Official Protocol Title:	A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate Sotatercept When Added to Maximum Tolerated Background Therapy in Participants With Pulmonary Arterial Hypertension (PAH) World Health Organization (WHO) Functional Class (FC) III or FC IV at High Risk of Mortality
NCT number:	NCT04896008
Document Date:	16-Oct-2024

STATISTICAL ANALYSIS PLAN
PHASE III

VERSION: Amendment 3
DATE OF PLAN:
16-Oct-2024

BASED ON:

Protocol Revision 4.0 Amendment 03 of April 23, 2024 (MK-7962-006-07)

STUDY DRUG:

Sotatercept

PROTOCOL NUMBER (STUDY NAME):

A011-14/MK-7962-006(ZENITH)

STUDY TITLE:

A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate Sotatercept When Added to Maximum Tolerated Background Therapy in Participants With Pulmonary Arterial Hypertension (PAH) World Health Organization (WHO) Functional Class (FC) III or FC IV at High Risk of Mortality

SPONSOR:

Acceleron Pharma Inc., a wholly owned subsidiary of Merck & Co., Inc., Rahway, NJ, USA

120 East Lincoln Avenue

P.O Box 2000

Rahway, NJ 07065 USA

This study is being conducted in compliance with good clinical practice, including the archiving of essential documents.

TABLE OF CONTENTS

LIST OF TABLES

1 LIST OF ABBREVIATIONS

Table 1 List of Abbreviations

Abbreviation	Term
6MWD	Six-minute walk distance
6MWT	Six-minute walk test
ADA	Anti-drug Antibody
AE	Adverse Event
AEOI	Adverse Events of Interest
AESI	Adverse Events of Special Interest
ALT	Alanine Aminotransferase
ANC	Absolute Neutrophils Count
ANCOVA	Analysis of Covariance
AST	Aspartate Aminotransferase
BMI	Body Mass Index
CMH	Cochran Mantel-Haenszel
CRF	Case Report Form
CRP	C-reactive protein
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for Adverse Events
DBP	Diastolic Blood Pressure
DBPC	Double-blind Placebo-controlled
DMC	Data Monitoring Committee
DOB	Date of Birth
ECG	Electrocardiogram
EQ-5D-5L	EuroQol – 5 dimensions scale 5 levels
eGFR	estimated Glomerular Filtration Rate
FAS	Full Analysis Set
FCS	Fully Conditional Specification
GGT	Gamma-Glutamyl Transferase
Hgb	Hemoglobin
IA	Interim Analysis
ICH	International Conference on Harmonization
IRT	Interactive Response Therapy

Abbreviation	Term
LLN	Lower Limit of Normal
LTDB	Long Term Double-blind
MAR	Missing at Random
MedDRA	Medical Dictionary for Regulatory Activities Terminology
MI	Multiple Imputation
N	Total Sample Size
PAH	Pulmonary Arterial Hypertension
PC	Platelet Counts
PRO FAS	Patient Report Output Full Analysis Set
PT	Preferred Term
PVR	Pulmonary Vascular Resistance
REVEAL	Registry to Evaluate Early and Long-Term PAH Disease Management
RVSP	Right Ventricular Systolic Pressure
SAP	Statistical Analysis Plan
SAE	Serious Adverse Event
SAS	Statistical Analysis System
SBP	Systolic Blood Pressure
SD	Standard Deviation
SEM	Standard Error of the Mean
SMQ	Standard MedDRA Queries
SOC	System Organ Class
TAPSE	Tricuspid Annular Plane Systolic Excursion
TEAE	Treatment Emergent Adverse Events
TTCW	Time to Clinical Worsening
ULN	Upper Limit of Normal
VAS	Visual Analogue Scale
WBC	White Blood Cell Count
WHO	World Health Organization
WHO FC	World Health Organization Functional Class

2 INTRODUCTION

The statistical analysis plan (SAP) is based on:

- Protocol No. A011-14 (MK-7962-006) (ZENITH), Global Amendment 03 (v4.0) approved on April 23, 2024

- ICH guidelines E9 (Statistical Principles for Clinical Trials) and E9(R1) (Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials)

The purpose of this document is to provide details on study populations and on how the variables will be derived, how missing data will be handled as well as details on statistical methods to be used to analyze the safety and efficacy data.

The SAP will be finalized and approved before the database is locked. Deviations from the final approved plan will be noted in the clinical study report.

2.1 Changes From Statistical Section of the Protocol

Table 2 Changes from Statistical Section of Protocol

CCI

[redacted]

2.2 Summary of Changes from Previous Versions of the SAP

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

3 STUDY OBJECTIVES AND ENDPOINTS

3.1 Study Objectives

The objective of this study is to evaluate the efficacy and safety of sotatercept (plus maximum tolerated background PAH therapy) versus placebo (plus maximum tolerated background PAH therapy) on time to first event of all-cause death, lung transplantation, or PAH worsening-related hospitalization of ≥24 hours, in participants with WHO FC III or FC IV PAH at high risk of mortality.

3.2 Study Endpoints

3.2.1 Primary Endpoint

The primary efficacy endpoint is the time to first event of all-cause death, lung transplantation, or PAH worsening-related hospitalization of ≥ 24 hours. An independent blinded adjudication committee will adjudicate all clinical worsening events, including death that occurred while the participants are in the study and deaths caused by any serious adverse events (SAE) during the study, up to the end of the study to determine whether these events are due to PAH. Only adjudication-confirmed lung transplantation and hospitalization of ≥ 24 hours will be included in the primary analysis. All deaths that are a first event for a participant, whether occurring during the study or following early discontinuation, will be included regardless of adjudication.

3.2.2 Secondary Endpoints

The following are the secondary efficacy endpoints, listed according to the order in which they will be tested:

1. Overall survival. The primary approach for this endpoint will include all deaths up to the data cutoff date, except for those that occurred after enrollment in the long-term follow-up study (SOTERIA) or after lung transplantation. Deaths that were obtained by the collection of vital status among participants who (1) completed the study or discontinued prematurely and (2) did not enroll in SOTERIA, will be included.

2. Transplant-free survival. The primary approach for this endpoint will include all events that occurred prior to the cutoff date, including deaths that were obtained by the collection of vital status for participants who completed the study or withdrew prematurely without lung transplantation. Deaths that occurred after enrollment in SOTERIA will not be included in the primary analysis.

3. Proportion of participants who experienced a mortality event at end of study (EOS). This endpoint will include the same events as defined for the primary approach for the overall survival.

4. Change from baseline in Registry to Evaluate Early and Long-Term PAH Disease Management (REVEAL) Lite 2.0 risk score at Week 24

a. The REVEAL Lite 2.0 risk score is calculated based on the values of the following variables: renal insufficiency (by estimated glomerular filtration rate [eGFR]), WHO FC, systolic blood pressure (SBP), heart rate, 6-minute walk distance (6MWD), and N-terminal prohormone B-type natriuretic peptide (NT-proBNP). Scores are assigned to each of these variables based upon their presentation and contribution to mortality risk, and a total score is obtained. See [Appendix 12.1] for the details of the REVEAL Lite 2.0 risk score calculator.

5. Proportion of participants achieving a low or intermediate (≤ 7) REVEAL Lite 2.0 risk score at Week 24: only participants with a REVEAL Lite 2.0 risk score > 7 at baseline will be included in the analyses.

6. Change from baseline in NT-proBNP levels at Week 24

7. Change from baseline in mean pulmonary artery pressure (mPAP) at Week 24

8. Change from baseline in PVR at Week 24

9. Proportion of participants who improve in WHO FC at the end of the double-blind placebo-controlled (DBPC) Treatment Period

10. Change from baseline in 6MWD at Week 24

11. Change from baseline in cardiac output (CO) at Week 24

12. Change from baseline in EuroQoL-5 dimensions scale 5 levels (EQ-5D-5L) index score at Week 24

3.2.3 Exploratory Endpoints

CCI

[redacted]

CCI

[redacted]

CCI [redacted]

3.2.4 Safety Endpoints

Safety endpoints include the following:

- Adverse events (AEs)
- Anti-drug antibodies (ADAs)
- Laboratory assessments (hematology, serum chemistry, and urinalysis)
- Vital signs (systolic blood pressure, diastolic blood pressure, pulse rate, respiratory rate, temperature, weight, body mass index)
- Physical examination
- 12-lead electrocardiogram (ECG)

4 STUDY DESIGN

4.1 Summary of Study Design

This is a Phase 3, randomized, double-blind, placebo-controlled, multicenter, parallel-group study to evaluate sotatercept versus placebo in participants with PAH WHO FC III or FC IV who are at high risk of mortality.

The study population includes participants with symptomatic PAH (WHO FC III or FC IV at high risk of mortality) who present with idiopathic or heritable PAH, PAH associated with connective tissue diseases (CTD), drug- or toxin-induced, post-shunt correction PAH, or PAH presenting at least 1 year following the correction of congenital heart defects (CHD). Participants must have a REVEAL Lite 2.0 risk score of ≥ 9 and be on maximum tolerated combination background PAH therapy.

A planned interim analysis (IA) will occur when approximately 59 participants have experienced a primary endpoint event (roughly 50% of the required number of events) and median participant time on study is at least 6 months. The time on study for each participant is calculated from the randomization date to the database cutoff date, study discontinuation date, or the onset date of the primary endpoint, whichever comes first. If the study continues after the IA, the final analysis will happen when approximately 118 participants have experienced a primary endpoint.

Each participant will be enrolled in the study for up to approximately 43 months as follows:

- Screening Period (up to 4 weeks)
- DBPC Treatment Period (up to approximately 40 months)
- Follow-up Period (up to 8 weeks)

Each participant will remain in the DBPC Treatment Period until one of the following occurs, whichever comes first: 1) they experience the first event of all-cause death, lung transplantation, or PAH worsening-related hospitalization of ≥24 hours; 2) the time when the required number of primary events are accrued for the final analysis; 3) the study is stopped early at the IA for either efficacy or futility. Study participants who have not experienced an event will remain in the DBPC Treatment Period until the required number of participants have experienced a first event of all-cause death, lung transplantation, or PAH worsening-related hospitalization.

The DBPC treatment period starts when the first participant receives the first dose of the treatment and ends when the first of the following occurs: 1) at least 118 participants have experienced a primary endpoint event for the final analysis; 2) the study is stopped early at the IA for either efficacy or futility.

4.2 Definition of Study Drugs

Investigational treatments include:

- Placebo administered subcutaneously (SC) every 21±3 days plus background PAH therapy

- Sotatercept at a starting dose of 0.3 mg/kg SC with a target dose of 0.7 mg/kg SC every 21±3 days plus background PAH therapy

Background PAH therapy refers to approved PAH-specific medications. Study participants must be stable on maximum tolerated double or triple combination background PAH therapy (per the investigator’s judgment) for at least 30 days prior to the Screening Visit. Adjustments in parenteral prostacyclin doses by up to 10% are permitted and should not affect therapy stability determination.

