Statistical Analysis Plan For the Final Statistical Report

PROTOCOL NUMBER: **OPTIMIZE-IP-12**

PROTOCOL TITLE: **The OPTIMIZE Multicenter, Placebo-Controlled, Double-Blind, Randomized Trial**

PRINCIPAL INVESTIGATORS:

Nicole Mayer-Hamblett, PhD
OPTIMIZE DCC

Bonnie Ramsey, MD
OPTIMIZE CCC

FUNDING AGENCIES: **National Institutes of Health/National Heart, Lung, and Blood Institute (NHLBI)**

PREPARED BY:

Margaret Kloster, MS
Shelly Ma, MS
Nicole Hamblett, PhD

RELEASE DATE: **April 22, 2015**

PREFACE

The Statistical Analysis Plan (SAP) as outlined in this document will be finalized prior to the completion of the first comprehensive DSMB interim report. Any modifications to the SAP after finalization will be documented. The SAP contains all modifications and updates to the planned analyses that were outlined in the original study protocol. This plan details all *a priori* specified analyses that will be performed upon study completion and database lock, with detailed specifications for all tables, figures, and statistical models.

Signature Page

Analysis plan approved by:

Bonnie Ramsey, M.D. Principal Investigator OPTIMIZE CCC	Date
Nicole Mayer Hamblett, PhD Principal Investigator OTPIMIZE DCC	Date

1. Overview

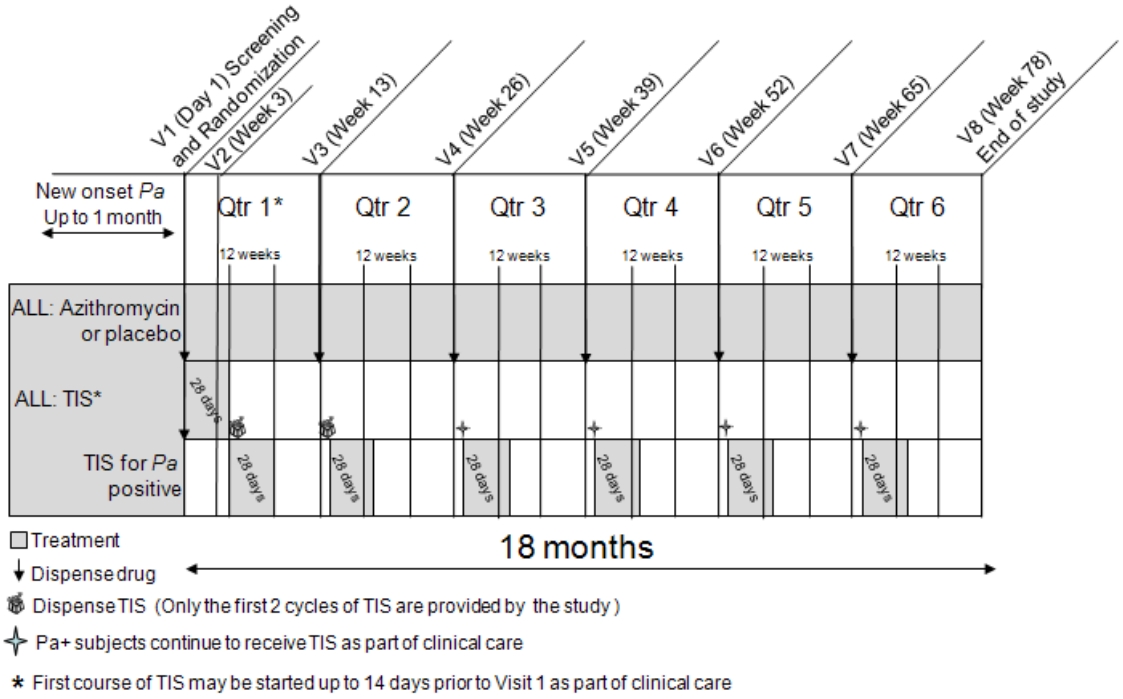
1.1 Study Rationale and Design

This is a multicenter, double-blind, randomized, placebo-controlled clinical trial in 274 children with CF ages 6 mos -18 years with early *Pa*, defined as either a first lifetime documented *Pa* culture or a *Pa* positive culture after at least two years of negative cultures. The study will assess the clinical and microbiologic efficacy and safety of azithromycin given three times weekly in combination with standardized TIS therapy among children with early *Pa*. TIS therapy is defined as an initial eradication treatment with 1-2 courses of 28 days TIS and subsequent 28 day treatments only at times a quarterly oropharyngeal culture is positive for *Pa*. Eligible participants will be randomized within one month of their *Pa* positive culture. Participants initiating TIS more than 14 days prior to the baseline visit for treatment of their *Pa* positive culture will be excluded. Participants will be randomized in a 1:1 fashion to receive one of the following two treatment strategies for 18 months: (1) oral placebo in addition to standardized TIS therapy, or (2) oral azithromycin in addition to standardized TIS therapy.

Three times weekly administration of oral azithromycin or matching placebo will be continued throughout the 18 month trial. All participants will receive a 28-day course of TIS therapy during the first quarter of the trial and have a follow-up culture obtained to document *Pa* clearance. Participants who remain *Pa* positive after the first 28-day course of TIS will receive a second 28-day course of TIS. Subsequently, participants will receive TIS therapy (administered as a single 28-day course) only if their quarterly cultures are *Pa* positive. Cultures will be obtained at the start of each quarter for the duration of the 18 month study (see Figure 1).

