Statistical Analysis Plan H6D-MC-LVHV (v4)

A Double-Blind Efficacy and Safety Study of the Phosphodiesterase Type 5 Inhibitor Tadalafil in Pediatric Patients with Pulmonary Arterial Hypertension

NCT01824290

Approval Date: 28-Mar-2019

1. Statistical Analysis Plan H6D-MC-LVHV: A Double-Blind Efficacy and Safety Study of the Phosphodiesterase Type 5 Inhibitor Tadalafil in Pediatric Patients with Pulmonary Arterial Hypertension

Tadalafil (LY450190) Pulmonary Arterial Hypertension

Study H6D-MC-LVHV (LVHV) is a Phase 3, international, randomized, multicenter, 2-period, double-blind, placebo-controlled (Period 1), add-on (in addition to the patient's current endothelin receptor antagonist, [ERA]) study to evaluate tadalafil efficacy, safety, and population pharmacokinetics (PK) in pediatric patients with pulmonary arterial hypertension (PAH).

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol H6D-MC-LVHV
Phase 3

Statistical Analysis Plan Version 1 electronically signed and approved by Lilly: 08 February 2017
Statistical Analysis Plan Version 2 electronically signed and approved by Lilly: 13 April 2017
Statistical Analysis Plan Version 3 electronically signed and approved by Lilly: 07 March 2019
Statistical Analysis Plan Version 4 electronically signed and approved by Lilly on date provided below

Approval Date: 28-Mar-2019 GMT

2. Table of Contents

Section **Page**

Table of Contents

Figure **Page**

Table of Contents

3. Revision History

Statistical Analysis Plan Version 1 was originally approved on 25 June 2013 with a later approval date of 08 February 2017 to allow appropriate storage of the PDF version.

Statistical Analysis Plan Version 2 was approved on 17 April 2017 prior to unblinding of the Study LVHV for efficacy interim analysis. The changes include:

- Added a minor clarification in Section 6.1 (General Considerations) for descriptive summary statistics to be provided for Period 1 and 2.
- Added definition for Period 2 population in Section 6.1.1.
- Removed items (prior pulmonary arterial hypertension [PAH[therapies, concomitant medications, screening labs) that will not be included in tables/listings of baseline disease characteristics from Section 6.7.
- Corrected definition of time to clinical worsening (CW) as being calculated relative to date of first dose (rather than randomization) in Section 6.9.2.1.
- Added analysis for Period 2 of change in World Health Organization (WHO) functional class to Section 6.9.3.2.
- Replaced eccentricity index with more specific terms in Section 6.9.3.3.
- Added statement specifying the denominator and percentage computation for gender-specific events in Section 6.11.2.
- Added new Section 6.11.4 describing the analysis of special adverse event follow-ups.
- Removed items that will not be summarized for the Data Monitoring Committee (DMC) from Section 6.14.
- Removed details concerning conduct of the efficacy interim analysis (eg, timing, enrollment, power) from Section 6.14.
- Added reference to Appendix 3 to Section 6.14.
- Added Appendix 3: Analysis Plan for Efficacy Interim Analysis.
- Removed details concerning conduct of the efficacy interim analysis in Section 7 to be consistent with Interim Analyses and Data Monitoring, Section 6.14.
- Made minor grammatical, spelling or formatting corrections as needed.

Statistical Analysis Plan Version 3 was approved on 07 March 2019 prior to unblinding of Study LVHV. Statistical Analysis Plan Version 3 was based on Protocol Amendment C approved by Lilly on 13 December 2018. The changes include:

- Revised some of the above bullet points in Section 3 due to re-numbering of some section because of changes in Statistical Analysis Plan.

- Revised Section 4.1.1 for primary objectives in Period 1. The analysis of 6-minute walk (6MW) is the only primary objective and removed EU assessment of time to CW.
- Revised Section 4.2.1 for secondary objectives in Period 1. Removed EU regulatory evaluation of 6MW and kept time to CW analysis.
- Revised Section 5.1 study design for sample size justification.
- Section 6.1: removed treatment comparison p-values and added that only confidence interval will be reported.
- Section 6.1: added language for pooling PAH etiology level with less than 1 patient per treatment group
- Section 6.1: added language to account for small sample size for when using mixed-effects model approach (MMRM).
- Section 6.1.1: clarified language of analysis patient population
- Revised Section 6.5 to remove no adjustments for multiple comparisons since there is only 1 primary objective.
- Revised Section 6.7 to remove concomitant from "baseline disease characteristics and therapies" summary.
- Revised Section 6.9.1, "Primary Outcome and Methodology" to plan for potential convergence issue, and moved previous 6.9.1.2 section "Time to clinical worsening" from primary section to Secondary Efficacy Analyses, Section 6.9.2.
- Revised 6.9.1.1 MMRM analysis to account for the small sample size.
- Added Section 6.9.1.2 sensitivity analysis for 6MW
- Updated 6.9.2 secondary efficacy analysis for CW and moved some content from Section 6.9.1.
- Removed Section 6.9.3.1 additional analysis of 6MW regarding missing data, and shifting the next corresponding section numbers accordingly.
- Revised Section 6.9.3.2 to add no formal treatment comparison for WHO function analyses
- Revised Section 6.9.3.3 due to small sample size, replaced MMRM with analysis of covariance (ANCOVA) for echocardiogram and removed between treatment p-values reported.
- Revised Section 6.9.3.4: removed between treatments p-values reported for magnetic resonance imaging (MRI).
- Revised Section 6.9.3.5 to add no formal treatment comparison for Clinical Global Impression of Improvement (CGI-I).

- Revised Section 6.9.3.6 due to small sample size, replaced MMRM with ANCOVA for N-terminal prohormone brain natriuretic peptide (NT-Pro-BNP) and removed between treatments p-values reported.
- Revised 6.9.3.7: No p-values reported from Child Health Questionnaire Parent Form 28 (CHQ-PF28) analysis.
- Revised Section 6.11.2 to add no treatment comparison for adverse events related analysis.
- Revised Section 6.11.5.1 to add no p-values reported for clinical laboratory test.
- Revised Section 6.11.6 to add no p-values reported for vital sign analyses.
- Revised Section 6.12 to indicate only descriptive summary provided for subgroup analysis of 6MWD due to small sample size.
- Revised Section 6.18 analysis requirements for the Japanese addendum since there are only 2 Japanese patients.
- Added Appendix 4, Analysis Plan for Extrapolation
- Made minor grammatical, spelling, or formatting corrections as needed.

Statistical Analysis Plan Version 4 was approved on (see date on cover page) prior to unblinding of Study LVHV.

- Revised Section 6.9.1.1 to change 95% CI to 80% CI for primary analysis to be align with the sample size calculation in Section 5.1.
- Added language to Section 6.9.3.4 to provide only listing if there are less than 3 patients per treatment arm.

4. Study Objectives

4.1. Primary Objective

4.1.1. Period 1

The primary objective of Period 1 is to evaluate the efficacy of tadalafil compared with placebo in improving 6-minute walk (6MW) distance from Baseline to Week 24, as assessed in a subset of patients ≥6 to <18 years of age who are developmentally capable of performing a 6MW test.

4.1.2. Period 2

The primary objective of Period 2 is to evaluate long-term safety of tadalafil while providing continued access to tadalafil for pediatric patients with pulmonary arterial hypertension (PAH) who participated in Period 1.

4.2. Secondary Objectives

4.2.1. Period 1

The secondary objectives of Period 1 are as follows:

- Assess the efficacy of tadalafil compared with placebo on time to clinical worsening (CW) and the incidence of CW.
- Characterize the population pharmacokinetics (PK) of tadalafil in pediatric PAH patients.
- Assess the safety of tadalafil compared with placebo.

4.2.2. Period 2

The secondary objective of Period 2 is to evaluate the incidence of, and time to CW.

4.3. Additional Objectives

4.3.1. Period 1

Additional objectives of Period 1 are as follows:

- Assess the efficacy of tadalafil compared with placebo on changes in World Health Organization (WHO) functional classification.
- Explore by cardiac magnetic resonance imaging (MRI), changes from Day 1 to Week 24 in the following cardiac MRI parameters:
    - Left-ventricular (LV) ejection fraction
    - Right-ventricular (RV) end diastolic volume
    - Right-ventricular end systolic volume
    - Right-ventricular ejection fraction

- Evaluate by echocardiography, changes from Day 1 to Week 24 in the following echocardiographic parameters:
  - tricuspid annular plane systolic excursion (TAPSE)
  - eccentricity index
  - pericardial effusion
  - maximal tricuspid regurgitant velocity
- Evaluate change from Day 1 to Week 24 in N-terminal prohormone brain natriuretic peptide (NT-Pro-BNP) concentrations.
- Assess physician- and caregiver-reported health outcome, as measured by Clinical Global Impression of Improvement (CGI-I), and in a subset of patients ≥5 years of age, Child Health Questionnaire Parent Form 28 (CHQ-PF28).

5. Study Design

5.1. Summary of Study Design

This is a Phase 3, international, randomized multicenter, 2-period, double-blind (Period 1), placebo-controlled (Period 1), add-on (ie, in addition to the patient's current endothelin receptor antagonist [ERA]) study to evaluate the efficacy, safety, and population PK of tadalafil in pediatric patients with PAH.

Study H6D-MC-LVHV (LVHV) will enroll pediatric PAH patients ≥6 months to <18 years of age with WHO functional class II or III and who are already receiving treatment with an ERA. Patients will be randomized to receive either placebo or active drug in a 1:1 ratio, based on weight cohort, PAH etiology, and type of ERA. Patients will receive study treatment for 6 months in the double-blind period (Period 1), and then will be eligible to enroll into an open label 2-year extension period (Period 2) during which all patients will receive tadalafil.

It is anticipated that 34 patients will be randomly assigned to treatment in Period 1 of this study and approximately 2 patients without 6MW test postbaseline. Therefore, a sample size of 32 randomized patients who could perform 6MW test will provide 71% power to detect a placebo-adjusted mean difference in change in 6MW distance (6MWD) at Week 24 of 40 meters with a standard deviation (SD) of 60 meters and a 2-sided significance level of 0.2. To achieve a representative distribution of patients' ages, enrollment will be monitored throughout the study to achieve ≥30% of all patients <12 years of age.

Patients entering the study will be stratified into 1 of 3 weight-cohorts, based on the patient's weight at the time of the Screening visit:

Heavy-weight: ≥40 kg

Middle-weight: ≥25 kg to <40 kg

Light-weight: <25 kg

If a patient's weight changes during Period 1, such that he/she falls into a different weight cohort, he/she will continue to receive the study drug dose appropriate to his/her original weight cohort.

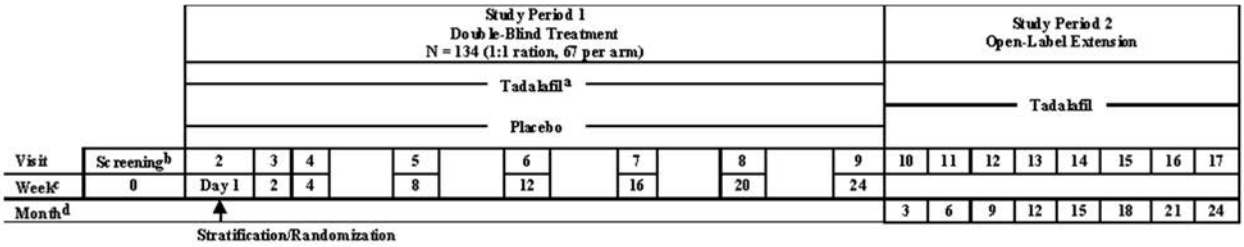
Patients will also be stratified by type of ERA (bosentan or other) and PAH etiology.

If a patient will be participating in Period 2, and if that patient's weight changes at the conclusion of Period 1 (at the Week 9 or Early Termination visit) or during Period 2, such that he/she falls into a different weight cohort (defined as at least 1 kg above or below the weight cohort thresholds of 25 kg and 40 kg), then the patient's dose of study drug may be adjusted so that he or she receives the appropriate weight cohort-related dose.

Dose selection for this study will be based on pediatric PK and safety data from Study H6D-MC-LVIG (LVIG) and the PK and safety data from the adult PAH development plan. The selected dose for each weight cohort will reflect exposures comparable to the approved 40-mg dose of

tadalafil in adults, unless unexpected safety concerns unique to the pediatric population are revealed.

The study design for protocol H6D-MC-LVHV is illustrated in Figure LVHV.5.1.



Visit	Screening[b]	2	3	4		5		6		7		8		9	10	11	12	13	14	15	16	17
Week[c]	0	Day 1	2	4		8		12		16		20		24								
Month[d]															3	6	9	12	15	18	21	24

a Final dose to be determined after the cohort completion in Study H6D-MC-LVIG.

b Screening period is days -28 to 0.

c Weeks = ±7days.

d Months = ±10 days. Month 3 is 3 months from Visit 9; all other months (6, 9, 12, 15, 18, 21, and 24) are in relation to Visit 10.

Figure LVHV.5.1. Illustration of study design for Protocol H6D-MC-LVHV.

6. A Priori Statistical Methods

6.1. General Considerations

The statistical analyses for this study are the responsibility of Eli Lilly and Company. Given the small sample size, no formal comparison will be made between treatment groups. Hence, in place of the overall treatment difference p-value and the visitwise p-values, the corresponding 95% confidence intervals (CIs) will be reported. Unless otherwise specified, analyses will be provided for the double-blind period only and across both the double-blind and open-label periods.

Descriptive summary statistics for continuous measurements will include the number of nonmissing observations, mean, standard deviation (SD) or standard error, median, and minimum, and maximum values by each treatment group for Period 1. Similar summary for Period 2 will be conducted overall and by treatment assignment during Period 1. The 25th percentile, median, and the 75th percentile will be presented for variables that are analyzed using rank-transformed data. As the study has a limited number of patients enrolled, p-value will not be reported. Categorical variables will be summarized with counts and percentages for each category.

