Actelion Pharmaceuticals Ltd*
(a Janssen Pharmaceutical Company of Johnson & Johnson)

Statistical Analysis Plan

A prospective, randomized, double-blind, multicenter, placebo-controlled, parallel group, adaptive Phase 3 study with open-label extension to evaluate efficacy and safety of macitentan 75 mg in inoperable or persistent/recurrent chronic thromboembolic pulmonary hypertension

Macitentan in inoperAble or persistent/reCurrent chronIc ThromboEmbolic Pulmonary Hypertension (MACiTEPH)

Protocol 67896062CTP3001; Phase [3]

JNJ-67896062/ACT-064992 (Macitentan)

*Actelion Pharmaceuticals Ltd ("Actelion") is a global organization that operates through different legal entities in various countries. Therefore, the legal entity acting as the sponsor for Actelion studies may vary, such as, but not limited to Janssen-Cilag International NV; Janssen, Inc; Janssen Pharmaceutica NV; or Janssen Research & Development, LLC. The term "sponsor" is used throughout this document to represent these various legal entities; the sponsor is identified on the Contact Information page that accompanies the protocol

Status:	Approved
Date:	18 March 2024
Prepared by:	Actelion Pharmaceuticals Ltd, Janssen Research & Development, a division of Janssen Pharmaceutical NV
Document No.:	EDMS-RIM-368296, 3.0

Compliance: The study described in this report was performed according to the principles of Good Clinical Practice (GCP).

TABLE OF CONTENTS

TABLE OF CONTENTS **2**

LIST OF TABLES **5**

LIST OF FIGURES **5**

VERSION HISTORY **6**

1. INTRODUCTION **7**
1.1. Objectives and Endpoints 7
1.1.1. Primary Objective and Endpoint 8
1.2. Main Study Design 8
1.2.1. Study Periods 9
1.2.2. Timing of Interim and Final Analyses 13
1.3. Hemodynamic Sub-Study Design 13

2. STATISTICAL HYPOTHESES **14**
2.1. Primary Hypothesis 14
2.2. Secondary Hypotheses 14
2.3. Overall Testing Strategy 15

3. SAMPLE SIZE DETERMINATION **15**

4. POPULATIONS (ANALYSIS SETS) FOR ANALYSIS **15**

5. STATISTICAL ANALYSES **17**
5.1. General Considerations 17
5.1.1. Study Day and Relative Day 17
5.1.2. Baseline 17
5.1.2.1. DB Period 17
5.1.2.2. OL Extension Period 17
5.1.2.3. Hemodynamic Sub-Study 17
5.1.3. Double-blind Treatment Phases 18
5.1.4. Study Completion 19
5.1.5. End of study 19
5.1.5.1. Hemodynamic Sub-Study 20
5.1.6. Visit Windows 21
5.1.7. Pooling of Countries 23
5.2. Participant Dispositions 23
5.2.1. Disposition Reporting 23
5.2.1.1. DB Period 23
5.2.1.2. OL Extension Period 26
5.3. Primary Endpoint Analysis 27
5.3.1. Definition of Endpoint 28
5.3.1.1. The 6-Minute Walk Test 28
5.3.2. Estimand 28
5.3.2.1. Primary Estimand 28
5.3.3. Analysis Methods 30
5.3.3.1. Statistical Model 30
5.3.3.2. Handling of Intercurrent Events and Missing Data 31
5.3.4. Primary, Supplementary and Subgroup Analyses 32
5.3.4.1. Primary Endpoint Analysis (Main Analysis) 32
5.3.4.2. Supplementary Analysis 32
5.3.4.3. Subgroup Analyses 33
5.4. Secondary Endpoint(s) Analysis 34
5.4.1. Time to First Clinical Worsening up to EODBT 34

5.4.1.1. Definition of Endpoint 34
5.4.1.2. Estimand 35
5.4.1.3. Analysis Methods 36
5.4.2. Improvement in WHO FC From Baseline to Week 28 (Yes/No) 37
5.4.2.1. Estimand 37
5.4.2.2. Analysis Methods 38
5.4.3. Percent of Baseline NT-proBNP Over time up to EODBT 38
5.4.4. Hemodynamic Endpoints 39
5.4.4.1. PVR Ratio 39
5.4.4.2. Other Hemodynamic Endpoints 40
5.5. Safety Analyses 41
5.5.1. Extent of Exposure 42
5.5.1.1. Extent of Exposure during the DB period 42
5.5.1.2. Extent of Exposure during the OL period 43
5.5.2. Adverse Events during DB period 43
5.5.2.1. Adverse Events of Special Interest (AESI) 44
5.5.2.2. Potential Risks 46
5.5.2.3. Deaths 46
5.5.3. Adverse Events during OL period 47
5.5.4. Additional Safety Assessments 47
5.5.4.1. Clinical Laboratory Tests 47
5.5.4.2. Vital Signs and Physical Examination Findings 50
5.5.4.3. Electrocardiogram 51
5.6. Other Analyses 52
5.6.1. Pharmacokinetics 52
5.6.2. Pharmacodynamics 53
5.6.3. Definition of Subgroups 53
5.7. Interim Analyses 54
5.7.1. Futility and Sample Size Re-estimation at the Interim Analysis 54
5.7.1.1. Stage 1 54
5.7.1.2. Stage 2 55
5.7.1.3. Futility Criteria 55
5.7.1.4. Sample Size Re-estimation (SSRE) Criteria 55
5.7.2. Combination Testing for Control of Type I Error Inflation 56
5.7.2.1. The Median-Unbiased Point and Interval Estimates for the Primary and Secondary Endpoints 57
5.7.3. Determination of the Operational Characteristics via Simulation 58
5.7.4. Data Monitoring Committee (DMC) or Other Review Board 59
5.7.4.1. Independent Data Monitoring Committee 59
5.7.4.2. Other Review Boards 59

6. SUPPORTING DOCUMENTATION 60
6.1. Appendix 1: List of Abbreviations 60
6.2. Appendix 2: Changes to Protocol-Planned Analyses 62
6.3. Appendix 3: Demographics and Baseline Characteristics 63
6.4. Appendix 4: Protocol Deviations 66
6.4.1. Per-Protocol Set 66
6.4.2. Hemodynamic Sub-Study 67
6.5. Appendix 5: Prior and Concomitant Medications 68
6.6. Appendix 6: Procedures 69
6.7. Appendix 7: Medical History 70
6.8. Appendix 8: Intervention Compliance 71
6.9. Appendix 9: Adverse Events of Special Interest 72
6.9.1. Potential Risks 72
6.10. Appendix 10: Medications of Special Interest 73
6.11. Appendix 11: Laboratory Toxicity Grading 75
6.12. Appendix 12: Schedule of Activities 81
6.13. Appendix 13: COVID-19 and Regional Crisis 86

6.14. Appendix 14: Modelling and Simulation Report 89
6.15. Appendix 15: Hemodynamic Variables 109

7. REFERENCES 110

LIST OF TABLES

Table 1: SAP Version History Summary 6
Table 2: Study Documents 7
Table 3: Overview of Analysis Sets 15
Table 4: Analysis Sets and Key Usage 16
Table 5: DB Treatment Phases 18
Table 6: EOS Visit Definition 20
Table 7: Visit Windows: 6MWD, and NT-proBNP 22
Table 8: Visit Windows for WHO FC, Vitals Signs, Body Weight, Central Laboratory 22
Table 9: Visit Windows for 12 Lead ECG 23
Table 10: Visit Windows for Hemodynamic Assessments 23
Table 11: Intercurrent Events and Related Strategy for Primary Estimand 29
Table 12: Primary Analysis: Handling of IEs and Missing Data During DB Period up to Week 28 31
Table 13: Summary of Planned Analyses Supporting Primary Endpoint 32
Table 14: Intercurrent Events and Related Strategy for Treatment Policy Estimand 33
Table 15: Markedly Abnormal Vital Signs 51
Table 16: Criteria for Abnormal QTc Values and Changes from Baseline for QTcB and QTcF 51
Table 17: Definition of Subgroups 53
Table 18: Summary of Significant Changes from Protocol 62
Table 19: Demographic Variables 63
Table 20: Baseline Characteristics Variables 63
Table 21: PH-Specific Medications 73
Table 22: Common Terminology Criteria for Adverse Events (CTCAE) v5.0 76
Table 23: Schedule of Activities 81
Table 24: Visit and Assessment Schedule during the Open-Label Extension Period 83
Table 25: Visit and Assessment Schedule for Participants Entering the Post-treatment Observation Period (PTOP) 85
Table 26: Hemodynamic Variables Collected/Derived in the eCRF 109

LIST OF FIGURES

Figure 1: Schematic Overview of the Study – Participants Who Complete the DB Treatment Period as per Protocol version 6 Amendment 4 11
Figure 2: Participants Who Prematurely Discontinue DB Study Intervention (Example of a Premature Discontinuation at Week 9) as per Protocol version 6 Amendment 4) 11
Figure 3: Schematic Overview of the Study – Participants Who Complete the DB Treatment Period as per Protocol versions prior to Protocol version 6 Amendment 4 12
Figure 4: Participants Who Prematurely Discontinue DB Study Intervention (Example of a Premature Discontinuation at Week 9) as per Protocol versions prior to Protocol version 6 Amendment 4 12

VERSION HISTORY

Table 1: SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1	15 Mar 2021	N/A	Initial Release
2	30 Mar 2023	Updates and clarifications to Disposition, Demographics and Baseline Characteristics, exposure and compliance, prior and concomitant medications, safety endpoints (AEs, laboratory endpoints, ECG and vital signs), visit windowing. Added summary of procedures Updated subgroups, Clarified Intercurrent events PAH-SYMPACT Null and alternative hypothesis Added details for Per Protocol Set Clarified participants included in stage 1 interim analysis. Updated formulae for conditional power and sample size re-estimation. Minor clarification to primary estimand intercurrent event handling and to supplementary and sensitivity analyses for primary endpoint.	Clarifications following DPS review (Dispositions and Safety). New countries added. New definitions of AE of potential risk added Subgroup of interest added Clarification for interim analysis and sample size re-estimation with fixed weights (w) Clarification
3	18 Mar 2024	Abbreviated CSR: reduction/change in scope	Early termination of the study

1. INTRODUCTION

This statistical analysis plan (SAP) describes the planned analyses for the abbreviated clinical study report (CSR) of study 67896062CTP3001. Plans for the reporting of the double-blind (DB), and open-label (OL) periods will be presented. Details for the reporting of the hemodynamic sub-study will also be provided.

On 30 August 2023, the results of the planned interim analysis (IA) on the primary endpoint, six-minute walk distance (6MWD) versus placebo at Week 28, were presented to the independent data monitoring committee (IDMC) by the independent statistical support group (SSG). The IDMC recommended to prematurely stop for futility (see IDMC Charter Section 7 for details). Based upon this recommendation from the IDMC, the sponsor decided to discontinue the study prematurely. This was communicated to the sites on 05 September 2023.

Following the decision to prematurely terminate the study, it was decided that an abbreviated CSR would be produced to summarize key efficacy and all safety data. Such analyses are described in this SAP. The study team remains blinded to the results of the IA conducted by the SSG, until database lock occurs for the abbreviated CSR planned analyses.

This SAP refers to documents listed in Table 2. The SAP contains definitions of analysis sets, key derived variables and statistical methods for analysis of efficacy and safety for reporting of the 67896062CTP3001 MACiTEPH study.

Table 2: Study Documents

Document (EDMS Number)	Date, Version
Study protocol 67896062CTP3001, Amendment 4 (EDMS-RIM-265157, 5.0)	16 December 2020, Version 6.0
IDMC Charter (EDMS-RIM-381306, 3.0)	19 June 2023, Version 3.0

The SAP for IDMC regular data review meetings is provided in a separate document and is written and maintained by the SSG (see IDMC Charter Section 8 for details).

Titles, mock-ups, and programming instructions for all statistical output (tables, listings, and figures) are provided in a separate data presentation specifications document.

Source data for the analyses are provided as Statistical Analysis Software (SAS®) data sets according to Clinical Data Interchange Standards Consortium Study Data Tabulation Model (SDTM). All descriptive or formal statistical analyses will be performed using SAS statistical software (Version 9.4), unless otherwise specified.

1.1. Objectives and Endpoints

Section 3 of the study protocol outline all objectives and endpoints of the study.

Changes to protocol planned analyses are described in Section 6.2, Appendix 2.

1.1.1. Primary Objective and Endpoint

Primary Objective	Primary Endpoint
To evaluate the effect of macitentan 75 mg versus placebo on exercise capacity at Week 28 in participants with chronic thromboembolic pulmonary hypertension (CTEPH).	Change from baseline to Week 28 in exercise capacity (6-minute walk distance [6MWD], as measured by the 6-minute walk test [6MWT]).

1.2. Main Study Design

MACiTEPH is a randomized, double-blind (DB), placebo-controlled, multicenter, two-stages, group sequential, adaptive study to evaluate efficacy and safety of macitentan 75 mg in men and women between 18 and 80 years of age, with a diagnosis of inoperable or persistent/recurrent chronic thromboembolic pulmonary hypertension (CTEPH). CTEPH diagnosis and eligibility is confirmed by an adjudication committee (AC) prior to randomization.

At least 144 (up to a maximum of 230) participants will be randomized in a 1:1 ratio to macitentan 75 mg or placebo given once daily. Study intervention allocation is stratified by use of PH-specific therapy (riociguat /other PH-specific therapy/none) and by inoperable (with or without balloon pulmonary angioplasty (BPA)) vs persistent/recurrent after pulmonary endarterectomy (PEA) surgery (with or without BPA).

The DB treatment period of the study includes a 28-week fixed duration part followed by a variable duration part which will last until all randomized participants have either completed their Week 28 visit or prematurely discontinued the study. Participants can enter the open-label (OL), single-arm treatment extension period with macitentan 75 mg, after having completed the DB period or based on the investigator's approval for those who had a clinical events committee (CEC) confirmed clinical worsening event and remained in the study until at a minimum up to 28 weeks (either on treatment or in PTOP). See details in Section 1.2.1.

Prior to protocol amendment 4, the DB treatment period of the study included a 52-week fixed duration period (See Figure 3 and Figure 4). The assessment of the primary endpoint nor the analysis as described next did not change.

This study is designed as a two-stage adaptive design with one interim analyses (IA). At the IA, a futility analysis, and a sample size re-estimation (SSRE) will be performed in an unblinded manner by an external independent Statistical Support Group (ISSG). Results will be communicated directly to the Independent Data Monitoring Committee (IDMC).

MACiTEPH also includes a hemodynamic sub-study (see Section 1.3) and a male reproductive system safety sub study. No participant was enrolled in the male reproductive system safety sub-study; hence there are no data to report for the CSR.

1.2.1. Study Periods

The study comprises the following periods as depicted in Figure 1:

- **Screening period (at least 14 days and up to 60 days**): starts with the signature of the informed consent form and ends with the participant's randomization at the beginning of the DB period (first dose of study intervention). Individuals who do not meet the criteria for participation in this study (screening failure) may be re-screened twice if the reason for non-eligibility is transient.
- **DB treatment period:** The overall DB treatment period is divided into a 28-week fixed duration part and a variable duration part which will end when all randomized participants in the study have either completed the Week 28 visit or prematurely discontinued the study. DB treatment period starts on Day 1 (Randomization / first dose of study intervention) and ends on the day of the last DB study intervention intake. The DB treatment period consists of an 8-week up-titration phase and a maintenance phase. During the 8-week up-titration phase participants will receive 10 mg for 4 weeks and then 37.5 mg for 4 weeks before receiving the target dose of 75 mg during the maintenance phase.

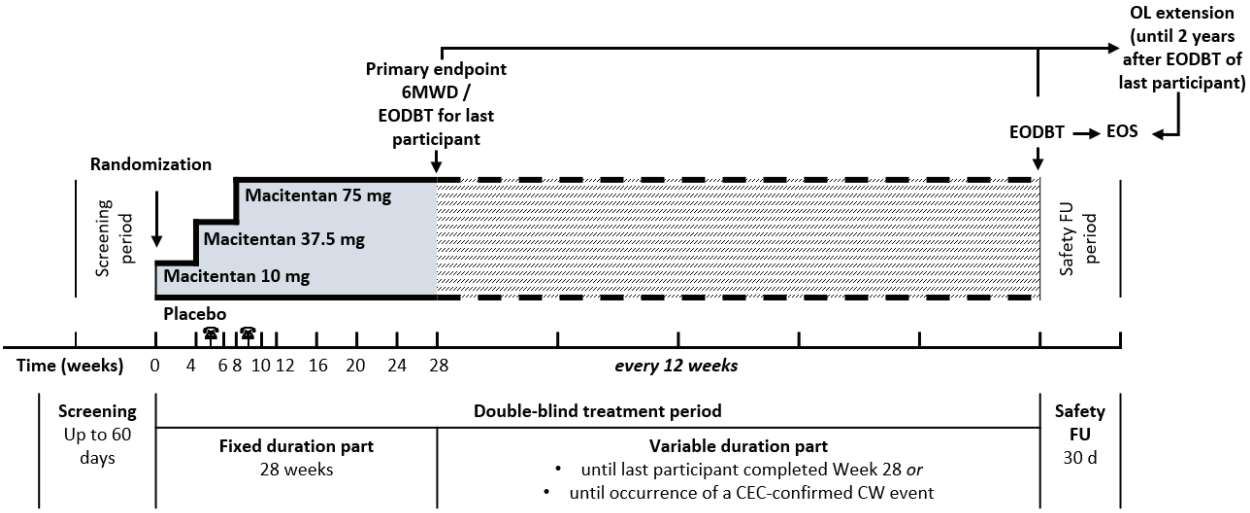
  Participants assigned to the placebo arm will receive matching placebo tablets during the up-titration and maintenance phases.
- **Post-treatment observation period (PTOP):** up to Week 28 for participants who prematurely discontinue DB study intervention but do not withdraw consent (Figure 2). In case of premature discontinuation, the DB treatment period ends on the day of intake of the last dose and the end of double-blind treatment (EODBT) visit is performed within 7 days of intake of the last dose of study intervention (EODBT + 7 days). These participants will enter the PTOP and continue to perform the visits and assessments as scheduled until Week 28 (see schedule of activities in Section 6.12, Appendix 12). The PTOP visits to be performed depend on the timepoint of discontinuation. The PTOP1 visit corresponds to the post-treatment safety follow-up visit to be performed 30 (+ 5) days after last study intervention and may be combined with any of the subsequent PTOP visits if it falls within the same time window. Participants who completed PTOP 6 up to Week 28 are eligible to enter the open-label (OL) extension period.

  Participants who prematurely discontinue the study intervention and choose not to enter /were not able to enter the OL extension will be contacted by phone approximately every 6 months to get details on their vital status and hospitalization (including date and reason, if available). Information will be recorded in the case report form (CRF), provided their consent was not withdrawn.
- **Open-label (OL) extension period:** starts on the day of the EODBT visit (i.e., EODBT + 7 days, for those premature withdrawals outside of a scheduled visit) and ends with the 'End of

open-label treatment' (EOLT) visit (i.e., EOLT + 7 days, for those premature withdrawals outside of a scheduled visit). Participants who have reached the target dose of 75 mg and completed the DB period as per protocol or experienced a CEC-confirmed clinical worsening and remained in the study at minimum up to Week 28 (either on treatment or in PTOP 6) are eligible for transitioning into the OL extension period receiving macitentan 75mg daily. Participants randomized to macitentan will continue to receive macitentan 75 mg daily. Participants randomized to placebo will receive macitentan at a dose of 10 mg for 4 weeks followed by a dose of 37.5 mg for 4 weeks prior to receiving the target dose of 75 mg. To preserve the blind of the DB treatment period, a double-dummy up-titration is performed. Both participants and sites will remain blinded to the previous DB treatment allocations during the OL extension.

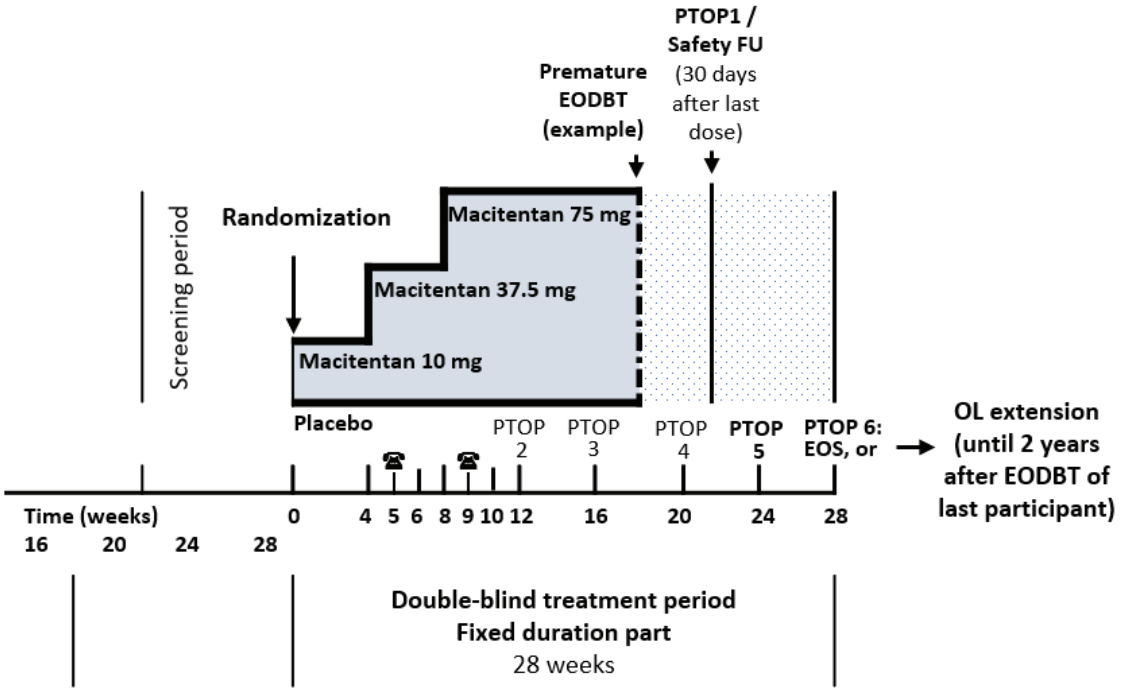
- **Safety follow-up period:** starts on the day after the last dose of study intervention and ends with a post-treatment safety follow-up visit (30 [+5] days thereafter i.e., at EODBT+30 days / EOLT+30 days. Participants who prematurely discontinue or complete DB intervention will have their safety follow-up visit done 30 days after the EODBT unless they transit directly to OL extension.

Figure 1: Schematic Overview of the Study – Participants Who Complete the DB Treatment Period as per Protocol version 6 Amendment 4



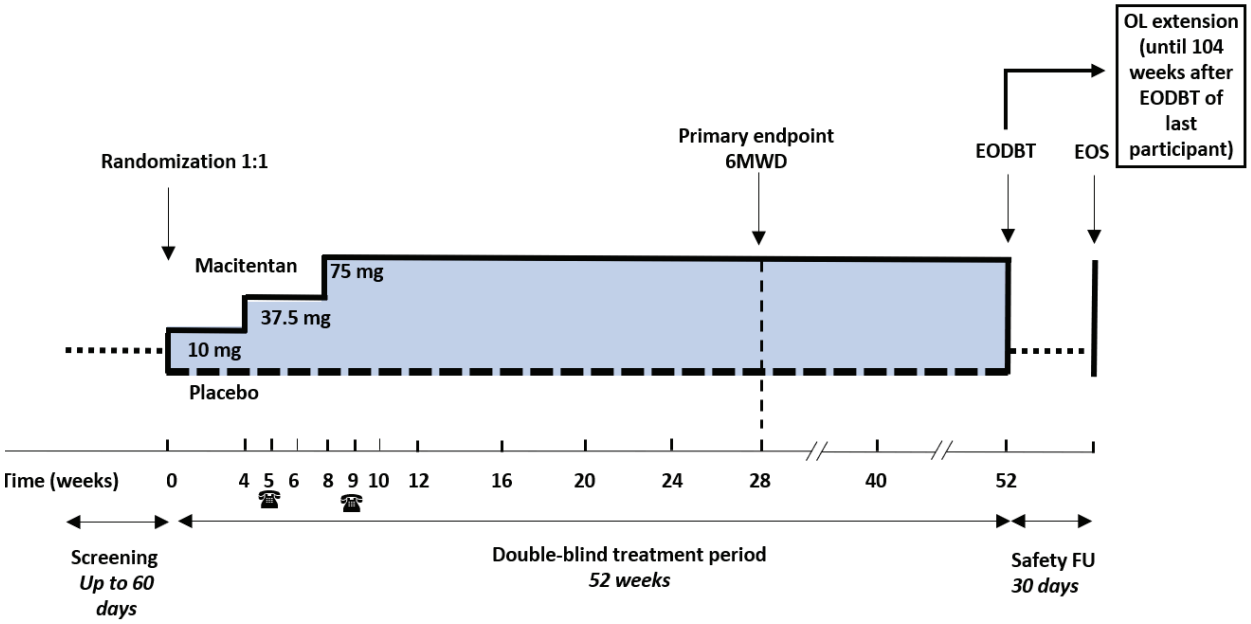
6MWD = 6-minute walk distance; EODBT = End of double-blind treatment; EOS = End of study; FU = follow-up; OL = open-label.

Figure 2: Participants Who Prematurely Discontinue DB Study Intervention (Example of a Premature Discontinuation at Week 9) as per Protocol version 6 Amendment 4)



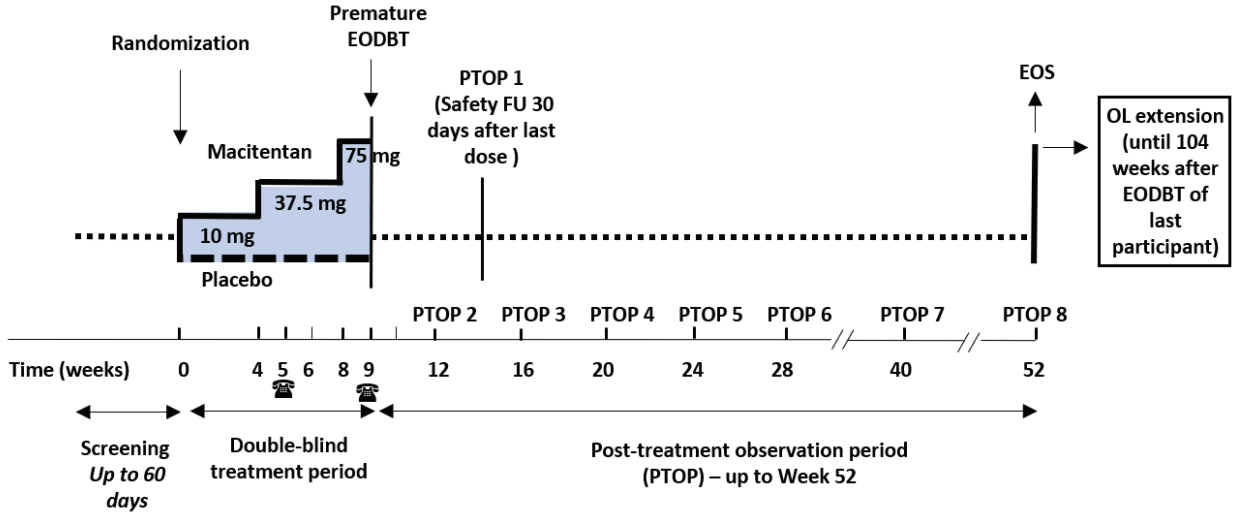
DB = double-blind; EODBT = End of double-blind treatment; EOS = End of study; FU = follow-up; PTOP = post-treatment observation period.

Figure 3: Schematic Overview of the Study – Participants Who Complete the DB Treatment Period as per Protocol versions prior to Protocol version 6 Amendment 4



6MWD = 6-minute walk distance; EODBT = End of double-blind treatment; EOS = End of study; FU = follow-up, OL = open-label.

Figure 4: Participants Who Prematurely Discontinue DB Study Intervention (Example of a Premature Discontinuation at Week 9) as per Protocol versions prior to Protocol version 6 Amendment 4



DB = double-blind; EODBT = End of double-blind treatment; EOS = End of study; FU = follow-up; PTOP = post-treatment observation period.

The schedule of activities and procedures of the study are illustrated in Table 23 and Table 24.