4.3 Sample Size Considerations

4.3.1 Sample Size Justification

The sample size determination is based on the primary efficacy endpoint of time to first event of all-cause death, lung transplantation, or PAH worsening-related hospitalization of ≥24 hours using EAST® version 6.4. In STELLAR, the hazard ratio (HR) in the sotatercept group compared with the placebo group was 0.16 (95% CI: 0.08 to 0.35) [Hoeper, M. M., et al 2023]. Given the differences in the populations and definitions of endpoints between STELLAR and this study, the HR is assumed to be 0.55 in this study.

Assuming a HR of 0.55, a 1:1 randomization, a 1-sided 0.025 Type 1 error rate, 90% power, and with a planned IA at approximately 50% of the required number of events with the option to stop the study for futility, approximately 118 events will be required based on the log-rank test.

Given that approximately 166 participants are planned to be enrolled in this study, the accrual period is approximately 26 months, assuming an accrual rate of approximately 6.5 participants per month. In addition, assuming a dropout hazard rate of 0.04% per month (0.5% per year), and the probability of observing an event for placebo is 0.45 for the first year, 0.60 for the second year, and 0.90 for the third year and later, the projected time of the IA will occur around 26 months. If the study continues after IA, the final analysis will happen around 40 months. Median participant time on study must be at least 6 months in order for analyses following the occurrence of the required number of events.

4.4 Randomization

The randomization schedule is stratified by REVEAL Lite 2.0 risk score (9 to 10 or ≥ 11) and PAH subtype (CTD-associated or not CTD-associated) at screening. Additional details on the randomization schedule can be found in the randomization specifications document in the study trial master file.

Participants who have signed the informed consent and meet all eligibility criteria will be stratified by REVEAL Lite 2.0 risk score and PAH subtype and then randomized in a 1:1 ratio to receive placebo plus maximum tolerated background PAH therapy or sotatercept plus maximum tolerated background PAH therapy.

Randomization assignments will be generated through a computerized system, provided by an Interactive Response Technology (IRT).

4.5 Clinical Assessments

The schedule of clinical assessments can be found in the study protocol (Section 2).

5 PLANNED ANALYSES

5.1 Interim Analyses

One IA of the primary efficacy endpoint is planned to occur when approximately 59 participants have experienced a primary endpoint event (roughly 50% of the required number of events). The IA will include data only up to a cutoff date defined prior to the interim database lock. The stratified log-rank test with randomization factors as strata will be used for the analysis of the primary efficacy endpoint. The point estimate of the HR with 95% CI will be estimated by a Cox regression model stratified by the randomization factors.

The IA will be performed by an unblinded independent statistics provider and will be presented to the data monitoring committee (DMC) where a recommendation will be communicated to the Executive Oversight Committee (EOC), which is comprised of members of Sponsor Senior Management. The EOC will receive and decide upon any recommendations made by the DMC regarding the study. [Table 3] shows the boundary properties for the planned interim and final analysis of the primary endpoint. The efficacy boundary is derived using a Lan-DeMets spending function approximating O'Brien-Fleming bounds and the futility boundary is derived using a gamma family spending function approximating Hwang-Shi-Decani bounds with gamma = -7.

Table 3 Efficacy and Futility Boundaries and Properties of the Primary Endpoint

Analysis	Value	Efficacy	Futility
IA: 50% [a] information fraction **Required events: 59** **Timing: 26 months** **N: 166**	Z	2.963	-0.458
	p (1-sided)[b]	0.0015	0.677
	HR at boundary[c]	0.461	1.127
Final Analysis: **Required events: 118** **Timing: 40 months** **N: 166**	Z	1.969	NA
	p (1-sided)	0.0245	NA
	HR at boundary	0.695	NA

HR = hazard ratio; IA = interim analysis. The number of events and timings are estimated.

a Percentage of total planned events at the IA.

b p (1-sided) is the nominal α for group sequential testing.

c The HR at boundary is the approximate HR required to reach an efficacy/futility bound.

In the scenario that the event accumulation at IA is different from expected, i.e., the number of observed events at the time of database lock (DBL) is more than 59, the alpha spending at the IA will be based on the information fraction calculated as the actual number of events at IA over the target number of events at final analysis.

If the efficacy boundary is crossed for the primary endpoint at the IA, then analyses of secondary endpoints will be performed using a gatekeeping method. The 1-sided type 1 error rate for the evaluation of secondary endpoints will be the same as that used for the primary hypothesis at the IA. More details of the analysis methods for secondary endpoints are described in (Section 8.7).

The DMC has responsibility for assessment of overall risk/benefit. As such, the DMC may request to look at efficacy data at times other than the prespecified IA. If an unplanned efficacy look is prompted by safety concerns without the potential to stop for efficacy before the prespecified interim analysis, this will not require a multiplicity adjustment typically associated with the planned efficacy interim analysis; however, to account for any multiplicity concerns in this case, a sensitivity analysis that reduces the final alpha by 0.0001 (1-sided) will be conducted. If an unplanned efficacy look is conducted with the potential to stop the study for (positive) efficacy before the prespecified interim analysis, an alpha of 0.0001 (1-sided) will be applied to the primary endpoint at the time of the analysis, and if the study is not stopped, the same alpha will be deducted from the alpha specified for the planned interim analysis and the final analysis.

5.2 Final Analyses

If the study continues after the IA, the final analysis is planned to occur when approximately 118 participants have experienced a primary endpoint event. The final analysis will use the remaining type I error that was not spent at the earlier analysis. The p-value bound at the final primary analysis will be calculated by considering the type I error spending using the information fraction as determined by the actual number of events at IA of the study and the expected number of events (118) at the final analysis, with the correlations of the test statistics between IA and FA as well as the HR boundary adjusted if the actual number of final events differs from 118.

If the efficacy boundary is crossed for the primary endpoint at the final analyses, then analyses of secondary endpoints will be performed using a gatekeeping method. For all secondary endpoints except for the proportion of participants who experienced a mortality event at EOS (secondary endpoint 3) and the proportion of participants who improve in WHO FC at the end of DBPC (secondary endpoint 9), the p-value boundary for the secondary endpoints will be updated using the same remaining type I error spending as used for the primary endpoint, with the correlations determined by the observed information at the final and interim analyses. For the secondary endpoints 3 and 9, for which the information fraction and correlation are not clearly defined, the p-value boundary at the final analysis will be adjusted using the Bonferroni adjustment, i.e., 0.025 minus the p-value boundary from the IA. This is a conservative adjustment for secondary endpoints 3 and 9 that does not attempt to take advantage of the correlation between the interim and final analyses.

6 GENERAL CONSIDERATIONS FOR DATA ANALYSES AND HANDLING

6.1 General Summary Table and Individual Participant Data Listing Considerations

All summary tables and figures as well as individual participant data listings will include a "footer" that will include the data source.

6.2 General Post Text Summary Table, Figure, and Individual Participant Data Listing Format Considerations

The default convention is to number summary tables, figures, and listings using a decimal system to reflect main levels of unique tables, figures, and listings and sub-levels of replicate tables and listings with two digits per level (e.g., Table XX. Y. Z. …). In general, summary tables and figures will occupy Appendix 14 of the CSR so that the table or figure number should start with 14 (e.g., Table 14. Y. Z. …). Individual participant data listings will occupy Appendix 16 of the CSR so that the listing number should start with 16 (e.g., Listing 16. Y. Z. …).

6.3 Data Management

Derived datasets will be created using SAS® software. Derived datasets, summary tables, summary figures, statistical analyses, and individual participant data listings will be generated using SAS version 9.4 or above.

6.4 Data Presentation Conventions

Continuous variables (e.g., age) are summarized using descriptive statistics (the number of participants with available data, the mean, standard deviation (SD), median and minimum and maximum). Categorical variables (e.g., race) are summarized using counts and percentages. Percentages are calculated using the total participants per treatment group.

The following conventions are applied to all data presentations and summaries.

- For continuous variables, all mean and median values are formatted to one more decimal place than the measured value. Standard deviation values are formatted to two more decimal places than the measured value. Minimum and maximum values are presented with the same number of decimal places as the measured value.

- For categorical variables, the number and percentage of responses are presented in the form XX (XX.X) where the percentage is in the parentheses.

- Date variables are formatted as DDMMMYYYY for presentation. Time is formatted in military time as HH:MM for presentation.

- Wherever possible, data will be decimal aligned.

- P-values, if applicable, will be presented to 3 decimal places. If the p-value is less than 0.001 then it will be presented as <0.001. If the rounded result is a value of 1.000, it will be displayed as >0.999.

- Unless otherwise stated, any statistical tests performed will use 2-sided tests at the 5% significance level.

6.5 Analysis Populations

6.5.1 Screen Failures

Screen failures are defined as participants who consent to participate in the clinical study but are not subsequently randomized to receive study treatment.

Individuals who do not meet the criteria for participation in the study (screen failure) may be rescreened once with the approval of the study medical monitor. A rescreened participant will be assigned a new participant number. The two participant numbers are linked in the data so that each participant will be counted once in the summaries. A participant is considered as randomized if the participant is screen failed the first time and meet the criteria for participating in the study in the re-screening process.

6.5.2 Safety Set

The Safety Set is defined as all participants who receive at least one dose of study treatment. All participants will be analyzed according to the treatment they received.

6.5.3 Full Analysis Set

The Full Analysis Set (FAS) is defined as all randomized participants, regardless of whether study treatment was administered, with the exception of one participant who was randomized in error and immediately discontinued by the site. This participant was never dosed and it is the only participant that did not receive any dose of the study treatment. Participants in the FAS will be analyzed according to the treatment group to which they were randomized.

6.5.4 Patient Report Outcome Full Analysis Set (PRO FAS)

The EQ-5D-5L analyses will be performed in the PRO FAS population. This population is the subset of FAS participants who had at least one dose of study medication and completed at least one baseline or post-baseline PRO assessment. Participants will be included in the treatment group to which they are randomized for the analyses of PRO data using the PRO FAS population.

6.6 Baseline Definition

6.6.1 Secondary Efficacy Endpoints

The baseline REVEAL Lite 2.0 risk score is calculated based on the values of the following variables measured prior to the first dose: eGFR, WHO FC, SBP, heart rate, 6MWD, NT-proBNP.

For the eGFR, WHO FC, systolic BP, and heart rate, measurements taken at Visit 1 will be used as the baseline values; if the measurements at Visit 1 are missing, the corresponding values at Screening will be used.