The PAH etiology level with less than 1 patient per treatment group will be pooled together with a PAH etiology level with next smallest number of patients per treatment. If this results in a combined PAH etiology level still having less than 1 patient per treatment group, this data will be pooled together with the data from the next smallest PAH etiology level, if one exists, otherwise; no further pooling is needed. The pooled PAH etiology will be included in all analyses. The actual PAH etiology level will be included in the listing.

For the mixed-effects model approach (MMRM) models, the effect of pooled PAH etiology will only be included if there are at least 3 patients per treatment arm at each pooled PAH etiology level. Similarly, the effect of ERA therapy will only be included if there are at least 3 patients per treatment arm at each ERA therapy level. As the study has a limited number of patients enrolled, an inspection of the model mean change from baseline and corresponding CIs will be used in place of significance testing to ensure whether they are trending in the right direction. Hence, the overall treatment difference p-value and the visitwise p-values will not be reported.

Any changes to the data analysis methods described in the protocol will require an amendment only if it changes a principal feature of the protocol. Any other change to the data analysis methods described in the protocol, and the justification for making the change, will be described in the clinical study report (CSR). Additional exploratory analyses of the data may be conducted as deemed appropriate.

6.1.1. Definitions of Analysis Population

Efficacy analyses will be conducted on the Primary Analysis Population. This population includes all data from all randomized patients who receive at least 1 dose of the study drug

according to the randomized treatment. Patients with baseline and at least 1 postbaseline data for a particular efficacy endpoint will be included in the analysis of that endpoint.

The 6MW Analysis Population will include the subset of randomized patients ≥6 to <18 years of age who take at least 1 dose of study medication and were capable of performing a 6MW test.

Safety analyses will be conducted on the Primary Analysis Population.

Analyses for Period 2 will only include patients who entered Period 2.

6.2. Adjustments for Covariates

Randomization at Visit 2 (Day 1) will be stratified by the following variables:

- weight cohort (Heavy-weight: ≥40 kg; Middle-weight: ≥25 kg to <40 kg; Light-weight: <25 kg)
- endothelin receptor antagonist medication (bosentan or other)
- pulmonary arterial hypertension etiology (idiopathic, connective tissue disease, anorexigen use, and associated to surgical repair)

These stratification factors, in addition to the baseline value of the analysis variable, will be included as covariates in all the numerical models, unless otherwise specified.

6.3. Handling of Dropouts or Missing Data

Baseline values for a patient will be derived from the last set of measurements collected prior to first dose of study drug at Visit 2. Change from baseline to a specific follow-up visit will be calculated for each patient as the visit value minus the baseline value. In change from baseline to endpoint analyses, endpoint values for patients who discontinued the study early or did not have data at the last double-blind treatment visit values will be imputed using the last non-missing postbaseline data as the endpoint values (last observation carried forward [LOCF] data imputation methodology). Patients with no postbaseline data for a particular efficacy endpoint will be excluded from the analysis of that endpoint.

For the purpose of calculating duration, the following imputation rules will be used for an incomplete date record: impute 15 for the day if only the day is missing; impute 07 for the month and 01 for the day if both the day and month are missing; if the entire date is missing, the duration will be missing.

When expanding the adverse event (AE), concomitant therapy, and previous therapy records across visits, the following imputation methods will be used:

For start date:

- If only the day component is missing, then the first day of the month will be used to complete the date.
- If only the month component is missing or both the day and month components are missing, then January 1 will be used to complete the date.

- If only the year component is missing, then the year part of the patient's consent date will be used.

For end date,

- If only the day component is missing, then the last day of the month will be used.
- If only the month component is missing or both the day and month components are missing, then December 31 will be used to complete the date.
- If the year component is missing, then the patient's study end date year will be used.

For the purpose of deriving CHQ-PF28 scores, missing items will be handled as documented in the scoring manual.

Missing AE severities will be imputed as follows: if missing at baseline, then 'Mild' will be assigned; if missing at postbaseline, then 'Severe' will be assigned.

6.4. Multicenter Studies

This is a multicenter study and results will be provided for specific study sites or geographic regions where deemed appropriate.

6.5. Multiple Comparisons/Multiplicity

No adjustments for multiple comparisons will be made.

6.6. Patient Disposition

The number and percentage of patients screened and randomly assigned to treatment will be presented by investigative site and weight cohort. Reasons for screen failure will be summarized. All patients who discontinue from the study will be identified, and the extent of their participation in the study will be reported. If known, a reason for their discontinuation will be given. Frequencies and percentages of all patients randomized, discontinuing the study, and completing the study will be presented for each of the treatment groups during the double-blind treatment period and open-label extension. The completion status for this study is based on the designation on the case report form (CRF). A summary of discontinuations will be presented by treatment group and by visit.

6.7. Demographics and Baseline Characteristics

Patient demographics and baseline characteristics will be summarized for each treatment group and overall using summary statistics for continuous and categorical data, as appropriate.

Patient characteristics at baseline will include:

- age, age category, gender, race, ethnicity
- supine vital signs (systolic blood pressure [SBP], diastolic blood pressure [DBP], and heart rate [HR])

Baseline disease characteristics and therapies to be summarized by treatment group include:

- body mass index (BMI), height, weight, weight category

- pulmonary arterial hypertension etiology
- duration of PAH
- endothelin receptor antagonist therapy
- World Health Organization functional classification
- 6-minute walk distance
- Tanner Score
- Clinical Global Impression of Severity (CGI-S)

A by-patient listing of demographic data and baseline clinical and disease characteristics will be presented for all randomized patients.

Demographics and baseline clinical and disease characteristics will also be summarized for each treatment group and overall as above for the 6MW analysis population.

6.8. Treatment Compliance

6.8.1. Study Drug Exposure and Compliance

Study drug exposure and compliance will be reported for all randomized patients. Study drug exposure will be listed by patient and summarized by treatment group.

For all randomized patients in the double-blind treatment period (Period 1) and open-label extension period (Period 2), the total number of days of exposure, the cumulative number of doses taken, and the number of average doses taken per week will be summarized using descriptive statistics by treatment group.

Patients in this study will receive tadalafil or matching placebo once daily for 24 weeks in Period 1. In Period 2, all patients will receive tadalafil once daily in an open-label fashion. Patients will receive their treatment in the form of tablets or suspension.

The doses are prescribed based on weight cohort, and the number of tablets or volume of suspension are given to a patient as one dose are described in the LVHV study protocol Section 9.4, "Rationale for Selection of Doses in the Study." For example, patients assigned to a 40-mg dose are expected to take 2 tablets (20 mg each) per day; patients assigned to a 10-mg dose are expected to take 1 tablet (10 mg) per day; and patients assigned to a 5-mg dose are expected to take 2.5 mL of oral suspension per day.

The cumulative number of doses taken over a study period or the entire study is defined as:
(total number of doses dispensed since randomization – total number of doses returned since randomization), where doses dispensed and returned are calculated from pill counts or bottle weights collected on the case report forms and the dosage definitions above.

The total number of days of exposure is defined as:
(last dose date – first dose date) + 1.
The average doses taken per week is defined as:

([the cumulative number of doses taken] / ([total number of days of exposure]/7).

If a patient is lost to follow-up during the double-blind treatment period, the total number of days of exposure, the cumulative number of doses taken, and the number of doses taken per week, will be calculated based on data available from the patient's last documented visit during the double-blind treatment period. If the patient has no visit after randomization and use of study drug is unknown, the total number of days of exposure, the cumulative number of doses taken, and the number of doses taken per week will be missing. If a patient is lost to follow-up during the open-label treatment period, the total number of days of exposure, the cumulative number of doses taken, and the number of doses taken per week, will be calculated based on data available from the patient's last documented visit during the open-label treatment period. If the patient has no visit after entering the open-label treatment period and use of study drug is unknown, the total number of days of exposure, the cumulative number of doses taken, and the number of doses taken per week will be missing for the open-label period.

Patient compliance with study drug will be monitored at each visit starting at Day 1 (Visit 2). Treatment compliance will be assessed by direct questioning, counting returned tablets, or suspension volume reconciliation.

Treatment compliance will be assessed in Period 1 by reconciling the number of doses of study treatment dispensed at Visits 2 through 8 with the number of doses returned at Visits 3 through 9. The dates of first and last doses in Period 1 will be used to estimate exposure duration during Period 1. If either of these dosing dates is missing, visit dates will be used to estimate exposure duration. Treatment compliance in the open label period (Period 2) will be assessed by reconciling the number of doses of study treatment dispensed at Visits 9 through 16 with the number of doses returned at Visits 10 through 17. The dates of first and last doses in Period 2 will be used to estimate exposure duration during Period 2. If either of these dosing dates is missing, visit dates will be used to estimate exposure duration.

Treatment compliance for a study period is estimated as the cumulative number of doses taken during the study period divided by the total exposure duration in the study period, expressed as a percentage.

The study period compliance rate is defined as follows: $100 \times$ (cumulative number of doses dispensed – cumulative number of doses returned)/total days of exposure in the study period, where the number of days of exposure is calculated as the difference between the dates of first dose in the study period and last dose in the study period. If the last dose date is missing, the last documented visit date will be used to calculate the total number of days of exposure and the study period compliance rate. If the patient is lost to follow up, the last available visit will be used to calculate the study period compliance rate.

Treatment compliance at each visit may be estimated as the number of doses of treatment taken during a visit interval divided by the exposure duration in the visit interval, expressed as percentage.

The visit-wise compliance rate is defined as follows: 100 × (number of doses dispensed - number of doses returned)/days of exposure, where the number of days of exposure is calculated as the difference between the dates of 2 consecutive visits with drug dispensed or returned except during the last visit interval, for which exposure is calculated as [the last dose date - the date of the prior visit] + 1.

If the patient returns to the investigative site with study drug packaging but no remaining tablets for a study visit, a zero value will be recorded for the number of tablets returned. For the last visit interval, if the last dose date is missing, the date of the last visit will be used to calculate exposure.

Compliance will be reported up to the time the patient either completes the trial or to the time the study drug is discontinued. Number of observations, mean, median, SD, minimum and maximum exposure will be used to summarize compliance for each treatment group.

6.8.2. Previous and Concomitant Therapy

Previous therapies are those therapies that started and stopped prior to the first dose of study medication. Concomitant therapies are those therapies that started on or after the first dose of study medication or those therapies that started prior to the first dose of study medication and were ongoing when the first dose of study medication was given.

For Study LVHV, only previous therapies for PAH are recorded in the clinical database. Previous PAH therapy usage will be summarized by the WHO drug substance name in the randomized population by treatment group.

Concomitant medications will be summarized by the WHO drug substance name. Patients may report the use of a concomitant medication more than once within a period. For each WHO drug substance group, a patient is counted only once if the same medication is reported multiple times. The number and percent of patients using at least 1 concomitant therapy as well as each individual concomitant therapy by WHO drug name will be used to summarize each treatment group.

Concomitant medications usage will be presented by the following study periods:

- endothelin receptor antagonist therapies (Period 1 and Period 2)
- non-ERA therapy medications used during the double-blind treatment period (Period 1)
- non-ERA therapy medications used during the open-label extension period (Period 2)

The duration of use of each type of ERA therapy at baseline will also be summarized.

6.9. Efficacy Analyses

6.9.1. Primary Outcome and Methodology

The primary endpoints will be evaluated in this study during Period 1. The primary efficacy measure is 6MWD.

6.9.1.1. 6 Minute Walk

The primary outcome of Period 1 is change in 6MW distance from baseline through 24 weeks. The analysis set will include data from all visits through Week 24. This analysis will include only patients who are ≥6 and <18 years of age and are developmentally able to complete 6MW testing during Period 1. Patients with no postbaseline data will be excluded from the primary analysis. The comparison of change in 6MWD between tadalafil and placebo treatment groups will be performed using a restricted maximum likelihood (REML)-based, MMRM. Factors in the MMRM model may include visit, baseline 6MWD, PAH etiology, type of ERA therapy, treatment group, treatment-by-visit, and treatment-by-baseline. The effect of PAH etiology will only be included if there are at least 3 patients per treatment arm at each PAH etiology level. Similarly, the effect of ERA therapy will only be included if there are at least 3 patients per treatment arm at each ERA therapy level.

The MMRM analysis will be based on an assumption of normality, independence between patients, and independently collected yet correlated data for a patient at different time points. An unstructured covariance matrix will be used to model the within-subject errors. If the model for unstructured covariance matrix fails to converge, the heterogeneous Toeplitz covariance structure, followed by the heterogeneous autoregressive covariance structure, will be used. Additionally, if the model fails to converge even once a heterogeneous autoregressive covariance structure is assumed, a standard Toeplitz covariance structure, followed by the standard autoregressive level 1 covariance structure, will be used. Should these fail to converge due to small sample size, then the covariates, PAH etiology, ERA will be removed from the model. If failure due to converge persist, only summary statistics by treatment and visit will be provided.

The Kenward-Roger approximation will be used to estimate denominator degrees of freedom. The least squares (LS) means and standard error for each treatment, the LS mean for the treatment difference, and the 80% CI of the difference will be estimated for each visit. As the study has a limited number of patients enrolled, an inspection of the numerical mean change in 6MWD and corresponding CIs will be used instead of significance testing to ensure whether they are trending in the right direction. The overall treatment difference p-value and the visit-wise p-values will not be reported.

The 6MWD data in Period 2 will be summarized for each visit, overall, and by original randomized treatment assignment in Period 1.

6.9.1.2. 6 Minute Walk (Sensitivity Analysis)

In addition, a supportive (sensitivity) analysis for a Bayesian MMRM model that leverages data from the adult study (Study H6D-MC-LVGY [LVGY]) will be conducted to increase precision in confirming the 6MWD efficacy endpoint. Factors in the Bayesian MMRM model may include visit, baseline 6MWD, PAH etiology, type of ERA therapy, treatment group, interaction terms treatment-by-visit and treatment-by-baseline. Similar to the above frequentist approach, the factors included in this model may change if there are too few patients for a given PAH etiology or a given ERA therapy. Bayesian posterior probability of active treatment arm being superior to placebo will be calculated. Details on the Bayesian model enabling partial

extrapolation from the adult data are provided in Appendix 4. The simulation plan and results to assess the performance of the model are also provided in Appendix 4.