The double-blind study is expected to last approximately 3.5 years based on current recruitment projections. The OL extension period will end 2 years after the last participant has completed the DB period. The total study duration from first participant enrolled until last participant completing is therefore estimated to be approximately 6 years. The maximum duration of the OL extension

period and the 30-day safety follow-up period is expected to be 5 years, i.e., first participant treated with OL macitentan to last participant's end of safety follow-up period.

1.2.2. Timing of Interim and Final Analyses

It was planned to conduct one interim analysis (IA) during the DB study period and one final analysis after the end of DB study period. Similar, for the OL extension period one interim analysis and a final analysis after the end of open label extension were planned.

The IA during the DB study period was performed on the primary endpoint (change from baseline in 6MWD at week 28) by an independent SSG when 72 sequentially randomized participants had completed week 28 or discontinued the study prematurely. At the time of this IA data cut-off, all available data including those ongoing participants not having yet completed week 28 but who had (or were expected to have) conducted their first 6MWT on maintenance dose at week 12 were used to perform the futility analysis and unblinded SSRE. Overall, 84 participants were included into the interim analysis and formed stage 1 of the adaptive design.

Due to the premature termination of the study for futility following the IA during the DB study period, the final analysis for the DB period and OL extension period have been combined and constitute the final reporting of the study.

1.3. Hemodynamic Sub-Study Design

This sub-study to assess right heart hemodynamics using right heart catheterization (RHC) is conducted at selected sites who comply with study specific RHC guidelines (see Protocol Appendix 16). Enrollment in the sub-study is optional for eligible participants enrolled at the selected sites, except for participants enrolled in Japan where the participation in the sub-study is mandatory based on an agreement with the pharmaceuticals and medical devices agency (PMDA).

There is no target sample size pre-defined in the protocol due to the optional nature of the sub-study. Enrollment into the sub-study will continue until the enrollment into the main study is completed or there is a sponsor decision to stop the main study or sub-study earlier.

Hemodynamic parameters are assessed at screening and at week 28 following the guidelines detailed in Appendix 16 of the protocol. If a historical RHC was performed in the 60-day period prior to randomization, and the criteria for historical RHCs are met (see study protocol, Appendix 16), the RHC does not need to be repeated at screening. PH-specific therapy must be stable for at least 90 days prior to the baseline RHC. It is recommended to keep the dose of diuretics stable for 7 consecutive days prior to the screening and week 28 RHC.

Participants may withdraw consent from the sub-study at any time and continue in the main study. Conversely, if a participant withdraws from the main study they are also withdrawn from the sub-study.

It is noted that randomization is performed for all participants in the main study, stratifying by PH-specific therapy and inoperable vs persistent/recurrent after PEA surgery. However, entry into the sub-study is voluntary for participants from selected sites (except Japan) and as such the treatment balance within the sub-study may not be maintained.

2. STATISTICAL HYPOTHESES

2.1. Primary Hypothesis

The primary efficacy endpoint is the change in 6MWD from baseline to Week 28. The null hypothesis to be tested to address the primary objective of the trial is that there is no difference between macitentan 75 mg and placebo in exercise capacity (as measured by the change from baseline in 6MWD to Week 28), or that the difference favors placebo. Then, the alternative hypothesis is that the difference in the changes from baseline is greater in macitentan 75 mg compared to placebo i.e., treatment effect is in favor of macitentan 75 mg.

The null and alternative hypotheses are formulated in terms of the difference between study intervention groups at Week 28, as estimated from an MMRM approach utilizing changes from baseline to post-baseline 6MWD values (Weeks 4, 8, 12, 16, 20, 24, and 28).

If $\mu_{MACI,28}$ and $\mu_{PBO,28}$ denote the mean changes from baseline in 6MWD at Week 28 for the participants treated with macitentan 75 mg and placebo, respectively, then the following null (H_0) and alternative (H_1) hypotheses are considered:

$$H_0: \delta = \mu_{MACI,28} - \mu_{PBO,28} \leq 0$$
$$H_1: \delta = \mu_{MACI,28} - \mu_{PBO,28} > 0$$

The null hypothesis is tested at a study-wise 1-sided significance level of 0.025.

To ensure control of type I error in presence of an adaptive design with unblinded SSRE it was planned to conduct an inverse normal combination test as described in Section 5.7.2. Due to limited data being available in stage 2 of the adaptive design following declaration of futility, analyses irrespective of stage will be presented.

2.2. Secondary Hypotheses

Secondary Endpoints	Hypotheses
Time to first clinical worsening (CEC confirmed) up to EODBT	The null hypothesis (H_{01}) is that there is no difference between the macitentan 75 mg and placebo survival distributions i.e., S(t) or that the difference favors placebo [H_{01}: $S_{MACI}(t) \leq S_{PBO}(t)$, $t \geq 0$]. The alternative hypothesis (H_{11}) is that there is a difference between study intervention groups in favor of macitentan 75 mg [H_{11}: $S_{MACI}(t) > S_{PBO}(t)$, $t \geq 0$].
Improvement (decrease) in WHO FC from baseline to Week 28 (yes/no)	The null hypothesis (H_{02}) is that the proportion of participants who improve is the same for macitentan 75 mg and placebo or is higher for placebo [H_{02}: $P_{MACI,28} \leq P_{PBO,28}$]. The alternative hypothesis (H_{12}) is that the proportion of participants who improve is higher for macitentan 75 mg [H_{12}: $P_{MACI,28} > P_{PBO,28}$].

CEC = Clinical Event Committee; EODBT = end of double-blind treatment; WHO-FC = World health Organization Functional Class

All analyses planned for the OL extension period and the hemodynamic sub-study are exploratory. No statistical hypotheses will be tested.

2.3. Overall Testing Strategy

This study implements a two-stage, group sequential adaptive design with one IA with hypotheses tested at a study-wise 1-sided significance level of 0.025.

Details of the overall testing strategy for the primary and secondary endpoints are provided in Section 9.1.3 of the study protocol.

3. SAMPLE SIZE DETERMINATION

Full details regarding sample size determination for the main study are provided in the protocol Section 9.2.

At least 144 participants (up to a maximum of 230) were planned to be randomized in a 1:1 ratio to macitentan 75 mg or placebo. At the time of the decision to terminate the study, a total of 127 participants had been randomized. 7 participants entered the open-label treatment extension period.

For the hemodynamic sub-study, there is no target sample size pre-defined in the protocol due to the exploratory nature of the study. At the time of the decision to terminate the study, a total of 65 participants had been entered into the hemodynamic sub-study.

4. POPULATIONS (ANALYSIS SETS) FOR ANALYSIS

The overview of analysis sets is provided in Table 3.

Table 3: Overview of Analysis Sets

Analysis Sets	Description
Screened Analysis Set (SCR)	All screened participants who are assigned a participant identification number.
Full Analysis Set (FAS)	The FAS includes all randomized participants who are assigned to a study intervention. Participants will be analyzed according to the intervention they have been assigned to via interactive web response system (IWRS)
Safety Analysis Set (SS)	The safety analysis set includes all participants who received at least 1 dose of study intervention. Participants will be analyzed according to the intervention they actually received (which may be different to the study intervention group they are assigned to via IWRS).
Pharmacokinetics/Pharmacodynamic (PK/PD) Analysis Set	The PK/PD analysis set is defined as participants who received at least 1 dose of study intervention and have at least 1 valid blood sample drawn for PK and/or ET-1 analysis.
Open Label Analysis Set (OLAS)	All participants who received at least one dose of OL study intervention in the OL treatment extension period.
RHC Set (RHCS)	All randomized participants who took part in the hemodynamic sub-study, i.e., who agreed to participate in the hemodynamic sub-study and have at least one valid baseline or post-baseline right heart catheterization (RHC) assessment

Table 4 defines an overview of analysis sets and their key usage.

Table 4: Analysis Sets and Key Usage

Analysis	Analysis Set					
	SCR	**FAS**	**Safety Analysis Set**	**PK/PD**	**OLAS**	**RHCS**
Number of participants by analysis set	x	x	x	x		x
Participant disposition		x	x		x	x
Demographics and baseline characteristics		x			x	x
Previous and concomitant diseases		x				
Intervention compliance		x	x[2]			
Extent of exposure			x			x
Protocol deviations		x				
Prior and concomitant medication		x				
Primary and secondary efficacy endpoints[1]		x				
Exploratory efficacy[2]		x				
Safety endpoints			x		x	
PK/PD				x		
Hemodynamic measures						x

[1] Efficacy endpoints include 6MWD, clinical worsening, and WHO FC
[2] Exploratory efficacy includes NT-proBNP
PD = Pharmacodynamic; PK = Pharmacokinetic; PK = PK Set; SCR = Screened Analysis Set

5. STATISTICAL ANALYSES

5.1. General Considerations

5.1.1. Study Day and Relative Day

Study Day 1 refers to the date of the first dose of DB study intervention intake. For participants who were randomized and never received study intervention, the Study Day 1 is imputed with the randomization date. Unless otherwise specified, all efficacy and safety assessments at all visits will be assigned a day relative to Study Day 1.

Study day for a visit or a measurement or an event occurrence is defined as:

- Visit/measurement/event date – Study Day 1 date +1 if visit/measurement/event date is ≥ date of Study Day 1.
- Visit/measurement/event date – Study Day 1 date if visit/measurement/event date < Study Day 1 date.

There is no 'Day 0'. The double-blind treatment period ends on the day of the last DB study intervention intake (EODBT).

Day 1 of the open label treatment extension period (OL Day 1) refers to the date of the first OL treatment intake after the EODBT visit for DB treatment completers or at PTOP 6 visit for those who discontinued DB study intervention prematurely during the fixed DB part and will be defined for participants treated in the OL extension only.

5.1.2. Baseline

5.1.2.1. DB Period

Baseline is defined as the last non-missing observation prior to or on the first DB treatment intake (i.e., Study Day 1), unless otherwise specified.

In case of multiple assessments on the same day, then the last assessment (taking into consideration date/time of assessment) will be used to define baseline.

5.1.2.2. OL Extension Period

Baseline for the OL extension period is defined as the last non-missing observation obtained prior or on the first OL study treatment.

5.1.2.3. Hemodynamic Sub-Study

Baseline is defined as the last assessment obtained prior to or on day of first intake of study intervention in the DB treatment period (including unscheduled and historical assessments).

All participants in the sub-study must have a baseline RHC to be performed during the screening period unless the criteria for the historical RHC are met and may be used as the baseline RHC (see Section 1.3). Data from historical assessments may be used as entered on the eCRF.

5.1.3. Double-blind Treatment Phases

The double-blind treatment period is divided into 3 phases according to the start / end date of intervention for each dose in the up-titration phase (10 mg and 37.5 mg macitentan / matching placebo) and the maintenance phase (macitentan 75 mg / matching placebo) as summarized in Table 5 below.

Table 5: DB Treatment Phases

Phase # (Dose level)	Start date	End date
Up-titration Phase 1 (10 mg macitentan/matching placebo)	Date of first intake of 10 mg dose[1]	Date of last intake of 10 mg dose and before intake of first 37.5 mg dose
Up-titration Phase 2 (37.5 mg macitentan/matching placebo)	Date of first intake of 37.5 mg dose	Date of last intake of 37.5 mg dose and before intake of first 75 mg dose
Maintenance Phase 3 (75 mg macitentan/matching placebo)	Date of first intake of 75 mg dose	Date of last intake of 75 mg dose and before intake of first macitentan OL dose
DB treatment phase	Date of first intake of 10 mg dose[1]	Date of last intake of DB dose[2]

[1] Start of DB treatment period = Study Day 1
[2] For the DB treatment completers, EODBT=date of last intake of 75 mg dose and before intake of first macitentan OL dose.

When all randomized participants have either completed the Week 28 visit or prematurely discontinued the DB period of the study, the entire DB period of the study is considered as completed. That means, at this time, ongoing participants who have completed at least 28 weeks of the DB study intervention period are considered to have completed their DB treatment period (except participants on protocol version 5 or earlier who are considered to have completed their DB treatment period if they complete 52 weeks of the DB study intervention). The EODBT visit will occur within 7 days after the last intake of study intervention (EODBT+7 days) (see Section 6.12 Appendix 12). For participants who continue in the OL extension period, EODBT assessments are done before their first dose of OL intervention.

The reasons for premature discontinuation of study intervention are described in Protocol Section 7.

A participant who prematurely discontinues DB study intervention during the fixed duration part is not considered as withdrawn from the study and will be asked to enter the PTOP which lasts until Week 28, provided that the participant's consent for this limited participation in the study has not been withdrawn.

The participant with premature discontinuation outside of a regular visit will be asked to return for a premature EODBT visit within 7 days after the last intake of study intervention (EODBT+7 days) and for a safety follow-up visit 30 (+5) days after the last intake of study intervention (EODBT+30 days, i.e., at PTOP1). If the decision to stop study intervention is taken at a regular visit, this visit may become the premature EODBT visit. If the premature EODBT visit or safety follow-up visit falls within the time-window of any of the PTOP visits, the visits are combined i.e., a participant who discontinues at Week 16 will have his/her PTOP 1[Safety follow-up] combined with PTOP 4 at Week 20 and then continue with PTOP 5 at Week 24 and PTOP 6 at Week 28.

5.1.4. Study Completion

Protocol version 6

A participant will be considered to have completed the entire double-blind period if he or she has remained in the DB period until the end of the DB period is announced and either entered the OL extension period or had a safety follow-up visit (i.e., at EODBT+30 days). The end of the DB period of the study will be announced once all participants have completed the 28-week fixed duration period.

Participants who prematurely discontinue study intervention for any reason before completion of the double-blind phase will not be considered to have completed the study.

A participant will be considered to have completed the OL Treatment Extension Period if he or she remained in the study until the end of the OL period is announced and has completed the safety follow-up / EOS visit or entered a continued access program.

Protocol version 5 or earlier

Participant who completes 52 weeks of follow-up (± time-window indicated in schedule of assessments) including the safety follow-up period (if the participant does not continue in the OL extension period) is considered to have completed the DB period of the study as per protocol, regardless of whether he/she has remained on treatment. A participant will be considered to have completed the OL Treatment Extension Period if he or she did not prematurely withdraw from the study and has completed the safety follow-up / EOS visit.

5.1.5. End of study

The end of study (EOS) is considered as the last visit for the last participant in the overall study.

The definition of the EOS visit on the participant level depends on the status of the participant as described in Table 6 below.

Participants who do not enter the OL treatment extension phase:

- Participants who completed the DB treatment per protocol: EOS corresponds to the Safety FU visit.

- Participants who discontinued DB treatment during the fixed duration part and entered PTOP: EOS corresponds to the PTOP 6 visit
- Participants who discontinued DB treatment during the fixed duration part and did not enter PTOP: EOS corresponds to the Safety FU visit 30 (+5) days after the last intake of DB study treatment
- Participants who discontinued DB treatment at the end of the fixed duration part, or during the variable duration part of the DB treatment period: EOS corresponds to the Safety FU visit 30 (+5) days after the last intake of DB study treatment

Participants who entered the OL extension:

- Participants who either discontinue OL treatment prematurely or complete it per protocol: EOS corresponds to the Safety FU visit 30 (+5) days after the last intake of OL study treatment
- Participants who transition to a post-trial access option (post-trial access program, other OL study): EOS corresponds to the EOLT visit

Withdrawal of consent:

- EOS will be the last visit performed

Table 6: EOS Visit Definition

DB treatment period			**OL extension period**		**EOS visit definition for an individual participant[1]**
Fixed duration part		**Variable duration part**			
Completed (Y/N)?	**Entered & completed PTOP (Y/N)?**	**Completed (Y/N)?**	**Entered (Y/N)?**	**Completed (Y/N)?**	
Y	N/A	Y	N	N/A	EOS=EODBS=EODBT +30 days
Y	N/A	N	N	N/A	EOS=EODBS=EODBT +30 days
N	Y	N/A	N	N/A	EOS= EODBS =PTOP6
N	N	N/A	N	N/A	EOS= EODBS =EODBT +30 days
Y	N/A	Y	Y	Y	EOS=EOLT +30 days[2]
Y	N/A	Y	Y	N	EOS=EOLT +30 days[2]
Y	N/A	N	Y	Y	EOS=EOLT +30 days[2]
Y	N/A	N	Y	N	EOS=EOLT +30 days[2]
N	Y	N/A	Y	Y	EOS=EOLT +30 days[2]
N	Y	N/A	Y	N	EOS=EOLT +30 days[2]

[1] Participants who withdrew consent, EOS=Last Visit performed
[2] Participants who transition to a post-trial access option (post-trial access program, other OL study): EOS = EOLT visit
EODBS = end of DB study, EODBT = end of double-blind treatment; EOLT = end of open-label treatment; EOS = end of study; OL = open-label; N/A = not applicable; PTOP = post-treatment observation period up to week 28.

5.1.5.1. Hemodynamic Sub-Study

A participant is considered to have completed the hemodynamic sub-study if they completed both a baseline and Week 28 RHC assessment and the participant did not withdraw from the sub-study

and did not discontinue from the main study prior to the Week 28 assessment, irrespective of DB study intervention discontinuation.

A participant is considered to have completed the hemodynamic sub-study to Week 28 on DB study intervention if they did not discontinue study intervention prior to the Week 28 RHC assessment, where the Week 28 RHC assessment was performed on the protocol defined window of Week 28 +/- 5 days (study 192 to 202).

Participants are considered to have discontinued the DB study intervention during the hemodynamic sub-study if they discontinued from DB study intervention prior to Week 28 RHC assessment (performed in the protocol defined window of Week 28 ± 5 days), or prior to study day 197 if the Week 28 RHC assessment was not performed or performed out of window.

5.1.6. Visit Windows

In case of non-adherence to the protocol visit schedule, in order to minimize missing data and to analyze the data at the relevant planned (scheduled) visits, all recorded assessments visits (including unscheduled visit results) for each subject up to EODBT+30 days (for safety endpoints) will be reassigned to the most appropriate visit. Listed below are the visit windows and the target days for each visit. The reference day is Study Day 1 for the DB period. If a participant has two or more visits in each visit window, the visit closest to the target day will be selected as the analysis visit for that visit window. The other additional visit(s) will not be used in the summaries or analyses unless specified otherwise (for example, they can be used for determination of clinically important endpoints). If two visit assessments are equidistant from the target day within a visit window, the later visit assessment is selected, unless the later visit is off-treatment (after EODBT) and the earlier visit is on DB treatment, in which case the earlier visit will be selected to capture on treatment assessment.

All assignments will be made in chronological order. Once a visit date has been assigned to a visit window, it will no longer be eligible to be used for a later time point. The analysis visit windows and the target days for each visit defined in the protocol are given below.

All the assessments considered for the visit windows as described below must occur prior to first dose of open label macitentan. Date of first dose of OL treatment is included in the upper limit where OL Day 1 visit (and first dose OL) is combined with EODBT visit and assessments are to be performed prior to start of OL treatment. Inclusion /exclusion of assessments occurred during the PTOP is analysis dependent and therefore, must be evaluated for inclusion prior to performing the analysis. The assessments performed up to the safety follow-up visit are in general used for safety analysis and for certain efficacy analysis where up to EODBT + 30 days is mentioned.

Table 7: Visit Windows: 6MWD, and NT-proBNP

Parameter	Analysis Period	Scheduled Visit Number	Analysis visit (label on output)	Time Interval (Day)* #	Target Time Point** (Day)
6MWD	Screening[1]	1	Screening	< 1	-14 (up to -60) to -1
NT-proBNP	Baseline[2]	2	Baseline/ Randomization / D1	≤1	1
	Double Blind	3	Week 4	2 to 42	29
		6	Week 8	43 to 70	57
		9	Week 12	71 to 98	85
		10	Week 16	99 to 126	113
		11	Week 20	127 to 154	141
		12	Week 24	155 to 182	169
		13	Week 28	183 to 238	197
		14	Week 40	239 to 322	281
		15	Week 52	323 to 406	365
		16	Week 64	407 to 490	449
		…	… (every 12 weeks)	…	…

* Relative to [Study Day 1], ** based on the protocol defined schedule.
1 Not applicable for NT-proBNP.
2 The last non-missing assessment prior to or on the day of the first DB study intervention intake (see Section 5.1.2).
# Up to last non-missing assessment prior to the first OL intervention intake (includes Day 1 OL as EODB/ Day 1 OL visits are combined and assessments are to be performed prior to start of OL treatment).
EODBT=End of double-blind treatment; 6MWD = 6-minute walk distance; DB = Double-blind.
NT-proBNP = N-terminal prohormone of Brain natriuretic peptide.

Table 8: Visit Windows for WHO FC, Vitals Signs, Body Weight, Central Laboratory

Parameter	Analysis Period	Scheduled Visit Number	Analysis visit (label on output)	Time Interval (Day)* #	Target Time Point** (Day)
WHO FC	Screening	1	Screening	< 1	-14 (up to -60) to -1
Vital signs Body weight	Baseline[1]	2	Baseline / Randomization / D1	≤1	1
Central laboratory	Double Blind	3	Week 4	2 to 35	29
		5	Week 6	36 to 49	43
		6	Week 8	50 to 63	57
		8	Week 10	64 to 77	71
		9	Week 12	78 to 98	85
		10	Week 16	99 to 126	113
		11	Week 20	127 to 154	141
		12	Week 24	155 to 182	169
		13	Week 28	183 to 238	197
		14	Week 40	239 to 322	281
		15	Week 52	323 to 406	365
		16	Week 64	407 to 490	449
		…	… (every 12 weeks)	…	…

* Relative to study Day 1 ** based on the protocol defined schedule.
1 The last non-missing assessment prior to or on the day of the first DB study intervention intake (see Section 5.1.2).

[#] Up to last non-missing assessment prior to the first OL intervention intake (includes Day 1 OL as EODB/ Day 1 OL visits are combined and assessments are to be performed prior to start of OL treatment).
WHO FC = World Health Organization functional class.

Table 9: Visit Windows for 12 Lead ECG

Parameter	Analysis Period	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)[*]	Target Time Point[**] (Day)
12 Lead ECG	Baseline[1]	2	Baseline/ Randomization / D1	≤1	1
	Double Blind	6	Week 8	2 to 70	57
		9	Week 12	71 to 140	85
		13	Week 28	141 to 280	197
		15	Week 52	≥281[#]	365

[*] Relative to [Study Day 1], [**] based on the protocol defined schedule.
[1] The last non-missing assessment prior to or on the first DB study intervention intake (includes Day 1 OL as EODB/ Day 1 OL visits are combined and assessments are to be performed prior to start of OL treatment).
[#] Up to last non-missing assessment prior to the first OL intervention intake (includes Day 1 OL as EODB/ Day 1 OL visits are combined and assessments are to be performed prior to start of OL treatment).
ECG = Electrocardiogram.

Table 10: Visit Windows for Hemodynamic Assessments

Parameter	Analysis Period	Scheduled Visit Number	Time Interval* (label on output)	Time Interval (Day)[*]	Target Time Point[**] (Day)
Hemodynamic Endpoints	Baseline	1	Baseline	≤1	1
	Double Blind	13	Week 28	169 to 225	197

[*]Time interval and target time point are relative to study day 1 (first dose of DB study intervention) [**] based on the protocol defined schedule.
Analysis time windows (+/- 4 weeks) are wider than the assessment schedules (+/- 5 days specified in protocol)

5.1.7. Pooling of Countries

Five regions will be defined by pooling participating countries (see Table 17) to reflect both geographical and adherence of current local clinical practices to the PH treatment guidelines.

5.2. Participant Dispositions

5.2.1. Disposition Reporting

Screened participants, screening failures and reason for screen failures will be summarized overall (n and %) based on the screened analysis set. Participants who are subsequently randomized following a re-screening are not considered screening failures.

5.2.1.1. DB Period

The number and percentage of participants in the following disposition categories will be summarized by intervention group and overall based on the FAS and the safety analysis set as specified.

Study disposition will be summarized on FAS and will show:

- Participants randomized

- Participants received at least one dose of DB study intervention
- Participants who completed the 28-week fixed duration part of the DB study period irrespective of whether on or off study intervention
- Reasons for discontinuation of the 28-week fixed duration part of the DB study prematurely irrespective of whether on/off treatment
- Participants who entered PTOP
- Participants who completed PTOP: Participants who entered PTOP are considered to have completed PTOP if they completed visit PTOP 6 (Week 28) under protocol version 6 or visit PTOP 8 (Week 52) under protocol version 5 or earlier.
- Participants who completed the DB study period
  - Completed DB study intervention (Protocol v6)
  - Participant completed 52-weeks fixed study period irrespective of whether on/ off study intervention (protocol v5 or earlier, identified as signing the last inform consent prior to 16 December 2020)
- Reasons for discontinuing DB study period prematurely (irrespective of whether on/off treatment **).

Treatment disposition will be summarized on SS and will show:

- Participants treated with at least one dose of DB study intervention
- Participants who completed the 28-week, fixed duration part of the DB study intervention
- Participants who discontinued DB study intervention prematurely during fixed duration part (up to Week 28)
- Reasons for prematurely discontinuation of DB study intervention during the 28-week fixed duration part of the DB study, and possibly related to study intervention (yes/no)*
- Participants who completed the DB (fixed and variable duration part) study intervention period
- Participants who completed fixed 52-week DB study intervention (protocol v5 or earlier)
- Participants who discontinued DB study intervention prematurely:
  - Overall
  - During the 10 mg up-titration phase
  - During the 37.5 mg up-titration phase
  - During the 75 mg maintenance phase
- Reasons for prematurely discontinuation DB study intervention (during the fixed and variable duration part), and possibly related to DB study intervention (yes/no)*

DB Study Intervention Completion and Discontinuation

DB study intervention completion is defined according to protocol version (see Section 5.1.3).

Participants who completed/ prematurely discontinued the DB study intervention are based the information recorded in the DB "Treatment Disposition (End of treatment)" form on the eCRF. Participants who completed the fixed 52-week DB study intervention are the subset of participants that completed study intervention according to the DB "Treatment Disposition (End of treatment)" form under protocol version 5 or earlier.

Premature treatment discontinuations during the up-titration / maintenance treatment phases are assigned according to the date of the premature treatment discontinuation recorded in the DB "Treatment Disposition (End of treatment)" form on the eCRF.

Participants who discontinued DB study intervention prior to Week 28 (before or on study day 197), are considered to prematurely discontinued during the 28-week fixed duration part of the DB study intervention. Participants who either completed DB intervention or prematurely discontinued after study day 197 are considered to have completed the 28-week fixed duration part of the DB study intervention.

* The following reasons for premature discontinuation of study intervention will be considered as **possibly related to study intervention**:

- adverse event, death, lack of efficacy
- withdrawal by subject with sub-reason lack of Improvement; subject refused further study treatment with sub-reason lack of improvement.
- physician decision, subject refused further study treatment, withdrawal by subject, or withdrawal by legally authorized representative, if accompanied by sub-reason "Treatment Related"

In addition, all participants who discontinued study intervention prematurely who experienced a CEC-confirmed clinical worsening event up to EODBT + 7 days irrespective of recorded reason for premature treatment discontinuation will be considered as having reason for study intervention discontinuation as possibly related to study intervention.

DB Study Completion and Discontinuation

DB study disposition is defined for participants according to the protocol version at time of their completion / discontinuation (see Section 5.1.4). Participant disposition is recorded in the "Trial Disposition (Completion / Discontinuation)" form on the eCRF**.

- **Protocol v6:** a participant is considered to have completed the DB period if they have either
  - 1) not entered OL AND completed the study (as recorded on the "Trial Disposition (Completion / Discontinuation)' form on the eCRF)

OR

- 2) continued into OL extension AND completed DB study intervention as recorded on the DB "Treatment Disposition" eCRF.

- **Participants Protocol version 5 or earlier:** a participant is considered to have completed the DB period of the study if they have either:
  - 1) not entered OL AND completed the study as recorded on the "Trial Disposition (Completion / Discontinuation)' form on the eCRF (not entered OL).

  OR

  - 2) received OL after 52 weeks of follow-up in DB period [DB visit ≥ study day 360; includes the lower limit (-5 days) of the protocol time window]

**Reasons for discontinuing DB study prematurely are per the investigator reported reason on the "*Trial Disposition (Completion / Discontinuation)"* form when participants have not entered OL. Where participants have entered the OL prior to the EODB being declared then the reason for not completing the DB study = reason for discontinuation of DB treatment as reported by the investigator on the "T*reatment Disposition (Completion / Discontinuation"* form. Per protocol version 6 a participant should not be considered as "completed" for DB treatment or DB study if they discontinue treatment for any reason prior to EODB being declared.

