Official Title:

Phase III Multicenter Open-label Randomized Clinical Trial Comparing Everolimus and Low Dose Tacrolimus to Tacrolimus and Mycophenolate Mofetil at 6 mo Post-Transplant to Prevent Long-term Complications After Pediatric Heart Transplantation

Brief Title:

Tacrolimus/Everolimus vs. Tacrolimus/MMF in Pediatric Heart Transplant Recipients Using the MATE Score (TEAMMATE)

NCT Number:

NCT03386539

Document Date:

April 12, 2023

STATISTICAL ANALYSIS PLAN

Protocol PR160574

Phase III Multi-Center Open Label Randomized Clinical Trial Comparing Everolimus and Low Dose Tacrolimus to Tacrolimus and Mycophenolate Mofetil at 6 Months Post Heart Transplant To Prevent Long-term Complications After Pediatric Heart Transplantation

Version Date	12 April 2023
Protocol Number:	PR160574
ClinicalTrials.gov ID:	NCT03386539
IND Number	#127980
Name of Test Drug(s):	Everolimus
Phase:	3
Methodology:	The primary aim of the TEAMMATE Trial is to improve the long-term survival of children undergoing heart transplantation by reducing transplant- and immunosuppression- related major adverse transplant events. The study is a Phase 3, open-label, 2-arm trial with randomization of 210 subjects in a 1:1 ratio to determine whether the combination of everolimus and low-dose tacrolimus compared with the combination of tacrolimus and mycophenolate myofetil reduces the risk of coronary artery vasculopathy, chronic kidney disease and biopsy proven-acute cellular rejection without an increase in graft loss due to all causes (e.g. infection). Each patient will be followed for 2.5 years. The trial has a group sequential design with two interim analyses, the results of which will be used by an independent Data and Safety Monitoring Board to recommend early termination for futility or efficacy, or trial continuation.
IND Sponsor:	Stanford University Christopher Almond, MD, MPH Co-Chair, TEAMMATE Trial Clinical Coordinating Center
Contacts:	Lynn Sleeper, ScD PI, TEAMMATE Trial Data Coordinating Center

Kevin Daly, MD
Co-Chair, TEAMMATE Trial Clinical Coordinating Center
Christopher Almond, MD, MPH
Co-Chair, TEAMMATE Trial Clinical Coordinating Center
Darrell Ellsworth PhD
U.S. Dept. of Defense Senior Science Officer

Document Date: 12 April 2023

Document Version: 1.0

SIGNATURE PAGE

Protocol Title:	Phase III Multi-Center Open Label Randomized Clinical Trial Comparing Everolimus and Low Dose Tacrolimus to Tacrolimus and Mycophenolate Mofetil at 6 Months Post Heart Transplant To Prevent Long-term Complications After Pediatric Heart Transplantation
IND Sponsor:	Stanford University Lucile Packard Children's Hospital 725 Welch Road Suite 325 Palo Alto, CA 94304 USA
Document Date/Version:	12 April 2023/v1.0

Author:

Lynn Sleeper, ScD

Signature:___ 

Date:____4/12/23_____

PI, TEAMMATE Data Coordinating Center
Boston Children's Hospital
Dept. of Cardiology
300 Longwoood Avenue
Boston, MA 02115

TABLE OF CONTENTS

Section **Page**

STATISTICAL ANALYSIS PLAN 1
Signature Page 3
TABLE OF CONTENTS 4
ABBREVIATIONS 6
1. INTRODUCTION AND OBJECTIVES OF ANALYSIS 8
1.1. Introduction 8
1.2. Objectives of Statistical Analysis 8
1.2.1. Primary Objective 8
1.2.2. Secondary Objective 8
2. STUDY DESIGN 10
2.1. Synopsis of Study Design 10
2.2. Randomization 10
2.2.1. Methodology 10
2.2.2. Unblinding 10
2.3. Stopping Rules and Unblinding 10
2.4. Study Procedures 11
2.5. Endpoints 11
2.5.1. General Comments on Analytic Approach 11
2.5.1. Primary Efficacy Endpoint 8
2.5.2. Secondary Efficacy Endpoints 8
2.5.4. Primary Safety Endpoint 10
2.5.4. Secondary Safety Endpoints 10
3. SUBJECT POPULATIONS 16
3.1. Population Definitions 16
4. STATISTICAL METHODS 17
4.1. Sample Size Justification 17
4.2. General Statistical Methods and Data Handling 15
4.2.1. General Method 15
4.2.2. Computing Environment 16
4.2.3. Standardized Coding 16

Section **Page**

4.2.4. Adjustments for Covariates....16
4.2.5. Multiple Comparisons/Multiplicity....16
4.2.6. Subpopulations....16
4.2.7. Withdrawals, Discontinuations, Lost to Follow-up....16
4.2.8. Missing and Unused Data....17
4.2.9. Change in Value from X to Y....176
4.3. Interim Analyses....17
4.4. Subject Disposition....18
4.5. Demographic and Baseline Characteristics....18
4.6. Treatment Exposure....18
4.7. Efficacy Evaluation....18
4.7.1. Primary Efficacy Endpoint....19
4.7.2. Secondary Efficacy Endpoints....19
4.8. Safety Evaluation....20
4.8.1. Primary Safety Endpoint....20
4.8.2. Secondary Safety Endpoints....21
4.9. Subgroup Analyses....22
5. CHANGES TO PLANNED ANALYSES....23
6. REFERENCES....24

LIST OF ABBREVIATIONS

AE	Adverse event
AMR	Antibody-mediated rejection
ATG	Anti-thymocyte globulin
BP-ACR	Biopsy-proven acute cellular rejection
CAV	Cardiac allograft vasculopathy
CK	Creatine kinase
CKD	Chronic kidney disease
CMV	Cytomegalovirus
DCC	Data coordinating center
DSMB	Data and Safety Monitoring Board
EBV	Epstein-Barr virus
EQ-5D-Y	European Quality of Life Scale, 5 dimensions, youth version
EVL	Everolimus
FDA	Food and Drug Administration
GFR	Glomerular filtration rate
HDL	High-density lipoprotein
HT	Heart transplantation
ISHLT	International Society of Heart and Lung Transplantation
ITT	Intention-to-Treat
LDL	Low-density lipoprotein
LDTAC	Low-dose tacrolimus
MATE	Major Adverse Transplant Event
mg	Milligram
mL	Milliliter
MMF	Mycophenolate mofetil
MPA	Mycophenolic acid
pAMR	Pathologic diagnosis of cardiac antibody-mediated rejection
PI	Principal investigator
PTLD	Post-transplant lymphoproliferative disease
SAE	Serious adverse event

