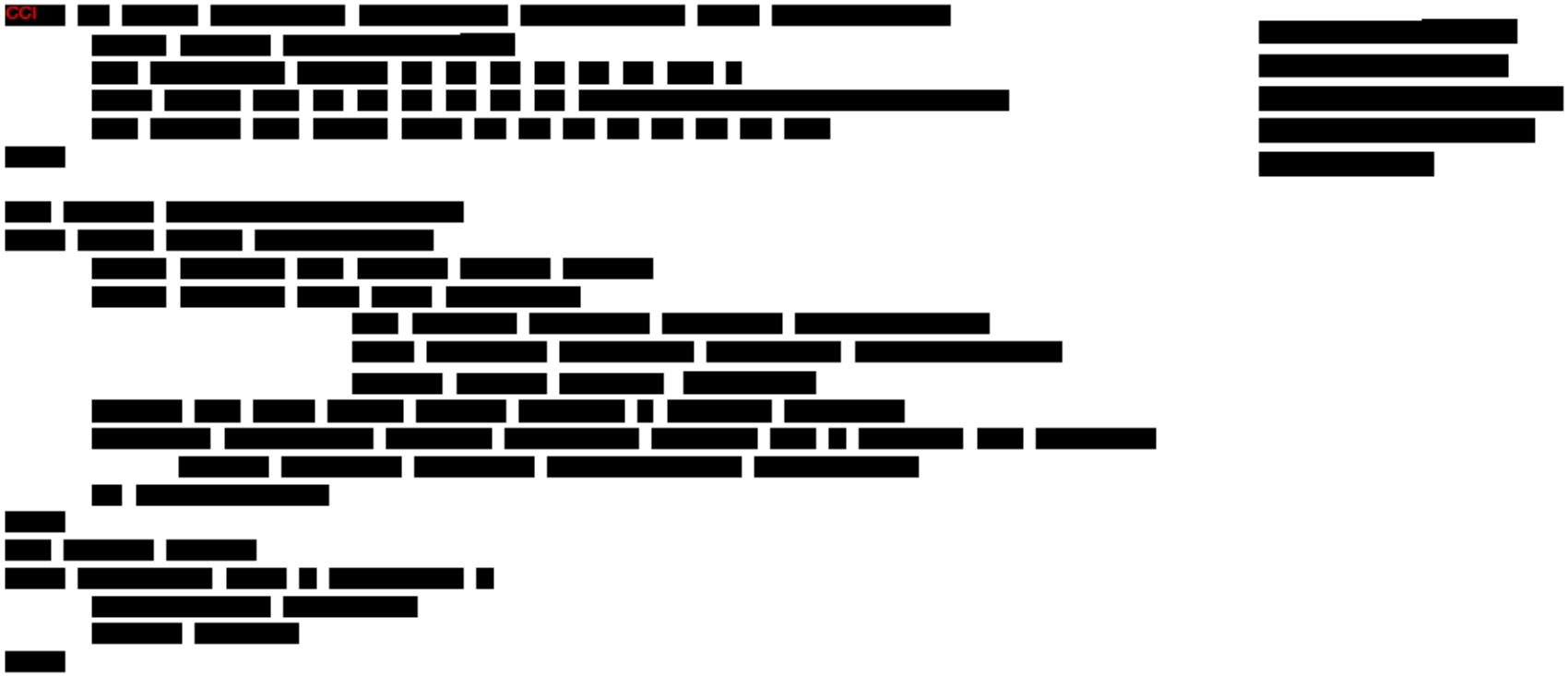
Variable Name	Description
CCI [redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

17.1 Primary analysis

CCI

17.2 Sensitivity analysis - MNAR

CCI



17.3 Sensitivity analysis – Tipping point

[redacted]

CCI [redacted]

[redacted]

CCI [redacted]

17.4 Sensitivity analysis - MMRM

CCI

17.5 Supplementary analysis - Piecewise regression

CCI

CCI

[redacted]