6.9.2. Secondary Efficacy Analyses

The secondary efficacy endpoint of Period 1 time to first occurrence of CW during Period 1 will be analyzed using a Cox proportional hazard model. Time to CW will be measured from the date of first dose to the date of the CW event. Patients with no known date of CW will be censored at the date of current contact or last valid date. The analysis model will include terms for weight cohort, PAH etiology, type of ERA therapy, and treatment group. Estimates of hazard ratios (HR) and corresponding 95% CIs will be provided. The proportional hazards assumption will be checked with plotting methods. Due to small sample size, similar consideration for PAH etiology and type of ERA therapy as covariates,will be applied.

Kaplan-Meier plots of the proportion of patients without CW over time will be presented for all patients and by treatment group and weight cohort. As the study has a limited number of patients enrolled, p-value will not be reported.

Analyses of CW during Period 1 will use adjudicated data reported by the Clinical Endpoints Committee (CEC).

Patients meeting any of the following 5 major criteria would be considered to have met the definition of CW:

- all-cause mortality
- lung or heart lung transplantation
- atrial septostomy or potts shunt
- hospitalization for PAH progression. Hospitalization for PAH progression should not be due to a potentially precipitating event such as pneumonia hemoptysis, etc.; however, if after the hospitalization is completed, the patient is discharged and remains worse, then the patient can be assessed for clinical worsening.
- worsening of PAH. Patient has any of the following criteria:
    - new-onset syncope
    - addition of new PAH-specific concomitant therapy including, but not restricted to epoprostenol or treprostinil, sildenafil, vardenafil, or increase in dose of existing PAH specific concomitant therapy (for example, ERA)
    - increase of 1 or more in WHO Functional Class (Protocol Attachment 8) (except for patients already in Class IV) only for patients unable to perform the 6MW test
    - worsening of WHO functional class and a decrease of 20% in the 6MW test (confirmed 5 to 10 days later) for those patients who are ≥6 years of age and are developmentally capable of performing the 6MW test

The analyses for CW will only be conducted if there are in total at least 5 patients with adjudicated CW.

The time to first occurrence of CW during Period 2, and during the entire study (including both Periods 1 and 2) will be analyzed separately using Cox proportional hazard models with weight cohort, PAH etiology, and type of ERA therapy as factors in the model, similar to that described above without comparison between treatment groups. For the analysis of the CW events during Period 2 only, the time to first occurrence of CW is defined as the time between entering Period 2 of the study and the time of the first occurrence of CW in Period 2, regardless of prior events in Period 1. Any patients who had a CW event in both Period 1 and Period 2 will be listed separately; Kaplan-Meier estimates will be summarized and time from entering Period 2 to first CW event in Period 2 will be plotted for the patients with prior CW event in Period 1 versus the patients without prior CW event in Period 1. The analyses to first occurrence of CW during Period 2 if only there are at least 5 patients have experienced CW during Period 2.

The time to first occurrence of CW during the entire study is the time between first dose date and the time of the first occurrence of CW in the entire study.

To assess the incidence of CW during Periods 1 and 2, the number and percentage of patients in each treatment group and overall who experience at least 1 criterion of the CW definition during the study periods will be summarized. Summary tables will be provided for Period 1, Period 2 and the entire study (including both Periods 1 and 2).

6.9.3. Other Efficacy Analyses

The following efficacy analyses will be assessed during Period 1.

6.9.3.1. Additional Analyses of CW Data

Exploratory analyses of time to CW may be performed if sufficient data are available. These analyses may involve examining the impact of additional clinical factors such baseline WHO functional class, PAH duration at baseline, duration of ERA used at baseline, gender, and baseline 6MW distance for those patients who could perform the test on the Cox proportional hazard model described in Section 6.9.2.

6.9.3.2. WHO Functional Class

The proportion of patients who experience a change from baseline to endpoint in WHO functional class will be summarized. An additional analysis will be performed with changes categorized as “worsening,” “no change,” or “improving” over the study period. The percentages of patients in these categories will be summarized at endpoint of Periods 1 and 2. In addition, change from Day 1 of Period 1 in WHO functional class will be reported for all patients who participated in Period 2 including changes categorized as “worsening,” “no change,” or “improving.”

6.9.3.3. Echocardiogram

The following parameters will be assessed by echocardiography:

- tricuspid annular plane systolic excursion (TAPSE)

- left ventricular eccentricity index systolic
- left ventricular eccentricity index diastolic
- pericardial effusion
- maximal tricuspid regurgitant velocity

For the above parameters, an analysis of covariance (ANCOVA) model will be used to compare the changes from baseline to endpoint (Visit 9) between treatment groups. The model will include terms for baseline (Day 1) value, weight cohort, PAH etiology, type of ERA therapy, and treatment group. Least-squares mean estimates for the change from baseline with corresponding standard errors, and LS mean estimates of the treatment group differences with corresponding standard errors, CIs

As the study has a limited number of patients enrolled, an inspection of the numerical mean change from baseline and corresponding CIs will be used in place of significance testing to ensure whether they are trending in the right direction. Treatment difference p-value will not be reported.

In addition, descriptive statistics will be presented by treatment group for the changes from baseline in echocardiography parameters at Visits 5, 7, and 9.

6.9.3.4. Cardiac Magnetic Resonance Imaging

Cardiac MRI will be collected at Visit 2 and 9. These data will be collected at selected sites that have been using MRI as routine PAH patient management. The analyses will be based on available data. The following parameters will be assessed by cardiac MRI:

- left ventricular ejection fraction
- right ventricular end diastolic volume
- right ventricular end systolic volume
- right ventricular ejection fraction

For the above parameters, an ANCOVA model will be used to compare the changes from baseline to endpoint (Visit 9) between treatment groups. The model will include terms for baseline (Day 1) value, weight cohort, PAH etiology, type of ERA therapy, and treatment group. Least-squares mean estimates for the change from baseline with corresponding standard errors, and LS mean estimates of the treatment group differences with corresponding standard errors, CIs.

As the study has a limited number of patients enrolled, an inspection of the numerical mean change from baseline and corresponding CIs will be used in place of significance testing to ensure whether they are trending in the right direction. Treatment difference p-value will not be reported.

If there are less than 3 patients per treatment arm, only the listing of the Cardiac MRI will be provided.

6.9.3.5. Clinical Global Impression of Improvement (CGI-I)

Patient outcome will be assessed using the CGI-I at Weeks 16 and 24 (Visits 7 and 9). These measures have 7 discrete categories of response:

- very much improved
- much improved
- minimally improved
- no change
- minimally worse
- much worse
- very much worse

In addition, the 7 categories of responses for the CGI-I at each visit will be grouped into 3 derived categories:

- worse includes responses of "minimally worse", "much worse", or "very much worse"
- no change
- better includes responses of "minimally improved," "much improved," or "very much improved"

Proportions of patients in each of the 7 response categories and each of the 3 derived categories of the CGI-I will be summarized by treatment groups.

6.9.3.6. N-terminal Prohormone Brain Natriuretic Peptide

N-terminal prohormone brain natriuretic peptide will be measured at baseline (Visit 2) and Visits 7 and 9 postbaselines. Changes from baseline (or last lab test prior to randomization) to each postbaseline visit in NT-Pro-BNP measurements will be analyzed using an ANCOVA model with terms for baseline (Day 1) value, weight cohort, PAH etiology, type of ERA therapy, and treatment group. Least-squares mean estimates for the change from baseline with corresponding standard errors, and LS mean estimates of the treatment group differences with corresponding standard errors, CIs.

As the study has a limited number of patients enrolled, an inspection of the numerical mean change from baseline and corresponding CIs will be used in place of significance testing to ensure whether they are trending in the right direction. The treatment difference p-value will not be reported.

6.9.3.7. Child Health Questionnaire Parent Form 28

The Child Health Questionnaire Parent Form 28 will be measured in patients ≥5 years old. Patients <5 years will not be included in this analysis.

For the purpose of this study, the CHQ-PF28 items have 4, 5, or 6 response options divided over 9 multi-item scales (child's global health, physical activities, everyday activities, pain, behavior,

well-being, self-esteem, your child's health, you and your family). Responses to each of the CHQ-PF28 items are scored and expressed on a 0 (worst possible score) to 100 (best possible score) scale for each of the questions. The values recorded on the electronic case report form (eCRF) will be scored and standardized using algorithms documented in the scoring manual. The summary scores will be summarized and analyzed.

Changes from Day 1 to Weeks 16 and 24 in CHQ-PF28 scores will be analyzed with an ANCOVA model. The model will include terms for baseline (Day 1) score, weight cohort, PAH etiology, type of ERA therapy, and treatment group. Least-squares mean estimates for the change from baseline with corresponding standard errors, and LS mean estimates of the treatment group differences with corresponding standard errors, CIs. As the study has a limited number of patients enrolled, p-value will not be reported.

6.10. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

Analysis methods used to examine any plasma tadalafil concentration data collected during this study will be documented in a separate PK analysis plan. These analyses will be carried out by the PK department at Eli Lilly and Company.

6.11. Safety Analyses

Safety during Period 1 will be assessed through AEs which will include any abnormalities detected by electrocardiogram (ECG) or physical examination, clinical chemistry and hematology panels, urinalysis, vital signs, eye examinations, and concomitant medications. During Period 2, safety will be monitored using AEs, changes in body weight and height, inhibin B biomarker (male patients only), eye examinations, Tanner scale, and intelligence tests. The analysis of safety will include all patients who took at least 1 dose of study medication.

6.11.1. Pre-Existing Conditions and Overview of Adverse Events

Pre-existing conditions and AEs will be summarized using Medical Dictionary for Regulatory Activities (MedDRA, Version 15.1) Preferred Terms (PTs), and/or System Organ Classes (SOCs).

6.11.2. Adverse Events

A treatment-emergent adverse event (TEAE) is defined as an event that first occurred or worsened in severity after baseline (Day 1, Visit 2). The primary definition of baseline for assessing treatment-emergent status in both Period 1 and Period 2 of Study LVHV is prior to the first dose of study drug at Visit 2. The maximum severity for each MedDRA PT prior to first dose (Visit 2) will be used as baseline. Events occurring at Visit 2 or after will be defined as postbaseline for the analysis. The postbaseline severity will be compared to the baseline maximum severity and if the PT has a higher severity postbaseline, the event will be considered a TEAE. Note that missing severities will be imputed as follows: if missing at baseline, then 'Mild' will be assigned; if missing at postbaseline, then 'Severe' will be assigned. An additional summary of events that first occur or worsen in Period 2 compared to the baseline period from Visits 1 through 9 may also be presented.

The number and percentage of patients who experienced a TEAE, serious adverse event (SAE), TEAEs related to treatment, TEAEs related to procedure, died, or discontinued from the study due to an AE will be summarized by treatment groups. Patient incidence of TEAEs summarized by PT will be presented by decreasing frequency of occurrence in the tadalafil group. The patient incidence of TEAEs will also be presented alphabetically within SOC and by maximum severity within SOC. For events that are gender-specific, the denominator and computation of the percentage will include only patients from the given gender.

These analyses will be carried out for the double-blind treatment period and the entire study (including open label extension), separately.

6.11.3. Deaths, Serious Adverse Events, and Discontinuations Due to Adverse Events

All SAEs, deaths, and discontinuations due to an AE will be listed. Some events reported in these categories may not be defined as treatment-emergent. Deaths will also be counted as SAEs and AEs leading to discontinuation.

An SAE is any AE from this study that results in 1 of the following outcomes:

- death
- initial or prolonged inpatient hospitalization
- a life-threatening experience (that is, immediate risk of dying)
- persistent or significant disability/incapacity
- congenital anomaly/birth defect
- considered significant by the investigator for any other reason

6.11.4. Special Adverse Event Follow-ups

The number and percentage of patients with hearing, visual, prolong erection, and pediatric PAH uterine bleeding abnormality follow up will be summarized by treatment group during Period 1. A patient is counted only once if the same abnormality follow-up is reported multiples times. In addition, an individual patient listing will be provided.

6.11.5. Clinical Laboratory Evaluation

6.11.5.1. Clinical Laboratory Tests

For the double-blind treatment period, baseline for clinical laboratory values will be the last available lab values recorded prior to Visit 2, and endpoint will be the last non-missing postbaseline data collected at or prior to Visit 9. Laboratory data will be listed by laboratory parameter for each patient. Summary statistics (including number of patients, mean, SD, minimum, and maximum) of the raw and change-from-baseline values for these parameters will be computed for each treatment group at each visit during the double-blind treatment period. Changes from baseline in laboratory results will be analyzed at each visit and endpoint using a

ranked ANOVA with treatment in the model. As the study has a limited number of patients enrolled, p-value will not be reported. For the Primary Analysis Population, baseline values, endpoint values, and the change from baseline to endpoint during the double-blind treatment period for hematology (except red blood cell [RBC] morphology) and chemistry parameters will be presented using box plots. Descriptive statistics for baseline, endpoint, and change from baseline to endpoint will be presented by treatment group in a table.

In addition, shift tables will be presented for each of the laboratory parameters. The tables will show the percentages of patients with shifts in laboratory results from baseline to each postbaseline visit using categories (low, normal, high) based on the central laboratory reference ranges.

6.11.5.2. Inhibin b Biomarker

Inhibin b biomarker concentrations will be collected at baseline (Visit 2), Year 1 (Visit 13), and Year 2 (Visit 17) for male patients. Inhibin B values collected in patients who are less than 9 years of age are considered exploratory. As such, separate analyses of Inhibin B values will be conducted for patients <9 years of age, ≥9 and <13 years of age, and patients ≥13 years of age. Changes from baseline to Years 1 and 2 will be summarized descriptively.