The baseline 6MWD is derived using the data from the 6-minute walk test (6MWT) performed at Visit 1 or Screening, prior to the first dose. If a participant discontinues the 6MWT prematurely, the total distance walked at the time of discontinuation will be the 6MWD used in the analysis. The Screening 6MWT is performed twice at least 4 hours apart, but no longer than 1 week apart. The following rules will describe how the baseline 6MWD is to be derived in general scenarios:

- If at least one 6MWD measurement is present prior to the first study drug administration at Visit 1, then the corresponding last measurement prior to the time of the first dose will be used as the baseline.

- If the 6MWD measurement at Visit 1 is not available then:
  - If the two 6MWD screening measurements are present, then the average of the two screening 6MWD measurements will be used as the baseline.
  - If one of the two 6MWD screening measurements is missing and there is no other 6MWT done prior to the first study drug administration, then the non-missing 6MWD measurement will be used as the baseline.

The baseline for mPAP, PVR, CO is the assessment taken at Screening. If the Screening assessment is missing, then an unscheduled assessment may be used if it was done prior to the first dose of study medication. No imputations will be performed for missing data.

For all other variables such as NT-proBNP, WHO FC, EQ-5D-5L index score, the baseline is the last value on or before Visit 1 (pre-dose).

6.6.2 Exploratory Efficacy Endpoints

CCI

[redacted]

6.6.3 Safety data

For safety data (vital signs, laboratory data, ECG), the baseline is defined as the last observation prior to the first dose of study treatment.

The first dose date will serve as the reference from which the non-missing pre-treatment measurements would be identified. For participants that are randomized but do not receive any study drug administration, the date of randomization will be the baseline reference.

6.7 Derived and Transformed Data

6.7.1 Study Day

If the date of interest occurs on or after the first dose/randomization date, then study day will be calculated as (date of interest – date of first dose/randomization) + 1. If the date of interest occurs prior to the first dose/randomization date, then study day will be calculated as (date of interest – date of first dose/randomization). There is no study day 0.

6.7.2 Change from Baseline

Change from baseline is calculated as (post-baseline result – baseline result).

6.7.3 Analysis Visit Windows

Depending on the frequency of the measurements of the endpoints [Table 4], the analyses windows are different across endpoints. [Table 5] gives a summary of the endpoints measured at each visit per schedule of events (Section 2) in the protocol.

Table 4 Evaluation Frequency for Each Endpoints

Visit schedule	Analysis Visit	Efficacy Endpoints	Safety Endpoints
Visit Schedule 1	Every 4 visits starting from Visit 1	• REVEAL List 2.0 risk score (Consists of eGFR, WHO FC, BP, heart rate, 6MWD, NT-proBNP) • EQ-5D-5L	• Serum chemistry • Urinalysis
Visit Schedule 2	Visit 1, Visit 2, Visit 3, Visit 4, Visit 5 and every 4 visits after Visit 4	• NT-proBNP • WHO FC	• ADA • Vital Signs • Hematology
Visit Schedule 3	Visit 1, Visit 2, Visit 5, and every 4 visits after Visit 5	• 6MWD	
Visit Schedule 4	Screening Visit, Visit 9	• mPAP • PVR • CO	• ECG

The analysis window for each endpoint is defined by following the rule below:

- Target day of each visit is defined as the first day of the week the participant should come in for that visit. For example, Visit 1 is defined as Day 1, correspondingly Visit 2 should happen in Week 3, therefore target day of Visit 2 is Day 22 ($1 + 3 \times 7$).
- The starting day of the analysis window is defined as the floor (average of the target day of current visit and previous visit according to the respective visit schedule). For example, the start day of Visit 3 is $floor(\frac{22+43}{2}) = 32$, where $43 = 1 + 6 \times 7$ is the target day of Visit 3 (if the endpoint is measured at Visit 3).
- To avoid any gap between the analysis window for two consecutive analysis visits, the end day of the analysis window is defined as the starting day of the analysis window for the next visit – 1. For example, the end day of Visit 2 is $floor\left(\frac{43+22}{2}\right) - 1 = 31$.

Based on the above definition, [Table 5] provides an example of the analysis windows for REVEAL Lite 2.0 risk score.

For measurements that are analyzed/summarized at specific time points, the value of measurements at each time point used in the analysis will be determined using visit windows as defined above rather than the name of the visit in database.

If there are multiple measurements for a participant within an analysis window, the measurement that is closest to the target day will be used. If days to the target day are the same between the two measurements, the measurement that is taken later will be used. If there are multiple measurements collected on the same day, the average measurement will be used.

Table 5 Analysis Visit Windows

Visit Schedule 1		
Analysis Visit	**Target Day**	**Analysis Window [Study Day Relative to First Dose]**
Screening	-28	Screening visit
Baseline	1	Last observation prior to first dose[a]
Visit 5 (Week 12)	85	2 to 126
Visit 9 (Week 24)	169	127 to 210
Visit 13 (Week 36)	253	211 to 294
Visit 17 (Week 48)	337	295 to 378
Visit 21 (Week 60)	421	379 to 462
Visit 25 (Week 72)	505	463 to 546
Visit Schedule 2		
Analysis Visit	**Target Day**	**Analysis Window [Study Day Relative to First Dose]**
Screening	-28	Screening visit
Baseline	1	Last observation prior to first dose
Visit 2 (Week 3)	22	2 to 31
Visit 3 (Week 6)	43	32 to 52
Visit 4 (Week 9)	64	53 to 73
Visit 5 (Week 12)	85	74 to 126
Visit 9 (Week 24)	169	127 to 210
Visit 13 (Week 36)	253	211 to 294
Visit 17 (Week 48)	337	295 to 378
Visit 21 (Week 60)	421	379 to 462
Visit 25 (Week 72)	505	463 to 546
Visit Schedule 3		
Analysis Visit	**Target Day**	**Analysis Window [Study Day Relative to First Dose]**
Screening	-28	Screening visit
Baseline	1	Last observation prior to first dose
Visit 2 (Week 3)	22	2 to 52
Visit 5 (Week 12)	85	53 to 126
Visit 9 (Week 24)	169	127 to 210
Visit 13 (Week 36)	253	211 to 294
Visit 17 (Week 48)	337	295 to 378
Visit 21 (Week 60)	421	379 to 462
Visit 25 (Week 72)	505	463 to 546
Visit Schedule 4		
Analysis Visit	**Target Day**	**Analysis Window [Study Day Relative to First Dose]**
Screening (Baseline)	-28	Last observation prior to first dose
Visit 9 (Week 24)	169	127 to 210[b]

[a] The analysis window for baseline starts at Day 1 but after the first dose.

[b] For the endpoints that are measured under visit schedule 4, if there is any measurement conducted at Visit 9 that is after Day 210 (visit out of window), the measurements will be included in the analyses given there is no future visits scheduled for the measurements.

6.8 Handling of Missing Data

6.8.1 Missing Efficacy Endpoints

6.8.1.1 Primary Efficacy Endpoint

Participants who are in the study and do not experience any of the components of the primary endpoint at the time of the data cutoff will be censored at the time of the data cutoff. Participants who discontinue from the study or are lost to follow-up before experiencing any of the components of the primary endpoint will be censored at the last known study contact record. This can be 1) the study withdrawal date or 2) the database cutoff date, whichever comes first. Other censoring rules of the primary endpoint for the sensitivity analyses are described in [Sec. 8.6.3].

6.8.1.2 Secondary endpoints

Time-to-event endpoints

For the primary analysis of the overall survival endpoint, participants who have a lung transplantation in ZENITH will be censored at the date of lung transplantation; participants who enroll to SOTERIA will be censored at the ZENITH study completion date; other participants who do not report a death at the time of the data cutoff will be censored at the earlier of the data cutoff date and last known alive. The last known alive date can be 1) the last contact date in the vital status follow-up (if available) or 2) the study discontinuation date, whichever is later.

For transplant-free survival, the censoring rule for the primary analysis is the same as the primary endpoint.

Other censoring rules of for the sensitivity analyses for the above two endpoints are described in [Table 8].

Continuous endpoints

For continuous endpoints, multiple imputation (MI) will be used to impute missing data for reasons other than death or a non-fatal clinical worsening event. For those with non-existent data due to death or missing data due to a non-fatal clinical worsening event, [Sec. 8.7] describes details of handling these data.

For the IA, data will not be imputed for ongoing participants who had not completed Week 24 at the time of the database cutoff as the non-existent change from baseline at Week 24 is missing completely at random.

The Missing at Random (MAR) assumption is made to perform Multiple Imputation (MI). It has been shown [Mogg, R. and Mehrotra, D. V. 2007] that MAR-based imputation under non-MAR conditions is unlikely to impact the overall treatment-level mean ranks. For missing points, Fully Conditional Specification (FCS) regression [van Buuren, S., et al 2006] [van Buuren, S. 2007] is used to fill in the missing points in the order of timepoints using measurements calculated at the previous timepoints. The analysis involves the following steps:

1. The missing data are filled in *m* times to generate *m* (where m = 100) complete datasets using an FCS regression model accounting for the baseline measurement, treatment group, and prior to the efficacy assessment at time point "X" that is to be imputed. For each participant requiring imputation, this imputation will be performed within the relevant stratum (defined by the participant's screening REVEAL Lite 2.0 risk score and PAH subtype).

2. The *m* complete datasets are analyzed by using the analysis described in [Sec. 8.7.2]

3. The results from the *m* complete datasets are combined for the inference.

If there are not enough non-missing observations for imputation when using 2 randomization factors as strata, only the screening REVEAL Lite 2.0 risk score will be used in the FCS regression if this does not cause any issues in the missing data imputation procedure. Otherwise, only the baseline PAH subtype will be used in the FCS regression.

For participants who die, the worst-rank score will be assigned for the non-existent change from baseline at Week 24. For participants with missing values at Week 24 on account of an event other than all-cause mortality, the next worst-rank score will be used to impute the missing change from baseline at Week 24 for continuous endpoints [Lachin, J. M. 1999].

The random number seed used for all imputations will be 7962006.

Sensitivity analyses are described in [Sec. 8.7.2].

Additionally, for the NT-proBNP endpoint, a log transform, y = log(x), will be applied to the data before using the standard MI method.

Categorical / Qualitative endpoints

The handling of missing data for categorical / qualitative endpoints is described in [Sec. 8.7] for each of the applicable endpoints.

6.8.2 Missing Dates for Prior and Concomitant Medications and Adverse Events

Incomplete Start Date

Missing day and month

- If the year is the same as the year of the first dosing date, then the day and month of the first dosing date will be assigned to the missing fields.

- If the year is prior to the year of first dosing date, then December 31 will be assigned to the missing fields.

- If the year is after the year of first dosing, then January 1 will be assigned to the missing fields.

Missing day only

- If the month and year are the same as the year and month of the first dosing date, then the first dosing date will be assigned to the missing day.

- If either the year of the partial date is before the year of the first dosing date or the years of the partial date and the first dosing date are the same, but the month of the partial date is before the month of the first dosing date, then the last day of the month will be assigned to the missing day.