SAP	Statistical analysis plan
TAC	Tacrolimus
TG	Trigylcerides
UNOS	United Network for Organ Sharing

1. INTRODUCTION AND OBJECTIVES OF ANALYSIS

1.1. Introduction

The study is a Phase 3, multi-center open-label, two-arm trial with randomization of a target 210 subjects in a 1:1 ratio to compare the effect of everolimus/ low-dose tacrolimus (EVL/LDTAC) vs. tacrolimus/ mycophenolate mofetil (TAC/MMF) on the primary safety and efficacy endpoints. The trial design includes two prespecified interim analyses, the results of which will be used by an independent Data and Safety Monitoring Board (DSMB) to recommend early termination for efficacy, or trial continuation. Safety data will also be reviewed at DSMB meetings held at least twice a year. The funding agency has no involvement in the data analysis, interpretation, or early termination decisions.

1.2. Objectives of Statistical Analysis

This statistical analysis plan (SAP) is designed to outline the methods to be used in the analysis of study data for the interim analyses and the final study report. Populations for analysis, data handling rules, statistical methods, and formats for data presentation are provided.

Recommendations from ICH-E9[1] Guideline on Statistical Principles for Clinical Trials will be followed.

This SAP also will outline any differences in the currently planned analytical objectives relative to those stated in the study protocol.

1.2.1. Primary Objectives

Efficacy: To test the hypothesis that among children who survive to 6 months post-HT, those randomized to EVL/LDTAC for 30 months will have a lower total burden of three major adverse transplant events: CAV, CKD and BP-ACR, as defined by the MATE-3 score, compared to children randomized to standard therapy with TAC/MMF.

Safety: To test the hypothesis that among children who survive to 6 months post-HT, those randomized to EVL/LDTAC for 30 months will not have a higher risk of graft loss due to all causes as defined by the MATE-6 score, compared to children randomized to standard therapy with TAC/MMF.

Medication Tolerability: To determine the tolerability of EVL/LDTAC compared to TAC/MMF for prevention of major post-transplant complications by characterizing the frequency and incidence of AEs that are at least possibly related to EVL/LDTAC or TAC/MMF.

1.2.2 Secondary Objectives

- To determine whether EVL/LDTAC is superior to TAC/MMF at delaying time to patient death
- To determine whether EVL/LDTAC is superior to TAC/MMF at delaying time to cardiac allograft loss
- To determine whether EVL/LDTAC is superior to TAC/MMF at preserving kidney function (eGFR)
- To determine whether EVL/LDTAC is superior to TAC/MMF at preventing CAV
- To determine whether EVL/LDTAC is superior to TAC/MMF at preventing BP-ACR

- To compare quality of life between subjects taking EVL/LDTAC and TAC/MMF as assessed by the EuroQOL EQ-5D-Y (Youth version)
- To compare functional status between subjects taking EVL/LDTAC and TAC/MMF as assessed by Lansky (<16 years) or Karnofsky scores (≥16 years)
- To compare the frequency and incidence of medication-related adverse events in subjects taking EVL/LDTAC and TAC/MMF

2. STUDY DESIGN

2.1. Synopsis of Study Design

This is a Phase 3 multi-center open label randomized clinical trial. The primary aim of the study is to test the hypothesis that, among children who survive to 6 months post-HT, children randomized to EVL/LDTAC will have a lower total burden of three major adverse transplant events —CAV, CKD and BP-ACR as defined by MATE-3—compared to children randomized to TAC/MMF, without a higher predicted risk of death/graft loss due to all causes as defined by MATE-6. The rationale for this approach is that while we wish to focus on reducing the three transplant complications that EVL has been shown to affect, we would not want to omit from consideration death/graft loss events caused by other major adverse transplant events, such as infection, as was found in an adult heart transplant trial of EVL and cyclosporine (reference population cyclosporine/MMF).[2] Thus MATE-3 is the primary efficacy endpoint whereas MATE-6, which is predictive of all-cause graft loss, is the primary safety endpoint. Patients will be randomized at 6 months post-HT if they meet the study eligibility criteria and provide informed consent. Subjects will be followed for a total of 30 months post-randomization.

The target sample size is 210 subjects enrolled at up to 25 study sites. Randomization will occur in a 1:1 ratio to EVL/LDTAC or TAC/MMF, using a 2 x 2 stratified randomization scheme according to center volume and sensitization staus (presence vs. absence of donor specific antibody).

As requested by the FDA, after the first 10 patients randomized to EVL/LDTAC reach 180 days follow-up (Study Visit 3), a Data and Safety Monitoring Board (DSMB) safety review will be performed. The study has a group sequential design. Two formal interim analyses for efficacy will be performed. The funding agency will have no involvement in interim data analysis, interpretation, or study continuation decisions.

2.2. Randomization

2.2.1. Methodology

Consenting subjects who meet all eligibility criteria will be randomized in a 1:1 ratio to EVL/LDTAC and TC/MMF. Randomization will be stratified according to 2 factors (4 total strata):

- Sensitization status (presence vs. absence of donor specific antibody) and
- Study site annual cardiac transplant volume (≤12/year vs. >12/year).

The randomization list will be generated by the statistician and research computing unit at the Data Coordinating Center (DCC).

2.2.2. Unblinding

Not Applicable. This is an open-label study. However, the Endpoint Adjudication Committee members are blinded to treatment assignment.

2.3. Stopping Rules

The first interim analysis for the purposes of safety will be performed after the first 10 patients randomized to EVL/LDTAC reach 180 days follow-up (Study Visit 3); a DSMB safety review will be performed. In addition, two formal interim analyses will be performed by the DCC

when follow-up (Visit 8, 30 months) is complete for 50% and 80% of subjects (Section 4.3). For each interim analysis, the database will be cleaned and locked The DCC will then produce the results to be presented to the DSMB. The DSMB will hold regularly scheduled meetings at least twice a year.

The DSMB will be responsible for making a recommendation to Trial leadership to either a) continue with the study or b) stop the study if any of the following have occurred:

1. unacceptable safety
2. overwhelming efficacy
3. unacceptable efficacy (futility)

A formal stopping rule is in place only for premature termination due to overwhelming efficacy. A DSMB recommendation of stopping for safety or futility will be made based on the totality of evidence provided in the DSMB report.