5.2.1.2. OL Extension Period

Disposition status for the OL treatment extension period will be listed and summarized on the OL analysis set. The number and percentage of participants in the following disposition categories will be presented overall and partitioned by DB study intervention group (received).

Treatment disposition during OL will show:

- Participants treated with OL macitentan intervention
  - Overall
  - Transitioned to OL macitentan after CEC-confirmed clinical worsening
  - Transitioned to OL macitentan at end of DB study (protocol version 6) Transitioned to OL macitentan at Week 52 of DB study (up to protocol version 5)
- Participants prematurely discontinued OL macitentan with reasons for discontinuation
- Participants who completed OL treatment

Participants who completed/ prematurely discontinued the OL study intervention are based on the investigator reported record in the OL "Treatment Disposition (End of treatment)" form on the eCRF.

Study disposition during OL will show:

- Participants who entered the OL extension period
- Participants treated with OL macitentan intervention

- Participants who completed the OL extension period (irrespective of whether they completed study intervention)
- Participants prematurely discontinued OL extension period with reason for discontinuation

A participant will be considered to have completed/ prematurely discontinued the OL extension period if they have completed / discontinued the study per the 'Trial Disposition' form on the eCRF after entering the OL study period.

Survival Status at the end of 28-week fixed duration period

Survival status at end of the Week 28 fixed duration period in the FAS will be summarized. The number and percentage of participants in the following categories will be provided by study intervention arm and overall, on the FAS:

- Participants known to be alive
- Participants who died
- Participants with unknown survival status

The following listings will be provided (if applicable):

- Participants with temporary study intervention interruptions, duration of interruptions, last dose taken and reasons (see Protocol Section 7.1.2 for details)
- Participants with permanent discontinuation of study intervention, last dose taken, and reasons (see Protocol Section 7.1.1 for details) and whether possibly related to study intervention (yes/no)
- Participants who prematurely discontinued the study and reasons for discontinuation
- Participants who were randomized but did not receive study intervention
- Participants who received wrong study intervention despite their randomized assignment
- Participants who were unblinded during the double-blind study period
- Participants who were mis-stratified i.e., stratification factors as collected on eCRF do not match those entered in IwRS (PH therapy and CTEPH population)
- Screen failures and reason for screening failures

5.3. Primary Endpoint Analysis

The primary endpoint, primary estimand and the primary analysis methods are described below in subsequent Sections 5.3.1, 5.3.2 and 5.3.3, respectively.

The primary endpoint analysis is performed when all participants have completed the 28-week DB period or prematurely discontinued the study. In this SAP, this analysis is referred as the main analysis and will be performed on the FAS.

5.3.1. Definition of Endpoint

The primary efficacy endpoint is defined as the change from baseline to Week 28 in exercise capacity (6-minute walk distance [6MWD], as measured by the 6-minute walk test [6MWT]).

Baseline for 6MWD is defined as the last non-missing assessment obtained prior to or on day of the first intake of DB study intervention.

5.3.1.1. The 6-Minute Walk Test

The 6MWT is a standardized test that measures the distance walked in 6 minutes. It will be performed at the timepoints shown in the Schedule of Activities (see Section 6.12 Appendix 11). During the Screening period, an eligibility 6MWT, followed by a baseline 6MWT is performed in the 1-week period prior to Randomization or at Randomization. Further details of the 6MWT are described in study protocol section 8.2.1.

5.3.2. Estimand

5.3.2.1. Primary Estimand

Primary trial objective:

To evaluate the effect of macitentan 75 mg versus placebo on exercise capacity at Week 28 in participants with CTEPH.

Estimand scientific question of interest:

What is the effect on change from baseline to Week 28 in the exercise capacity [6MWD], when participants diagnosed with CTEPH, as defined by the study eligibility criteria, are assigned to macitentan 75 mg versus placebo?

Estimand

The estimand is described according to the following four attributes:

A. **Population:** Participants diagnosed with CTEPH, meeting the study eligibility criteria.
B. **Variable:** 6MWD changes from baseline to Weeks 4, 8, 12, 16, 20, 24, and 28.
C. **Intercurrent events (IE):** IE and related strategies are summarized in Table 11 below. IEs up to Week 28 are defined as those that occur up to study day 197.

Table 11: Intercurrent Events and Related Strategy for Primary Estimand

Intercurrent Event	Strategy
Death	A composite strategy will be used for deaths occurring up to 7 days after study intervention discontinuation (EODBT +7 days) without administration of rescue CTEPH therapy[1]. All scheduled assessments following the date of death will be imputed by a 'worst-case' imputation, i.e., the sum of the participant's baseline value and the lowest[2] observed value of 6MWD change from baseline across both intervention groups and over all assessments up to Week 28[3]. All assessments prior to the death, as well as all imputed assessments, thereafter, will be considered in the analysis.
Premature treatment discontinuation / Administration of rescue therapy[1]	A composite strategy will be employed for premature discontinuation of study intervention that are possibly treatment related (e.g., clinical worsening, lack of response, adverse events, informed consent withdrawal due to tolerability issues see details in section 5.3.3.2) and/or administration of rescue CTEPH therapy[1]. All 6MWD assessments performed after end of intervention (possibly treatment related) plus 7 days (EODBT + 7 days) or after administration of rescue therapy will be ignored in the analysis and for each visit occurring after the IE, the sum of the participant's baseline 6MWD value and the minimum of the following values will be considered instead: (1) the last observed value (observed 6MWD change from baseline) prior to intercurrent event at the participant level (2) the 25th percentile of observed 6MWD change from baseline values (both study intervention groups pooled and over all assessments up to Week 28)[3] For premature discontinuations of study intervention that are not treatment related, no penalty is imposed (not considered as an IE) and the values are assumed to be missing at random, as they are implicitly handled by the model. i.e., observed values after EODBT + 7 days, including during PTOP are not discarded.
6WMT not performed due to clinical worsening or PH related	For 6MWT that are not performed with reason not done recorded as "clinical worsening", or "Other" with reason not done specified as "PH related", the value for the missing assessment will be imputed in the analysis following the same imputation algorithm as specified for Premature treatment discontinuation /Administration of rescue therapy.

[1] Initiation or dose escalation of rescue medication such as ERAs (e.g,, ambrisentan, bosentan), Riociguat, PDE-5 inhibitors, prostanoids, prostacyclin analogs, prostacyclin receptor agonist and rescue procedures such as BPA and PEA (See Protocol Section 6.5.1 for a description of rescue therapy. See Section 6.10 for full details of PH therapies / rescue medication).

[2] To avoid imputing implausible 6MWD values (i.e., 6MWD < 0 meters), the imputation of change from baseline will be capped at minus the baseline value so that any imputed 6MWD value will be ≥ 0.

[3] Excluding assessments occurring after EODBT + 7 days / administration of rescue therapy.

Note: if a participant has an assessment within the analysis visit window (see section 5.1.6), this assessment will be taken for the analysis irrespective of a later IE occurring after the in-window assessment and prior to the target day.

D. Population-level summary (estimator): difference in mean 6MWD change from baseline at Week 28 between macitentan 75 mg and placebo along with 95% CLs, and p-value as estimated from a mixed effect repeated measurement model (MMRM) based on the FAS.

5.3.3. Analysis Methods

5.3.3.1. Statistical Model

Hypotheses:

The null hypothesis (H_0) is that there is no difference between the macitentan 75 mg and placebo in mean changes from baseline in 6MWD at Week 28 or that the difference favors placebo. The alternative hypothesis (H_1) is that there is a difference between intervention groups in favor of macitentan 75 mg.

$$H_0: \delta = \mu_{MACI,28} - \mu_{PBO,28} \leq 0$$
$$H_1: \delta = \mu_{MACI,28} - \mu_{PBO,28} > 0$$

The null hypothesis will be tested by means of an MMRM. An MMRM model adjusted for covariates intrinsically handles missing data that are not study intervention-related i.e., missing at random (MAR).

Mixed effect repeated measurement model (MMRM):

A general MMRM (linear mixed effects model) is considered to evaluate response outcome as change from baseline (specified as change in the model) which is expected to be approximately normally distributed and measured at scheduled visits up to Week 28.

The model includes intervention (via an indicator variable for randomized study intervention), time (via visit [analysis visit]) as a categorical variable for visit, intervention-by-visit interaction, strata (one indicator variable for each randomization stratification factor i.e., use of PH-specific therapy [riociguat / other PH-specific therapy / none] and CTEPH diagnosis [inoperable with or without BPA vs persistent/recurrent after PEA surgery with or without BPA]) as fixed effects, and baseline (baseline 6MWD measures) as a covariate. An unstructured variance-covariance matrix for within-subject (participant) errors is used (type=un).

From this model, least squares means, treatment intervention differences in least squares means, standard errors and 95% confidence limits (CLs) will be estimated for each visit time point. Primary inference will be based on the study intervention comparison of least squares means at Visit Week 28 obtained from this model. Treatment effect at a specific time point is derived using contrasts and t-distribution.

If convergence issues arise from the above, a heterogeneous first order autoregressive covariance structure will be considered for the primary endpoint analysis.

5.3.3.2. Handling of Intercurrent Events and Missing Data

Handling of missing data arising from IEs & missing data due to other reasons are described in Table 12 below.

Any assessment post IEs will be handled by a composite strategy. For the primary endpoint analysis, these data are replaced by using imputation strategies described in Table 11 above.

Missing data arising from events that are not treatment-related (e.g., life events such as relocation, scheduling of transportation, inadequate reimbursement of associated expenses) are assumed to be missing at random (MAR), and as such, will require no further pre-processing (i.e., imputation), since the MMRM approach will implicitly impute these values. In Table 12 below, missing data arising from events that are not treatment-related are identified as "Other: Premature discontinuations of study intervention, not related to study intervention".

Table 12: Primary Analysis: Handling of IEs and Missing Data During DB Period up to Week 28

Event type (IE/Other) and time point	Strategy (method for handling IE / Other missing data)	Strategy applied from
IE: Deaths occurring **up to** EODBT +7 days **without** intake of rescue medication	Composite strategy[1] for deaths.	Date of death
IE: Deaths occurring **up to** EODBT+7 days **after** intake of rescue medication	Composite strategy[1] for treatment related intervention discontinuation / intake of rescue medication	Date of rescue medication administration
IE: Premature discontinuation of study intervention **possibly related** to study intervention[2] and /or intake of rescue medication	Composite strategy[1] for treatment related intervention discontinuation / intake of rescue medication	EODBT +7 days or date of rescue medication administration, whichever occurs first
Other: Premature discontinuations of study intervention, not related to study intervention are not considered as IEs.	All available 6MWD assessments up to Week 28 done after EODBT (including during PTOP) will be retained. MAR assumption: process of handling missing data is inherent to the MMRM model.	Not applicable, MAR imputation applied for missing data.
Other: 6MWD assessments not performed with reason given as "clinical worsening" or "PH related"	Composite strategy[1] for treatment related intervention discontinuation / intake of rescue medication	Applied to associated missing time point. Not applicable for repeat 6MWD to confirm clinical worsening

[1] Composite strategy is described in Table 12
[2] Possibly study intervention related as indicated in the "*treatment disposition (end of treatment)"and CEC-confirmed clinical worsening.*

Reasons for premature discontinuation of study intervention considered as possibly related to study intervention are described in Section 5.2.1.

5.3.4. Primary, Supplementary and Subgroup Analyses

Table 13 provides a summary of the main, supplementary and subgroup analyses which are planned to support the primary efficacy endpoint. All analyses will be performed based on the FAS.

Table 13: Summary of Planned Analyses Supporting Primary Endpoint

Analysis type	Analysis strategy	Section #
Primary estimand (main analysis)	MMRM (MAR)	Sections 5.3.2.1 and 5.3.4.1
Supplementary analysis	Treatment policy estimand with composite strategy ordered on survival time	Section 5.3.4.2
Subgroup analyses	Subgroup analyses	Section 5.3.4.3

5.3.4.1. Primary Endpoint Analysis (Main Analysis)

The main analysis refers to the primary endpoint analysis performed when all participants have completed the 28-week DB fixed duration period or prematurely discontinued the study beforehand.

For the main estimand defined in Section 5.3.2.1, the main analysis refers to the change from baseline to Week 28 in 6MWD, tested on the FAS at a 1-sided alpha = 0.025 level of significance using the MMRM (see Section 5.3.3.1). Missing data and IEs will be handled according to the rules defined in Section 5.3.3.2.

The analysis will be conducted irrespective of the stages (as opposed to the protocol planned inverse normal combination test, see section 5.7.2) and will be considered explorative. Rationale is that following the decision to premature stop the study for futility only limited data in stage 2 are available.

In addition, the treatment effect will be estimated at 4-weekly intervals from baseline to Week 28 based on the same model along with their associated 95% CLs.

A line plot for mean change from baseline (LS means ± SE) over time will be generated.

A listing of observed and imputed values and changes from baseline to Week 28 for 6MWD will be provided for the FAS by study intervention group. All available 6MWD data throughout the study (DB + OL) will be listed. Observed data after EODBT (PTOP)/imputed data will be flagged.

5.3.4.2. Supplementary Analysis

In support of the primary endpoint analysis (main analysis), a supplementary analysis using the treatment policy estimand with composite strategy ordered on survival time will be applied. The aim of this supplementary estimand is to apply a ‘treatment policy’ approach, i.e., include 6MWD results irrespective of treatment discontinuation or administration of rescue therapy. The intercurrent event death is handled with an alternative strategy, imputing 6MWD for participants with death as approximately 0m, ordering the imputed values as per time to death.

The estimand will be analyzed using an MMRM as in the primary estimand main analysis.

Estimand: the estimand components A and B are described in Section 5.3.2.1.

C. Intercurrent events (IE): IE and related strategies are summarized in Table 14 below.

Table 14: Intercurrent Events and Related Strategy for Treatment Policy Estimand

Intercurrent Event	Strategy
Deaths irrespective of treatment discontinuation or administration of a rescue CTEPH therapy (up to Week 28)	A composite strategy will be used for deaths occurring up to Week 28 irrespective of study intervention discontinuation or administration of rescue therapy. All scheduled 6MWD assessments following the date of death will be imputed with a value ranging from 0m to -1m, increasing with survival time (e.g., 0m for a participant surviving up to Week 28, -1m for a participant dying on Day 1) Derived as Imputed 6MWD W28 value = 0m – delta, where delta = 1 – time to death (weeks) / 28. Obtain the imputed W28 change score. Note: the small deviation from a 0m imputation applied is expected to have limited impact for a parametric analysis but for a non-parametric analysis allows for ranking based on survival time.
Premature discontinuation of study intervention possibly treatment related	Treatment policy strategy: observed 6MWD values after IE (i.e., during PTOP) are included
Administration of rescue CTEPH therapy	Treatment policy strategy: observed 6MWD values after rescue (i.e., during PTOP) are included
Handling of missing data	Any occurrence of missing data will be handled using the MMRM approach.

Population-level summary: Difference in mean 6MWD change from baseline at Week 28 between macitentan 75 mg and placebo along with 95% CLs and p-value as estimated from an MMRM based on FAS. A line plot for mean change from baseline (LS means ± SE) over time will also be generated.

5.3.4.3. Subgroup Analyses

In order to assess the consistency of the treatment effect across different participant subgroups for the primary efficacy variable, subgroup analyses will be performed according to the demographic and baseline disease characteristics at randomization defined in Table 17 on the FAS. Subgroup analyses will be carried out for the primary estimand (Section 5.3.2) and for the treatment policy estimand (Section 5.3.4.2) respectively. No multiplicity adjustment will be introduced; the subgroup analysis is descriptive in nature.

Estimand:

This follows the primary estimand described in Section 5.3.2 and for the treatment policy estimand in Section 5.3.4.2 respectively.

Analyses:

Within each of the subgroup level, a separate MMRM will be fitted (without including baseline 6MWD and randomization stratification factors) to estimate change from baseline to week 28 in 6MWD. In case of major imbalances in baseline covariates, additional models will be fitted to explore the treatment effect by including baseline covariates.

The intervention-by-subgroup interaction p-value is estimated for each subgroup variable from a separate model including intervention, subgroup, and intervention-by-subgroup interaction term(s).

The treatment effect within subgroups will primarily be displayed with their corresponding 95% CLs and presented in a forest plot. The forest plot will also include as a reference the 'overall' treatment intervention effect based on the primary efficacy endpoint analysis.

For CTEPH diagnosis and use of specific therapies at baseline subgroup analyses, only descriptive statistics will be provided.

Participants with undetermined subgroups due to missingness will not be included in the subgroup analyses.

5.4. Secondary Endpoint(s) Analysis

Following the decision to prematurely terminate the study due to futility at the interim analysis, the two key secondary endpoints of interest for the purposes of the abbreviated CSR SAP are time to first CEC confirmed clinical worsening up to EODBT, and improvement in WHO FC from baseline to Week 28. These analyses are considered exploratory in nature and will be performed on the FAS.

5.4.1. Time to First Clinical Worsening up to EODBT

5.4.1.1. Definition of Endpoint

Clinical worsening (CW) is defined as the occurrence of at least one of the following events:

- All-cause death**
- Heart and/or lung transplantation
- Unplanned[a] pulmonary hypertension (PH)-related hospitalization
- PH-related deterioration identified by at least one of the following:
  - Persistent[b] increase in World Health Organization functional class (WHO FC) that cannot be explained by another cause (e.g., viral infection)

[a] Not planned at study entry

[b] Confirmed by a second measurement performed on a different day within 2 weeks

  - Persistent[a] deterioration by at least 15% in exercise capacity, as measured by the 6MWD
  - New or worsening signs or symptoms of right heart failure
- Rescue pulmonary endarterectomy (PEA) and/or balloon pulmonary angioplasty (BPA) procedure due to worsening of PH.

**Including deaths occurring within 30 days of last dose of study intervention caused by an AE that occurs within 7 days of intake of last dose of DB study intervention.

CW is assessed at the timepoints indicated in Table 23 and at unscheduled visits in case of a suspected CW.

5.4.1.2. Estimand

A 'while on treatment' strategy is followed for time to first CW up to EODBT. An event will be considered 'while on treatment' if it occurs up to 7 days after study intervention discontinuation, where 7 days corresponds to the time window in which an effect can still be attributed to macitentan, based on its half-life.

The estimand is described according to the following attributes:

A. Population

Participants diagnosed with CTEPH meeting the study eligibility criteria.

B. Variable

C. Time (days) from randomization to first occurrence of CEC confirmed clinical worsening up to EODBT. Only events occurring up to 7 days after DB study intervention discontinuation and prior to start of OL intervention are considered.

Participants who do not experience any CEC confirmed clinical worsening will have their time to clinical worsening right censored at earliest among the following times and prior to first dose of OL extension intervention:

- Study intervention discontinuation + 7 days (i.e., EODBT +7 days)
- EODBT +7 days for DB intervention completers.

Note: censoring should occur prior to first dose of OL extension intervention.

D. Intercurrent events

Not applicable.

E. Population-level summary

Median time to CW (stratified log rank test p value) and Hazard ratio [95% CLs] of macitentan 75 mg versus placebo estimated from the Kaplan-Meier survival curves and the Cox proportional hazards model in the FAS.

5.4.1.3. Analysis Methods

Hypotheses:

The null hypothesis (H_{01}) is that there is no difference between the macitentan 75 mg and placebo survival distributions or that the difference favors placebo. The alternative hypothesis (H_{11}) is that there is a difference between intervention groups in favor of macitentan 75 mg.

Time to first clinical worsening (CEC confirmed) up to EODBT:

H_{01}: $S_{MACI}(t) \leq S_{PBO}(t)$ vs. H_{11}: $S_{MACI}(t) > S_{PBO}(t)$, $t \geq 0$
H_{01} is tested at α=0.025 (1-sided)

Main Analysis:

A stratified log-rank test will be performed on the FAS, stratifying by intervention group and randomization stratification factors. Over the double-blind treatment period at pre specified time points, the number of participants at risk, number with events and number censored per intervention group will be graphically displayed via Kaplan-Meier curves. The curves will be truncated when there are fewer than 10 participants at risk. The median time to event (95% CLs) including the number of participants in the FAS per intervention group as well as the p-value corresponding to stratified log rank test will be presented.

The stratified log-rank test is conducted with SAS Proc LIFETEST stratifying for the randomization stratification factors, and group (treatment intervention). Furthermore, the intervention effect will be estimated based on a Cox proportional hazards model adjusting for the randomization stratification factors (strata). SAS Proc PHREG is used to estimate the hazard ratio (HR) and the corresponding 95% CLs (using Wald based methods). The EXACT method will be used to handle ties. An investigation into the assumption of proportional hazards for treatment is performed informally using a plot of the complementary log-log of the survival against the log of time (for each intervention group). If the hazards are proportional, the lines should be approximately parallel.

Listings will be provided including a description of events with flags to indicate CEC confirmation status, most significant event, and occurrence of any events during PTOP, event and censoring times.

Additionally, listings of clinical events according to adjudicator and investigator opinion, respectively will be provided. All available CW data throughout the study (DB + OL) will be listed. A participant is considered to have had an investigator assessed CW event if any event is

documented on the "suspected clinical worsening event" eCRF, regardless of the CEC review. The date of the event corresponds to the earliest "date of onset of first component" documented on that form.

5.4.2. Improvement in WHO FC From Baseline to Week 28 (Yes/No)

The WHO Function Classification (see Protocol Appendix 9) which contains 4 levels of functioning, i.e., Class I (Participants from this class are not included in the study), Class II (slight limitations), Class III (marked limitations) and Class IV (most advanced limitations), is assessed at the timepoints indicated in the schedule of activities (Table 23). For the analysis purpose, these WHO FC class values are transformed to a scale with scores ranging from 1–4; where a score of 1 corresponds to WHO FC Class I and a score of 4 corresponds to WHO FC Class IV. The higher scores indicate greater symptom severity or worse impact.

Improvement (decrease) in WHO FC from baseline to Week 28 is calculated for each participant.

5.4.2.1. Estimand

The estimand is defined for consistency with the primary 6MWD estimand. IEs up to Week 28 are defined as those that occur up to study day 197.

A. Population: Male/female participants between 18 and 80 years of age with a diagnosis of CTEPH.

B. Variable: Proportion of participants with an improvement in WHO FC from baseline to Week 28.

C. Intercurrent events (IE) and imputation for missing data:

- **Deaths:**

  For deaths occurring up to 7 days after study intervention discontinuation (EODBT +7 days) without administration of rescue CTEPH therapy, all scheduled assessments following the date of death will be imputed by a 'worst-case' imputation (a score of 4 corresponding to WHO FC=IV) to calculate the change from baseline to post-baseline assessment timepoints up to Week 28.

- **Premature discontinuation (possibly intervention related)/Administration of rescue therapy:**

  Assessments performed after (possibly intervention related) discontinuation of study intervention + 7 days (EODBT +7 days) and/or administration of rescue CTEPH therapy will be excluded from the analysis. For each visit occurring after the IE, the sum of the participant's baseline WHO FC score and the 75th percentile of the observed WHO FC change score from baseline across both intervention groups combined and over all assessments up to Week 28 will be used for imputation.

- Missing values at Week 28 (for reasons other than described above) will be imputed by carrying forward the last assessment score before discontinuation of study intervention + 7 days.
- For non-treatment related discontinuations, WHO FC assessments observed post EODBT +7 will be excluded as this follows a while on treatment strategy.
- If imputation indicates an improvement, then baseline value will be imputed to Week 28.

D. Population-level summary: Odds ratio [95% CLs] of macitentan 75 mg versus placebo estimated from the Cochran Mantel Haenszel (CMH) χ^2 test (based on FAS).

5.4.2.2. Analysis Methods

Hypotheses

The null hypothesis (H_{02}) is that the proportion of participants who improve is the same for macitentan 75 mg and placebo or is higher for placebo. The alternative hypothesis (H_{12}) is that the proportion of participants who improve is higher for macitentan 75 mg.

Improvement in WHO FC to Week 28:

H_{02}: $P_{MACI,28} \leq P_{PBO,28}$ vs. H_{12}: $P_{MACI,28} > P_{PBO,28}$
H_{02} is tested at α=0.025 (1-sided)

Main Analysis:

Descriptive statistics on WHO FC response (n, %) at each visit up to Week 28 will be provided by intervention group. Shift tables from baseline to each visit will be provided for observed WHO _FC classes up to Week 28

WHO_FC response i.e., improved versus stable or worsened at Week 28 will be analyzed by using a Cochran-Mantel-Haenszel test adjusted by WHO FC at baseline, intervention group, and randomization stratification factors.

All available WHO FC data throughout the study (DB + OL) will be listed.

5.4.3. Percent of Baseline NT-proBNP Over time up to EODBT

NT-proBNP is assessed at timepoints indicated in Table 23. NT-proBNP is assumed to follow a log-normal distribution. Therefore, data are log transformed before analysis.

The percent of baseline NT-proBNP at post-baseline, i.e., at Week 28 is calculated as a ratio of the Week 28 NT-proBNP value over baseline value, expressed as a percentage, i.e., as 100 times the Week 28 value divided by the baseline value. Reduction in NT-proBNP (ratio < 100) is associated with clinical improvement.

Percent of baseline NT-proBNP over time up to EODBT will be summarized using descriptive statistics using observed data. A listing of observed NT-proBNP values over time will also be provided.

5.4.4. Hemodynamic Endpoints

Unless otherwise indicated, all analyses related to endpoints of the hemodynamic sub-study will be performed on the RHCS.

5.4.4.1. PVR Ratio

5.4.4.1.1. Definition of Endpoint

The ratio of Week 28 pulmonary vascular resistance (PVR) over baseline PVR is derived as follows, using PVR values in dyn·sec·cm^{-5} as derived in the eCRF:

$$PVR\ ratio = \left(\frac{PVR\ at\ Week\ 28}{PVR\ at\ Baseline}\right)$$

The Week 28 value is the one identified for the Week 28 analysis time window (see Table 10).

5.4.4.1.2. Analysis Methods

The PVR ratio is assumed to follow a log-normal distribution and so the analysis will be performed on the natural log of this ratio.

To evaluate the effect of macitentan 75 mg versus placebo on PVR at Week 28, an ANCOVA model will be applied for the $\log_e$-transformed PVR ratio, including randomized intervention group, and the $\log_e$-transformed baseline PVR value as covariates. Depending on the number of participants enrolled, the stratification factors, PH-specific therapy (riociguat/other PH-specific therapy/none) and inoperable (with or without BPA) vs persistent/recurrent after PEA surgery (with or without BPA) may also be included in the model.

The resulting least squares (LS) means, and 95% CLs obtained in each intervention group, and the LS-means differences (95% CLs) for macitentan 75 mg versus placebo will be inversely transformed using the exponential function and multiplied by 100 to provide:

- The adjusted geometric mean of the ratios of Week 28 to baseline PVR values and corresponding 95% CLs, expressed in percent, in each intervention group, and
- The adjusted geometric mean ratio and corresponding 95% CLs for macitentan 75 mg versus placebo

A geometric mean ratio ($GM_{maci75}/GM_{placebo}$) < 1 favours macitentan 75 mg.

Absolute values at baseline and at Week 28 as well as absolute changes from baseline to Week 28 in PVR will also be summarized using descriptive statistics.

Handling of Missing Data

Imputation methods for PVR will be specific to the reason for missing data.

In case of a missing PVR value at Week 28, the last available post-baseline value obtained before the Week 28 analysis time window is carried forward. No RHC assessment is performed in case of premature discontinuation. Therefore, it is assumed that this will occur in the rare case a participant had an assessment value prior to the Week 28 analysis time window. This imputation will be performed except in the following cases:

- If a participant dies prior to Week 28 (≤ Day 197 corresponding to target day of protocol defined window (Table 10) without a Week 28 value, then the missing PVR ratio is imputed with the highest observed individual PVR ratio value at Week 28 amongst all participants in the same analysis set across both intervention groups. The resulting imputed Week 28 PVR is the product of this imputed PVR ratio value and the respective baseline PVR value. Participants who die before the Week 28 assessment are likely to have experienced considerable deterioration prior to death. This imputation approach assigns the worst observed value across both intervention groups to such participants; hence presenting the potentially worst-case scenario.

All observed in-window (Section 5.1.6) Week 28 PVR values and imputed LOCF values will be used to derive the imputed values.