Clinical and microbiologic outcomes will be assessed throughout the 18-month trial, including the occurrence of pulmonary exacerbations and *Pa* recurrence.

OPTIMIZE Trial Study Schematic



The total duration of the study is expected to be forty-nine (49) months: thirty-one (31) months for participant recruitment and eighteen (18) for final participant follow-up.

1.2 Study Objectives and Endpoints

The primary endpoint in the study is the difference between treatment groups in the time to a protocol-defined pulmonary exacerbation, comparing oral placebo with inhaled tobramycin versus oral azithromycin with inhaled tobramycin. Pulmonary exacerbations will be defined according to an *a priori* sign and symptom-based definition (Appendix II).

Secondary Endpoints

1. Time to *Pa* recurrence over the 18-month study period
2. Safety as measured by the incidence of adverse events and laboratory abnormalities over the 18-month study period
3. Changes from baseline in plasma inflammatory markers

Secondary Exploratory Endpoints

4. Proportion of participants with initial *Pa* eradication success, defined as a *Pa* negative respiratory culture at week 13
5. Proportion of all cultures obtained that are positive for *Pa* during the 18 month study period
6. Prevalence of treatment-emergent *Staphylococcus aureus* (methicillin susceptible and resistant), *Haemophilus influenzae*, *Burkholderia cepacia* complex, *Stenotrophomonas maltophilia*, and *Alcaligenes xylosoxidans* (from oropharyngeal

and sputum cultures), and non-tuberculous mycobacteria (in sputum only) during the 18-month period
7. Proportion of participants hospitalized over the 18-month study
8. Proportion of participants prescribed acute oral, inhaled, and IV acute antibiotics over the 18-month study
9. Change from baseline to the end of the 18-month study in pulmonary function (FVC, FEV_1, $FEF_{25-75\%}$) in those old enough to perform spirometry (≥4 years of age)
10. Change from baseline to the end of the 18-month study in weight and height
11. Change from baseline to the end of the 18-month study in the self-report and parent-completed Cystic Fibrosis Respiratory Diary scores
12. Changes from baseline in the respiratory microbiome
13. Association of baseline inflammatory marker values, microbial diversity measures and other baseline clinical and demographic characteristics with clinical and microbiologic outcomes over the 18-month study

1.3 Interim Data Monitoring Committee Reviews

The DSMB will review quarterly abbreviated safety reports beginning 3 months after the first participant is randomized. Summaries of enrollment, drug discontinuations, and AEs and SAEs tabulated by treatment group will be included in the report.

In addition to the quarterly safety reports, comprehensive interim reports will be generated and reviewed on a semi-annual basis beginning one year after the first participant enrolled. These report will include summaries of enrollment, randomization, demographic and baseline characteristics, safety (SAEs and AEs (including abnormal changes in audiology and electrocardiograms), hospitalizations, laboratory abnormalities, drug discontinuations), and key efficacy data (pulmonary exacerbations, microbiology, anthropometrics, spirometry, acute antibiotic usage).

Formal stopping rules for efficacy, futility and safety and the timing of their evaluation are outlined in section 1.3.

1.4 Statistical Monitoring Guidelines and Stopping Rules for Planned Efficacy and Futility Analyses

The DSMB will be guided by a formal stopping rule based on the relative risk of the primary endpoint, time to pulmonary exacerbation. With the sample size of approximately 137 per group and assuming a two-sided type 1 error of 0.05 and that 50% of participants will experience an exacerbation in the culture-based TIS with placebo group, the study has approximately 90% power to detect a 47% or greater reduction in exacerbations in the group receiving culture-based TIS with azithromycin (hazard ratio of 0.53 or lower) and approximately 80% power to detect a 42% or greater reduction (hazard ratio of 0.58 or lower). These estimates take into account six planned interim analyses, which will be included in the comprehensive interim report, and one final analysis, included in the final report for the study. The first formal interim analysis will

be described in the first comprehensive interim analysis generated one year after the first participant is randomized.

Among the 137 participants enrolled in the culture-based TIS with placebo group, at a rate of 50%, we would expect approximately 69 primary endpoint events during the 18 month randomized trial. Assuming that the addition of azithromycin conservatively results in a 40% reduction in exacerbations (relative risk ≤ 0.6), we would expect approximately 41 events among the 137 subjects enrolled in the azithromycin group for a total of 110 events during the study. The timing of the 7 analyses does not need to be equally spaced with respect to the accrual of PE events. However, if equally spaced analyses are performed, these would occur after 16, 31, 47, 63, 79, 94, and 110 events. Assuming an O'Brien-Fleming boundary1, the two-sided stopping rule would suggest that the trial be stopped early for safety if the observed hazard ratio were greater than or equal to 15.71, 3.96, 2.51, 1.88, 1.74, 1.58, and 1.48 at each analysis time (i.e. significantly greater number of events in the azithromycin group as compared to the placebo group) and the stopping rule for efficacy would suggest that the trial be stopped early if the observed hazard ratio were equal to or less than 0.06, 0.25, 0.39, 0.50, 0.58, 0.63, and 0.68 at each analysis time (i.e. significantly less exacerbations in the azithromycin group as compared to placebo). With this example of equally spaced analyses, on the error spending function scale this corresponds to the following cumulative error spent: <0.00001, 0.023, 0.0333, 0.1384, 0.3396, 0.6323, and 1.0. The actual stopping boundaries will be computed at each analysis and in conjunction with comprehensive semi-annual report to the DSMB. Thus, the actual stopping boundaries may differ than those provided above depending on the number of events that have occurred. Splus SeqTrial software will be used to calculate the formal stopping boundary and adjust for this interim analysis in the final analysis.