- If either the year of the partial date is after the year of the first dosing date or the years of the partial date and the first dose date are the same, but the month of the partial date is after the month of the first dosing date, then the first day of the month will be assigned to the missing day.

- If the stop date is not missing and the imputed start date is after the stop date, the start date will be imputed by the stop date.

Missing day, month, and year

- No imputation needed. The corresponding AE will be included as a TEAE.

Incomplete Stop Date: If the imputed stop date is before the start date, then the imputed stop date will be equal to the start date.

Missing day and month

- If the year of the incomplete stop date is the same as the year of the last dosing date, then the day and month of the last dosing date will be assigned to the missing fields.

- If the year of the incomplete stop date is prior to the year of the last dosing then December 31 will be assigned to the missing fields.

- If the year of the incomplete stop date is after the year of the last dosing date, then January 1 will be assigned to the missing fields.

Missing day only

- If the month and year of the incomplete stop date are the same as the month and year of the last dosing date, then the day of the last dosing date will be assigned to the missing day.

If either the year of the partial date is not equal to the year of the last dosing date or the years of the partial date and the last dosing date are the same, but the month of the partial date is not equal to the month of the last dosing date, then the last day of the month will be assigned to the missing day.

6.8.3 Missing Dates for Disease Diagnosis Date

For disease diagnosis dates, the imputation rules are:

a. If day is missing, use 15th of the month

b. If both day and month are missing, impute as January 1st

c. If month is missing, impute as January

d. If year is missing, set to missing.

6.8.4 Handling of Data Limits

The following rules will be applied:

- Measurements reported as less than Lower Limits of Quantification (LLOQ) will be imputed to a value of LLOQ/2 for purposes of summarization.

- Measurements reported as less than some numerical value "X" will be imputed to "0.5 * X" for purposes of summarization.

- Percentage measurements reported as greater than 99% will be assigned a value of 99.5% during summarization.

Actual measurements will appear in the individual participant data listings, not the imputed measurements used for summarization.

7 STUDY POPULATION

7.1 Participant Disposition

Individual participant disposition data will be listed for all screened participants.

The number and percentage of participants receiving study treatment who completed the DBPC treatment periods along with the associated reasons for discontinuation from treatment and/or withdrawal from study will be presented.

The disposition summary table will also include a count of the number of participants who terminated the study as screen failures.

7.2 Screen Failures

A summary of the number of participants screened will be provided in the disposition table and the following information will be available in a separate table:

- Participants who failed the first screening
- Participants who were rescreened
- Participants who were rescreened and then randomized
- Participants who screen failed

  This includes the number of participants who failed a second rescreening as well as those who failed initial screening and not subsequently rescreened. In such cases, these participants terminate the study as screen failures.

7.3 Protocol Deviations

A listing of all protocol deviations by type of deviation will be provided.

7.4 Demographic, Baseline Characteristics, and Disease History

Demographic and baseline characteristics data, medical history, and disease history data will be listed for each participant.

Demographic and baseline characteristics will be summarized by descriptive statistics for all the randomized participants as randomized by treatment group.

7.5 Listing of Participant Inclusion and Exclusion Criteria

A listing of participants meeting all eligibility criteria for entry into the study will be provided. If a participant did not meet all eligibility criteria, then the individual inclusion and exclusion criteria that the participant did not meet will be listed.

7.6 Medical History and Medical Conditions Present at Entry

A listing of past medical history and medical conditions present at entry will be provided. A summary of medical history conditions by MedDRA preferred term for each treatment group will also be provided.

8 EFFICACY

8.1 General Considerations

All efficacy endpoints will be analyzed using the FAS unless otherwise specified. In the change from baseline analyses, participants who do not have baseline measurements will be excluded from the analyses.

Per protocol, randomization is stratified by REVEAL Lite 2.0 risk score (9 to 10 or ≥11) at screening and PAH subtypes (CTD-associated or not CTD-associated). Participants who are mis-stratified at the time of randomization will be analyzed using the “as intended”/“correct” stratum to which they were supposed to be randomized for all analyses, unless otherwise specified. Participants with screening REVEAL Lite 2.0 risk scores ≤ 9 will be grouped into the 9 to 10 category. If there are more than 10% participants mis-stratified, sensitivity analyses for all endpoints with a statistically significant outcome may be conducted using the stratum assigned at the time of randomization.

8.2 Testing Statistical Assumptions

The proportional hazards assumption of the Cox model will be examined using both graphical and analytical methods if warranted. The log[-log] of the survival function vs. time for first event will be plotted for the comparation between sotatercept and the placebo arm. If the curves are not parallel, indicating the hazards are not proportional, supportive analyses may be conducted to account for the possible non-proportional hazards effect, for example, using the Restricted Mean Survival Time (RMST) method [Anderson, K. M. 1991], and parametric method [Mehrotra, D. V., et al 2012].

The RMST is the population average of the amount of event-free survival time experienced during the study follow up time. This quantity can be estimated by the area under the KM curve up to the follow up time. The difference of two RMSTs for two treatment groups will be estimated and 95% confidence interval will be provided.

In addition, it is likely that a parametric model will fit data well and can be used as an alternative approach to comparing two group event rates over time. A Weibull model [Anderson KM. 1991] that allows the shape parameter to be a function of the covariates can also be used to examine the proportional hazards assumption where event rates change over time; these will be fit with the gamlss.cens R package (package and reference are available from CRAM the R library at http://cran.r-project.org).

One assumption for the stratified Cox proportional hazard model is that the treatment hazard ratio (HR) is constant across the strata. In case of a strong deviation from the assumption, which can result in a notably biased and/or less powerful analysis, a sensitivity analysis may be performed based on a two-step weighted Cox model approach by Mehrotra (2012) [Uno, H., et al 2014]. The first step is to estimate the treatment effect for each stratum and then the stratum specific estimates are combined to make overall inference using sample size weights.

8.3 Statement of the Null and Alternate Hypotheses

The null and alternate hypotheses for the primary and secondary efficacy endpoints are as follows:

- H_o: Sotatercept does not have a differential effect compared to placebo
- H_A: Sotatercept does have a differential effect compared to placebo

8.4 Subgroup Analyses

Subgroups are defined as follows:

- Age (<65 vs ≥65 years)
- Sex (male and female)
- PAH subtype (CTD-associated or not CTD-associated)
- Screening WHO functional class (III or IV)
- Screening REVEAL Lite 2.0 risk score (9 to 10 or ≥11): Participants with baseline REVEAL Lite 2.0 risk scores < 9 will be grouped into the 9 to 10 category.
- Double vs. Triple combination therapy at Screening
- Prostacyclin Infusion Therapy vs. Non-Prostacyclin Infusion at Screening
- Screening PVR (≤ 800 or >800 dynes*sec/cm5)
- eGFR at baseline (0-30; >30-60; > 60 ml/min/1.73m^2)

Subgroup analyses will be performed on the primary efficacy endpoint and any secondary efficacy endpoints that demonstrate statistical significance.

For the primary endpoint, the between-group treatment with a nominal 95% CI by treatment group will be estimated using Cox regression and plotted within each category of the subgroups described above.

For the secondary endpoints, the consistency of the treatment effect will be assessed using the unstratified analysis based on the corresponding analysis method for the endpoint as specified in [Sec. 8.7] for each category of the subgroup variables listed above. If the number of participants in subgroup category is less than 10% of FAS, the subgroup analysis will not be performed for this category of the subgroup variable, and this subgroup variable will not be displayed in the forest plot.

8.5 Multiple Comparisons and Multiplicity

A gatekeeping method will be used to control the Type I error rate in the primary and secondary efficacy endpoints by testing starting with the primary efficacy endpoint and then proceeding in the order of the secondary efficacy endpoints as listed in [Sec. 3.2.2].The p-value boundaries to be used at the IA and FA are described in Section 5.1 and Section 5.2, respectively.

8.6 Analysis of the Primary Efficacy Endpoint

8.6.1 Estimand

Following ICH E9(R1) , the estimand for the primary efficacy endpoint contains the following attributes:

Treatment: Sotatercept or placebo on top of background PAH therapy.

Population: Adults with PAH WHO FC III or IV.

Endpoint: Time to first event of all-cause death, lung transplantation, or PAH worsening-related hospitalization of ≥24 hours.

Intercurrent events: Changes in treatment (dose reduction, dose delay, discontinuation from sotatercept or placebo, or changes to background PAH therapy); a treatment policy strategy will be used. Thus, the endpoint is of interest regardless of changes in treatment.

Population-level summary: HR (sotatercept relative to placebo)

8.6.2 Primary Efficacy Analyses

The primary efficacy endpoint will be evaluated by comparing sotatercept to placebo with respect to time to first event. The stratified log-rank test with randomization factors as strata will be used to calculate the p-value. The treatment difference in survival will be assessed by the stratified log-rank test with randomization stratification factors as strata. The HR and its 95% confidence interval (CI) will be estimated using a stratified Cox proportional hazard model with Efron's method of tie handling. The non-parametric Kaplan-Meier method without stratification will be used to estimate the survival curve in each treatment group.

For the first event meeting the primary composite endpoint event definition, the onset date reported by the site in the CRF will be used for analyses. Specifically, for participants who die, the date of death will be used; for participants who report a PAH worsening-related hospitalization of ≥24 hours, the start date of the AE that leads to hospitalization will be used as the onset date of the event; for participants who had a lung transplantation, the date of the procedure will be used as the onset date of the event.

PAH-related hospitalization and lung transplant data are collected only during the study, whereas deaths are collected both during the study and post-study. In participants who do not have a primary event during the study, follow-up time for the primary analysis will be censored at the earlier of the data cutoff date and the study discontinuation (or study completion date) except for participants who died post-study and prior to the data cutoff date. For such participants, the death (and follow-up time up to the death) will be included in the primary analysis. Follow-up time (and events) in SOTERIA are ineligible for the primary analysis because all participants who enter SOTERIA had a primary event in ZENITH.

The number and proportion of participants with a primary endpoint event, as well as incidence rates per 100 patient-years of follow up, will be provided by treatment group.

To support the primary analysis, a summary of the individual component events contributing to the first event of the composite primary endpoint will be provided by treatment group and overall. In the event multiple first component events occur on the same day for a participant the unique combination of these first component events will be summarized as separate categories.

Table 6 Primary Analysis Strategy for the Primary Efficacy Endpoint

Statistical Method	Event Inclusion Criteria and Censoring Rules	Missing Data Approach
Testing: stratified log-rank test Estimation: stratified Cox regression	Include the first event of adjudication-confirmed PAH worsening-related hospitalization ≥24 hours, lung transplantation, or all-cause death prior to the cutoff date. All pre-cutoff deaths will be eligible, regardless of adjudication and regardless of whether they occurred during or post ZENITH. Follow-up time will be censored at the earlier of the data cutoff date and the date of study discontinuation (or study completion) except that post-study deaths (and follow-up time) before the data cutoff date will be included for participants whose first event was a post-study death occurring before the cutoff.	No imputation of missing data.

