CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

050814Orig1s000

STATISTICAL REVIEW(S)

Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA # :	50814 (Amendment S/N 040)
Drug Name:	Aztreonam lysine for inhalation (75mg TID)
Indication(s):	To improve respiratory symptoms and pulmonary function in cystic fibrosis patients with *P.aeruginosa*
Applicant:	Gilead Sciences, Inc.
Stamp Date:	August 13, 2009
PDUFA Goal Date:	February 13, 2010
Reviewer Completion Date:	January 27, 2010
Biometrics Division:	Division of Biometrics IV (HFD-725)
Medical Division:	Division of Anti-Infective and Ophthalmology Drug Products (HFD-520)
Documents Reviewed:	\\CDSESUB1\evsprod\Nda050814\0040
Statistical Reviewer:	Christopher Kadoorie, PhD
Concurring Reviewers:	Thamban Valappil, PhD
Clinical Team Leader:	John Alexander, MD, MPH
Clinical Reviewers:	Menfo Imoisili, MD, MPH, John Alexander, MD, MPH
Project Manager:	Kyong Hyon, BSN, MA

TABLE OF CONTENTS

List of Tables

List of Figures

1. EXECUTIVE SUMMARY

1.1 Introduction

The Applicant had originally submitted NDA 50814 on November 16, 2007 to support approval of CAYSTON™ (75mg of aztreonam inhalation (AZLI)) as a three times daily (TID) treatment to improve respiratory symptoms and pulmonary function in patients with cystic fibrosis due to *P.aeruginosa*. The Agency did not recommend approval for this submission and had issued a complete response letter on September 16, 2008. On November 24, 2008 the Applicant made a formal dispute resolution request (FDRR) to the Office of Antimicrobial Products which was denied on February 18, 2009. On March 13, 2009, the Applicant made another FDRR to the Office of New Drugs (OND) and had provided additional analyses on April 8 and 22, 2009 to address the OND concerns. OND had denied the Applicant's appeal on the basis that the FDRR Guidance specifically proscribes new data from consideration. OND recommended that the Applicant provide a complete response to the Agency's CR letter including the new analyses relevant to NDA 50814. The Applicant has followed this recommendation and has provided a complete response including the new post-hoc analyses in the current submission. The Applicant had also requested an anti-infective advisory committee (AIDAC) to discuss the safety and efficacy of AZLI 75mg TID. The AIDAC was held on December 10, 2009 and committee members voted in favor of AZLI 75mg TID (15 to 2) as being safe and effective.

This Review will address statistical issues with the evidence provided by the Applicant in Study 005 and Study 007 as well as the Applicant's new post-hoc analyses.

1.2 Conclusions and Recommendations

Study 005 was an uninformative study providing very limited supportive evidence. Study 005 primary analyses failed to demonstrate the efficacy of AZLI TID in prolonging time to need for IV or inhaled anti-pseudomonal antibiotics due to predefined symptoms predictive of pulmonary exacerbations. There was also substantial dependent missing data significantly favoring the TID vs. BID regimens as well as AZLI vs. Placebo during the TOBI run-in period. Among randomized patients without a primary analysis time to need event, AZLI TID patients had the highest early termination rate that was substantially higher than the rate for Placebo BID patients (58.6% vs. 37.9%). Since there is no evidence to suggest that missing data was uninformative or not treatment related, there is potential for biases in the primary analysis favoring AZLI TID.

In Study 005, there was additional uncertainty due to a detrimental TID vs. BID regimen effect which was more influential than the AZLI drug effect in the primary analysis. Due to this uncertainty, AZLI TID comparisons against Pooled Placebo or Placebo TID would be problematic since these placebo groups may fail to reliably estimate the true placebo rate due to a detrimental TID regimen effect. Therefore, AZLI TID vs. Placebo BID was considered to be the most informative comparison. In the primary analysis, AZLI TID patients fared worse than Placebo BID patients (p=.5377). Furthermore, in sensitivity

analyses, AZLI TID comparisons against Placebo BID were even less favorable than in the primary analysis due to a highly robust regimen effect that was consistently stronger than the AZLI drug effect. In a sensitivity analysis using a broad definition for 'event' to control for potential biases from informative dropouts (i.e. time to early termination), AZLI TID patients fared substantially worse than Placebo BID patients, p=.0941 (Table 3). In other Study 005 analyses, AZLI TID patients fared similar to or possibly worse than placebo over longer time periods. For example, over the entire 84 day AZLI/placebo study period, AZLI TID patients had a substantially shorter 'time to hospitalization' compared to Pooled Placebo patients (p=.085). Although the Applicant did provide post-hoc analyses in this submission attempting to explain biases in the primary endpoint of 'time to need' using FEV_1 findings, these analyses were not considered to be adequate especially given a highly significant and robust regimen effect compounded with dependent missing data.

In Study 005, the Cystic Fibrosis Questionnaire-Revised Respiratory Symptoms Score (CFQ-R RSS) and FEV_1 endpoints showed marginal AZLI TID benefits when considering only the Day 0 to Day 28 time period, however this evidence was limited by dependent missing data (both before and after AZLI/placebo treatment), regimen effects in the CFQ-R RSS, multiplicity issues, an unclear primary analysis and other unfavorable study findings. CFQ-R RSS changes at Day 28 for AZLI TID vs. Pooled Placebo also failed to meet the minimum clinically important difference (MCID) of ± 5 points at 4.37 (-0.94, 9.69), p=.1061, while changes (worsening) from Placebo TID vs. Placebo BID did meet the MCID at -5.66 (-11.92, 2.61), p=.2078. Therefore a substantial portion of the observed AZLI TID treatment benefit over Pooled Placebo may be due to detrimental TID regimen effects included in the Pooled Placebo estimate as well as potential biases from higher rates of missing data in AZLI TID patients. Sensitivity analyses controlling for these factors also failed to show any trend towards an AZLI TID benefit in CFQ-R RSS at Day 28. Furthermore, after Day 28, the CFQ-R RSS was not assessed until the Day 84 visit at which AZLI TID patients fared significantly worse than placebo patients based on imputed data, p=.020 (Table 6).

Study 007 demonstrated its primary endpoint, changes in the CFQ-R RSS from Day 0 to Day 28, showing an improvement over placebo of 9.71 points (95% CI: 4.31, 15.11) which reduced to 6.33 points (95% CI: 1.22, 11.43) at the Day 42 visit. Although Study 007 provided some evidence of sustained improvement in respiratory symptoms, this evidence was limited because AZLI TID mean improvements from baseline had dropped substantially from Day 28 (7.08 points) to Day 42 (0.62 points). These findings were further limited by uncertainty in the validity and reliability of the CFQ-R RSS instrument in demonstrating clinical improvement based on a pre-defined MCID of a 5 point increase. There were also limitations in the robustness of primary analysis findings. For example, the CFQ-R RSS at Day 28, was primarily driven by patients <18 years of age (n=37) who had substantially larger improvements over placebo compared to patients ≥18 years of age (n=126) at 18.92 (95% CI: 8.78, 29.05), p=.0006 vs. 6.35 (95% CI: 0.02, 12.69), p=.0495.

(b) (4)

However, there is some evidence of an improvement in respiratory symptoms from Day 0 to Day 28 from Study 007, but this evidence is limited in patients aged 18 years and older. Due to the nature of Cystic Fibrosis due to *P. aeruginosa*, clinical and other considerations should also be taken into account when evaluating the evidence of improvement in respiratory symptoms.

We recommend that additional studies are conducted to better address many of the limitations stated in this Review. These limitations were often related to observed regimen effects, regimen dependent dropout rates, respiratory improvements in patients in the < 18 and ≥ 18 year age groups and sustained AZLI treatment effects beyond Day 28 and over multiple cycles.

1.3 Brief Overview of Clinical Studies

There were no clinical studies presented in this submission. A brief overview of Studies 005 and 007 which were included in the original submission is shown below.

Table 1: Overview of Studies 005 and 007

Comparison of Studies 005 & 007	Study 005	Study 007
Type of Study:	Phase 3 randomized, double-blind study designed to assess the safety and efficacy of a 28 day course of 75mg aztreonam inhalation (AZLI) versus placebo in cystic fibrosis (CF) patients aged 6 years or older with lung disease due to *PA*.	
Objective:	Demonstrate efficacy and safety of 28 day course of 75mg AZLI versus placebo in cystic fibrosis (CF) patients aged 6 years or older with lung disease due to *PA*.	
Treatment Arms:	Four arms: AZLI BID, AZLI TID, Placebo BID, Placebo TID with 2:2:1:1 Randomization	Two arms: AZLI TID & Placebo TID, with 1:1 Randomization
Sample Size:	211 ITT patients	164 ITT patients
Primary Endpoints:	Time to Need for Inhaled or IV Antibiotics After Initial AI/Placebo Dosing (Day 0)	Actual Change in CFQ-R Respiratory Domain Scores at End of Treatment (Day 28)
Study Design:	Patients randomized at Day -28 and receive 28 day course of TSI Patients start study therapy on Day 0 and receive a 28 day course of AZLI/Placebo, Patients followed to Day 84	Patients randomized at Day 0 and receive 28 day course of AZLI/Placebo, Patients followed to Day 42

1.4 Major Statistical Issues and Findings

The following are major statistical issues and findings relating to study evidence. Issues and findings were considered major if they were likely to have a substantial impact on overall findings. The four major issues identified in this review all related to Study 005: 1) substantial dependent missing data, 2) strong regimen effect, 3) lack of significant AZLI TID comparisons, 4) inadequate post-hoc analyses.

1.4.1 Substantial Dependent Missing Data (Study 005)

Study 005 results may be biased from substantial missing data dependent upon the treatment, regimen and other unknown factors. These biases may favor AZLI TID since dropouts, who may be sicker than those remaining in the study, were significantly greater in the AZLI vs. Placebo arms (Day -28 to 0, TSI run-in period) as well as in TID vs. BID arms (Day 0 to 84). The notion that dropouts are sicker is supported by the high rate of dropouts who withdrew due to an adverse event or treatment intolerance (i.e. 88% of ITT dropouts). Due to the relatively high drop out rates for AZLI TID patients, there is a potential for substantial biases in comparisons of AZLI TID vs. placebo.

It should be noted that five patients (2 AZLI BID, 3 AZLI TID) had early termination events on the day they completed the study and were not considered dropouts. However, since potential study biases may still arise from these patients, they were included in Reviewer analyses of missing data based on early termination rates.

Overall Early Termination Rates (Change from Baseline Endpoints)

Early termination rates were significantly higher in patients randomized to AZLI vs. Placebo arms (Day -28 to 0, TSI run-in period) at 17.9% (AZLI) vs. 7.3% (Placebo), p=.026 and significantly higher in TID vs. BID arms (Day 0 to 84) at 69.2% (TID) vs. 50.5% (BID), p=.005. Early termination rates in randomized patients for the four treatment arms were as follows: Placebo BID (56.1%), AZLI BID (57.3%), Placebo TID (80.5%) and AZLI TID (70.7%).

Primary Analysis Early Termination Rates (Time to Need)

Since the primary outcome accounts for those dropouts who withdrew due to a primary analysis time to need event, it is more relevant to consider only those dropouts without a time to need event in estimating potential biases from missing data. Although potential biases were somewhat reduced in primary analyses, there are still concerns due to more AZLI TID dropouts withdrawing for reasons other than a primary analysis time to need event. Early termination rates in randomized patients without primary analysis time to need for the four treatment arms were as follows: Placebo BID (37.9%), AZLI BID (44.4%), Placebo TID (46.7%) and AZLI TID (58.6%).

1.4.2 Strong Regimen Effects (Study 005)

- Study 005 results may have further biases due to the presence of a strong and detrimental TID vs. BID regimen effect (i.e. TID regimen effect – BID regimen effect) in many analyses. This regimen effect outweighed the drug effect in the primary analysis and even more so in the sensitivity analyses after accounting for missing data.

- Sensitivity analyses, as pre-specified by the Applicant, considered the primary endpoint of 'time to need' in a hierarchical manner with loosening restrictions, S1: time to need (any pre-defined symptoms), S2: time to need for specified anti-psuedomonal antibiotics (any administration route), S3: time to early termination (any reason). These analyses aimed to show the robustness of primary analysis findings as well as to control for potential biases from informative censoring. The Reviewer also considered an additional post-hoc sensitivity analysis, S4, of time to early termination in all randomized patients to control for potential biases occurring prior to the AZLI/placebo treatment period.

- Sensitivity analyses (S1-S4) showed that the regimen effect observed in the primary analysis remained highly significant and robust to various definitions of event. In contrast, AZLI drug effects in the primary analysis were not robust in the sensitivity analyses. Differences between Placebo BID and AZLI TID favored Placebo BID as more restrictions in time to need event definition were loosened with p-values decreasing from .5377 to .0941 (Table 3). This trend was due to larger numbers of AZLI TID dropouts without events who were now counted as having an event in an analysis using a broader definition of 'event'.

1.4.3 Unfavorable AZLI TID Comparisons (Study 005)

Study 005 results failed to show significant AZLI TID benefits over placebo for most clinical endpoints even without controlling for multiplicity.

- As discussed in 1.4.2, AZLI TID patients fared worse than Placebo BID patients in time to need for antibiotics (regardless of symptoms or route of administration) and time to early termination.
- AZLI TID patients showed further possible detrimental effects based on shorter time to first hospitalization compared to Pooled Placebo patients (p=.085) and other hospitalization measures.
- AZLI TID patients achieved only marginal benefits vs. Pooled Placebo in the CFQ-R RSS & FEV_1 endpoints from Day 0-28 with no sustained benefit by Day 42 and worsening by Day 84.
- AZLI TID patients had significant worsening relative to Placebo BID in the CFQ-R RSS at Day 84 based on imputed data (p=.020).

- AZLI TID patients failed to achieve significant improvements versus placebo based on weight, BMI, FVC, FEF_{25-75} and CFQ-R (non-respiratory) from Day 0-28 or missed school/work from Day 0-84.

1.4.4 Inadequate Post-Hoc Analyses (Study 005)

The Applicant provided post-hoc sensitivity analyses in this submission (subsequent to the original NDA application) to address Agency concerns relating to the regimen effect. These post-hoc analyses aimed to show that the primary endpoint of time to need had an unanticipated bias since AZLI patients with events tended to react to relative FEV_1 declines following improved FEV_1 levels whereas placebo patients with events tended to react to absolute FEV_1 declines from baseline predictive of pulmonary exacerbation. According to the Applicant, biases would be expected to be most pronounced in comparisons involving AZLI TID patients since AZLI TID patients had steep initial gains. The Applicant's post-hoc analyses included two key sensitivity analyses.

- Sensitivity analysis #1 (S_1) was based on Cox regression analyses of time to need controlling for the FEV_1 slope covariate.
- Sensitivity analysis #2 (S_2) censored time to need events (i.e. changed 'event' to 'no event') if FEV_1 improved from baseline at Day -28.

However, these analyses were not considered adequate in clarifying Study 005 findings. There were several concerns as stated below.

- The Applicant's re-analyses were conducted in a post-hoc manner and not based on pre-specified hypotheses, there are concerns that assumptions used in the analysis could have been determined retrospectively and primarily driven based on the significance of findings observed (after the data was unblinded).

- In addition to the post-hoc nature of the analyses, the strength of evidence from the analyses was weak given the high degree of uncertainty in the primary analysis (e.g. regimen effects, dropouts). FEV_1 improvements over placebo were not significant at Day 28 ($p=0.057$) and had decreased rapidly after Day 28 (Figure 14). During the time period from Day 42 to Day 84 in which the majority of time to need events had occurred, FEV_1 improvements over placebo were negligible.

- The Applicant's re-analyses assumed FEV_1 changes can be carried forward (LOCF) or represented by a linear trend using a mixed model repeated measures (MMRM) approach in imputing missing data. However, LOCF assumptions are problematic because changes in FEV_1 for AZLI TID vs. Pooled Placebo generally sloped downward representing a decreasing AZLI benefit over placebo over time (Figure 14). Since these downward slopes were also not constant over the Day 28-84 period, tending to be steepest from Day 28-42, an MMRM approach may not account for these steep declines accurately.

- FEV1 (% change) is problematic for use as a post-hoc 'gold standard' to 'judge' for biases in the time to need endpoint. The association of FEV_1 & 'time to need' is unclear since these endpoints measure different symptoms in different manner.

- The Applicant's first sensitivity analysis, 'S_1: Time to Need Adjusting for FEV_1 Slope', inappropriately uses FEV_1 slope as a covariate. FEV_1 slope is treatment related (confounded) and controlling for this factor can result in serious biases.

- The Applicant's second sensitivity analysis, 'S_2: Changing 'event' to 'no event' if FEV_1 improved from baseline', inappropriately uses a Day -28 vs. Day 0 baseline and includes hospitalized patients. Findings from our analysis, using a Day 0 baseline and not changing 'event' to 'no event' in those hospitalized, were similar to findings in the primary analysis and showed no benefit from AZLI TID treatment.

1.5 Other Statistical Issues and Findings

The following are other statistical issues and findings which were important to consider in evaluating the evidence of efficacy provided by each study.

Study 005

- There is subjectivity in the primary endpoint since investigators may assess patient reporting of time to need events due to pre-defined symptoms differently. In addition, prescribing physician of antibiotics in patients with time to need events were not always trained trial investigators.

- There were several cases in which time to need events were reported after the nominal 84 study period or after the patient had discontinued from the study. Therefore, it is questionable as to whether these patients should be reported in the primary analysis as having an event.

- The randomization of Study 005 was on Day -28 when patients started a 28 day course of TSI. The start of study drug, AZLI/Placebo, was not until Day 0. This may be problematic because of possible imbalances at Day 0 among the treatment groups with respect to patient dropout rates and risk of clinical worsening in the AZLI/placebo study period (Day 0-84).

- Study 005 did not include a "washout" period after the 28 day course of TSI. This may result in biases favoring AZLI if there is a synergistic effect between the AZLI and the residual TSI such that AZLI efficacy is overestimated.

Study 007

- Study 007 included only 14 days of follow-up (i.e. Day 29 to Day 42) and failed to observe a critical time period after Day 42 (e.g. Day 43-56). Note that the majority of time to need events in Study 005 occurred after Day 42.

- There are Agency concerns regarding the validity and reliability of the CFQ-R instrument due to the potential for recall bias. This may limit the strength and interpretability of study findings.

- Improvements in the CFQ-R RSS at Day 28 from AZLI TID, though highly significant at Day 28, could not establish a clinically meaningful improvement of 5 points based on the lower bound of the 95% CI of the mean.

- Primary analysis results may not be robust since patients 18 years or older demonstrated only a marginal benefit from AZLI TID therapy.

- Study 007 failed to show a significant AZLI TID treatment benefit in 'time to need for IV/inhaled antibiotics' and 'time to hospitalization. The Agency considers these endpoints to be clinically most meaningful. Note that significance was not found for either endpoint in either Study 005 or Study 007.

- Study 007 involved a sample size re-estimation (SSR) performed by an unblinded third party. Biases can result from SSR especially when not executed according to plan. The actual sample size planned (n=140) differed from the sample size recommended by the third party (n=150). A sample size of 164 subjects was actually used which indicates Study 007 may have been overpowered.

2. INTRODUCTION

2.1 Overview

2.1.1 Class and Indication

According to the Applicant, AZLI is a novel formulation of a synthetic monobactam antibiotic aztreonam, which has been used extensively as parenteral therapy for infections caused by a wide range of gram-negative bacteria. The parenteral form of aztreonam, aztreonam for infection (Azactam®) has been approved for use in the US since the mid 1980s. Azactam is indicated for the following infections caused by gram-negative organisms: urinary tract infections, lower respiratory tract infections, septicemia, skin and skin structure infections, intra-abdominal infections, and gynecological infections. It is also indicated for adjunctive therapy to surgery caused by susceptible organisms.

2.1.2 Applicant's Response to Agency's CR Letter

The Applicant has provided a complete response to the Agency's CR letter of September 16, 2007 which attempts to provide persuasive evidence of efficacy of the AZLI TID treatment. The Applicant believes that the new exploratory analyses would support efficacy of AZLI TID in Study 005 due to an inherent bias in the time to need endpoint.

2.2 History of Drug Development (ONLY for Internal Agency Use)

The following is a timeline of some of the key events (not all) in the history of drug development. Some key points from these meetings related to this review are listed below for each date. These descriptions do not attempt to summarize the meeting discussions.

March 2002: FDA granted orphan designation for aztreonam as inhalation therapy for control of gram-negative bacteria in the respiratory tracts of patients with CF.

April 29, 2004: Corus and Division met to discuss the proposed study design. The Division mentioned the following:

- Change in FEV_1 and its relationship to clinical benefit is not well defined and may be difficult to interpret. The expected benefit of AZLI in terms of clinical symptoms is of greater interest.
- The Pulmonary Division views FEV_1 as a surrogate marker for lung function, not an endpoint, esp. with chronic obstructive pulmonary diseases (COPD).
- TOBI trials demonstrated additional benefits to FEV_1 improvements, such as less hospitalization.
- The Division suggestions included a three-arm placebo-controlled trial in TOBI resistant patients comparing 2-week vs. 4-week AZLI treatment

July 1, 2004: Corus and the Division met to gain concurrence on the design of the phase 3 trials.

- The Division suggested using improvement in respiratory symptoms as the primary endpoint and FEV as a secondary endpoint.
- The Division also suggested that Corus consider a large Phase 3 trial, which along with supportive information from the Phase 2 trial, would answer the efficacy and safety issues.

August 24, 2004: Corus and the Division had a telecon.

- A Phase 2 trial to determine the minimum clinically important difference (MCID) for the CFQ-R was discussed.
- The Division noted that Corus should be careful in Study 005 about investigator bias in the primary endpoint of time to need because different patients assess patients differently.
- The Division suggested that Corus concentrate on symptoms only (not findings) as they directly related to how a patient feels. It was agreed that the following 4 symptoms be used: decreased exercise tolerance, increased cough, increased sputum/chest congestion and decreased appetite.

September 13, 2004: Corus met with the Division to discuss revised Phase 3 designs. Gilead proposed to modify the program to include three Phase 3 studies:

- Study 004 (pivotal) consisted of 14 days of BID AZLI/placebo treatment with 14 days of follow-up. The primary endpoint was change in predefined CFQ-R domains at Day 28.
- Study 005 (supportive) consisted of a 14 day screening, 28 day course of TSI, 14 days of BID AZLI/placebo treatment with 14 days of follow-up. The primary endpoint was time to need for IV or inhaled anti-pseudomonal antibiotics following AZLI/placebo treatment.
- Study 006 (safety) was an open-label, follow-on study for patients from Study 004 and 005. Patients could receive up to six 14-day courses of AZLI over 12 months provided each was separated by at least 56 days.

November 23, 2004: End of Phase 2 meeting, Corus presented Phase 2 data, Phase 3 development plan and MCID study design.

- The Division was concerned that the Phase 2 study didn't achieve the objective of showing a difference between placebo and AZLI 225mg group in change in FEV_1 (the primary endpoint).
- Corus proposed changes to the Phase 3 designs based on the Phase 2 results. The main change provided for two dosing regimens (75mg BID or TID) to be given for a longer duration (28 rather than 14 days).
- The Division expressed concern that there is no break between TOBI and AZLI administration is Study 005.
- The Division commented supportive study (Study 005) could be considered as an alternate pivotal study but it would depend on the "robustness" of the 005 results and would be limited to the design/endpoints of the study.

September 13, 2005: Teleconference with Division to discuss the preliminary CFQ-R results from the Phase 2 and Phase 3 study for AZLI. The Division stated:

- It is acceptable to continue to pursue the respiratory domain of the CFQ-R
- Depending on the Sponsor's ability to demonstrate the MCID and change in FEV_1 between groups, the 007 study (formally study 004) could serve as a pivotal study.
- Evaluating different doses in the trials is acceptable, but the labeling recommendations would need to identify when a TID dose should be used. These labeling recommendations need to be based on the clinical trial data.

November 29, 2005: The Division reviewed additional data concerning the MCID results in Studies MCID-001 and 005 and discussed Phase 3 statistical analysis plans.

- An MCID of 5 was considered reasonable for adolescents/adults, ages 14 and above.
- The Division questioned the planned AUC analysis of the primary endpoint (FEV_1) in Study 007, preferring to use Day 42 FEV_1 as a single time point to demonstrate a sustained effect.
- The Division also recommended that CFQ-R be used as the primary endpoint as they are more confident it is clinically meaningful
- The Division would like to see a sensitivity analysis for the primary endpoint of Study 005. Censoring of missing data in the primary analysis may not be appropriate because the reason for withdrawal many not be random and could represent clinically significant worsening.

February 3, 2005: The Division provided Study 005 comments on the time of randomization, data collection tool for the clinician assessment of need for antibiotics, the need for a pre-specified statistical framework to control for the type I error rate and other issues. Key items included:

- The Division inquired about the interpretation of the Study 005 primary analysis.
- The Division asked the question "What if significance is found in the pooled AZLI group but is not found in the 75mg AZLI BID or the 75mg AZLI TID groups individually? "
- The Division requested sensitivity analyses for the primary analysis of time to need to handle potential biases from informative dropouts.

July 11, 2006: Corus met with Division and gave updates on the status of the Phase 3 studies and described the planned CFQ-R analyses of Study 007. For Study 007, the Division recommended against using average change at Day 14 and Day 28 from baseline for the CFQ-R respiratory domain score (primary endpoint).

August 2006: Corus was acquired by Gilead who continued the development of AZLI

February 14, 2007: Pre-NDA Meeting. The main discussion involved efficacy and safety results of Study 005.

- The Division stated the Sponsor needs to examine the issue of hospitalization imbalance in pediatric AZLI group.

- The Division emphasized in Study 005 the need for alternative strategies to LOCF for imputing missing observations, appropriate multiplicity adjustments for all secondary endpoints, unadjusted analyses and a CFQ-R analysis at Day 42
- The Division noted that claims related to improvement of respiratory symptoms should be supported by both studies and expressed concerns about the corroborative evidence from Study 007 for time to antibiotics use..

November 16, 2007: NDA 50814 for AZLI TID filed by Gilead

July 25, 2008: Teleconference. The purpose of this teleconference was to inform the Sponsor of the Division's concerns with the application. It was indicated that Study 005 was likely not going to be considered as one of the two positive trials supporting efficacy.

September 16, 2008: CR letter was issued by the Agency

August 28, 2008: Meeting with Gilead to discuss the review status of their application.
- In response to the Sponsor's question regarding one pivotal study as the basis for approval, the Division explained that the standard is more than one study except only with rare cases. Since there is one more drug available for the same treatment as the study product, the more appropriate number of studies should be more than one study.
- The Division stated that study 007 had a few design issues such as: 1) the unblinded interim analysis for sample size re-estimation which may have resulted in over-powering and over-enrollment and 2) differences in efficacy results based on age and geographic location of the study site which could have favored the treatment drug.
- The Division expressed concern over the validation and adequacy of the CFQ-R tool due to potential recall bias in measuring the primary endpoint in CP-AI-007 trial.
- The Division stated that the CP-AI-006 trial can not be viewed as a separate trial because it used the same patients as the other trials.
- The Division stated that they will continue with the review and will make a determination of our action and will be communicated to the Sponsor via an action letter at least by the PDUFA goal date.