In addition to monitoring for the primary endpoint, the DSMB will also be guided by a safety threshold for which to monitor safety of the SAE defined as sensorineuronal hearing loss. Based on data from the EPIC clinical trial among children ages 1 – 12 years of age receiving TIS for newly acquired *Pa*, we expect approximately 8.6% of children to have an abnormal hearing test post-baseline (with a comparable definition to sensorineuronal hearing loss above) (17/197 children in EPIC, 95% CI 5.5%,13.4%). Based on this data, we recommend a stopping rule for the SAE sensorineuronal hearing loss as the following:

> (1) The proportion of patients with sensorineuronal hearing loss post-baseline is at least 8% greater in the azithromycin group as compared to placebo. This assumes, for example, at the end of the study that 12/137 participants (8.8%) of participants in the placebo group and 23/137 (16.8%) in the azithromycin group experience this SAE (an 8.0% difference, 95% CI: 0.06%, 16.1%)

AND

> (2) The proportion of participants in the azithromycin group with sensorineural hearing loss is at least 13% or greater.

2. Report Generation

2.1 Data Flow

An electronic data capture system, Medidata Rave®, will be utilized for collection of study data. The Investigator will prepare and maintain adequate and accurate source documents designed to record all observations and other pertinent data for each participant who signs informed consent.

Study personnel at each site will enter data from source documents corresponding to a participant's visit or assessment into the protocol-specific electronic Case Report Form (eCRF). Participants will not be identified by name in the study database or on any study documents to be collected by the Sponsor (or designee), but will be identified by a site number, participant number and initials. If a correction is required for an eCRF, the time and date stamp will track the person entering or updating eCRF data and creates an electronic audit trail. The Investigator is responsible for all information collected on participants enrolled in this study. All data collected during the course of this study must be reviewed and verified for completeness and accuracy by the Investigator. A copy of the CRF will remain at the Investigator's site at the completion of the study.

2.2 Report Generation

The final statistical report will describe and justify any deviations from the original statistical plan described herein. If necessary, such deviations will be treated as a protocol amendment.

The final statistical report will be accompanied by a description of the data cleaning process and will provide summaries of key findings. All programs used to produce the statistical reports will be documented, validated, and archived. Statistical analyses will be generated by the OPTIMIZE DCC.

The statistical reports will be prepared using current versions of SAS, R and/or S-PLUS statistical software. All statistical tests are two-sided and will be evaluated at a 0.05 level of significance unless otherwise specified.

2.3 Definition of the Analysis Populations

Intent-to-Treat Population

The intent-to-treat population is defined as all randomized participants. Participants who are discontinued from study drug early are encouraged to complete all remaining study visits and will remain in the analysis population according to ITT.

Per Protocol Population

The per-protocol population is defined as all randomized participants who have completed ≥ 80% of expected doses of azithromycin or placebo, were not determined to

be ineligible, did not incur any protocol violations, or who at any point in the study received IV anti-pseudomonal therapy in the absence of accompanying symptoms (as discouraged in the protocol). Additionally, the per-protocol population will be analyzed according to the treatment they received (not how they were randomized).

All analyses included in this report will be performed using the intent-to-treat (ITT) population unless otherwise specified. The primary efficacy analyses evaluating time to protocol-defined pulmonary exacerbation, and key secondary efficacy endpoints evaluating time to *Pa* recurrence and changes in inflammatory markers, will also be performed using the per-protocol population (see sections 4, 5 and 6, respectively).

Missing Data

Site training, automatic queries in the electronic data capture system, and thorough clinical site monitoring will be performed in order to minimize the occurrence of missing data. Participants who discontinue treatment will be encouraged to remain in the study while completing all study procedures. Details regarding the handling of missing data for analyses evaluating the primary efficacy endpoint, time to pulmonary exacerbation, and for analyses of the key secondary efficacy endpoints evaluating time to *Pa* recurrence and change in inflammatory markers will be provided in each of their respective sections of the statistical analysis plan (Section 4, 5 and 6). Observed case analyses will be performed for all other secondary and exploratory endpoints.

2.4 Definitions

Baseline – Baseline refers to the first study visit at which randomization occurs.

Study drug – Study drug refers to the drug the participant was randomized to (either azithromycin or placebo)

1. Summary of Enrollment

The cumulative enrollment of participants entering the study is graphically presented. The number of participants screened, eligible, randomized, treated, and withdrawn post-Baseline are summarized by both treatment arm and site. Reasons for screen failure are summarized. Screen failures are participants that signed informed consent but did not meet eligibility criteria or chose not to participate in the study. The number of participants who withdrew early from the study are tabulated by treatment arm and reason for withdrawal is summarized.

2. Summary of Participant Demographics and Baseline Characteristics

Participant demographics and clinical characteristics at baseline are descriptively summarized by treatment arm and overall. Characteristics summarized are gender, age, race, genotype (homozygous, heterozygous, other), FEV_1, height, weight, and sweat chloride. *P. aeruginosa* (Pa) eligibility status (first lifetime *Pa* culture or history of *Pa*) as well as *Pa* and *Staphylococcus aureus (SA)* status at baseline are also summarized by treatment arm and overall. The number and percent of participants using either a nebulizer or a podhaler for administration of tobramycin at baseline is also summarized, in addition to season of enrollment (winter, spring, summer and fall). The number of participants using hypertonic saline and pulmozyme within thirty days of baseline is also summarized. Fisher Exact tests are used to test at a two-sided 0.05 level of significance for differences in categorical variables across treatment arms. A two-sided t-test assuming unequal variance is used to test the difference between treatment arms for all continuous variables at a two-sided 0.05 level of significance.