8.6.3 Sensitivity Analyses of the Primary Efficacy Results

The following sensitivity analyses are planned for the primary efficacy endpoints. [Table 7] provides an overview of each method and the missing data approach. Missing data are defined as follow-up time and potential events between the date of discontinuation and the data cutoff date among participants who discontinued prematurely without having had a primary endpoint.

Table 7 Sensitivity Analyses for the Primary Efficacy Endpoint

Sensitivity Analysis	Statistical Method	Event Inclusion Criteria and Censoring rules	Missing Data Approach
Sensitivity #1	Testing: stratified log-rank test Estimation: stratified Cox regression	Same as in [Table 6] except that deaths obtained after study discontinuation (that are not adjudicated) will be excluded, and post-study follow-up time will be excluded.	No imputation of missing data.
Sensitivity #2		Same as in [Table 6]	Impute missing follow-up time and events using the retrieved dropout method[1]
Sensitivity #3		Same as in [Table 6]	Impute missing follow-up time and events using the jump to reference method.
Sensitivity #4		Same as in [Table 6]	Impute missing follow-up time and events using the tipping point method.

[1] This method will be performed only if there are more than 5 such participants with retrieved dropout data.

A sensitivity analysis (Sensitivity #1) will be conducted restricted to events that are confirmed via adjudication. Post-study events (i.e., deaths) are the only events not subject to adjudication, so this analysis can equivalently be described as an analysis that excludes post-study events and post-study follow-up time.

A simulation approach will be used to assess the influence of missing data on the primary endpoint. The following is a detailed description of the planned analyses (Sensitivity #2-Sensitivity #4):

1. Counting the number of participants with missing follow-up data

Any participant who did not have a reported event prior to the data cutoff date and who discontinued the study prior to the cutoff date will be considered as having missing follow-up. The number of participants with missing follow up data will be summarized by treatment group.

2. Counting missing follow up time for the primary efficacy endpoint

Lost follow-up time will be calculated as the time from the last visit with full assessment of efficacy endpoints to the data cutoff date. The total amount of lost follow-up time will be summarized by treatment group and compared to the total potential follow-up time in a complete study where every patient is observed to have a primary outcome event, died while being actively followed or followed-up until the end of study.

3. Multiple imputation analysis of lost follow-up time

Missing follow-up data on the primary endpoint will be simulated based on the assumption that the time to first primary event will have a similar survival distribution pattern as the observed data. Parametric regression analysis will be used to estimate the survival curve using all available follow-up data. Parametric distributions such as Weibull distribution and exponential distribution will be explored to fit the distribution of the time to first primary efficacy event.

The following scenarios will be explored on the hazard rate of the primary endpoint in missing follow-up data:

1. Missing at random (MAR), Sensitivity #2: the missing follow-up data from both treatment groups have the same hazard rate as the observed data from patients from retrieved dropouts. Retrieved dropouts are participants who do not develop an event prior to treatment discontinuation and remain in study until the end of DBPC [He, J., et al 2023]. This method will only be performed only if there are more than 5 such participants with retrieved dropout data.

2. Missing not at random (MNAR), using the following approaches:

   i. Sensitivity #3: The missing follow-up data from both treatment groups have the same hazard rate as the observed data from placebo group (Jump to Reference approach);

   ii. Sensitivity #4: Tipping point analyses to assess the degree of robustness of statistical significance of the observed treatment effect, as follows:

      - The hazard rate in the placebo group missing data will be fixed at the placebo group's observed rate. The hazard rate in the sotatercept group missing data will be varied over a range of values and will be compared to the hazard rate in the placebo group with missing data to determine the point at which the log-rank test (based on imputed + observed data) is no longer significant.

The time-to-event analysis on the combined data (observed + imputed) using the same one-sided log-rank test stratified by stratification factors for primary efficacy analysis will be performed, and p-values and confidence intervals will be obtained using Rubin's approach for multiple imputation analysis.

8.7 Analysis of the Secondary Efficacy Endpoints

8.7.1 Estimand

The Treatment and Population attributes of all secondary endpoints estimands are as follows:

Treatment: sotatercept or placebo on top of background PAH therapy.

Population: Adults with PAH WHO FC III or IV.

The Endpoint, Intercurrent Events, and Population-level Summary attributes for each estimand are provided below.

8.7.1.1 Estimand for Secondary Time-to-Event Endpoints

Endpoints:

- Overall survival, defined as the time to date of death due to any cause
- Transplant-free survival, defined as the time to the first lung transplantation or death from any cause

Intercurrent events: Changes in treatment (dose reduction, dose delay, discontinuation from sotatercept or placebo, or changes to background PAH therapy); a treatment policy strategy will be used. Thus, the endpoint is of interest regardless of changes in treatment.

Population-level summary: Hazard ratio (sotatercept relative to placebo)

8.7.1.2 Estimand for Continuous Secondary Endpoints

Endpoints: Change from baseline at Week 24 in each of the following, with death prior to Week 24 represented quantitatively by any fixed worst-rank change from baseline to reflect the worst clinical outcome:

- REVEAL Lite 2.0 risk score
- NT-proBNP
- mPAP
- PVR
- 6MWD
- CO
- EQ-5D-5L index score

Intercurrent events:

- Changes in treatment: Same as for the estimand for the time-to-event endpoints.
- Death: A composite strategy will be implemented, in which the occurrence of death is incorporated into the definition of the endpoint.

Population-level summary: The midpoint of the distribution of the variable/endpoint noted above, compared between treatment conditions using a difference (sotatercept minus placebo) in midpoints; this between-treatment difference is referred to in statistical terms as the location-shift parameter.

8.7.1.3 Estimand for Binary Secondary Endpoints

Endpoints: Indicator (yes/no) of meeting each of the following:

- Achievement of a low or intermediate [≤ 7] REVEAL Lite 2.0 risk score at Week 24, where death prior to Week 24 is defined as not having met the criteria
- Mortality event
- Improvement in WHO FC, where death prior to the end of the DBPC Treatment Period is defined as not having met the criteria

Intercurrent events:

- Changes in treatment: Same as the estimand for the time-to-event endpoints
- Death (applicable to only the first and third binary endpoints): A composite strategy will be used, such that anyone who dies prior to Week 24 without having had the endpoint is considered to be a failure.

Population-level summary: The difference (sotatercept minus placebo) in proportions of patients achieving responses.

8.7.2 Secondary Efficacy Analyses

[Table 8] and [Table 9] present an overview of the analysis strategy for the secondary efficacy endpoints. Details of the analyses can be found below.

Table 8 Analysis Strategy for the Time-to-Event Secondary Efficacy Endpoints

Endpoint	Type	Statistical Method	Event Inclusion Criteria and Censoring Rules	Missing Data Approach
Overall survival	P	Testing: stratified log-rank test Estimation: stratified Cox regression	Include all deaths up to the data cutoff date, except for those occurring after lung transplantation or enrollment in SOTERIA. -Participants who have a pre-cutoff lung transplantation will be censored at the date of lung transplantation; -Participants who enroll in SOTERIA will be censored at date of ZENITH study completion; -Other participants who do not report a pre-cutoff death in ZENITH will be censored at the earlier of the data cutoff date and the last known alive date (study discontinuation date or last vital status contact date, whichever is later).	No imputation for missing data.
	S1		Include all deaths occurring prior to the data cutoff date (even those occurring post lung transplantation or in SOTERIA). -Participants without a reported pre-cutoff death will be censored at the earlier of the data cutoff date and the last known alive date.	No imputation for missing data.
	S2		Include all deaths occurring prior to the data cutoff date (even those occurring post lung transplantation or in SOTERIA). -Participants without a reported pre-cutoff death and being followed up as of the data cutoff date will be censored at the earlier of the data cutoff date and the last known alive date.	Impute the missing follow-up time and events using tipping point method.
	S3		Same as S2.	Impute the missing follow-up time and events using jump to reference method.

Endpoint	Type	Statistical Method	Event Inclusion Criteria and Censoring Rules	Missing Data Approach
Transplant-free survival	P	Testing: stratified log-rank test Estimation: stratified Cox regression	Include all lung transplantations reported up to the data cutoff. Include all pre-cutoff deaths reported in the study or after study discontinuation. Exclude deaths that happened after enrollment in SOTERIA. Participants who do not have pre-cutoff event will be censored at the earlier of the data cutoff and the last known event-free date (data study discontinuation or study completion). Post-study follow-up time will not be included.	No imputation for missing data.
	S1		Same as the primary approach (P) except that post-study events and follow-up time will be excluded.	No imputation for missing data.
	S2		Include all lung transplantations; Include all deaths reported in the study, after study discontinuation/completion, or after enrollment in SOTERIA. Participants who do not have a pre-cutoff event will be censored at the earlier of the data cutoff and the last known event-free date (data study discontinuation or study completion). Post-study follow-up time (whether in, or not in, SOTERIA, will be included.	No imputation for missing data.
	S3		Include all lung transplantations; Include all deaths reported in the study, after study discontinuation/completion, or after enrollment in SOTERIA. -Participants who do not have a pre-cutoff event and being in the study will be censored at the data cutoff date.	Impute the missing follow-up time and events using jump to reference method.
P = Primary; S = Sensitivity				

Table 9 Analysis Strategy for the Non-Time-to-Event Secondary Efficacy Endpoints

Endpoint	Type	Statistical Method	Missing Data Approach
Participants who experienced a mortality event at the end of the study	P	Stratified CMH	For participants that do not consider as a death as the overall survival primary analysis, -Impute as a non-death case if a participant is lost to follow-up prior to the date cutoff of IA or the end of DBPC
Change from baseline in REVEAL Lite 2.0 risk score at Week 24	P	ARSW	MI[1]
	S	ARSW	Pattern mixture control-based
Participants achieving a low or intermediate (≤7) REVEAL Lite 2.0 risk score at Week 24	P	Stratified CMH	Impute as a non-responder
Change from baseline in NT-proBNP levels at Week 24	P	ARSW	MI[1]
	S	ARSW	Pattern mixture control-based
Change from baseline in mPAP at Week 24	P	ARSW	MI[1]
	S	ARSW	Pattern mixture control-based
Change from baseline in PVR at Week 24	P	ARSW	MI[1]
	S	ARSW	Pattern mixture control-based
Participants who improve in WHO FC at the end of DBPC treatment period	P	Stratified CMH	Impute as a non-improver
Change from baseline in 6MWD at Week 24	P	ARSW	MI[1]
	S	ARSW	Pattern mixture control-based
Change from baseline in CO at Week 24	P	ARSW	MI[1]
	S	ARSW	Pattern mixture control-based

Endpoint	Type	Statistical Method	Missing Data Approach
Change from baseline in EQ-5D-5L index score at Week 24	P	ARSW	MI[1]

P = Primary; S = Sensitivity; CMH = Cochran-Mantel-Haenszel; ARSW = Aligned Rank Stratified Wilcoxon, MI = Multiple Imputation;

[1] Change from baseline at Week 24 for participants who died will be assigned a value that will receive the worst rank. Change from baseline at Week 24 for participants who have missing data due to a non-fatal clinical worsening event will be imputed to receive the next worst-rank.