October 2, 2008: Meeting to discuss the complete response letter, dated September 16.
- The Division noted for Study 005 that since placebo groups performed differently, it would be inappropriate to pool the data as was done. The Sponsor noted the reviewed protocol and SAP called for pooled comparisons. The Division responded sometimes it may not be appropriate to follow the SAP, as in this situation.
- The Division stated that unless the primary endpoint was met, any consideration of secondary endpoint data would not be appropriate or meaningful.
- The Division reiterated that the issue of Study 005 is the imbalance of results between placebo and treatment groups and this trial alone could not answer the reason for this imbalance. There was a larger difference among the placebo groups (placebo BID and Placebo TID) than the difference among treatment groups (AZLI BID and AZLI TID). Therefore, there is a need for an additional study.
- The Division pointed out that issues of Study 005 might be 1) timing of randomization relative to that actual start of AZLI/Placebo dosing, 2) in terms of

regimen effect, the BID regimen demonstrated better response than TID. The Sponsor also indicated the choice primary endpoint may have been problematic for this study.

- The Division was concerned about the robustness of Study 007 results as well as potential biases in the primary endpoint (CFQ-R RSS) due to the 14-day recall period. The current PRO guidance emphasizes a patient's current state without comparing the current state to an earlier time period or recalling an earlier time period.
- The Division had reservations about basing a claim on one study alone (Study 007) if primary endpoint evidence was derived from this less than adequate PRO tool.
- The Division believed that each of the Study 005 and 007 primary endpoints should be corroborated by the other study.
- The Division indicated options for the next studies which included repeating Study 005 correcting the problems with randomization, using an updated PRO tool and evaluating a 150mg BID dose.

November 24, 2008: Formal dispute resolution request (FDRR) to the Office of Antimicrobial Products. Gilead included responses to the Division's comments from the complete response letter.

December 22, 2008: The Division met with Gilead to discuss the NDA submission.

- The Agency mentioned that efficacy in Study 007 was not seen in patients with baseline FEV_1 below 50% of predicted and was attenuated in patients ≥ 18 years.
- The Agency asked what magnitude of change in FEV_1 is clinically meaningful. Gilead replied that it remains to be established; however physicians tend to regard 5% as adequate. A decrease of 10% will generally lead to a course of IV antibiotics. The MCID-like value for FEV_1 is probably less than 10%.
- The Agency asked whether Gilead had explored any factors that explained the imbalances between the placebo groups in Study 005. Gilead commented analyses adjusting for baseline imbalances did not account for the observed differences.
- The Agency asked about the longer-term clinical significance of a 28-day improvement in symptoms. Gilead responded they are only asking for an indication of improved symptoms and lung function based on 28 days of dosing.

March 13, 2009: Gilead's 2nd formal dispute resolution request (FDRR): FDRR to the Office of New Drugs (OND).

April 24, 2009: Meeting to discuss questions in Agency's April 10, 2009 letter to Gilead related to 1) the regimen effect and its influence on interpretations and pooling, 2) justification of TID vs. BID dosing over multiple cycles and 3) EMEA discussions. Gilead gave a presentation to address these issues. Gilead included analysis of "the rollercoaster" effect of AZLI on FEV_1 along with the response shift to explain the underperformance of AZLI TID with regard to time to need. Issues included for 1) were:

- Since placebo mean baseline values do not shift and there is no difference at Day 28 for secondary endpoints, placebo group difference in time to need are due to chance
- Efficacy results interpretable despite regimen effect, effect of AZLI adjusted for regimen (p=.0083, Cox PH Regression). No treatment by regimen interaction.

- Regimen effect is biologically implausible
- Time to Need is a surrogate for FEV_1. Classification tree analysis showed that physicians assessed time to need through FEV_1 changes: Decreased PFT: 80% use antibiotics, No decreased PFT: 31%.
- AZLI TID had largest increase and decline in FEV_1 from Day -28 baseline
- Response shift bias in time to need endpoint.
    - Response shift is worsening relative to post-treatment effects. The patient's perceived baseline is continually reset based on treatment effects
    - Treatments (e.g. AZLI TID) which induce a greater response are paradoxically subject to greater penalty, since when treatment is stopped, the decline in FEV_1 is more dramatic as the patient returns to their initial status
    - Time to need is measuring true pulmonary exacerbation, worsening of disease (Placebo) but loss of treatment benefit relative to baseline for (AZLI)
- AZLI patient with events had better pulmonary function vs. placebo in an analysis using LOCF with a Day -28 baseline

The Agency requested additional information regarding the observed regimen effect.

May 08, 2009: Gilead provided a preliminary response to questions from the April 24, 2009 meeting with OND. This included 1) information from CF trials that antibiotic use is triggered by FEV1 decline and 2) Analysis of the percent of treatment effect explained by FEV_1 in Study 005

May 21, 2009: Gilead provided a submission to further address the outstanding issues from the April 24, 2009 meeting. This included assessment of time to need as a function of FEV1 and further work to analyze "the rollercoaster" effect. Gilead also examined whether any aspect of FEV_1 changes may constitute a surrogate marker for time to need. According to the Applicant, Sensitivity analyses S_1 'Time to need: controlling for FEV_1 slope' and S_2 'Time to need: change 'event' to 'no-event' if FEV_1 improved from baseline' showed significantly delayed time to need in AZLI TID vs. Pooled Placebo.

June 17, 2009: In response to the 2nd FDRR with OND, the Agency recommended that Gilead submit new statistical analyses of data relevant to NDA 50814 and that the Division should present the full application to the AIDAC. It was also stated "what remains at issue is whether that exploratory work is robust enough to stand on its own as an explanation for the treatment regimen effect or it should simply be considered hypothesis generating. The decision can only be made on the basis of a full review of these new data analyses.

August 11, 2009. Gilead submitted a formal response to Agency's 9/16/08 CR letter

December 10, 2009: An anti-infective advisory committee (AIDAC) meeting was held to discuss the safety and efficacy of AZLI 75mg TID. The majority of AIDAC members voted in favor of AZLI 75mg TID as being safe and effective (15 to 2 votes).

3. EVALUATION OF EFFICACY (STUDY 005)

3.1 Review of Study 005

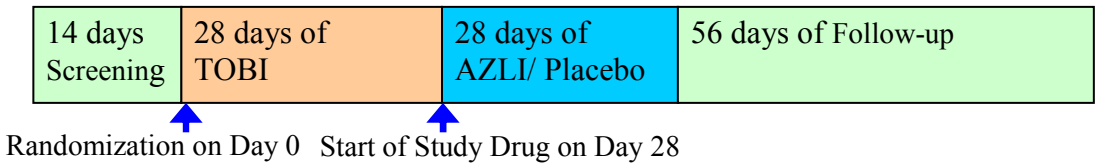
Study Design: Study 005 was a placebo controlled study in CF patients 6+ yrs with *PA* & FEV_1 25%-75% of predicted.

Study Objective: The study objective was to demonstrate efficacy & safety of AZLI (75 mg BID or TID).

In comparison to Study 007, Study 005 patients were more extensively treated:

- Had to have ≥ 3 TSI courses in prior 12 months.
- Could continue use of chronic azithromycin therapy if no change in regimen in the previous months and patient had a need for TSI and/or additional antipseudomonal therapy since initiation of azithromycin.
- Could use macrolide or antipseudomonal antibiotics within 28 days of study drug

Figure 1: Study 005 Design



Study Arms: Study 005 included 4 study arms: AZLI BID, AZLI TID, Placebo BID, Placebo TID. Randomization was 2:2:1:1 at Day -28 prior to start of study drug at Day 0.

- N=246 randomized at Day -28 (82 per AZLI arm, 41 per placebo arm)
- N=211 ITT patients receiving AZLI/Placebo at Day 0 (69 in AZLI BID, 66 in AZLI TID and 38 per placebo arm)

Hypotheses Tested: Study 005 tested three primary hypotheses. $H_{0Pooled}$ served as the gatekeeper hypothesis, if $H_{0Pooled}$ was rejected then H_{0BID} & H_{0TID} were then tested in parallel without multiplicity adjustments.

- $H_{0Pooled}$: Pooled AZLI = Pooled Placebo
- H_{0BID}: AZLI BID = Pooled placebo, H_{0TID}: AZLI TID = Pooled placebo

Primary endpoint: Time to need for IV/inhaled antibiotics due to one or more pre-defined symptom(s) from Day 0-84

- Decreased exercise tolerance
- Increased cough
- Increased sputum/chest congestion
- Decreased appetite

Key secondary endpoints:

- Changes in CFQ-R respiratory symptoms
- Changes in FEV_1(L)
- Change in log 10 *PA* CFU

Other secondary endpoints included:

- Other CFQ-R (e.g. non-respiratory)
- Other spirometry (e.g. changes in FVC and FEF)
- Hospitalization
- Number of school/work days missed
- % Change in Weight/BMI

For more detailed information regarding Study Design and Endpoints, Subject Disposition, Demographic and Baseline Characteristics and Statistical Methodologies refer to our earlier statistical review of the Study 005 protocol (Sections 3.1.1-3.1.3).

3.2 Uncertainty/Subjectivity in Primary Analysis

There is a high degree of uncertainty/subjectivity in the primary analysis due to the nature of the time to need endpoint and the manner in which it is measured. The time to need endpoint aims to assess patient differences in pulmonary exacerbations as predicted by the need for IV/inhaled antibiotics due to predefined symptoms. However, there is uncertainty in that symptoms may not be reliably reported by the patient or assessed by the investigator which may lead to inappropriate use of antibiotics. There is further uncertainty as to whether antibiotics usage would have been predictive of a pulmonary exacerbation even if prescribed correctly.

In the Study 005 primary analysis, there are further uncertainties. For example, some prescribing physicians of antibiotics in patients with time to need events were not trained trial investigators. There was also uncertainty as to which patients should be included as having an event. Based upon a review of prescriptions for IV or inhaled antibiotics in the two weeks following trial termination or completion, exacerbations were identified in several patients, and these patients were classified and meeting the primary endpoint. This included 11 patients with time to need events reported after termination from the study (2 Placebo BID, 3 Placebo TID, 3 AZLI BID, 3 AZLI TID) and 5 cases reported after nominal 84 study period (1 Placebo TID, 2 AZLI BID, 2 AZLI TID).

In addition there is uncertainty related to missing data and strong regimen effects as discussed further in sections 3.3 and 3.4. For these reasons, conclusions based only on time to need for IV or inhaled antibiotics may not be reliable and consideration of several sensitivity analyses is necessary. However, regardless of whether sensitivity analyses define an 'event' using a very broad definition such as early termination or a very strict definition such a hospitalization, AZLI TID patients fared substantially worse than Placebo BID patients, with p-values approaching significance (p=.085 for time to hospitalization and p=.0941 for time to early termination in randomized patients).

3.3 Missing Data

High rates of missing data led to difficulties with interpreting study evidence, especially since the missing data were dependent in nature and based on both the drug (AZLI or placebo) and regimen (BID or TID) assignment. Early termination rates were significantly higher in patients randomized to AZLI vs. Placebo arms (Day -28 to 0, TSI run-in period) at 17.9% (AZLI) vs. 7.3% (Placebo), p=.026 and significantly higher in TID vs. BID arms (Day 0 to 84) at 69.2% (TID) vs. 50.5% (BID), p=.005. Early termination rates in randomized patients for the four treatment arms were as follows: Placebo BID (56.1%), AZLI BID (57.3%), Placebo TID (80.5%) and AZLI TID (70.7%).

These trends in dropouts would favor AZLI TID if dropouts are sicker than those remaining in the study. Study evidence suggests dropouts may be sicker since 88% of dropouts were due to adverse events or treatment intolerance. Based on the observed dropout patterns, there is a potential for substantial biases in comparisons of AZLI TID vs. placebo.

Section 3.3.1 considered early termination rates and time to early termination in ITT and all randomized patients in order to evaluate potential study biases due to missing data. Section 3.3.2 addressed missing data specifically in time to event analyses and considered only those patients who terminated early without the event (i.e. censored early). Note that those patients terminating early with the event were not considered because they were already counted in the time to event analysis and would not introduce potential biases. Early termination rates in patients without primary analysis time to need events were highest in AZLI TID patients: Placebo BID (37.9%), AZLI BID (44.4%), Placebo TID (46.7%) and AZLI TID (58.6%).

In Section 3.3.3, the reviewer then describes various sensitivity analyses and imputation strategies which were conducted to account for missing data in the three types of endpoints reported in Study 005: 1) ‘time to event’, 2) ‘event-free rate’ and 3) ‘change from baseline’ endpoints.

3.3.1 Early Termination Rates

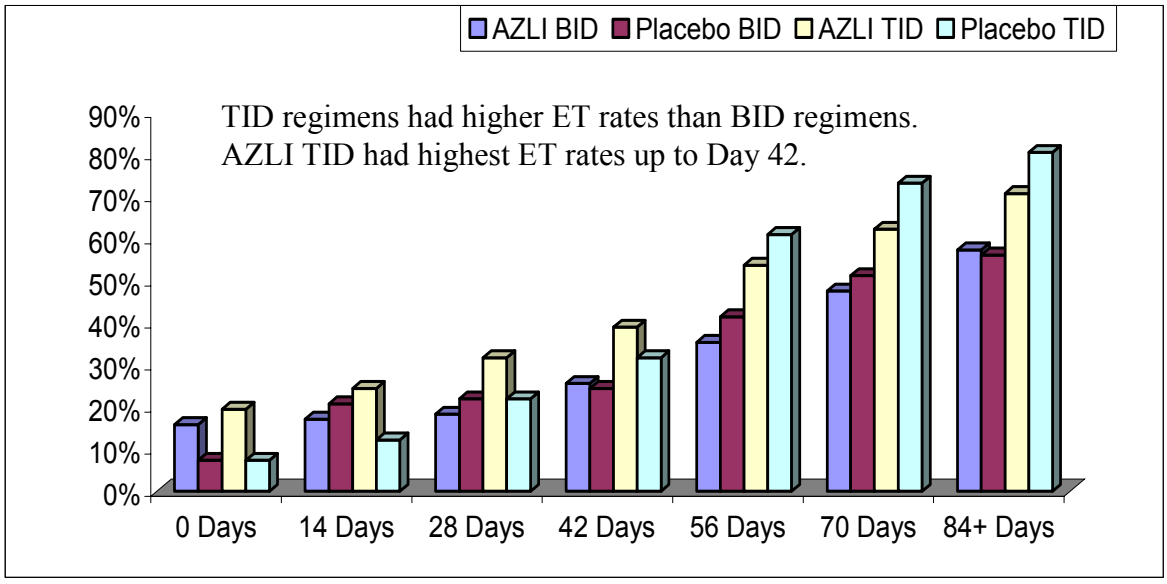
Figures 2-4 show early termination rates and time to early termination in randomized subjects. (Corresponding figures in ITT subjects are provided in the Appendix, Figures 19-21). Figures 2-4 show higher early termination rates and shorter time to early termination in TID vs. BID regimens as well as for AZLI TID vs. AZLI BID and AZLI TID vs. Placebo BID comparisons. These results indicate potential biases in treatment group comparisons for change from baseline endpoints favoring TID vs. BID regimens as well as AZLI TID.

Figure 2 shows a trend towards higher rates of early termination in TID vs. BID regimens by Day 28 that becomes more pronounced over time. Early termination was also highest in AZLI TID patients from 0 to 42 days after start of study therapy. Figures 3-4 show

shorter time to early termination in patients randomized to TID vs. BID regimens: TID vs. BID (p=.0030), AZLI TID vs. AZLI BID (p=.0260) and AZLI TID vs. Placebo BID (p=.0941). Additionally, AZLI TID patients had shorter time to early termination than Placebo TID patients up to 45 days after start of study therapy. Analyses considering only patients terminating due to an AE or treatment intolerance show similar or stronger trends of shorter time to early termination in TID regimens: TID vs. BID (p=.0003), AZLI TID vs. AZLI BID (p=.0059) and AZLI TID vs. Placebo BID (p=.1022) as shown in Figures 22-23 of the Appendix.

As noted earlier, time to event endpoints (e.g. time to need, time to first hospitalization) are not necessarily biased by TID vs. BID differences in ET. In time to event endpoints, only differences in ET rates among those patients not reporting an event may lead to potential biases. Missing data in time to event endpoints is discussed in Section 3.3.3.

Figure 2: Early Termination Rates (All Randomized Patients)



Early Termination Rates (Days After Start of Study Therapy)

Figure 3: Time to Early Termination by Regimen (All Randomized)

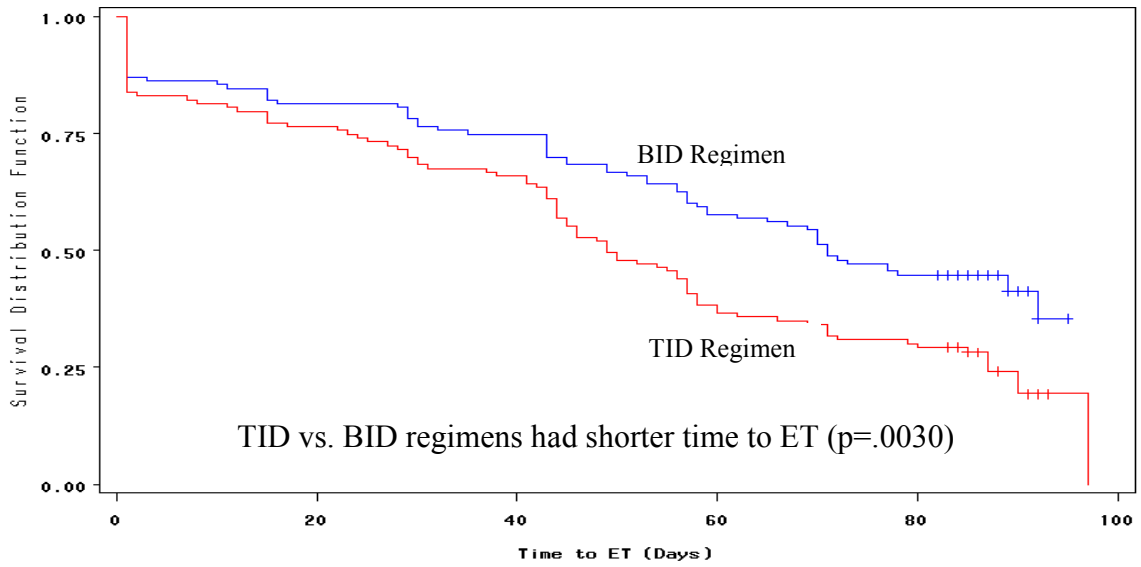
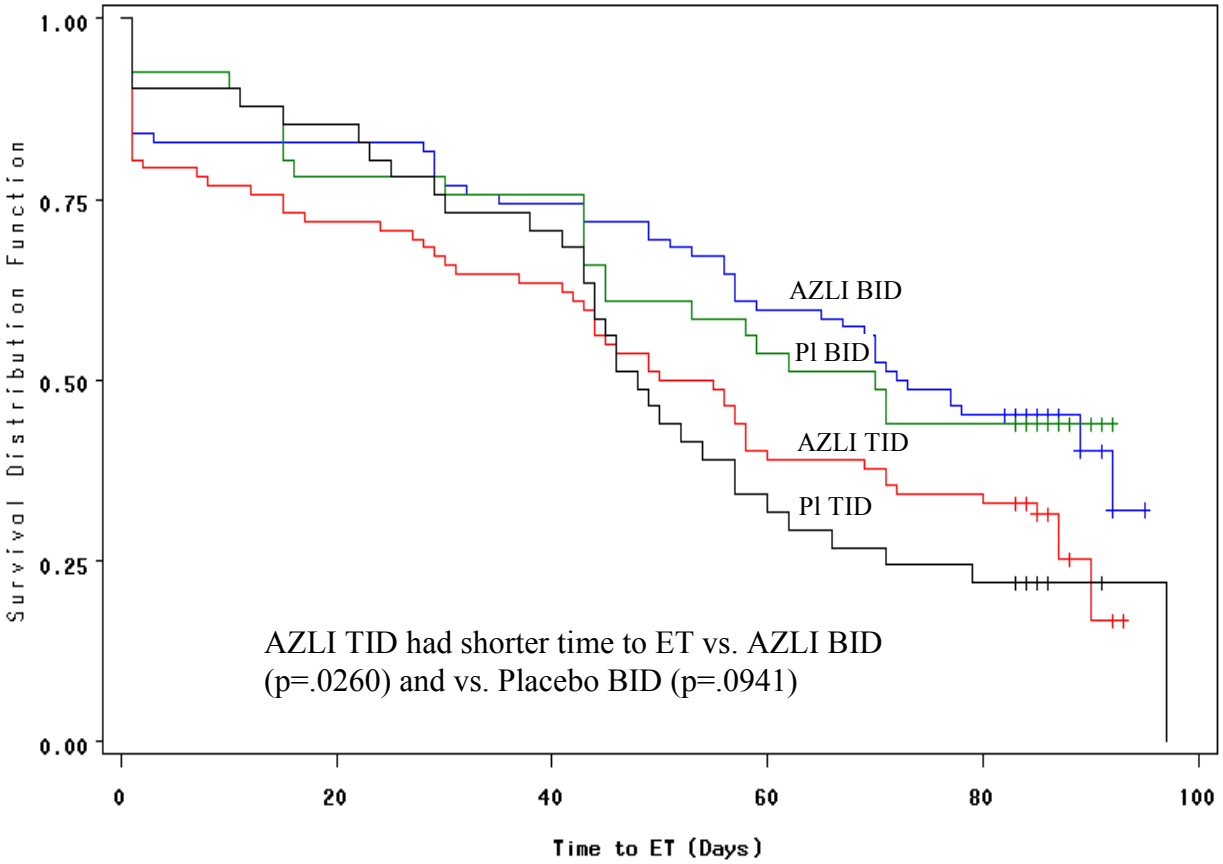


Figure 4: Time to Early Termination (All Randomized)



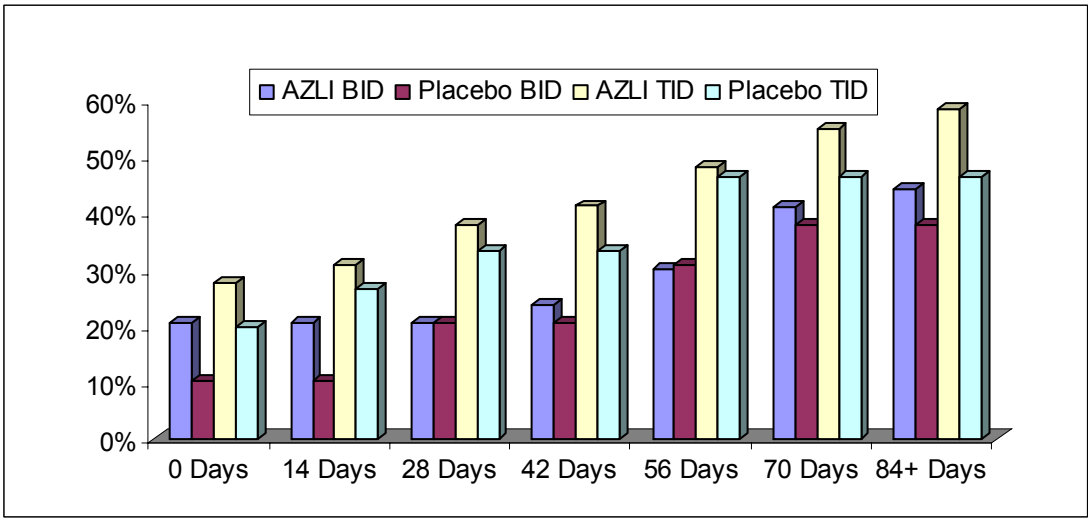
3.3.2 Early Termination Rates (Patients without Events)

As discussed earlier, time to event endpoints such as the primary endpoint are not necessarily biased by differences in BID vs. TID early termination. Differences in early termination rates among patients not reporting an event, however, may lead to potential biases. This section considers early termination in patients without events for two 'time to event' endpoints of interest: 1) time to need for IV/inhaled antibiotics due to pre-defined symptoms (primary endpoint) and 2) time to first hospitalization.

Time to need for IV/inhaled antibiotics due to pre-defined symptoms (primary endpoint)

Figure 5 compares early termination rates in randomized patients without a primary analysis time to need event. (The corresponding figure in ITT patients is provided in the Appendix, Figure 24.) Figure 5 shows there is still a trend towards higher early termination in the TID regimens, though less pronounced than when including all dropouts (Figure 2). Figure 5 also shows that early termination rates were highest for patients randomized to AZLI TID at all time points. This suggests potential biases in the time to need analysis favoring AZLI TID. Early termination rates in patients without primary analysis time to need events were as follows: Placebo BID (37.9%), AZLI BID (44.4%), Placebo TID (46.7%) and AZLI TID (58.6%).

Figure 5: ET Rates in Patients without a 1° Analysis Event (All Randomized)



Early Termination Rates (Days after Start of Study Therapy)

Figures 6-7 consider time to early termination in randomized patients without time to need events in BID vs. TID regimens and among treatments. (Corresponding figures for ITT patients are provided in the Appendix, Figures 25-26). Figures 6-7 show potential biases in the primary analysis for BID vs. TID regimens including AZLI TID. These biases can occur because among patients censored early (i.e. without primary analysis events), those randomized to a TID vs. BID regimen had a shorter time to early termination and may be sicker. Potential biases in the primary analysis were not as severe as suggested by comparisons of early termination rates in all patients since many patients terminating early were already counted in the primary analysis as having an event.