6.11.5.3. Hepatic Monitoring

Chemistry test results related to liver function (alanine transaminase [ALT], aspartate transaminase [AST] and total bilirubin) will be classified using Covance conventional ranges and summarized by treatment group in a table that presents counts and percentages of randomized patients who met the following conditions at any postbaseline visit but not at baseline in the double-blind treatment period:

- ALT or AST >3x upper limit of normal (ULN)
- ALT or AST >3x ULN and bilirubin >2x ULN
- Total bilirubin > 2x ULN

For randomized patients who met at least 1 of the above criteria at any postbaseline visit but not at baseline in the double-blind treatment period, a by-patient listing displaying ALT, AST and total bilirubin values by visit will also be presented.

6.11.6. Vital Signs

For the double-blind treatment period, baseline for vital signs will be the last available recorded prior to Visit 2 and endpoint will be the last non-missing postbaseline data collected at or prior to Visit 9.

Vital signs data (supine SBP, supine DBP, supine heart rate) will be collected at each visit in the double-blind treatment period (Period 1) and will be listed for each patient. Summary statistics (including mean, SD, and median) of the raw and change-from-baseline values for these parameters (including height and weight) will be computed for each treatment group at each visit and at endpoint (LOCF). Changes from baseline in vital signs will be analyzed at each visit and

endpoint (double-blind treatment period) using a ranked ANOVA with treatment in the model. As the study has a limited number of patients enrolled, p-value will not be reported.

6.11.7. Intellectual Ability and Cognitive Functioning Assessment

The patient's intellectual ability (intelligence quotient, IQ) will be assessed at Day 1 (Visit 2, prior to first dose of study drug), and after 1 year and 2 years following treatment initiation. The Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV) will be the preferred instrument for IQ assessment. Due to age restrictions of the WISC-IV, the Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV) and Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III) test may also be used. Patients may be assessed with a different scale at subsequent visits depending on their age. Scales will not be pooled for analysis purposes. Patients with available data will be summarized. Due to the possibility of different scales being used between and within patients, no comparisons are planned; all available data will be listed.

If the recommended versions of the IQ scales listed above are not available in the patients' primary language, the site may use the most recent version of the available scale in that geography. Investigator or site study personnel should ensure the instrument administrator/examiner and interpreter, either at the Investigator site or from an external evaluation service, meet the qualification, training, and interpretation requirements per the instrument manual. There is also the possibility that data may be unavailable for certain countries/regions, if none of the recommended instruments are available in the patient's primary language or if no qualified examiner is available to conduct the evaluation.

6.11.8. Tanner Score

Shift tables will be presented for each of the Tanner stage parameters (pubic hair and breast score for females; pubic hair and genital score for males) for the patients. The tables will show the percentages of patients with shifts in Tanner scores from baseline (Day 1) to each postbaseline visit (Years 1 and 2 during the open label extension). No comparisons are planned for Tanner score.

6.11.9. Eye Examination

Fundoscopy examinations will be recorded at screening and at the end of each study period (Week 24 and Year 2). The results of these examinations are reported as normal or abnormal for each eye with abnormal results designated as clinically significant or not. Shift tables will be presented for each of the postbaseline eye examination results. The tables will show the percentages of patients with shifts in eye examination results from baseline to each postbaseline visit (Week 24 and Year 2).

6.11.10. Right heart catheterization

Right heart catheterization data will be collected under the LVHV protocol addendum for right heart catheterization. Descriptive summary statistics of collected parameters including mPAP, MRAP, CO, PCWP, Scv02 and SaO2 will be presented by treatment group and time of collection.

6.12. Subgroup Analyses

The following is a list of subgroups which may be analyzed for differential treatment effects, if the sample size is appropriate:

- endothelin receptor antagonist type
- pulmonary arterial hypertension etiology
- weight cohort

As the study has a limited number of patients enrolled, change in 6MWD results within these subgroups will be displayed using descriptive summaries.

6.13. Protocol Violations

Important protocol violations are defined as deviations from the protocol that could reasonably have an impact on patient safety, data integrity, or conclusions drawn from the study. The following categories of protocol violations will be summarized from the clinical database using statistical programming:

- inclusion criteria not met/exclusion criteria met
- informed consent date missing, or obtained after Visit 1

Statistical output summarizing protocol violations by these categories will be generated using the criteria defined in Appendix 2. This list may not include all inclusion/exclusion criteria from the protocol.

The number of randomized patients in each protocol violation category will be summarized by treatment group as well as for all patients by study period for the study through the double-blind treatment period and for the open-label period.

It is important to note that situations not listed in Appendix 2 may be identified as protocol violations during scheduled study data reviews; these protocol violations will be summarized separately and will be discussed in the CSR.

If any patients receive incorrect study medication, a by-patient listing displaying investigator, randomized treatment, actual treatment received, randomization date and total number of days of exposure will be presented.

6.14. Interim Analyses and Data Monitoring

A Data Monitoring Committee (DMC) will review available safety and efficacy data at specified times during the study. The DMC comprises individuals external to Lilly Research Laboratories (LRL) who will be responsible for the evaluation and interpretation of safety and efficacy data available at the time of the review. The LVHV DMC Charter describes the objectives, membership, and procedures of the DMC for Study H6D-MC-LVHV.

For the DMC review, the following will be listed and/or summarized (based on data available at the time of review):

- patient disposition, demographics, baseline characteristics

- historical illness
- concomitant medications
- adverse events
- vital signs
- laboratory results
- clinical worsening
- 6-minute walk
- World Health Organization functional class
- IQ test results (during Period 2)
- Tanner scores (during Period 2)
- N-terminal prohormone brain natriuretic peptide concentration

As mentioned in Section 12.2.12 of the LVHV protocol, an interim analysis of available efficacy data is planned for this study. This interim will occur before the final analysis for Period 1.

The results of this interim analysis will be reviewed by the Data Monitoring Committee (DMC) for Study LVHV to assess evidence of efficacy, and to confirm the accuracy of study design assumptions. These results may also be reviewed by a committee of Lilly representatives who are independent of the study team and may be reported in a confidential manner to regulatory agencies. However, members of the LVHV study team, patients, and site personnel will not view the results of the interim analysis and will remain blinded to treatment assignments.

The detailed description of the analyses for this efficacy interim analysis are available in Appendix 3.

In addition to the above DMC reviews, the database will be locked and the data collected during Period 1 will be analyzed when all randomized patients have ended participation in Period 1. These results will be reported in a CSR. An additional CSR will be prepared at the conclusion of Period 2 to present analyses of the Open-label extension data.

Periodic blinded reviews of the safety data (AEs, vital signs, ECGs, and safety laboratory tests) from all enrolled patients will occur throughout the study. Details of safety monitoring will be documented in the trial-level safety-review plan.

6.15. Trial Level Safety Review Reports

Trial level safety reporting (TLSR) will be performed 4 months after the first patient visit with blinded data. The following listings or data summaries will be produced:

- non-serious AEs
- Serious adverse events
- protocol-related SAEs
- laboratory data
- electrocardiograms (abnormalities only)
- discontinuations
- vital signs

- concomitant medications
- historical/pre-existing conditions
- other study data if available at the time of analysis: 6MW, WHO functional class (DMC review only)

6.16. Development Safety Update Report

Based on requirements for the annual report, the following should be produced (if not already available from the study CSR):

- summary of estimated cumulative patient exposure to tadalafil or placebo
- summary of estimated patient exposure to tadalafil by gender and age
- summary of estimated patient exposure to tadalafil by race
- summary of estimated patient exposure of all ongoing and completed trials during the reporting period
- listing of patients who discontinued due to AEs during the reporting period
- listing of patients who died during the reporting period

6.17. Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements.

Analyses provided for the CTR requirements include the following:

Summary of AEs, provided as a dataset which will be converted to an XML file. Both SAEs and 'Other' AEs are summarized by treatment group and MedDRA PT.

- An AE is considered 'Serious' whether or not it is a TEAE.
- An AE is considered in the 'Other' category if it is both a TEAE and is not serious. For each SAE and 'Other' AE, for each term and treatment group, the following are provided:
    - the number of participants at risk of an event
    - the number of participants who experienced each event term
    - the number of events experienced.
- Consistent with www.ClinicalTrials.gov requirements, 'Other' AEs that occur in fewer than 5% of patients/patients in every treatment group may not be included if a 5% threshold is chosen (5% is the minimum threshold).
- Adverse event reporting is consistent with other document disclosures for example, the CSR, manuscripts, and so forth.

6.18. Analysis Requirements for the Japanese Addendum

Since limited number of Japanese patients (2 patients) will be included in the analyses at the final database lock, subgroup analyses of the Japanese patients for the planned efficacy, health outcomes, and safety measures planned in the LVHV Japan addendum will not be performed. Additional data presentation for Japanese patients may be performed upon requests after database lock.

7. Unblinding Plan

The purpose of this unblinding plan is to maintain the scientific integrity of the study and document the personnel with access to information regarding individual randomized treatment assignment and summary (group) level data. Access to this information will be limited in order to minimize bias.

The final data analysis for Period 1 of the study to be presented in the CSR will require unblinding of the study data when all patients have finished treatment and visits for Period 1. Data collected during Period 2 of the study will be unblinded, that is, open-label treatment.

Unblinding authorization is inherent in the datalock authorization for final analysis. The designated systems analyst applies the patient treatment assignments to the blinded database and maintains the records as each individual is given access to the unblinded data.

One interim analysis of efficacy data is planned before the final analysis for Period 1 of this study. The core study team, patients, and site personnel will remain blinded to treatment assignment during this interim analysis.

Study LVHV will use an external DMC whose members will have access to unblinded data. While LVHV is on-going, only the Statistical Analysis Center (SAC) responsible for producing DMC reports will have access to the randomization codes and will unblind the data for reporting to the DMC.

No one other than the SAC or DMC will have access to these unblinded data, unless an internal review is needed to make an informed decision based on a DMC recommendation to stop or modify the study. In that case, the individuals involved in the internal review and LRL Senior Management may also view the unblinded or partially unblinded data.

The SAC statistician will maintain a list of all unblinded individuals, the date and level of their unblinding, and a description of what subset of data, if not all the data, was shared.

7.1. Site-Level Unblinding

The site monitor is responsible for verifying compliance with the blinding procedures at the investigator site and verifying that access to the patients' treatment assignments remains restricted from the investigator and site personnel in direct contact with patients. The documentation of emergency unblinding reported to Lilly is filed in the Clinical Trial Management study files. A final Study Unblinding Summary will be prepared at the end of the study (at the study closeout).

7.2. Sponsor / Trial-Level Unblinding

A DMC, composed of members external to Lilly, will be used for this study to review safety data during the study. More detail can be found in the DMC Charter for this study.

The Lilly study team will remain blinded during the study until datalock for the primary endpoint. If the LRL Senior Management Designee receives a recommendation from the

Advisory Committee (AC) to modify the study, an internal review could be deemed necessary. In this event, the internal review committee (IRC) will request the AC unblinded reports for review. If more analyses are necessary, the statistician will receive the new analysis code and the unblinding treatment codes from Lilly CT-SMS in order to provide the unblinded reports to the IRC.

If an emergency interim analysis is performed, the statistician is responsible for authorizing the access of Lilly personnel to unblinded data. Every attempt should be made to contact the statistician and document the authorization before access is given to unblinded data. A designated systems analyst maintains the records as each individual is given access to the patient treatment assignment or unblinded data. The documentation is filed in the study files.

A designated study team member, collaborating with the statistician, will be responsible for keeping a running log of individuals given access to any unblinded study data. This log will include the person's name, title, and date of unblinding.

8. References

Dallow N, Best N, Montague TH. Better decision making in drug development through adoption of formal prior elicitation. *Pharma Stat*. 2018;17(4):301-306.

[FDA] Food and Drug Administration. Guidance for the use of Bayesian statistics in medical device clinical trials. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health. Available at: https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071072.htm. Published February 05, 2010. Accessed July 9, 2018.

Fu H, Manner D. Bayesian adaptive dose-finding studies with delayed responses. *J Biopharam Stats*. 2010;20(5):1055-1070.

Morita S, Thall PF, Müller P. Determining the effective sample size of a parametric prior. *Biometrics*. 2008;64(2):595-602.

Ye J, Travis J. 2017. A Bayesian approach to incorporating adult clinical data into pediatric clinical trials. Food and Drug Administration, Office Biostatics (DB V and II). Available at: https://www.fda.gov/downloads/Drugs/NewsEvents/UCM576645.pdf. Accessed July 9, 2018.