A listing of observed and imputed values will be provided for the RHCS by intervention group. Observed data after DB study intervention discontinuation and imputed data will be flagged.

5.4.4.2. Other Hemodynamic Endpoints

5.4.4.2.1. Definition of Endpoints

All hemodynamic variables collected in the hemodynamic sub-study are summarized in Appendix 6.15, Table 26.

Continuous endpoints for the hemodynamic sub-study are as follows:

- Change from baseline to Week 28 in:
    a. Mean right atrial pressure (mRAP)
    b. Mean pulmonary artery pressure (mPAP)
    c. Cardiac index (CI)
    d. Cardiac output (CO)
    e. Total pulmonary resistance (TPR)
    f. Mixed venous oxygen saturation (SvO_2)

5.4.4.2.2. Analysis Methods

Change from baseline to Week 28 is derived for all hemodynamic variables (Table 26 in Appendix 6.15) as follows:

$$Change\ from\ baseline = value\ at\ Week\ 28 - value\ at\ baseline$$

Descriptive summary statistics for absolute values at baseline and at Week 28 as well as absolute changes from baseline to Week 28 will be provided by intervention group for the continuous endpoints defined in Section 5.4.4.2.1. The analysis of all endpoints will consider participants with a Week 28 value within the window described in Section 5.1.6 only.

A participant listing of all hemodynamic variables collected or derived is also provided.

5.5. Safety Analyses

All safety analyses on the safety analysis set (SS) will be based on actual intervention received.

Safety analyses on OL set will be presented overall and by actual DB study intervention received.

For all continuous safety variables, descriptive statistics by intervention group will include the N, mean, standard deviation, median, range [minimum, and maximum]. Categorical variables will be summarized by intervention group using frequency counts and percentages.

The safety endpoints described in this section will be presented for the DB and OL treatment emergent period, as specified. Additionally, if applicable, selected safety events will be described separately for the up-titration and maintenance phases. The DB and OL treatment emergent period and phases are defined below.

Unless otherwise specified, listings will present all study data including the OL period.

DB Treatment Emergent Period

Unless otherwise specified, the DB treatment emergent period is defined from first intake of intervention in the DB period up to the start of OL treatment intervention (as participants should enter OL without interruption), or up to 30 days after the end of DB intervention (EODBT +30 days) for premature DB treatment discontinuations.

DB Treatment Emergent Phases

The following three phases are defined for both intervention groups:

- up-titration 10 mg
- up-titration 37.5 mg
- maintenance 75 mg.

Each phase is defined from first intake of respective intervention dose in DB period up to but excluding the start of the next phase intervention dose / OL treatment intervention if participants complete that particular titration phase, or up to 30 days after the end of DB intervention dose if they prematurely discontinue DB intervention, respectively.

OL Treatment Emergent Period

The OL treatment emergent period is defined from first intake of intervention in the OL period up to 30 days after the end of OL intervention.

5.5.1. Extent of Exposure

5.5.1.1. Extent of Exposure during the DB period

Extent of exposure will be summarized for the DB treatment period and for the DB treatment phases (see Section 5.1.3) by study intervention on the SS set.

The number and percentage of participants who received any DB study intervention will be summarized for each DB treatment phase: up-titration (10 mg and 37.5 mg) and maintenance (75 mg).

Descriptive statistics for duration of study intervention (N, mean, SD, median, and range [minimum, maximum]) will be summarized.

Participant-Years (PY) of intervention exposure is calculated by summing up each individual participant's exposure to study intervention (in days), irrespective of interruptions or dose titrations, and dividing it by 365.25. For n participants, PY is derived using:

$$\frac{1}{365.25} * \sum_{1}^{n} individual\ subject\ treatment\ exposure\ duration$$

PY will be presented by intervention group.

Cumulative duration of study treatment intervention will be summarized in the following duration categories:

≥4 weeks, ≥8 weeks, ≥12 weeks, ≥16 weeks, ≥20 weeks, ≥24 weeks, ≥28 weeks, ≥40 weeks, ≥52 weeks, up to study EODBT period (every 12 weeks).

Study intervention duration during DB (in weeks) is defined as [(date of last dose of study intervention in DB – date of first dose of study intervention) +1] / 7.

The total dosing weeks of intervention (defined as the total number of days that study intervention was administered to the participant (excluding days off study intervention)/7) will be also summarized.

The number (%) of participants with a dose not administered (interrupted) will be summarized by intervention group and DB treatment phases. Reasons for doses not administered will also be summarized by DB treatment phases.

Maximum dose titration received by an individual participant (n, %) will be summarized. Note a participant is counted only in the highest dose category:

- 10 mg
- 37.5 mg
- 75 mg

The number (%) of participants with a dose not administered (interrupted) will be summarized by intervention group and DB treatment phases. Reasons for doses not administered will also be summarized by DB treatment phases. Participants with a dose not administered (interrupted) will be listed together with the duration of the interruption, last dose taken, and reason for interruption.

Participant listings of DB period exposure data will be provided on the SS.

Study intervention compliance will be summarized descriptively. See Section 6.8 Appendix 8 for further details.

5.5.1.2. Extent of Exposure during the OL period

Extent of exposure will be summarized for the OL treatment period overall and by DB study intervention group (received) on the OL set.

The number and percentage of participants who received any OL study intervention will be summarized.

Descriptive statistics for duration of study intervention (N, mean, SD, median, and range [minimum, maximum]) will be provided.

Study intervention duration during OL (in weeks) is defined as [(date of last dose of OL study intervention – date of first dose of OL study intervention) +1] / 7.

Participant listings of OL period exposure data will be provided on the OL set.

5.5.2. Adverse Events during DB period

The verbatim terms used in the CRF by investigators to identify adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA).

Any AE occurring at or after the initial administration of DB study intervention through up to the day of last dose in the double-blind plus 30 days (EODBT + 30) and prior to start of OL study intervention (as participants transition without interruption in the OL study) is considered to be treatment emergent. If the event date is recorded as partial or completely missing, then the event will be considered to be treatment emergent unless it is known to be prior to the first administration of study intervention based on partial onset date or resolution date. All reported treatment emergent AEs (TEAEs) will be included in the analysis. For each adverse event, the number and percentage of participants who experience at least one occurrence of the given event will be presented by

MedDRA system organ class (SOC) and preferred term (PT) by descending frequency in the macitentan arm.

Summary tables will be provided for TEAEs by intervention group for:

- AEs
- Serious AEs (SAEs)
- AEs leading to discontinuation of study intervention
- AEs with fatal outcome
- AEs by worst severity
- AEs related to study intervention
- SAEs related to study intervention
- AEs leading to dose interruptions of study intervention.

In addition, TEAEs, TESAEs and AEs leading to discontinuation of study intervention will be also summarized by preferred term only.

An overview summary of the number and percentage of participants experiencing at least one TEAE during the DB period in each of the above categories will also be provided, including the number of COVID-19 associated AEs/SAEs.

Summary of AEs by preferred term and by frequency of at least 5% will also be presented. In addition to the summary tables, listings will be provided for participants who had an AE, as well as separate listings for participants who had SAEs, AEs leading to discontinuation of study intervention, AEs with fatal outcome and COVID-19 associated AEs. Listings will include the dose level at the time associated with each TEAE (up-titration 10 mg & 37.5 mg, and maintenance 75 mg) for both intervention groups.

A listing of AEs for not treated participants will be also provided on the SCR set.

The overview summary table and summaries of AEs, SAEs, AEs leading to discontinuation of study intervention, AEs with fatal outcome, AE leading to dose interruption will be provided for AEs emergent during the DB treatment phases (up-titration 10 mg and 37.5 mg, and maintenance 75 mg, see Section 5.1.3). AEs with a start date in a given phase will be counted in that phase only.

5.5.2.1. Adverse Events of Special Interest (AESI)

AESI include hypotension, edema/fluid retention, hemoglobin decrease/anemia and liver events. See further details on AESI in Section 6.9 Appendix 6.

The number and percentage of participants experiencing at least one TE AESI will be presented by descending PT frequency for each AESI category. Within each category will be also displayed (by descending PT frequency):

- Number and % of participants with at least one event leading to permanent study intervention discontinuation
- Number and % of participants with at least one event leading to interruption of study intervention
- Number and % of participants with at least one serious event
- Number and % of participants with at least one fatal event
- Number and % of participants with at least one severe event

In addition, overview of AESI tables will be provided by AESI category and will include:

- Number and % of participants with at least one AESI
  - Relative risk with 95% CLs for macitentan 75 mg vs placebo
- Number and % of participants with at least one event within the worst severity category where following hierarchical order is considered to determined worst severity: Severe, Moderate, Mild, Missing
- Number and % of participants with at least one recovered/resolved event
- Number and % of participants with at least one event that recovered/resolved with sequelae
- Number and % of participants with at least one event that did not recover/did not resolve
- Number and % of participants with at least one event that is recovering/resolving
- Number and % of participants with at least one event with unknown outcome
- Number and % of participants with at least one fatal event

For recurrent events:

- Cumulative number of recurrent TEAESI per AESI category (note: this includes the total number of events as participants may have multiple PTs under an AESI, and multiple events in a given PT)

Participant-Years (PY) of intervention exposure is calculated as indicated in Section 5.5.1.1.5.5.1

  - Average annualized event rate (AER)

    where AER (per 100-PY) $= \left(\frac{Cumulative\ number\ of\ recurrent\ events}{PY\ intervention\ exposure} x100\right)$

  - Number and % participants with 1, 2, 3 or more events

Separate listings for each AESI category will be provided which will include PT terms, treatment emergent status, outcome, seriousness, severity, relationship to study treatment, event start and stop dates, intervention dose at onset, relative days to onset since the first day of study intervention.

TE AESI emerging during the different DB treatment phases (up-titration 10 mg and 37.5 mg and maintenance 75 mg) will be summarized by intervention group. AEs with a start date in a given phase will be counted in that phase only.

Time of first TE AESI will be derived for each participant and each AESI category. Participants who did not have such AEs will be censored at earliest of 30 days after the last dose in DB period (EODBT + 30 days) / day prior to first dose of OL intervention.

For each of the AESI categories, the following graphs will be displayed:

- KM curves by intervention group
- HR (macitentan 75 mg vs placebo) with 95% CLs estimated using a Cox model will be presented in the KM plots.
- Recurrent AESI will be presented using mean cumulative function (MCF) graphs by intervention group.

5.5.2.2. Potential Risks

Potential risks include the following categories:

- Menstrual disorders/vaginal hemorrhage, (excluding males)
- Ovarian cyst (excluding males)
- Testicular disorders and male infertility (excluding females)
- PVOD with event of pulmonary edema.

See further details on macitentan potential risks in Section 6.9 Appendix 9.

TEAEs under macitentan potential risks will be analyzed, and summary tables presented in the same way as described for TE AESI above for the overall DB period. A separate listing for all AEPR will be provided.

5.5.2.3. Deaths

Deaths will be displayed on the safety analysis set.

- A summary will be provided by study intervention arm for all deaths during the DB treatment emergent period and all deaths in the DB study period (including those occurring during PTOP up to start of OL treatment) showing number of participants who died
- Cause of death

The summary will be based on the Death Information eCRF page.

A listing of all participants who died (including those during the OL) will be provided including the study day of death in relation to Study Day 1, days off treatment and study period.

5.5.3. Adverse Events during OL period

Any AE occurring at or after the initial administration of OL study intervention through up to the day of last dose in the open label extension period plus 30 days is considered to be treatment emergent in the OL period.

For AEs and SAEs, the number and percentage of participants who experience at least one occurrence of the given event will be presented by MedDRA system organ class (SOC) and preferred term (PT) by descending frequency overall.

An overview summary table of AE will be provided for AEs emergent during the OL treatment period.

A listing will be provided for participants in the OL analysis set who had an AE.

5.5.4. Additional Safety Assessments

5.5.4.1. Clinical Laboratory Tests

Clinical laboratory tests (as described below) will be analyzed for the participants included in the safety or OL analysis set as further specified.

List of parameters:

Hematology:

- Platelet count
- Red blood cell (RBC) count
- Hemoglobin
- Hematocrit
- INR
- Prothrombin time

RBC Indices:

- MCV
- MCH
- % Reticulocytes

White Blood Cell (WBC) count with Differential:

- Neutrophils
- Lymphocytes
- Monocytes
- Eosinophils
- Basophils

Clinical chemistry:

- Sodium

Potassium
Chloride
Bicarbonate
Blood urea nitrogen (BUN)
Creatinine
Glucose [nonfasting]
Aspartate aminotransferase (AST)
Alanine aminotransferase (ALT)
Total and direct bilirubin
Alkaline phosphatase
Uric acid
Calcium
Phosphate
Albumin
Total protein
Magnesium
Thyroid stimulating hormone (TSH)
Iron
Ferritin
Glomerular filtration rate (GFR) from creatinine adjusted for body surface area (BSA)

Descriptive statistics (n, mean, SD median and range [minimum, maximum]) will be presented for all chemistry and hematology laboratory tests data provided by the central laboratory at scheduled time points (see visit windows in Table 8). Summaries will include baseline values, observed values and changes from baseline values to each timepoint and displayed by intervention group.

Laboratory toxicity grades will be derived from the Common Terminology Criteria for Adverse Events (CTCAE) v5.0 (see Section 6.11 Appendix 11) using central and local laboratory data.

Marked abnormalities for hematology and liver aminotransferases abnormalities based on normal ranges are defined below.

Abnormality criteria based on toxicity grades and normal ranges/ and or criteria will be applied to baseline and post-baseline values. All central and local laboratory data will be used to evaluate abnormalities.

- Post-baseline toxicity grades will be compared with their corresponding baseline toxicity grades and treatment emergent (TE) will be concluded if the post-baseline toxicity grade is worse than the baseline toxicity grade.
- For abnormalities based on normal range and/or criteria: If the post-baseline value is above the upper limit and the baseline value is below the upper limit (e.g., Normal or Low), then the post-baseline abnormality will be considered TE. The same applies to the post-baseline value being below the lower limit with the baseline value being above the lower limit (e.g., Normal or High).
- If the baseline value is missing, a post-baseline abnormality will always be considered as TE.

A flag for the worst toxicity grade during the DB treatment emergent period will be created. The worst TE toxicity grade during DB treatment emergent period will be summarized by intervention group for all hematology and chemistry tests with CTCAE. Shift summaries from baseline laboratory toxicity grade to the worst toxicity grade during DB treatment emergent period will be presented by intervention group for all chemistry and hematology tests with CTCAE.

Markedly abnormal hematology and liver aminotransferases abnormalities:

Hematology abnormalities:

- Hemoglobin < 80 g/L
- Hemoglobin ≥ 80 g/L and < 100 g/L
- Hemoglobin < 100 g/L and a decrease from baseline >20 g/L
- A decrease in hemoglobin from baseline of >20 g/L
- A decrease in hemoglobin from baseline of >20 g/L and ≤ 50 g/L
- A decrease in hemoglobin from baseline of >50 g/L.

Liver aminotransferases abnormalities (i.e., alanine aminotransferase [ALT] and/or aspartate aminotransferase [AST]):

- ALT and / or AST ≥3xULN
- ALT and / or AST ≥5xULN
- ALT and / or AST ≥8xULN
- ALT and / or AST ≥3 and <5 × ULN
- ALT and / or AST ≥5 and <8 × ULN
- ALT and/or AST ≥3xULN and total bilirubin ≥2xULN on same date
- ALT and/or AST ≥3xULN and total bilirubin ≥2xULN at any time during the DB/OL treatment emergent period

A value is a marked abnormality if it meets any of the criteria described above.

The number and percentage of participants with TE markedly abnormal liver aminotransferase and hemoglobin abnormalities values at any time during the DB and OL treatment emergent period will be presented. For both periods, the relative baseline is considered.

An evaluation of drug-induced serious hepatotoxicity (e-DISH) plot of the maximum treatment-emergent ALT (in multiples of ULN) by maximum treatment-emergent bilirubin (in multiples of ULN) will be presented to detect possible cases of drug-induced liver toxicity in the DB period. Reference lines of 3×ULN for ALT and 2×ULN for bilirubin will be drawn and number and percentage of participants in each quadrant will be displayed. Plot will be repeated replacing ALT with AST.

Data listings will be prepared separately for hematology and biochemistry. All laboratory data will be listed including CTCAE grade together with worst TE toxicity grade flag, dose level at onset of the TE abnormality with corresponding relative days from start of intervention. Pregnancy test data will be listed for positive tests only. A separate listing of all laboratory results will be provided for all participants in the OLAS.

In addition, separate listings of hemoglobin and LFT abnormalities during the DB period will be provided. The listings will include all participants with any marked abnormality together with investigator reported description of associated AEs, concomitant medications, and medical history. The listing of LFT abnormalities will also include the ALT/AST and bilirubin values in multiples of ULN corresponding to the e-DISH plot.

5.5.4.2. Vital Signs and Physical Examination Findings

The vital signs and physical examination assessed are described in protocol Section 8.3.1 and 8.3.2, respectively. The schedule of assessment is shown in Table 23. Rules for windowing analysis visits are described in Table 8.

Descriptive statistics of vital sign parameters including pulse rate, blood pressure (systolic and diastolic), and weight, and the changes from baseline will be summarized at each time point for baseline values and all values measured during the DB period.

Blood pressure and pulse are measured in triplicate, and the mean of the triplicate measures calculated in the EDC will be used in all respective analyses.

Body mass index (BMI) will be calculated in the EDC as weight (kg)/(height (m))^2. Descriptive statistics (mean, standard deviation, median, range [minimum and maximum]) for BMI and height will be presented for the baseline values only (among baseline characteristics) The weight and height measurements collected at screening will be used in the calculation of BMI at baseline.

Abnormality criteria (as defined in Table 15 below) will be applied to post-baseline values. Post-baseline values will be considered TE if they meet both value and change criteria in the table below and if the abnormality occurs during the DB treatment emergent period.

Incidence of treatment emergent (TE) markedly abnormal vital signs during DB treatment emergent period will be summarized for participants who had a baseline assessment and at least 1 post-baseline assessment for that vital sign in the period. A listing of participants with treatment emergent markedly abnormal vital signs during the DB period will be presented, along with a listing of all vital signs measurements including BMI, height, weight pulse rate and blood pressure (systolic and diastolic) at all visits. Only the mean values of triplicate measurements for pulse rate and blood pressure will be provided, as calculated in the EDC.

Table 15: Markedly Abnormal Vital Signs

Vital Sign	Criteria
Pulse	>120 bpm and with >30 bpm increase from baseline
	<50 bpm and with >20 bpm decrease from baseline
Systolic blood pressure	>180 mm Hg and with >40 mm Hg increase from baseline
	<90 mm Hg and with >30 mm Hg decrease from baseline
Diastolic blood pressure	>105 mm Hg and with >30 mm Hg increase from baseline
	<50 mm Hg and with >20 mm Hg decrease from baseline

Any physical examination abnormalities worsening post-baseline are reported as AEs. No summary or listing of physical examination findings is provided.

5.5.4.3. Electrocardiogram

A single standard 12-lead ECG is performed at timepoints indicated in the schedule of assessments as shown in Table 23 and results are interpreted by a central reader (eRT). Visit windows for 12-lead ECG are described in Table 9. ECG will be summarized by intervention group.

The ECG parameters that will be analyzed are heart rate, PR interval, RR interval, QRS interval, QT interval, and corrected QT (QTc) interval using the following correction methods: Bazett's formula (QTcB), Fridericia's formula (QTcF).

Descriptive statistics (mean, standard deviation, median, range [minimum and maximum]) will be presented at baseline and for observed values and changes from baseline at each timepoint during the DB period.

Criteria for ECG abnormalities are defined in Table 16 below.

The number and percentage of participants with QTc interval increases from baseline to the maximum post-baseline value during the DB treatment emergent period will be summarized.

Table 16: Criteria for Abnormal QTc Values and Changes from Baseline for QTcB and QTcF

QTc Abnormality criteria	QTc Categorization	Marked abnormality criteria flag
QTc value (msec)	≤450	
	>450 – 480	H
	>480 – 500	HH
	>500	HHH
QTc change from baseline (msec)	≤30	
	>30 – ≤60	HH
	> 60	HHH
QTc value > 450 msec and QTc interval increase from baseline > 30 msec		

A shift table will be provided summarizing the shift from baseline to maximum post-baseline QTc interval classification, including all measurements falling in the DB treatment emergent period.

If ECG measurements are repeated at a visit on same day, they will be averaged. The averaged value will be considered the 'Visit' ECG result and will be used for continuous measures summaries by time point. For abnormalities defined above both are considered.

Abnormality criteria will be applied to baseline and post-baseline values falling in the DB treatment emergent period.

Post-baseline abnormalities will be compared with their corresponding baseline result:

- TE will be concluded if the post-baseline value is above the upper limit and the baseline value is below the upper limit (e.g., Normal or Low). The same applies to the post-baseline value being below the lower limit with the baseline value being above the lower limit (e.g., Normal or High).
- If the baseline value is missing, a post-baseline abnormality will always be considered as TE.

The number and percentage of participants with DB treatment emergent ECG values outside predefined limits/abnormal post-baseline values (relative to baseline) will be presented by intervention group over the DB treatment emergent period of the study:

- Heart rate (bpm): <50 and >100
- PR interval (msec): <120 and >200
- RR interval (msec): <600 and >1200
- QRS interval (msec): >120

An abnormality/finding is considered TE if it was not present at baseline. If the baseline value is missing, a postbaseline abnormality is considered as TE. If the baseline value is missing, a postbaseline abnormality is considered as TE.

TE findings during will be also summarized.

Listings will be produced for all ECG data including unscheduled visit data. A listing of ECG abnormalities and findings during the DB period will also be provided.

5.6. Other Analyses

5.6.1. Pharmacokinetics

All participants will undergo sparse PK sampling, performed pre-dose and 2-10 hours post dose at timepoints indicated in the schedule of assessments as shown in Table 23.

Descriptive statistics (N, mean, geometric mean, SD, median, range, CV (%), IQ range) will be used to summarize macitentan / ACT-132577 concentrations at each sampling time point.

Macitentan /ACT-132577 concentrations below the LLOQ will be imputed as zero in the summary statistics. Furthermore, a steady-state flag will be created to capture either no change in dose (flag=yes) or dose interrupted (flag=no) occurring within the 7-day window before PK sampling.

5.6.2. Pharmacodynamics

ET-1 samples are collected at timepoints indicated in the schedule of assessments as shown in Table 23.

Descriptive statistics (N, mean, SD, median, range, CV (%) and IQ range) will be used to summarize ET-1 concentrations at each sampling time point by study intervention.

5.6.3. Definition of Subgroups

Table 17: Definition of Subgroups

Subgroup	Definition
Region 1[!]	North America and Australia: Canada, United States and Australia Latin America: Argentina, Colombia and Mexico Eastern Europe and Russia*: Bulgaria, Czech Republic, Hungary, Lithuania, Poland, Romania, Serbia, Slovakia, Turkey, Ukraine, Russia Western Europe, Israel and Saudi Arabia *: Austria, Denmark, France, Germany, Italy, Portugal, Spain, United Kingdom, Israel and Saudi Arabia Asia: China, Japan, South Korea, Singapore, Taiwan and Thailand
Region 2	• US • Non-US
Age Group	• 18-64 • ≥65-75 • >75
Sex	• Male • Female
Race 1	• White • Asian (includes Japanese and non-Japanese) • Black or African American* • Others (includes any other race and multiple race)
Race 2	• White • Asian (includes Japanese and non-Japanese) • Others (includes Black or African American, and any other race including multiple race)
CTEPH Diagnosis (randomization)	• Inoperable (with or without BPA) • Persistent/recurrent after PEA surgery (with or without BPA)
Use of PH specific therapies at baseline (randomization)	• Riociguat • Other PH specific therapy • None
CTEPH classification at Baseline**	• Inoperable with BPA

Subgroup	Definition
	• Inoperable without BPA • Persistent/recurrent after PEA (with BPA) • Persistent/recurrent after PEA (without BPA)
WHO FC at Baseline 1	• I/II • III/IV
WHO FC at Baseline 2	• I • II • III • IV*
mPAP (mmHg)	• <25 • ≥25

* If there are more than 20 participants in a subgroup then the subgroups will remain as defined. Otherwise combine categories (e.g., "Black or African American" with the race category "Others" and/or combine Eastern and Western Europe to create one category "Europe".

** As captured on Sub-etiology eCRF

! SAP template guidance: For the region subgroup, Variant 1 maps country to the UN defined geographic region as per the M49 standard (https://unstats.un.org/unsd/methodology/m49). If region is not defined based on the UN guidance, then the user must provide the study-specific definition of region.

5.7. Interim Analyses

5.7.1. Futility and Sample Size Re-estimation at the Interim Analysis

This study implements a two-stage, group sequential adaptive design with one IA. The "adaptive" aspect of this study focusses either on the early termination of the study for futility or performing an unblinded sample size re-estimation (SSRE), as discussed below.

Decision rules will be applied based on the primary endpoint variable change from baseline in 6MWD.

5.7.1.1. Stage 1

An IA will be planned when the first 72 sequentially randomized participants have completed Week 28 or prematurely discontinued the study. Because this study plans to randomize a total of 144 participants, and complete primary endpoint information will be available at Week 28 the amount of information i.e., information fraction (IF_1) at the time of the interim analysis is 0.5.

$$IF_1 = {}^{72}/_{144} = 0.50$$

In addition to the first 72 sequentially randomized participants, any randomized participant with a partial 6MWD profile, i.e., ongoing participants on maintenance dose (details below), even though not having yet completed Week 28 at the time of IA will be considered in the IA to provide more precision to SSRE.

Participants with a partial profile include any additional sequentially randomized participants who have (or were expected to have) conducted their 6MWT at Week 12, which is the first scheduled 6MWT on maintenance dose. They may or may not have received maintenance dose or have an assessment done on or post Week 12. The last sequentially randomized participant to include in the IA is determined as being randomized up to 20 February 2023, i.e., at least 12 weeks prior to

the cut-off date (the latest date when all 72 sequentially randomized participants have either completed Week 28 or prematurely discontinued prior to Week 28 taking into account the analysis window). Subsequent randomized participants will be excluded from the IA.

This total number of sequentially randomized participants included in the IA is denoted by n_1 and forms the stage 1 of the adaptive design.

At the IA, an MMRM analysis (described in Section 5.3.3.1) for the primary endpoint will be conducted on all available 6MWD data up to Week 28 from these n_1 participants. The test statistic for the treatment contrast (i.e., Z_1) from the MMRM analysis will be used to determine futility as well as to conduct the sample size calculation as detailed in Section 5.7.1.4.

The weights in the combination test for the final analysis are defined based on the information fraction $IF_1 = 0.5$ for participants with complete profiles. Hence, the weights $w_1 = 0.5$ is used for the calculation of conditional power (CP) (see Section 5.7.1.4).

5.7.1.2. Stage 2

All participants who did not provide data for the IA will be considered in Stage 2, irrespective of whether they were enrolled before or after the IA.

At the end of Stage 2, a combination test using the inverse normal method will be applied to control for the possible type I error inflation due to performing an unblinded SSRE by combining two test statistics Z_1 and Z_2 as described in Section 5.7.2 using pre specified weights. Z_1 for the final analysis is calculated from all accumulated data, including those collected after the IA from n_1 participants who contributed to the IA in Stage 1. Z_2 is calculated from the rest of n_2 or n_{2new} participants' accumulated data in Stage 2.

5.7.1.3. Futility Criteria

At the end of the first stage when the IA is performed, the study may be prematurely stopped for futility if the treatment effect (δ) is in favor of placebo, i.e., if the test statistic for the treatment contrast Z_1 from the MMRM analysis at the IA (see Section 5.7.1.1) is < 0.

5.7.1.4. Sample Size Re-estimation (SSRE) Criteria

At the IA, if the study is not stopped for futility, the sample size for the second stage of the study will be adjusted (up to a maximum of 230 participants) if the conditional power (CP) is below 80%.

The IA is planned when the first 72 sequentially randomized participants have completed Week 28 or prematurely discontinued the study. The planned information fraction $IF_1 = 0.5$, and a fixed weight $w_1 = 0.5$ will be used for CP calculation, irrespective of the actual number of participants n_1 in Stage 1.