Figure 6 compares randomized patients without primary analysis time to need events by regimen (i.e. TID vs. BID) and shows significantly shorter time to early termination in TID patients (p=.0438). Figure 7 compares randomized patients by treatment and shows shorter time to early termination in AZLI TID vs. AZLI BID (p= .0713) & AZLI TID vs. Placebo BID (p=.0631) comparisons. Analyses of patients terminating early showed similar or stronger trends of shorter time to early termination due to an AE or treatment intolerance in TID regimens: TID vs. BID (p=. 0090), AZLI TID vs. AZLI BID (p= .0189) and AZLI TID vs. Placebo BID (p= .0672) as shown in Figures 27-28 of the Appendix. These findings indicate biases may exist in time to need analyses favoring TID vs. BID regimens including the AZLI TID treatment

Figure 6: Time to ET in Patients without 1° Analysis Event by Regimen (All Randomized)

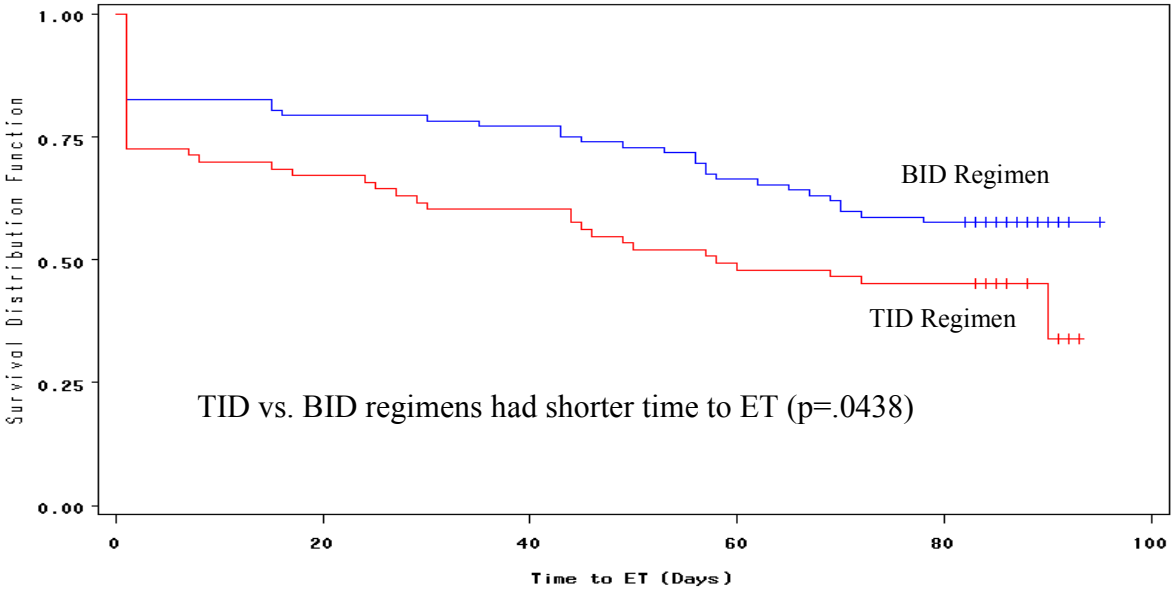
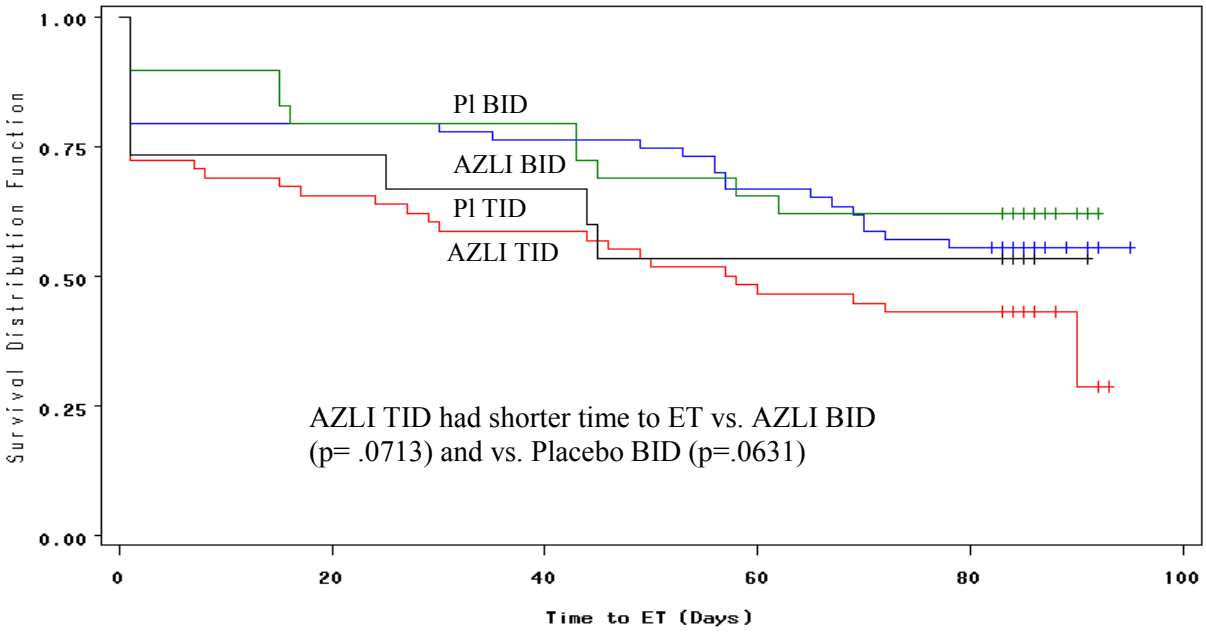


Figure 7: Time to ET in Patients without 1° Analysis Event (All Randomized)



Time to first hospitalization

Compared to the primary analysis of time to need, potential biases in time to first hospitalization analysis were more severe. There were more dropouts without a hospitalization event vs. dropouts without a time to need event. This led to more pronounced TID vs. BID regimen differences. Due to the small number of dropouts with

a primary analysis hospitalization event (n=17), regimen differences in those dropping out without a hospitalization event were similar to regimen differences among all dropouts.

Comparisons in randomized patients without a hospitalization event during the AZLI/Placebo period showed shorter time to early termination in those using a TID vs. BID regimen: TID vs. BID (p= .0079), AZLI TID vs. AZLI BID (p=.0533) and for AZLI TID vs. Placebo BID (p =.1667) as shown in the Appendix in Figures 29-30. (Corresponding analyses in ITT patients shown in Figures 31-32). Analyses of patients terminating due to an AE or treatment intolerance showed stronger trends of shorter time to early termination in TID regimens: TID vs. BID (p=. 0007), AZLI TID vs. AZLI BID (p= .0083) and AZLI TID vs. Placebo BID (p= .1471) as shown in Figures 33-34 of the Appendix. These findings indicate biases may exist in time to first hospitalization analyses favoring TID vs. BID regimens.

3.3.3 Strategies Used for Controlling Biases due to Missing Data

As mentioned previously, imputing for missing data presented challenges not only due to the high overall rate of missing data but due to the dependent nature of the missing data on regimen (BID or TID) or drug (AZLI or Placebo) received. Further limitations arose from the fact that much of the missing data in the AZLI arms occurred from Day -28 to Day 0 which was prior to the start of study drug on Day 0. Imputing missing data from this period is unclear because similar rates of missing data may not have resulted if AZLI rather than TSI were taken from Day -28 to 0.

To control for biases due to missing data, various sensitivity analyses and imputation strategies were conducted in the three types of endpoints reported in Study 005: 1) 'time to event', 2) 'event-free rate' and 3) 'change from baseline' endpoints.

- In time to event and event-free rate endpoints, potential biases resulted only from those dropouts without events who were censored early as not having the event. This may be problematic since dropouts may have a less favorable time to event upon censoring as those continuing in the study. In the case of event-free rates, there is an additional bias related to the dropout's shorter risk period of an event compared to those continuing in the study. To assess these potential biases in these endpoints, the Reviewer conducted sensitivity analyses which considered 1) dropouts excluded or counted as failures, 2) various definitions of 'event', 3) various populations.

- In change from baseline endpoints, potential biases may result since data using last observation carried forward (LOCF) imputation or data using only observed cases were considered. Observed cases may favor AZLI TID by excluding relatively more dropouts in the AZLI TID arm. LOCF, however, may be especially problematic because it assumes AZLI TID dropouts could have maintained their effects over time. Observed AZLI TID patients generally did not maintain their effects over time even against placebo (especially for FEV_1 from Day 28 to Day 42) and therefore LOCF

may favor AZLI TID even if dropouts are not sicker. To minimize potential biases from more TID or AZLI dropouts in 'change from baseline' endpoints, the Reviewer conducted sensitivity analyses in which the Placebo BID group mean was used to impute missing data from BID patients and the Placebo TID group mean was used to impute missing data from TID patients.

3.4 Regimen Effects

Compounded by the uncertainties created from missing data in the primary analysis, there were also strong regimen effects which made interpretations in the primary analysis unclear. Although a significant AZLI drug effect was observed in the Pooled AZLI vs. Pooled Placebo analysis (p=.0070), a detrimental TID regimen effect was also observed in the Pooled TID vs. Pooled BID analysis (p=.0012). This may support a benefit from the AZLI drug, but does not clearly indicate a benefit from the AZLI TID treatment which requires increased dosing via a TID vs. BID regimen. In fact, the primary analysis suggests that any potential gains from use of 225mg AZLI drug would be more than offset by the losses from use of a TID regimen.

Due to the presence of regimen effects, comparisons based on pooled estimates across the BID and TID regimens could not be clearly interpreted, especially in drawing inferences regarding AZLI TID treatment which requires use of a significantly less favorable TID vs. BID regimen. To analyze primary analysis results, only comparisons based on individual treatment groups and not pooled treatment groups could potentially be considered as meaningful. This included the AZLI TID vs. Placebo TID and AZLI TID vs. Placebo BID comparisons. However, the AZLI TID vs. Placebo BID comparison is considered the most appropriate comparison because it fully accounts for the likely presence of a detrimental TID vs. BID regimen effect.

Section 3.4.1 provides a more detailed analysis of the regimen effect observed in the primary analysis. Comparisons of time to need for IV/inhaled antibiotics and event-free rates are made among possible individual vs. individual, individual vs. pooled, pooled vs. pooled treatment arms. Section 3.4.2 illustrates the robustness of the observed regimen effect. These findings confirm the presence of a regimen effect in the primary analysis and support the consideration of more conservative comparisons which fully take into the regimen effect (e.g. AZLI TID vs. Placebo BID). Section 3.4.3 discusses the TID vs. BID regimen in a general context with consideration to all study endpoints considered. This section explores the question of whether the regimen effect occurs in only the primary analysis and whether it is likely to be a real study effect vs. a chance finding.

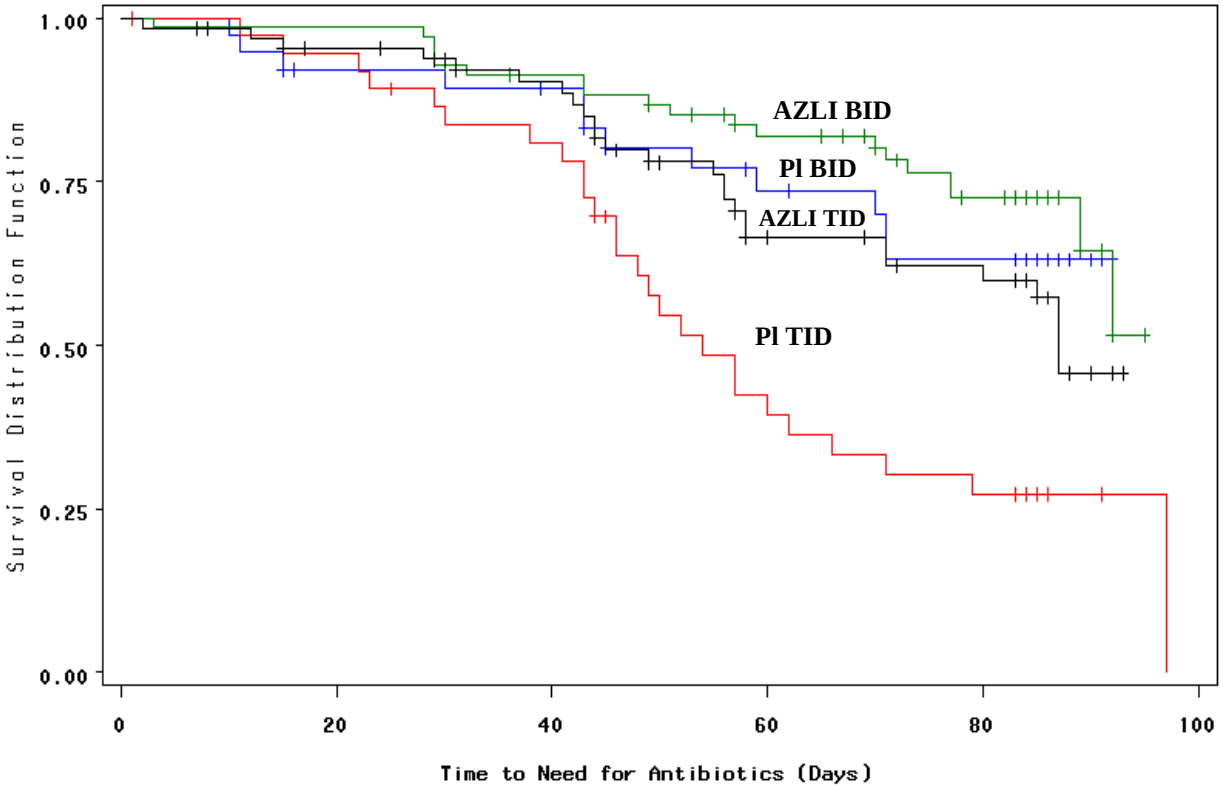
3.4.1 Regimen Effects in the Primary Analysis

Time to Need for IV/Inhaled Antibiotics

The primary analysis showed that the most influential factor on time to need for IV/inhaled antibiotics most was the regimen used, BID vs. TID. This factor was even more influential than AZLI vs. Placebo drug which was expected to be the only factor affecting time to need. The effect of regimen use had been assumed to be negligible or non-existent in designing Study 005, however this assumption was clearly not met. Due

to the detrimental TID regimen effect, the pre-specified Applicant analyses were not considered to be appropriate in drawing meaningful inferences regarding the efficacy of AZLI TID. The regimen effect is illustrated in the figures below.

Figure 8: Primary Analysis of Time to Need for IV/Inhaled Antibiotics due to Pre-defined Symptoms (ITT)



In the above figure, the treatment arms with highest survival (i.e. not needing IV/inhaled study antibiotics) were the treatments using BID regimens regardless of the drug received (i.e. AZLI vs. placebo). Placebo TID patients had significantly lower survival compared to the other three treatment arms. From Day 42 to Day 56, there were steep drops in survival in the TID treatment arms (esp. Placebo TID) but not the BID treatment arms. Although patients in the AZLI TID arm fared significantly better than patients in the Placebo TID arm, they fared worse in comparison to patients in the Placebo BID arm due to a substantial regimen effect (i.e. Placebo TID – Placebo BID). Since we do not know whether Placebo BID or Placebo TID is providing the more appropriate estimate of the true placebo rate, the efficacy of AZLI TID compared to placebo is uncertain.

Table 2: Primary Analysis, P values of Treatment Comparisons (ITT)

AZLI vs. *Placebo* (p value)	***Placebo BID***	***Placebo TID***	***Pooled Placebo***
AZLI BID	.4269	< .0001	.0019
AZLI TID	.5377	.0043	.1816
Pooled AZLI	.9240	< .0001	.0070
Regimen Effect	**AZLI**	**Placebo**	**Pooled**
BID vs. TID	.0835	.0043	.0012

Table 2 shows the p-values of various comparisons in the primary analysis. Based on this table, there is a significant Pooled AZLI vs. Pooled Placebo drug effect p=.0070 but a more significant detrimental Pooled TID vs. Pooled BID regimen effect p=.0012. The TID regimen effect was significant in the placebo arms (Placebo TID vs. Placebo, p= .0043) and trending towards significance in the AZLI arms (AZLI TID vs. AZLI BID, p=.0835). Note that the AZLI TID vs. AZLI BID comparison likely provides an overly conservative estimate of the regimen effect since patients on a TID vs. BID regimen had the benefit of receiving more AZLI drug (225 mg AZLI vs. 150 mg AZLI daily).

In the Applicant's three pre-specified primary analysis comparisons of AZLI arm(s) against Pooled Placebo, Pooled AZLI vs. Pooled Placebo and AZLI TID vs. Pooled Placebo comparisons were not considered appropriate for drawing inferences regarding the efficacy of AZLI TID because they fail to fully account for the TID regimen effect. The only comparison that fully accounts for the TID regimen effect is the AZLI TID vs. Placebo BID comparison which is estimating the combined effect of both the AZLI TID drug (i.e. 225mg daily) and use of the TID vs. BID regimen. This is considered to be least biased estimate of the AZLI TID treatment effect. The AZLI TID vs. Placebo BID comparison favored Placebo BID with a p value = .5377 and may also include biases favoring AZLI TID due to the higher rates of missing data in AZLI TID vs. Placebo BID patients as described in Section 3.3.2.

The AZLI BID vs. Placebo BID comparison was considered an appropriate comparison for assessing the efficacy of AZLI BID but failed to show a significant benefit from use of AZLI BID. This finding would similarly imply no significant benefit in AZLI TID which was substantially less favorable compared to AZLI BID (p=.0835) as given in Table 2. AZLI TID patients also generally fared worse than AZLI BID patients in most analyses and had significantly higher early termination rates.

Event-Free Rates in Primary Analysis

Unlike time to event comparisons, comparisons of event-free rates (i.e. rates of patients not needing IV/inhaled antibiotics) do not take into account the time at which the event occurred, only whether or not the event occurred (i.e. binary response). Event-free rate comparisons may allow for a clearer interpretation of treatment benefits and regimen effects. Since there were no strong trends related to the treatment and the timing of event (i.e. early vs. late), comparisons among treatment arms based on time to event and event-free rates were generally similar.

Figure 9: Event Free Rates in Primary Analysis (ITT)

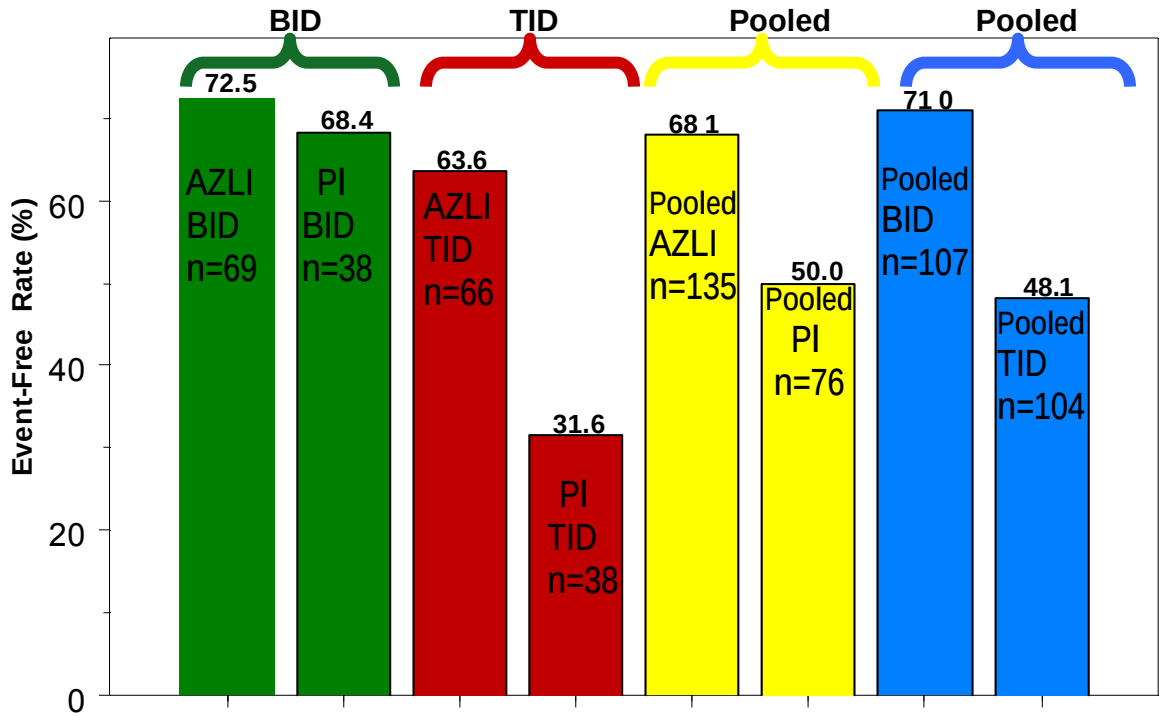
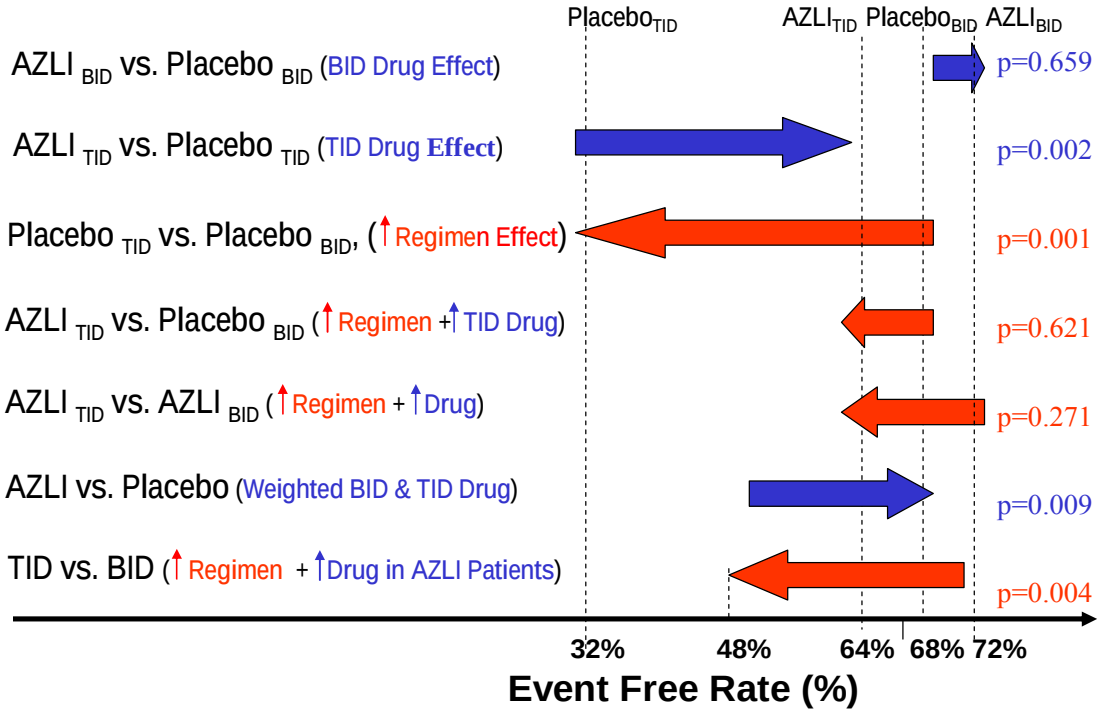


Figure 9 further illustrates the regimen effect in event-free rates. There is a substantial AZLI benefit over placebo among patients on a TID regimen (32.0%, 63.6% vs. 31.6%) but not on a BID regimen (4.1%, 72.5% vs. 68.4%). Patients on AZLI TID also fared nearly 5% worse than those Placebo BID (63.6% vs. 68.4%) in event-free rates. As with the time to need, the most influential factor on event-free rates is the regimen used (BID or TID) and not the drug used (AZLI or Placebo). This can be observed by the

comparison of Pooled AZLI vs. Pooled Placebo (18.1% difference, 68.1% vs. 50.0%, p=.0009) which is not as significant as the comparison of Pooled TID vs. Pooled BID (-22.9% difference, 48.1% vs. 71.0%, p=0.004).

Since the proposed treatment of AZLI TID requires use of a TID regimen, it is highly problematic that the regimen effect (TID vs. BID) is the most influential factor and detrimental towards TID regimens. The presence of an AZLI TID treatment benefit which includes both a positive effect from use of 225mg AZLI drug (daily) as well as a detrimental effect from an increase in regimen from BID to TID has not been clearly shown. Furthermore, based on primary analysis data, it appears likely that patients taking AZLI TID would be more likely to have less favorable AZLI TID effects relative to placebo, being influenced more by the detrimental effects of the TID vs. BID regimen than by the positive effects of the AZLI drug. The following figure further illustrates the effects of the regimen, drug and treatment on event-free rates.

Figure 10: Event-Free Rates by Drug & Regimen Effects (ITT)



In Figure 10, we see a minimal BID drug effect (i.e. 150 mg AZLI daily), a significant TID drug effect (i.e. 225 mg AZLI daily) at p=0.002 but a more significant detrimental TID regimen effect (Placebo TID vs. Placebo BID) at p=0.001. Patients taking AZLI TID vs. Placebo BID were influenced by a combination of two effects, increased BID to TID regimen (negative effect) and 225mg daily of AZLI drug (positive effect). Patients taking AZLI TID vs. Placebo BID lost nearly 5% in event-free rates from this combination of effects indicating the detrimental effects from increasing regimen outweighed the benefits of the AZLI drug. Patients taking AZLI TID vs. AZLI BID were influenced by the combined effect of increasing AZLI drug from BID to TID (i.e.

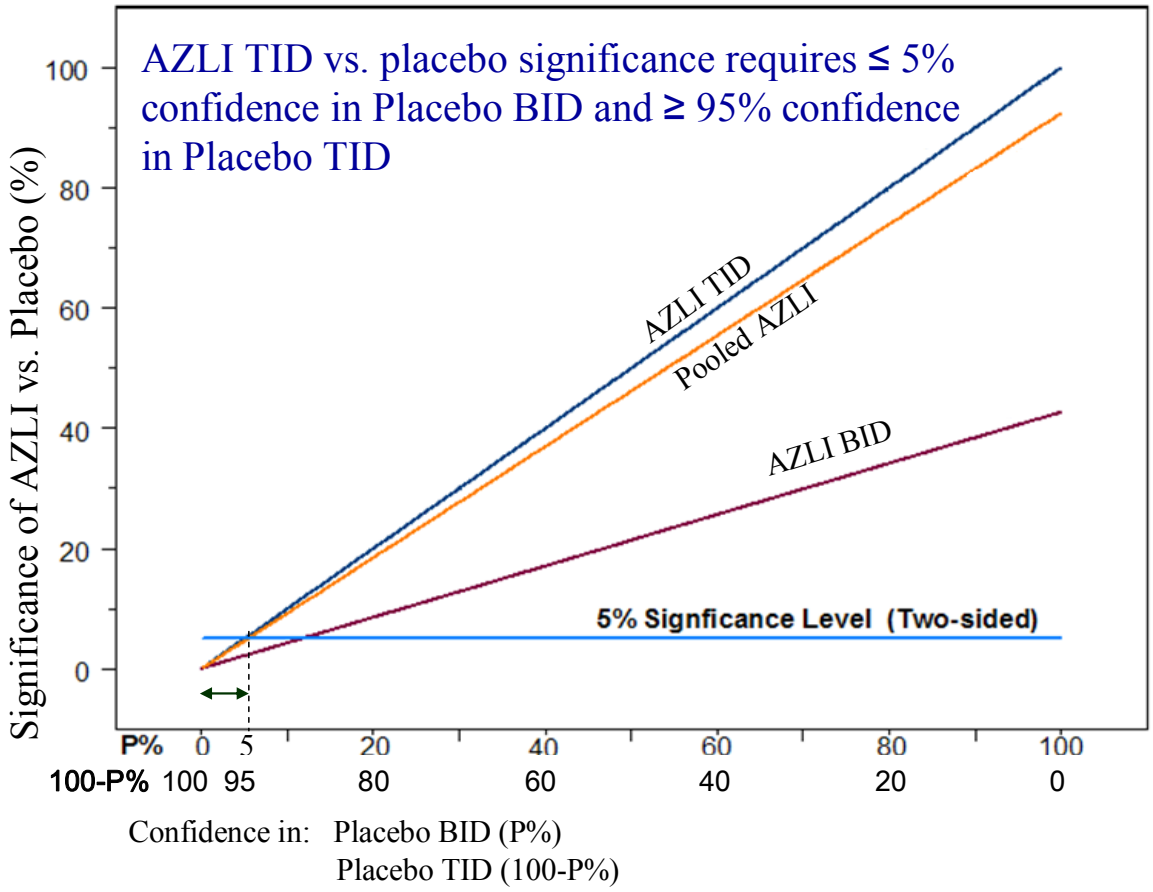
150 mg to 225mg) and increasing the regimen from BID to TID. These patients lost 9% in event-free rates indicating the detrimental effects from TID vs. BID regimen outweighed the benefits of TID vs. BID AZLI (150mg vs. 225mg daily). Pooled comparisons of AZLI vs. Placebo and TID vs. BID further show regimen use as the most influential factor on event-free rates.

Assay Sensitivity

Assay sensitivity relates to the ability of a study drug to distinguish itself from an ineffective drug. It is not clear that AZLI TID has distinguished itself from placebo in the primary analysis especially if a regimen effect is assumed. In Study 005, it is not clear what the true placebo rate since estimates in Placebo BID and Placebo TID patients were not consistent, AZLI TID vs. Placebo BID ($p = 0.9999$) and AZLI TID vs. Placebo TID ($p = 0.0043$). In order to conclude a significant AZLI TID benefit, a high degree of confidence in the Placebo TID vs. Placebo BID estimate would be needed. Since we do not have such confidence, we cannot conclude efficacy from AZLI TID. Figure 11 shows that under the assumption that either Placebo BID or Placebo TID estimates the true placebo rate, $\geq$ 95% confidence in the Placebo TID estimate (or $\leq$ 5% confidence in the Placebo BID estimate) would be required to show a significant AZLI TID benefit. We do not have such a confidence in Placebo TID.

It should be further noted that Placebo BID does not necessarily provide a conservative estimate of the true placebo rate. Due to Study 005 primary analysis findings which suggest possible detrimental effects associated with more frequent dosing (e.g TID vs. BID), there may be detrimental effects associated with using a BID regimen vs. no regimen, even though those detrimental effects are smaller than those associated with using a TID regimen vs. no regimen.