3. Summary of Protocol Adherence and Per Protocol Population Definition

The number and percentage of participants who completed each visit is summarized by treatment arm. A study visit is considered completed if a participant completed any study related procedures.

Follow-up time is summarized as well as the average length of follow-up time for each participant. Follow-up time is defined as time between baseline and the last study visit (either end of study Visit 8 [Week 78] or withdrawal date).

Study drug (azithromycin/placebo) could either be stopped permanently or temporarily at any time at the discretion of the subject, site investigator or the Sponsor. If a participant stopped study drug permanently, they were encouraged to remain in the study and complete all remaining study visits. If a participant experienced moderate to severe intolerance of study drug, dosage adjustments could be made at the discretion of the investigator. The number and proportion of participants temporarily or permanently stopping study drug at the discretion of the site investigator are summarized, as well as the reason for study drug stoppage. The number and proportion of participants modifying study drug dosage frequency, at the discretion of the site investigator, is also summarized as well as reasons. The total number and proportion of participants with

physician initiated dose modifications is summarized and the difference in proportions between treatment arms is reported. A corresponding 95% confidence interval is calculated using Newcombe-Wilson method and p-values are from a Fisher's exact test. Study drug compliance is measured based on participant self-report. Participants were instructed to maintain a daily diary in which they recorded doses taken. The diaries were reviewed at each study visit by the site Research Coordinator (RC) and any deviation from the expected dose will be discussed with the participant and reason for deviation noted. To determine compliance, the total number of expected doses for each participant in the study is calculated. Per protocol, 3 doses of study drug are to be taken each week until the end of the study. For participants who completed Visit 8, the number of expected doses is calculated as the number of full weeks between the first dose of study drug and Visit 8 multiplied by 3. For participants who withdrew from the study prior to Visit 8, the number of expected doses is based on the expected time on study per protocol (78 weeks); therefore the number of expected doses is 234. Study drug compliance is calculated as: (Number of doses reported taken / Number of expected doses) x 100. Compliance is not dependent on the number of bottles dispensed or returned. Study drug compliance is summarized by treatment arm and overall as well as the number and percentage of participants with ≥ 80% compliance.

Adherence to the protocol as it pertains to inhaled tobramycin use throughout the trial is also summarized. All participants should have initiated at least a first course of inhaled tobramycin no more than fourteen days prior to Baseline. Subsequently, if a culture resulted in a *Pa* positive result at Visit 2 (Week 3) or any of Visits 3-8 (quarterly visits), an inhaled tobramycin course would be initiated. If the study visit culture resulted in a *Pa* negative result, no inhaled tobramycin would be initiated unless indicated clinically for symptoms. Additionally, at the discretion of the primary care clinician, treatment with TIS may be revised to a 28-day on/28-day off regimen if the *Pa* infection is persistent. A persistent *Pa* infection is defined as any two respiratory cultures collected at a study visits, positive for *Pa* from Visit 3 onward. Sites were encouraged to prescribe TIS, however, a participant could choose to use the TOBI podhaler (TIP) or other commercially available nebulized tobramycin formulations in place of TOBI. The total number and percent of participants initiating the first course of inhaled tobramycin and the time from Baseline to initiating of the first course is summarized. The total number of subsequent *Pa* positive quarters is summarized as well as the number of *Pa* positive quarters for which inhaled antibiotics were initiated and the type of inhaled antibiotic used (TIS, TIP, other inhaled tobramycin, other inhaled antibiotic). Finally, the number of subsequent *Pa* negative quarters is summarized as well as the number of *Pa* negative quarters for which inhaled antibiotics were initiated. A quarter is considered *Pa* positive when at least one *Pa* culture was positive at or between quarterly visits.

A listing of all protocol violations is provided. The number of participants included in the per-protocol population is provided by treatment arm as well as reasons for exclusion. The per-protocol population is defined as all randomized participants who have completed ≥ 80% of expected doses of azithromycin or placebo, were not determined to be ineligible, did not incur any protocol violations, or who at any point in the study

received IV anti-pseudomonal therapy in the absence of accompanying symptoms (as discouraged in the protocol). The per-protocol population will be analyzed according to the treatment they received (not how they were randomized).

4. Summary of Pulmonary Exacerbations

The primary endpoint is time to a protocol-defined pulmonary exacerbation. In this study, we used an *a priori* definition for exacerbation based on signs and symptoms characterized as major criteria or minor criteria (protocol Appendix II). Implementation of this definition was consistent with its definition in the randomized, placebo controlled trial of azithromycin in pediatric patients with CF culture negative for *Pa* (1, 2). At the initiation of each new antibiotic, site investigators completed an electronic case report form and indicated the presence or absence of signs and symptoms comprising the PE definition that prompted antibiotic treatment. However, the decision to utilize antibiotics was not dependent on whether or not the protocol definition for exacerbation was met.

Time to exacerbation is presented using Kaplan-Meier survival estimation and is compared between treatment groups using a Cox proportional hazards model with effects for treatment group and randomization strata (≥6 months – 3 years, ≥3 -6 years, ≥6 – 12 years, ≥12 years), with a test of the treatment effect at a two-sided significance level of 0.05. Participants who withdraw from the study early will be censored at the date of last contact. The results from the model, including the hazard ratio and its 95% confidence interval are presented using both the ITT population and the per-protocol population.