The CP will be calculated based on the observed treatment effect estimate (δ) at IA by using the below formula:

$$CP = 1 - \phi\left(\frac{Z_\alpha - Z_1\sqrt{w_1}}{\sqrt{1-w_1}} - \frac{Z_1\sqrt{n_2}}{\sqrt{n_1}}\right)$$

Z_α: the $1-\alpha$ standard normal quantile where $\alpha = 0.025$

Z_1: test statistic for Stage 1 from MMRM, see Section 5.7.1.1.

n_1: number of participants in Stage 1

n_2: initially planned number of participants in Stage 2, $144 - n_1$

$w_1 = 0.5$

ϕ: Standard normal cumulative density function

SSRE will not be performed if the calculated CP is ≥ 80%. Then, the initially planned number of participants (i.e., $n_2 = 144 - \mathrm{n}_1$) will be randomized in stage 2. Otherwise, the required number of randomized participants in the second stage will be determined as the minimum between:

- $\mathrm{n}_{2\mathrm{new}} = 230 - \mathrm{n}_1$

  and

- $\mathrm{n}_{2\mathrm{new}} = \left[\frac{\mathrm{n}_1}{\mathrm{z}_1^2}\right]\left[\frac{\mathrm{z}_\alpha - \mathrm{z}_1\sqrt{\mathrm{w}_1}}{\sqrt{1-\mathrm{w}_1}} + \mathrm{z}_\beta\right]^2$ with z_β the 0.8 standard normal quantile

so that the overall maximum sample size will be 230.

5.7.2. Combination Testing for Control of Type I Error Inflation

To account for the inflation of type I error due to performing an unblinded SSRE, a combination test using the inverse normal method, i.e., the weighted Z-test will be used for each comparison of interest (i.e., for the primary and secondary endpoints) at the final analysis. The combination test statistics Z_C will be defined by an approach which defines the test statistic as a weighted sum of the test statistics Z_1 and Z_2 where $Z_1 = \Phi^{-1}(1-p_1)$ and $Z_2 = \Phi^{-1}(1-p_2)$ denote the Z-values corresponding to the one-sided stage-wise p-values p_1 and p_2, respectively. For each of the hypothesis of interest, Z_1 and Z_2 are obtained by repeating analyses using Stage 1 and Stage 2 participants separately.

Using the pre-specified combination test weights w_1 and w_2,

$$Z_C = w1 \cdot Z_1 + w2 \cdot Z_2$$

For the combination test the two stages will be weighted equally, i.e., $w_1 = \sqrt{0.5}$ and $w_2 = \sqrt{0.5}$ as shown below:

$$Z_C = \sqrt{0.5} \times Z_1 + \sqrt{0.5} \times Z_2$$

Resulting Z_C will be compared to the standard critical value corresponding to the type I error rate of 0.025 for the primary endpoint analysis as no type I error rate will be spent for the option of an early efficacy stop.

Similarly, conditional on rejecting the null hypothesis for the primary efficacy endpoint, resulting p values corresponding to Z_C obtained from each of the secondary endpoint analysis will be formally evaluated according to the testing hierarchy following the order of endpoints defined in the study protocol.

For the secondary endpoint time to clinical worsening main analysis, a participant wise combination of the p-values from the log-rank test statistics will be applied utilizing the entire follow-up time of the participants, including all clinical worsening events in that time frame. It is recognized that such a participant-wise combination for time to event type endpoints in designs with adaptions based on short term endpoints may result in a type I error inflation in specific situations (see Magirr 2016). However, in a simulation study provided in Section 6.14 Appendix 14, type I error control was demonstrated with this approach in a stress test type scenario relevant for this study (e.g., adaption based on short-term primary endpoint that is tested earlier in the hierarchical testing strategy) between primary and secondary endpoint.

As a sensitivity analysis, an analysis restricting follow-up time for the participants from stage 1 to an upper limit that is independent from the adaption is applied. The p-value for participants from stage 1, Z_1, will be derived based on follow-up time being restricted to the randomization date of the 144th participant plus 28 weeks. This represents the minimum follow-up time for stage 1 patients irrespective of the unblinded sample size re-assessment outcome at the interim analysis. Restricting the follow-up time of stage 1 participants controls the type I error rate irrespective of the specific study scenario (see Jörgens et al. 2019).

5.7.2.1. The Median-Unbiased Point and Interval Estimates for the Primary and Secondary Endpoints

The median-unbiased estimate of the treatment effect δ will be used for point estimation (Brannath 2006).

$$\hat{\delta} = \frac{\frac{w_1}{\mathrm{SE}(\hat{\delta}_1)}\hat{\delta}_1 + \frac{w_2}{\mathrm{SE}(\hat{\delta}_2)}\hat{\delta}_2}{\frac{w_1}{\mathrm{SE}(\hat{\delta}_1)} + \frac{w_2}{\mathrm{SE}(\hat{\delta}_2)}}$$

where $\hat{\delta}_1$ and $\hat{\delta}_2$ denote the estimated treatment effects at Stage 1 and Stage 2 and $\mathrm{SE}(\hat{\delta}_1)$ and $\mathrm{SE}(\hat{\delta}_2)$ are the standard errors of the estimates obtained from Stage 1 and Stage 2, respectively. $W_1 = \sqrt{0.5}$ and $w_2 = \sqrt{0.5}$ are the pre-specified combination test weights.

The two-sided flexible confidence interval at level *(1-2α)* will be symmetric about the median-unbiased point estimator. To have a proper coverage probability in the design with SSRE, the flexible confidence interval will be based on the duality of confidence intervals and hypothesis

tests. That is, for the parameter δ of interest, the one-sided flexible interval with a coverage probability 1- α is defined by the set $\{\delta : Z_C<z_\alpha\}$. For analysis of the primary and key secondary endpoints, the two-sided confidence interval for the treatment effect at level (1-2α) will be defined by

$$\left[\hat{\delta} - \frac{z_\alpha}{\frac{w_1}{\mathrm{SE}(\hat{\delta}_1)} + \frac{w_2}{\mathrm{SE}(\hat{\delta}_2)}}, \quad \hat{\delta} + \frac{z_\alpha}{\frac{w_1}{\mathrm{SE}(\hat{\delta}_1)} + \frac{w_2}{\mathrm{SE}(\hat{\delta}_2)}}\right]$$

where z_α denotes the $100(1-\alpha)^{th}$ percentile of the standard normal distribution.

For each of primary and secondary endpoints, the median-unbiased estimate of the treatment effect and the corresponding confidence interval will be provided.

5.7.3. Determination of the Operational Characteristics via Simulation

A modelling and simulation report (MSR) is provided in Section 6.14 (Appendix 14) where design characteristics used for the simulation are specified in the report Table 1 and the longitudinal profiles under consideration are given in Figure 1. Report Tables 2 and 3 contain alternative and null hypothesis scenarios, respectively with variance-covariance matrix (Table 4). Two separate sets of simulations were performed to obtain the operational characteristics of the primary analysis as described below.

1. Simulation for the primary endpoint only, using MMRM; these were performed in SAS (referenced as “longitudinal”)
2. Simulation for the primary and secondary endpoints, using an ANOVA approach for the primary endpoint; these were performed in R (referenced as “ANOVA”). This simplified approach to the primary analysis was chosen for reasons of simulation speed.

The power, futility stops and ASN for null and alternative scenarios are given in the report Tables 7 and 8, respectively. Results of interim point estimation are given in Table 10 (Appendix 14) whereas Table 11 (Appendix 14) shows the performance of the median unbiased estimator and the corresponding 95% confidence interval for the ANOVA case only.

In summary, simulations performed indicate that both approaches considered, i.e., MMRM and ANOVA, have the desired statistical inference properties. As there is no downside against the ANOVA approach, and the MMRM is expected to provide better performance in situations with missing data, the MMRM is recommended to be used as the primary analysis model.

The rejection probabilities for key secondary endpoint TTCW are provided in the modelling and simulation report provided in Section 6.14 Appendix 14, Section 3.5.3 where assumptions for Hazard ratio and baseline median TTCW are given in Table 6. Even though the power for TTCW is highly dependent on the performance of the primary endpoint, there seems to be sufficient power for an HR of 0.4 such that the power for the key secondary endpoint is not much less than the

power for the primary endpoint in these cases. In case of a true HR of 0.6, the power does not exceed 60% for any scenario considered.

5.7.4. Data Monitoring Committee (DMC) or Other Review Board

5.7.4.1. Independent Data Monitoring Committee

An Independent Data Monitoring Committee (IDMC) has overall responsibility for safeguarding the interests of participants by monitoring unblinded safety and efficacy data obtained during the study and for reviewing the results of the interim analysis), thereby ensuring that the study is conducted in accordance with the highest scientific and ethical standards.

The IDMC is tasked to organize and conduct the interim analysis with the assistance of the independent Statistical Support Group (SSG). The IDMC will make appropriate recommendations based on statistical analyses summaries provided by the independent statistician at SSG based on the IDMC SAP, prepared in accordance with the IDMC charter.

All communication between the IDMC and the Sponsor will be directed through the Sponsor Committee, which includes senior medical and statistical members.

The Sponsor Committee will make decisions based on these recommendations.

5.7.4.2. Other Review Boards

A Steering Committee (SC) is commissioned for this study.

An Adjudication Committee (AC) is appointed to confirm the initial judgment of inoperability as well as the persistence/recurrence of CTEPH assessed by the investigational site. Anonymized participant screening data will be submitted to the AC for their review and confirmation of eligibility prior to randomization. Each eligibility case will be read in parallel by at least 2 members of the AC. The AC is an external independent committee composed of multidisciplinary experts with extensive experience in the fields of PH, in particular with CTEPH and in clinical trials in these domains. Full details of the composition of the AC, their roles and responsibilities, and adjudication process are described in the AC Charter.

A Clinical Event Committee (CEC) is appointed to review and adjudicate in a blinded fashion each clinical worsening event and the time of its first occurrence. In addition, primary reason for hospitalization and mode of death, will be adjudicated. Any relevant supporting information (hospital discharge summaries, local laboratory or imaging data, etc) will be provided to the CEC. Full details for the procedure, as well as the composition and operation of the CEC is described in the CEC charter.

The program-level established Independent Liver Safety Data Review Board (ILSDRB) will be appointed for this study to provide ongoing assessment and advice regarding hepatic events of special interest (HAESI) and cases of abnormal transaminases laboratory values.

6. SUPPORTING DOCUMENTATION

6.1. Appendix 1: List of Abbreviations

6MWD	6-minute walk distance
AC	adjudication committee
AE	adverse event
AEPR	adverse events of potential risk
AER	average annualized event rate
AESI	adverse events of special interest
ALT/SGPT	alanine aminotransferase
ANCOVA	analysis of covariance
AST/SGOT	aspartate aminotransferase
ATC	anatomic and therapeutic class
BMI	body mass index
BPA	balloon pulmonary angioplasty
BSA	body surface area
BUN	blood urea nitrogen
CEC	clinical events committee
CHMP	committee for medicinal products for human use (European Union)
CL	confidence limit
CRF	case report form
CSR	clinical study report
CTCAE	common terminology criteria for adverse events
CTEPH	chronic thromboembolic pulmonary hypertension
CV	coefficient of variation
CW	clinical worsening
CWE	clinical worsening event
DB	double-blind
DPS	data presentation specifications
ECG	electrocardiogram
eCRF	electronic case report form
EODBT	end of double-blind treatment
EOLT	end of open-label treatment
EOS	end of study
EODBT	end of treatment
FAS	full analysis set
FDA	food and drug administration
GFR	glomerular filtration rate
HR	hazard ratio
IA	interim analysis
ICH	international council for harmonisation
IDMC	independent data monitoring committee
IE	intercurrent event
IF	information fraction
IQ	interquartile
IWRS	interactive web response system
LLOQ	lower limit of quantification
LOCF	last observation carried forward
LFT	liver functional test
MAR	missing at random
MCF	mean cumulative function
MedDRA	medical dictionary for regulatory activities
MMRM	mixed model with repeated measures
OL	open-label
OLAS	open-label analysis set
PAH	pulmonary arterial hypertension
PEA	pulmonary endarterectomy

PFT	pulmonary function test
PH	pulmonary hypertension
PD	pharmacodynamic(s)
PI	principal investigator
PK	pharmacokinetic(s)
PMDA	pharmaceuticals and medical devices agency
PT	preferred term
PVOD	pulmonary veno occlusive disease
PY	patient years
RBC	red blood cell
RHC	right heart catheterization
SAE	serious adverse event
SAP	statistical analysis plan
SAS	safety analysis set
SC	steering committee
SD	standard deviation
SMQs	standardized MedDRA queries
SDG	Standardized drug groupings
SOC	system organ class
SS	safety set
TEAE	treatment emergent adverse event
Tmax	time to maximum concentration
TSH	thyroid stimulating hormone
TTCW	time to clinical worsening
ULN	upper limit of normal
US NCI	united states national cancer institute
WBC	white blood cell
WHO	world health organization
WHO-DD	world health organization drug dictionary

6.2. Appendix 2: Changes to Protocol-Planned Analyses

The following changes to the per-protocol planned analysis are made in this SAP:

Table 18: Summary of Significant Changes from Protocol

Section Number and Name	Description of Significant Change	Brief Rationale
4 Populations for Analyses	Safety initiated Set (SIS) is replaced by the newly defined Open Label Analysis Set (OLAS) Analysis set for the hemodynamic sub-study was not defined in the protocol. The analysis set RHCS and its usage is added in Section 4.	To consider only participants who receive OL study intervention in the OL treatment extension period.
5 Statistical Analyses	Change/reduction in scope of statistical analyses for an abbreviated CSR	Early termination of study for futility
5 Statistical Analyses	Efficacy analyses are conducted irrespective of stages as opposed to the protocol planned inverse normal combination test	Following the decision to prematurely stop the study for futility, only limited data in stage 2 are available.
5.3.2.1 Primary Estimand	Added a new intercurrent event definition for 6MWTs not performed due to participant being unable to perform test due to 'clinical worsening PAH related'.	To apply a conservative imputation in this situation rather than allowing the model to treat as MAR.

6.3. Appendix 3: Demographics and Baseline Characteristics

The number of participants in each analysis set will be summarized and listed by intervention group and overall. In addition, the distribution of participants by region, country, and site ID will be presented unless otherwise noted. Region will be further categorized by US versus Non-US.

Table 19 and Table 20 present a list of the demographic variables and baseline characteristics that will be summarized by intervention group and overall, for the FAS, OLAS and RHCS analysis sets. Demographics will also be summarized by region using the FAS analysis set.

By participant listing of demographic data will be provided based on FAS only.

Table 19: Demographic Variables

Continuous Variables:	**Summary Type**
Age (years)	Descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum], and IQ range).
Weight (kg)	
Height (cm)	
Body Mass Index (BMI) (kg/m^2)	
Categorical Variables	
Age group (18-<65, ≥65-75, >75 years)	Frequency distribution with the number and percentage of participants in each category.
Sex (male, female, unknown, undifferentiated)	
Race[a] (American Indian or Alaska Native, Asian (Asian- Japanese and Asian- Non-Japanese), Black or African American, Native Hawaiian or other Pacific Islander, White, Not reported, Unknown, Multiple)	
Ethnicity (Hispanic or Latino, not Hispanic or Latino, not reported, unknown)	
BMI ([underweight <18.5 kg/m^2, normal 18.5-<25 kg/m^2, overweight 25-<30 kg/m^2, obese ≥30 kg/m^2])	
Country	
Region (See Section 5.6.3)	
Region (USA, Non-USA)	

[a] Where multiple race categories are indicated for a participant, the Race is displayed as 'Multiple'

Table 20 presents a list of the baseline characteristic variables that will be summarized by intervention group and overall, for the FAS, OLAS and RHCS analysis sets. By-participant listings of baseline data will be provided based on FAS.

Table 20: Baseline Characteristics Variables

Continuous Variables:	**Summary Type**
Time since the initial CTEPH diagnosis (months)*	Descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum], and IQ range).
Time since most recent CTEPH hospitalization (days)*	
Number of previous hospitalizations for CTEPH within 6 months prior to Screening	
6MWD (m)	
Borg dyspnea category-ratio 10 scale post-6MWT	
Systolic blood pressure (mmHg)	
Diastolic blood pressure (mmHg)	
Pulse/Heart rate (beats per minute)	
Mean Pulmonary Arterial Pressure (mPAP)	
Categorical Variables	
CTEPH diagnosis (as randomized):	
• Inoperable (with or without BPA)	

• Persistent/recurrent after PEA surgery (with or without BPA)	Frequency distribution with the number and percentage of patients in each category.
CTEPH Classification (Sub-Etiology) • Inoperable with BPA • Inoperable without BPA • Persistent / recurrent after PEA (with BPA) • Persistent / recurrent after PEA (without BPA)	
WHO Functional Class: I, II, III, IV	
Mean Pulmonary Arterial Pressure (mPAP) (<25 mmHg, ≥25 mmHg)	
Use of PH specific therapies at baseline (as the randomization stratum in IRT):	
• Riociguat	
• Other PH specific therapy	
• None	
6MWD (<165 m, ≥165-≤440 m, >440 m)	
NT-proBNP (<300 ng/L, 300-1400 ng/L, >1400 ng/L)	
PAH Therapy at Screening (ERA, PDE5i, sGC stimulator, Prostacyclin receptor agonist, Prostanoids and prostacyclin analogues, Other PAH therapies and all the combination of these therapies)	

m = meter; bpm = beats per minute; WHO FC =World Health Organization functional class; PAH = Pulmonary arterial hypertension; ERA = Endothelin receptor antagonist; NT-proBNP = N-terminal prohormone of Brain natriuretic peptide; IRT = Interactive Response Technology; 6MWD = 6-minute walk distance; PDE-5i = Phosphodiesterase type-5 inhibitor; sGC = Soluble guanylate cyclase; BPA = balloon pulmonary angioplasty; PEA = pulmonary endarterectomy
* Relative to randomization date

Pulmonary Function Tests

Pulmonary function tests (PFT) including Forced vital capacity (FVC), Forced expiratory volume in 1 second (FEV1), FEV1 percent of predicted and FEV1/FVC) are performed at Screening for participants with a known or suspected history of significant lung disease. Historical PFT data obtained within 12 months prior to Screening are accepted where participant's pulmonary status has been stable/unchanged during this time and the results are reliable by the investigator.

Baseline PFT will be summarized by intervention group and overall, for the FAS. Participant listings will be provided based on FAS.

Heart Catheterization

The following heart catheterization variables at baseline will be summarized on the FAS:

- Heart rate
- Pulmonary Vascular Resistance (PVR)
- Pulmonary Artery Wedge Pressure (PAWP)
- Mean Right Atrial Pressure (mRAP)
- Systolic Pulmonary Arterial Pressure (sPAP)
- Diastolic Pulmonary Arterial Pressure (dPAP)
- Mean Pulmonary Artery Pressure (mPAP)
- Cardiac Output (CO)
- Mixed venous oxygen saturation (SvO2)
- Systolic Systemic Arterial Pressure (sSAP)

- Diastolic Systemic Arterial Pressure (dSAP)
- Cardiac index (CI)
- Total Pulmonary Resistance (TPR)
- Left Ventricular End Diastolic Pressure (LVEDP)

In addition, the time since catheterization was performed (weeks) relative to date of randomization will be derived and summarized. A listing will be provided based on FAS.

Other Characteristics
Signs and symptoms of right heart failure will be listed only.

Riociguat use at screening (yes/no) with reasons will be listed only.

6.4. Appendix 4: Protocol Deviations

Protocol deviation will be analyzed during both the DB and OL periods.

In general, the major protocol deviations will have the potential to impact participants' rights, safety or well-being, or the integrity and/or result of the clinical study.

Participants with major protocol deviations will be identified during the conduct of the trial in an ongoing manner using a comprehensive list of criteria indicated in the Cross-Pharma TV-SOP-04282: Identification and Management of Clinical Trial Issues and Protocol Deviations for the definition of major protocol deviations. A separate document, the Major Protocol Deviation (MPD) Criteria form, contains details of major and potentially major protocol deviations criteria.

Participants with major protocol deviations will be identified prior to database lock and the participants with major protocol deviations will be summarized by following categories/codes (DVDECOD):

- Developed withdrawal criteria but not withdrawn
- Entered but did not satisfy criteria
- Received a disallowed concomitant treatment
- Received wrong treatment or incorrect dose
- Other

Major protocol deviations will be summarized and listed using the FAS and OLAS.

Protocol Deviations resulting from COVID-19 impact and regional crisis are described in Section 6.13 (Appendix 13).

Participants who did not meet the inclusion criteria or who met exclusion criteria will be summarized and listed on the FAS. A summary of participants who did not meet the inclusion criteria or who met exclusion criteria will be provided overall for the Screened failures.

6.4.1. Per-Protocol Set

The per protocol analysis set (PP) includes a subset of participants in the full analysis set (FAS) who receive at least one dose of study intervention and who comply adequately with the protocol to be likely to exhibit the intervention effects.

Criteria for sufficient compliance include exposure to intervention, availability of measurements and absence of major protocol deviations [MPD] that have an impact on the treatment effect.

Participants will be excluded from the PPS if:

- **Participant is randomized and did not satisfy certain eligibility criteria:**
  - Participant did not have CTEPH (WHO group 4) fulfilling criteria specified in inclusion criteria 3 [MPD3, MPD3a]

  - Inclusion criteria 4 not met: 6MWD was not ≥ 100 m AND ≤ 450 m at Screening and baseline OR baseline 6MWD differs more than 15% from eligibility test [MPD4]
  - Inclusion criterion 5 not met: WHO FC is not II, III or IV [MPD5]
  - Exclusion criteria 20 / 21 met: Treatment with a strong CYP3A4 inducer or strong CYP3A4 inhibitor within 1 month prior to baseline; [MPD28 and 29]
  - Exclusion criteria 11 met: Participant was treated with ERAs, intravenous prostacyclins / prostacyclin analogs, or investigational treatment within 90 days prior to Randomization [MPD 19, 19a and 19b]

- **Participant compliance to study intervention < 80%:**
  - Intervention compliance < 80 % during the 28-week fixed period, as per compliance based on number of tablets dispensed/ returned (as described in section 6.8), derived from the eCRF

- **Participant received wrong study intervention during the 28-week fixed period:**
  - Wrong study intervention (i.e., received placebo when assigned macitentan and vice versa) taken for > 14 days during the maintenance 75 mg dose of the 28-week fixed period.

- **Participant received unauthorized concomitant medication during the 28-week fixed period:**
  - Administration of strong CYP3A4 inhibitor, strong CYP3A4 inducer concomitantly with study intervention [MPD 45 & 46]
  - Administration of any other investigational drug concomitantly with study intervention [MPD 47]
  - Participant had an elective BPA and/ or PEA procedure before completion of Week 28 [MPD 50: "BPA PEA procedures without clinical worsening (until Week 28)"]

- **Occurrence of major protocol deviation that could confound the interpretation of analysis conducted on the FAS:**
  - 6WMT assessments not conducted at any visit during the 28-week fixed duration period
  - Participant's randomized study intervention allocation was unblinded by the investigator / site (eCRF *Treatment Unblinding*)

6.4.2. Hemodynamic Sub-Study

The number and percentages of participants with MPDs related to the hemodynamic sub-study will be summarized using the RHCS.

The MPDs related to the hemodynamic sub-study are:

- Exclusion criteria 26 met: Previous RHC with serious complications

- Inclusion criteria 3 not met: participant who does NOT have CTEPH (WHO group 4) fulfilling either:

    a. Confirmed inoperable by AC or
    b. Persistent/recurrent CTEPH after BPA deemed inoperable confirmed by AC or
    c. Persistent/recurrent CTEPH after PEA

The above MPDs related to the hemodynamic sub-study will be listed on the RHCS.

6.5. Appendix 5: Prior and Concomitant Medications

Prior and Concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO-DD). Prior medications are defined as any therapy used and stopped before the day of first dose of study intervention. Concomitant medications during the DB period are defined as any therapy used on or after the same day as the first dose of study intervention, including those that started before and continue on / after the first dose of study intervention, excluding those that started on or after start of OL intervention.

Summaries of concomitant medications will be presented by Anatomical Therapeutic Chemical (ATC) class, standardized medication name and intervention group on the FAS. The proportion of participants who receive each concomitant medication will be summarized as well as the proportion of participants who receive at least 1 concomitant medication.

Prior medications will be summarized by intervention group and ATC class, standardized medication name and study intervention group on the FAS.

A participant listing of prior medication data will be provided on the FAS. Participant listings of concomitant medication data will also be provided on the FAS for the DB period. Additionally, a listing of concomitant medication data will be provided for the OL period including therapies that started on or after the start of OL treatment.

In addition, separate summaries of concomitant medications of special interest will be presented. These include PH therapy. See Section 6.10 Appendix 10 for list of medications in each category.

6.6. Appendix 6: Procedures

Procedures will be coded using the Medical Dictionary for Regulatory Activities (MedDRA).

Prior procedures are defined as any procedure performed before the day of first dose of study intervention and include surgical and medical procedures reported as medical history (general and MH of interest).

Treatment emergent procedures in the DB period are defined as any procedure occurring at or after the initial administration of DB study intervention through up to the day of last dose in the double-blind plus 30 days (EODBT + 30 days) and prior to start of OL study intervention. Summaries of prior CTEPH related procedures (BPA/ PEA), and treatment emergent CTEPH related procedures (BPA/ PEA) and other treatment emergent procedures in the DB period will be presented by preferred term and intervention group on the FAS. The proportion of participants who had each procedure will be summarized as well as the proportion of participants who had at least 1 procedure, and the proportion of participants who had BPA or PEA will be presented.

Participant listings of all procedures will be provided on the FAS.

6.7. Appendix 7: Medical History

All relevant medical history/current medical conditions based on the investigator's judgment (e.g., chronic and ongoing acute conditions, serious past conditions) present before and/or at the time of signing informed consent are collected in the electronic case report form (eCRF) and will be summarized by intervention group and overall, for the FAS. Includes medical history of interest and COVID-19 related medical history.

Participant listing of previous and concomitant disease at screening will also be provided on the FAS.

In addition, medical history of interest will be summarized separately according to the eCRF prespecified terms and by the sub-level coded terms for "Thrombophilic risk factors" and "Indwelling catheters and leads".

Medical history of interests:
• Pulmonary Embolism
• Deep vein thrombosis
• Thrombophilic risk factors o Term 1 o Term 2
• Right heart failure
• Splenectomy
• Ventriculoatrial shunt
• Inflammatory bowel disease
• Osteomyelitits
• Indwelling catheters and leads o Term 1 o Term 2

6.8. Appendix 8: Intervention Compliance

Compliance will be summarized descriptively (N, mean, SD, median, and range [minimum, maximum]) by study intervention group for the FAS and actual intervention group for the Safety analysis set, respectively, over the DB period. Compliance to randomized intervention versus actual intervention will be presented in a summary table.

The number and percentage of participants who have at least 80% compliance through the **double-blind treatment period** in the FAS and safety analysis set will be summarized by intervention group.

Compliance during **double blind treatment period** irrespective of the temporary intervention discontinuations will be calculated as follows:

$$Compliance\ (\%) = \frac{(number\ of\ tablets\ dispensed\ \ - number\ of\ tablets\ returned) \times 100}{expected\ number\ of\ tablets\ \ taken\ during\ the\ DB\ treatment\ period}$$

As tablets are taken once daily,

$$Expected\ number\ of\ tablets\ taken\ during\ the\ DB\ treatment\ period = DB\ treatment\ end\ date\ (EODBT) -\ DB\ treatment\ start\ date\ (Day\ 1)\ + 1$$

The number and percentage of participants who have at least 80% compliance through the **fixed study period up to Week 28** treatment period in the FAS analysis set will be summarized by intervention group.

Compliance during **fixed study period up to Week 28 treatment period** irrespective of the temporary intervention discontinuations will be calculated as follows:

$$Compliance\ (\%) = \frac{(number\ of\ tablets\ dispensed\ prior\ to\ Week\ 28 - number\ of\ tablets\ returned\ **) \times 100}{expected\ number\ of\ tablets\ \ taken\ up\ to\ Week\ 28}$$

***tablets returned with same reference ID as tablets dispensed prior to Week 28*

As tablets are taken once daily,

$$Expected\ number\ of\ tablets\ taken\ up\ to\ Week\ 28 = Week\ 28\ visit\ date\# -\ DB\ treatment\ start\ date\ (Day\ 1)\ + 1$$

 # or DB treatment end date if participant discontinued treatment prior to Week 28 Visit

If the FAS is identical to the SS and randomized intervention arm is identical to actual intervention arm for all participants, these summaries and listings will be produced on the FAS only.