Overall Survival

The overall survival (time to death) will be analyzed using the same approach as the primary endpoint. In addition, Kaplan-Meier curves will be generated for each arm with randomization factors as strata.

[Table 8] describes the primary and supportive analyses for the overall survival. Details of the multiple imputations are the same as the corresponding sensitivity analyses for primary endpoint, see [Sec. 8.6.3] for details.

Any participant with unknown vital status as of the data cutoff date will be considered as having missing follow-up.

Transplant-free survival

The analysis approach for the time to lung transplantation is analogous to that for the overall survival.

The details for the primary and supportive analyses for this endpoint are described in [Table 8]

Any participant who discontinued the study prior to the data cutoff date and did not have a reported lung transplantation or death prior to the data cutoff will be considered as having missing follow-up.

Participants who experienced a mortality event at the end of the study

This endpoint is defined identically to the overall survival endpoint.

The Cochran-Mantel-Haenszel test, stratified by the randomization factors, will be used to provide the p-value. The treatment difference will be provided with the 95% CI using the Miettinen and Nurminen method [Miettinen, O. and Nurminen, M. 1985] stratified by the randomization factors with Cochran-Mantel-Haenszel weights.

Change from baseline in REVEAL Lite 2.0 risk score at Week 24

Descriptive statistics will be provided for the baseline, Week 24, and the change from baseline at Week 24 for REVEAL Lite 2.0 risk score.

At the IA, the analysis will include only those participants who were randomized more than 24 weeks prior to the database cutoff. At the final analysis, all participants will be included in the analyses.

The change in REVEAL Lite 2.0 risk score at Week 24 from baseline will be analyzed using the aligned rank stratified Wilcoxon test [Hodges, J. L., Jr. and Lehmann, E. L. 1962] [Mehrotra, D. V., et al 2010] with the randomization stratification factors as strata (PAH subtype only since REVEAL Lite 2.0 risk score is the other stratification factor). In this test, the endpoint values are first aligned across the randomization strata using the stratum-level Hodges-Lehmann location shift estimates, and the aligned values are then analyzed using a Wilcoxon rank sum test. The output from this analysis will be used to provide a 2-sided p-value and corresponding Hodges-Lehmann location-shift estimate of the overall treatment difference with 95% CI. SAS implementation code for the aligned rank stratified Wilcoxon test is provided in [Appendix 12.3].

A sensitivity analysis using a control-based pattern mixture model will be performed if the result from the primary analysis approach is statistically significant. For the IA, data will not be imputed for ongoing participants who had not completed Week 24 at the time of the database cutoff as the non-existent change from baseline at Week 24 is missing completely at random. Non-existent REVEAL Lite 2.0 risk score at Week 24 due to death and missing data following a non-fatal primary endpoint event will be handled in the same way as the primary analyses described in [Sec. 6.8.1.2]. Missing data due to other reasons for both treatment arms will be imputed using placebo data only. Missing data will be filled in m (m=100) times using FCS regression accounting for the baseline measurement within each stratum. The complete datasets will be analyzed using the aligned rank test stratified Wilcoxon test with randomization factors as strata. The results from the *m* complete datasets will be combined for the inference.

Participants achieved a low or intermediate (≤ 7) REVEAL Lite 2.0 risk score at Week 24

The analysis approach for this endpoint is analogous to that for participants experienced a mortality event at the end of DBPC Treatment Period. PAH subtype is the only stratification factor to be used since REVEAL Lite 2.0 risk score is the other stratification factor). At the IA, the analysis will include only those participants who were randomized more than 24 weeks prior to the database cutoff. At the final analysis, all participants will be included in the analyses.

Achieving a low or intermediate risk score at Week 24 means that the criterion outline above are satisfied at Week 24, regardless of the score calculated at baseline. Participants who do not have the risk score at Week 24 will be considered as non-responder. Participants who had a REVEAL Lite 2.0 risk score ≤ 7 at baseline will not be included in the analyses.

Change from baseline for the following endpoints at Week 24

- NT-proBNP
- mPAP
- PVR
- 6MWD
- CO

The analysis approach for these endpoints is analogous to that for change from baseline in REVEAL Lite 2.0 risk score at Week 24. For 6MWD, if a participant discontinues the 6MWT prematurely, the total distance walked at the time of discontinuation will be the 6MWD used in the analysis.

Participants improved in WHO FC at the end of DBPC Treatment Period

The analysis approach for this endpoint is analogous to that for participants who experienced a mortality event at the end of DBPC Treatment Period.

Participants who die prior to the date for the data cutoff will be considered as non-improvers. For participants who complete or discontinue the study prior to the cutoff date, the last measurement that was taken prior to the end of the study will be used. For the participants that are in the study at the time of DBL, the last measurement prior to cutoff date will be used.

Change from baseline in EQ-5D-5L index score

The EQ-5D-5L consists of 5 dimensions (Mobility, Self-Care, Usual Activities, Pain/Discomfort, and Anxiety/Depression). Each dimension has 5 response levels (no problems, slight problems, moderate problems, severe problems, and unable to/extreme problems). Each response level is coded from 1 to 5 with ‘1’ assigned to “No problems”, ‘2’ assigned to “Slight problems”, and so on up to ‘5’ assigned to “Unable to/extreme problems”.

Individual responses from subjects at each time the questionnaire is completed are coded as single-digit numbers expressing the severity level selected in each dimension. For example, “21111” means slight problems in the mobility dimensions and no problems in any of the other dimensions. This 5-digit code is often referred to as a health state. Such 5-digit codes should not be added to obtain any kind of overall score.

The summary index score is derived from an appropriate “value set”. Value sets represent the average measurements of a sample of people. Usually this is for the general public of a particular country/region.

CCI

[redacted]

9 SAFETY ANALYSIS

[Table 10] below summarizes the analysis strategy for safety endpoints.

Table 10 Analysis Strategy for Safety Parameters

Analysis Part	Safety Endpoint	Descriptive Statistics	95% Between-group CI[a]
Overall Safety Assessment	Any TEAE	X	X
	Any serious TEAE	X	X
	Any suspected drug related TEAE	X	X
	Any serious and drug related TEAE	X	X
	Discontinuation due to TEAE	X	X
	Death	X	X
	SOC (incidence ≥4 participants in any treatment group)	X	X
	AE PT (incidence ≥4 participants in any treatment group)	X	X
	SOC, AE PT (incidence <4 in both treatment groups)	X	
	Change from baseline results (laboratory tests, vital signs, ECG)	X	X[b]
Assessment of safety topics of special interest	telangiectasia	X	X
	AEOI	X	X

AE=adverse event; AEOI=adverse events of interest; CI=confidence interval, SOC=system organ class, PT=preferred term, ECG=electrocardiogram, X=results to be provided

[a] 95% between-treatment group CI will be provided using the Miettinen and Nurminen method [Miettinen, O. and Nurminen, M. 1985].

[b] Only to be provided for selected vital sign summaries described in [Sec. 9.4].

The safety endpoints will be summarized using the Safety Set. The safety endpoints include treatment emergent adverse events, laboratory tests, vital signs, immunogenicity, and ECGs. In the change from baseline analyses, both measurements at baseline and the relevant post-baseline timepoint will be required for a participant to be included in the analyses.

For participants who have either discontinued or completed study intervention, all measurements within 56 days (8 weeks) following the last dose of study intervention will be considered to be on-treatment measurements. For participants who are on study intervention (i.e., have not discontinued or completed study intervention), all measurements will be considered to be on-treatment measurements.

9.1 Adverse Events

All adverse events (AEs) and SAEs reported from the signing of the informed consent form to the end-of-study visit will be reported on the AE CRF and present in the study database. All AEs that started or worsened from the time of first dosing of study medication to 8 weeks after the last dose of study medication will be considered as treatment emergent adverse events (TEAE).

Any partial dates will be imputed based on the rules described in [Sec. 6.8.2].

A drug related TEAE is defined as any TEAE that is "suspected" to be related to study treatment as reported on the CRF or with missing assessment of the relationship to study treatment.

The following summaries will be presented for each treatment group:

- Overall summary of TEAEs
- Number and percentage of participants reporting each TEAE, categorized by System Organ Class (SOC) and Preferred Term (PT)
- Number and percentage of participants reporting each TEAE with incidence ≥4 by SOC and PT
- Number and percentage of participants with TEAE that were suspected to be related to study drug by SOC and PT.
- Number and percentage of participants reporting SAE, categorized by SOC and PT
- Number and percentage of participants reporting SAE that are suspected to be related to study drug by SOC and PT
- Number and percentage of participants reporting TEAE leading to death by SOC and PT
- Number and percentage of participants reporting severe TEAE, categorized by SOC and PT
- Number and percentage of participants reporting TEAE leading to study drug withdrawal, categorized by SOC and PT
- Number and percentage of participants with treatment emergent adverse events of special interest (AESI) and adverse events of interest (AEOI)
- Number and percentage of participants with AESI that are suspected to be related to study drug by SOC and PT
- Number and percentage of participants with TEAE by SOC and PT indicating severity of the TEAE
- Number and percentage of TEAE/AEOI/AESI categorized by 1) background therapy at baseline (Double vs. Triple combination), 2) prostacyclin therapy at

baseline (Prostacyclin Infusion Therapy vs. Non-Prostacyclin Infusion), and 3) Age (<65 years of old vs ≥65 years of old).

Counting will be by numbers of participants, not events, and participants will be counted once within each applicable SOC or PT. If a participant experiences the same AE at more than one severity or with more than one relationship to study drug, the severity rating or relationship that is more severe or stronger to study drug will be given precedence in summaries that consider severity or drug relationship. Any missing severity, causality, or outcome will not be imputed and classed as unknown.

In addition to the summaries described above, point estimates and 95% CIs for the differences between treatment groups in the percentages of participants will be provided for selected AE summaries in accordance with what is outlined in [Table 10] above that occur in at least 4 participants in any treatment group. This threshold was chosen because the 95% CI for the between-group difference in percent incidence will always include zero when fewer participants per group have events and thus would add little to the interpretation of potentially meaningful differences.

Confidence intervals for between-treatment group differences will be provided using the unstratified Miettinen and Nurminen method [Miettinen, O. and Nurminen, M. 1985].

No adjustments for multiplicity are planned.

Adverse events of interest (AEOI) and the AESI (Telangiectasia) are presented in [Table 11] with the corresponding search strategies.