Several sensitivity analyses will also be performed to assess the robustness of study results as outlined in the study protocol. The first sensitivity analysis will utilize a composite endpoint of time to pulmonary exacerbation or repeated use of acute antibiotics (at least 2 acute antibiotic courses prescribed that do not meet the PE definition) (3). Furthermore, the potential influence of acute respiratory events not meeting the primary endpoint definition will also be assessed by censoring participants with acute antibiotic usage prior to meeting the primary endpoint. To investigate the impact of missing data, results from a model will be presented using only participants with complete follow up data. A model will also be presented assuming any censored participant experienced a pulmonary exacerbation at the time of withdrawal or loss to follow-up. Hazard ratios and 95% confidence intervals from each model adjusting for treatment group and randomization strata are presented.

Exploratory subgroup analyses of the primary endpoint will also be performed. Estimates of the hazard ratio and associated 95% confidence interval along with p-values for the primary endpoint will be reported by subgroups defined by pre-specified demographic and disease specific baseline characteristics of interest. No adjustments for multiple comparisons will be made. Additional exploratory analyses include summarizing the frequency of signs/symptoms presented during all pulmonary exacerbations events as well as comparing treatment groups with respect to the average

number of PEs. The results from an Anderson-Gill Cox Model for the mean number of exacerbations with effects for randomization strata and treatment group is also presented with corresponding hazard ratio and 95% confidence interval.

Additional exploratory analyses will be performed utilizing alternative definitions of the PE endpoint incorporating additional data collection during the study. In particular, additional signs and symptom data were collected at scheduled study visits independent of antibiotic use for exploratory purposes and will be used to investigate frequency of signs and symptoms consistent with the PE definition regardless of antibiotic initiation (defined a "significant respiratory event") as well as the frequency of physician defined (rather than protocol defined) exacerbation events.

References

1. Saiman L, Anstead M, Mayer-Hamblett N, Lands LC, Kloster M, Hocevar-Trnka J, Goss CH, Rose LM, Burns JL, Marshall BC, Ratjen F; AZ0004 Azithromycin Study Group. Effect of azithromycin on pulmonary function in patients with cystic fibrosis uninfected with Pseudomonas aeruginosa: a randomized controlled trial. *JAMA.* 2010 May 5;303(17):1707-15.
2. Anstead M, Saiman L, Mayer-Hamblett N, Lands LC, Kloster M, Goss CH, Rose LM, Burns JL, Marshall BC, Ratjen F. Pulmonary Exacerbations in CF Patients with Early Lung Disease. *Journal of Cystic Fibrosis.* 2014 Jan;13(1):74-9. Epub 2013 Sep 10. PMID: 24029220.
3. Mayer-Hamblett N, Saiman L, Lands LC, Rosenfeld M, Kloster M, Fisher L, Ratjen F. Impact of acute antibiotic therapy on the pulmonary exacerbation endpoint in cystic fibrosis clinical trials. *Contemporary Clinical Trials.* 2013 Sep;36(1):99-105. doi: 10.1016/j.cct.2013.06.004. Epub 2013 Jun 14. PMID:23770109.

5. Summary of *Pseudomonas aeruginosa* (*Pa*) Endpoints

Respiratory culture results are obtained at scheduled study visits with additional cultures obtained between visits only if clinically necessary. Analyses include only cultures obtained during scheduled study visits unless otherwise specified. Respiratory cultures could be collected via oropharyngeal (OP) swab, sputum (expectorated or induced) or bronchoalveolar lavage (BAL). In the event that more than one culture is obtained for the same visit from the same participant, a positive result for *Pa* in any of the specimens designates the participant positive for *Pa* at that visit. The number and percent of cultures obtained and by which method is summarized by visit and treatment arm.

A key secondary efficacy endpoint is time to *Pa* recurrence after initial treatment with anti-pseudomonal therapy received during the first quarter of the study. Time to *Pa* recurrence from weeks 13 through 78 is presented using Kaplan-Meier survival estimation and compared between treatment arms using a Cox proportional hazards model with effects for treatment and age group (≥6 months – 3 years, ≥3 – 6 years, ≥6 – 12 years, ≥12 years), with a test of the treatment effect at a two-sided significance level

of 0.05. The results from the model, including hazard ratios and 95% confidence intervals, are presented. This main model will be performed for both the ITT population and the PP population (see Introduction). In the event that there are missing culture results at a scheduled study visit for a participant, the following imputation method will be used. If the participant has culture results available from the quarter directly following the missing culture quarter, the result of that culture will be used for the preceding missing quarter. If there are two or more missing quarters, the participant will be assumed to be a microbiologic "failure" and assumed positive for *Pa.* Further, a sensitivity model of time to *Pa* recurrence will be performed including culture results obtained both at acute study visits and scheduled study visits.

Exploratory subgroup analyses of the time to *Pa* recurrence endpoint will also be performed. Estimates of the hazard ratio and associated 95% confidence interval along with p-values for the primary endpoint will be reported by subgroups defined by pre-specified demographic and disease specific baseline characteristics of interest. No adjustments for multiple comparisons will be made.

The proportion of *Pa*-positive respiratory cultures among all respiratory cultures taken at scheduled study visits from weeks 13 through 78 is also modeled. The response is binary (positive or negative culture) at each visit. A generalized estimating equation (GEE) model using a logit link is used to model this data with robust variance estimation and an independence working correlation matrix, including effects for both treatment and age group (≥6 months – 3 years, ≥3 – 6 years, ≥6 – 12 years, ≥12 years) in addition to the number of available cultures at scheduled study visits over the 18 month study period. The significance of the treatment arm variable is tested using a two-sided 0.05 level of significance. The treatment associated odds ratio from this model and corresponding 95% confidence interval is the measure of treatment effect. The estimated treatment effect is interpreted as the marginal odds ratio of *Pa* positive respiratory cultures during the 18 months. A sensitivity model will also be performed including culture results obtained both at acute study visits and scheduled study visits.