Participant listing of compliance data will be provided on the FAS for the fixed study period up to Week 28.

6.9. Appendix 9: Adverse Events of Special Interest

The following categories of AESI have been considered as events of interest for this study. These events will be analyzed based on the MedDRA definitions provided in the AESI definition file.

Any modifications of terms (according to the MedDRA SMQs/PTs) may occur based on later dictionary updates and/or external guidance and the latest definitions will be used at the time of analyses.

Details will be provided in an AESI definition file that is saved at the macitentan compound level and version controlled statistical programming environment.

- Hypotension
  - Symptomatic hypotension
- Anemia/hemoglobin decrease
- Oedema and fluid retention
- Liver events:
  1. Hepatic disorders (excl. Liver-related coagulation and bleeding disturbances and Ascites not associated to any other hepatic disorder)
  2. Drug related hepatic disorders – comprehensive search (excl. Liver-related coagulation and bleeding disturbances and Ascites not associated to any other hepatic disorder)

6.9.1. Potential Risks

In addition, the following potential risks for macitentan are identified as follows:

- Menstrual disorders/vaginal hemorrhage
- Ovarian cyst
- Testicular disorders and male infertility
- Pulmonary veno occlusive disease (PVOD) with event of pulmonary oedema.

SMQs/PTs to identify these potential risks will be contained within the AESI definition file.

6.10. Appendix 10: Medications of Special Interest

Medications of special interest include pulmonary hypertension (PH) Specific therapies which are detailed as follows based on the Standardized Drug Groupings (SDG) B3/C3 March 1, 2023, applying the group "*Drugs for pulmonary arterial hypertension (PAH)".*

Table 21: PH-Specific Medications

Drug code	Class	Product name	ATC code	ATC description
01766201001	ERA	Ambrisentan	C02KX	Antihypertensives for pulmonary arterial hypertension
13988401001	ERA and PDE5i	Ambrisentan; Tadalafil	C02KX	Antihypertensives for pulmonary arterial hypertension
01208701001	Prostanoids and prostacyclin analogues	Beraprost	B01AC	Platelet aggregation inhibitors excl. heparin
01208702001	Prostanoids and prostacyclin analogues	Beraprost Sodium	B01AC	Platelet aggregation inhibitors excl. heparin
01587701001	ERA	Bosentan	C02KX	Antihypertensives for pulmonary arterial hypertension
01587702001	ERA	Bosentan Monohydrate	C02KX	Antihypertensives for pulmonary arterial hypertension
00652701001	Prostanoids and prostacyclin analogues	Epoprostenol	B01AC	Platelet aggregation inhibitors excl. heparin
00652702001	Prostanoids and prostacyclin analogues	Epoprostenol Sodium	B01AC	Platelet aggregation inhibitors excl. heparin
00944801001	Prostanoids and prostacyclin analogues	Iloprost	B01AC	Platelet aggregation inhibitors excl. heparin
00944802001	Prostanoids and prostacyclin analogues	Iloprost Trometamol	B01AC	Platelet aggregation inhibitors excl. heparin
06301501001	ERA	Macitentan	C02KX	Antihypertensives for pulmonary arterial hypertension
15953501001	ERA; PDE5i	Macitentan; Tadalafil	C02KX	Antihypertensives for pulmonary arterial hypertension
08813301001	Prostacyclin receptor agonist	Ralinepag	C02KX	Antihypertensives for pulmonary arterial hypertension
08813302001	Prostacyclin receptor agonist	Ralinepag Sodium	C02KX	Antihypertensives for pulmonary arterial hypertension
08813303001	Prostacyclin receptor agonist	Ralinepag Sodium Monohydrate	C02KX	Antihypertensives for pulmonary arterial hypertension
06329101001	sGC stimulator	Riociguat	C02KX	Antihypertensives for pulmonary arterial hypertension
08970101001	Prostacyclin receptor agonist	Selexipag	B01AC	Platelet aggregation inhibitors excl. heparin
01367501001	PDE5i	Sildenafil	C02KX	Antihypertensives for pulmonary arterial hypertension
01367502001	PDE5i	Sildenafil Citrate	C02KX	Antihypertensives for pulmonary arterial hypertension
01635901001	ERA	Sitaxentan	C02KX	Antihypertensives for pulmonary arterial hypertension
01635902001	ERA	Sitaxentan Sodium	C02KX	Antihypertensives for pulmonary arterial hypertension
01579201001	PDE5i	Tadalafil	C02KX	Antihypertensives for pulmonary arterial hypertension
01537001001	Prostanoids and prostacyclin analogues	Treprostinil	B01AC	Platelet aggregation inhibitors excl. heparin

Drug code	Class	Product name	ATC code	ATC description
01537003001	Prostanoids and prostacyclin analogues	Treprostinil Diolamin	B01AC	Platelet aggregation inhibitors excl. heparin
01537004001	Prostanoids and prostacyclin analogues	Treprostinil Sodium	B01AC	Platelet aggregation inhibitors excl. heparin
01536601001	PDE5i	Vardenafil	C02KX	Antihypertensives for pulmonary arterial hypertension
01536603001	PDE5i	Vardenafil Dihydrochloride	C02KX	Antihypertensives for pulmonary arterial hypertension
01536602001	PDE5i	Vardenafil Hydrochloride	C02KX	Antihypertensives for pulmonary arterial hypertension
01536604001	PDE5i	Vardenafil Hydrochloride Trihydrate	C02KX	Antihypertensives for pulmonary arterial hypertension
15228401001	Other PAH medications	Seralutinib	C02KX V98	Antihypertensives for pulmonary arterial hypertension Investigational drug
15846601001	Other PAH medications	Sotatercept	C02KX V98	Antihypertensives for pulmonary arterial hypertension Investigational drug

ATC= anatomic and therapeutic class; ERA = Endothelin receptor antagonist; PDE-5i = Phosphodiesterase type-5 inhibitor; sGC = Soluble guanylate cyclase.
Standardized Drug Groupings B3/C3 March 1, 2023, [Grouping (Drugs for pulmonary arterial hypertension (PAH)]. Medications must match on both Preferred Term and ATC code.

Note: Any modifications of terms may occur based on later dictionary updates and/or external guidance.

PH Specific Therapy at Screening includes any PH therapy specified above that started prior and ongoing at date of screening or started during the screening period prior to Day 1. PH Specific Therapy at Screening are summarized by PH therapy class as part of the Baseline Characteristics. Changes to PH therapy and / or initiation of new PH therapy after screening and prior to Day 1 will be flagged.

Concomitant PH specific medications will be summarized by PH therapy class, standardized medication name and study intervention group on the FAS. Changes in concomitant PH therapies assessed by the investigator as “initiation of new PH therapy”, “switch”, or “dose increase” compared to baseline will be summarized.

A listing of medications of special interest will also be provided including changes in PH therapy compared to baseline according to investigator assessment and reasons (efficacy, safety, other: specify) where applicable. Initiation of CTEPH rescue therapy (defined as an initiation of new PH therapy, switch or dose increase) on or after first dose of study intervention will also be flagged in the listing.

6.11. Appendix 11: Laboratory Toxicity Grading

The grading scale use for lab assessments is based on 'Common Terminology Criteria for Adverse Events (CTCAE) v5.0'.

If a laboratory value falls within the grading as specified below but also within the laboratory normal limits, the value is considered to be normal and will be reset to grade 0.

Pre-baseline measurements will use the same grading ranges as applied to baseline measurements. In case a test has two sets of ranges – one for baseline normal and one for baseline abnormal, the one for baseline normal will be applied for all measurements taken pre-baseline and on baseline.

Text in gray italic in Table 22 is present in the grading scale but is not applied by Janssen when grading lab data.

Table 22: Common Terminology Criteria for Adverse Events (CTCAE) v5.0

CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Janssen implementation notes
Blood and lymphatic system disorders					
Anemia	Hemoglobin (Hgb) <LLN – 10.0 g/dL; <LLN – 6.2 mmol/L; <LLN – 100 g/L	Hemoglobin (Hgb) <10.0 – 8.0 g/dL; <6.2 – 4.9 mmol/L; <100 – 80g/L	Hemoglobin (Hgb) <8.0 g/dL; <4.9 mmol/L; <80 g/L; *transfusion indicated*	*Life-threatening consequences; urgent intervention indicated*	Clinical signs and symptoms are not taken into consideration for grading.
Leukocytosis	-	-	Leukocytes >100,000/mm3; >100 x 10e9 /L	*Clinical manifestations of leucostasis; urgent intervention indicated*	Clinical signs and symptoms are not taken into consideration for grading; Added ranges in SI unit (x 10e9 /L)
Investigations					
Alanine aminotransferase increased	>ULN – 3.0 x ULN if baseline was normal; 1.5 – 3.0 x baseline if baseline was abnormal	>3.0 – 5.0 x ULN if baseline was normal; >3.0 – 5.0 x baseline if baseline was abnormal	>5.0 – 20.0 x ULN if baseline was normal; >5.0 – 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal	Ranges defined for "abnormal baseline" are applied only if baseline > ULN. If baseline < LLN, then ranges for "normal baseline" are applied.
Alkaline phosphatase increased	>ULN – 2.5 x ULN if baseline was normal; 2.0 – 2.5 x baseline if baseline was abnormal	>2.5 – 5.0 x ULN if baseline was normal; >2.5 – 5.0 x baseline if baseline was abnormal	>5.0 – 20.0 x ULN if baseline was normal; >5.0 – 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal	Ranges defined for "abnormal baseline" are applied only if baseline > ULN. If baseline < LLN, then ranges for "normal baseline" are applied.
Aspartate aminotransferase increased	>ULN – 3.0 x ULN if baseline was normal; 1.5 – 3.0 x baseline if baseline was abnormal	>3.0 – 5.0 x ULN if baseline was normal; >3.0 – 5.0 x baseline if baseline was abnormal	>5.0 – 20.0 x ULN if baseline was normal; >5.0 – 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal	Ranges defined for "abnormal baseline" are applied only if baseline > ULN. If baseline < LLN, then ranges for "normal baseline" are applied.

CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Janssen implementation notes
Blood bilirubin increased	>ULN – 1.5 x ULN if baseline was normal; > 1.0 – 1.5 x baseline if baseline was abnormal	>1.5 – 3.0 x ULN if baseline was normal; >1.5 – 3.0 x baseline if baseline was abnormal	>3.0 – 10.0 x ULN if baseline was normal; >3.0 – 10.0 x baseline if baseline was abnormal	>10.0 x ULN if baseline was normal; >10.0 x baseline if baseline was abnormal	Ranges defined for "abnormal baseline" are applied only if baseline > ULN. If baseline < LLN, then ranges for "normal baseline" are applied.
Creatinine increased	>ULN – 1.5 x ULN	>1.5 – 3.0 x baseline; >1.5 – 3.0 x ULN	>3.0 x baseline; >3.0 – 6.0 x ULN	>6.0 x ULN	
Hemoglobin increased	Increase in >0 – 2 g/dL; Increase in >0 – 20 g/L	Increase in >2 – 4 g/dL; Increase in >20 – 40 g/L	Increase in >4 g/dL; Increase in >40 g/L	-	The increase indicates the level of increase above normal (above ULN). Applied as, e.g. grade 1 (g/dL): >ULN – ULN+2 g/dL; Added ranges in SI unit (g/L).
INR increased	>1.2 – 1.5; *>1 – 1.5 x baseline if on anticoagulation; monitoring only indicated*	>1.5 – 2.5; *>1.5 – 2.5 x baseline if on anticoagulation; dose adjustment indicated*	>2.5; *>2.5 x baseline if on anticoagulation; bleeding*	-	Concomitant therapy or clinical signs and symptoms are not taken into consideration for grading.
Lymphocyte count decreased	<LLN – 800/mm^3; <LLN – 0.8 x 10e9/L	<800 – 500/mm^3; <0.8 – 0.5 x 10e9 /L	<500 – 200/mm^3; <0.5 – 0.2 x 10e9 /L	<200/mm^3; <0.2 x 10e9 /L	
Lymphocyte count increased	-	>4000/mm^3 – 20,000/mm^3; >4 – 20 x 10e9 /L	>20,000/mm^3; >20 x 10e9 /L	-	Added ranges in SI unit (x 10e9 /L).
Neutrophil count decreased	<LLN – 1500/mm^3; <LLN – 1.5 x 10e9 /L	<1500 – 1000/mm^3; <1.5 – 1.0 x 10e9 /L	<1000 – 500/mm^3; <1.0 – 0.5 x 10e9 /L	<500/mm^3; <0.5 x 10e9 /L	Both Neutrophils and segmented neutrophils are graded using these criteria.
Platelet count decreased	<LLN – 75,000/mm^3; <LLN – 75.0 x 10e9 /L	<75,000 – 50,000/mm^3; <75.0 – 50.0 x 10e9 /L	<50,000 – 25,000/mm^3; <50.0 – 25.0 x 10e9 /L	<25,000/mm^3; <25.0 x 10e9 /L	
White blood cell decreased	<LLN – 3000/mm^3; <LLN – 3.0 x 10e9 /L	<3000 – 2000/mm^3; <3.0 – 2.0 x 10e9 /L	<2000 – 1000/mm^3; <2.0 – 1.0 x 10e9 /L	<1000/mm^3; <1.0 x 10e9 /L	

CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Janssen implementation notes
Metabolism and nutrition disorders					
Hypercalcemia	Calcium of >ULN – 11.5 mg/dL; >ULN – 2.9 mmol/L;	Calcium of >11.5 – 12.5 mg/dL; >2.9 – 3.1 mmol/L; *symptomatic*	Calcium of >12.5 – 13.5 mg/dL; >3.1 – 3.4 mmol/L; *hospitalization indicated*	Calcium of >13.5 mg/dL; >3.4 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading. Note: For Hypercalcemia or Hypocalcemia in the CTCAE Lab Toxicity grading scale, the result (reference ranges) from serum calcium (ULN /LLN) from the local or central lab are being utilized.
Hyperkalemia	Potassium >ULN – 5.5 mmol/L	Potassium >5.5 – 6.0 mmol/L; *intervention initiated*	Potassium >6.0 – 7.0 mmol/L; *hospitalization indicated*	Potassium >7.0 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.
Hypermagnesemia	Magnesium >ULN – 3.0 mg/dL; >ULN – 1.23 mmol/L	-	Magnesium >3.0 – 8.0 mg/dL; >1.23 – 3.30 mmol/L	Magnesium >8.0 mg/dL; >3.30 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.
Hypernatremia	Sodium >ULN – 150 mmol/L	Sodium >150 – 155 mmol/L; *intervention initiated*	Sodium >155 – 160 mmol/L; *hospitalization indicated*	Sodium >160 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.
Hypoalbuminemia	Albumin <LLN – 3 g/dL; <LLN – 30 g/L	Albumin <3 – 2 g/dL; <30 – 20 g/L	Albumin <2 g/dL; <20 g/L	*Life-threatening consequences; urgent intervention indicated*	Clinical signs and symptoms are not taken into consideration for grading.

CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Janssen implementation notes
Hypocalcemia	Calcium of <LLN – 8.0 mg/dL; <LLN – 2.0 mmol/L;	Calcium of <8.0 – 7.0 mg/dL; <2.0 – 1.75 mmol/L; *symptomatic*	Calcium of <7.0 – 6.0 mg/dL; <1.75 – 1.5 mmol/L; *hospitalization indicated*	Calcium of <6.0 mg/dL; <1.5 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading. Note: For Hypercalcemia or Hypocalcemia in the CTCAE Lab Toxicity grading scale, the result (reference ranges) are from serum calcium (ULN /LLN) from the local or central lab are being utilized.
Hypoglycemia	Glucose <LLN – 55 mg/dL; <LLN – 3.0 mmol/L	Glucose <55 – 40 mg/dL; <3.0 – 2.2 mmol/L	Glucose <40 – 30 mg/dL; <2.2 – 1.7 mmol/L	Glucose <30 mg/dL; <1.7 mmol/L; *life-threatening consequences; seizures*	Clinical signs and symptoms are not taken into consideration for grading. Urine glucose is not graded.
Hypokalemia	*Potassium <LLN – 3.0 mmol/L*	*Symptomatic with* Potassium <LLN – 3.0 mmol/L; *intervention indicated*	Potassium <3.0 – 2.5 mmol/L; *hospitalization indicated*	Potassium <2.5 mmol/L; *life-threatening consequences*	"Symptomatic" ranges are applied for grade 2, grade 1 not assigned, i.e. worst case applied. Clinical signs and symptoms are not taken into consideration for grading of grade 3 and 4.
Hypomagnesemia	Magnesium <LLN – 1.2 mg/dL; <LLN – 0.5 mmol/L	Magnesium <1.2 – 0.9 mg/dL; <0.5 – 0.4 mmol/L	Magnesium <0.9 – 0.7 mg/dL; <0.4 – 0.3 mmol/L	Magnesium <0.7 mg/dL; <0.3 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading.

CTCAE Term	**Grade 1**	**Grade 2**	**Grade 3**	**Grade 4**	**Janssen implementation notes**
Hyponatremia	Sodium <LLN – 130 mmol/L	*Sodium 125-129 mmol/L and asymptomatic*	*Sodium 125-129 mmol/L symptomatic; 120-124 mmol/L regardless of symptoms* Sodium <130-120 mmol/L	Sodium <120 mmol/L; *life-threatening consequences*	Clinical signs and symptoms are not taken into consideration for grading. Worst case ("<130-120 mmol/L" for grade 3 added by Janssen) is applied across grade 2/3 ranges: 120-129 mol/L assigned to grade 3, grade 2 not used.
Classification for renal function					
Chronic Kidney Disease	GFR 60 -< 90 mL/min/1.73m^2	GFR 30 -< 60 mL/min/1.73m^2	GFR 15 -< 30 mL/min/1.73m^2	GFR < 15 mL/min/1.73m^2	Rename Chronic Kidney disease as "Estimated glomerular filtration rate" GFR ≥90 = Grade 0. Note classifications are based on FDA guidance and CTCAE V5

* Grade 0 is assigned to a lab assessment when the lab test is described in the table, but the lab value is not assigned a grade 1 or higher.

6.12. Appendix 12: Schedule of Activities

Table 23: Schedule of Activities

PERIOD	Name	SCREENING	DOUBLE-BLIND TREATMENT PERIOD											Unscheduled[1]	Safety follow-up
PHASE	Name		Up-titration Phase				Maintenance Phase 75 mg								
			10 mg	37.5 mg			Fixed duration part						Variable duration part		
VISITS	Number/ Name	1	2	3	4☎	5	6	7☎	8	9	10-12	13	14-X	EODBT[2]	Anytime in case of clinical worsening
	Timing	At least 14 and up to 60 days prior to D1[1]	D1 / Randomization	W 4 (±3 d)	W 5 (±3 d)	W 6 (±3 d)	W 8 (±3 d)	W 9 (±3 d)	W10 (±3 d)	W12 (±3 d)	W16/20/24 (±5 d)	W 28 (±5 d)	W 40, every 12 weeks thereafter (±5d)	End of DB treatment (±5 d)	30 (+ 5) days after study interv. Disc.
Screening/Administrative															
Informed Consent (ICF)[3]		X													
Eligibility		X													
Demographics		X													
Medical history		X													
RHC (historical allowed[4])		X													
V/Q scan, PA, CTPA, MRA[5]		X													
PFTs[6]		X													
DB Study intervention. Dispensing			X	X			X			X	X	X	X		
Safety assessments															
Physical examination[7]		X	X	X		X	X		X	X	X	X	X	X	X
Vital signs (BP, PR)		X	X	X		X	X		X	X	X	X	X	X	X
Body weight		X	X	X		X	X		X	X	X	X	X	X	X
Home body weight monitoring			At least once per week until Week 12												
12-lead ECG			X				X			X		X	W 52		
Clinical Laboratory tests															
Central laboratory test		X	X	X		X	X		X	X	X	X	X	X	X
Serum pregnancy test[8]		X	X	X			X			X	X	X	X	X	X
Liver function tests		X	X	X		X	X		X	X	X	Monthly until W 52, every 12 weeks thereafter[9]			X
Urine pregnancy test[10]			X									Monthly in between visits starting at Week 32			
Efficacy assessments															
6MWT / BDI		X	X	X			X			X	X	X	X	X	X
Accelerometry[11]		Daily during waking hours													
WHO FC		X	X	X		X	X		X	X	X	X	X	X	X
Clinical worsening			X	X		X	X		X	X	X	X	X	X	X
EQ-5D-5L©			X	X			X			X	X	X	Week 40, 52		
PAH-SYMPACT®		(X[12])	X	X			X			X	X	X	Week 40, 52		
PHQ-8, WPAI, SF36 v2			X	X			X				W 16	X	Week 52		
PGA-S			X	X			X				W 16	X	Week 52		
Ongoing participant review															
Concomitant therapy		X	X	X	X	X	X	X	X	X	X	X	X	X	X

PERIOD	Name	SCREENING	DOUBLE-BLIND TREATMENT PERIOD											Unscheduled[1]	Safety follow-up
PHASE	Name		Up-titration Phase				Maintenance Phase 75 mg								
			10 mg	37.5 mg			Fixed duration part						Variable duration part		
VISITS	Number/ Name	1	2	3	4☎	5	6	7☎	8	9	10-12	13	14-X	EODBT[2]	Anytime in case of clinical worsening	
	Timing	At least 14 and up to 60 days prior to D1[1]	D1 / Randomization	W 4 (±3 d)	W 5 (±3 d)	W 6 (±3 d)	W 8 (±3 d)	W 9 (±3 d)	W10 (±3 d)	W12 (±3 d)	W16/20/24 (±5 d)	W 28 (±5 d)	W 40, every 12 weeks thereafter (±5d)	End of DB treatment (±5 d)		30 (+ 5) days after study interv. Disc.
SAEs/AEs		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Pharmacokinetics / Pharmacodynamics																
PK samples[13]			X	X			X			X		X				
ET-1 samples[13]			X	X			X			X		X				
NT-proBNP			X	X			X			X	X	X	X	X	X	
Research biomarker			X	X			X			X		X		X		
Optional Substudies																
RHC [14]		X										X				
Male reproductive system safety sub-study		Please refer to the schedule of assessments in Appendix 17, Section 4.														

Footnotes:

1 During the double-blind period, an unscheduled visit must be performed in case of a suspected clinical worsening or initiation/dose escalation of PH-specific therapy. The assessments performed at the unscheduled visit include, but are not limited to concomitant therapy, physical examination, vital signs (BP/PR) weight, laboratory assessment, post-dose 6MWT/Borg CR10 Scale® and WHO FC.

2 In case of premature discontinuation of double-blind study intervention, the EODBT visit must be performed within 7 days of taking the decision to stop study intervention. If the premature EODBT visit falls within the time window of a regular scheduled visit, then the visits can be combined, and all assessments pertaining to these visits will be performed.

3 Must be signed before first study-related activity.

4 If historical RHC data from the 24-week period prior to Randomization are available, these can be used. Otherwise, the RHC needs to be repeated at Screening.

5 Confirmation of diagnosis and inoperability of CTEPH based at least 2 of these assessments performed in the 14-month period prior to Randomization. Confirmation of diagnosis of persistent recurrent CTEPH after BPA based on at least one of these assessments in the 14-month period prior to Randomization.

6 Historical PFTs accepted if performed within 12 months prior to Screening and judged reliable by the investigator, provided the participant's pulmonary status remained unchanged during this time. Only necessary for participants with a known or suspected history of significant lung disease.

7 Physical examination at Screening will include assessment of the Child Pugh score for participants with known hepatic impairment.

8. For women of childbearing potential. Additional serum pregnancy test to be performed in case of a positive urine pregnancy test.

9 Monthly AST/ALT monitoring required until Week 52. Thereafter, AST/ALT monitoring can be performed every 12 weeks, unless continued monthly monitoring is mandated based on local regulatory requirements. Local laboratory can be used. If study treatment is interrupted/discontinued due to AST/ALT elevations, weekly tests are to be performed until to the values return to pre-treatment levels or within normal ranges. *Requirement for Japan:* Monthly AST/ALT monitoring is required. Whenever participants have monthly liver function tests, the following assessments should be conducted and reported in source notes: physical examination, body weight, blood pressure/pulse rate, and hematology (hemoglobin, hematocrit, and RBC count). These assessments must be recorded in the eCRF only if they are related to AEs.

10 Women of childbearing potential only. To be performed monthly (at site or at home if the interval between 2 regular site visit exceeds 4 weeks).

11 The accelerometer will be provided to the participant at Screening and worn daily during waking hours up to Week 28.

12 The participant will receive the mobile device at Screening and will be trained on its use. Completion of the PAH-SYMPACT® will start 7 days prior to Visit 2 (Randomization).

13 PK and ET-1 samples to be drawn pre-dose and 2-10 hours post-dose.

14 Participation in RHC sub-study is optional, except for Japan, where the participation in the sub-study is mandatory based on an agreement with PMDA. A historical RHC performed within 60 days of Randomization can be used as the baseline RHC, provided the criteria for historical RHCs described in the RHC guidelines (Protocol Appendix 16) are met.

Table 24: Visit and Assessment Schedule during the Open-Label Extension Period

PERIOD	Name	OPEN-LABEL (OL) TREATMENT EXTENSION PERIOD										Unscheduled[2]	SAFETY FOLLOW-UP
PHASE	Name	OL Double-Dummy Uptitration Phase				OL Maintenance Phase							
		10 mg	37.5 mg			75 mg							
VISITS	Number/ Name	EOBDT 6 /OL 1	OL 2	OL 3☎	OL 4	OL 5	OL 6 ☎	OL 7	OL 8	OL 9-x		Anytime in case of clinical worsening	
	Timing	OL D1[1]	OL W4 (±3 days)	OL W5 (±3 days)	OL W 6 (±3 days)	OL W 8 (±3 days)	OL W 9 (±3 days)	OL W10 (±3 days)	OL W 12 (±3 days)	OL W 24, 48, 72, 96, … (q 24 weeks) (±5 days)	EOLT[3] (±5 days)		30 (+ 5) days after last OL dose
Screening/Administrative													
Eligibility		X											
OL study interv. dispensing		X	X			X			X	X			
Safety assessments													
Physical examination		X	X		X	X		X	X	X	X	X	X
Vital signs (BP, PR)[7]		X	X		X	X		X	X	X	X	X	X
Body weight		X	X		X	X		X	X	X	X	X	X
12-lead ECG		X				X				Weeks 24, 48			
Clinical Laboratory tests													
Central laboratory test		X	X		X	X		X	X	X	X	X	X
Serum pregnancy test[4]		X	X			X			X	X	X		X
Liver function tests[5]		X	X		X	X		X	X	monthly until 52 weeks exposure; every 12-weeks after 52 weeks exposure	X	(X)	X
Urine pregnancy test[6]										Monthly in between visits			
Efficacy assessments													
6MWT / Borg CR10 Scale®		X	X			X			X	X	X	X	
WHO FC		X	X		X	X		X	X	X	X	X	X
Clinical worsening		X	X		X	X		X	X	X	X	X	X
Ongoing participant review													
Concomitant therapy		X	X	X	X	X	X	X	X	X	X	X	X
SAEs/AEs		X	X	X	X	X	X	X	X	X	X	X	X
Optional sub-study													
Male reproductive system safety sub-study		Please refer to the schedule of assessments in Appendix 17, Section 4.											

Footnotes:

[1]The OL Day 1 visit (and associated assessments) will be combined with the EODBT visit for participants who remained on study intervention, or with PTOP 6 for participants who prematurely discontinued study intervention before Week 28 and completed the PTOP, had a CEC-confirmed CW event 28 and didn't meet any study-specific criterion for permanent discontinuation of study treatment.

[2] During the open-label extension period, an unscheduled visit must be performed in case of a suspected clinical worsening or initiation/dose escalation of PH-specific therapy. The assessments performed at the unscheduled visit include, but are not limited to: concomitant therapy, physical examination, vital signs (BP/PR) weight, post-dose 6MWT/Borg CR10 Scale® and WHO FC.

[3] EOLT will take place 2 years after the last participant has completed the fixed duration part of the DB period. In case of premature discontinuation of open-label study intervention, the EOLT visit must be performed within 7 days of taking the decision to stop study intervention.

[4] For women of childbearing potential. Additional serum pregnancy test to be performed in case of a positive urine pregnancy test.