9. Appendices

Appendix 1. Protocol LVHV Study Schedule

Protocol LVHV Study Schedule

		Period 1								Period 2				
Visit	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9/ET[a]**	**10-12**	**13**	**14-16**	**17**	
Description of event LVHV	**Screening Day -28 to 0**	**Day 1**	**Wk2 ± 7days**	**Wk 4 ± 7days**	**Wk8 ± 7days**	**Wk12 ± 7days**	**Wk16 ± 7days**	**Wk20 ± 7days**	**Wk 24 ± 7days**	**Every 3 months ±10 days**	**1 Year ±10 days**	**Every 3 months ±10 days**	**2 Year ±10 days**	**Follow -up[b]**
Informed Consent	X													
Medical History	X													
PAH etiology	X													
OB/GYN History[c]	X													
CXR (within 6 months of screening)	X													
WHO Functional Class	X			X	X	X	X	X	X	X	X	X	X	
Physical Examination	X	X				X			X					
Eye Examination[d]	X								X				X	
6MW Test[e]		X		X	X	X	X	X	X		X		X	
Concomitant Medications	X	X	X	X	X	X	X	X	X	X	X	X	X	
Height	X								X		X		X	
Weight	X	X	X	X		X	X		X	X	X	X	X	
Vital Signs	X	X	X	X	X	X	X	X	X					
ECG (single)[f]	X				X				X					
Urinalysis	X			X			X		X					
Urine Pregnancy Test[g]	X													
Urine Drug Screen[f]	X													
Safety Lab Tests: Chemistry, hematology, Coagulation[h]	X[i]			X			X		X					
NT-Pro-BNP		X					X		X					
Inhibin B biomarker (for male patients)		X									X		X	
CHQ-PF28 (≥ 5 yrs. old)		X					X		X					
DNA (PGx) Sample		X[j]												
PK (tadalafil concentration)[k,l]			X[m]	X[m,n]			X[m,n]		X[m.n]					

		Period 1								Period 2				
Visit	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9/ET[a]**	**10-12**	**13**	**14-16**	**17**	
Description of event LVHV	**Screening Day -28 to 0**	**Day 1**	**Wk2 ± 7days**	**Wk 4 ± 7days**	**Wk8 ± 7days**	**Wk12 ± 7days**	**Wk16 ± 7days**	**Wk20 ± 7days**	**Wk 24 ± 7days**	**Every 3 months ±10 days**	**1 Year ±10 days**	**Every 3 months ±10 days**	**2 Year ±10 days**	**Follow -up[b]**
CGI-S		X												
CGI-I							X		X					
Intelligence Test (WISC-IV, WAIS-IV, or WPPSI-III[o])		X									X		X	
Tanner Score		X									X[p]		X	
Echocardiography		X			X		X		X					
Cardiac MRI[q]		X							X					
Pre-existing Conditions and Adverse Events	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Dispense Study Drug		**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**		
Drug Return and Accounting			X	X	X	X	X	X	X	X	X	X	X	

Abbreviations: 6MW = 6-minute walk; CHQ-PF28 = Child Health Questionnaire Parent Form 28; CGI-I = Clinical Global Impression of Improvement questionnaire; CGI-S = Clinical Global Impression of Severity questionnaire; CXR = chest radiography; DNA (Pgx) = deoxyribonucleic acid pharmacogenetics; ECG = 12 –lead electrocardiogram; ERA = endothelial receptor antagonists; ET = early termination; NT-Pro-BNP = N-terminal prohormone brain natriuretic peptide; MRI = magnetic resonance imaging; OB-GYN = obstetrics-gynecology; PAH = pulmonary arterial hypertension; PK = pharmacokinetics; SAE = serious adverse event; WAIS = Wechsler Adult Intelligence Scale; WHO = World Health Organization; WISC = Wechsler Intelligence Scale for Children; Wk = week; WPPSI = Wechsler Preschool and Primary Scale of Intelligence.

a Patients who continue to Period 2 (because they completed or discontinued Period 1), will have all of the Week 24 (Visit 9/ET Visit) assessments performed before proceeding to Period 2. Patients, who discontinue Period 1 and do not participate in Period 2, will also have all of the Week 24 (Visit 9/ET Visit) assessments performed.

b This follow-up visit will be conducted only for patients who discontinue from the study during Period 1 and will not participate in Period 2. If a patient discontinues prior to or at Visit 8, the follow-up visit will be performed 24 weeks after the patient's initial study drug dosing (Visit 9). If a patient discontinues after Visit 8, the follow-up visit will occur 30 days after the patient has taken the last dose of study drug. This visit can be done by phone.

c Including family history of menarche.

d Eye examination includes patient medical eye history, external eye examination and retinal examination using ophthalmoscopy.

e 6MW test will be performed for those patients ≥6 years of age and who are, in the opinion of the Investigator developmentally capable (mentally and physically) of performing a 6MW test. Patient with worsening of WHO functional class by 1 class or more and a decrease of ≥20% in the 6MW distance, another 6MW will be repeated 5 to 10 days later to confirm the change. An unencouraged 6-MW will be used to ensure that patients are not pressured during the test. A separate "practice" 6MW test must be done before or during Visit 2 (Day 1).

f To be performed locally. Patients must be supine for approximately 5 to 10 minutes before ECG collection and remain supine but awake during ECG collection.

g Local pregnancy test for females of child bearing potential; may be repeated at Investigator's discretion throughout the trial.

h Additional samples may be collected as needed at time of SAE reporting and clinical worsening. Digoxin, warfarin, ERA and coagulation tests should be carried out using the Investigator's standard of care.

i Screening laboratory exam includes measured or estimated creatinine clearance (See Section 8.2 of the LVHV Protocol).

j If not collected at this visit, the sample could be collected at a following visit.

k The sampling times relative to dosing should vary as much as possible across the PK sampling visits.

l At the time of any SAE, a blood sample for tadalafil concentration analysis may be collected.

m Obtain and record the patient's weight at each PK sampling visit.

n The PK blood sample should be obtained prior to the 6MW test.

o WISC-IV is to be administered for patients ranging from 6 years 0 months through 15 years 11 months, WAIS-IV is to be used for patients 16 years 0 months and older, and WPPSI is to be used for patients 2 years 6 months to 5 years 11 months at Visit 2 and up to 7 years and 3 months for the follow up visit. The Intelligence test may be performed prior to Visit 2.

p If patient has Tanner Score 5 on all criteria, the following Tanner Score evaluation will not be required.

q Participation for MRI assessment will be based on selection of specific sites that have been using MRI as routine PAH patient management.

Appendix 2. Pre-Specified Listing of Protocol Violations Summarized from the Clinical Database

Major protocol violations that will be summarized from the clinical database using pre-specified statistical programming are as follows:

Related to inclusion/exclusion criteria not met:

- informed consent date missing or after date of Visit 1
- age at Visit 1 <6.0 months or ≥18.0 years
- World Health Organization functional class not equal to II or III at Visit 1
- at Visit 1, aspartate transaminase (AST) ≥ 3x upper limit of normal (ULN) or alanine transaminase (ALT) ≥3x ULN
- Concomitant endothelin receptor antagonist (ERA) start date <12 weeks prior to Visit 1 date (Visit 1 date – ERA start date from electronic case report form (eCRF) CCT-ERA <84 days). NOTE: This violation is based on inclusion criteria #5 [All subjects must be receiving an ERA (such as Bosentan or ambrisentan) and must be on a maintenance dose with no change in dose (other than weight-based adjustments) for at least 12 weeks prior to screening and have a screening AST/ALT <3 times the ULN]
- estimated creatinine clearance at Visit 1 <30 mL/min
- concomitant therapy (eCRF CCT) with phosphodiesterase type 5-inhibitor (dictionary term = "SILDENAFIL" or "VARDENAFIL" or "TADALAFIL") on or after Visit 1 date.
- Prior pulmonary arterial hypertension (PAH) therapy (eCRF parathyroid hormone [PTH]) with dictionary term = "SILDENAFIL" or "VARDENAFIL" or "TADALAFIL" and drug stop date less than 12 weeks prior to Visit 2 date (ie, Visit 2 date – drug stop date ≤84 days. If the stop date is missing for either of these drug names such that the duration between the stop date and Visit 2 date cannot be established, this incidence of use will be flagged as a protocol violation.
- concomitant therapy with prostacyclin or its analogues on or after Visit 1 date and prior to Visit 9 date [WHO dictionary terms:
  - epoprostenol (Flolan) (Veletri)
  - iloprost (Ventavis)
  - treprostinil (Remodulin) (Tyvaso)
  - beraprost (Prostalin) (oral in Japan)]

Major protocol deviations and patients belonging to each population will be finalized before database lock.

Appendix 3. Analysis Plan for Efficacy Interim Analysis

Given the small number of patients enrolled at the time of this interim analysis of efficacy, no formal comparison will be made between treatment groups. No stopping for efficacy is planned; therefore, there is no alpha adjustment due to this interim efficacy analysis.

Patient Disposition:

The number and percentage of patients screened and randomly assigned to treatment will be presented by investigative site and weight cohort. Frequencies and percentages of all patients randomized, discontinuing the study, and completing the study will be presented for each of the treatment groups during the double-blind treatment period and open label extension. A summary of discontinuations will be presented by treatment group and by visit.

Demographics and Baseline Characteristics:

Patient demographics and baseline characteristics will be summarized for each treatment group and overall using summary statistics for continuous and categorical data, as appropriate.

A by-patient listing of demographic data and baseline clinical and disease characteristics will be presented for all randomized patients.

The demographics and baseline clinical and disease characteristics will also be summarized for each treatment group and overall as above for the 6-minute walk (6MW) analysis population.

Previous and Concomitant Therapy:

Previous pulmonary arterial hypertension (PAH) therapy usage will be summarized in the randomized population by treatment group.

Concomitant medications will be summarized by endothelin receptor antagonist (ERA) therapies and non-ERA therapy medication separately. The number and percent of patients using at least 1 concomitant therapy as well as each individual concomitant therapy will be presented.

Concomitant medication usage will be presented by the following study periods:

- endothelin receptor antagonist therapies (Period 1 and Period 2 separately)
- non-ERA therapy medications used during the double-blind treatment period (Period 1)
- non-ERA therapy medications used during the open-label extension period (Period 2)

The duration of use of each type of ERA therapy at baseline will also be summarized.

A by-patient listing of previous PAH therapies and concomitant medications will be presented for all randomized patients.

Study Drug Exposure:

Study drug exposure will be listed by patient and summarized by treatment group. For all randomized patients in the double-blind treatment period (Period 1) and open-label extension period (Period 2), the total number of days of exposure, the cumulative number of doses taken,

and the average number of doses taken per week will be summarized using descriptive statistics by treatment group.

Efficacy Analyses:

6MW:

The analyses as described in Section 6.9.1.1 will be provided with the possible following exceptions. Within the model, the effect of PAH etiology will only be included if there are at least 3 patients per treatment at each PAH Etiology level. Similarly, the effect of ERA therapy will only be included if there are at least 3 patients per treatment at each ERA therapy level. Additionally, if the model fails to converge even once a heterogeneous auto regressive covariance structure is assumed, a standard Toeplitz covariance structure, followed by the standard auto regressive level 1 covariance structure, will be used. Should these fail to converge, only summaries by treatment and visit will be provided. The overall treatment difference p-value and the visit-wise p-values will not be reported. In addition, a similar analysis using the methods described in Section 6.9.1.1 will be provided for percentage change from baseline.

The 6MW data in Period 2 will be summarized for each visit, overall, and by original randomized treatment assignment in Period 1.

Clinical Worsening (CW):

Summary tables for the incidence of CW will be provided for Period 1 and Period 2.

World Health Organization (WHO) Functional Class:

The proportion of patients who experience a change from baseline to endpoint in WHO functional class will be provided. An additional table will be performed with changes as categorized by "worsening," "no change," or "improving" over the study period.

Echocardiogram:

The analyses included in Section 6.9.3.3 will be provided. The overall treatment difference p-value and the visit-wise p-values will not be reported.

In addition, descriptive statistics will be presented by treatment group for the changes from baseline in echocardiography parameters at Visits 5, 7, and 9.

N-terminal prohormone brain natriuretic peptide:

The analyses included in Section 6.9.3.6 will be provided. The overall treatment difference p-value and the visit-wise p-values will not be reported.

Child Health Questionnaire Parent Form 28:

The analysis included in Section 6.9.3.7 will be provided. The p-value will not be provided.

Safety Analyses:

Adverse Events (AEs):

The following AE summary tables will be provided:

- pre-existing conditions
- an overview of AEs, including deaths, serious adverse events (SAEs)
- adverse events leading to discontinuation, treatment-related AEs and treatment-emergent adverse events (TEAEs)
- treatment-emergent adverse events by Preferred Term (PT) by descending incidence
- treatment-emergent adverse events by System Organ Class (SOC) and PT
- treatment-emergent adverse events by descending incidence
- treatment-emergent adverse events by severity
- serious adverse events by descending incidence

All SAEs, deaths, and discontinuations due to an AE, and AEs of special interest will be listed. Some events reported in these categories may not be defined as treatment-emergent. Deaths will also be counted as SAEs and AEs leading to discontinuation.

Clinical laboratory tests:

Laboratory data will be listed by laboratory parameter for each patient. Summary statistics (including number of patients, mean, standard deviation [SD], minimum, and maximum) of the raw and change-from-baseline values for these parameters will be computed and will be presented using box plots for each treatment group at each visit during the double-blind treatment period.

In addition, shift tables will be presented for each of the laboratory parameters. The tables will show the percentages of patients with shifts in laboratory results from baseline to each postbaseline visit using categories (low, normal, high) based on the central laboratory reference ranges.

Inhibin B biomarker:

Inhibin B values collected in patients who are less than 9 years of age are considered exploratory. As such, separate analyses of Inhibin B values will be conducted for patients <9 years of age, ≥9 and <13 years of age, and patients ≥13 years of age. Changes from baseline to Years 1 and 2 will be summarized descriptively. A by-patient listing will be provided.

Vital signs:

Vital signs data (supine systolic blood pressure (SBP), supine diastolic blood pressure (DBP), supine heart rate) will be collected at each visit in the double-blind treatment period (Period 1) and will be listed for each patient. Summary statistics (including mean, SD, and median) of the raw and change-from-baseline values for these parameters (including height and weight) will be computed for each treatment group at each visit and at endpoint (LOCF). A by-patient listing will be provided.

Intellectual Ability and Cognitive Functioning Assessment:

Summary statistics (including mean, SD, and median) will be computed at each visit for raw values and change-from-baseline for each treatment group. A by-patient listing will be provided.

Tanner Score:

Shift tables will be presented for each of the Tanner stage parameters (pubic hair and breast score for females; pubic hair and genital score for males). The tables will show the percentages of patients with shifts in Tanner scores from baseline (Day 1) to each postbaseline visit (Years 1 and 2 during the open label extension). No comparisons are planned for Tanner score.

Eye Examination

Shift tables will be presented for each of the postbaseline eye examination results. The tables will show the percentages of patients with shifts in eye examination results from baseline to each postbaseline visit (Week 24 and Year 2).