2.4. Study Procedures

The schedule of assessments is outlined in Section 6.1 of the study protocol (Table 6-1).

2.5. Endpoints

2.5.1. General Comments on Analytic Approach

- Endpoints based on specific study visits utilize the information submitted for the specific study visit, even if obtained outside of the protocol-recommended visit window.
- Endpoints related to MATEs will be analyzed according to the Adjudication Committee endpoint determination, unless otherwise specified.
- Endpoints related to the CAV MATE will be analyzed according to the Angiography Core Laboratory determination, unless otherwise specified.

2.5.2. Primary Efficacy Endpoint

The primary efficacy endpoint of the trial is a validated score ranging from 0 to 12, denoted by the MATE-3 score.[3] The score represents the cumulative burden between 0 and 30 months post-randomization (Study Visit 8) of three major adverse transplant events (see Table 1): CAV, CKD, and BP-ACR, as defined by published international guidelines/consensus statements.[4-10]

The MATE-3 score is a composite ordinal score that captures elements of both major adverse transplant event frequency and severity for CAV, CKD and BP-ACR, major adverse transplant events that may be reduced by everolimus based on previous studies. Key calculation features are:

- For each component, the maxium MATE score component (see Table 1) adjudicated for a participant across the study period is assigned, and then the components are summed.
- CKD is based on serum creatinine values for the participant, but excluding from consideration any values noted by the Site PI as being associated with an acute renal event.
- To avoid several mild complications experienced by a subject summing to a value greater than graft loss (i.e., death or re-transplant), graft loss due to CAV, CKD, or BP-ACR is automatically assigned a MATE-3 score of 12.

The protocol specifies an allowable window of 30±2 months for study Visit 8. In addition, the reporting guidelines require serious adverse events up to 30 days beyond the final study visit to be reported. Hence, for purposes of analyses, any graft loss (death or re-transplant) occurring by 33 months post-randomization will qualify as a trial endpoint.

Table 1. The MATE Scores: MATE-3 and MATE-6[3]

	0 (No Disease)	**1 (Mild)**	**2 (Moderate)**	**3 (Severe)**	**4* Death or Graft Loss**	**MATE**	
						3	**6**
Chronic Kidney Disease	**Stage 1-2** ≥60 ml/min/ 1.73 m^2	**Stage 3A** 45-59 ml/min/ 1.73 m^2	**Stage 3B** 30-44 ml/min/1.73 m^2	**Stage 4-5** ≤29 ml/min/1.73 m^2, dialysis or renal transplant	**Death or Graft Loss**	**X**	**X**
Coronary Allograft Vasculopathy	**ISHLT CAV_0**	**ISHLT CAV_1**	**ISHLT CAV_2**	**ISHLT CAV_3**	**Death or Graft Loss**	**X**	**X**
Treated Biopsy Proven-Acute Cellular Rejection	**ISHLT Grade 0**	**ISHLT Grade 1R**	**ISHLT Grade 2R**	**ISHLT Grade 3R**	**Death or Graft Loss**	**X**	**X**
Antibody Mediated Rejection (AMR)	**ISHLT Grade pAMR0**	**ISHLT Grade pAMR1**	**ISHLT Grade pAMR2**	**ISHLT Grade pAMR3**	**Death or Graft Loss**		**X**
Infection	**None**	**Low-risk** PO/IM Antibiotics	**Moderate-Risk** Not life threatening, IV antibiotics	**High-Risk** Life threatening, IV antibiotics	**Death or Graft Loss**		**X**
PTLD	**None**	**Low-risk** Polymorphic PLTD	**Moderate-risk** Monomorphic no "high-risk" features	**High-risk** CNS or Marrow Involvement	**Death or Graft Loss**		**X**

Additional details on categorical classification can be found in the published manuscript.[3]

*Patients experiencing death or graft loss (re-transplant) due to a MATE-3 event will be assigned a score of 4 in each category resulting in a MATE-3 score of 12.

Patients experiencing death or graft loss (re-transplant) due to a MATE-6 event will be assigned a score of 4 in each category resulting in a MATE-6 score of 24.

pAMR: Pathologic diagnosis of cardiac antibody-mediated rejection

2.5.3. Secondary Efficacy Endpoints

The secondary efficacy endpoints are measured from the time of randomization, where applicable.

- Overall patient survival from randomization to 30 months post-randomization (36 months post-HT)
- Overall graft survival (freedom from death or re-transplant) through 30 months post-randomization (36 months post-HT)
- Change in kidney function after start of study drug as assessed by change in estimated glomerular filtration rate (eGFR) using the modified Schwartz equation from month 0 to 6, month 0 to 12 and month 0 to 30 post-randomization (i.e., 12, 18, and 36 months post-HT).
- Freedom from event for the composite MATE-3 events: CKD, CAV and BP-ACR
- Freedom from event for each of the 3 MATE-3 components: CKD, CAV and BP-ACR.
- Freedom from composite efficacy failure, defined as freedom from first occurrence of death, repeat HT, 2R/3R ACR rejection or rejection with HD compromise. HD compromise is derived from the local site determination.
- The composite score consisting of the following elements: MATE CAV, MATE CKD, MATE BP-ACR, and CMV infection (as defined by the MATE infection sub-score and where any CMV infection is assigned a value of at least 1).
- Freedom from CAV, CKD, BP-ACR, or any CMV infection.
- The change in CKD stage from prior to randomization to 30 months post-randomization. The rationale for this endpoint is that change in CKD stage can take on a negative value to reflect the scenario where renal function improves if the CKD stage at trial completion is lower than the CKD stage at enrollment.
- The MATE-3 score where the MATE CKD score component is calculated by the change in the MATE CKD score from prior to randomization to 30 months post-randomization. The MATE CKD score may take on a negative value to reflect the scenario where renal function can improve if the MATE CKD score at trial completion is lower than the MATE CKD score at enrollment.
- The MATE-3 score where the CKD MATE score component is replaced by the change in the CKD stage from prior to randomization to 30 months post-randomization. The change in CKD stage may take on a negative value to reflect the scenario where renal function can improve if the CKD stage at trial completion is lower than the CKD stage at enrollment.
- The composite score consisting of the following four elements: MATE CAV, MATE BPACR, MATE ΔCKD, and CMV infection (as defined by the MATE infection sub-score and where any CMV infection is assigned a value of at least 1); The MATE ΔCKD score is calculated by the change in the CKD MATE score from prior to randomization to 30 months post-randomization. The CKD MATE score may take on a negative value to reflect the scenario where renal function can improve if the CKD stage at trial completion is lower than the CKD stage at enrollment.