[5] AST/ALT monitoring required monthly until Week 52 if interval between 2 regular site visits exceeds 4 weeks. After 52 weeks exposure in the OL period AST/ALT monitoring required every 12 weeks if interval between 2 regular site visits exceeds 12 weeks unless continued monthly monitoring is mandated based on local regulatory requirements . Local laboratory can be used. *Requirement for Japan:* Monthly AST/ALT monitoring is required. Whenever participants have monthly liver function tests, the following assessments should be conducted and reported in source notes: physical examination, body weight, blood pressure/pulse rate, and hematology (hemoglobin, hematocrit, and RBC count). These assessments must be recorded in the eCRF only if they are related to AEs.

[6] Women of childbearing potential only. To be performed monthly (at site or at home if the interval between 2 regular site visit exceeds 4 weeks).

[7] Phone calls to obtain information about vital status and hospitalization will be done approximately every 6 months until completion of the study for all participants who prematurely discontinued DB or OL study intervention.

Table 25: Visit and Assessment Schedule for Participants Entering the Post-treatment Observation Period (PTOP)

Period	POST TREATMENT OBSERVATION PERIOD (PTOP)								
Duration	Up to 28 Weeks								
Timeframe[2]	PTOP 1 (Safety FU visit)[1] 30 (+5) days after last dose	PTOP 2[4] W 12 (±5 d)	PTOP 3[4] W 16 (±5 d)	PTOP 4[4] W 20 (±5 d)	PTOP 5[4] W 24 (±5 d)	PTOP 6[4] W 28 (±5 d)	PTOP 7[4,6] W 40 (±5 d)	PTOP 8[4,6s] W 52 (±5 d)	Phone call W 28 (±5 d)[5]
Safety assessments									
Physical examination	X	X	X	X	X	X	X	X	
Vital signs (BP, PR)	X	X	X	X	X	X	X	X	
Body weight	X	X	X	X	X	X	X	X	
Clinical laboratory tests									
Central laboratory test[3]	X								
Liver function tests	X								
Efficacy assessments									
6MWT / Borg CR10 Scale®	X	X	X	X	X	X	X	X	
WHO FC	X	X	X	X	X	X	X	X	
Clinical worsening	X	X	X	X	X	X	X	X	
EQ-5D-5L©		X	X	X	X	X	X	X	
PAH-SYMPACT®		X	X	X	X	X	X	X	
PHQ-8, WPAI, SF36® v2 Acute			X			X		X	
Ongoing participant review									
Concomitant therapy	X	X	X	X	X	X	X	X	
SAEs/AEs	X	X	X	X	X	X	X	X	
Vital Status									X

Footnotes:

1 If PTOP 1 (safety follow-up visit) to be performed 30 (+5) days after intake of last dose. If PTOP 1 falls within the time-window of any other PTOP visit, the visits can be combined.

2 From Randomization.

3 Including Serum pregnancy test for women of childbearing potential.

4 PTOP visits to be performed depend on the timepoint of premature discontinuation.

5 Phone call obtain vital status only applies to participants who prematurely discontinued study intervention prior to Week 28 and chose not to enter or complete PTOP, provided consent was not withdrawn. Phone calls to obtain information about vital status and hospitalization will be done approximately every 6 months until completion of the study for all participants who prematurely discontinued DB or OL study intervention.

6 PTOP 7 and 8 visits are not applicable in current protocol version 6. However, participants who consented to an earlier version of the protocol will be given the choice to maintain the previous visit schedule and enter the OL period at Week 52, PTOP 7 and 8 visits may still be performed for such participants if they prematurely discontinued study intervention.

6MWT = 6-minute walk test; AE = Adverse Event; ALT = Alanine aminotransferase; AST = Serum aspartate aminotransferase; BP = Blood pressure; CTPA = computed tomography pulmonary angiogram; D = Day; ECG = Electrocardiogram; EODBT = End of double-blind treatment; EQ-5D-5L© = Euro Quality of life-5-Dimension-5-Level; FU = Follow-up; OL = open-label; MRA = magnetic resonance angiography; NT-proBNP = N-terminal prohormone of Brain natriuretic peptide or N-terminal pro B-type natriuretic peptide; PA = pulmonary angiography; PAH-SYMPACT® = Pulmonary Arterial Hypertension Symptoms and Impact; PFT = pulmonary function test; PHQ-8 = Patient Heath Questionnaire; PGA-S = Patient Global Assessment of Severity; PK = pharmacokinetic; PMDA = Pharmaceuticals and Medical Devices Agency; PR = Pulse rate; PTOP = Post-treatment Observation Period; RHC = right heart catheterization; SF36 = 36-item Short Form survey; SAE = Serious adverse event; V/Q = Ventilation / Perfusion; W = Week; WHO FC = World Health Organization functional class; WPAI = Work Productivity and Activity Impairment Questionnaire.

6.13. Appendix 13: COVID-19 and Regional Crisis

In the protocol Appendix 18, guidance is given on the conduct of the study during the COVID-19 pandemic. This also includes guidance specific to the protocol sections: study visits, treatment, laboratory assessments, PK, efficacy and safety assessments. The guidance was updated on 22 May 2022 (EDMS-RIM-734890, 1.0) to include guidance during natural disasters / major disruptions (eg, regional crisis) and pandemic (e.g., COVID-19).

COVID-19

In terms of handling safety and efficacy data, COVID-19 pandemic related protocol deviations will be captured and their impact on study safety and efficacy outcomes will be evaluated.

Statistical considerations to evaluate the impact of COVID-19 pandemic:

- Monitor missing data to collect reason for missingness due to COVID-19,
- Apply central statistical surveillance (CSS) to detect patters of issues/impact across sites and regions in an ongoing manner,
- Evaluate the quality of data collected at the time of the first IA (futility decision and impact on SSRE). Consider additional IA (SSRE) if COVID-19 impact is high i.e., due to heterogeneity of collected data,
- Perform sensitivity/descriptive analyses to assess the impact of COVID-19:
  - These may include re-evaluation of
    - estimand definitions
    - imputation methods for handling of missing values

    and/or excluding partial data such as visits, patients and/or centers.
- Evaluate the impact of mask usage when performing 6MWT.

Participant Disposition:

Following summary tables and listings will be added to the disposition reporting described in Section 5.2.1.

Summary tables:

- Participants who prematurely discontinued the study due to COVID-19 pandemic
- Reasons for termination of study, overall and by COVID-19 pandemic relatedness
- Participants who prematurely discontinued the study intervention due to COVID-19 pandemic
- Reasons for termination of study intervention, overall and by COVID-19 pandemic relatedness.

Exposure:

Study intervention compliance during the DB period will be calculated and descriptively presented by intervention group and COVID-19 Pandemic status.

Adverse Events:

Following will be tabulated in addition to those described in Section 5.5.2.

- AEs of interest for COVID-19 infection during DB period (by special interest category and preferred term)

The following AEs of interest for COVID-19 infection* will be summarized:

- Asymptomatic COVID-19,
- COVID-19,
- COVID-19 pneumonia,
- SARS-Cov-2 test positive,
- Suspected COVID-19.

*[Note: This list of COVID-19 preferred terms is derived from the latest MedDRA version 23.0 release date (1st of March 2020). Any modifications of terms based on later dictionary updates and/or external guidance will be documented in the DPS.]

A listing of COVID-19 Associated Adverse Events will be also provided on the SS set.

Subgroup analysis:

Subgroup analyses will be performed for the primary endpoint for exploratory purposes only.

Medical history:

Participants With COVID-19 Medical History (as collected in the specific eCRF form) will be listed and summarized based on the FAS.

Concomitant medication and therapies:

Concomitant medications and therapies used for COVID-19 infection during the DB period will be summarized and detailed listings will be provided.

Major Disruptions due to Regional Crisis

The impact of regional crisis on the collection and/or missingness of key study data and analyses will be evaluated. Enrollment into the study from the impacted countries (Russia and Ukraine) was stopped when only few participants (5) had been enrolled.

In terms of handling safety and efficacy data, regional crisis related protocol deviations will be captured and listed. Their impact on study safety and efficacy outcomes will be evaluated.

Discontinuations of study intervention and withdrawal for the study due to regional crisis will be summarized.

6.14. Appendix 14: Modelling and Simulation Report

Actelion Pharmaceuticals Ltd*
(a Janssen Pharmaceutical Company of Johnson & Johnson)

Simulation Report to Support

Study Protocol 67896062CTP3001; Phase 3

JNJ-67896062 (macitentan)

*Actelion Pharmaceuticals Ltd ("Actelion") is a global organization that operates through different legal entities in various countries. Therefore, the legal entity acting as the sponsor for Actelion studies may vary, such as, but not limited to Janssen-Cilag International NV; Janssen, Inc; Janssen Pharmaceutica NV; or Janssen Research & Development, LLC. The term "sponsor" is used throughout this document to represent these various legal entities; the sponsor is identified on the Contact Information page that accompanies the protocol

Status:	Approved
Version:	1.0
Date:	9 March 2021
Prepared by:	Actelion Pharmaceuticals Ltd, Janssen Research & Development, a division of Janssen Pharmaceutical NV
Document No.:	EDMS-RIM-368285

Compliance: The study described in this report was performed according to the principles of Good Clinical Practice (GCP).

TABLE OF CONTENTS

TABLE OF CONTENTS **2**

LIST OF TABLES **3**

LIST OF FIGURES **3**

VERSION HISTORY **4**

1. INTRODUCTION **5**

2. STUDY DESIGN **5**
2.1. Interim Analysis 5
2.2. Final Analysis 6

3. SIMULATION **6**
3.1. Simulation Specifications 7
3.2. Simulation Scenarios 7
3.2.1. Longitudinal approach 7
3.2.2. ANOVA approach 10
3.3. Data Generation 10
3.3.1. Longitudinal approach 10
3.3.2. ANOVA approach 12
3.4. Performance of simulations 12
3.5. Results 13
3.5.1. Results on Power, type I error rate and sample size 13
3.5.2. Results on point estimate and confidence interval 16
3.5.3. Results on key secondary endpoint 17

4. SUMMARY **19**

5. REFERENCES **20**

LIST OF TABLES

Table 1: Design Parameters 7
Table 2: Alternative Hypothesis Scenarios 8
Table 3: (Near) Null Hypothesis Scenarios 9
Table 4: Variance-covariance Matrix 10
Table 5: Treatment Effect in Change From Baseline for Primary Endpoint (in m) 10
Table 6: Hazard Ratio for the Key Secondary Endpoint 10
Table 7: Power, Futility Stops and ASN for Null Scenarios 13
Table 8: Power, Futility Stops and ASN for Alternative Scenarios 14
Table 9: Types of Increase, and Proportions Leading to Rejection (Non-futile Cases Only) 15
Table 10: Results of Interim Point Estimation 16
Table 11: Performance of Median Unbiased Estimator and Confidence Interval 17
Table 12: Rejection Probabilities for Key Secondary Endpoint 18

LIST OF FIGURES

Figure 1: Longitudinal Profiles Under Consideration 8

VERSION HISTORY

Version History Summary

Version	Approval Date	Change	Rationale
1	9 March 2021	Not Applicable	Initial release

1. INTRODUCTION

This appendix describes the simulations performed to obtain the operational characteristics of the primary analysis and key secondary analysis proposed for Macitentan in inoperAble or persistent/reCurrent chronIc ThromboEmbolic Pulmonary Hypertension (MACiTEPH).

2. STUDY DESIGN

MACiTEPH is a randomized, double-blind (DB), placebo-controlled, multicenter, two-stage, group sequential, adaptive study of macitentan 75 mg in men and women between 18 and 80 years of age, with a diagnosis of inoperable or persistent/recurrent CTEPH. At least 144 participants will be randomized in a 1:1 ratio to macitentan 75 mg or placebo. Treatment allocation will be stratified by

- inoperable (with or without BPA) vs persistent/recurrent after PEA surgery (with or without BPA), and
- use of PH-specific therapies (riociguat, other PH-specific therapy, none) at baseline.

The primary efficacy endpoint is change in 6MWD from baseline to Week 28. The secondary endpoint is a time to event endpoint with variable duration. A hierarchical testing strategy is applied.

The study uses an adaptive design with one interim analysis. At the interim analysis, the following decisions may be taken:

- stop the study early for futility, or
- keep sample size at originally planned number of 144 participants
- increase the sample size for the second stage of the study, up to a maximum of 230 participants.

The proposed sample size reassessment procedure can help mitigate the risk of misspecification of the treatment effect assumptions in any of the operability subgroups (inoperable/post-BPA vs. post-PEA), given the proof of concept study for macitentan in CTEPH (MERIT-1) was only conducted in inoperable patients.

The interim analysis will be conducted by an Independent Statistical Support Group (ISSG) and the results will be reviewed by an IDMC. See Clinical Protocol 67896062CTP3001 for further details.

2.1. Interim Analysis

The interim analysis (IA) is planned when a total of 72 patients have been have either completed Week 28 or dropped out early (but would have completed Week 28 had they continued), which corresponds to an information fraction $IF = 50\%$. Given the observation period of 28 weeks and under a recruitment rate of 1 patient per week, there will be 99 patients enrolled at the time of the IA: 72 with completed assessments and 27 with partial assessments only. The IA will be conducted

on 97 patients (two of the patients with partial assessments will have no post-baseline measurement).

Longitudinal data on 6MWD change from baseline will be analysed using an MMRM approach. The test statistic for the first interim analysis will be derived from the contrast for the treatment difference at Week 28. At the interim analysis, this test statistic is preliminary as not all participants will have complete data yet. It will not be used for hypothesis testing decisions but only for sample size reassessment.

Sample size reassessment will be based on the conditional power approach. The sample size needed will be calculated to maintain 80% conditional power at the final analysis, assuming the treatment effect estimate and its variability from the MMRM approach will be observed for the remainder of the trial. The new sample size is bounded from below by 144, the preplanned sample size, and from above by 230, the preplanned maximum number of participants.

2.2. Final Analysis

At the final analysis, two test statistics will be calculated, using the same approach as for the interim analysis. The first test statistic will be computed on all participants who already provided data for the interim analysis; this is the update mentioned before. All participants who did not provide data for the interim analysis will contribute to the second test statistics, irrespective of whether they were enrolled before or after the interim analysis.

The test statistics for both groups of participants will be combined using the inverse normal method (Lehmacher, Wassmer [1999]) with equal weights, corresponding to the prespecified information rate of 50% at the interim analysis.

The key secondary endpoint, the time to clinical worsening (TTCW), will be tested in a similar approach. Although there will be no interim analysis on the key secondary endpoints, care needs to be taken to protect the type I error rate as the number of clinical worsening events depends on the interim analysis results for the primary endpoint. Therefore, the inverse normal combination test approach will be employed for these endpoints as well, with the information rate as above and groups defined by timepoint of study entry (before/after interim analysis).

In addition to the hypothesis testing result, median-unbiased point and interval estimates for the primary endpoint will be provided based on an ordering of the sample space (Tsiatis, Rosner and Mehta [1987]).

3. SIMULATION

Two separate sets of simulations were performed:

1. Simulation for the primary endpoint only, using MMRM; these were performed in SAS (referenced as “longitudinal” from now on)
2. Simulation for the primary and secondary endpoints, using an ANOVA approach for the primary endpoint; these were performed in R (referenced as “ANOVA” from now on). This simplified approach to the primary analysis was chosen for reasons of simulation speed.

Simulations scenarios and data generation are therefore described separately in the following sections. Using both the ANCOVA and the MMRM approach will allow to assess the added value of utilizing longitudinal data for sample size reassessment. In addition to this assessment, the different approaches also serve to verify each other.

3.1. Simulation Specifications

All relevant design parameters are defined and specified in Table 1.

Table 1: Design Parameters

Parameter	Description	Value(s)
n_{plan}	Pre-planned total sample size	144
IF	Information fraction at the IA	40%, 50%, 60%
n_{IA}	Pre-planned sample size for the IA (stage 1)	58 (for $IF = 40\%$), 72 (for $IF = 50\%$), 86 (for $IF = 60\%$)
$n_{2,plan}$	Pre-planned sample size for stage 2	86 (for $IF = 40\%$), 72 (for $IF = 50\%$), 58 (for $IF = 60\%$)
n_{max}	Maximum total sample size	230
α	One-sided type I error rate	0.025
Δ	Targeted treatment effect (difference in mean 6MWD change from baseline at Week 28 between treatment and placebo groups)	33 m
SD	Standard deviation of 6MWD change from baseline at Week 28	70 m*
RR	Recruitment rate	1 patient/week
p_1	Proportion of subjects in inoperable/post-BPA subgroup	0.7

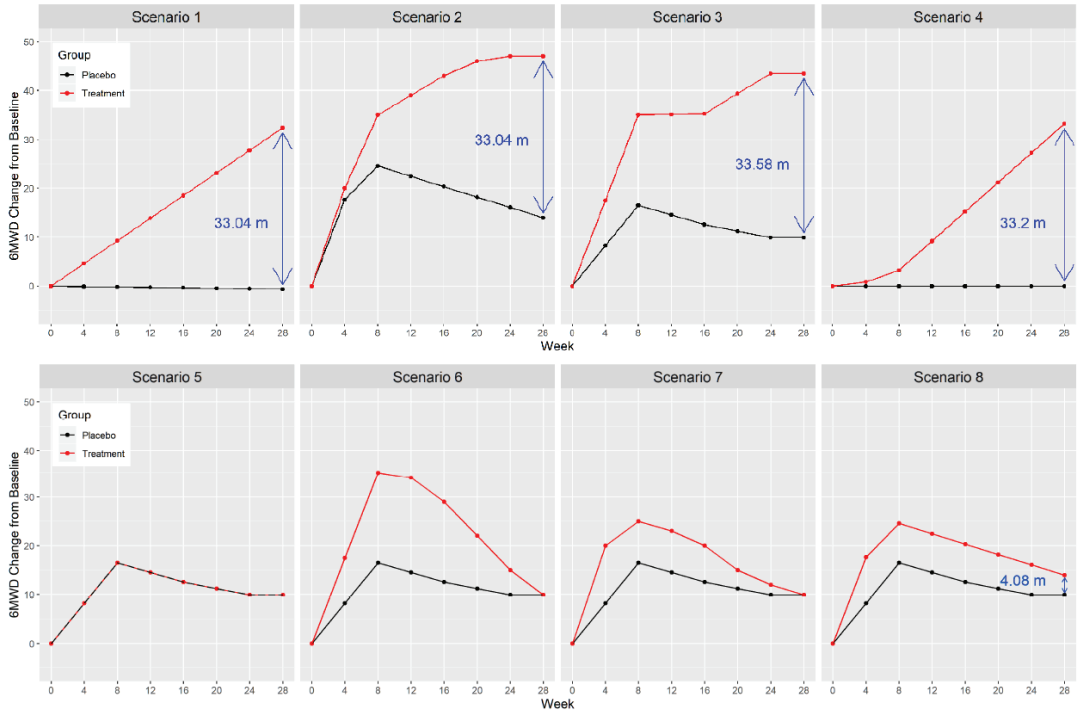
* as given in Table 4

3.2. Simulation Scenarios

3.2.1. Longitudinal approach

The scenarios considered for the simulations were based on 8 pairs of longitudinal profiles for 6MWD change from baseline for the 2 treatment arms shown in Figure 1.

Figure 1: Longitudinal Profiles Under Consideration



Scenarios 1, 2, 3 and 4 were chosen to reflect the protocol assumption of a treatment effect of $\Delta = 33$ m at Week 28 but differ in the shape of the longitudinal profiles. They are detailed in Table 2. Scenario 1 assumes a linear increase for the treatment arm and a slight linear decrease for the placebo arm. Scenario 2 was patterned after the average longitudinal profiles observed in CHEST-1 (Ghofrani 2013), while Scenario 3 was patterned after those observed in MERIT-1 (Ghofrani 2017), both adjusted to account for a longer treatment duration and to attain the targeted treatment effect. Scenario 4 has a flat time profile assumption for placebo and a linear increase for the treatment group starting from Week 8.

Table 2: Alternative Hypothesis Scenarios

Week	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	PBO	TRT	PBO	TRT	PBO	TRT	PBO	TRT
0	0	0	0	0	0	0	0	0
4	-0.1	4.62	17.68	20	8.268	17.5	0	0.8
8	-0.2	9.24	24.56	35	16.536	35	0	3.2
12	-0.3	13.86	22.44	39	14.548	35.125	0	9.2
16	-0.4	18.48	20.32	43	12.56	35.25	0	15.2
20	-0.5	23.1	18.2	46	11.222	39.359	0	21.2
24	-0.6	27.72	16.08	47	9.884	43.468	0	27.2
28	-0.7	32.34	13.96	47	9.884	43.468	0	33.2

Scenarios 5, 6, 7 and 8 were chosen to reflect various null scenarios resulting in no treatment effect at Week 28 (in one case, Scenario 8, a very small effect not worthwhile detecting). They are detailed in Table 3. Scenario 5 consists of the placebo arm profile in Scenario 3 and an identical

profile for the treatment arm. Scenarios 6 and 7 consider an initial treatment effect, which gradually reduces to a null effect at Week 28; this initial treatment effect was formulated to be larger in Scenario 7 than in Scenario 6.

Finally, Scenario 8 reflects a treatment effect at Week 28 that is smaller (4 m) than the targeted treatment effect. The placebo arm profile is the same as that in Scenario 3, while the treatment arm profile is the placebo arm profile in Scenario 2.

Table 3: (Near) Null Hypothesis Scenarios

Week	Scenario 5		Scenario 6		Scenario 7		Scenario 8	
	PBO	TRT	PBO	TRT	PBO	TRT	PBO	TRT
0	0	0	0	0	0	0	0	0
4	8.268	8.268	8.268	17.5	8.268	20	8.268	17.68
8	16.536	16.536	16.536	35	16.536	25	16.536	24.56
12	14.548	14.548	14.548	34	14.548	23	14.548	22.44
16	12.56	12.56	12.56	29	12.56	20	12.56	20.32
20	11.222	11.222	11.222	22	11.222	15	11.222	18.2
24	9.884	9.884	9.884	15	9.884	12	9.884	16.08
28	9.884	9.884	9.884	9.884	9.884	9.884	9.884	13.96

For the alternative hypothesis scenarios 1-4 in Table 2, different assumptions regarding the treatment effect in the operability subgroups (inoperable/post-BPA vs. post-PEA) were considered.

These subgroup-specific average longitudinal profiles per treatment arm were derived assuming:

1. an overall incidence of 70% for the inoperable/post-BPA subgroup and
2. the treatment effect in the inoperable/post-BPA subgroup is twice that in the post-PEA subgroup.

Letting $\boldsymbol{D}, \boldsymbol{D_1}$ and $\boldsymbol{D_2}$ respectively denote the overall, subgroup 1 (inoperable/post-BPA) and subgroup 2 (post-PEA) average longitudinal profile for a particular treatment arm, under the 2 assumptions above, the subgroup-specific profiles $\boldsymbol{D_1}$ and $\boldsymbol{D_2}$ can be obtained as:

$$\left.\begin{array}{r} 0.7\boldsymbol{D_1} + 0.3\boldsymbol{D_2} = \boldsymbol{D} \\ \boldsymbol{D_1} = 2\boldsymbol{D_2} \end{array}\right\} \quad \Rightarrow \quad \left\{\begin{array}{l} \boldsymbol{D_1} = 2\boldsymbol{D}/1.7 \\ \boldsymbol{D_2} = \boldsymbol{D}/1.7 \end{array}\right. \qquad (1)$$

For the null hypothesis scenarios, equal profiles were assumed for both subgroups. Using the unequal profile approach as detailed above was briefly investigated but showed minor impact on the results only. These results are therefore not presented in this simulation report.

For all scenarios described, the assumed variance-covariance matrix was formulated based on exploration of various longitudinal models fitted to the MERIT-1 data, adjusted to reflect a target standard deviation of $SD = 70$ m at Week 28. Table 4 provides the variance-covariance matrix used for all scenarios.

Table 4: Variance-covariance Matrix

	Week 4	Week 8	Week 12	Week 16	Week 20	Week 24	Week 28
Week 4	1088.61	1088.61	884.50	680.40	703.19	726.00	546.61
Week 8	1088.61	2177.22	1769.01	1360.79	1406.39	1451.99	1093.14
Week 12	884.50	1769.01	2886.43	2478.21	2145.71	1813.22	1474.40
Week 16	680.40	1360.79	2478.21	3595.63	2885.03	2174.44	1831.46
Week 20	703.19	1406.39	2145.71	2885.03	4219.25	3508.66	2109.94
Week 24	726.00	1451.99	1813.22	2174.44	3508.66	4842.89	2351.70
Week 28	546.61	1093.14	1474.40	1831.46	2109.94	2351.70	4900.95

3.2.2. ANOVA approach

Simulation scenarios are less variable for the analysis using ANCOVA as only the change from baseline for the primary endpoint and the hazard ratio and baseline median TTCW for the secondary endpoint need to be specified. Table 5 and Table 6 show assumptions for both endpoints; all 36 possible combinations will be investigated in Section 3.5.3. The primary endpoint scenarios from Table 5 will be referenced as ANOVA 1 through ANOVA 6. For results in Sections 3.5.1 and 3.5.2, results are averaged over key secondary endpoint scenarios.

Table 5: Treatment Effect in Change From Baseline for Primary Endpoint (in m)

Scenario	Inoperable	Post-PEA	Overall difference
1	0.00	0.00	0.00
2	19.40	0.00	5.82
3	0.00	19.40	13.58
4	19.40	19.40	19.40
5	0.00	38.80	27.16
6	19.40	38.80	32.98

Standard deviation for change from baseline is assumed as 70m throughout

For the key secondary endpoints, assumptions are uncertain. Therefore, a wide range will be investigated for the hazard ratio, ranging from reduction of the risk by 60% (HR=0.4) up to no effect (HR=1).

Table 6: Hazard Ratio for the Key Secondary Endpoint

Scenario	Hazard ratio	Median TTCW (months) PBO	Median TTCW (months) TRT
1	0.4	20	50
2	0.5	20	40
3	0.6	20	33.33
4	1	20	20

3.3. Data Generation

3.3.1. Longitudinal approach

The following section outlines the procedure used to generate the required data for a single simulated trial under a given scenario and for a fixed set of design parameters. This procedure was done in parallel to obtain 10,000 simulated trials per scenario for a fixed set of design parameters.

Under the proposed adaptive design, the maximum total sample size is $n_{max} = 230$. For a 1:1 randomization ratio, this implies that for each treatment arm, the maximum number of subjects is 230/2 = 115. Thus, for each subgroup and for each treatment arm, 115 full longitudinal profiles for 6WMD change from baseline at Weeks 4, 8, 12, 16, 20, 24 and 28 were generated from a multivariate normal distribution, with parameters corresponding to a particular scenario (see Table 2 and Table 3 for the mean vectors and Table 4 for the variance-covariance matrix) and pre-specified random seeds. Note that this data generation step provides the maximal required data for a single simulated trial but depending on the results of the SSR/futility rule at the IA, only a subset may be used. There were no drop-outs simulated for this effort.

For a given set of design parameters, n_{IA} subjects with complete profiles (to Week 28) are required for the IA. Under a recruitment rate of 1 subject per week, the implied timing of the IA is $t_{IA} = n_{IA} + 28$. To identify which subjects from the maximal generated data will be used for the IA, the number of subjects, $n_{IA,S1}$, in the inoperable/post-BPA subgroup (subgroup 1) at the time of the IA was simulated using a binomial distribution with parameters $\pi = 0.7$ and $m = n_{IA}$, with the latter possibly varying over different choices of information fraction (IF). The number of subjects in the post-PEA subgroup (subgroup 2) at the time of the IA was determined as:

$$n_{IA,S2} = n_{IA} - n_{IA,S1}.$$

Thus, $n_{IA,S1}$ subjects from subgroup 1 and $n_{IA,S2}$ subjects from subgroup 2 are identified for the IA. In addition to these, all subjects enrolled up to the time of the IA, t_{IA}, but with only partial profiles will also be used at the IA to assess futility and to re-estimate the sample size. This subset of subjects, with complete and partial profiles, is then taken from the maximal generated data and used for the IA.