Figure 11: Significance of AZLI vs. Placebo Assuming P% Confidence in Placebo BID vs. 100-P% in Placebo TID (ITT)



3.4.2 Robustness of Regimen Effects in the Primary Analysis

The robustness of regimen effects in the primary analysis is an important consideration in assessing evidence from the primary analysis. If the regimen effect is not robust, it may be more likely the result of a spurious or chance finding. If the regimen effect is robust, it is likely a real study effect that must be taken into account in assessing the efficacy of AZLI TID. The robustness of the AZLI TID vs. Placebo BID comparison is also of interest, if AZLI TID is in fact an effective treatment, we would expect AZLI TID to become more favorable vs. Placebo BID in the sensitivity analyses and trend towards significance (i.e. an AZLI TID benefit).

To demonstrate the robustness of primary analysis results as well as address Agency concerns of potential biases from informative censoring, the Applicant had pre-specified three sensitivity analyses (S1, S2, S3) conducted in a hierarchical fashion with loosening restrictions of the definition of 'event' in the primary analysis. The Reviewer further considered a post-hoc comparison similar to S3 except including all randomized subjects

instead of only ITT subjects. S4 was conducted to control for possible biases resulting from the higher pre-treatment dropout rate in AZLI vs. placebo patients. Primary (1°) and sensitivity analyses (S1-S4) are defined below and treatment comparisons for these analyses are provided in Table 3.

- 1° - Time to inhaled/IV antipsuedomonal antibiotics due to pre-defined symptoms
- S1 - Time to inhaled/IV antipsuedomonal antibiotics (all symptoms)
- S2 - Time to specified antipsuedomonal antibiotics (all routes of administration)
- S3 - Time to early termination (all reasons for termination in ITT subjects)
- S4 - Time to early termination (all reasons for termination in all randomized subjects)

Table 3: Comparisons in Primary (1°) & Sensitivity Analyses (S1- S4) (ITT)

AZLI vs. Placebo (p value)		Placebo BID	Placebo TID	Pooled Placebo
AZLI BID	(1°)	.4269	< .0001	.0019
	(S1)	.6400	< .0001	.0078
	(S2)	.5017	.0004	.0178
	(S3)	.4237	.0003	.0122
	(S4)	.9205	.0082	.1152
AZLI TID	(1°)	.5377	.0043	.1816
	(S1)	.5184	.0069	.2254
	(S2)	.4655	.0800	.5806
	(S3)	.3217	.1367	.7923
	(S4)	.0941	.5629	.4725
Pooled AZLI	(1°)	.9240	< .0001	.0070
	(S1)	.9124	< .0001	.0198
	(S2)	.9635	.0026	.0801
	(S3)	.9203	.0036	.0947
	(S4)	.3973	.0749	.6029
		AZLI:	**Placebo:**	**Pooled:**
BID vs. TID	(1°)	.0835	.0043	.0012
	(S1)	.1317	.0066	.0031
	(S2)	.0576	.0517	.0055
	(S3)	.0268	.0431	.0023
	(S4)	.0260	.0557	.0030

*** P values in red are for testing a Placebo vs. AZLI benefit since comparison favored Placebo.**

Table 3 shows that treatments using a TID vs. BID regimen performed relatively worse under broader definitions of event. This trend was due to larger numbers of AZLI TID dropouts without events who were now counted as having an event in an analysis using a broader definition of 'event'. S4 also shows that AZLI vs. Placebo treatments performed relatively worse under a broader definition of analysis population (i.e. all randomized vs. ITT patients). This is due to more AZLI vs. Placebo dropouts prior to AZLI/Placebo treatment that were excluded in the ITT analysis but included in the all randomized analysis as having an 'early termination' event.

In assessing potential regimen effects, Pooled TID vs. Pooled BID comparisons remained highly significant regardless of the definition used for event. Findings were generally consistent in both Placebo and AZLI patient groups. These comparisons were consistent with the assumption of a TID vs. BID regimen effect in the primary analysis. In contrast, Pooled AZLI vs. Pooled Placebo comparisons were not significant in the majority of the sensitivity analyses and trended away from significance in S4 after controlling for higher rates of AZLI vs. Placebo and TID vs. BID dropouts (p=.6029).

As previously stated, due to the strong evidence of a regimen effect, AZLI TID vs. Placebo BID was considered the most meaningful comparison since it controlled for the TID vs. BID regimen effect. For the primary endpoint of time to need, the AZLI TID vs. Placebo BID comparison favored Placebo BID (p=.5377). Since this comparison may include biases from the relatively high rates of missing data in AZLI TID patients (without time to need events), further sensitivity analyses should also be considered. After controlling for biases related to more TID vs. BID dropouts in the ITT analysis using the broadest definition of event (i.e. early termination) as in S3, comparisons became more favorable to Placebo BID (p=.3217) . After further controlling for more AZLI vs. Placebo dropouts prior to AZLI/Placebo treatment as in S4, comparisons became even more favorable to Placebo BID and approached significance (p=.0941).

3.4.3 Regimen Effects in Study 005 Analyses

Sensitivity Analyses

Sensitivity analyses showed that the regimen effects observed in the primary analysis were not unique to the time to need endpoint since the same trends favoring BID vs. TID were also observed for broader definitions of events such as early termination. The AZLI vs. Placebo drug benefit did appear to be highly dependent upon the definition of event and population considered. This tells us that the regimen used was a more real and robust study effect than the AZLI drug effect. This raises serious concerns in evaluating treatment efficacy because AZLI TID requires use of an adverse TID regimen and whatever benefit gained from use of AZLI is likely to be outweighed by losses from use of a detrimental TID vs. BID regimen.

Secondary Analyses

Regimen effects in secondary analyses could not be clearly observed due to lack of power. Placebo TID vs. Placebo BID, the only comparison directly estimating the TID vs. BID regimen effect, was severely underpowered as it included only 76 of the 211 ITT subjects. Furthermore, there is a high degree of missing data especially at later visits to complicate assessment of a regimen effect. Overall, secondary analyses measured across the entire AZLI/placebo study period (Day 0 to Day 84) did not appear inconsistent with regimen effects observed in the primary analysis. However, it did appear that compared to regimen effects observed over the Day 0-84 period, regimen effects observed from Day 0-28 were generally smaller. However, overall, it is not clear that regimen effects in secondary analyses can be ruled out or be considered as inconsistent with primary analysis regimen effects. It is further noted that AZLI TID patients generally fared worse than AZLI BID patients across most endpoints and had significantly higher early termination rates despite having the benefit of receiving 50% more AZLI drug (225mg vs. 150mg). This observation would appear to be consistent with a general TID vs. BID regimen effect.

CFQ-R Respiratory Symptoms Score (CFQ-R RSS)

In the CFQ-R RSS endpoint, the Placebo TID vs. Placebo BID comparison based on observed cases trended towards significance at Day 28 (p=.206) and was at borderline significance by Day 84 (p=.056). (Note that CFQ-R RSS endpoint was not observed after Day 28 except for the Day 84 visit). Furthermore, when imputing dependent missing BID and TID data using the Placebo BID and Placebo TID (or BID and TID) group means, these p-values reduced to p=.164 at Day 28 and p=.015 at Day 84 (Table 6). These findings appeared to be consistent with the regimen effect observed in the time to need analysis although they were not always significant. However, as stated earlier, this lack of significance may be due to the limited numbers of placebo subjects and high rates of dependent missing data.

Subgroup Analyses

Regimen effects in the primary analysis were compared across various subgroups related to age (< 18, ≥18), gender (male, female) and disease severity (FEV_1% pred. ≤ 50%, FEV_1% pred. > 50%). Patients on TID regimen fared numerically worse than patients on a BID regimen regardless of the subgroup or the regimen effect comparison considered (e.g. Placebo BID vs. Placebo TID or AZLI BID vs. AZLI TID). Regimen differences were more pronounced in placebo regimens and especially strong in patients with lesser disease severity (FEV_1 > 50% of predicted), female patients and adult patients (Table 9). There was also regimen effects in subgroup analyses in the CFQ-R RSS in which female patients on a TID regimen fared worse than female patients on a BID regimen (p=.0239) as shown in Table 11. These findings further support the robustness of the regimen effect.

Unclear Dose Response Relationship due to Regimen Effects

If a treatment is indeed effective, we expect to see a clear dose response relationship with more favorable responses resulting from more drug received. In Study 005, the dose response relationship is not clear in AZLI BID vs. Placebo BID comparisons for most endpoints since the effect of the 150 mg of AZLI daily is often negligible. For primary and sensitivity analyses of time to event, AZLI BID vs. Placebo BID differences failed to even approach significance (p=.4237 to p=.9204). The lack of significance of AZLI BID vs. Placebo BID is problematic in making inferences of significance for AZLI TID, especially since AZLI TID patients fared worst than AZLI BID patients for most endpoints and had significantly higher early termination rates compared to AZLI BID patients (p=.0268).

Unfavorable AZLI TID vs. AZLI BID comparisons further supported an unclear dose response relationship. We would expect AZLI TID patients to fare better than AZLI BID patients due to their having received 225 mg AZLI vs. 150 mg AZLI daily. However, we actually observed the opposite with AZLI BID patients generally performing better than AZLI TID patients. One explanation for this would be a consistent TID vs. BID regimen effect which outweighs possible benefits from additional AZLI drug. This is further supported in placebo patients where estimated differences in Placebo TID vs. Placebo BID also consistently favored the BID regimen (i.e. Placebo BID).

Real Effect vs. Chance Finding

The regimen effect appears to be a real study effect rather than just a chance finding. The probability of the event of a detrimental TID vs. BID effect such that primary analysis findings would be called into question, as in Study 005, is only 0.0006 (.0012/2; Table 3 using one-sided test). In addition to this finding, the probability of the event of a shorter time to early termination in TID vs. BID patients among those censored (i.e. not having a time to need event) is only 0.066 (0.1319/2; Figure 23- one sided test). These two events favoring TID vs. BID with probabilities of 0.0006 and 0.066 are considered to have little or no positive correlation and it would be highly unlikely for both of these events to occur solely by chance.

Since the regimen effect was considerably more robust in primary and sensitivity analyses relative to the AZLI drug effect, it is not clear how we can objectively conclude that AZLI drug effects are real effects while regimen effects are not real effects that were solely due to chance and can be ignored. It should further be noted that after controlling for biases from more AZLI vs. Placebo and more TID vs. BID early terminations as performed in S4 (Table 3), the Pooled AZLI vs. Pooled Placebo effect is not significant (p=.6029) while Pooled BID vs. Pooled TID remains highly significant (p=.0030).

3.5 AZLI TID Comparisons for Secondary Endpoints

Due to lack of assay sensitivity in the primary analysis in which comparisons of AZLI TID against Placebo BID were highly unfavorable especially after controlling for missing data, the primary study hypothesis could not clearly be rejected. Consequently, there may not be adequate power for statistical testing of any secondary endpoints. However, Study 005 results failed to show significant AZLI TID benefits over placebo for most clinical endpoints even without controlling for multiplicity.

Section 3.5.1 shows comparisons for AZLI TID and AZLI BID against Pooled Placebo for the Applicant's key secondary endpoints as well as other secondary endpoints. These comparisons show:

- Lack of a significant AZLI TID benefit over placebo for all clinical endpoints considered (i.e. changes in CFQ-R (respiratory), CFQ-R (non-respiratory), FEV_1, FVC, FEF_{25-75}, hospitalization, weight, BMI, missed school/work)
- AZLI TID as generally worse than AZLI BID
- Greater hospitalization in AZLI vs. Placebo Patients

Sections 3.5.2 provides a more detailed analysis of clinical endpoints of interest which included hospitalization endpoints, changes in CFQ-R (respiratory) and changes in FEV_1. These findings show the following for AZLI TID patients:

- Greater hospitalization rates (p=.113), shorter time to first hospitalization (p=.085) and more days hospitalized (p=.083)
- Marginal benefits vs. placebo in CFQ-R (respiratory) & FEV_1 from Day 0-28 with no sustained benefit by Day 42 and worsening by Day 84.
- AZLI TID patients had significant worsening relative to Placebo BID in the CFQ-R RSS at Day 84 based on imputed data (p=.020).

3.5.1 Secondary Analyses

Table 4: Key Secondary Endpoints Reported by Applicant (ITT)

Study 005 Comparison:			AZLI BID vs. Pooled Placebo	AZLI TID vs. Pooled Placebo
Endpoint (Change from Day 0):	**At Day:**	**Using:**	**Difference (95% CI), p-value**	
FEV_1 (L)* (Actual Change)	Day 28	Observed Cases	.069 (-.015, .154), p=.106	.084 (-.003, .171), p=.057
	Day 42		.079 (-.009, .166), p=.077	.040 (-.050, .130), p=.381
	Day 28	LOCF	.105 (.022, .188), p =.013	.090 (.006, .173), p =.036
	Day 42		.111 (.033, .189), p =.005	.065 (-.014, .144), p =.105
CFQ-R** (Respiratory Domain)	Day 28	Observed Cases	5.29 (.18, 10.39), p =.043	4.37 (-.94, 9.69), p =.106
		LOCF	5.77 (.89, 10.64), p =.021	4.22 (-.07, 9.14), p =.092
		Categories, LOCF	18.2% improved, p =.030	10.7% improved, p =.141
Log_{10} PA CFUs	Day 28	Observed Cases	-.72 (-1.26, -.17), p =.011	-.60 (-1.15, -.06), p =.031
	Day 42		.29 (-.29, .86), p >.999	-.15 (-.73, .43), p =.610

* FEV_1 (L) endpoint included actual and percent change (not shown), ** CFQ-R not administered at Day 42

In Table 4, using observed cases, AZLI TID patients had a marginal benefit in FEV_1 from Day 0 to Day 28 of .084 Liters (L) with a 95% CI of (-.003, .171) and p=.057. However, this improvement was not sustained at Day 42, dropping to .040 L (-.050, .130), p=.381. In contrast, AZLI BID patients actually increased their treatment improvement from Day 28 to Day 42 (i.e. 0.69 L to 0.79 L). In the CFQ-R RSS endpoint, AZLI TID patients had a marginal benefit at Day 28 with 4.37 (-0.94, 9.69) point increase in respiratory symptoms score, p =.106. However, since CFQ-R RSS had not been measured at Day 42, the sustained benefit is not clear. In the Log_{10} *PA* CFU endpoint, there did appear to be a reduction in CFUs at Day 28 for both the AZLI BID and AZLI TID treatments.

Table 5: Other Secondary Endpoints Reported by Applicant (ITT)

Study 005 Secondary Comparison:			AZLI BID vs. Pooled Placebo	AZLI TID vs. Pooled Placebo
Endpoint:	**Period**	**Using**	**Difference, 95% CI or p-value** *p-values in red for test of placebo benefit vs. AZLI*	
Time to First Hosp. (days)	Day 0-84	Observed cases	n/a, p=.410	n/a, p=.085
% Patients Hospitalized			8.7 vs. 3.9 , 4.8%, p=.310	12.1 vs. 3.9, 8.2%, p=.113
Number of hospital days			0.8 vs. 0.5, 0.3, p=.254	1.0 vs. 0.5, 0.5, p=.083
% Patients missing School/Work			23.2 vs. 22.4, 0.8, p=.860	19.7 vs. 22.4, -2.7, p=.837
CFQ-R (non-respiratory)	Day 0-28 Change	LOCF	1.80 (-.68, 4.29), p=.153	1.46 (-1.05, 3.97), p=.254
FEF_{25-75} (L/sec)			.131 (.030, .232), p=.012	.095 (-.008, .197), p=.069
FVC (L)			.125 (.008, .242), p=.036	.093 (-.025, .211), p=.120
Weight (kg)		Observed cases	1.03 (0.15, 1.91) p=.023	.48 (-.44, 1.39), p=.305
BMI (kg/m^2)			224 (.039, .409), p=.018	.116 (-.077, .309), p=.237

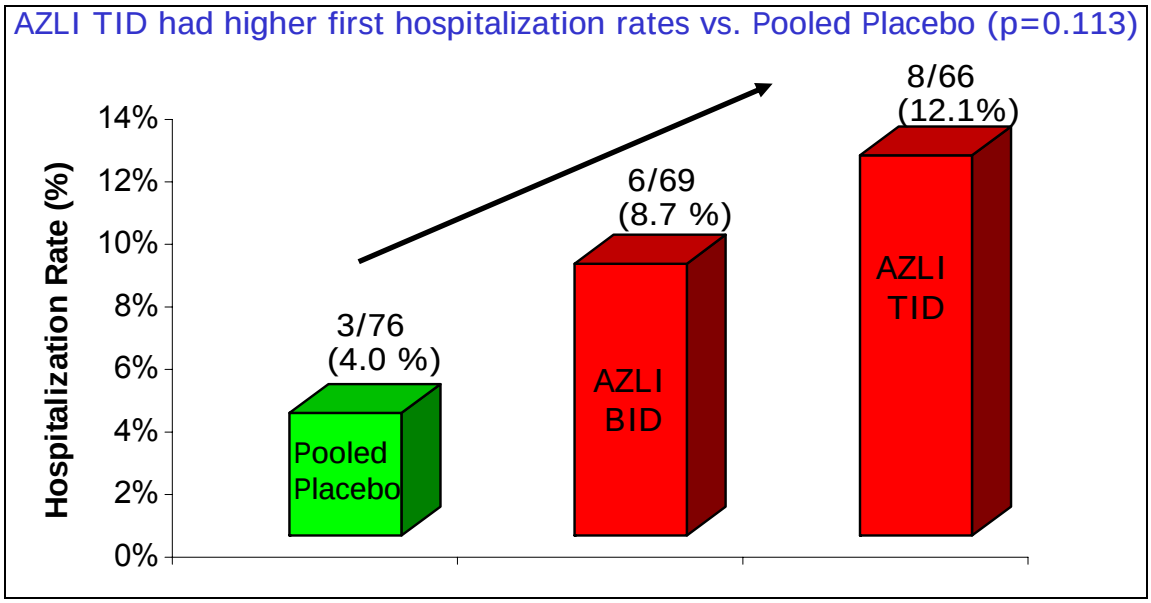
In Table 5, we can observe a lack of significance of AZLI TID for all comparisons against placebo. We can also see a trend towards greater hospitalization rates in AZLI vs. Pooled Placebo (p=.113) as well as shorter time to first hospitalization (p=.085) and more days hospitalized (p=.083). AZLI BID patients generally fared better than AZLI TID patients for the endpoints listed.

3.5.2 Clinical Endpoints of Interest

The clinical endpoints of most interest included those related to hospitalization as well as the CFQ-R (respiratory) and FEV_1 measures. The evidence of a potential AZLI TID benefit in these endpoints is unclear, especially after Day 28. In the case of hospitalization, there are concerns of potential detrimental effects from AZLI use.

Changes in Hospitalization

Figure 12: Hospitalization Rates (ITT)



In Figure 12, AZLI TID patients had substantially higher first hospitalization rates in comparison to placebo patients (12.1% vs. 4.0%), p=.113. Figure 12 also shows a 'dose-response' relationship between AZLI drug and hospitalization. Patients taking 75 mg of AZLI twice daily vs. placebo more than doubled their hospitalization rate while patients taking AZLI three times daily vs. placebo had more than tripled their hospitalization rate. Hospitalization rates for Placebo BID and Placebo TID (not shown) were 1/38 (2.6%) and 2/38 (5.3%), respectively. Although there was greater hospitalization in Placebo TID vs. Placebo BID and AZLI TID vs. AZLI, regimen effects were unclear due to the small number of hospitalizations.

In addition to higher hospitalization rates, AZLI TID patients also had a shorter time to first hospitalization vs. Pooled Placebo (p=.085) and spent more days hospitalized, 1.05 days per AZLI TID patient vs. 0.46 days per Pooled Placebo patient, p=.083 using a non-parametric rank sum test (Figures 35-36 of the Appendix).

Changes in CFQ-R Respiratory Symptoms

The change from baseline in the CFQ-R respiratory symptoms score (CFQ-R RSS) at Day 28 was the primary endpoint of Study 007 and a key secondary endpoint in Study 005. However, evidence in Study 005 from this endpoint was limited because the CFQ-R was only measured at the Day 28 and Day 84 visits. Additionally, CFQ-R RSS changes in observed patients suggested a possible detrimental TID regimen effect for this endpoint (especially at Day 84). Differences in dropout rates also suggested a detrimental TID regimen effect. While regimen differences were smaller at Day 28 vs.

Day 84, there appeared to be a trend at Day 28 towards less favorable CFQ-R RSS changes in the TID regimens.

Figure 13: Changes in CFQ-R RSS (ITT)

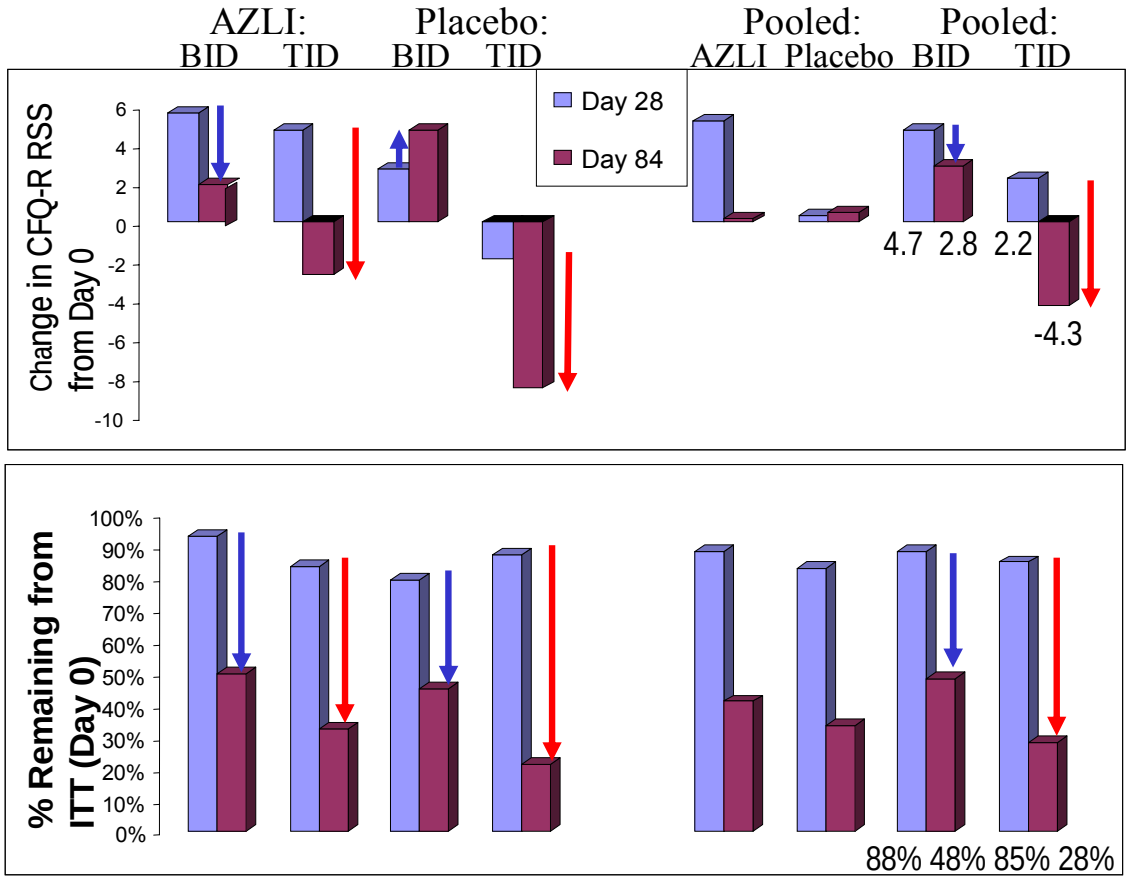


Figure 13 (top) shows that at Day 28, there are marginal benefits in CFQ-R RSS for AZLI TID patients and that Placebo BID patients are faring better than Placebo TID patients indicating a possible regimen effect. However, at Day 84, there are substantially greater drops in CFQ-R RSS in the TID regimens (red arrows) in comparison to the BID regimens (blue arrows). From Day 28 to Day 84, CFQ-R RSS in patients on BID regimens dropped 1.9 points from a 4.7 point improvement from baseline at Day 28 to a 2.8 point improvement at Day 84. However, CFQ-R RSS in patients on TID regimens dropped 6.5 points from a 2.2 point improvement at Day 28 to a -4.3 point worsening at Day 84.

Figure 13 (bottom) shows that similar trends also occurred with respect to patients dropping out of the CFQ-R RSS analysis. From Day 28 to Day 84, drops were greater in the TID regimens (red arrows) than in the BID regimens (blue arrows). In BID regimens from Day 28 to Day 84, the percentage of ITT patients remaining in the analysis fell from 88% (Day 28) to 48% (Day 84). In TID regimens, the percentage of ITT patients remaining in the analysis fell from 85% (Day 28) to 28% (Day 84). The combination of substantial observed TID vs. BID effects in the CFQ-R RSS and greater numbers of

patients dropping out of the CFQ-R RSS analysis from TID vs. BID regimens further supports the robustness of the regimen effect.

Table 6: Observed (Imputed) P values for CFQ-R RSS Changes (ITT)

AZLI vs. Placebo **p values:**		Placebo BID Obs. (imputed)	Placebo TID Obs. (imputed)	Pooled Placebo Obs. (imputed)
AZLI BID	Day 28	.377 (.368)	.017 (.013)	.043 (.020)
	Day 84	.565 (.694)	.099 (.020)	.737 (.146)
AZLI TID	Day 28	.556 (.775)	.042 (.064)	.106 (.145)
	Day 84	.156 (.020)	.383 (.667)	.504 (.165)
Pooled AZLI	Day 28	.413(.497)	.014 (.014)	.033 (.013)
	Day 84	.308 (.115)	.156 (.097)	.928 (.936)
BID vs. TID		**AZLI:**	**Placebo:**	**Pooled:**
	Day 28	.735 (.473)	.206 (.164)	.256 (.157)
	Day 84	.292 (.023)	.056 (.015)	.054 (.001)

Table 6 shows comparisons among the treatment arms in changes from baseline in CFQ-R RSS at Day 28 and at Day 84 using observed and imputed data. To control for potential biases from dependent missing data, Placebo BID and Placebo TID group means at Day 28 and BID and TID group means (Day 84) were used to impute missing BID and TID data. We see less favorable CFQ-R RSS changes in TID vs. BID patients which trended towards significance at Day 28 and were significant at Day 84 using observed (imputed) data, p=.054 (p=.001). This indicates that pooled comparisons may not be justified, especially at Day 28. Using the more conservative AZLI TID vs. Placebo BID comparison shows no improvement from AZLI TID at Day 28 and possible worsening at Day 84, p=.156 (p=.020).

Changes in FEV_1

Changes in FEV_1 were used in the original TOBI trials as a co-primary endpoint and were considered as another clinical endpoint of interest. However, Study 005 changes in FEV_1 did not show a clear AZLI TID benefit, especially after controlling for potential biases from dependent missing data. Furthermore, any potential benefit from AZLI TID vs. placebo treatment in FEV_1 at Day 28 was not sustained at Day 42 and there was worsening vs. placebo at Day 84.