- The composite score consisting of the following four elements: MATE CAV, MATE BPACR, ΔCKD Stage, and CMV infection (as defined by the MATE infection sub-score and where any CMV infection is assigned a value of at least 1); The ΔCKD score is calculated by the change in the CKD stage from prior to randomization to 30 months post randomization.
  The change in CKD stage may take on a negative value to reflect the scenario where renal function can improve if the CKD stage at trial completion is lower than the CKD stage at enrollment
- Lansky (< 16 years at randomization) and Karnofsky (≥ 16 years at randomization) Scores at 18 and 30 months post-randomization (6, 24, 36 months post-HT).
- EuroQOL EQ-5D-Y (Youth version)[11, 12] at 18 and 30 months post-randomization (24 and 36 months post-HT) for subjects ages ≥ 8 years at randomization; EQ-5D-Y proxy version 1 will be used for children ≥ 4 years but less than 8 years at randomization or children ≥ 8 years who are unable to complete the EQ-5D-Y at randomization.

2.5.4. Primary Safety Endpoint

The primary safety endpoint is a validated score ranging from 0 to 24, denoted the MATE-6 score. The score represents the predicted risk of graft loss based on the cumulative burden between 0 and 30 months post-randomization (6 to 36 months post-HT) of all six major adverse transplant events as defined published international guidelines/consensus statements[4-10] (Table 1). The MATE-6 score is a composite ordinal score that captures elements of both major adverse transplant event frequency and severity for six major adverse transplant events and has been shown to predict long-term graft loss (CPE-statistic 0.74)[3]. Key calculation features include:

- For each component, the maximum MATE score component (see Table 1) adjudicated for a participant across the study period is assigned, and then the components are summed.
- CKD is based on serum creatinine values for the participant, but excluding from consideration any values noted by the Site PI as being associated with an acute renal event.
- To avoid several mild complications summing to a value greater than graft loss (death or re-transplant), graft loss due to any MATE-6 event is automatically assigned a MATE-6 score of 24.

2.5.5. Secondary Safety Endpoints

The secondary efficacy endpoints are measured from the time of randomization, where applicable.

The purpose of the secondary safety endpoints is to examine individually the MATE-6 score endpoints that are not included in the MATE-3 efficacy endpoint, and to complete the safety profile of the drugs under study. They include:

- Freedom from event for the composite MATE-6 events: CKD, CAV and BP-ACR, pAMR, PTLD, and infection
- Freedom from event for each of the 3 MATE-6 components not captured in MATE-3: pAMR, infection, and PTLD.

- The frequency and incidence of AEs including, but not limited to, hyperlipidemia, anemia, thrombocytopenia, interstitial lung disease, aphthous stomatitis, proteinuria, and rash that are unlikely related, possibly related, or related to EVL/LDTAC compared to TAC/MMF. These AEs will be reported as individual endpoints as well as a composite.
- Freedom from composite of any major adverse transplant event (CKD, CAV, BP-ACR, pAMR, PTLD and infection) as defined in the MATE score.[3]
- Freedom from first occurrence of any level 2 severity CKD major adverse transplant event as defined in the MATE score.[3]
- Freedom from first occurrence of any level 2 severity CAV major adverse transplant event as defined in the MATE score.[3]
- Freedom from first occurrence of any level 2 severity BP-ACR major adverse transplant event as defined in the MATE score.[3]
- Freedom from first occurrence of any level 2 severity pAMR major adverse transplant event as defined in the MATE score.[3]
- Freedom from first occurrence of any level 2 severity PTLD major adverse transplant event as defined in the MATE score.[3]
- Freedom from first occurrence of any level 2 severity infection major adverse transplant event as defined in the MATE score.[3]

3. SUBJECT POPULATIONS

3.1. Population Definitions

The following subject populations will be constructed and used for analysis of the study data:

(1) **Intention-to-Treat (ITT) Population:** All patients who are randomized who were assessed at least once after the administration of at least one dose of study medication.
(2) **Modified ITT Population:** The subset of patients in the ITT Population who were not found post-randomization to have failed any trial eligibility criterion that was measured prior to randomization (i.e., the population excludes ineligible patients who were randomized in error).
(3) **Per-Protocol Population:** The subset of participants in the Modified ITT Population who: a) do not have major deviations from the protocol; and b) have at least two post-baseline assessments (i.e., have at least 6 months of data post-randomization). Major protocol deviations were pre-specified prior to the start of the trial. A less restrictive Per-Protocol Population will also be examined: this cohort will exclude surviving non-retransplanted participants who specifically have a major deviation due to missing one or more of the following from the final study visit (Visit 8): coronary angiography, endomyocardial biopsy, and serum creatinine value.
(4) **As Treated Population:** All randomized patients who received at least one dose of study medication and undergo at least one safety assessment. For the purposes of analysis, Treated with EVL/LDTAC is defined as: recorded as having taken EVL/LDTAC for at least 80% of time on study. Treated with TAC/MMF is defined as: recorded as having taken EVL/LDTAC for at least 80% of time on study. This population will exclude subjects who do not meet the compliance criterion for either of the two trial treatments. The definition of Treated will not utilize participant medication diaries due to anticipated high rates of incomplete data.

Primary efficacy analyses (for all endpoints) will be conducted on data from the ITT population, with subjects remaining in their assigned treatment group regardless of actual treatment received.

Secondary efficacy analyses (for the primary efficacy endpoint and its components) will be conducted on data from the modified ITT population and the Per-Protocol population, with subjects remaining in their assigned treatment group regardless of actual treatment received.

Tertiary efficacy analyses (for the primary efficacy endpoint and its components) will be conducted on the As Treated Population, with subjects classified into the group corresponding to treatment actually received (i.e., non-ITT).

Primary safety and tolerability analyses (for all endpoints) will be conducted on data from the ITT population.

Secondary safety and tolerability analyses (for all endpoints) will be conducted on data from the As Treated population, with subjects classified into the group corresponding to treatment actually received.

A review of the treatment data will be completed to determine if there were any subjects who did not receive clinical trial material, that is, who received no study drug. All subjects not receiving clinical trial material will be fully documented in the clinical study report. If the

reason for not receiving study drug can be confirmed to be independent of treatment assignment, then their data will be excluded from all analyses.