18 REFERENCES

1. Kieser M, Friede T (2003). Simple procedures for blinded sample size adjustment that do not affect the type I error rate. Stat Med, 22(23):3571-3581.
2. Rubin DB. Multiple imputation for nonresponse in surveys. John Wiley and Sons, Inc.: New York, 1987.
3. Permutt T, Li F (2017). Trimmed means for symptom trials with dropouts. Pharm Stat, 16(1):20-28.
4. Verleden GM, Glanville AR, Lease ED, et al. Chronic lung allograft dysfunction: Definition, diagnostic criteria and approaches to treatment. A consensus report from the Pulmonary Council of the ISHLT, J Heart Lung Transplant 2019;5:493-503.
5. de Zwart AES, Riezebos-Brilman A, Lunter GA, Neerken ECU, van Leer-Buter CC, Alffenaar JC, van Gemert AP, Erasmus ME, Gan CT, Kerstjens HAM, Vonk JM, Verschuuren EAM. Impact of COVID-19 social distancing measures on lung transplant recipients: decline in overall respiratory virus infections is associated with stabilisation of lung function. Eur Respir J. 2022 Nov 24;60(5):2200085. doi: 10.1183/13993003.00085-2022. PMID: 35896214; PMCID: PMC9301935.
6. Fine JP and Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. J Am Stat Assoc 1999; 94:496–509.
7. https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/valuation-standard-value-sets/.
8. https://www.lshtm.ac.uk/research/centres-projects-groups/missing-data#dia-missing-data.
9. Interim Analysis documents: Interim Analysis Summary Dated 01Dec2021, and Interim Analysis Communication Plan, Final Version 1.0 dated 23 August 2021.
10. Liu J, Jackson K, Weinkauf J, Kapasi A, Hirji A, Meyer S, Mullen J, Nagendran J, Lien D, Halloran K. Baseline lung allograft dysfunction is associated with impaired survival after double-lung transplantation. J Heart Lung Transplant. 2018 Jul;37(7):895-902. doi: 10.1016/j.healun.2018.02.014. Epub 2018 Feb 27. PMID: 29602706.
11. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. Am J Respir Crit Care Med. 1999 Jan;159(1):179-87. doi: 10.1164/ajrccm.159.1.9712108. PMID: 9872837.
12. Hankinson, J. L., Kawut, S. M., Shahar, E., Smith, L. J., Stukovsky, K. H., & Graham Barr, R. (2010). Performance of american thoracic society-recommended spirometry reference values in a multiethnic sample of adults the multi-ethnic study of atherosclerosis (MESA) lung study. *Chest*, *137*(1), 138–145. https://doi.org/10.1378/chest.09-0919

19 APPENDICES

CCI

INTERIM ANALYSIS SUMMARY

Interim Analysis Summary

Study Title and Version:	A Phase III, Prospective, Multicenter, Randomized, Controlled Clinical Trial to Demonstrate the Efficacy and Safety of Liposomal Cyclosporine A (L-CsA) Inhalation Solution [CCI redacted] Standard of Care versus Standard of Care Alone in the Treatment of Chronic Lung Allograft Dysfunction / Bronchiolitis Obliterans Syndrome in Patients post Double Lung Transplantation Protocol version 4.0 dated 19 January 2021 Protocol Number: BT – L-CsA – 302 – DLT
Sponsor:	Zambon SpA Via Lillo del Duca, 10 Bresso (MI) 20091 Italy
Date of Document Distribution:	Data Snapshot Date: 30 Sep 2021 Delivered Date: 01Dec2021

CONFIDENTIALITY STATEMENT

19.2 Interim analysis communication plan

Zambon SpA

Study Title:	Interim Analysis Communication Plan
Name of Test Product:	Liposomal Cyclosporine A (L-CsA) CCI
Protocol Number:	BT – L-CsA – 302 – DLT (BOSTON-2)
Protocol Version:	Version: 4.0
Protocol Date:	19 January 2021
Covance Study Number	000000178421
Communication Plan Version:	Final Version 1.0
Communication Plan Date:	23 August 2021
Communication Plan Author:	PPD