The proportion of participants with *Pa* positive cultures is summarized by visit and treatment. Differences and 95% confidence intervals are reported for each visit.

The proportion of participants who achieved initial eradication at week 3 and after the first quarter of treatment at week 13 is summarized and compared between treatment groups, and additional summaries are provided for the proportion who were *Pa* positive at week 3 but were negative by week 13.The proportion of patients with at least one *Pa* positive culture result after the first quarter of treatment (on or after week 13) is also summarized for each treatment arm with the difference between treatment arms tested using a 0.05 level of significance Chi-square test. The number of participants with persistent *Pa* infection at some point during the trial is summarized. Persistence is defined in the protocol as any two respiratory cultures collected at study visits, positive for *Pa* from Week 13 onward. The difference between treatment arms is tested using a two-sided 0.05 level Chi-square test. Additionally, among those who turned persistent,

the inhaled tobramycin formulation used and frequency of treatment (i.e. 28 days on/28 days off, quarterly) after persistent infection is summarized.

Time to persistent *Pa* infection is presented using Kaplan-Meier survival estimation and compared between treatment arms using a Cox proportional hazards model with effects for treatment and age group (≥6 months – 3 years, ≥3 – 6 years, ≥6 – 12 years, ≥12 years), with a test of the treatment effect at a two-sided significance level of 0.05. The results from the model, including hazard ratios and 95% confidence intervals, are presented.

Exploratory analyses of the frequency of mucoid *Pa* cultures were also performed. *Pa* mucoid status at each visit and the change for each visit from baseline is summarized by treatment arm. *Pa r*esistance to aminoglycosides, quinolones and β-lactams were also tested when indicated clinically. The number and proportion of participants with at least one emergent drug-resistant *Pa* culture post-baseline is summarized by treatment arm for each drug class among those participants with at least one result available for testing. Differences and 95% confidence intervals are presented.

A multivariable logistic regression model using stepwise regression is developed to identify key baseline predictors associated with initial eradication failure, determined using *Pa* status at Week 13. Baseline demographic, clinical, microbiologic and inflammatory variables will be assessed for inclusion of the model. Cross-validation techniques are used to determine a model. The model coefficients and standard errors (SE) are presented as well as corresponding 95% confidence intervals and the p-values. The positive predictive value, negative predictive value, sensitivity and specificity of the model are presented.Summary of Pseudomonas aeruginosa (Pa) Endpoints

6. Summary of Inflammatory Markers

Serum inflammatory markers (myeloperoxidase (MPO), C-reactive Protein (CRP) and calprotectin) are transformed using $\log_{10}$ and summarized at each visit (Baseline, Week 39 and Week 78) by treatment arm. For values below the limit of detection, ½ the lower limit of detection is used. For values greater than the limit of detection, the upper limit of detection is used.

Change from baseline to end of study in each inflammatory marker is a key secondary efficacy endpoint. Linear regression will be used to estimate the difference between treatment groups in the 78 week change from baseline in each inflammatory marker, adjusted for randomization strata. These analyses will be performed on both the ITT and PP population. Additional sensitivity analyses will be performed to address missing data. Specifically, the least favorable treatment arm imputation method will be used which imputes that missing value with the mean change from the treatment arm with the worst change in the observed case analysis.

Observed changes from baseline at Weeks 39 and 78 are also descriptively summarized by treatment arm. The change from baseline in inflammatory markers at Weeks 39 and 78 are also graphed for each treatment arm along with 95% confidence intervals.

7. Summary of Adverse Events

Treatment emergent adverse events (AEs) and serious adverse events (SAEs) will be comprehensively summarized. Treatment emergent adverse events are those events that occurred on or after Baseline as reported by the participant. The safety population is used in all summaries of SAEs and AEs. Note that all tables summarizing AEs also include SAEs (which are a subset of the AEs) unless otherwise specified.

Included in the summaries are tables displaying: 1) the proportion of participants with at least one (S)AE, 2) the average number of (S)AEs per patient, and 3) the rate of (S)AEs per participant-month of follow-up. Poisson regression is used to compare the rate of (S)AEs by treatment arm. Rate ratios and corresponding 95% confidence intervals and p-values are reported. The difference in the proportion of participants experiencing at least one (S)AE between treatment arms is also reported, with corresponding 95% confidence intervals calculated using the Newcombe-Wilson method and p-values derived from a Fisher's exact test.

A histogram showing the number of (S)AEs per participant in each treatment arm is displayed. Detailed summary tables of individual (S)AEs grouped by SOC will be included to show the number and percent of participants experiencing each (S)AE by treatment arm. The number and rate of events for each (S)AE is also reported by treatment arm. Poisson regression is used to derive rate ratios and corresponding 95% confidence intervals comparing treatment arms for each SOC. Treatment arms will be compared by plotting rates of (S)AEs and the proportion of participants experiencing (S)AEs by System Organ Class (SOC) as well as the rate ratio and difference in proportions with corresponding 95% confidence intervals to compare treatment arms. The most frequently occurring adverse events, defined as those events occurring in ≥ 5% of participants, will be separately summarized by SOC and treatment arm.