4. STATISTICAL METHODS

4.1. Sample Size Justification

A target of 210 subjects will be enrolled in this study (105 per group). The sample size justification is based on the analytic approach proposed prior to trial initiation.

Aim 1 (Efficacy): To determine if the cumulative burden of domain-specific MATE scores for CAV, CKD, BP-ACR (MATE-3) is superior for EVL/LDTAC compared to TAC/MMF our null hypothesis is that the mean MATE-3 scores at 30 months are equal for the EVL/LDTAC and TAC/MMF groups. The alternative hypothesis is that the mean MATE-3 scores at 30 months differ for the EVL/LDTAC and TAC/MMF groups. With the assumptions shown in Table 2,

Table 2. Sample Size Assumptions

- Two-sided superiority test, $\alpha = 0.05$
- Comparison of Month 30 means, adjusted for baseline score
- 15% crossover from EVL/LDTAC to TAC/MMF
- 5% crossover from TAC/MMF to EVL/LDTAC
- SD of MATE-3 = 2.4
- Within-subject correlation R=0.5 (per PHTS data)
- 5% loss to follow-up
- 2% inflation for interim look for efficacy

a total of 210 subjects is required to detect with 90% power a mean difference of 1.2 in the 30-month change score between the two independent treatment groups. A difference of 1.2 points is one-half SD, which is typically considered to be a minimum clinically important difference.[13,14] The calculation assumes, however, a reduced mean difference of 0.96 to account for the treatment crossover shown in Table 2, that will incur dilution of treatment effect in the primary intention-to-treat analysis. The treatment group comparison will be adjusted for the baseline MATE-3 score, thus the required sample size is $N^*=N(1-R^2)$, where N=130 and R=0.5. Hence N*=97. With the inflation for loss to follow-up and an interim look, the required sample size is (97/0.95)*1.02 = 105 per group (total 210 subjects).

If the treatment crossover from TAC/MMF to EVL/LDTAC is 15% rather than 5%, the planned sample size will still yield 80% power to detect the desired mean difference in a baseline-adjusted comparison of Month 30 MATE-3 scores.

Aim 2 (Safety): With 194 analyzable subjects (97x2; see above), a 95% two-sided CI for the 30-month treatment group difference in MATE-6 score assessing safety will have width ± 1.24, thus is able to rule out the pre-specified acceptable difference of less than 3 points with high confidence.

4.2. General Statistical Methods and Data Handling

4.2.1. General Method

Tabulations will be produced for appropriate demographic, baseline, efficacy and safety parameters. For categorical variables, summary tabulations of the number and percentage within each category (with a category for missing data) of the parameter will be presented. For

continuous variables, the N, mean, standard deviation, median, interquartile range, and some instances minimum and maximum values will be presented.

Formal statistical hypothesis testing will be performed on the primary endpoint with all tests being two-sided. For the primary endpoint, the significance level will be according to the nominal level resulting from the performance of two interim looks (i.e., less than 0.05 level of significance). The primary safety endpoint will be reported as a two-sided confidence interval as described in the FDA IND and below in Section 4.8.1.1.

4.2.2. Computing Environment

All statistical analyses will be performed using SAS statistical software (Version 9.4 or higher) and R (version 4.03 or higher).

4.2.3. Standardized Coding

Adverse events will be coded using MedDRA version 12.0 and graded using Version 4.03 of the National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE v4.03).

4.2.4. Adjustments for Covariates

Adjustment for baseline MATE score will be performed in selected analyses.

4.2.5. Multiple Comparisons/Multiplicity

No adjustments for multiple comparisons will be applied.

4.2.6. Subpopulations

The following pre-specified subgroups which are defined according to status at the time of randomization will be utilized for analysis of the primary efficacy and safety endpoints and medication tolerability.

- Center transplant volume (Lower vs. Higher, ≤12 vs. >12 annually)
- Age <1 year vs. ≥1 year
- Sensitization status (presence vs. absence of donor-specific antibody).
- Race (2 comparisons): White, Non-Hispanic vs. All Others; White vs. Black vs. Asian vs. Other
- History of treated rejection (present vs. absent)
- CKD stage (1 vs. 2 vs. 3)
- Males vs. Females
- Ethnicity: Hispanic vs. Non-Hispanic
- Type of induction therapy (3 comparisons):
  - None vs. IL-2 receptor blocker (basiliximab, daclizumab) vs. thymoglobulin (ATG);
  - None vs. IL-2 receptor blocker vs. low-dose ATG (<5 mg/kg total dose) vs. high-dose ATG (≥5 mg/kg total dose)
  - Low-dose ATG (<5 mg/kg total dose) vs. high-dose ATG (≥5 mg/kg total dose) – excludes patients with no induction therapy or IL-2 receptor blocker therapy
- Type of maintenance immunosuppression prior to randomization: TAC/MMF vs. cyclosporine/MMF

In addition, the treatment effect on the primary endpoints (MATE-3, MATE-6 and freedom from MATE-3, MATE-6) and medication tolerability will be examined according to the presence vs. absence (determined post-randomization) of additional immunosuppression (e.g., prednisone) during the study period.

4.2.7. Withdrawals, Missed Final Visit, Lost to Follow-up

The data of subjects who withdraw consent, do not complete a final study visit (30 months post-randomization) for any reason or are lost to follow-up will be included in analyses according to the definitions of the Population Definitions in Section 3.1. For the purpose of primary endpoint MATE score calculation, data from the subject's last response and latest procedures (e.g., biopsy, coronary angiography, creatinine) will be used to determine final outcomes in place of 30-month data if none is available (see Section 4.2.8). Such subjects will not be replaced. See Section 3.1 regarding definition of a study population that excludes subjects with unavailable 30-month data.

4.2.8. Missing Data

For the purpose of the primary efficacy and primary safety MATE score calculations, the maximum value observed for each MATE score component will be utilized in the score, even if data were not obtained for the final study visit or an intermediate visit.

For the purpose of creatinine clearance calculation in the absence of a height measurement, height will be imputed as the time-weighted mean height derived from the two closest available height measurements prior to and after the visit with a missing height. If a height that has no later value is missing, then the height will be imputed as the prior height if the following condition is met: female, age>15.0 years; male, age>16.0 years.

There will be no other substitutions made or imputation performed to accommodate missing data points.

4.2.9. Change in Value from X to Y

Change in Value from X date to Y date = Value on Y date minus Value on X date.