Appendix 4. Analysis Plan for Extrapolation

The Bayesian paradigm provides a natural framework for extrapolating information accumulated in adult Study LVGY to the current pediatric Study LVHV. In this section, the Bayesian model is presented including the model and mixture prior specification as well as decision criteria. The simulation plan and results to assess the performance of the model are also presented here which included virtual patient scenarios, prior specification, and performance or operating characteristics of the study and model (power and false positive rates). Furthermore, to better understand the impact of various prior distributions using the Bayesian model, we provided the resulting posterior distribution plots for 3 simulated trials.

Model specification

The mathematical details regarding the planned Bayesian mixed model repeated measures (MMRM) approach are as follows: In the primary analysis, let $y_{ij}, i = 1, \dots, n, j = 1, \dots, m_i$ denote the *j*th repeated measurement of the response (change in 6-minute walk distance (6MWD) from baseline at Months 1, 2, 3, 4, 5, and 6) for the ith patient receiving at the time point of assessment. The model may include covariates corresponding to factors including visit, baseline (Day 1) 6MWD, pulmonary arterial hypertension (PAH) etiology, type of endothelin receptor agonist (ERA) therapy, treatment group, and interaction terms treatment-by-visit and treatment-by-baseline.

Then y_{ij} is modeled as,

$$\begin{aligned} y_{ij} = {} & \beta_1 t_{i1} + \beta_2 t_{i2} + \beta_3 t_{i3} + \beta_4 t_{i4} + \beta_5 t_{i5} + \beta_6 t_{i6} \\ & + (\beta_7 t_{i1} + \beta_8 t_{i2} + \beta_9 t_{i3} + \beta_{10} t_{i4} + \beta_{11} t_{i5} + \beta_{12} t_{i6})\mathrm{trt}_i + \beta_{\mathrm{PAH}}\mathrm{PAH}_i \\ & + \beta_{\mathrm{ERA}}\mathrm{ERA}_i + (\beta_{b1} t_{i1} + \beta_{b2} t_{i2} + \beta_{b3} t_{i3} + \beta_{b4} t_{i4} + \beta_{b5} t_{i5} + \beta_{b6} t_{i6})\mathrm{baseline}_i \\ & + \varepsilon_{ij}. \end{aligned}$$

Here t_{ij} are indicator variables for month *j*, $\mathrm{trt}_i = \{0,1\}$ is the treatment indicator for patient *i* (0 = placebo, 1 = tadalafil), PAH_i is the indicator variable for PAH etiology for patient *i*, ERA_i is the indicator variable for type of ERA therapy, $\mathrm{baseline}_i$ is the numeric baseline score for patient *i*, and ε_{ij} is the normally distributed residual error accounting for repeated measures within patient *i*, month *j*.

The Bayesian MMRM model follows the generalized standard linear model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

The residual error vector, $\boldsymbol{\varepsilon}_i$, of patient i over time (ie, $\boldsymbol{\varepsilon}_i = [\boldsymbol{\varepsilon}_{i1}, \boldsymbol{\varepsilon}_{i2}, \boldsymbol{\varepsilon}_{i3}, \boldsymbol{\varepsilon}_{i4}, \boldsymbol{\varepsilon}_{i5}, \boldsymbol{\varepsilon}_{i6}]$) is modeled using a multivariate normal distribution with zero mean and covariance matrix, $\boldsymbol{\Sigma}$. Vector $\boldsymbol{\beta}$ consists of all of the effects of each *p* variable described in the model, and $\boldsymbol{X}_i$ is a matrix with 6 rows, each containing the values of the variables for patient *i*.

Full specification of the Bayesian MMRM model requires prior distributions for each element of effect vector $\boldsymbol{\beta}$, as well as the covariance matrix $\boldsymbol{\Sigma}$. This section describes the prior distributions for all parameters, including details of the informative mixture prior approach utilized for tadalafil's effect over time.

Diffuse Prior Distributions

The prior on $\boldsymbol{\Sigma}$, is,

$$\boldsymbol{\Sigma} \sim \text{Inv} - \text{Wishart}_6(\mathbf{I}),$$

where **I** is the 6x6 identity matrix and the degrees of freedom is set to 6, which resulted in a diffuse prior that yields essentially equivalent inference to an unstructured covariance matrix in a traditional MMRM analysis.

Diffuse independent normal priors were used on all β parameters except those parameters associated with the treatment effect at each time point (ie, β_7, β_8, β_9, β_{10}, β_{11}, β_{12}). The prior variance of these priors were chosen to be diffuse over the range of plausible effects such that the resulting inference for these parameters would almost entirely be driven by the observed data in pediatric Study LVHV.

Informative Mixture Prior Distribution on the Tadalafil Effect

In order to borrow treatment information from available adult data (Study LVGY), we used the mixture prior approach described in Ye and Travis (2017). Given limited sample sizes of the pediatric data in Study LVHV, we increased precision of inferences via careful specification of prior distributions for the β parameters associated with the treatment effect across time variable (ie, β_7, β_8, β_9, β_{10}, β_{11}, β_{12}). We fit a Bayesian model to the adult Study LVGY and used the resulting posterior distribution as a component for the mixture prior distribution.

The mixture prior approach is well suited for a pediatric trial where partial extrapolation from adult data is acceptable. The mixture prior distribution is a weighted combination of 2 component prior distributions, a skeptical prior and an adult prior. The skeptical prior is centered around no effect, and the adult prior is centered around the adult effect.

This approach does not assume that the adult and pediatric effects are the same. It borrows from the adult component of the mixture prior if the pediatric effect is similar to the adult effect. Alternatively, if the pediatric effect is not favorable (and therefore not comparable to the adult data), the skeptical component of the mixture prior shrinks the pediatric effect toward zero (no effect).

This prior specification of certain parameters of interest β associated with the treatment effect across time variable is given the following form with w being the weight of the adult component prior, skeptical prior being normally distributed with mean 0 and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{S}}$, and adult prior with mean $\boldsymbol{\mu}_{\boldsymbol{A}}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{A}}$:

$$\begin{bmatrix} \boldsymbol{\beta_7} \\ \boldsymbol{\beta_8} \\ \boldsymbol{\beta_9} \\ \boldsymbol{\beta_{10}} \\ \boldsymbol{\beta_{11}} \\ \boldsymbol{\beta_{12}} \end{bmatrix} \sim (\mathbf{1} - \boldsymbol{w})\, \boldsymbol{N}(\mathbf{0}, \boldsymbol{\Sigma_s}) + \boldsymbol{w}\, \boldsymbol{N}(\boldsymbol{\mu_A}, \boldsymbol{\Sigma_A})$$

Adult Component of the Mixture Prior Distribution

The adult study (Study LVGY) was utilized to build the adult component prior distribution $N(\mu_A, \Sigma_A)$. A total of 405 patients were randomized equally to placebo, LY 2.5 mg, LY 10 mg, LY 20 mg, and LY 40 mg. A subset of 216 patients had Bosentan as ERA therapy and would be similar to the patient population in the pediatric study (Study LVHV). The highest 2 doses in the adult study are most relevant to the doses which are being studied in the pediatric (Study LVHV) and approximately 81**%** of the observed tadalafil area under the concentration versus time curve at steady state (AUC_{SS}) from Study LVGY predicted following 40-mg once-daily administration were within the fifth to 95th percentiles of those estimated following 20 mg. We pooled the data from these doses to get an estimate of the mean and standard error of the treatment difference versus placebo in 6MWD change from baseline across the time points. Hence, a subset of 132 patients was the focus of the analyses to form the adult component of the mixture prior distribution.

Since the adult study (Study LVGY) collects 6MWD endpoints up to Month 4, we need to provide estimates for the change from baseline in 6MWD at Months 5 and 6. We used a model-based approach, Integrated Two-Component Prediction (ITP) model, to provide mean and variance estimates across all time points up to 6 months.

The ITP model used for this analysis is based on Fu and Manner (2010), with some modifications. The model is:

$$Y_{ijl} = \lambda_i \frac{1 - \exp(-k_i t_{ijl})}{1 - \exp(-k_i d)} + s_j + \epsilon_{ijl}$$

where Y_{ijl} represents an observation from dose level i, subject j, at time l. Here, we assume that the between subject random effect is $s_j \sim N(0, \sigma_s^2)$ and the usual error term $\epsilon_{ijl} \sim N(0, \sigma^2)$. This model assumes one random effect per subject and a constant variance over time, which is consistent in what is seen in the adult tadalafil data. We used the following priors on the ITP model:

$\lambda_i \sim N(0,100^2)$

$k_i \sim$Uniform(0,0.75)

$1/\sigma_s^2 \sim$Gamma(1,0.01)

$1/\sigma^2 \sim$Gamma(1,0.01)

where Gamma(a,,b) is a gamma distribution with shape a and rate b. The priors are generally diffuse and non-informative. However, the model is fairly sensitive to the selection of the

hyperparameters for k_i. Therefore, using data from the literature, it was determined through simulation that the selected prior produced mean curves which are consistent in shape to previously observed studies.

The posterior distribution of the parameters was sampled from using Markov chain Monte Carlo (MCMC) methods. The subject level random effects were marginalized analytically, therefore only the k_i, λ_i, σ_S, and σ need to be sampled. A Metropolis-Hastings algorithm was used to generate samples from the posterior. The MCMC algorithm was run for 10^7 iterations following a burn-in period of 10^4 iterations. A thinning interval of 10 was used, yielding a total of 10^6 MCMC samples.

The resulting adult component of the mixture prior distribution is:

$$N(\mu_{\boldsymbol{A}}, \boldsymbol{\Sigma_A}) = N\left(\begin{bmatrix} 8 \\ 15 \\ 19 \\ 21 \\ 23 \\ 24 \end{bmatrix}, \begin{bmatrix} 67 & 69 & 63 & 56 & 50 & 45 \\ 69 & 78 & 75 & 71 & 67 & 63 \\ 63 & 75 & 77 & 76 & 75 & 73 \\ 56 & 71 & 76 & 78 & 79 & 79 \\ 50 & 67 & 75 & 79 & 82 & 85 \\ 45 & 63 & 73 & 79 & 85 & 89 \end{bmatrix}\right)$$

We also looked at another variation to this ITP model by assuming nonconstant independent variance over time. The resulting adult component of the mixture prior distribution has a mean vector similar to the mean vector above; the variance component is slightly higher by up to 5% compared to the variance component above.

Skeptical Prior Distribution

The skeptical prior distribution was developed to balance the informativeness of the adult prior component and therefore control the probability of a false positive result. It is centered at the null hypothesis of no effect to represent the possibility that the beneficial effect of tadalafil in adults with PAH does not translate to pediatrics.

The variance of the skeptical prior distribution σ_s^2 was chosen to control the false positive probability of the study. We looked at a variety of σ_S ranging from 9 to 100 and found that this parameter impacted the simulation results significantly such that increasing σ_S led to higher false positive rate. For example, for a given set of parameters in the mixture prior, changing the σ_S from 9 to 15 increased the false positive rate from 10% to 25%.

The resulting skeptical component of the mixture prior distribution is:

$$N(\mathbf{0}, \boldsymbol{\Sigma_s}) = N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_s^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_s^2 \end{bmatrix}\right)$$

where σ_S was chosen to be 9 to keep the standard deviation (SD) of the mixture prior and ultimately the false positive rate at a reasonable level. This choice of σ_S is also similar to the SD of the adult component of the mixture prior as shown previously.

Weight Component of the Mixture Distribution

Although empirical data is the preferred method to justify the prior, limitations of the data and/or interpretation of the data as well as translational gap between the empirical data and constructed model can prohibit one from constructing a prior distribution directly (Dallow et al. 2018). Formal expert elicitation is a reasonably robust technique to quantify the key parameter in the mixture prior distribution that governs the amount of extrapolation from adult to pediatric patients. Therefore, the weight of the adult component prior was determined using elicitation from medical experts.

The exercise included 4 experts (2 external pediatric PAH experts and 2 internal experts). Two external pediatric PAH experts (one from the EU and another from the US) are well-known pediatric cardiologists who are specialized in pediatric PAH for many years. A short survey consisting of 2 questions (adapted from Ye and Travis 2017 [WWW]) was completed by each expert. The outcome of this survey is provided below.

Question 1: On a scale of 0 to 10, how much confidence do you have in applying adult PAH clinical trial data to make decisions on PAH treatment effect for pediatric patients?

0	1	2	3	4	5	6	7	8	9	10
Completely ignore adult data and demand that all evidence come from specific studies conducted within the pediatric patient population										Fully trust the adult patient data as applicable to pediatric patient population

Question 2: On a scale of 0 to 10, how much confidence do you have in applying tadalafil adult PAH prevention clinical trial data to make decisions on PAH tadalafil treatment effect for pediatric patients?

0	1	2	3	4	5	6	7	8	9	10
Completely ignore adult data and demand that all evidence come from specific studies conducted within the pediatric patient population										Fully trust the adult patient data as applicable to pediatric patient population

Below are the tabulated results of the 4 experts for each question:

	Question 1	Question 2
Expert 1	9	8
Expert 2	8	8
Expert 3	8	8
Expert 4	7	7
Median	8	8

The median for Questions 1 and 2 is 8. This provides an estimate of the mixture prior weight *w* of 0.8. To understand the impact of the weight mixture, we also looked at the properties of the mixture prior distribution using a weight of 0.5 as shown below.

Resulting Mixture Prior Distribution

Mixture prior distribution with a weight of 0.8

By combining the adult and skeptical prior components and a weight of 0.8 on the adult prior distribution, we have the following mixture prior distribution for Month 6 (Figure APP.4.1).