Figure 14: FEV_1 Differences Using Observed Cases (ITT)

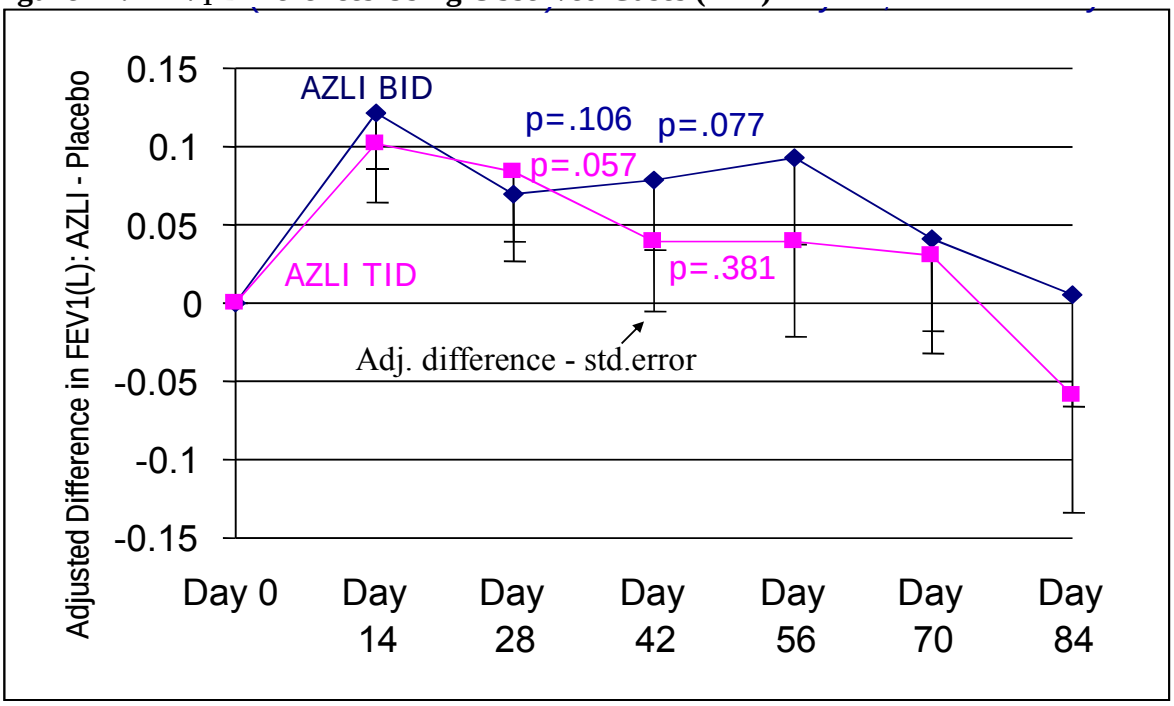
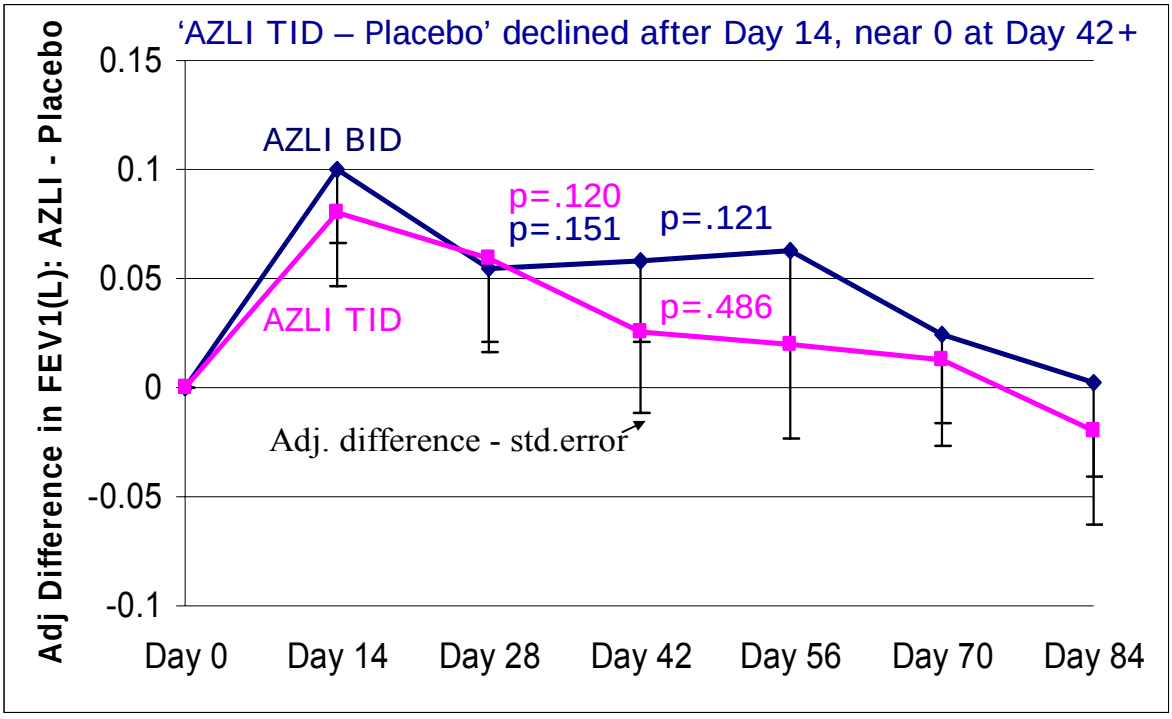


Figure 15: Changes in FEV_1 using Imputed Data (All Randomized)



Figures 14-15 show that changes in FEV_1 in AZLI TID patients were not significant at Day 28 and were near 0 from Day 42 to Day 84, especially after imputing missing data in randomized patients based on the Pooled Placebo mean.

3.6 Summary and Conclusions

Study 005 may have provided some evidence of an AZLI drug effect but there appeared to be stronger evidence of a TID vs. BID regimen effect, especially in the analyses over the entire study period such as the primary analysis. Since the AZLI TID treatment effect is a combined effect from both the AZLI drug and the TID regimen (detrimental effect), there is high degree of uncertainty as to whether AZLI TID patients would receive an overall treatment benefit. Primary and sensitivity analyses suggest that patients taking AZLI TID would not receive a treatment benefit but may have detrimental effects, especially when controlling for possible biases from dependent missing data and regimen effects. For example, time to early termination in randomized patients clearly favored Placebo BID over AZLI TID (p=.0941).

Secondary analyses were more severely limited by dependent missing data which was significantly higher in TID vs. BID regimens and AZLI vs. Placebo randomized patients prior to AZLI/Placebo treatment. At Day 28, in which the majority of secondary endpoints were assessed, missing data rates were substantially higher for AZLI TID vs. other treatment arms. This led to possible directional biases favoring AZLI TID. However, even when ignoring possible regimen effects, dependent missing data favoring AZLI TID and issues related to multiple testing, secondary analysis comparisons failed to clearly show a significant AZLI TID benefit over placebo in any of the clinical endpoints considered. When controlling for potential biases from dependent missing data, AZLI TID comparisons trended further away from statistical significance. Furthermore, hospitalization endpoints clearly favored placebo patients over AZLI TID patients with p-values approaching statistical significance.

Study 005 failed to provide clear and reliable evidence of an AZLI TID treatment benefit. There is uncertainty even when considering AZLI BID based on primary and sensitivity analysis comparisons of AZLI BID vs. Placebo BID which failed to approach significance. Hospitalization rates also clearly favored placebo patients. This is notable because AZLI BID patients fared better than AZLI TID in nearly all analyses, especially when controlling for possible biases from higher rates of TID vs. BID drop outs. For example, AZLI BID patients had a significantly longer time to early termination vs. AZLI TID patients (p=.0268). This further confirms the lack of efficacy with the AZLI TID treatment as well as possible adverse effects from use of a TID vs. BID regimen.

4. EVALUATION OF EFFICACY (STUDY 007)

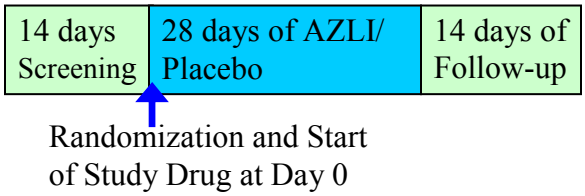
4.1 Review of Study 007

Design: Study 007 was a placebo controlled study in CF patients 6+ yrs with *PA* & FEV_1 25%-75% of predicted.

Study Objective: The study objective was to demonstrate efficacy & safety of AZLI (75 mg TID).

In comparison to Study 005, Study 007 patients were less extensively treated:
- TSI courses in prior 12 months not required
- Use of chronic azithromycin therapy not allowed
- Use of macrolide or antipseudomonal antibiotics within 28 days of study drug not allowed

Figure 16: Study 007 Design



Study Arms: Study 007 included 2 study arms: AZLI TID and Placebo TID. Randomization was 1:1 at the start of study drug at Day 0.
- N=166 randomized at Day 0 (83 AZLI, 83 placebo)
- N=164 treated in ITT (84 AZLI, 80 Placebo)

Hypotheses Tested: Study 007 tested the primary hypothesis: H_0: AZLI TID = Placebo TID

Primary endpoint: Actual Change in CFQ-R Respiratory Domain Scores from Day 0 to Day 28

Key secondary endpoints which were statistically controlled:
- % Change in FEV_1 at Day 28 (tested at α=.025)
- Change in log 10 *PA* CFU at Day 28 (tested only if above comparison significant)
- Proportion of patients receiving IV/inhaled antibiotics (tested at α=.025)
- Proportion of patients hospitalized (tested only if above comparison significant)

Other secondary endpoints included:
- Other CFQ-R (e.g. non-respiratory)
- Other spirometry (e.g. changes in FVC and FEF)
- Other hospitalization/time to need for antibiotics endpoints
- Number of school/work days missed

- % Change in Weight/BMI

Description of Primary Endpoint*:* The primary endpoint was change from Day 0 to Day 28 in the respiratory domain of the CFQ-R. The CFQ-R for adults and adolescents consists of several demographic questions followed by 50 questions intended to measure health-related quality of life for CF patients in multiple domains (physical functioning, role, vitality, emotional functioning, social, body image, eating disturbances, treatment burden, health perception, weight, respiratory symptoms, and digestive symptoms). A total of seven questions (#40, 41, 42, 43, 44, 45, and 46) are related to the respiratory symptoms domain of the CFQ-R. The questions and response categories are as follows:

Indicate how you have been feeling during the past two weeks.
40. Have you been congested?
41. Have you been coughing during the day?
42. Have you had to cough up mucus?
(The response categories for these questions are: A great deal, Somewhat, A little, or Not at all. Patients who answer not at all for question 42 are instructed to go to question 44.)

43. Has your mucus been mostly: Clear, Clear to yellow, Yellowish-green, Green with traces of blood, or Don't know

How often during the past two weeks:
44. Have you been wheezing?
45. Have you had trouble breathing?
46. Have you woken up during the night because you were coughing?
(The response categories for these three questions are: Always, Often, Sometimes, or Never.)

For more detailed information regarding Study Design and Endpoints, Subject Disposition, Demographic and Baseline Characteristics and Statistical Methodologies refer to our earlier statistical review of the Study 007 protocol (Sections 3.2.1-3.2.3).

4.2 Uncertainty in Primary Analysis

There is some uncertainty in the Study 007 primary analysis due to the patient reported outcome (PRO) instrument used (i.e. CFQ-R respiratory symptoms score, CFQ-RSS) For example, the CFQ-R RSS relies on a 2 week recall period to assess respiratory symptoms. This creates uncertainty as to what is being measured and the interpretation of the minimal important clinical difference (MCID) if patients do not have any recollection of respiratory symptoms experienced in the last two weeks. This also creates concerns that patients may be more easily influenced by possible study biases if they feel they cannot answer the questions confidently. Note that the current PRO guidance emphasizes capturing a patient's current state without comparing the current state to an earlier time period or recalling an earlier time period. Instrument reliance on a 2 week recall period would be especially problematic with younger patients. In Study 007 approximately 20%

of patients were under 18 years of age, however, this subgroup of patients had a strong influence on study results.

Other PRO instrument concerns related to the responder burden for the CFQ-R which seems high when interest is limited to only 7 questions relating to respiratory symptoms. Some response categories appeared unclear (e.g. "Somewhat" vs. "A little"). Some questions appeared vague (e.g. "trouble breathing" could represent dyspnea, or problems with cough and mucus production). It is also not clear if the questions capture all respiratory symptoms important to patients and why there is a strong emphasis on "cough" (three questions).

There are other concerns regarding the reliability of the MCID which was estimated to be a 5 point change. This estimate was primarily based on patients aged 14 and above who were observed during the Study 005 TSI run-in period. However, compared to Study 007 patients, Study 005 patients were more extensively treated than with less of a response to the AZLI treatment. Therefore, the MCID of 5 may underestimate the true MCID in the Study 007 population in which treatment effects were greater. Further underestimation of the MCID of 5 could result from the inclusion of some patients under the age of 14 in Study 007, a group in which the MCID could be established but would be expected to be higher than in patients over 14.

While these uncertainties were unlikely to lead false findings of statistical significance, they do raise questions regarding the interpretability of findings and whether the observed changes in CFQ-R could be considered as 'clinically meaningful'. Note also that primary analysis results, though highly significant, could not establish a clinically meaningful improvement in AZLI TID patients which had been defined as at least 5 points.

4.3 Efficacy Results

In Table 7, the primary endpoint of 'Actual Change in CFQ-R RSS, Day 0-28' was significant (p=.0005). Secondary comparisons further showed that both FEV_1 and CFQ-R RSS were significant at Days 28 and Day 42. However, two of the four key secondary endpoints that were pre-specified to control the overall type I error rate failed to show significance. These comparisons were based on proportion of patients needing antibiotics and the proportion of patients needing hospitalization during the course of the study. The lack of significance in these secondary endpoints was likely influenced by short follow-up period in Study 007 which limited power in detecting significant differences.

As noted in Section 4.2, although the primary analysis was significant (p=.0005) it failed to establish a clinically meaningful difference in CFQ-R RSS based on the lower bound of 4.31 which was below the MCID of 5. Also, although AZLI TID vs. Placebo TID CFQ-R RSS differences remained significant at Day 42, there was a steep loss from Day 28 to Day 42, a 7.08 to a 0.62 point mean improvement from baseline. This created some concerns regarding the sustainability of benefits after the AZLI TID treatment course. FEV_1 improvements from baseline also showed steep decreases from Day 28 to Day 42 (i.e. 10.29% to a 3.14% mean improvement).

Table 7: AZLI vs. Placebo Comparisons for Study Endpoints (ITT)

Study 007 Endpoints	AZLI TID vs. Placebo TID Difference (95% CI)	Observed P-values
Actual Change in CFQ-R RSS, Day 0-28 (primary)	7.08 vs. -2.63 9.71 (4.31, 15.11)	.0005[b]
Actual Change in CFQ-R RSS, Day 0-42	0.62 vs. -5.71 6.33 (1.22, 11.43)	.0154[b]
Categorized Change in CFQ-R RSS, Day 0-28	Improved: 45 (57.5%) vs. 31 (37.3%) No change: 15 (18.8%) vs. 15 (18.1%) Worsening: 20 (25.0%) vs. 37 (44.6%)	.0055
FEV_1 (L) % Change from Day 0-28	7.89% vs. -2.41% 10.29% (6.29, 14.30)	.0001[a,b]
FEV_1 (L) % Change from Day 0-42	3.14% vs. -2.59%, 5.73% (2.07, 9.40)	.0024[b]
Change in sputum log_{10} *PA* CFUs, Day 0-28	0.069 vs. -1.384 -1.453 (-2.115, -0.791)	.0001[a]
Time to Need for IV/Inhaled Antibiotics Due to Pre-Defined Symptoms	N/A	.0949
Proportion of Patients Not Using IV/Inhaled Antibiotics, Day 0-42	68 (85.0%) vs. 65 (77.4%) 7.6%	.2364[a]
Proportion of patients hospitalized, Day 0-42	4 (5.0%) vs. 12 (14.3%) -7.3%	.0640[a]
Proportion of patients with school/work missed, Day 0-28	7 (8.3%) vs. 13 (16.3%), -8.0%	.2201
Mean change in Weight (kg), Day 0-28	1.085 vs. 0.074 1.010 (.330, 1.691)	.0039[b]
Mean change in BMI (kg/m^2), Day 0-28	.213 vs. 011 .202 (.061, .344)	.0054[b]

[a] Secondary comparisons pre-specified under a statistical framework to control the overall type I error rate, [b] Means adjusted by disease severity at Day -14 and baseline measurement for respective endpoint. Missing data imputed by LOCF.

4.4 Summary and Conclusions

Study 007 was an adequate and well controlled study which demonstrated its primary endpoint, an improvement from Day 0 to Day 28 in CFQ-R RSS (p=.0005). This result was supported by significant findings at Day 28 in the 'changes in FEV_1' and 'changes in *PA* CFUs' endpoints. However, there are some concerns with the quality of evidence presented from Study 007.

- **Limited Patient Follow-up**: Study 007 included only 14 days of follow-up and failed to evaluate a critical time period in the 56 day cycle of 28 days on/28 days off AZLI TID therapy (i.e. Day 43 to Day 56). In Study 005, for example, the majority of time to need events occurred after Day 42 with a large number of events occurring between Days 43 and Day 56. The short study follow-up in Study 007 also limited comparisons of 'time to need for IV/inhaled antibiotics' and 'time to first hospitalization', both of which were not found to be significant in Study 007 but appeared to trend towards significance.

- **Unblinded sample size re-estimation (SSR):** Biases can result from SSR especially when unblinded (even if performed by a third party) and not executed according to plan. The actual sample size planned (n=140) differed from the sample size recommended by the third party (n=150). A sample size of 164 subjects was actually used due to over enrollment which indicates Study 007 may have been overpowered.

- **Validity/reliability of the CFQ-R RSS instrument**: Some questions used to compute the CFQ-R RSS relied on a 14-day recall period which is especially problematic in younger patients. Some of the language used in the instrument was vague or unclear. Some questions may have emphasized certain symptoms too much (e.g. cough included in 3 questions). There is also a responder burden in which subjects must answer 51 questions with only 7 questions included in the CFQ-R RSS.

- **Interpretations of clinical benefits and the appropriateness of the MCID:** MCID estimates of a 5 point difference in CFQ-R RSS were based on a more extensively treated study population (Study 005 TSI period) in which less pronounced treatment effects were observed. The MCID estimate was also based on a specific age group (patients 14 and over) while Study 007 included patients aged 6 and over. The MCID in patients between the ages of 6 and 13 is not clear but is thought to be higher than 5 points. There were 17 patients in Study 007 between the ages of 6 and 13 with CFQ-R RSS measurements at Day 28 who showed a mean treatment difference of 6.53 points, 13.74 in AZLI TID vs. 7.21 in Placebo.

- **Robustness of primary analysis findings:** Robustness of primary analysis findings was a concern in Study 007 especially if findings are considered to provide evidence from two independent studies. The CFQ-R RSS finding was mostly driven by patients under 18 years of age (n=37) who had substantially larger improvements over placebo compared to patients aged 18 or older (n=126) at 18.92 (95% CI: 8.78, 29.05), p=.0006; vs. 6.35 (95% CI: 0.02, 12.69), p=.0495.

5. SPECIAL/SUBGROUP POPULATIONS

5.1 Study 005

Table 8 examines subgroups based on AZLI TID treatment effects. Table 8 shows that AZLI TID was not significantly different from Placebo BID in any of the subgroups considered. However, AZLI TID vs. Placebo TID comparisons were significant or trended towards significance in many subgroups. The interpretation of the AZLI TID treatment effect depended heavily on whether Placebo BID or Placebo TID was used for comparison. Note that the AZLI TID vs. Placebo TID comparison is not considered to be informative if there is a detrimental TID regimen effect. Under this scenario, Placebo TID would then fail to provide an appropriate estimate of the true placebo effect.

Table 9 examines regimen effects based on event-free rates in Placebo TID vs. Placebo BID, AZLI TID vs. AZLI BID and TID vs. BID comparisons for various subgroups.

Event was defined according to the primary analysis as a need for IV or inhaled antibiotics due to pre-defined symptoms. It should be noted that the AZLI TID vs. AZLI BID comparison likely underestimates the regimen effect since AZLI TID patients may have benefitted by having received more AZLI drug than AZLI BID patients (225mg vs. 150 mg). Regardless of the comparison and subgroup considered, the TID group had numerically lower event-free rates than the BID group. Regimen differences were more pronounced in placebo regimens and especially strong in patients with lesser disease severity (FEV_1> 50% of predicted), female patients and adult patients.

Table 8: Event-Free Rates in Need for IV/Inhaled Antibiotics due to Predefined Symptoms by Subgroup, Treatment Effects (ITT)

Subgroup Category	**AZLI TID vs. Placebo BID**	**AZLI TID vs. Placebo TID**
Disease Severity		
$FEV_1 \leq$ 50% pred.(n=76)	12/22 (55%) vs. 8/15 (53%) p=.999	12/22 (55%) vs. 4/15 (27%) p=.176
FEV_1> 50% pred.(n=134)	30/44 (68%) vs. 18/23 (78%) p=.569	30/44 (68%) vs. 8/23 (35%) p=.011
Age Group		
≥ 18 (n=165)	30/49 (62%) vs. 21/30 (70%) p=.476	30/49 (62%) vs. 10/34 (29%) p=.007
< 18 (n=46)	12/17 (71%) vs. 5/8 (63%) p=.999	12/17 (71%) vs. 2/4 (50%) p=.574
Gender		
Male (n=121)	24/38 (63%) vs. 16/26 (62%) p=.895	24/38 (63%) vs. 7/19 (37%) p=.091
Female (n=90)	18/28 (64%) vs. 10/12 (63%) p=.999	18/28 (64%) vs. 5/19 (26%) p=.017
Overall (n=211)	42/66 (64%) vs. 26/38 (68%) p=.621	42/66 (64%) vs. 12/38 (32%) p=.002

P values in red are for tests of a Placebo BID benefit, Fisher's exact test used in cases where normal approximation to the binomial may be invalid, such as when $np(1-p) < 5$ in either sample.

Table 9: Event-Free Rates in Need for IV/Inhaled Antibiotics due to Predefined Symptoms by Subgroup, Regimen Effects (ITT)

Subgroup Category	**Placebo BID vs. Placebo TID**	**AZLI BID vs. AZLI TID**	**BID vs. TID**
Disease Severity at Day -14			
$FEV_1 \leq 50\%$ pred.(n=76)	8/15 (53%) vs. 4/15 (27%) p=.264	15/24 (63%) vs. 12/22 (55%) p=.584	23/39 (59%) vs. 16/37 (43%) p=.170
$FEV_1 > 50\%$ pred.(n=134)	18/23 (78%) vs. 8/23 (35%) p=.007	34/44 (77%) vs. 30/44 (68%) p=.338	52/67 (78%) vs. 38/67 (57%), p=.010
Age Group			
≥ 18 (n=165)	21/30 (70%) vs. 10/34 (29%) p=.002	37/52 (71%) vs. 30/49 (68%) p=.752	58/82 (71%) vs. 40/83 (48%), p=.003
< 18 (n=46)	5/8 (63%) vs. 2/4 (50%) p=.999	13/17 (76%) vs. 12/17 (71%) p=.999	18/25 (72%) vs. 14/21 (67%) p=.755
Gender			
Male (n=121)	16/26 (62%) vs. 7/19 (37%) p=.136	28/38 (74%) vs. 24/38 (63%) p=.324	44/64 (69%) vs. 31/57 (54%) p=.104
Female (n=90)	10/12 (63%) vs. 5/19 (26%) p=.003	22/31(71%) vs.18/28 (64%) p=.583	32/43 (74%) vs. 23/47 (48%) p=.013
Overall (n=211)	26/38 (68%) vs. 12/38 (32%) p=.001	50/69 (72%) vs. 42/66 (63%) p=.271	76/107 (71%) vs. 54/104 (52%) p=.004

Fisher's exact test used in cases where normal approximation to the binomial may be invalid, such as when $np(1-p) < 5$ in either sample.

Tables 10-11 examines AZLI TID treatment and regimen effects in Study 005 for CFQ-R RSS changes from Day 0 to Day 28 by subgroup based on observed cases in ITT patients. Table 11 shows that treatment differences (AZLI TID vs. Pooled Placebo) were not especially strong in any of the subgroups. Table 11 shows regimen effects were strong in female patients in which those on a Placebo BID vs. Placebo TID regimen fared better at 15.07 points (2.06, 28.10), p=.0239.

Table 10: Study 005 CFQ-R RSS Changes from Day 0 to Day 28 by Subgroup Based on Observed Cases (ITT)

Subgroup Category	**AZLI TID vs. Pooled Placebo Difference (95% CI)**	**p-value**
Disease Severity		
$FEV_1 \leq$ 50% predicted (n=76)	4.77 vs. -1.81 6.58 (-2.18, 15.35)	p=.1383
$FEV_1 >$ 50% predicted (n=134)	4.64 vs. 2.17 2.47 (-4.35, 9.31)	p=.4740
Age Group		
≥ 18 (n=165)	5.61 vs. 0.99 4.62 (-1.64, 10.89)	p=.1470
< 18 (n=46)	1.98 vs. -1.83 3.81 (-5.77, 13.38)	p=.4252
Gender		
Male (n=121)	0.82 vs. -0.95 1.77 (-4.79, 8.33)	p=.5942
Female (n=90)	9.06 vs. 2.52 6.54 (-2.81, 15.89)	p=.1676
Overall (n=211)	4.67 vs. 0.30 4.37 (-0.94, 9.69)	p=.1061

Table 11: Study 005 CFQ-R RSS Regimen Effects from Day 0 to Day 28 by Subgroup Based on Observed Cases (ITT)

Subgroup Category	**Placebo BID vs. Placebo TID Difference (95% CI)**	**p-value**
Disease Severity		
$FEV_1 \leq$ 50% predicted (n=76)	1.06 vs. -4.11 5.17 (-6.50, 16.84)	p=.3793
$FEV_1 >$ 50% predicted (n=134)	4.18 vs. 0.18 4.00 (-5.50, 13.50)	p=.4062
Age Group		
≥ 18 (n=165)	4.03 vs. -1.42 5.46 (-2.96, 13.88)	p=.2018
< 18 (n=46)	-1.04 vs. -3.21 2.17 (-12.89, 17.23)	p=.7716
Gender		
Male (n=121)	-1.13 vs. -0.74 -0.39 (-9.06, 8.28)	p=.9999
Female (n=90)	11.95 vs. -3.12 15.07 (2.06, 28.10)	p=.0239
Overall (n=211)	2.73 vs. -1.92 5.66 (-2.61, 11.92)	p=.2078

5.2 Study 007

Table 12 examines treatment effects in Study 007 for CFQ-R RSS changes from Day 0 to Day 28 by subgroup based on imputed values in ITT patients. Table 12 shows that treatment differences (AZLI TID vs. Placebo TID) were significant or nearly significant in all of the subgroups with treatment effects above the MCID of 5 points. However, treatment differences were weakest in patients aged 18 and over. This subgroup of patients improved 4.82 points from baseline which was below the MCID of 5 points. Improvements relative to placebo in the CFQ-R RSS for patients aged 18 and over were also only marginal but were slightly more favorable at 6.35 points (0.02, 12.69), p=.0495.

Table 13 examines CFQ-R RSS changes from Day 0 to Day 42 by age group. At Day 42, AZLI TID patients improved 1.28 points from baseline and 6.33 points (1.22, 11.43) relative to placebo, p=.0154. However, at Day 42, patients aged 18 and over did not improve from their baseline and did not significantly improve relative to placebo. Improvements over placebo were 4.43 points (-1.61, 10.48), p=.1491.