Table 11 Search Criteria for AESI / AEOI

AESI / AEOI	SMQ(s) [MedDRA 27.0]	Scope (if applicable)	Description
Increased hemoglobin (increased hematocrit, increased RBC count)	N/A	N/A	Preferred terms: • Haemoglobin increased • RBC count increased • Full blood count increased • Haematocrit increased • Polycythaemia • Stress polycythaemia
Thrombocytopenia	Haematopoietic thrombocytopenia	Narrow	N/A
Immunogenicity	• Anaphylactic reaction • Hypersensitivity	Narrow	SMQ and Preferred term: • Preferred terms: Drug specific antibody • Preferred terms: Drug specific antibody present • SMQ Anaphylactic reaction • SMQ Hypersensitivity
Increased blood pressure / hypertension	Hypertension	Broad+Narrow	N/A
Thrombo-embolic events	Embolic and thrombotic events	Narrow	N/A
Bleeding events	Haemorrhages		SMQ and Preferred term: • SMQ Haemorrhages (excluding laboratory terms) • Preferred term: Anemia
Renal toxicity	• Acute renal failure • Proteinuria • Chronic kidney disease	Narrow	N/A
Telangiectasia	N/A	N/A	Preferred Terms: • Telangiectasia • Spider vein • Spider naevi • Nasal mucosal telangiectasia
Hepatic toxicity	Hepatic disorders	Narrow	N/A
Cardiac events	Ischaemic heart disease	Narrow	SMQ and HLGT: • SMQ Ischaemic heart disease • HLGT Heart failures • HLGT Pericardial disorders

All AEs will be listed.

9.2 Anti-drug Antibodies

Individual anti-drug antibody (ADA) data will be listed.

The frequency and percentage of ADA responses will be summarized by treatment group and scheduled time.

The frequency and percentage of all patients testing positive for ADA (anti-sotatercept) at any point during the study (i.e., ADA prevalence) will be summarized by treatment group. In addition, for sotatercept ADA, a summary of the prevalence of sotatercept ADA and titer summary (median, minimum, and maximum value) will be provided by scheduled visit and antibody follow-up visit (as applicable).

The frequency and percentage of patients with neutralizing antibodies (NAb) will also be summarized by treatment group.

In addition, the following tables will be provided by ADA status

- Primary efficacy endpoint
- Serum concentration
- Adverse event summary, including a summary of the safety events selected based on the [Table 12]
- Number and percentage of participants with AESI/AEOI

Table 12 SMQ Information for the Selected Adverse Events

Selected Safety Events	SMQ(s) [MedDRA 27.0]	Scope (if applicable)	Description
Hypersensitivity-like reactions	• Anaphylactic reaction • Hypersensitivity	Narrow	
Administration site reactions (related to sotatercept)	N/A	N/A	HLGT: Administration site reactions

Additional analyses may be performed as appropriate.

9.3 Laboratory Evaluations

The following laboratory parameters will be analyzed over time, for urinalysis, only a listing will be provided:

Hematology

Hematology data consists of complete blood counts of red blood cells, absolute white blood cells, hemoglobin, hematocrit, and platelet counts. Such data will be collected and analyzed locally at the investigative sites.

Serum Chemistry

Serum chemistry data consists of blood urea, creatinine, total bilirubin, direct bilirubin, AST, ALT, alkaline phosphatase, sodium, potassium, chloride, calcium, phosphorous, glucose, magnesium, CO_2, and FSH. Such data will be analyzed at a central laboratory.

Actual measurements and changes in laboratory measurements from baseline will be summarized by timepoint.

Shift tables based on CTCAE criterion [Table 13] by the worst post-baseline grade value for the following parameters will be presented by treatment group:

Hematology: Platelets, and Hemoglobin

Serum Chemistry: ALT, AST, creatinine, alkaline phosphatase, total bilirubin, and eGFR

Table 13 CTCAE Version 4.03 Severity Grade Classifications for Selected Laboratory Parameters

Category	Parameter	Grade 1	Grade 2	Grade 3	Grade 4
Hematology	Hemoglobin (Hgb) [Anemia]	10.0 g/dL ≤ Hgb < LLN	8.0 ≤ Hgb < 10.0 g/dL	Hgb < 8.0 g/dL	*Life-threatening consequences; urgent intervention indicated* [a]
	Hemoglobin increased	0 < Increase ≤ 2 g/dL above ULN or above baseline if baseline is above ULN	2 < Increase ≤ 4 g/dL above ULN or above baseline if baseline is above ULN	Increase > 4 g/dL above ULN or above baseline if baseline is above ULN	N/A
	Platelet counts (PC)	75.0 x 10e9 /L ≤ PC < LLN	50.0 x 10e9 ≤ PC < 75.0 x 10e9 /L	25.0 x 10e9 ≤ PC < 50.0 x 10e9 /L	PC < 25.0 x 10e9 /L
Serum Chemistry	Creatinine	>ULN and ≤1.5x ULN	>1.5xULN and ≤3x ULN	>3xULN and ≤6x ULN	>6x ULN
	Total bilirubin	>ULN and ≤1.5x ULN	>1.5x ULN and ≤3x ULN	>3x ULN and ≤10x ULN	>10x ULN
	Direct bilirubin	>ULN and ≤1.5x ULN	>1.5x ULN and ≤3x ULN	>3x ULN and ≤10x ULN	>10x ULN
	AST	>ULN and ≤3x ULN	>3x ULN and ≤5x ULN	>5x ULN and ≤20x ULN	>20x ULN
	ALT	>ULN and ≤3x ULN	>3x ULN and ≤5x ULN	>5x ULN and ≤20x ULN	>20x ULN
	Alkaline Phosphatase Increased	>ULN - 2.5 x ULN	>2.5 - 5.0 x ULN	>5.0 - 20.0 x ULN	>20.0 x ULN
	eGFR	< LLN – 60 ml/min/1.73m^2	30-59 ml/min/1.73m^2	15-29 ml/min/1.73m^2	<15 ml/min/1.73m^2

[a] No grade 4 will be assigned given the assignment of the severity grade is based on the numeric values of hemoglobin.
Grade 0 will be assigned if the values are outside the defined ranges above.

A boxplot of the raw data and change from baseline will be provided for the following hematology parameters: hemoglobin, leukocytes, neutrophils, platelets, and hemoglobin (by gender) and chemistry parameters: ALT, alkaline phosphatase, AST, total bilirubin, calcium, chloride, creatinine, direct bilirubin, glucose, phosphorus, potassium, sodium, urea nitrogen.

A plot of the peak total bilirubin versus peak ALT/AST will be provided.

All laboratory measurements will be listed for all participants.

9.4 Vital Signs

Vital sign parameters include temperature, pulse rate, respiratory rate, and blood pressure. For each parameter at each timepoint, the change from baseline will be summarized. Vital signs will also be listed for all participants including height at screening and weight at all dosing visits.

The number and percentage of participants with the following changes in systolic blood pressure (SBP) and diastolic blood pressure (DBP) will be summarized by visit:

- Change from baseline SBP >20 mmHg and SBP ≥140 mmHg
- Change from baseline SBP >40 mmHg and SBP ≥140 mmHg
- Change from baseline DBP >10 mmHg and DBP ≥90 mmHg
- Change from baseline DBP >20 mmHg and DBP ≥90 mmHg

BMI will be calculated using the formula: BMI (kg/m^2) = weight (kg) / [height (m)]^2.

A boxplot of raw data as well as a boxplot of change from baseline will be provided for each vital sign parameter and treatment group.

Shift tables using vital sign parameters according to version 5.0 of the CTCAE criteria will be provided by treatment group as outlined in [Table 14].

Table 14 CTCAE Version 4.03 Severity Grade Classification for Vital Signs

Classification	Grade 1	Grade 2	Grade 3
Hypertension	Systolic BP 120 - 139 mm Hg or diastolic BP 80 - 89 mm Hg	Systolic BP 140 - 159 mm Hg or diastolic BP 90 - 99 mm Hg	Systolic BP >=160 mm Hg or diastolic BP >=100 mm Hg

9.5 Electrocardiogram (ECG)

ECG parameters include heart rate (HR), QRS, QT, and QTcF and will consist of a single 12-lead ECG that will be centrally read.

ECG parameters will be summarized at each timepoint.

ECG interpretation (normal, abnormal) will be presented for actual measurements and changes from baseline to each post baseline visit [expressed as Improvement, No Change, and Deterioration].

Note that:

- Improvement = Abnormal to Normal
- Deterioration = Normal to Abnormal
- No Change = Abnormal to Abnormal or Normal to Normal

If either result is missing or unknown for any patient, then the corresponding 'Missing'/'Unknown' category will also be presented.

ECG results will be listed for all participants.

10 STUDY MEDICATION

10.1 Compliance (Medication Adherence)

Percent compliance will be calculated according to the following formula and summarized using descriptive statistics.

Compliance (%) is calculated by the following:

$$100* \frac{\text{Number of visits where study medication was administered}}{\text{Number of visits in the Treatment Period where study medication should have been administered}}$$

The study medication is administered every 3 weeks. For a participant who is followed for the entire study treatment period, the "Number of Visits in the Treatment Period Where Study Medication Should Have Been Administered" is the total number of visits that should be done from randomization to the last scheduled day for treatment administration for that participant, excluding the number of dose delays/holds per protocol. For a participant who discontinues from the study treatment, the "Number of Visits in the Treatment Period Where Study Medication Should Have Been Administered" is the total number of visits that should be done from randomization to the date of the last visit, excluding the number of dose delays/holds per protocol.

10.2 Extent of Exposure

The duration of exposure in days and number of treatment visits will be summarized by treatment group with descriptive statistics. The total dose administered in mg will also be summarized by treatment group with descriptive statistics and is calculated as a function of the individual participant's weight in kg.