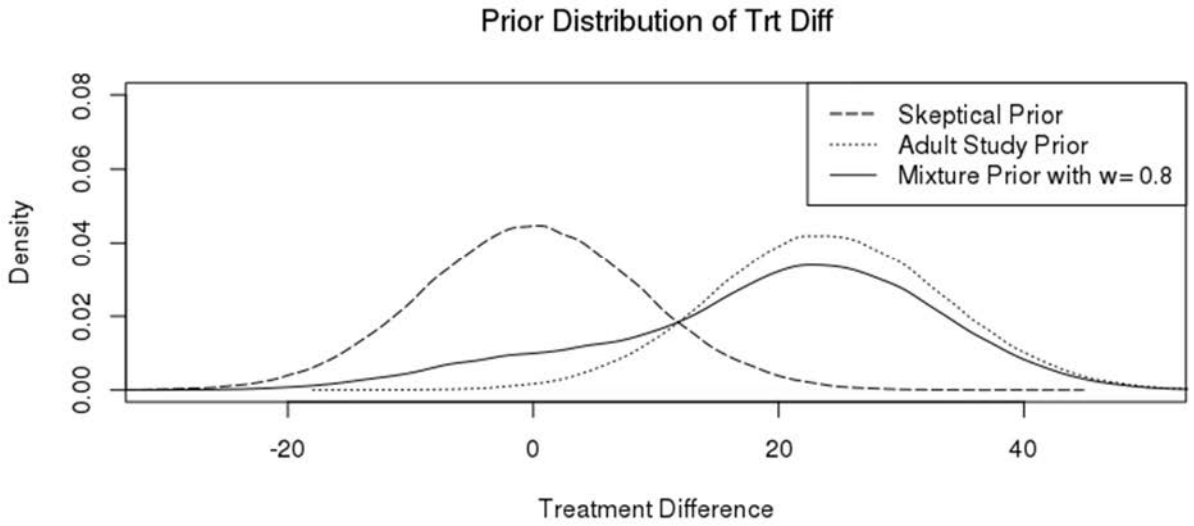


Figure APP.4.1. Mixture prior distribution of treatment difference for Month 6 (weight 0.8 on the adult prior distribution).

The mixture prior distribution with a weight of 0.8 has a treatment difference mean of 19.2, standard deviation of 13.4, and median of 21. In the FDA 2010 [WWW] Bayesian guidance document, the prior probability of success is defined as the probability of the study claim before seeing any new data. For this study, we define success as the probability the treatment difference is greater than 0. Using the mixture prior distribution, the probability of treatment difference being greater than 0 is 89.56%. Furthermore, to understand the amount of information contained in the prior distribution, we calculated the prior effective sample size (ESS) of the parameters of

interest β_{12}. The ESS was calculated using a moments-based approach (see Morita et al. 2008). The R function "ess" within the R Package RBesT (R Bayesian Evidence Synthesis Tools) was used to calculate the prior effective sample size for Month 6 to be 32.

Mixture prior distribution with a weight of 0.5

To understand the impact of the weight mixture, we also looked at the properties of the mixture prior distribution using a weight of 0.5. With this weight, we have the following mixture prior distribution for Month 6 (Figure APP.4.2).

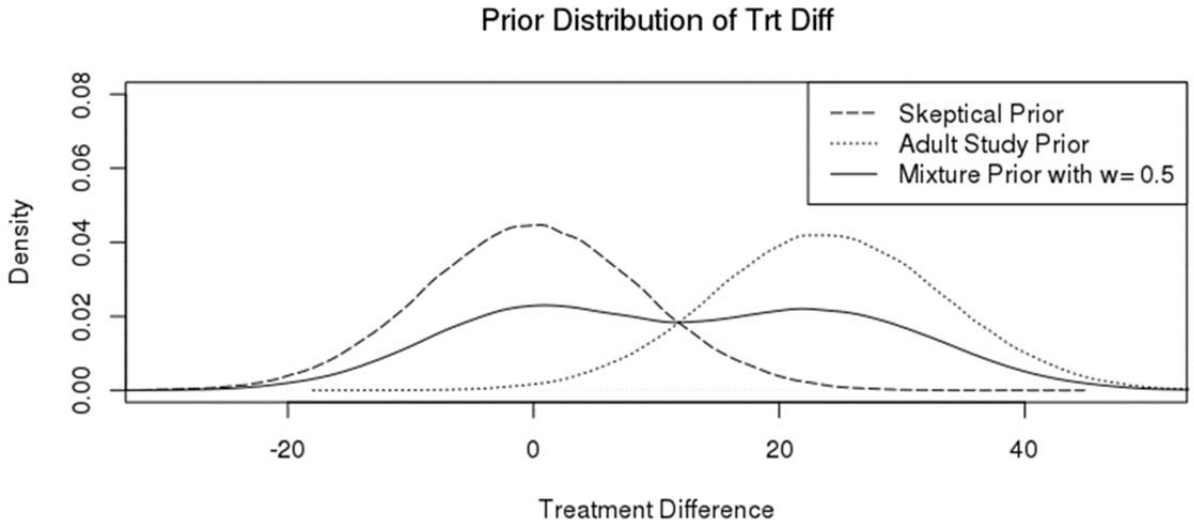


Figure APP.4.2. Mixture prior distribution of treatment difference for Month 6 (weight 0.5 on the adult prior distribution).

The mixture prior distribution with a weight of 0.5 has a treatment difference mean of 12, standard deviation of 15.1, median of 11.7, and 74.74% probability of treatment difference being greater than 0. The effective sample size is 24.

Sensitivity analyses were conducted to compare simulation results from our chosen mixture prior with a weight of 0.8 to a mixture prior with a weight of 0.5. Furthermore, sensitivity analyses were also conducted to assess impact on using diffuse priors across all parameters. The results from the model with the diffuse prior provide comparison to what one would expect under a frequentist analysis with no extrapolation.

Decision Criteria

The primary endpoint is difference in change from baseline in 6MWD between the treatment and the control group at month 6, β_{12}. In the simulation plan, the decision criterion for declaring that tadalafil is efficacious in children is:

$$\Pr(\beta_{12} > 0) > t,$$

where *t*, the probability threshold, is 0.95. The success criterion threshold was chosen based on simulation, with the goal of balancing the tradeoff between the false positive control rate and probability of study success. The posterior probability threshold was chosen to require posterior evidence similar to the traditional p-value approach based on a one-sided p-value of .05. We also looked at *t* of 0.90 and 0.975 to understand the impact of these values to the operating characteristics of the study.

Simulation Plan

Thirty-four pediatric patients randomized equally to placebo and tadalafil have been enrolled in Study LVHV. To evaluate the performance and operating characteristics of the Bayesian MMRM model using mixture prior distribution as described above, we conducted a simulation study consisting of 5000 trials per scenarios. Trial simulation requires complete specification of the virtual patient response, scenarios of truth, and analysis methods under which to summarize the operating characteristics.

Virtual Patient Response

Patient-level longitudinal data was generated from a multivariate normal distribution. For each pediatric patient, 6 longitudinal data points were simulated that represent the change from baseline in 6MWD at Months 1, 2, 3, 4, 5, and 6 respectively. The unstructured 6x6 variance-covariance matrix that was used to simulate both placebo and tadalafil patient data is given as:

$$\Sigma = \begin{bmatrix} 2078 & 1463 & 1323 & 1075 & 1000 & 950 \\ 1463 & 2266 & 1628 & 1543 & 1500 & 1450 \\ 1323 & 1628 & 2237 & 1814 & 1900 & 1950 \\ 1075 & 1543 & 1814 & 2858 & 1900 & 1950 \\ 1000 & 1500 & 1900 & 1900 & 2858 & 2500 \\ 950 & 1450 & 1950 & 1950 & 2500 & 2858 \end{bmatrix}$$

We conducted several scenarios for treatment effect at Month 6 for the simulation study (Table APP.4.1). These assumptions are based on a subset of 132 patients from the adult Study LVGY who had Bosentan as ERA therapy and who received either placebo, LY 20 mg, or LY 40 mg which would be similar to the pediatric patient population in Study LVHV. We also used the ITP model to provide estimates of treatment difference at Months 5 and 6. Standard deviation of 53 is based on the adult Study LVGY MMRM var-cov matrix estimates.

Table APP.4.1 Scenarios for Treatment Effect at Month 6

Scenario	Treatment difference (SD) at Month 6	Effect size at Month 6
Tadalafil is no different than placebo (Null)	0 (53)	0
25% effect of optimistic scenario	6 (53)	0.11
50% effect of optimistic scenario	12 (53)	0.23
75% effect of optimistic scenario	16 (53)	0.3
Optimistic	24 (53)	0.45

Abbreviations: SD = standard deviation.

For the longitudinal profile of the 6MWD, we assumed the following (Table APP.4.2):

Table APP.4.2. Proportion of Treatment Difference and SD at Month 6

Time points	Proportion of treatment difference at Month 6	Proportion of SD at Month 6
Month 1	33%	86%
Month 2	60%	90%
Month 3	77%	90%
Month 4	88%	100%
Month 5	96%	100%

Abbreviations: 6MWD = 6-minute walk distance; SD = standard deviation.

Analysis Methods and Simulation Results

Once a simulated dataset was generated, the Bayesian MMRM model with the mixture prior as described previously was fitted; covariate data was not utilized in the simulations because the purpose is to assess the performance of the key features of the proposed analysis model.

Sensitivity analyses were also conducted to assess impact on using diffuse priors across all parameters as well as using different weight parameter on the mixture prior such as 0.5. The results from the model with the diffuse prior provide comparison to what one would expect under a frequentist analysis with no extrapolation.

The primary statistic of interest from each simulation was the posterior probability that the treatment difference was greater than 0, ie, $\Pr(\beta_{12} > 0)$. The posterior probabilities were then compared with the probability threshold of 0.95. The percentage of times each simulated dataset achieved study success (ie, study power) is calculated under a range of effect sizes and various prior distribution, and shown in Table APP.4.3 below. The operating characteristics provided here are based off a total of 5000 simulations for each scenario.

Table APP.4.3. Results of the Simulation Study

Scenario	Diffuse prior	Mixture prior w=0.5	Mixture prior w=0.8
Null	0.05	0.07	0.20
25% effect of optimistic scenario	0.097	0.11	0.29
50% effect of optimistic scenario	0.16	0.18	0.41
75% effect of optimistic scenario	0.25	0.27	0.53
Optimistic	0.358	0.364	0.65

The false positive rate, which is the power under the null scenario, was higher when the mixture prior was utilized, compared with the diffuse prior. This false positive rate increased by 2 and 15 percentage points, respectively as the mixture prior weight increased from 0.5 to 0.8 compared to diffuse prior, respectively. The increase of false positive rate when using the mixture prior distribution is anticipated in most Bayesian analysis settings. Depending on the assumed effect size, there was up to a 29 percentage point increase in study power when the mixture prior with 0.8 weight was utilized, compared with the diffuse prior. In addition to the probability threshold of 0.95, we also assessed the false positive rate using probability thresholds from 0.90 to 0.975 under the 3 prior distributions as shown in Figure APP.4.3 below. As one would expect, increasing the probability threshold resulted in a decrease in the false positive rate for a given prior distribution.

In conclusion, through our simulation study we have shown that combining the relatively small pediatric sample size of 34 with the mixture prior chosen weight of 0.8 based on expert elicitation (which translates to a prior effective sample size of 32), resulted in a false positive rate of 0.2 and a study powered modestly at 0.65 (although greatly increased compared to 36% for diffuse prior).

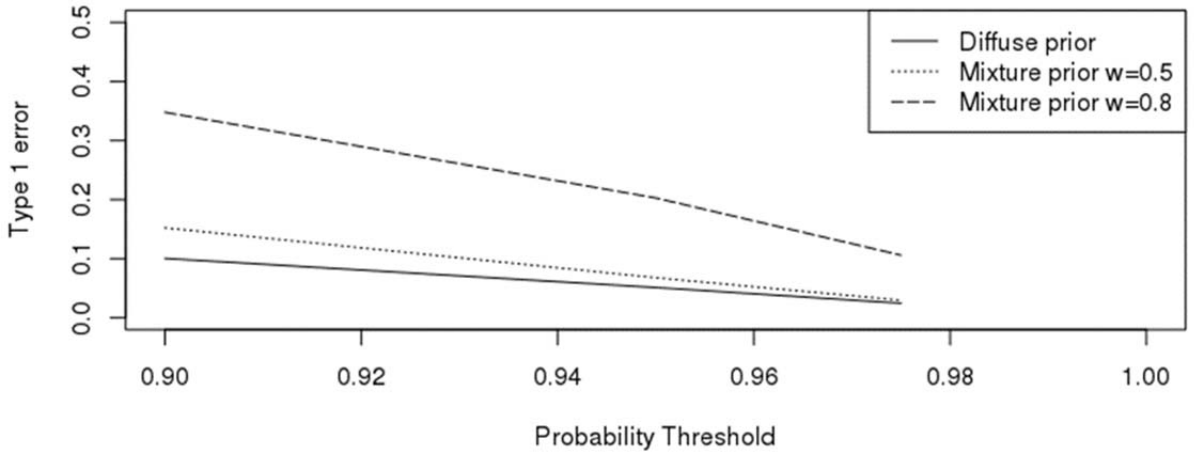


Figure APP.4.3. Assessment of type 1 error properties using probability thresholds of 0.90, 0.95, and 0.975 under the 3 prior distributions.

Individual Realization of Simulated Datasets

To better understand the impact of various prior distributions using the Bayesian MMRM model described previously, we provided the resulting posterior distribution plots for 3 simulated trials. The selected simulated trials are intended to show a broad range of individual realizations of simulated datasets including those with a treatment difference of approximately 0 (Figure APP.4.4), an effect size of 0.23 which is approximately half of the effect size of the adult study (Study LVGY) (Figure APP.4.5), and an effect size of 0.45 similar to Study LVGY (Figure APP.4.6). The plots include the posterior distributions using the diffuse prior and mixture priors with weights of 0.5 and 0.8. The posterior distribution with a diffuse prior can be viewed as the distribution of the likelihood (ie, the distribution of the observed data that is not influenced by the prior distribution).

The posterior distribution is influenced by how similar the observed data are to each of the components that comprise the mixture prior distribution: the adult prior component and the skeptical prior component.