The severity of AEs will be tabulated by SOC and treatment arm according to whether the AE was mild, moderate, severe or life-threatening. The number and proportion of participants experiencing at least one AE classified as severe or worse is summarized by treatment arm. The difference in the proportion of participants experiencing at least one severe AE or worse between treatment arms is reported. Corresponding 95% confidence intervals calculated using Newcombe-Wilson method and p-values derived from a Fisher's exact test. The proportion of participants experiencing AEs and rates of AEs will be summarized by the relation to study drug (unrelated, possibly, probably, definitely) by treatment arm. Finally, the rate of AEs will be plotted by site.

Both sensorineural hearing loss and protocol-defined abnormal ECG results, defined *a priori* as key safety endpoints, will be reported as SAEs and therefore will be included in

SAE and AE tables. All SAE narratives will also be included in a listing appended to this report.

8. Safety Laboratory Parameters

Blood was drawn for serum chemistry and hematology at Baseline, Visit 5 (Week 39) Visit 8 (Week 78) and at an Early Withdrawal Visit (if participant withdrew early). In the case that a participant had labs drawn at an Early Withdrawal Visit, those labs are included in the summaries for the visit that should have occurred closest to and after their early termination visit.

Average lab values at each time point and average changes from Baseline are summarized by treatment arm. The difference between treatment arms in the mean change from Baseline to each visit are displayed with corresponding 95% confidence interval. If a participant did not have labs at a specified time point, they are excluded from that particular measure.

The number and percent of participants with lab measures below (low), within (normal) and above (high) the normal range at each visit and for each lab are displayed by treatment arm. Each site used their own standard normal range, therefore, normal ranges for each measure may vary by site. For each abnormal lab value (low or high), the site PI determined if the abnormal lab was considered clinically significant or not. Thus, clinically significant lab values are a subset of the abnormal values. The number and percent of lab values deemed to be clinically significant are summarized by visit and treatment arm. Changes in abnormal and clinically significant lab values from Baseline at Visits 5 and 8 are summarized as emergent high or low. For example, a lab value is determined to be emergent low at a visit if the lab value was not low at Baseline but low at the visit . Emergent high values and emergent clinically significant low and high values are similarly categorized. The difference in proportions between treatment arms of participants with emergent clinically significant lab measures at Visit 5 and Visit 8 are displayed with corresponding 95% confidence interval.

9. Summary of Hospitalizations and Antibiotic Usage

Descriptive statistics are used to summarize the percent of participants hospitalized (all cause) by treatment arm, including the difference between treatment groups with associated 95% confidence interval calculated using the Newcombe Wilson method. Differences between treatment groups with respect to the proportion of participants hospitalized is assessed using a logistic regression model adjusting for treatment and age category (≥6 months – 3 years, ≥3 – 6 years, ≥6 – 12 years, ≥12 years). The Odds ratio and 95% confidence interval from the model are presented along with a p-value. The rate of hospitalizations per month is also descriptively summarized and compared using Poisson regression, including follow-up time as an offset, with an associated rate ratio and 95% confidence interval

Similarly, the percent of participants initiating oral, IV, and inhaled antibiotics after Baseline and over the follow up period and the number of days of oral and IV antibiotic use are descriptively summarized by treatment arm. Differences in proportions and the number of days of usage are summarized with associated 95% confidence intervals. The proportion of participants initiating oral, IV and inhaled antibiotics are compared using logistic regression modes adjusting for treatment arm and age category (≥6 months – 3 years, ≥3 – 6 years, ≥6 – 12 years, ≥12 years). The Odds ratios and 95% confidence intervals from the models are presented along with associated p-values.

10. Summary of Non-*Pseudomonal* Microbiologic Endpoints

Respiratory culture results are obtained at each study visit with additional cultures obtained between study visits only if clinically necessary. . Analyses include only cultures obtained at scheduled study visits. The respiratory culture could be collected via oropharyngeal (OP) swab, sputum (expectorated or induced) or bronchoalveolar lavage (BAL). In the event that more than one specimen is obtained for the same visit from the same participant, a positive result for an organism in any of the specimens designates the participant positive for that organism at that visit. The number and percent of cultures obtained and by which method is summarized by visit and treatment arm.

Non-*pseudomonal* microbiologic organism culture results (*Staphylococcus aureus* [methicillin susceptible and resistant], *Haemophilus influenzae*, *Burkholderia cepacia* complex, *Stenotrophomonas maltophilia*, and *Alcaligenes (Achromobacter) xylosoxidans)* are summarized for each visit and the change for each visit from baseline by treatment arm Differences and 95% confidence intervals are presented. For each organism, the treatment emergent prevalence is summarized by treatment arm. Differences and 95% confidence intervals are presented as well as p-values from a Chi-square test. NTM results are available only when sputum was produced at study visits in a quantity that allows for testing. The number and percent of participants who were able to produce sputum at any visit post-baseline is summarized as well as the number and percent of participants with an NTM positive result by treatment arm. The difference and 95% confidence interval is presented Note that a positive NTM result triggered permanent study drug discontinuation for that particular participant although they were encouraged to stay in the study for full follow-up.

11. Spirometry Results

Spirometry was to be performed at Baseline for all participants 4 years and older who could reproducibly perform spirometry. Baseline clinical and demographic characteristics are summarized among the subset of participants with spirometry values available at the Baseline visit. All spirometry tables, figures and models include those participants with a Baseline measurement. FEV_1 (liters and percent predicted), FVC and $FEF_{25\%-75\%}$ are descriptively summarized for each study visit and change (absolute and relative) from Baseline by treatment. The difference between treatment arms in the change from Baseline at each visit is presented along with 95% confidence intervals. FEV_1 percent

predicted is calculated using the Wang (1) equations (Female ≤ 16 years, Male ≤ 18 years) or Hankinson (2) equations (Female > 16 years, Male > 18 years) at each visit. Participants crossing age thresholds for a particular equation (e.g. from Wang to Hankinson) during the study remained on the same reference equations determined at Baseline to maintain continuity. Relative change is defined as: [(Visit Measure – Baseline Measure) x100]/Baseline Measure. The mean absolute and relative change from Baseline are plotted by treatment for each spirometry measure with 95% confidence intervals at each time point.