Table 12: CFQ-R RSS Changes from Day 0 to Day 28 by Subgroup Based on Imputed Data (ITT)

Subgroup Category	**AZLI TID vs. Placebo TID Difference (95% CI)**	**P-value**
Disease Severity		
$FEV_1 \leq$ 50% predicted (n=60)	4.22 vs. -4.03 8.25 (-1.14, 17.64)	P=.0839
$FEV_1 >$ 50% predicted (n=103)	10.14 vs. -0.76 10.90 (4.16, 17.64)	P=.0018
Region		
U.S./ Canada (n=124)	5.89 vs. -1.47 7.36 (1.07, 13.65)	P=.0223
Australia/New Zealand (n=39)	11.65 vs. -5.64 17.29 (6.00, 28.58)	P=.0037
Age Group		
≥ 18 (n=126)	4.82 vs. -1.53 6.35 (0.02, 12.69)	P=.0495
< 18 (n=37)	12.73 vs. -6.19 18.91 (8.78, 29.05)	P=.0006
Gender		
Male (n=92)	5.09 vs. -2.27 7.36 (1.42, 13.29)	P=.0157
Female (n=71)	10.14 vs. 0.08 10.05 (0.21, 19.90)	P=.0455
Overall (n=163)	7.84 vs. -1.87 9.71 (4.31, 15.11)	P=.0005

Table 13: CFQ-R RSS Changes from Day 0 to Day 42 by Age Group Based on Imputed Data (ITT)

Age Group Category	**AZLI TID vs. Placebo TID Difference (95% CI)**	**P-value**
≥ 18 (n=126)	-0.24 vs. -4.68, 4.43 (-1.61, 10.48)	P=.1491
< 18 (n=37)	5.70 vs. -6.79, 12.48 (2.68, 22.29)	P=.0142
Overall (n=163)	1.28 vs. -5.05, 6.33 (1.22, 11.43)	P=.0154

6. APPLICANT'S NEW POST-HOC ANALYSES

The Applicant had submitted new post-hoc analyses to address Agency concerns regarding the regimen effect observed in the Study 005 primary analysis and the interpretability of findings. The Applicant's analyses attempted to show that the primary endpoint (i.e. time to need) was subject to a ' paradigm shift' in which patients with steep initial FEV_1 improvements (e.g. AZLI TID patients) were more likely to react to losses in FEV_1 relative to improved FEV_1 levels as opposed to absolute losses relative to baseline FEV_1 levels. This implied that AZLI TID patients had less severe time to need events than placebo patients which resulted in biases in time to need against AZLI TID.

The Applicant analyzed FEV_1 changes as a function of time to need to show that AZLI patients with events had improving FEV_1 from baseline not reflective of clinical worsening and placebo patients with events had worsening FEV_1 from baseline reflective of clinical worsening. Further analyses showed that FEV_1 slope, steepest in AZLI TID patients, had a significant effect on time to need comparisons. The Applicant's post-hoc analyses indicated more severe declines in FEV_1 in placebo arms (esp. if having an event). According to the Applicant, sensitivity analyses S_1 'Time to need: controlling for FEV_1 slope' and S_2 'Time to need: change 'event' to 'no-event' if FEV_1 improved from baseline' showed significantly delayed time to need in AZLI TID vs. Pooled Placebo.

Section 6.1 discusses some of the limitations of the Applicant's post-hoc analyses in assessing potential biases in the time to need endpoint. Section 6.2 provides some of the Reviewer analyses exploring the Applicant's notion of less severe AZLI TID events and potential biases in time to need.

6.1 Limitations of Applicant Analyses

The Applicant's new post-hoc sensitivity analyses made several problematic assumptions and failed to adequately clarify regimen effects in Study 005 findings. Due to the post-hoc nature of the Applicant's analyses, robustness of findings is essential in attempting to demonstrate potential biases in the primary analysis of Study 005 (i.e. a well controlled

confirmatory study). However, the Reviewer did not find the Applicant's analyses to be robust under varying assumptions. Applicant analyses failed to meaningfully influence primary analysis findings when considering appropriate assumptions. The following were some of the major limitations identified in the Applicant's post-hoc analyses.

The Applicant's re-analyses were conducted in a post-hoc manner and not based on pre-specified hypotheses, there are concerns that assumptions used in the analysis could have been determined retrospectively and primarily driven based on the significance of findings observed (after the data was unblinded).

In addition to the post-hoc nature of the analyses, the strength of evidence from the analyses was weak given the high degree of uncertainty in the primary analysis (e.g. regimen effects, dropouts). FEV_1 improvements over placebo were not significant at Day 28 (p=0.057) and had decreased rapidly after Day 28 (Figure 14). During the time period from Day 42 to Day 84 in which the majority of time to need events had occurred, FEV_1 improvements over placebo were negligible.

The Applicant's re-analyses rely on a problematic assumption that FEV_1 (% change) can be used as a post-hoc 'gold standard' to 'judge' for biases in the time to need endpoint. FEV_1 (% change) was not pre-specified as a primary endpoint in either Study 005 or Study 007. The association between FEV_1 & 'time to need' is also unclear since these endpoints measure different sets of symptoms in different manner. FEV_1 measures lung function (after bronchodilator use) for a few seconds in the 84 day study period whereas time to need measures a broader range of symptoms continuously over the entire 84 day study period. The association between time to need and FEV_1 is also not clear. CFQ-R RSS which measures similar symptoms as time to need has little correlation with FEV_1, r=0.24 in AZLI TID patients.

The Applicant's re-analyses assumed FEV_1 changes can be carried forward (LOCF) or represented by a linear trend using a mixed model repeated measures (MMRM) approach in imputing missing data. However, LOCF assumptions are problematic because changes in FEV_1 for AZLI TID vs. Pooled Placebo generally sloped downward representing a decreasing AZLI benefit over placebo over time (Figure 14). Since these downward slopes were also not constant over the Day 28-84 period, tending to be steepest from Day 28-42, an MMRM approach may not account for these steep declines accurately.

The Applicant's first sensitivity analysis, 'S_1: Time to Need Adjusting for FEV_1 Slope', inappropriately uses FEV_1 slope as a covariate. FEV_1 slope is treatment related (confounded) and controlling for this factor can result in serious biases in assessing the AZLI TID treatment effect. The covariates specified in S_1 (e.g. FEV_1 slope) also fail to meet any of the conditions recommended by ICH-E9 guidelines:

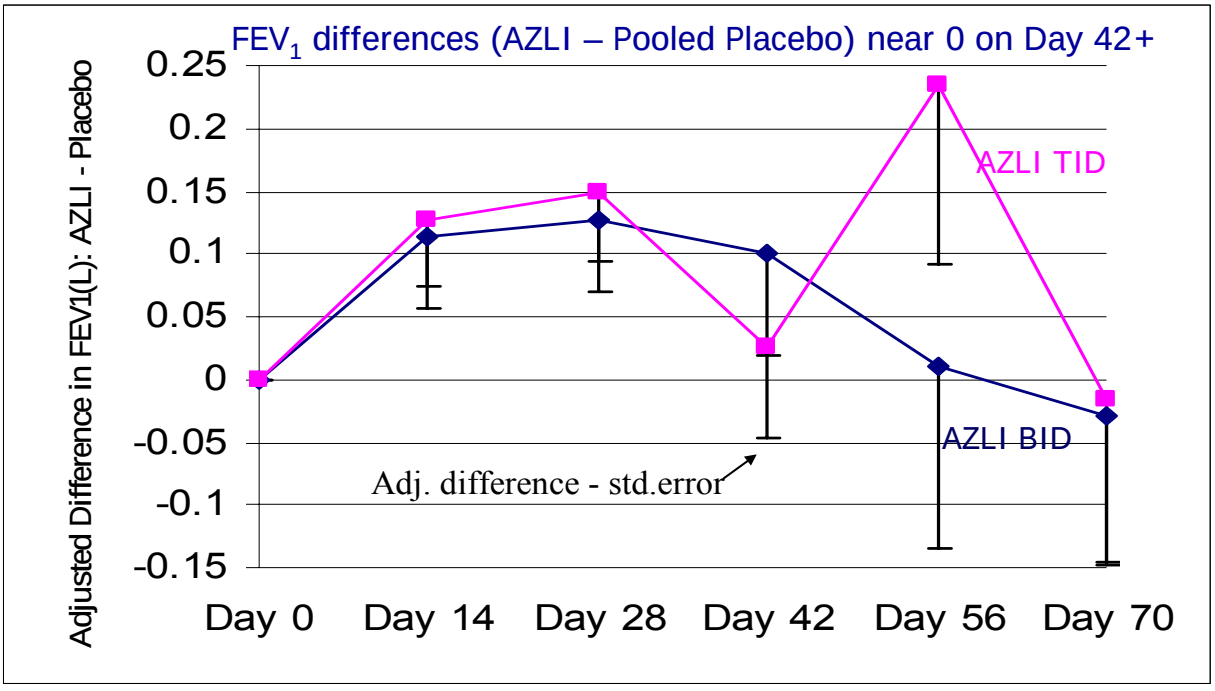
- Known to influence primary outcome (e.g. time to need)
- Pre-specified
- Used as stratification factors at randomization
- Measured prior to randomization
- Independent of treatment used

The Applicant's second sensitivity analysis, 'S_2: Changing 'event' to 'no event' if FEV_1 improved from baseline', inappropriately uses a Day -28 vs. Day 0 baseline. A Day -28 baseline is problematic in assessing time to need because the primary analysis of time to need was measured from Day 0. Additionally, FEV_1 changes from Day -28 combine AZLI/placebo effects from Day 0 to Day 84 and TOBI effects from Day -28 to Day 0 which may introduce biases in estimating AZLI TID effects since TOBI effects were most favorable in AZLI TID randomized patients. S_2 also changes 'event' to 'no event' if FEV_1 improved from baseline in hospitalized patients. It is unclear that patients with time to need events and improving FEV_1 did not have symptoms reflective of clinical worsening if they were also hospitalized.

6.2 Reviewer Analyses

The Reviewer conducted further analyses investigating possible differences in the severity of time to need events leading to potential biases in comparisons of time to need among AZLI and placebo treated patients. Based on the Reviewer analyses, there was no clear indication of less severe time to need events among AZLI treated patients that would lead to substantial biases affecting primary analysis findings.

Figure 17: FEV_1 Changes from Baseline in Patients with Events Based on Observed Cases (ITT)



In the above figure, we observe that among observed patients with events, FEV_1 differences (AZLI TID – Pooled Placebo) were near 0 at Day 42 and highly variable after

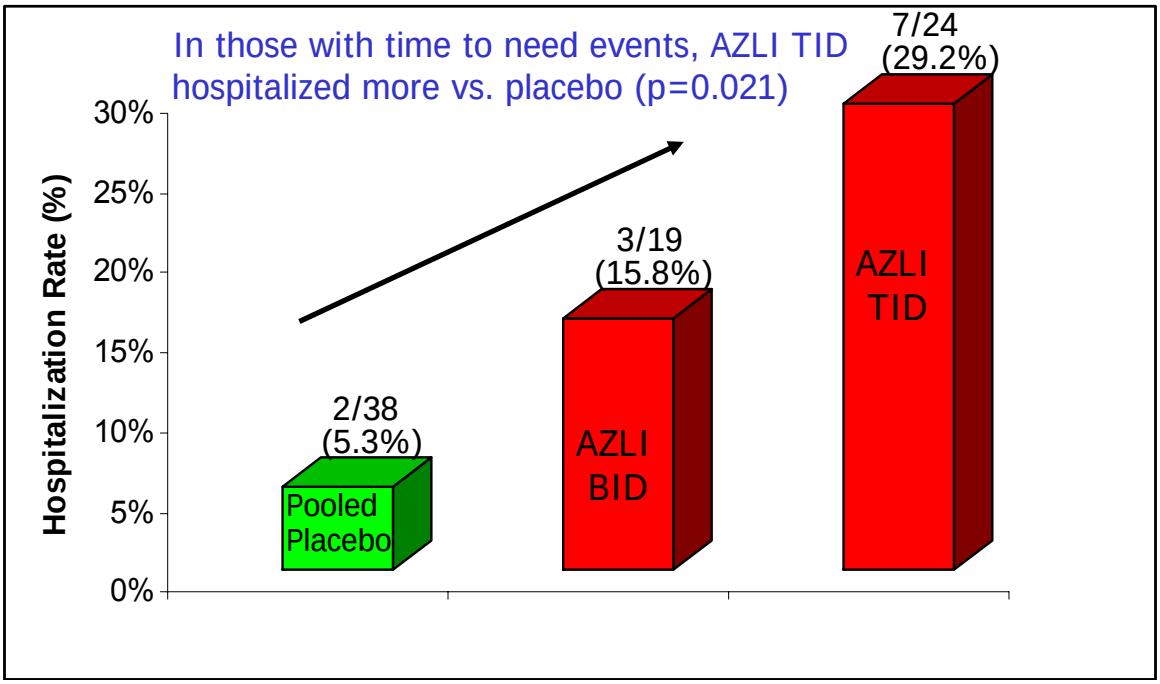
Day 42. Since the majority of time to need events occurred after Day 42, it is not clear that AZLI TID patients were having meaningfully higher FEV_1 at the time of the event that would lead to substantial biases affecting primary analysis findings.

Based on the Applicant's sensitivity analysis S_2 'Time to need Changing Event to No-event if FEV_1 Improved from Baseline', the Reviewer conducted a similar analysis (S_2^*) using a Day 0 vs. Day -28 baseline and excluding hospitalized patients from event re-classifications. In S_2^*, 18/81 events (1 Placebo BID, 7 Placebo TID, 5 per AZLI arm) were censored which did not affect primary analysis findings (Table 14). Like the primary analysis, S_2^* showed significant TID vs. BID regimen effects (p=.0086) with AZLI TID patients faring worse than Placebo BID patients (p=.9280). However, compared with primary analysis results, S_2^* results also showed less significant regimen effects in placebo patients (p=.0413 vs. p=.0043).

Table 14: S_2*- Time to need Changing 'Event' to 'No-event' if FEV_1 Improved from Baseline (ITT)

AZLI vs. *Placebo* (p value)		*Placebo BID*	*Placebo TID*	*Pooled Placebo*
AZLI BID	(1°)	.4269	< .0001	.0019
	(S_2*)	.1993	< .0001	.0025
AZLI TID	(1°)	.5377	.0043	.1816
	(S_2*)	.9280	.0233	.2020
Pooled AZLI	(1°)	.9240	< .0001	.0070
	(S_2*)	.5188	.0002	.0090
		AZLI:	**Placebo:**	**Pooled:**
BID vs. TID	(1°)	.0835	.0043	.0012
	(S_2*)	.1012	.0413	.0086

Figure 18: Hospitalization Rates in Patients with Time to Need Events (ITT)



In Figure 18, AZLI TID patients with time to need events had significantly higher rates of early hospitalization (p=.021) than Pooled Placebo patients. This shows a dose response relationship between AZLI use and hospitalization. Those patients taking AZLI BID (150 mg AZLI daily) vs. placebo had 3 times the hospitalization rate (15.8% vs. 5.3%) while those patients taking AZLI TID (225mg AZLI daily) vs. placebo had more than 5.5 times the hospitalization rate (29.2% vs. 5.3%). Figures 37-38 of the Appendix also show that AZLI TID patients had a significantly shorter time to early termination and had spent significantly more days hospitalized.

Among patients with time to need events, our analyses of FEV_1 changes and hospitalization rates failed to show less severe AZLI TID vs. placebo events nor show substantial biases in time to need against AZLI TID, especially biases that could meaningfully influence primary analysis findings. In fact, AZLI TID patients with events had significantly greater hospitalization rates than placebo patients with events suggesting greater rather than lesser severity of AZLI TID patient events. This may result in possible biases in the primary analysis of time to need favoring rather than hindering AZLI TID. Regardless of these findings, any possible biases in the time to need endpoint hindering AZLI TID would appear relatively small in comparison to the potential biases from robust regimen effects and dependent missing data favoring AZLI TID. Therefore, the Applicant's re-analyses would not be expected to meaningfully influence primary analysis findings.

7. ADVISORY COMMITTEE MEETING

An AIDAC was held on December 10, 2009 to discuss the issues and the committee members voted in favor of AZLI 75mg TID (15 to 2) as being safe and effective. However, concerns were expressed by committee members regarding the evidence presented from Study 005 and it remains unclear to the Reviewer as to whether Study 005 was viewed as an adequate and well controlled study supporting approval.

The Reviewer gave a presentation at the AIDAC based on findings of this review. However, one of the 17 AIDAC panelists questioned the rationale of some of the assumptions used in the statistical analyses. The Reviewer has provided detailed responses to this AIDAC panelist's comments as outlined below. After further consideration of the AIDAC panelist's comments, the Reviewer maintains that the statistical approaches used in this statistical review and the AIDAC presentation were most appropriate.

The AIDAC panelist suggested that an advantageous statistical approach for evaluating efficacy in Study 005 would be to consider a two factor factorial design with regimen (TID or BID) and drug (AZLI or placebo) as the two factors in the design. The panelist commented that using this factorial approach "you do get a strong regimen effect, but you still get a strong treatment effect'. The Reviewer disagrees with these comments for the following reasons:

Given the nature of the data and issues discussed earlier, a factorial design may not be informative due to its limitations with estimating main effect parameters. Additionally, a factorial design may fail to address the question of interest which is whether there is a significant benefit from use of an AZLI TID treatment.

- Factorial designs assume factors used in the model are independent. In Study 005, the 'regimen' and 'drug' factors, as proposed under a two factor factorial design, are not independent. The regimen factor is dependent on whether the patient is taking placebo or AZLI. If taking placebo, a regimen change from BID to TID measures the TID vs. BID regimen effect. However, if taking AZLI, a regimen change from BID to TID measures the combined effect of both a TID vs. BID regimen increase and the AZLI drug increase (150mg daily to 225 mg daily). Similarly, the drug factor depends upon whether the patient is on a BID or a TID regimen (i.e. if on BID, factor is 150 mg AZLI, if on TID factor is 225mg AZLI). These dependencies are problematic in interpreting the main effect estimates of the factorial design across all ITT patients.

- In addition to dependencies obscuring main effect interpretations, main effect estimates even if interpreted properly would not assess the significance of the AZLI TID treatment effect. Significant main effect estimates only indicate a significant drug effect controlling for regimen and a significant regimen effect controlling for AZLI drug. We cannot infer from this that there is a significant

AZLI TID treatment. This is because the AZLI TID treatment effect is not just the drug effect but rather a combined effect of both drug and regimen (i.e. beneficial 225mg AZLI drug + detrimental TID regimen) Note that the only comparison which assesses the AZLI TID treatment effect, fully accounting for both the drug and regimen effects, is AZLI TID vs. Placebo BID.

The panelist further commented that the 'AZLI TID vs. Placebo BID' comparison is underpowered and later recommended the pre-specified 'AZLI TID vs. Placebo TID' and 'AZLI BID vs. Placebo BID' comparisons. The Reviewer disagrees with these comments:

- Actually, the three above comparisons all include n=104 or n=107 ITT subjects with similar expected power. While these comparisons use approximately half the original ITT sample size, all would be expected to at least trend towards significance assuming an AZLI treatment benefit. However, 'AZLI TID vs. Placebo BID' clearly failed to approach significance (p=.9999) and even favored Placebo BID (p=.5377). Due to the high degree of failure in this comparison, the ability to show an AZLI TID benefit was unlikely to be affected by modest losses in power. In contrast, the ability to show a regimen effect (i.e. Placebo BID vs. Placebo TID) in secondary analyses such as the CFQ-R RSS at Day 28 would be more likely to be affected by modest losses in power. Note that in this comparison, there was a trend towards significance and greater limitations in subjects (76 vs. 104 patients).

- In comparison to AZLI TID vs. Placebo BID, the Placebo BID vs. Placebo TID comparison which assesses the regimen effect included only 76 vs. 104 (or 27% fewer) ITT subjects but showed a highly significant result favoring Placebo BID (p=.0043). Since AZLI treatment effects are presumably larger than regimen effects, we would expect at least a trend towards significance in the AZLI TID vs. Placebo BID comparison, especially with more subjects.

- The Agency had recommended that the Applicant have adequate power for showing an AZLI TID or AZLI BID treatment benefit to meet labeling requirements. Indeed, AZLI TID vs. Pooled Placebo (n=142) was included as a primary analysis comparison. Although AZLI TID vs. Placebo BID (n=104) comparisons based on 38/142 (26.7%) fewer subjects did result in a modest loss of power, this loss was expected based on the protocol which indicated that under the scenario of differences in the placebo arms, placebo arms would have to be considered individually.

- The Reviewer believes that the AZLI TID vs. Placebo BID comparison is the more appropriate comparison given the regimen effect observed in Study 005. Since AZLI TID requires use of a less favorable TID regimen, regimen effects have to be accounted for in the Study 005 primary analysis in order to draw

meaningful inferences. Only AZLI TID vs. Placebo BID fully accounts for the combined effect of both the AZLI vs. Placebo drug and the TID vs. BID regimen.

- AZLI TID vs. Placebo TID and AZLI BID vs. Placebo BID comparisons are not informative in assessing AZLI TID effects because they ignore the regimen effect, estimating only AZLI drug effects, and appear to be logically inconsistent. For example, they show AZLI BID as not significant and AZLI TID as significant. This is problematic inference because AZLI BID was shown to be more effective than AZLI TID across most endpoints and was clearly more favorable in primary and sensitivity analyses with significantly lower rates of early termination.

- AZLI TID vs. Placebo TID represents a more liberal comparison than the planned comparison of AZLI TID vs. Pooled Placebo. In the event of greater variability in the placebo arms, a more conservative comparison should be used, not a more liberal comparison.

- The panelist recommended the AZLI TID vs. Placebo TID comparison having earlier stated that he thinks there is a regimen effect. Assuming a regimen effect, Placebo TID is then inferior to the true placebo and AZLI TID vs. Placebo TID would fail to provide a valid or unbiased estimate of the AZLI TID treatment benefit.

- It is not clear to the Reviewer that AZLI TID vs. Placebo TID and AZLI BID vs. Placebo BID were pre-specified comparisons as stated by the panelist. The Study 005 SAP only made a general statement that "If there was a strong suggestion that the two placebo groups were different, they could have been analyzed separately as sensitivity analyses."

The panelist further commented that the 'AZLI TID vs. Placebo BID' comparison is basically inappropriate because the blinding is different.

- Since investigators were not blinded to the regimen received (BID or TID), there may be different sets of biases associated with knowing a subject is on a BID vs. TID regimen. However, these differences in biases should be minimal relative to the treatment effect in a well conducted study. The Reviewer considers the blinding issue as relatively minor in comparison to failing to fully account for the regimen effect.

- Most Study 005 comparisons performed for individual AZLI arms were against Pooled Placebo and also involved different blinding. The panelist's comments would therefore imply that these comparisons are also inappropriate and that only Pooled vs. Pooled or within regimen comparisons could be performed.

The panelist indicated that the Reviewer overstated the extent to which the time to need endpoint suffered from some substantial missing data by counting patients dropping out with an event as missing. The Reviewer disagrees.

- When analyzing potential biases due to missing data in the primary analysis, the Reviewer classified only those patients dropping out without a time to need event as missing (Section 3.3.2). Furthermore, all sensitivity analyses of time to need had made missing data imputations only for dropouts without a primary analysis time to need event.

- When considering overall rates of missing data in Study 005, the Reviewer considered dropouts with or without time to need events as missing because they can introduce biases in all other study analyses. However, when considering missing data in the primary analysis, dropouts with primary analysis time to need events were not considered as 'missing' because they still satisfied the time to need outcome. The Reviewer acknowledges that the rates of overall missing data can lead to false inferences about rates of missing data in the primary analysis. However, the distinction between overall missing data and primary analysis missing data had been made in the presentation and is made in this review.

8. SUMMARY AND CONCLUSIONS

Study 005 was an uninformative study providing very limited supportive evidence. Study 005 primary analyses failed to demonstrate the efficacy of AZLI TID in prolonging time to need for IV or inhaled anti-pseudomonal antibiotics due to predefined symptoms predictive of pulmonary exacerbations. There was also substantial dependent missing data significantly favoring the TID vs. BID regimens as well as AZLI vs. Placebo during the TOBI run-in period. Among randomized patients without a primary analysis time to need event, AZLI TID patients had the highest early termination rate that was substantially higher than the rate for Placebo BID patients (58.6% vs. 37.9%). Since there is no evidence to suggest that missing data was uninformative or not treatment related, there is potential for biases in the primary analysis favoring AZLI TID.

In Study 005, there was additional uncertainty due to a detrimental TID vs. BID regimen effect which was more influential than the AZLI drug effect in the primary analysis. Due to this uncertainty, AZLI TID comparisons against Pooled Placebo or Placebo TID would be problematic since these placebo groups may fail to reliably estimate the true placebo rate due to a detrimental TID regimen effect. Therefore, AZLI TID vs. Placebo BID was considered to be the most informative comparison. In the primary analysis, AZLI TID patients fared worse than Placebo BID patients (p=.5377). Furthermore, in sensitivity analyses, AZLI TID comparisons against Placebo BID were even less favorable than in the primary analysis due to a highly robust regimen effect that was consistently stronger than the AZLI drug effect. In a sensitivity analysis using a broad definition for 'event' to control for potential biases from informative dropouts (i.e. time to early termination), AZLI TID patients fared substantially worse than Placebo BID patients, p=.0941 (Table 3). In other Study 005 analyses, AZLI TID patients fared similar to or possibly worse than placebo over longer time periods. For example, over the entire 84 day AZLI/placebo study period, AZLI TID patients had a substantially shorter 'time to hospitalization' compared to Pooled Placebo patients (p=.085). Although the Applicant did provide post-hoc analyses in this submission attempting to explain biases in the primary endpoint of 'time to need' using FEV_1 findings, these analyses were not considered to be adequate especially given a highly significant and robust regimen effect compounded with dependent missing data.

In Study 005, the CFQ-R RSS and FEV_1 endpoints showed marginal AZLI TID benefits when considering only the Day 0 to Day 28 time period, however this evidence was limited by dependent missing data (both before and after AZLI/placebo treatment), regimen effects in the CFQ-R RSS, multiplicity issues, an unclear primary analysis and other unfavorable study findings. CFQ-R RSS changes at Day 28 for AZLI TID vs. Pooled Placebo also failed to meet the minimum clinically important difference (MCID) of ± 5 points at 4.37 (-0.94, 9.69), p=.1061, while changes (worsening) from Placebo TID vs. Placebo BID did meet the MCID at -5.66 (-11.92, 2.61), p=.2078. Therefore a substantial portion of the observed AZLI TID treatment benefit over Pooled Placebo may be due to detrimental TID regimen effects included in the Pooled Placebo estimate as well as potential biases from higher rates of missing data in AZLI TID patients. Sensitivity analyses controlling for these factors also failed to show any trend towards an

AZLI TID benefit in CFQ-R RSS at Day 28. Furthermore, after Day 28, the CFQ-R RSS was not assessed until the Day 84 visit at which AZLI TID patients fared significantly worse than placebo patients based on imputed data, p=.020 (Table 6).

Study 007 demonstrated its primary endpoint, changes in the CFQ-R RSS from Day 0 to Day 28, showing an improvement over placebo of 9.71 points (95% CI: 4.31, 15.11) which reduced to 6.33 points (95% CI: 1.22, 11.43) at the Day 42 visit. Although Study 007 provided some evidence of sustained improvement in respiratory symptoms, this evidence was limited because AZLI TID mean improvements from baseline had dropped substantially from Day 28 (7.08 points) to Day 42 (0.62 points). These findings were further limited by uncertainty in the validity and reliability of the CFQ-R RSS instrument in demonstrating clinical improvement based on a pre-defined MCID of a 5 point increase. There were also limitations in the robustness of primary analysis findings. For example, the CFQ-R RSS at Day 28, was primarily driven by patients <18 years of age (n=37) who had substantially larger improvements over placebo compared to patients ≥18 years of age (n=126) at 18.92 (95% CI: 8.78, 29.05), p=.0006 vs. 6.35 (95% CI: 0.02, 12.69), p=.0495.