This definition will exclude the data of subjects who do not have data at both time points.

The one exception to this definition is the change score for CKD Stage and for CKD MATE score component to be utilized in composite MATE score calculations. For these change scores, the subject's latest available creatinine value will be used to calculate change from baseline.

4.3. Interim Analyses

For efficacy, two interim looks by treatment arm were planned, at 50% and 80% statistical information. Statistical information is defined as the proportion of follow-up visits completed out of the total expected number of follow-up visits. An O'Brien-Fleming boundary[15] for premature trial termination will be applied (Table 3). The primary efficacy endpoint analysis will be performed per Section 4.7.1. There are no formal stopping rules for futility or safety.

Table 3. O'Brien-Fleming Early Stopping Rule for Treatment Efficacy (MATE-3)

Analysis	First Look	Second Look	Final
Fraction of total information	0.5	0.8	1.0
Follow-up time* (mo) accrued	3165 (50%)	5064 (80%)	6330 (100%)
Nominal p-value to reject null	0.0030	**0.0234**	0.0426
Nominal |z| to reject the null	2.963	2.266	2.028

*30 months x 211 subjects = 6330 months

4.4. Subject Disposition

Subject disposition will be tabulated by treatment group, including the number screened, the number in each subject population for analysis, reasons for subjects not meeting eligibility or providing informed consent, the number that received at least one dose of clinical trial material, the number who withdrew from study treatment but remained in the study, the number who withdrew or were lost from the study prior to completing the study, and reasons for withdrawal.

A by-subject listing of study completion information, including the reason for premature study withdrawal, if applicable, will be presented.

4.5. Demographic and Baseline Characteristics

Age at randomization, gender, race and ethnicity and medical history/cardiac transplant characteristics and baseline laboratory findings will be summarized by treatment group for each subject population, using descriptive statistics. Formal statistical comparisons will be performed, but will not be presented for publication unless requested by the journal.

4.6. Treatment Exposure

The following treatment exposure parameters will be summarized for the two treatment groups:

- Frequency and percentage of subjects experiencing any interruptions in assigned treatment or permanent withdrawal from assigned study treatment.
- Number of days reported to have received the study drug strategy (EVL/LDTAC or TAC/MMF)
- Number of days that trough levels for both EVL and TAC were within the protocol-specified target ranges for each treatment arm as estimated using the Rosendaal method.
- Number of days where MMF dose is greater than or equal to 500 mg/m^2/dose given BID OR mycophenolic acid level between 2-4 mcg/mL. If subject is taking mycophenolate sodium or Myfortic then exposure will is defined as the number of days where dose is greater than 360 mg/m^2/dose given BID.

Additional measures of treatment exposure will be reported based on FDA standards.

4.7. Efficacy Evaluation

The ITT population will be used for the primary analysis of all efficacy endpoints. The Modified ITT and PP populations will be used for the secondary analyses of all (primary and secondary) efficacy endpoints. The As Treated population will be used for tertiary analysis of the primary efficacy endpoint.

4.7.1. Primary Efficacy Endpoint

4.7.1.1 Primary Analytic Approach for the Primary Efficacy Endpoint

The null and alternative hypotheses for the primary efficacy endpoint are:

H_0: The distributions of MATE-3 scores for the EVL/LDTAC and TAC/MMF groups are equal.

H_A: The distributions of MATE-3 scores for the EVL/LDTAC and TAC/MMF groups are not equal.

The primary analytic approach for the primary efficacy endpoint will be a zero-truncated Wilcoxon rank sum test[16] performed at the two-sided α=0.0426 level (see Section 4.3 detailing the nominal p-value for the final analysis). If the two-sided p-value is less than 0.0426, then the null hypothesis of no difference in the distributions of MATE-3 scores will be rejected. The cumulative incidence of MATE-3 will be plotted over time to demonstrate the accrual of adverse events over time in the two treatment groups.

4.7.1.2 Secondary Analytic Approaches for the Primary Efficacy Endpoint

Two secondary analytic approaches for the primary efficacy endpoint will be implemented, although these may not be robust due to the departure from normality of MATE scores:

(1) a baseline adjusted comparison of means and

(2) mixed effects longitudinal regression appropriate for randomized trials[17] (i.e., parameterization that induces adjustment for baseline) will be used to model MATE-3 as a function of: a) continuous time post-randomization and b) interaction terms between the indicator for treatment assignment and time, with the variance-covariance matrix based on the sequence of the scheduled visits. The test of superiority for efficacy using this secondary analytic approach will be the test of whether the two slopes (i.e., MATE score trajectories) differ, as it represents the treatment effect on cumulative MATE-3 burden over time. If the time effects are not found to be linear hence imposing bias if a linear model is used, then the time effect will be fit using indicators for the study visits (7 degrees of freedom) and the categorical time x treatment assignment interaction will be tested, with the relative superiority of one treatment group over the other assessed using a time-weighted contrast of the mean MATE-3 score from the two treatment groups.

4.7.2 Secondary Efficacy Endpoints

4.7.2.1 Primary Analytic Approach for the Secondary Efficacy Endpoints

The treatment effect on the composite MATE score that includes CMV infection (MATE-3 + CMV) will be analyzed with a zero-truncated Wilcoxon test. The cumulative incidence of the MATE-3 + CMV infection score will also be plotted over time for the two treatment groups.

The cumulative incidence of MATE-3 + any CMV infection will also be plotted and compared as will the cumulative incidence of MATE-3 with any CMV infection where the MATE-3 CKD score is calculated using conventional CKD stages from the National Kidney Foundation.

The secondary MATE scores that incorporate change in CKD stage or change in CKD MATE component will be analyzed with a Wilcoxon rank sum test to assess treatment effect, if these distributions include values below zero.

The analytic approach for the time-to-event endpoints such as overall survival, time to first occurrence of a MATE-3 event, and time to occurrence of each component of the MATE-3 score will be a logrank test to evaluate treatment effect. Kaplan-Meier methodology will be used to estimate freedom from event with 95% CIs for the estimated freedom from event at 12, 24, and 30 months post-randomization. For endpoints other than survival and retransplant-free survival, subjects who die or undergo retransplantation from a cause unrelated to a MATE-3 event will remain in the analysis and have their follow-up censored at death or retransplant date. As a sensitivity analysis, the treatment effect for time to the composite MATE-3 endpoint will also be analyzed in a competing risks framework (e.g., subdistribution hazard model), to minimize bias from the competing risks of death and retransplant.