For each spirometry measure, a repeated measures linear regression model is used to model change over time among the subset of participants with spirometry at Baseline. The model adjusts for age group (≥6 months – 3 years, ≥3 – 6 years, ≥6 – 12 years, ≥12 years), time (days) and a time by treatment interaction. A generalized estimating equation framework with an independence working correlation matrix is utilized to calculate robust variance estimates. The coefficients of the model, along with corresponding standard error (SE), 95% confidence intervals (CI) and p-values are displayed. These models are utilized to estimate the 18-month change from Baseline with corresponding 95% CIs and p-values for each measure.

References:

1. Wang X, Dockery DW, Wypij D, Fay ME, Ferris BG, Jr. Pulmonary function between 6 and 18 years of age. Pediatr Pulmonol **1993**; 15(2): 75-88.
2. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. Am J Respir Crit Care Med **1999**; 159(1): 179-87.

12. Anthropometric Results

Height/length and weight were collected at each study visit. Height/length, height/length percentile, weight and weight percentile are descriptively summarized for each study visit and change (absolute and relative) from Baseline by treatment arm. WHO reference tables (1) were used to determine percentiles for participants < 2 years old at Baseline. CDC reference tables (2) were used to determine percentiles for participants ≥ 2 years old at Baseline. Participants crossing age thresholds for a particular equation (e.g. from WHO to CDC) during the study remained on the same reference equations determined at Baseline to maintain continuity. The difference between treatment arms in the change from Baseline at each visit is presented along with 95% confidence intervals. The mean absolute change from Baseline is plotted by treatment for each measure with 95% confidence intervals at each time point.

For each anthropometric measure, a repeated measures linear regression is to model change over time among the subset of participants with spirometry at Baseline. The model adjusts for the age randomization factor (age group: ≤3 years, >3-6 years, ≥6-12 years, >12-18 years), continuous time (in days) and a treatment by time interaction. A generalized estimating equation framework with an independence working correlation

matrix is utilized to calculate robust variance estimates. The coefficients of the model, along with corresponding standard error (SE), 95% confidence intervals (CI) and p-values are displayed. These models are utilized to estimate the 18-month change from Baseline with corresponding 95% CIs and p-values for each measure.

References

1. WHO, Apr 2006. Child Growth Standards based on length/ehight,weight and age. *Acta Paedreatr Suppl* Apr 2006; 450:76-85
2. Kuczmarski, Robert J., Cynthia L. Ogden, Laurence M. Grummer-Strawn, Katherine M. Flegal, Shumei S. Guo, Rong Wei, Zuguo Mei, Lester R. Curtin, Alex F. Roche, and Clifford L. Johnson. "CDC growth charts: United States." *Advance data* 314 (2000): 1-27.

13. Summary of the Cystic Fibrosis Respiratory Symptoms Diary (CFRSD)

The Cystic Fibrosis Respiratory Symptoms Diary (CFRSD) and the parent-completed CFRSD will be completed at each study visit for participants ≥12 years and <11 years, respectively.

The CFRSD asks a participant (or parent) to state the extent of their 8 respiratory symptoms: difficulty breathing, felt feverish, felt tired, had chills or sweats, coughed, coughed up mucus, had tightness in the chest, or wheezed. If a symptom is present, they are asked to rate the severity of the symptom using a 4 point scale with possible answers that include “a little”, “somewhat”, “a good deal”, “a great deal” or “slightly”, “moderately”, “very”, “extremely”. Each respiratory symptom is assigned a score from 0-4 based on the response (0 in the absence of symptom). For each participant, a summed score is then calculated, ranging from 0 to 24. In turn, this score is converted to a 0 to 100 scale. This composite score is referred to as the CFRSD-CRISS.

The CFRSD-CRISS is summarized at each visit and changes from Baseline by treatment. The difference between treatment arms in the change from Baseline for the CFRSD-CRISS at each visit are presented along with 95% confidence intervals. Summaries at each visit include scores from participants who responded to at least 7 of the respiratory questions.

A repeated measures linear regression model was used to model changes in the CFRSD-CRISS over time. The model adjusts for age group (≥6 months – 3 years, ≥3 – 6 years, ≥6 – 12 years, ≥12 years), continuous time and a time by treatment interaction. A generalized estimating equation framework with an independence working correlation matrix is utilized to calculate robust variance estimates. The coefficients of the model, along with corresponding standard error (SE), 95% confidence intervals (CI) and p-values are displayed. These models are utilized to estimate the 18-month change from Baseline with corresponding 95% CIs and p-values for the CFRSD-CRISS.

The severity of symptoms from the CFRSD at Baseline and Visit 8 (Week 72) are summarized by treatment as well as change in severity of symptoms from Baseline. The proportion of participants experiencing symptoms at Baseline and Visit 8 are summarized by treatment arm as well as differences between treatment arms with corresponding 95% confidence intervals. The proportion of participants with a decrease and increase in symptom severity from Baseline to Visit 8 is summarized by treatment arm as well as differences between treatment arms with corresponding 95% confidence intervals.