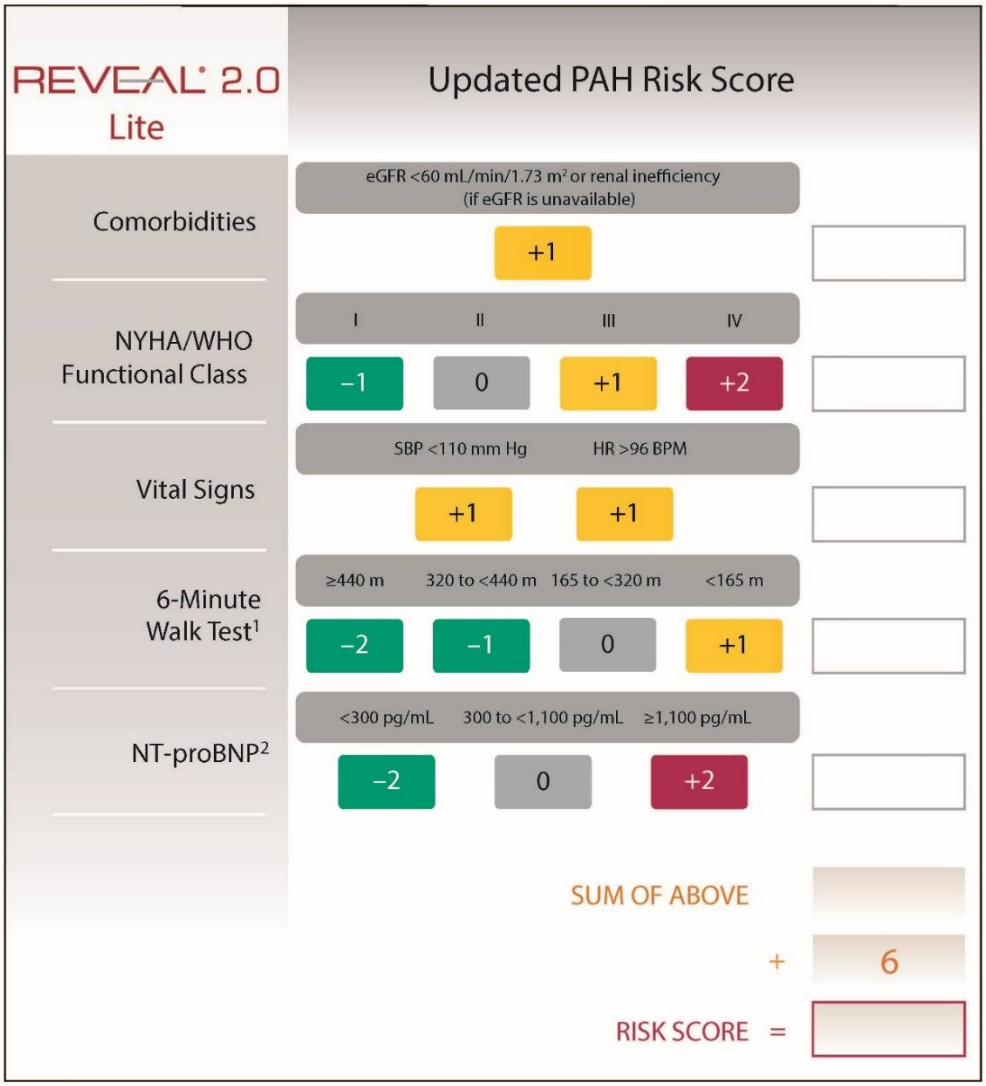
11 REFERENCES

[Anderson, K. M. 1991]	Anderson KM. A nonproportional hazards weibull accelerated failure time regression model. Biometrics 1991;47:281-8.	[00TH4Z]
[He, J., et al 2023]	He J, Crackel R, Koh W, Chen LW, Li F, Zhang J, et al. Retrieved-dropout-based multiple imputation for time-to-event data in cardiovascular outcome trials. J Biopharm Stat. 2023;33(2):234-52.	[089NSQ]
[Hodges, J. L., Jr. and Lehmann, E. L. 1962]	Hodges JL, Jr., Lehmann EL. Rank methods for combination of independent experiments in analysis of variance. The Annals of Mathematical Statistics 1962;33(2):482-97.	[03R277]
[Hoeper, M. M., et al 2023]	Hoeper MM, Badesch DB, Ghofrani HA, Gibbs JSR, Gomberg-Maitland M, McLaughlin VV, et al. Phase 3 trial of sotatercept for treatment of pulmonary arterial hypertension. N Engl J Med. In press 2023.	[088BQF]
[Humbert, M., et al 2022]	Humbert M, Kovacs G, Hoeper MM, Badagliacca R, Berger RMF, Brida M, et al. 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. Eur Heart J. 2022;43:3618-731.	[085V4Z]
[Lachin, J. M. 1999]	Lachin JM. Worst-rank score analysis with informatively missing observations in clinical trials. Control Clin Trials 1999;20:408-22.	[03PYVK]
[Mehrotra, D. V., et al 2010]	Mehrotra DV, Lu X, Li X. Rank-based analyses of stratified experiments: alternatives to the van Elteren test. Am Stat. 2010 May;64(2):121-30.	[055P90]
[Mehrotra, D. V., et al 2012]	Mehrotra DV, Su SC, Li X. An efficient alternative to the stratified Cox model analysis. Stat Med. 2012 Jul 30;31(17):1849-56.	[0422HQ]
[Miettinen, O. and Nurminen, M. 1985]	Miettinen O and Nurminen M. Comparative analysis of two rates. Stat Med 1985;4:213-26.	[00VMQY]

[Mogg, R. and Mehrotra, D. V. 2007]	Mogg R, Mehrotra DV. Analysis of antiretroviral immunotherapy trials with potentially non-normal and incomplete longitudinal data. Stat Med 2007;26:484-97.	[03RBNV]
[Mosteller, R. D. 1987]	Mosteller RD. Simplified calculation of body-surface area [Letter to the Editor] [Abstract]. N Engl J Med 1987;1098.	[03QGHG]
[Uno, H., et al 2014]	Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata T, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. J Clin Oncol. 2014 Aug 1;32(22):2380-5.	[045X5X]
[van Buuren, S. 2007]	van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res. 2007;16:219-42.	[0826M7]
[van Buuren, S., et al 2006]	van Buuren S, Brand JPL, Groothuis-Oudshoorn CGM, Rubin DB. Fully conditional specification in multivariate imputation. J Stat Comput Simul. 2006 Dec;76(12):1049-64.	[0826M5]

12 APPENDIX

12.1 REVEAL Lite 2.0 PAH Risk Score Calculator

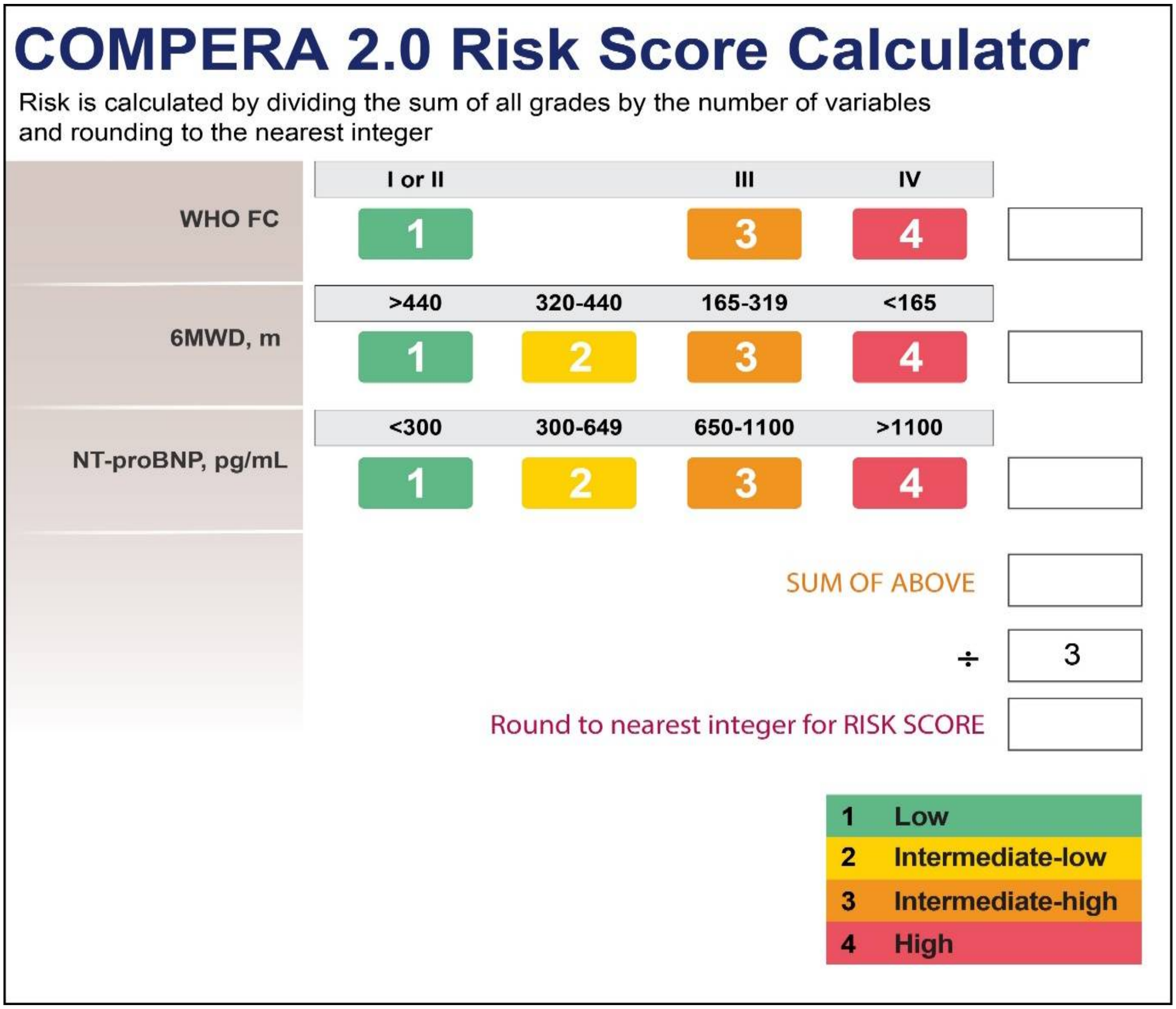


BPM = beats per minute; eGFR = estimated glomerular filtration rate; HR = heart rate; NT-proBNP = N-terminal prohormone B-type natriuretic peptide; NYHA = New York Heart Association; PAH = pulmonary arterial hypertension; REVEAL = Registry to Evaluate Early and Long-Term PAH Disease Management; SBP = systolic blood pressure; WHO = World Health Organization.

[1]The average of the two Screening 6MWDs should be used for score calculation;

[2]Central laboratory NT-proBNP result from Screening Visit should be used for score calculation.

12.2 COMPERA 2.0 Risk Score Calculator



6MWD = 6-minute walk distance; NT-proBNP = N-terminal prohormone B-type natriuretic peptide; WHO = World Health Organization.

12.3 SAS Code for the Aligned Rank Stratified Wilcoxon Test

```
PROC NPAR1WAY DATA=mimpdata WILCOXON ALIGN=STRATA(HL) HL
CORRECT=NO;
    CLASS trt;
    VAR resp;
    STRATA strat;
    BY impnumber;
    ODS OUTPUT HodgesLehmann=hlstats_a;
RUN;

PROC MIANALYZE DATA=hlstats_a;
    MODELEFFECTS shift;
    STDERR stderr;
     ODS OUTPUT ParameterEstimates=results;
RUN;
```

12.4 Approval Information

The SAP amendment 03 of Protocol MK7962-006-07 was approved by the BARDS TA head.

Name: PPD Date:16-Oct-2024