The analytic approach for comparison of change in eGFR in the two treatment groups will be a two-sided Student's t-test, for each post-baseline visit minus baseline. In addition, as described above for the primary endpoint, mixed effects longitudinal regression will be employed to estimate the trajectory of eGFR over time for the two treatment groups and in the absence of interaction, estimate an overall treatment effect.

The analytic approach for comparison of Lansky and Karnofsky scores in the two treatment groups will be a Wilcoxon rank sum test. The Lansky and Karnofsky scores will be analyzed separately as well as combined (i.e., all ages).

The analytic approach for comparison of the overall EQ-5D score in the two treatment groups at each study visit where measured will be a two-sided Student's t-test. If the the distribution of the score is highly skewed, then a Wilcoxon rank sum test will take precedence for significance testing. The EQ-5D scores (patient-reported and proxy-reported) will be analyzed separately as well as combined (i.e., all ages).

For both Lansky/Karnofsky and EQ-5D scores, although the distributions may depart from normality, trajectory analysis will be performed to examine treatment differences by time and, in the absence of time by treatment interaction, to estimate an overall treatment effect.

4.8. Safety Evaluation

Primary safety and tolerability analyses (for all endpoints) will be conducted on data from the ITT population (see Section 3.1). Secondary safety and tolerability analyses (for all endpoints) will be conducted on data from the As Treated population, with subjects classified into the group corresponding to treatment actually received, defined in Section 3.1.

4.8.1 Primary Safety Endpoint

4.8.1.1 Primary Analytic Approach for the Primary Safety Endpoint

The purpose of the primary safety endpoint is to test the hypothesis that among children who survive to 6 months post-HT, those randomized EVL/LDTAC for 30 months will not have a higher risk of graft loss due to all causes as defined by the MATE-6 score, compared to children randomized to standard therapy with TAC/MMF.

A 95% two-sided CI for the 2.5-year treatment group difference (EVL/LDTAC minus TAC/MMF) in MATE-6 score will be constructed, based on a baseline-adjusted comparison of the mean EVL/LDTAC vs. TAC/MMF MATE-6 scores at 30 months post-randomization. The estimation method for the baseline-adjusted mean group difference is analysis of covariance.

To avoid violations of the normality assumption for the endpoint and ensure valid inference, the 95% CI will be constructed using bootstrapping. If the upper limit of the CI is within the pre-specified clinically acceptable difference of less than 3 points (i.e., no large difference in MATE-6 score between treatment groups, which ranges from 0 to 24), then it will be concluded that EVL/LDTAC has no higher burden of death/graft loss than does TAC/MMF.

The cumulative incidence of MATE-6 will be plotted over time to demonstrate the accrual of adverse events in the two treatment groups.

4.8.2 Secondary Safety Endpoints

4.8.2.1 Time to Event Endpoints

The analytic approach for the time-to-event endpoints such as time to first occurrence of a MATE-6 event, and time to occurrence of each component of Infection, PTLD, and AMR separately, will be a logrank test to evaluate treatment effect. Kaplan-Meier methodology will be used to estimate freedom from event with 95% CIs for the estimated freedom from event at 12, 24, and 30 months post-randomization. Subjects who die or undergo retransplant from a cause unrelated to a MATE-6 event will remain in the analysis and have their follow-up censored at death or retransplant date. As a sensitivity analysis, the treatment effect for time to the composite MATE-6 endpoint will also be analyzed in a competing risks framework (e.g., subdistribution hazard model), to minimize bias from the competing risks of death and retransplant.

4.8.2.2 Medication Tolerability

Fisher exact tests will be used to compare the percentage of subjects in the two treatment groups who experience at least one occurrence of each prespecified adverse event (AE) of interest for this trial, namely: hyperlipidemia, anemia, thrombocytopenia, interstitial lung disease, aphthous stomatitis, proteinuria, and rash. This analysis will be performed for each event and for a composite of all the events (percentage of subjects experiencing at least one occurrence of any of the prespecified events).

4.8.2.3 Adverse Events

AEs are collected from the time of first clinical trial material administration through 30-month follow-up. Deaths are reported up to 30 days beyond the 30-month visit window (30±2 months). All analyses of AEs will utilize classifications with respect to expectedness and treatment relatedness that have been adjudicated by the Medical Monitor.

Fisher exact tests will be used to compare the percentage of subjects who experience at least one occurrence of a Serious Adverse Event (SAE) and any Adverse Event (AE). This analysis will be repeated to analyze events possibly or definitely related to study treatment. A secondary analysis of the SAEs and AEs will estimate the odds or an SAE/AE as a function of the average tacrolimus level for the participant.

Incidence rates for any SAE and for any AE will be calculated to account for multiple SAEs and AEs per subject, with Poisson regression utilized to test for differences between treatment groups. This analysis will be repeated to analyze events possibly or definitely related to study treatment.

Fisher exact tests will be used to compare, for each body system class, the percentage of subjects who experience at least one occurrence of a SAE or AE. vThis analysis will be repeated to analyze events possibly or definitely related to study treatment.

AEs associated with withdrawal from treatment or study will be noted.

All AEs occurring on study will be listed in subject data listings.

4.8.2.4 Safety Laboratory Data

Safety laboratory data are collected according to the Schedule in protocol Table 6-1. Summary statistics will be provided by treatment received for each visit. Where applicable, SI units will be used in reporting. A two-sided Student's t-test or Wilcoxon rank sum test, as appropriate for the laboratory data distribution, will be performed to identify treatment differences at each study visit. Mixed effects longitudinal regression will also be performed to estimate an overall treatment effect or to identify significant interaction of time and treatment.

Actual Values: Descriptive statistics (frequency, mean, standard deviation, median, interquartile range, and range) for hematology, blood chemistry, and urinalysis profile parameters will be provided. For any lab values that are comprised of a character and numeric value (e.g., <1.5), the average of 0 and the upper limit will be used for the descriptive statistics summaries. Graphical representations of mean or median laboratory results over time for selected parameters will also be provided.

Change Scores: Each laboratory measure will be reported as a change from baseline using descriptive statistics (frequency, mean, standard deviation, range). The last set of screening labs obtained prior to randomization, if more than one is recorded, will be assigned as the Baseline value. No statistical testing will be performed on the change scores.