In summary, this submission fails to provide adequate evidence of reduced pulmonary exacerbations in CF patients. The time to need for IV or inhaled antibiotics due to predefined symptoms and time to first hospitalization endpoints both failed to show significant AZLI TID benefits in each of the studies. However, there is some evidence of an improvement in respiratory symptoms from Day 0 to Day 28 from Study 007, but this evidence is limited in patients aged 18 years and older. Due to the nature of Cystic Fibrosis due to *P. aeruginosa*, clinical and other considerations should also be taken into account when evaluating the evidence of improvement in respiratory symptoms.

We recommend that additional studies are conducted to better address many of the limitations stated in this Review. These limitations often related to observed regimen effects, regimen dependent dropout rates, respiratory improvements in patients in the < 18 and ≥ 18 year age groups and sustained AZLI treatment effects beyond Day 28 and over multiple cycles.

SIGNATURES/DISTRIBUTION LIST

Primary Statistical Reviewer: Christopher Kadoorie, Ph.D.

Date: 1/27/2010

Concurring Reviewer: Thamban Valappil, Ph.D, Statistical Team Leader

cc:
HFD-520/Project Manager/ Kyong Hyon, B.S.N., M.A.
HFD-520/Medical Team Leader/John Alexander, M.D., M.P.H.
HFD-520/Medical Reviewer/Menfo Imoisili, M.D., M.P.H
HFD-520/Deputy Division Dir/ Katherine Laessig, M.D.
HFD-520/Acting Division Dir/ Wiley Chambers, M.D.
HFD-725/Primary Statistical Reviewer/ Christopher Kadoorie, Ph.D.
HFD-725/Statistical Team Leader/ Thamban Valappil, Ph.D.
HFD-725/Biometrics Deputy Division Director/ Daphne Lin, Ph.D.
HFD-725/Biometrics Division Director/ Mohammad Huque, Ph.D.
HFD-700/Mathematical Statistician/Lillian Patrician M.S., M.B.A

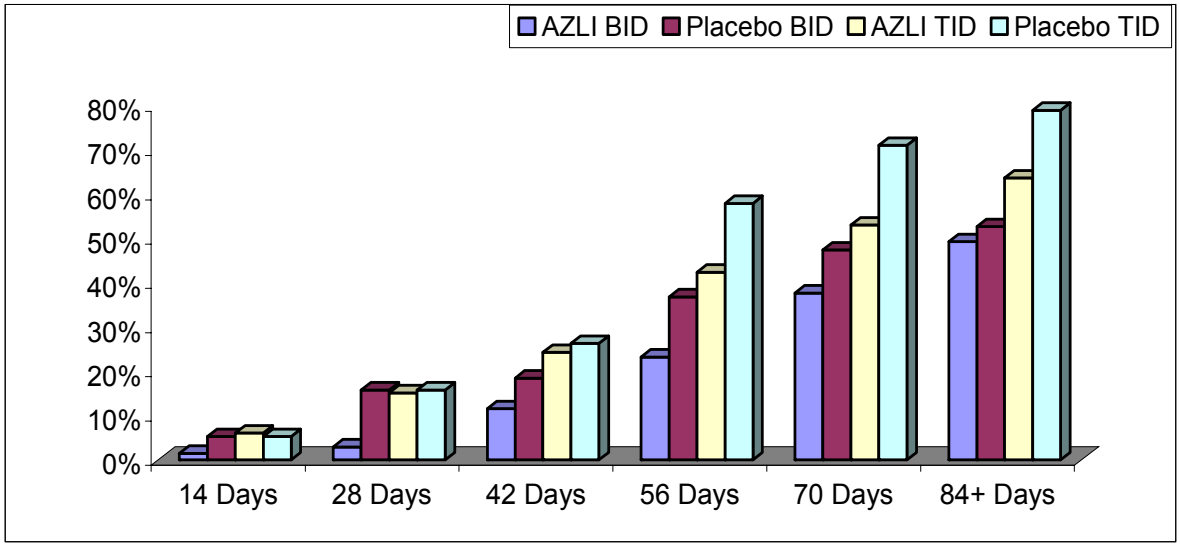
APPENDIX

The Appendix includes figures that were referenced but not included in the main body of the Review. These figures are listed in the Appendix according to the section of the review in which they were referenced.

Figures from 3.3.1: Early Termination Rates

Figures 19-21 show early termination rates and time to early termination in ITT patients for the four treatment arms. Figure 19 shows that early termination rates were highest in the TID regimens especially for later visits. Figure 20 shows a shorter time to early termination in TID vs. BID regimens (p=.0023). Figure 21 shows that AZLI TID patients had shorter time to early termination in comparison to treatments using a BID regimen. Time to early termination was shorter for AZLI TID vs. AZLI BID (p=.0268) and AZLI TID vs. Placebo BID (p=.3217).

Figure 19**: Early Termination Rates (ITT Patients)**



Early Termination Rates (Days after Start of Study Therapy)

Figure 20: Time to Early Termination in BID vs. TID Regimens (ITT)

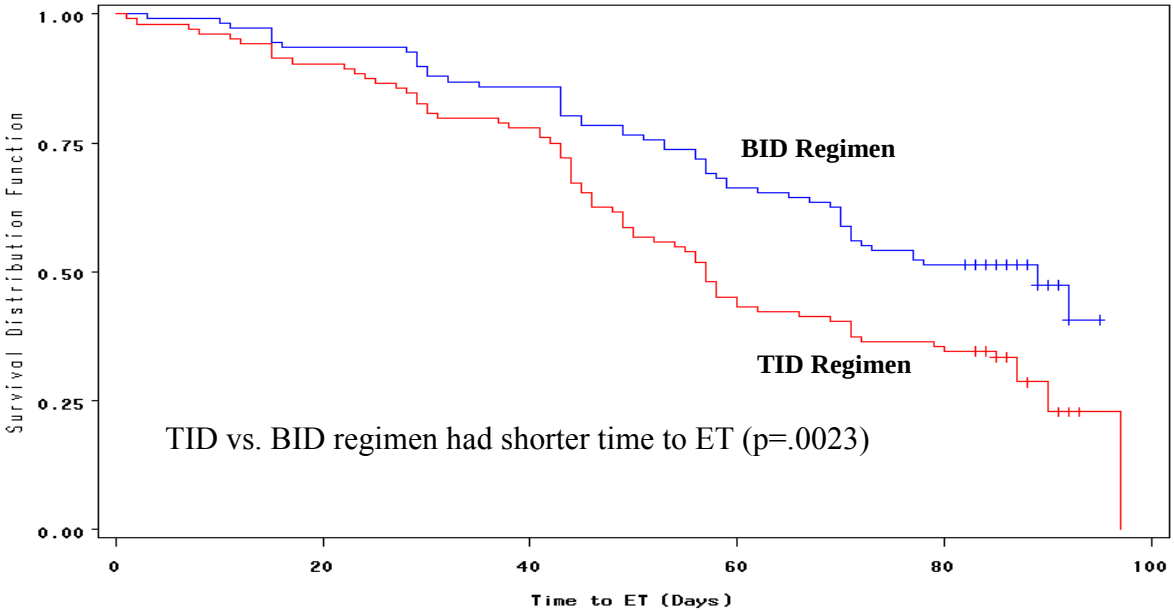
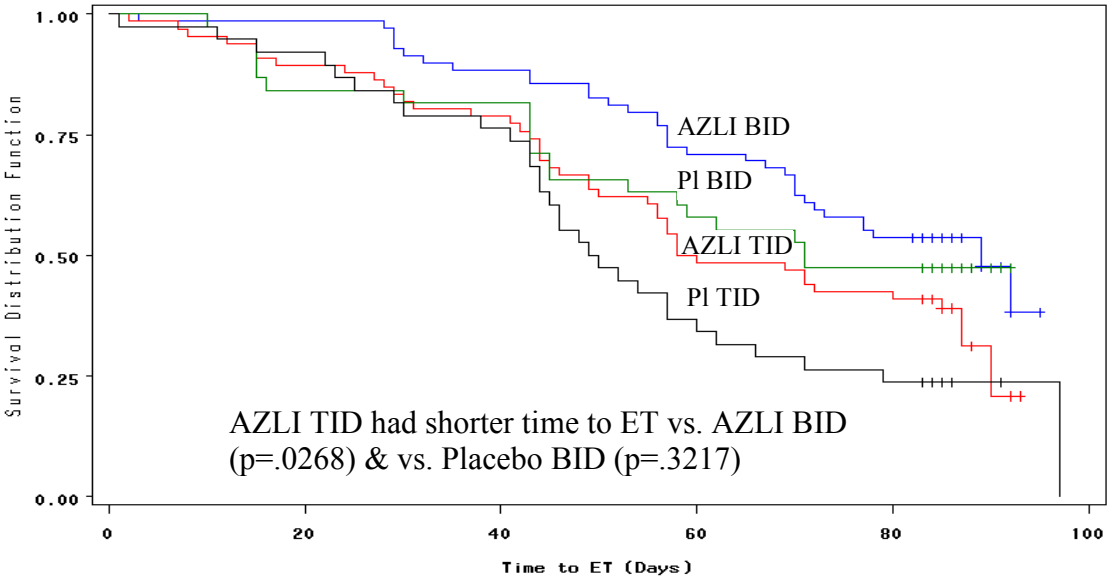


Figure 21: Time to Early Termination in Treatment Arms (ITT)



Figures 22-23 show for randomized patients shorter time to early termination due to an AE or treatment intolerance in TID vs. BID regimens (p=.0003), AZLI TID vs. AZLI BID (p=.0059) and AZLI TID vs. Placebo BID (p=.1022).

Figure 22: Time to ET due to an AE or Treatment Intolerance by Regimen (All Randomized)

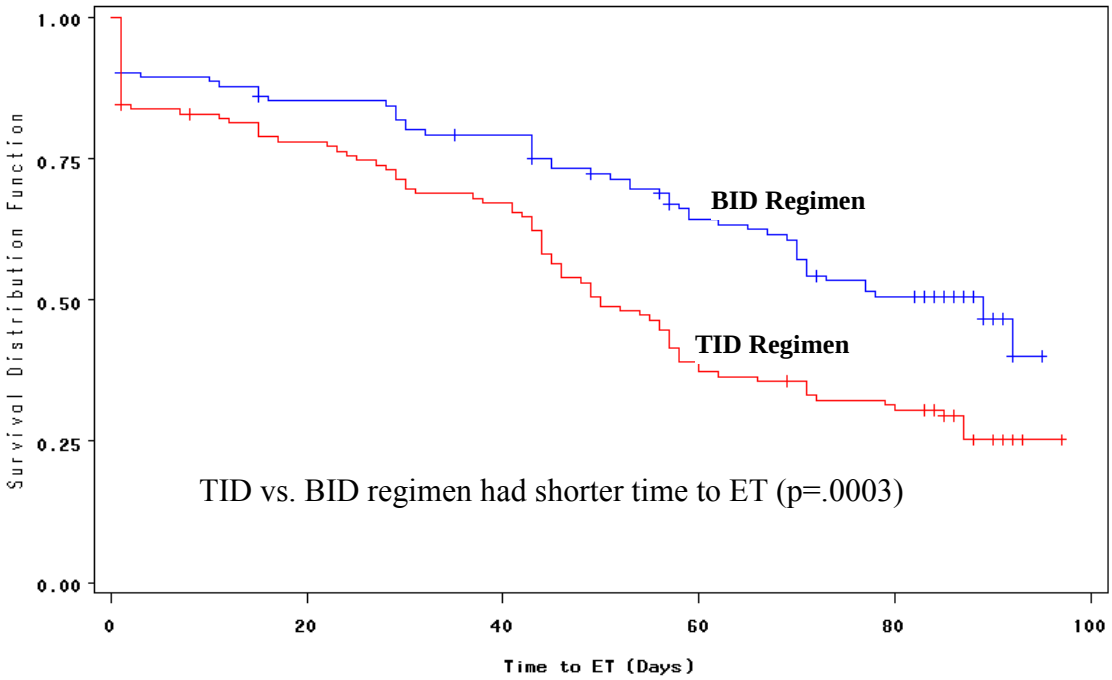
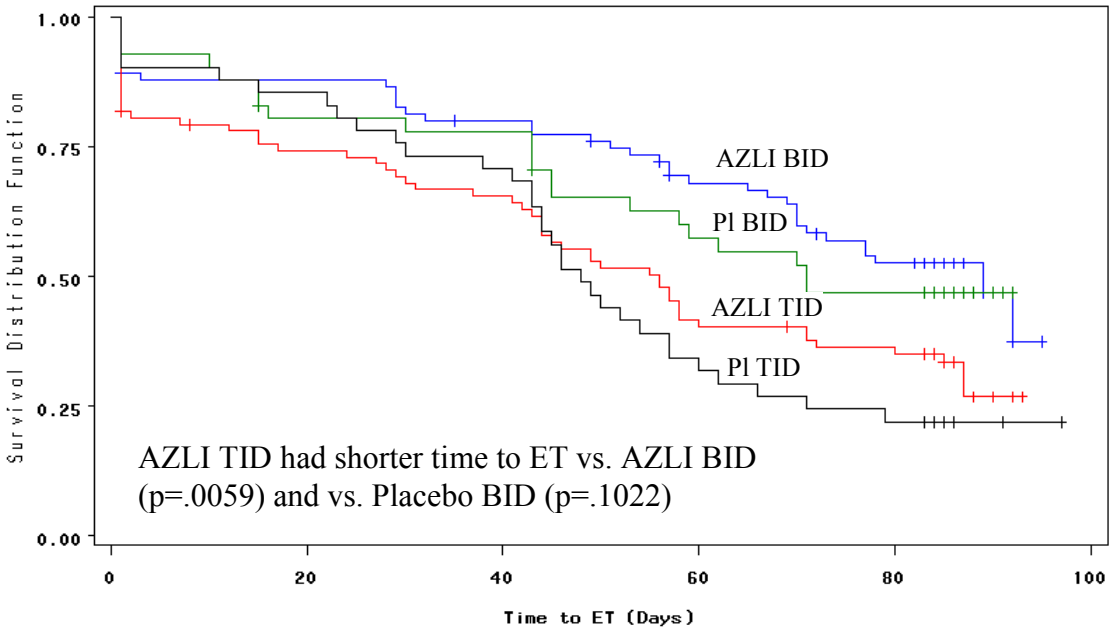


Figure 23: Time to ET due to an AE or Treatment Intolerance (All Randomized)



Figures from 3.3.2: Early Termination Rates (Patients without Events)

Early Termination Rates for Time to need (primary endpoint)

Figures 24-26 shows early termination rates in ITT patients without a time to need event were relatively high in TID regimens and AZLI TID. Figure 24 shows higher rates of early termination in the TID regimen. Figure 24 also shows that AZLI TID had the highest ET rate up to 42 days. Figures 25-26 show shorter time to ET in TID vs. BID regimens (p=.1319), AZLI TID vs. AZLI BID (p=.1082) and AZLI TID vs. Placebo BID (p=.3344). Differences in this subgroup were less pronounced than in the entire subject population but may suggest modest potential biases.

Figure 24: ET Rates in Patients without a Time to Need Event (ITT)

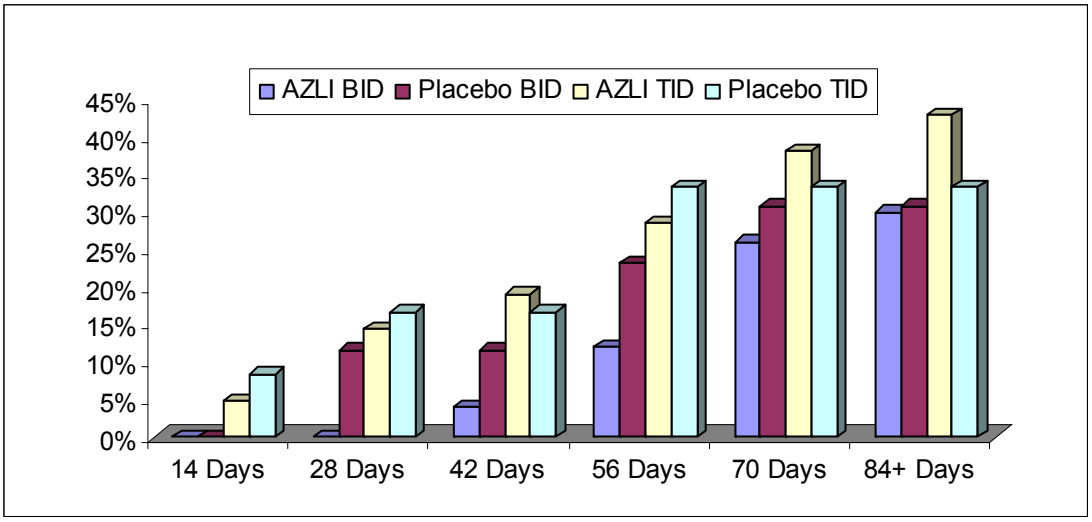


Figure 25: Time to ET in Patients without a Time to Need Event by Regimen (ITT)

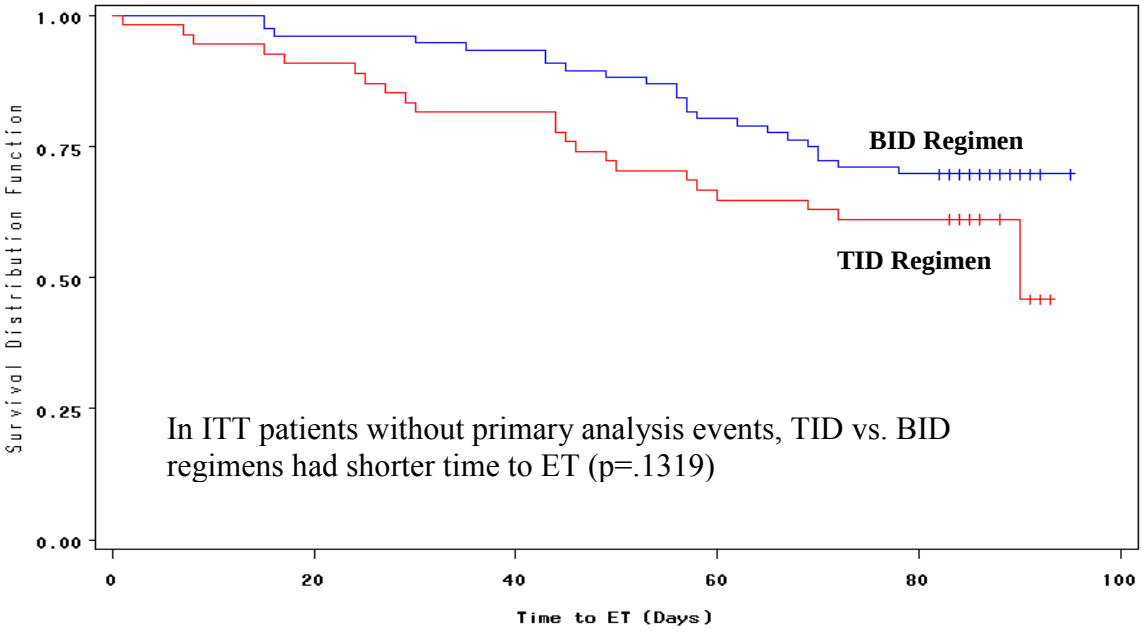
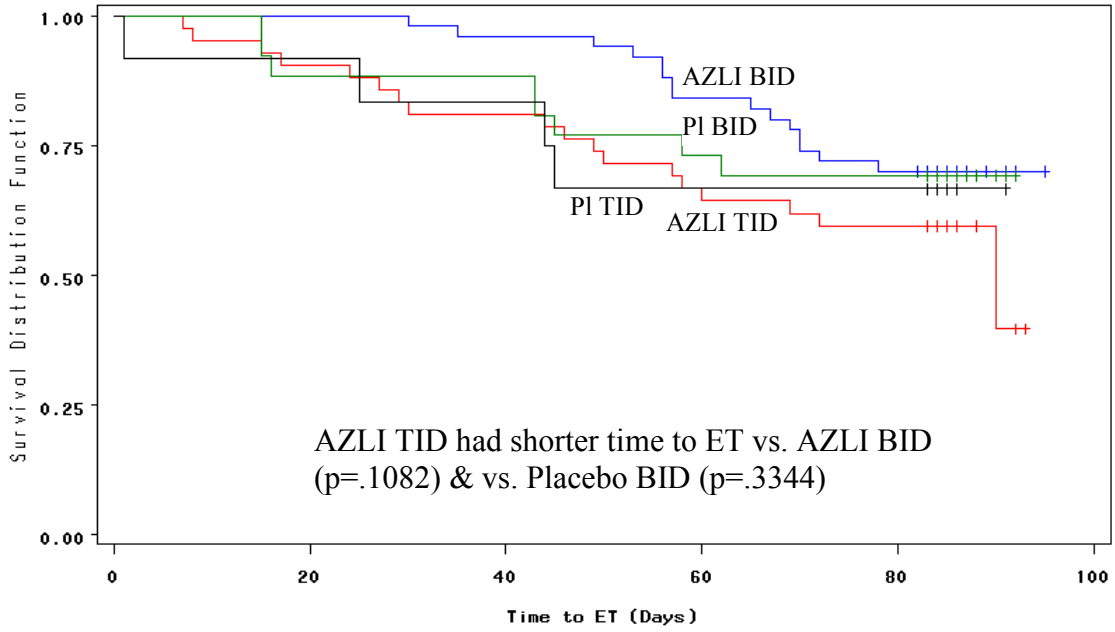


Figure 26: Time to ET in Patients without Time to Need Event (ITT)



Figures 27-28 show for randomized patients without a primary analysis time to need event shorter time to ET due to an AE or treatment intolerance in TID vs. BID regimens (p=.0090), AZLI TID vs. AZLI BID (p=.0189), AZLI TID vs. Placebo BID (p=.0672).

Figure 27: Time to ET due to an AE or Treatment Intolerance in Patients without a 1° Analysis Time to Need Event by Regimen (All Randomized)

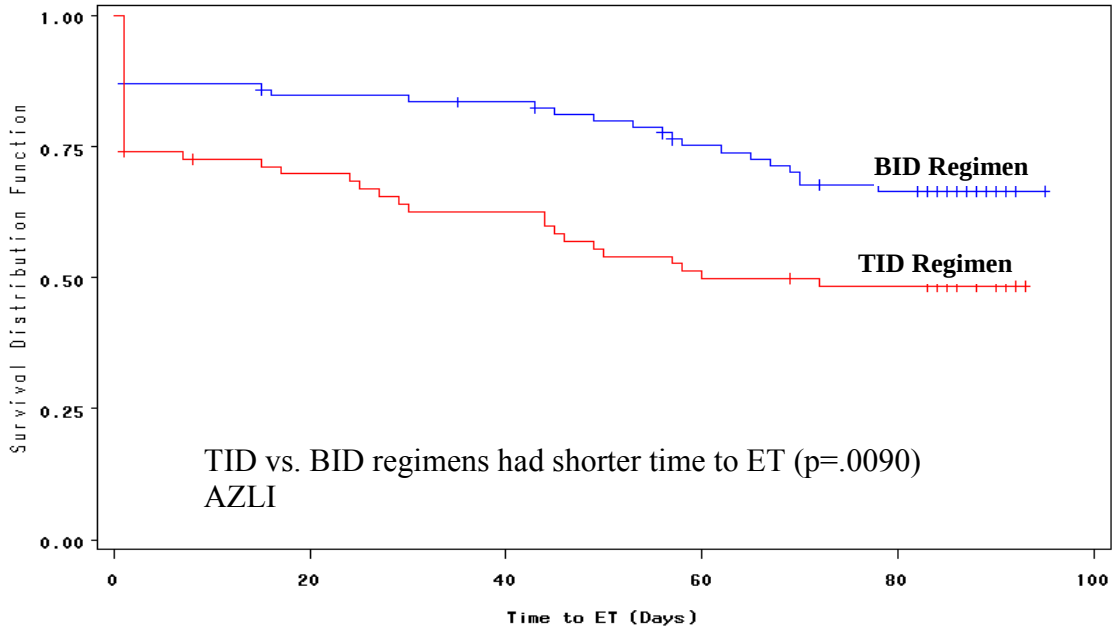
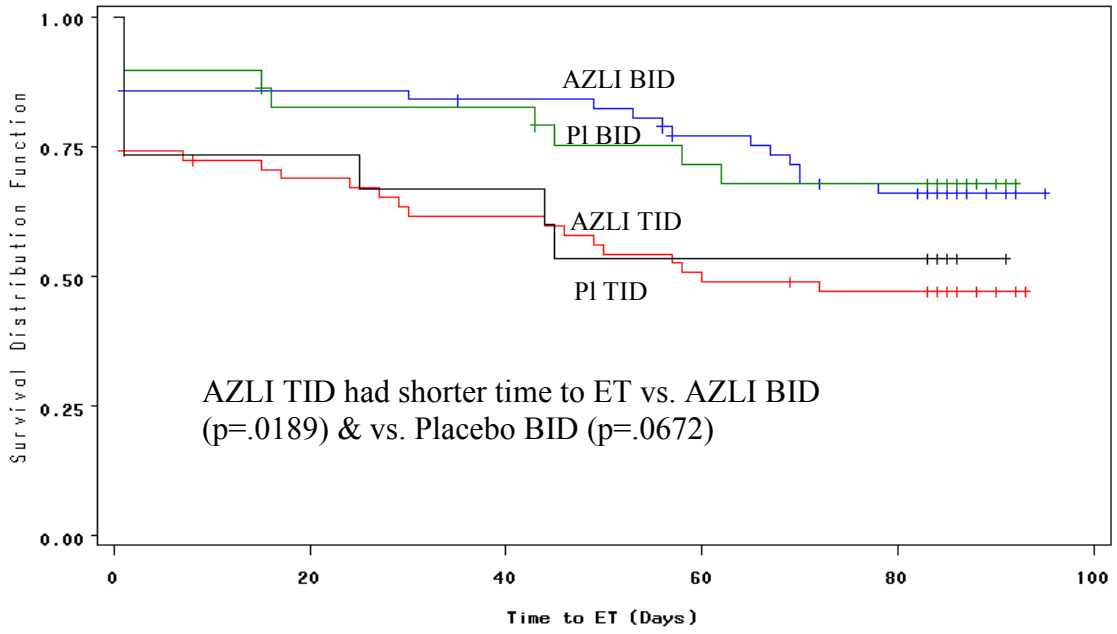


Figure 28: Time to ET due to an AE or Treatment Intolerance in Patients without 1° Analysis Time to Need Event (All Randomized)



Early Termination Rates for Time to First Hospitalization Endpoint

In Figures 29-30, comparisons in randomized patients without a hospitalization event during the AZLI/Placebo period showed shorter time to early termination in those using a TID vs. BID regimen: TID vs. BID (p= .0079), AZLI TID vs. AZLI BID (p=.0533) and for AZLI TID vs. Placebo BID (p= .1667).