For a simulated trial resulting in futility at the IA, no additional subjects are required, and the final analysis is performed on the same subset of subjects used at the IA, but all having completed profiles. For a non-futile simulated trial, however, the new sample size, n_{new} (≤ 230) is calculated at the IA and the corresponding recalculated sample size for stage 2 is:

$$n_2 = n_{new} - n_{IA},$$

which can further be formulated as $n_2 = n_{2,S1} + n_{2,S2}$, where $n_{2,S1}$ and $n_{2,S2}$ respectively denote the stage 2 sample sizes for the inoperable/post-BPA and post-PEA subgroups. To determine $n_{2,S1}$, a random number is generated from a binomial distribution with parameters $\pi = 0.7$ and $m = n_2$. Once $n_{2,S1}$ is identified, it follows that:

$$n_{2,S2} = n_2 - n_{2,S1} = n_{new} - n_{IA} - n_{2,S1}.$$

Thus, $n_{2,S1}$ subjects from subgroup 1 and $n_{2,S2}$ subjects from subgroup 2 are taken from remaining subjects in the maximal generated data and used for the stage 2 analysis.

The final analysis is based on the results obtained from the IA and the stage 2 analysis (see Section 2.2).

3.3.2. ANOVA approach

Data generation for the primary endpoint for the active treatment group participants will be simulated using a normal distribution with the mean set to the mean of the respective subgroup, while placebo participants will have simulated data based on a mean of 0 m.

Setup as to sample sizes, sample size per subgroup generally follows the approach described in Section 3.3.1 with the main difference that the sample size recalculation at the interim analysis uses participants with complete data only.

For the key secondary endpoint, exponential data will be simulated for both treatment groups, using λ_i = ln(2)/median to transform the median TTCW into a hazard rate for treatment group i. No difference is made between the strata for this endpoint. The intra-individual correlation between endpoints is unknown, and thus simulations are performed using values of ρ = 0.3; 0.5; 0.9. The correlation is taking into account while simulating data as follows:

Denote the variable simulated as primary endpoint by x. Then, a standardized normally distributed variable y is created via

$$y = \rho\left(\frac{x-\mu_i}{\sigma}\right) + \sqrt{1-\rho^2}u$$

where u is standard normal, independent of x, and μ_i denotes the primary endpoint mean per treatment group, with σ being the common standard deviation.

The TTCW variable z is then created by $z = \frac{-ln(1-\Phi(y))}{\lambda_i}$.

3.4. Performance of simulations

All simulations perform the following steps:

1. Use simulated data to calculate the interim test statistic depending on the information fraction specified
   a. In case of the longitudinal approach: Use the MMRM approach including partial data of those participants who did not reach study end before the interim analysis
   b. In case of ANOVA approach: Use only completed participants
   c. For the key secondary endpoint: Using logrank test
2. In case of an estimated treatment effect showing the wrong direction (i.e., harm): Stop for futility; else, go to 3.
3. Calculate the sample size needed for the second stage under consideration of the number of participants already recruited and the maximum number of participants based on the conditional power rule
4. Calculate second stage test statistic using the additional sample size as specified in 3. and perform the inverse normal combination test for the hypothesis test of the primary endpoint

a. In case of MMRM: For the first stage data, recalculate the test statistic using the now complete data from the participants used for the first stage

b. In case of MMRM: At the end of the trial, there are no missing data (no drop-outs have been simulated). Therefore, ANOVA on the final observation is equivalent to the MMRM and can be used to reduce simulation time.

5. Perform the inverse normal combination test for the key secondary endpoint (based on logrank test). This is performed in two different ways:

   a. All CW events are considered

   b. Only CW events up to the following time point are considered: 28 weeks after randomization of the 144th participant, irrespective of a sample size recalculation taking place

6. Calculate the 95% confidence interval and the median unbiased estimate for the primary endpoint

3.5. Results

3.5.1. Results on Power, type I error rate and sample size

Table 7 provides results for the null scenarios from both the longitudinal and the ANOVA approach. Included are also two near-null scenarios, namely scenario 8 for the MMRM (treatment effect of 4.08m) and scenario 2 for the ANOVA (treatment effect in the inoperable subgroup only; overall effect 5.82m). The type I error rate is preserved in all true null cases, with the observed conservativeness attributable to the stopping for futility boundary. For the near null scenarios, the rejection probability is higher than the targeted type I error rate, but it must not be forgotten that there is actually an effect there and they are therefore not true null scenarios.

The probability of stopping for futility is around 50% as expected for the true null scenarios. As to the sample size, the ASN under the null scenarios are between 5 and 25 participants in addition to the minimum number of participants. This calculation considers the futile cases, in which the number of participants stays equal to the preplanned number. Further details on the sample size is given further down in this section in Table 9.

Table 7: Power, Futility Stops and ASN for Null Scenarios

Scenario	Information rate	P(reject H0)	P(stop for futility)	ASN
Longitudinal 5	0.4	1.15	50.70	149.86
	0.5	1.13	50.68	158.41
	0.6	1.16	50.12	166.89
Longitudinal 6	0.4	1.23	49.46	151.28
	0.5	1.19	49.13	160.01
	0.6	1.19	49.76	167.04
Longitudinal 7	0.4	1.17	50.91	149.07
	0.5	1.22	50.83	158.13
	0.6	1.19	50.11	166.78
Longitudinal 8*	0.4	3.58	40.94	161.27
	0.5	3.58	40.30	169.35
	0.6	3.70	39.30	177.23

ANOVA 1	0.4	1.17	50.08	154.80
	0.5	1.16	49.87	162.77
	0.6	1.18	49.99	170.24
ANOVA 2*	0.4	4.88	37.40	168.89
	0.5	5.02	36.40	176.91
	0.6	5.06	35.27	184.11

* indicates near null scenario

The same characteristics are given in Table 8 below for the alternative scenarios. Those scenarios corresponding to the full effect size (treatment difference for change from baseline of approximately 33m) are indicated by an asterisk. The power is >85% for almost all simulations using the MMRM approach, and much the same for the ANOVA approach. Also the sample size is consistently around 173 participants for both approaches. For the ANOVA scenarios with less than full effect, the ASN goes up as the power goes down. Power is still in an acceptable region for the scenario where there is full effect in the post-PEA group and no effect in the inoperable group.

Table 8: Power, Futility Stops and ASN for Alternative Scenarios

Scenario	Information rate	P(reject H0)	P(stop for futility)	ASN
Longitudinal 1*	0.4	84.67	3.67	173.38
	0.5	86.94	2.30	173.40
	0.6	88.07	1.41	172.66
Longitudinal 2*	0.4	85.15	3.55	172.74
	0.5	88.02	2.29	173.00
	0.6	88.97	1.25	172.20
Longitudinal 3*	0.4	85.92	3.05	173.06
	0.5	87.45	2.18	172.93
	0.6	88.90	1.17	171.88
Longitudinal 4*	0.4	84.98	3.40	173.23
	0.5	87.12	2.19	173.05
	0.6	88.12	1.43	172.01
ANOVA 3	0.4	20.21	23.11	181.19
	0.5	21.00	20.47	188.55
	0.6	21.53	18.59	193.89
ANOVA 4	0.4	40.40	14.45	184.78
	0.5	41.80	11.79	190.07
	0.6	42.34	10.07	193.40
ANOVA 5	0.4	69.48	7.39	179.64
	0.5	71.48	5.26	182.12
	0.6	72.39	4.01	183.09
ANOVA 6*	0.4	85.68	3.72	173.8
	0.5	86.95	2.44	173.28
	0.6	88.19	1.50	171.89

*) indicates scenario with full effect

Table 9 gives the proportion of simulations leading to no sample size increase, maximal increase or in between values. Throughout the null scenarios, the majority of the cases shows an increase to the maximum. This is related to the futility bound, which stops the study only if the treatment effect goes in the wrong direction. For tANOVA, maximum increase is even more common. For the full effect scenarios, approximately 50% of the simulations do not require any sample size

increase and an increase to the maximum number was observed in <30% of the cases. MMRM and ANOVA perform similarly here. The in-between cases with less than full effect show intermediate results, as would be expected.

Table 9: Types of Increase, and Proportions Leading to Rejection (Non-futile Cases Only)

Scenario type	Scenario	Information rate	No increase (%) (of which rejections)	Submaximal increase (%) (of which rejections)	Maximal increase (%) (of which rejections)
Null	Longitudinal 5	0.4	7.12 (25.36)	9.29 (9.39)	83.59 (0.25)
		0.5	4.79 (34.32)	7.64 (14.06)	87.57 (2.20)
		0.6	3.26 (49.08)	6.72 (19.70)	90.02 (1.94)
	Longitudinal 6	0.4	7.70 (24.42)	9.60 (11.55)	82.71 (2.25)
		0.5	5.25 (31.46)	8.26 (12.14)	86.49 (2.49)
		0.6	3.78 (40.00)	7.15 (19.78)	89.07 (2.01)
	Longitudinal 7	0.4	7.78 (21.20)	9.80 (10.81)	82.42 (2.62)
		0.5	4.74 (30.04)	8.01 (17.26)	87.25 (2.59)
		0.6	3.81 (45.79)	6.67 (19.52)	89.52 (1.93)
	ANOVA 1	0.4	4.79 (20.11)	4.53 (10.91)	90.68 (0.99)
		0.5	3.53 (26.04)	3.77 (12.68)	92.69 (0.98)
		0.6	2.40 (37.40)	3.29 (15.77)	94.31 (1.00)
Full effect	Longitudinal 1	0.4	52.13 (94.46)	16.52 (90.76)	31.35 (86.16)
		0.5	53.04 (95.89)	17.55 (90.15)	29.41 (82.98)
		0.6	54.66 (97.01)	17.94 (91.86)	27.40 (77.01)
	Longitudinal 2	0.4	53.31 (94.52)	16.14 (90.56)	30.54 (86.83)
		0.5	54.07 (96.40)	16.77 (91.58)	29.16 (84.73)
		0.6	55.54 (96.90)	17.74 (91.67)	26.71 (79.19)
	Longitudinal 3	0.4	53.12 (94.70)	16.94 (89.46)	29.94 (86.67)
		0.5	54.31 (96.09)	16.98 (91.15)	28.71 (82.98)
		0.6	55.94 (96.87)	17.74 (91.44)	26.32 (78.28)
	Longitudinal 4	0.4	53.10 (94.64)	16.04 (88.96)	30.87 (86.02)
		0.5	53.69 (95.54)	17.66 (90.39)	28.66 (83.09)
		0.6	55.54 (96.93)	18.04 (91.06)	26.42 (77.34)
	ANOVA 6	0.4	53.74 (92.58)	13.44 (88.87)	32.82 (83.17)
		0.5	55.75 (94.15)	13.97 (88.98)	30.29 (79.95)
		0.6	58.17 (95.51)	14.24 (89.60)	27.59 (76.90)
Low effect	Longitudinal 8	0.4	10.41 (35.61)	10.57 (17.63)	79.02 (5.96)
		0.5	8.06 (43.66)	9.56 (25.92)	82.38 (4.90)
		0.6	6.01 (58.08)	8.95 (30.20)	85.04 (4.53)
	ANOVA 2	0.4	8.72 (31.84)	6.66 (19.30)	84.62 (4.48)
		0.5	7.19 (40.73)	6.29 (22.68)	86.53 (4.09)
		0.6	6.05 (50.39)	5.67 (28.59)	88.27 (3.56)
	ANOVA 3	0.4	18.07 (52.88)	9.90 (39.57)	72.03 (17.77)
		0.5	16.00 (60.55)	10.48 (43.85)	73.53 (16.48)
		0.6	15.10 (68.41)	10.43 (47.48)	74.47 (14.97)
	ANOVA 4	0.4	26.43 (68.17)	12.16 (57.58)	61.41 (36.16)
		0.5	25.28 (74.25)	12.78 (60.25)	61.94 (33.76)
		0.6	24.66 (79.36)	13.51 (62.13)	61.83 (30.93)
	ANOVA 5	0.4	42.20 (84.61)	13.51 (78.19)	44.29 (64.95)
		0.5	42.05 (88.07)	14.61 (80.16)	43.35 (61.60)
		0.6	42.48 (90.75)	15.19 (81.08)	42.33 (58.00)

3.5.2. Results on point estimate and confidence interval

There are two aspects of interest regarding point estimates here. Besides the final point estimate, also the interim estimate is of interest as the sample size reassessment can be expected to have better properties if the interim estimate is close to the true mean.

The first part of this section deals with the interim effect estimate. Table 10 has results for all groups of scenarios: Null, full effect and lower effect scenarios. The lower effect scenarios are not entirely comparable across methods, as the scenarios differ.

Table 10: Results of Interim Point Estimation

Scenario type	Scenario	True effect (m)	Information rate	Mean (est)	SD (est)
Null	Longitudinal 5	0	0.4	-0.35	18.17
			0.5	-0.32	16.22
			0.6	-0.23	14.92
	Longitudinal 6	0	0.4	0.16	18.33
			0.5	0.20	16.28
			0.6	0.10	14.92
	Longitudinal 7	0	0.4	-0.03	18.08
			0.5	-0.36	16.29
			0.6	-0.18	14.94
	ANOVA 1	0	0.4	0.00	18.52
			0.5	0.02	16.66
			0.6	-0.12	15.06
Full effect	Longitudinal 1	33.04	0.4	32.81	18.25
			0.5	32.71	16.36
			0.6	32.79	14.85
	Longitudinal 2	33.2	0.4	33.22	18.12
			0.5	33.33	16.60
			0.6	33.27	14.99
	Longitudinal 3	33.58	0.4	33.48	18.04
			0.5	33.39	16.35
			0.6	33.53	15.00
	Longitudinal 4	33.2	0.4	33.14	18.22
			0.5	33.19	16.45
			0.6	33.21	14.99
	ANOVA 6	32.98	0.4	33.03	18.56
			0.5	33.12	16.67
			0.6	33.18	15.27
Low effect	Longitudinal 8	4.078	0.4	4.06	17.99
			0.5	4.01	16.32
			0.6	3.97	15.01
	ANOVA 2	5.82	0.4	5.65	18.58
			0.5	5.72	16.77
			0.6	5.68	15.32
	ANOVA 3	13.58	0.4	13.80	18.68
			0.5	13.80	16.68
			0.6	13.84	15.33
	ANOVA 4	19.4	0.4	19.27	18.59
			0.5	19.42	16.52
			0.6	19.31	15.18
	ANOVA 5	27.16	0.4	27.56	18.89
			0.5	27.36	16.80
			0.6	27.20	15.33

Interim effect estimation can be seen to perform well across all scenarios and under both approaches. The standard deviation is also very similar across the MMRM and ANOVA approaches, reflecting the number of patients included in the interim analysis.

Table 11 shows the performance of the median unbiased estimator and the corresponding 95% confidence interval for the ANOVA case only.

Table 11: Performance of Median Unbiased Estimator and Confidence Interval

Scenario	True effect (m)	Information rate	Median (estimate)	Bias (estimate)	MAD (estimate)	CI coverage probability
ANOVA 1	0	0.4	-0.07	0.01	11.16	96.76
		0.5	0.03	0.09	10.43	96.52
		0.6	-0.11	-0.02	9.82	96.52
ANOVA 2	5.82	0.4	5.68	-0.04	10.35	96.85
		0.5	5.71	-0.10	9.69	96.78
		0.6	5.67	-0.20	9.22	96.69
ANOVA 3	13.58	0.4	13.48	-0.08	9.18	97.22
		0.5	13.45	-0.17	8.56	97.15
		0.6	13.44	-0.39	8.11	97.07
ANOVA 4	19.4	0.4	18.88	-0.67	8.24	97.24
		0.5	18.72	-1.00	7.54	97.28
		0.6	18.52	-1.42	6.98	97.33
ANOVA 5	27.16	0.4	25.75	-1.92	7.13	97.42
		0.5	25.24	-2.78	6.32	97.45
		0.6	24.39	-3.77	5.83	97.44
ANOVA 6	32.98	0.4	29.84	-3.88	6.44	97.31
		0.5	28.54	-5.33	6.30	97.38
		0.6	27.17	-6.67	6.97	97.39

Note: MAD=mean absolute deviation

The confidence interval attains its nominal level over all cases. Treatment effect estimation seems to be performing well with the exception of the full effect scenario, ANOVA 6. In this case, there is a consistent underestimation of the true difference of up to 6m.

3.5.3. Results on key secondary endpoint

As described above, primary and key secondary endpoint are tested in an a priori hierarchical fashion. The rejection probabilities for the key secondary endpoint therefore depends on the effect configuration of the primary endpoint. Table 12 gives rejection probabilities for both endpoints under the scenarios considered.

It is recognized that a participant-wise combination for time to event type endpoints in designs with adaptions based on short term endpoints may result in a type I error inflation in specific situations (see Magirr 2016). Therefore, in addition, a stress test scenario for evaluating the type I error rate was considered assuming effects of 100m for the primary endpoint, and a correlation of 0.9. This scenario will practically result in 100% power for the primary endpoint, hence the secondary endpoint in the hierarchical testing strategy will be always tested with no false negative decisions for the primary endpoint being made. The chance of rejecting the secondary endpoint hypothesis is known to increase with correlation (see also Table 12). Hence, the stress test scenario ensures that the type I error rate under the null scenarios are not underestimated. The stress test

observed type I error rates were 2.369% for the unrestricted approach, and 1.826% for the restricted approach.

In summary, the combination test with unrestricted follow-up does not result in type I error rate inflation in this setting and can be considered for the secondary endpoint analysis.

Table 12: Rejection Probabilities for Key Secondary Endpoint

	TTCW hazard ratio	Correlation	Rejection probability (all events considered)	Rejection probability (restriction on events considered)
ANOVA 1	0.4	0.3	1.05%	1.16%
		0.5	1.15%	1.16%
		0.9	1.05%	1.37%
	0.5	0.3	1.03%	1.1%
		0.5	1.23%	1.13%
		0.9	1.22%	1.26%
	0.6	0.3	0.97%	0.81%
		0.5	1.16%	0.98%
		0.9	1.13%	0.99%
	1	0.3	0.06%	0.07%
		0.5	0.23%	0.11%
		0.9	0.38%	0.44%
ANOVA 2	0.4	0.3	4.98%	4.61%
		0.5	4.96%	5.21%
		0.9	5.09%	4.61%
	0.5	0.3	4.87%	4.47%
		0.5	4.58%	4.93%
		0.9	4.94%	4.99%
	0.6	0.3	3.97%	3.68%
		0.5	4.45%	4.31%
		0.9	5.07%	5.02%
	1	0.3	0.34%	0.31%
		0.5	0.48%	0.39%
		0.9	1.05%	1.15%
ANOVA 3	0.4	0.3	20.4%	20.44%
		0.5	20.5%	20.9%
		0.9	20.69%	21.02%
	0.5	0.3	19.5%	19.09%
		0.5	19.28%	19.36%
		0.9	20.18%	20.41%
	0.6	0.3	15.51%	15.22%
		0.5	17.57%	16.48%
		0.9	19.97%	19.81%
	1	0.3	0.68%	0.88%
		0.5	1.22%	1.14%
		0.9	1.85%	1.95%
ANOVA 4	0.4	0.3	40.97%	40.71%
		0.5	40.96%	41.72%
		0.9	40.58%	42.74%
	0.5	0.3	37.35%	36.32%
		0.5	38.68%	37.9%
		0.9	40.71%	40.45%

	TTCW hazard ratio	Correlation	Rejection probability (all events considered)	Rejection probability (restriction on events considered)
	0.6	0.3	29.1%	28.8%
		0.5	31.51%	30.28%
		0.9	38.91%	36.94%
	1	0.3	1.38%	1.4%
		0.5	1.74%	1.85%
		0.9	2%	1.99%
ANOVA 5	0.4	0.3	69.4%	69.14%
		0.5	70.31%	70.54%
		0.9	69.18%	71.31%
	0.5	0.3	61.66%	60.53%
		0.5	63.87%	62.53%
		0.9	68.26%	67.72%
	0.6	0.3	46.29%	45.03%
		0.5	49.22%	47.73%
		0.9	55.67%	53.26%
	1	0.3	1.82%	1.92%
		0.5	1.92%	1.7%
		0.9	2.22%	2.05%
ANOVA 6	0.4	0.3	83.54%	83.13%
		0.5	84.75%	84.35%
		0.9	83.73%	85.61%
	0.5	0.3	72.43%	70.71%
		0.5	74.74%	73.01%
		0.9	79.83%	77.29%
	0.6	0.3	52.92%	51.48%
		0.5	54.37%	53.14%
		0.9	58.73%	56.41%
	1	0.3	2.24%	2.17%
		0.5	2.12%	2.31%
		0.9	2.32%	2.41%

4. SUMMARY

In summary, both approaches considered, MMRM and ANOVA, have the desired statistical inference properties. Both the specified type I error level and the targeted power are obtained under the appropriate scenarios. Interestingly, there is no big difference between the two approaches regarding the reassessed sample size, as the point estimates are performing equally well.

The MMRM simulations for the scenarios with an intermediate effect and no treatment effect at the final analysis actually show a slight overestimation of the treatment effect, which, in this situation, is beneficial – there is little probability for a final success and so the sample size increase is kept limited.

Regarding the key secondary endpoint, it is of course highly dependent on the performance of the primary test. There is sufficient power for an HR of 0.4 such that the power for the key secondary endpoint is not much less than the power for the primary endpoint in these cases. In case of a true HR of 0.6, the power does not exceed 60% for any scenario considered.

These simulations do not consider missing values except those missing by design at the interim analysis (which will then be available at the final endpoint). The MMRM can therefore not show its full potential compared to ANOVA. As there is no downside against the ANOVA approach, the MMRM should therefore be used for the primary analysis.

5. REFERENCES

Ghofrani HA, D'Armini AM, Grimminger F, Hoeper MM, Jansa P, Kim NH, et al., for the CHEST-1 Study Group. Riociguat for the treatment of chronic thromboembolic pulmonary hypertension. N Engl J Med. 2013;369:319-329.

Ghofrani HA, Simonneau G, D'Armini AM, Fedullo P, Howard LS, Jais X, et al., on behalf of the MERIT study investigators. Macitentan for the treatment of inoperable chronic thromboembolic pulmonary hypertension (MERIT-1): results from the multicentre, phase 2, randomised, double-blind, placebo-controlled study. Lancet Respir Med. 2017;5:785-794.

Magirr D, Jaki T, König F, Posch M. Sample size reassessment and hypothesis testing in adaptive survival trials. PLoS ONE. 2016;11(2):e0146465.

Lehmacher W, Wassmer G. Adaptive sample size calculations in group sequential trials. Biometrics. 1999;55;1286-1290.

Tsiatis AA, Rosner GL, Mehta CR. Exact confidence intervals following a group sequential test. Biometrics. 1987:40;797-803.

6.15. Appendix 15: Hemodynamic Variables

Table 26: Hemodynamic Variables Collected/Derived in the eCRF

Variables	Unit	Formula	Source/Derivation
Pulse/heart rate	Beats/min		Collected in eCRF
Systolic Systemic Arterial Pressure (sSAP) [peripheral SBP]	mmHG		Collected in eCRF
Diastolic systemic Arterial Pressure (dSAP) [peripheral DBP]	mmHG		Collected in eCRF
Pulmonary Artery Wedge Pressure (PAWP)	mmHG		Collected in eCRF
Left Ventricular End-Diastolic Pressure (LVEDP)[1]	mmHG		Collected in eCRF
Systolic Pulmonary Arterial Pressure (sPAP)	mmHG		Two measurements collected in eCRF, average derived in eCRF
Diastolic Pulmonary Arterial Pressure (dPAP)	mmHG		Two measurements collected in eCRF, average derived in eCRF
Mean Pulmonary Arterial Pressure (mPAP)	mmHG	((1 * sPAP) + (2 * dPAP))/3	Derived in eCRF
Mean Right Atrial Pressure (mRAP)	mmHG		Collected in eCRF
Cardiac Output (CO)	L/min		Three measurements collected in eCRF, average derived in eCRF (thermodilution method). One measurement for Fick method
Mixed venous Oxygen Saturation (SvO_2)	%		Collected in eCRF
Pulmonary Vascular Resistance (PVR)	Wood Units (calculated) $dyn \cdot sec/cm^5$	([mPAP – PAWP] / CO) ([mPAP – PAWP] / CO) * 80	Derived in eCRF
Cardiac Index (CI)	$L/min/m^2$	CO /BSA	Derived in eCRF
Total Pulmonary Resistance (TPR)	$dyn \cdot sec/cm^5$	(mPAP / CO) * 80	Derived in eCRF

[1]Only in the event that PAWP cannot be measured reliably it will be entered as 'NO' for the 'Was PAWP measured and deemed reliable?' field on the eCRF, and LHC may be required to obtain the LVEDP. In this case, LVEDP will replace PAWP in the calculation of PVR.

7. REFERENCES

(Brannath 2006) Brannath W, Konig F, Bauer P. Estimation in flexible two stage designs. Stat Med. 2006;25:3366-3381.

(Collett 1994) Collett D. Modelling survival data in medical research. London: Chapman & Hall; 1994.

(FDA 2020) FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID -19 Pandemic: Guidance for Industry, Investigators, and Institutional Review Boards. March 2020.

(Ghofrani 2013) Ghofrani HA, D'Armini AM, Grimminger F, et al. Riociguat for the Treatment of Chronic Thromboembolic Pulmonary Hypertension. N Engl J Med. 2013;369:319-29.

(Ghofrani 2017) Ghofrani HA, Simonneau G, D'Armini AM, et al; MERIT study investigators. Macitentan for the treatment of inoperable chronic thromboembolic pulmonary hypertension (MERIT-1): results from the multicentre, phase 2, randomised, double-blind, placebo-controlled study. Lancet Respir Med. 2017 Oct;5(10):785-794. doi: 10.1016/S2213-2600(17)30305-3. Epub 2017 Sep 11.

(Hettmansperger 2011) Hettmansperger TP, McKean JW. Robust Nonparametric Statistical Methods. 2nd ed. Boca Raton, LA: CRC Press; 2011.

(Hoeper 2017) Hoeper MM, Kramer T, Pan Z, et al. Mortality in pulmonary arterial hypertension: prediction by the 2015 European pulmonary hypertension guidelines risk stratification model. Eur Respir J. 2017;50(2).

(Jaïs 2008) Jaïs X, D'Armini AM, Jansa P, et al. Bosentan for treatment of inoperable chronic thromboembolic pulmonary hypertension: BENEFiT (Bosentan Effects in iNopErable Forms of chronIc Thromboembolic pulmonary hypertension), a randomized, placebo-controlled trial. J Am Coll Cardiol. 2008;52:2127-34.

(Jenkins 2016) Jenkins DP, Biederman D, D'Armini AM, et al. Operability assessment in CTEPH: Lessons from the CHEST-1 study. J Thorac Cardiovasc Surg. 2016;152(3):669-674.

(Jörgens 2019) Jörgens S, Wassmer G, Konig F, Posch M. Nested combination tests with a time-to-event endpoint using a short-term endpoint for design adaptations, Wiley 2019 DOI: 10.1002/pst.1926

(Kim 2017) Kim NH, D'Armini AM, Grimminger F, et al. Haemodynamic effects of riociguat in inoperable/recurrent chronic thromboembolic pulmonary hypertension. Heart 2017;103:599-606.

(Kloke 2012) Kloke JD, McKean JW. Rfit: Rank-based estimation for linear models. The R journal. 2012;4(2):57-64.

(Kroenke 2001) Kroenke K, Spitzer RL, Williams JBW. The PHQ-8. Validity of a brief depression severity measure. J Gen Intern Med. 2001;16:606-613.

(Lehmacher 1999) Lehmacher W, Wassmer G. Adaptive sample size calculations in group sequential trials. Biometrics. 1999;55:1286-1290.

(Mallinckrodt 2008) Mallinckrodt CH, Lane PW, Schnell D, et al. Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials. Drug Inf J. 2008;42:303-319.

(Magirr 2016) Magirr D, Jaki T, König F, Posch M. Sample size reassessment and hypothesis testing in adaptive survival trials. PLoS ONE. 2016;11(2):e0146465.

(Mathai 2012) Mathai SC, Puhan MA, Lam D, et al. The Minimal Important Difference in the 6-Minute Walk Test for Patients with Pulmonary Arterial Hypertension, Am J Respir Crit Care Med. 2012;186(5):428-433.

(NICE 2018) Position statement on use of the EQ-5D-5L valuation set for England (updated November 2018). Accessed on 30 April 2019. Available at: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l

(R Core Team 2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria. Available online at https://www.R-project.org/ (Accessed on 10 Mar 2021)

(Tsiatis 1987) Tsiatis AA, Rosner GL, Mehta CR. Exact confidence intervals following a group sequential test. Biometrics. 1987;40:797-803.

(Van Hout 2012) Van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health. 2012;15(5):708-715.