Clinical Abnormality: For total, HDL, and LDL cholesterol and trigylcerides, the laboratory measure will be reported as the frequency and percentage of subjects with a low clinical abnormality (LDL cholesterol below the normal range) or high clinical abnormality (total cholesterol, HDL and triglycerides above the normal range). This summary by visit will be presented in conjunction with the frequency and percentage of subjects receiving statin therapy (concomitant medications).

4.9 Subgroup Analyses

Analyses will be performed to determine if there is a differential treatment effect in prespecified patient subgroups with respect to the primary efficacy endpoint (MATE-3 score) and the secondary efficacy endpoint, time to occurrence of a MATE-3 event. A test of treatment group by subgroup interaction will be performed. The test of interaction for MATE-3 will be derived from a two-part logistic regression and cumulative logistic regression procedure that is equivalent to the zero-truncated Wilcoxon rank sum test. The test of interaction for time to MATE-3 will be generated from a Cox proportional hazards model. Because the trial is not powered to detect small to medium treatment differences within subgroups, any test with interaction $p<0.10$ will be considered to be hypothesis-generating, i.e., potential evidence of interaction.

5. CHANGES TO PLANNED ANALYSES

The primary analytic approach for the primary efficacy endpoint was changed from that described in the FDA IND and the original protocol based on recommendation from the trial DSMB.

6. REFERENCES

1. ICH-E9 Statistical Principles for Clinical Trials, September 1998. Issued by FDA Center for Drug Evluation and Research and Center for Biologics Evaluation and Research.
2. Eisen HJ, Kobashigawa J, Starling RC, Pauly DF, Kfoury A, Ross H, Wang SS, Cantin B, Van Bakel A, Ewald G, Hirt S, Lehmkuhl H, Keogh A, Rinaldi M, Potena L, Zuckermann A, Dong G, Cornu-Artis C and Lopez P. Everolimus versus mycophenolate mofetil in heart transplantation: a randomized, multicenter trial. *Am J Transplant*. 2013;13:1203-16.
3. Almond CS, Hoen H, Rossano JW, Castleberry C, Auerbach SR, Yang L, Lal AK, Everitt MD, Fenton M, Hollander SA, Pahl E, Pruitt E, Rosenthal DN, McElhinney DB, Daly KP, Desai M and Pediatric Heart Transplant Study Group R. Development and validation of a major adverse transplant event (MATE) score to predict late graft loss in pediatric heart transplantation. *J Heart Lung Transplant*. 2017.
4. Stewart S, Winters GL, Fishbein MC, Tazelaar HD, Kobashigawa J, Abrams J, Andersen CB, Angelini A, Berry GJ, Burke MM, Demetris AJ, Hammond E, Itescu S, Marboe CC, McManus B, Reed EF, Reinsmoen NL, Rodriguez ER, Rose AG, Rose M, Suciu-Focia N, Zeevi A and Billingham ME. Revision of the 1990 working formulation for the standardization of nomenclature in the diagnosis of heart rejection. *J Heart Lung Transplant*. 2005;24:1710-20.
5. Mehra MR, Crespo-Leiro MG, Dipchand A, Ensminger SM, Hiemann NE, Kobashigawa JA, Madsen J, Parameshwar J, Starling RC and Uber PA. International Society for Heart and Lung Transplantation working formulation of a standardized nomenclature for cardiac allograft vasculopathy-2010. *J Heart Lung Transplant*. 2010;29:717-27.
6. Hogg RJ, Furth S, Lemley KV, Portman R, Schwartz GJ, Coresh J, Balk E, Lau J, Levin A, Kausz AT, Eknoyan G, Levey AS and National Kidney Foundation's Kidney Disease Outcomes Quality I. National Kidney Foundation's Kidney Disease Outcomes Quality Initiative clinical practice guidelines for chronic kidney disease in children and adolescents: evaluation, classification, and stratification. *Pediatrics*. 2003;111:1416-21.
7. Levey AS, Eckardt KU, Tsukamoto Y, Levin A, Coresh J, Rossert J, De Zeeuw D, Hostetter TH, Lameire N and Eknoyan G. Definition and classification of chronic kidney disease: a position statement from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int*. 2005;67:2089-100.
8. Berry GJ, Burke MM, Andersen C, Bruneval P, Fedrigo M, Fishbein MC, Goddard M, Hammond EH, Leone O, Marboe C, Miller D, Neil D, Rassl D, Revelo MP, Rice A, Rene Rodriguez E, Stewart S, Tan CD, Winters GL, West L, Mehra MR and Angelini A. The 2013 International Society for Heart and Lung Transplantation Working Formulation for the standardization of nomenclature in the pathologic diagnosis of antibody-mediated rejection in heart transplantation. *J Heart Lung Transplant*. 2013;32:1147-62.
9. Husain S, Mooney ML, Danziger-Isakov L, Mattner F, Singh N, Avery R, Ison M, Humar A, Padera RF, Lawler LP, Fisher A, Drew RJ, Gould KF, Sole A, Studer S, Munoz P,

Singer LG, Hannan M and Definitions IIDCWGo. A 2010 working formulation for the standardization of definitions of infections in cardiothoracic transplant recipients. *J Heart Lung Transplant*. 2011;30:361-74.

10. Swerdlow SH WS, Chadburn A, et al . WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. Lyon: IARC Press; 2008. Post-transplant lymphoproliferative disorders.
11. Ravens-Sieberer U, Wille N, Badia X, Bonsel G, Burstrom K, Cavrini G, Devlin N, Egmar AC, Gusi N, Herdman M, Jelsma J, Kind P, Olivares PR, Scalone L and Greiner W. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. *Qual Life Res*. 2010;19:887-97.
12. Wille N, Badia X, Bonsel G, Burstrom K, Cavrini G, Devlin N, Egmar AC, Greiner W, Gusi N, Herdman M, Jelsma J, Kind P, Scalone L and Ravens-Sieberer U. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res*. 2010;19:875-86.
13. Cohen J. *Statistical power analysis for the behavioral sciences.* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.; 1988.
14. Norman GR, Sloan JA and Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. 2003;41:582-92.
15. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979 Sep;35(3):549-56.
16. Wanjie W, Chen EZ, Li H. Truncated Rank-Based Tests for Two-Part Models with Excessive Zeros and Applications to Microbiome Data. eprint arXiv:2110.05368. October 2021. DOI 10.48550/arXiv.2110.05368.
17. Fitzmaurice GM, Laird NM and Ware JH. *Applied Longitudinal Analysis, 2nd Edition*; 2011.