In Figure APP.4.4, the observed data centered at a treatment difference of 0 are similar to the skeptical component of the mixture prior (w=0.5) and the resulting posterior distribution heavily borrows from the skeptical component. The observed data centered at 0 is less similar to the mixture prior (w=0.8) and the resulting posterior distribution is a bimodal distribution impacted by the data and prior proportionately. The posterior probability of treatment difference being greater than 0 were 0.42, 0.6, and 0.75 under diffuse prior, mixture prior (w=0.5), mixture prior (w=0.8), respectively.

In Figure APP.4.5, the observed data centered at 12 (effect size of 0.23) combined with the mixture prior (w=0.5) resulted in posterior distribution which heavily borrows from the skeptical component; when combined with the mixture prior (w=0.8) resulted in posterior distribution which borrows more from the adult component. In the latter situation, the adult component essentially has impact on the posterior distribution. The posterior probability of treatment difference being greater than 0 were 0.64, 0.7, and 0.88 under diffuse prior, mixture prior (w=0.5), mixture prior (w=0.8), respectively.

Finally, in Figure APP.4.6, the observed data centered at 24 (effect size of 0.45) are similar to the adult component of both mixture priors (w=0.5 and 0.8) and hence the resulting posterior distributions heavily borrows from the adult component. The posterior probability of treatment difference being greater than 0 were 0.94, 0.96, and 0.99 under diffuse prior, mixture prior (w=0.5), mixture prior (w=0.8), respectively.

In conclusion, Figure APP.4.4, Figure APP.4.5, and Figure APP.4.6 showed the impact of the mixture prior on the posterior distribution in such a way that the posterior distribution shifts and borrows more from the adult component of the mixture prior when the observed data is similar to the adult data. Conversely, the posterior distribution shifts and borrows more from the skeptical component of the mixture prior when the observed data is similar to the skeptical component.

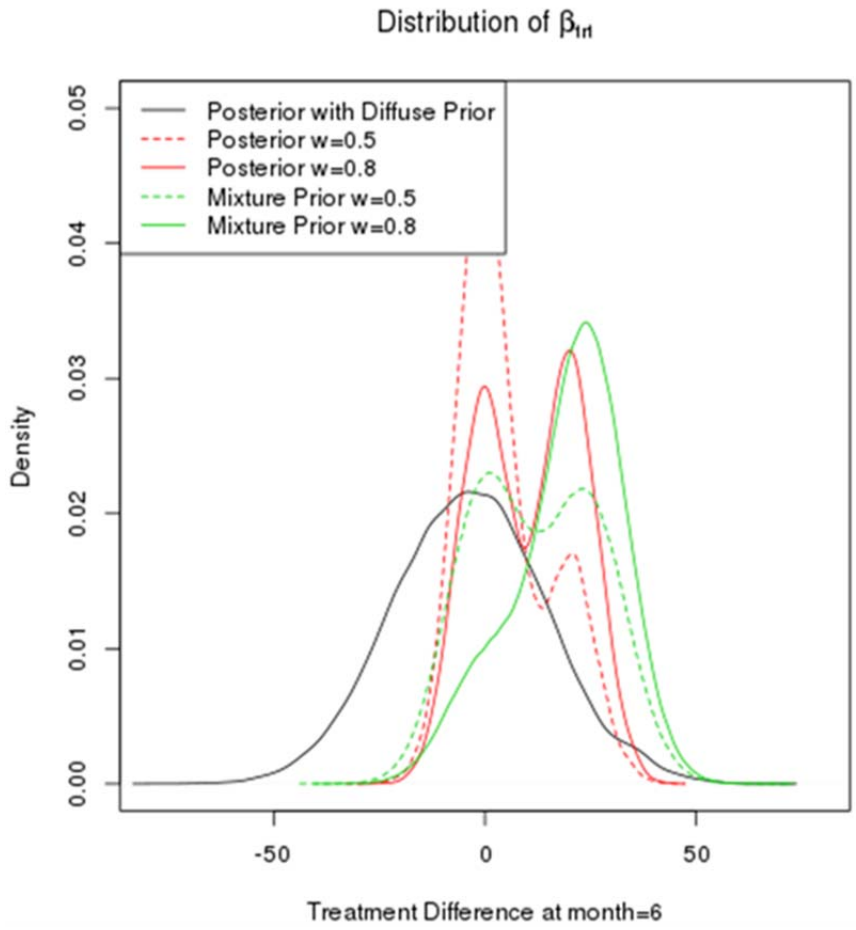


Figure APP.4.4. **Prior and posterior distributions from a simulated dataset under assumption of effect size=0.**

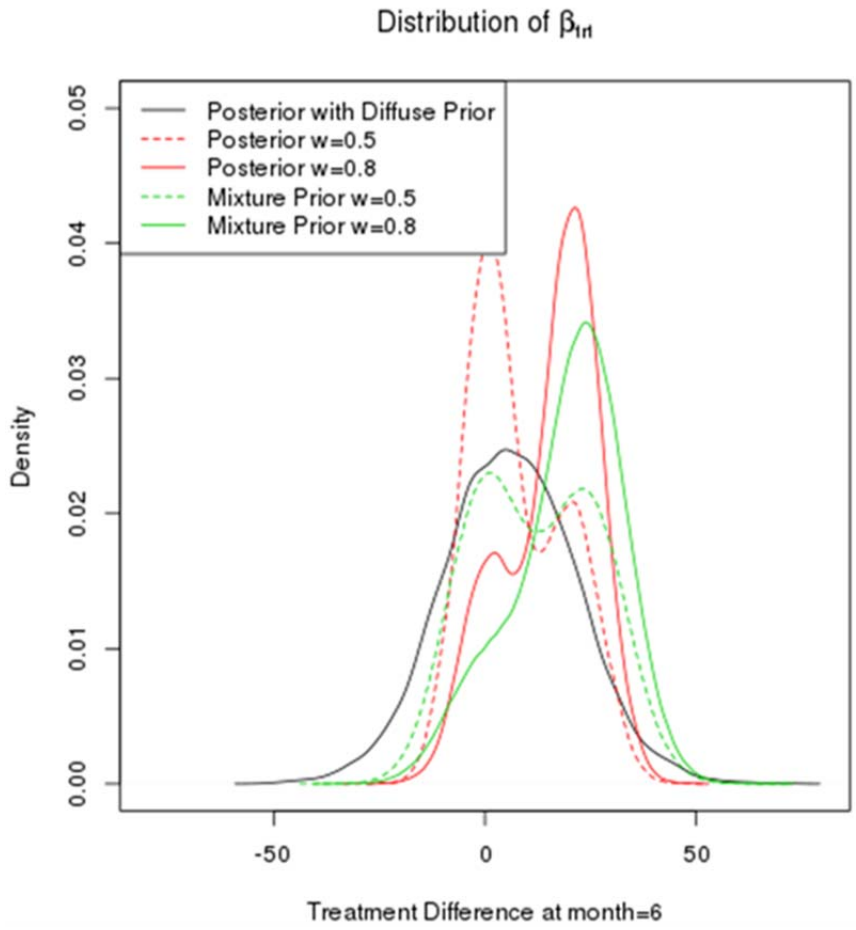


Figure APP.4.5. **Prior and posterior distributions from a simulated dataset under assumption of effect size=0.23.**

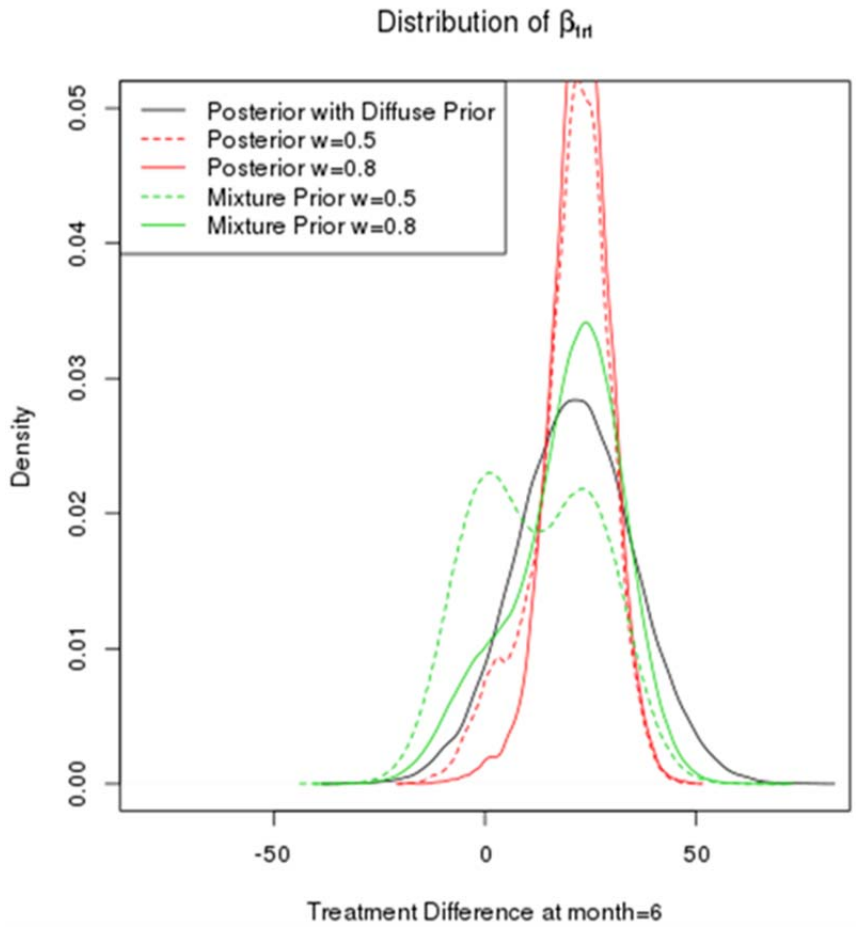


Figure APP.4.6. Prior and posterior distributions from a simulated dataset under assumption of effect size =0.45.

Conclusion

In the original pediatric study (Study LVHV) protocol, the planned sample size of 134 randomized patients was estimated to provide approximately 80% power to detect a hazard ratio of 3.6 in the time to clinical worsening in 6 months. This power estimate was based on a 2-sided 0.3 significance level. Utilizing the 2-sided 0.3 significance level would have led to a study of reasonable sample size that could have shown "trends" in time to clinical worsening.

Under the circumstances of difficult enrolment, the study will have recruited at least 34 patients (the current number randomized). We propose a Bayesian MMRM model with mixture prior approach, which dynamically and adaptively "borrows" evidence from adult data from Study LVGY based on the similarity of the pediatric data coming from the ongoing Study LVHV. Through our simulation study using the following assumptions

- Bayesian MMRM model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \varepsilon$$

$$\boldsymbol{\varepsilon}_i \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\Sigma} \sim \text{Inv} - \text{Wishart}_6(\mathbf{I})$$

- Diffuse independent normal priors were used on all β parameters except those parameters associated with the treatment effect at each time point (month 1, 2, 3, 4, 5, and 6) as shown below

$$\begin{bmatrix} \boldsymbol{\beta_7} \\ \boldsymbol{\beta_8} \\ \boldsymbol{\beta_9} \\ \boldsymbol{\beta_{10}} \\ \boldsymbol{\beta_{11}} \\ \boldsymbol{\beta_{12}} \end{bmatrix} \sim (\mathbf{1} - \boldsymbol{w})\, \boldsymbol{N}(\mathbf{0}, \boldsymbol{\Sigma_s}) + \boldsymbol{w}\, \boldsymbol{N}(\boldsymbol{\mu_A}, \boldsymbol{\Sigma_A})$$

- Adult component of the mixture prior

$$N(\mu_{\boldsymbol{A}}, \boldsymbol{\Sigma_A}) = N\left(\begin{bmatrix} 8 \\ 15 \\ 19 \\ 21 \\ 23 \\ 24 \end{bmatrix}, \begin{bmatrix} 67 & 69 & 63 & 56 & 50 & 45 \\ 69 & 78 & 75 & 71 & 67 & 63 \\ 63 & 75 & 77 & 76 & 75 & 73 \\ 56 & 71 & 76 & 78 & 79 & 79 \\ 50 & 67 & 75 & 79 & 82 & 85 \\ 45 & 63 & 73 & 79 & 85 & 89 \end{bmatrix}\right)$$

- Skeptical component of the mixture prior where σ_s was chosen to be 9

$$N(\mathbf{0}, \boldsymbol{\Sigma_s}) = N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_s^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_s^2 \end{bmatrix}\right)$$

- Decision criteria: treatment difference at month 6 was greater than 0 (ie, $\Pr(\beta_{12} > 0) > 0.95$)

We have shown that combining the relatively small pediatric sample size of 34 with the mixture prior chosen weight, *w*, of 0.8 based on expert elicitation (which translates to a prior effective sample size of 32), resulted in a false positive rate of 0.2 and a study powered modestly at 0.65. Under the circumstances of the study enrolment, we consider the properties of this study design and Bayesian model approach are reasonable to help ensure that we are able to detect that numerical mean change in 6MWD trending in the right direction.

Implementation Details

Once the datalock has occurred, the posterior distribution of the Bayesian model specified above in the conclusion section will be computed. Other model specification may be considered as deemed appropriate. The model will be performed using a MCMC algorithm with 3 chains, and initial values for the parameters drawn from their corresponding prior distributions. An appropriate burn-in will be used with at least 10,000 iterations to ensure that each MCMC chain has reached convergence to the posterior distribution. After the burn-in period , MCMC samples will be obtained from each chain for all parameters in the model. Gelman-Rubin diagnostics will

be used to assess convergence across the chains. If convergence is not attained, the length of the burn-in will be increased and convergence will be reassessed. If convergence is still not attained, initial values will be set based on a maximum likelihood estimate of the model. This process will be repeated until convergence is obtained. The exact settings and seed values will be retained to ensure reproducibility.