Figure 29: Time to ET in Patients without a Hospitalization Event in AZLI/Placebo Period by Regimen (All Randomized)

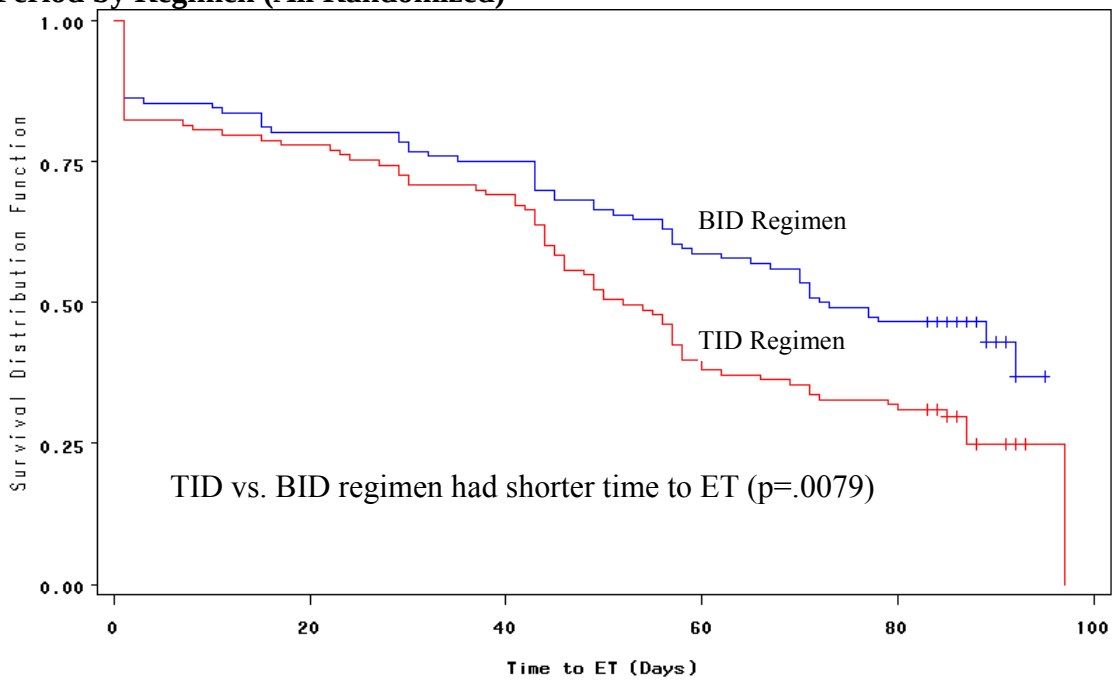
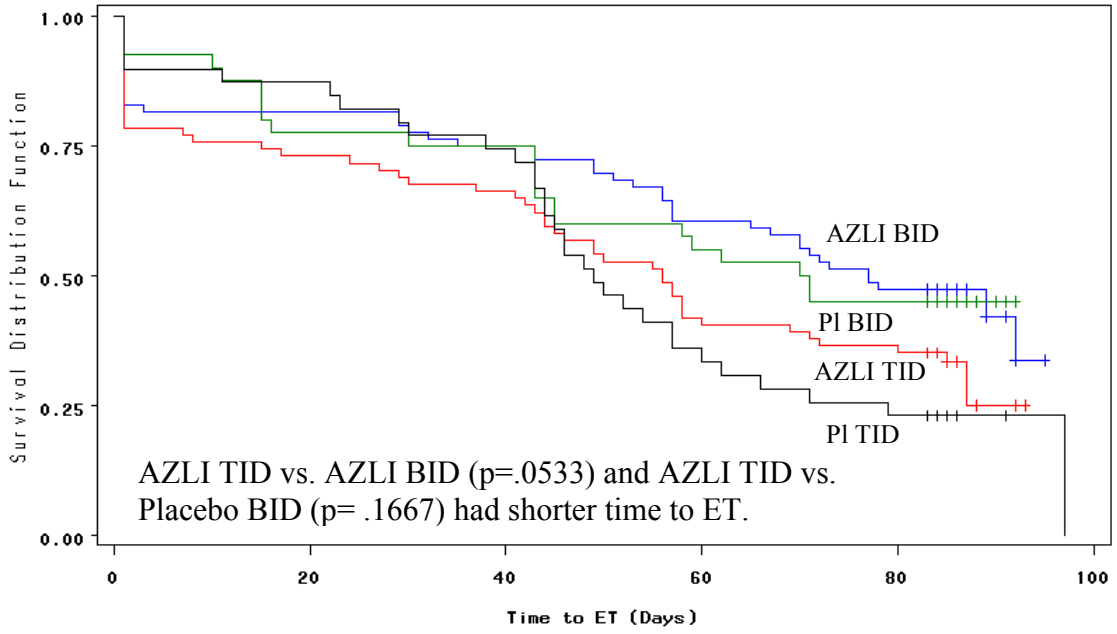


Figure 30: Time to ET in Patients without a Hospitalization Event in AZLI/Placebo Period (All Randomized)



Figures 31-32 show comparisons of early termination in ITT patients without a hospitalization event: TID vs. BID (p= .0076), AZLI TID vs. AZLI BID (p=.0653) and for AZLI TID vs. Placebo BID (p=.5961).

Figure 31: Time to ET in Patients without Hospitalization Event by Regimen (ITT)

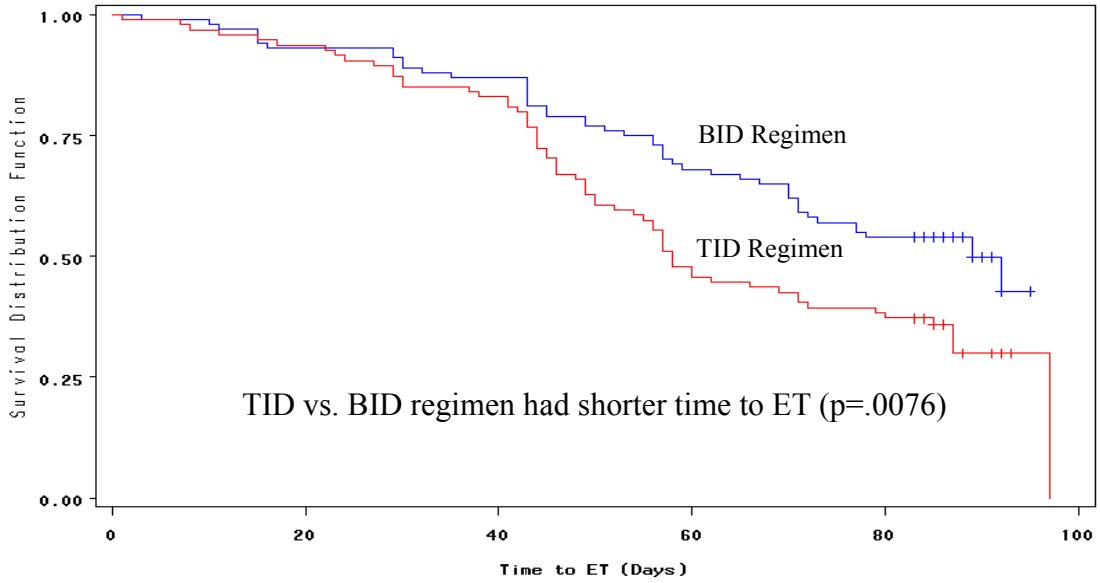
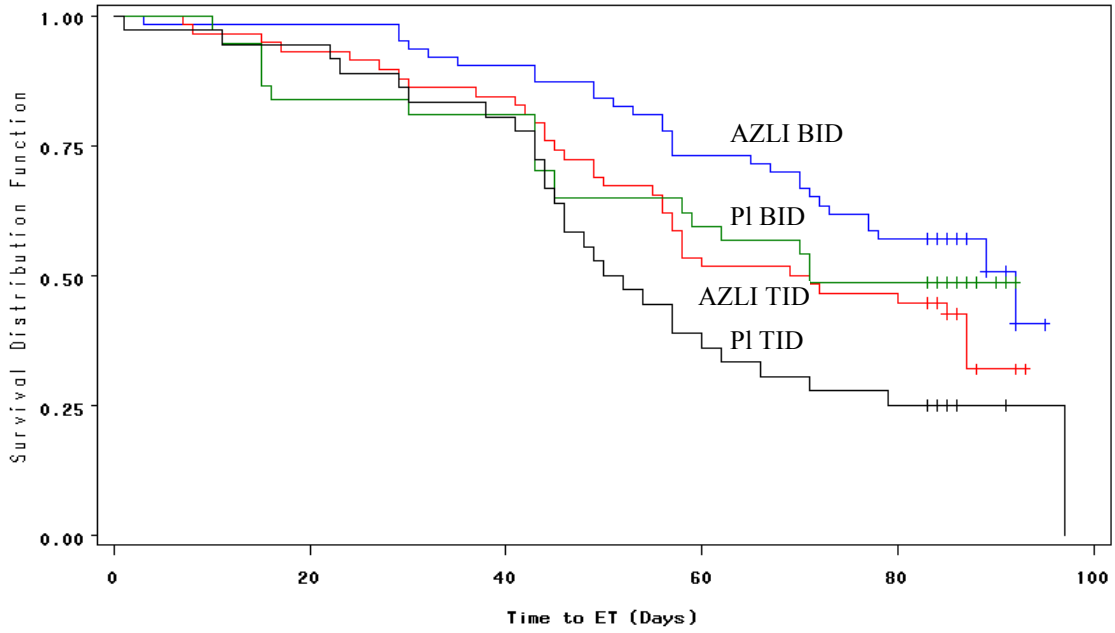


Figure 32: Time to ET in Patients without Hospitalization Event (ITT)



Figures 33-34 show comparisons of early termination due to an AE or treatment intolerance in randomized patients without a hospitalization event: TID vs. BID (p= .0076), AZLI TID vs. AZLI BID (p=.0653) and for AZLI TID vs. Placebo BID (p=.5961).

Figure 33: Time to ET due to an AE or Treatment Intolerance in Patients without a Hospitalization Event in the AZLI/Placebo Period by Regimen (All Randomized)

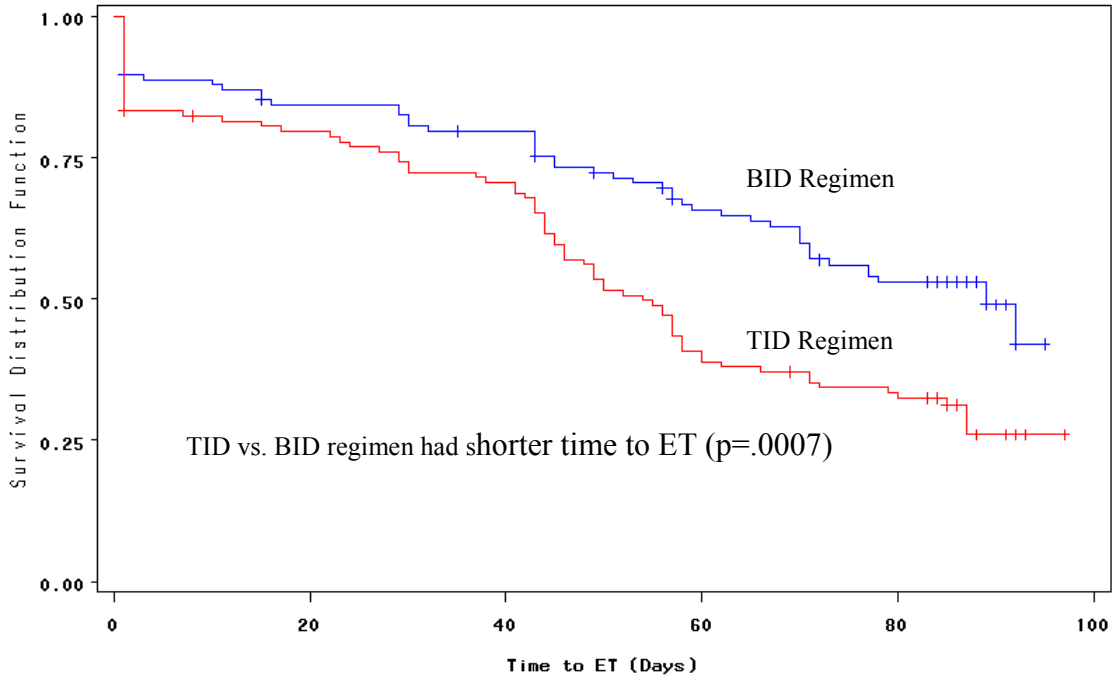
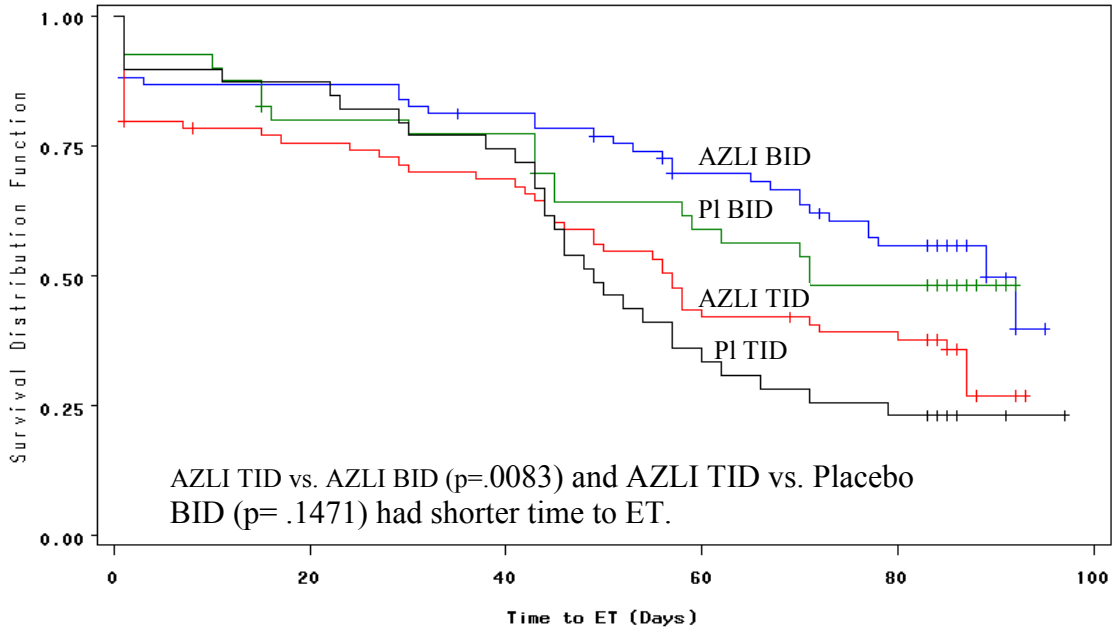


Figure 34: Time to ET due to an AE or Treatment Intolerance in Patients without a Hospitalization Event in the AZLI/Placebo Period (All Randomized)



Figures from 3.5.2 Clinical Endpoints of Interest

Changes in Hospitalization

Figure 35 shows that AZLI TID patients had a shorter time to first hospitalization (p=.085). Figure 36 shows that AZLI TID patients spent more days on average hospitalized at 1.05 days/patient vs. 0.46 days/patient for Pooled Placebo patients (p=.083). Comparisons were made using a non-parametric rank sum test due to the skewed distributions of days hospitalized.

Figure 35: Time to First Hospitalization (ITT)

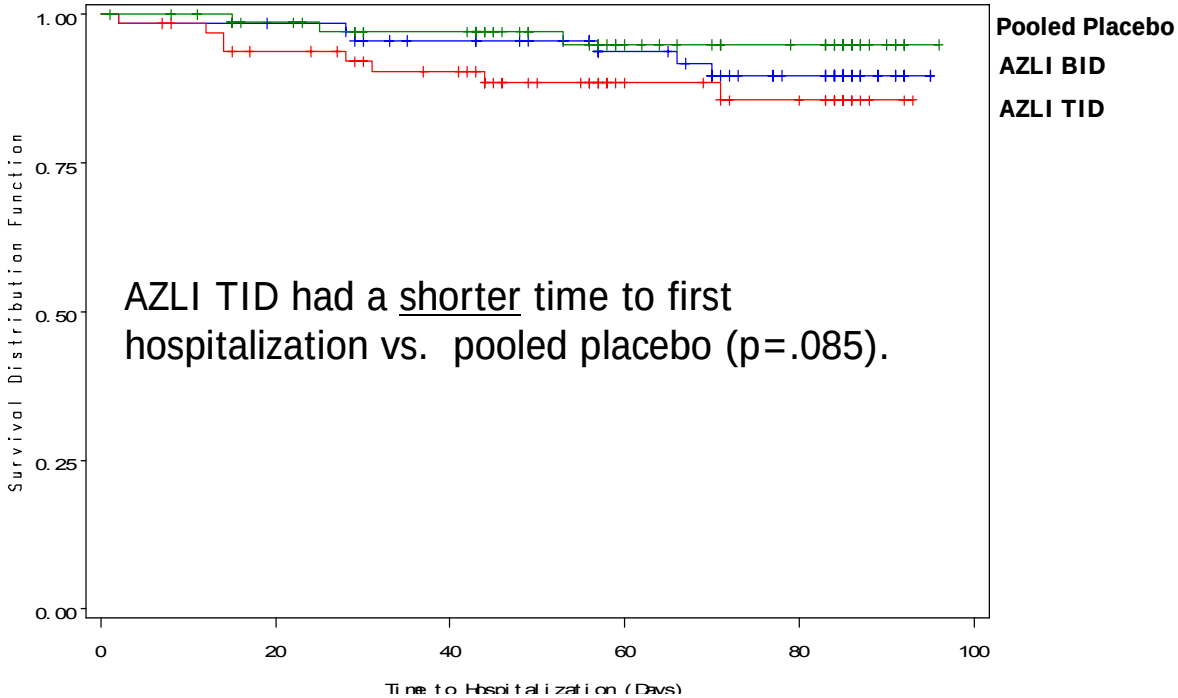
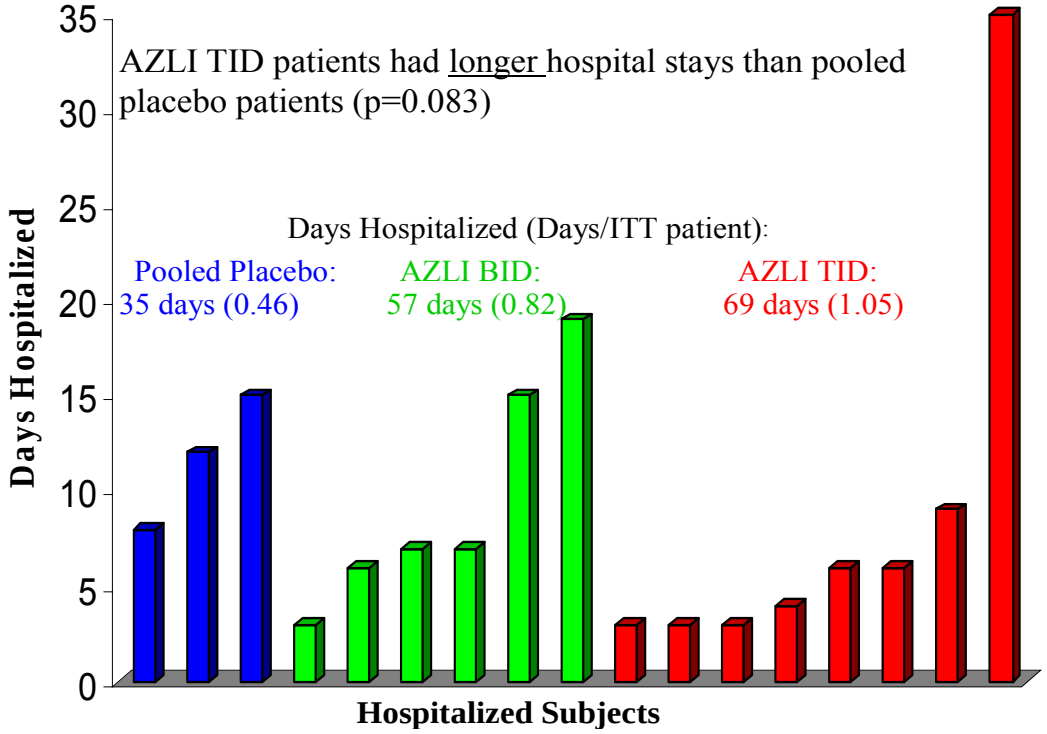


Figure 36: Days Hospitalized (ITT)



Figures 37-38 show that AZLI TID patients with time to need events had a significantly shorter time to first hospitalization (p=.024) vs. Pooled Placebo patients and had spent significantly more days on average hospitalized at 2.75 days/patient vs. 0.61 days/patient (p=.014). Comparisons were made using a non-parametric test due to the skewed distributions of days hospitalized.

Figures in 6.2 Reviewer Analyses

Figure 37: Time to First Hospitalization in Patients with Time to Need Events (ITT)

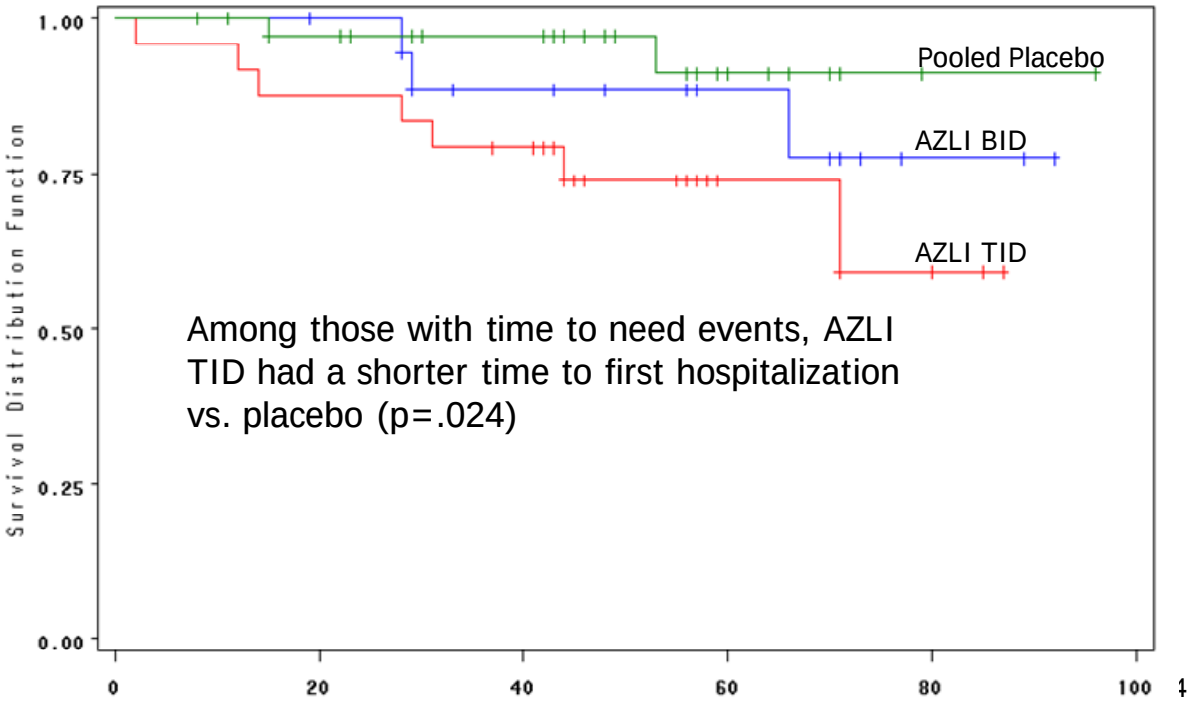
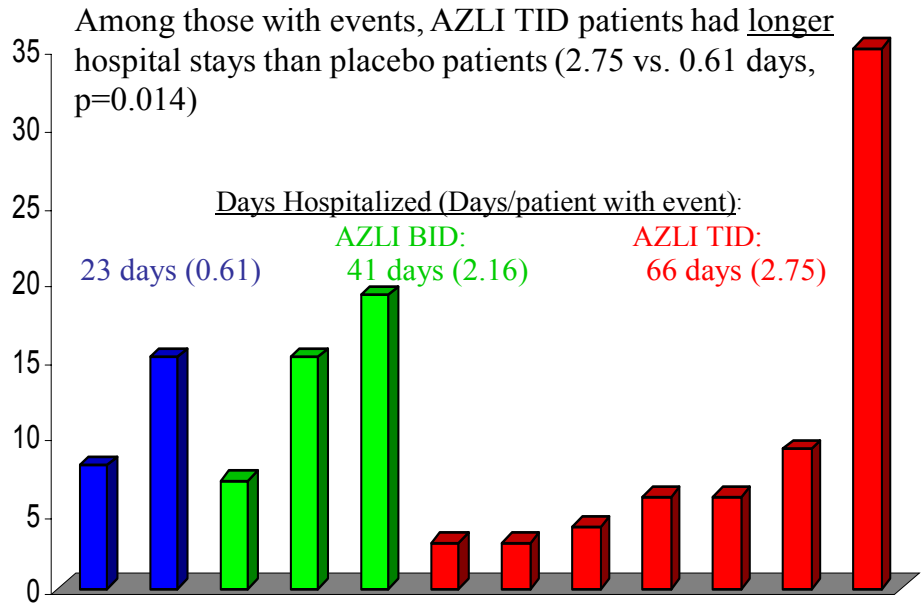


Figure 38: Days Hospitalized in Patients with Time to Need Events (ITT)



.

APPEARS THIS WAY ON ORIGINAL

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-50814	ORIG-1	GILEAD SCIENCES INC	CAYSTON(AZTREONAM FOR INHALATION SOL)

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

CHRISTOPHER E KADOORIE
01/27/2010

THAMBAN I VALAPPIL
01/27/2010
I concur

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Pharmacoepidemiology and Statistical Science
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CARCINOGENICITY STUDY

NDA Number:	50,814
Drug Name:	Cayston (Aztreonam)
Indication:	Cystic fibrosis patients with Pseudomonas aeruginosa
Applicant:	Gilead Pharmaceuticals, Inc.
Date:	Submitted 9 August 2007 (initial) 11 October 2007 (data sets)
Review Priority:	Standard
Biometrics Division:	Division 6
Statistical Reviewer:	Steve Thomson
Concurring Reviewer:	Team Leader: Karl Lin, Ph. D.
Medical Division:	Anti-Infective and Ophthalmology Products
Toxicologist:	Amy L. Ellis, Ph.D.
Project Manager:	Kyong Hyon, Ph.D.
Keywords:	Carcinogenicity, Cox regression, Kaplan-Meier product limit, Survival analysis, Trend test, Poly-k test

Table of Contents

1. EXECUTIVE SUMMARY

According to the report provided by the Sponsor: "Aztreonam, a monobactam antibiotic, is being developed by the Sponsor (b) (4) The objective of this study was to assess the toxicity and carcinogenicity of the aerosolised Aztreonam (as a lysine salt) formulation in IGS(CD) Sprague-Dawley rats following daily inhalation administration for 104 consecutive weeks." (page 9 of report)

The sponsor was Gilead Sciences Incorporated in Seattle, Washington. The studies were conducted by the (b) (4)

1.1. Conclusions and Recommendations

This submission summarizes the results of an inhalation study of aztreonam lysine in Sprague-Dawley rats. There were four treatment groups 1 through 4, labeled vehicle control, low, medium/intermediate, and high, each with 55 animals. Target inhalation doses were 0, 30, 60, and 120 mg Aztreonam/kg/day. The Sponsor claims achieved doses of 0, 31, 56, and 120 mg Aztreonam/kg/day.

In testing homogeneity in survival over the four dose groups with controls, the Cox test in Table 1 below is usually called the logrank test, while the K-W test, i.e., Kruskal-Wallis test, is more commonly called the Wilcoxon test or the generalized Wilcoxon test. Note that the Wilcoxon test places more weight on earlier events than does the logrank test. In male rats, the tests of homogeneity were statistically significant, i.e., there was evidence of heterogeneity over groups (both $p \leq 0.0091$), as was the test of trend (both $p \leq 0.0045$). There was no strong evidence of a departure from trend (both $p \geq 0.1080$). However, from the Kaplan-Meier plot in Appendix 1 and Table 9 in the report it seems clear that most of this lack of homogeneity in survival is due to higher mortality in the control group. The survival curves of the actual azertine doses are fairly closely intertwined, though there was be weak evidence of a decrease in mortality over increasing dose. In female rats there is no strong evidence of heterogeneity in over doses or a dose related trend. The actual p-values are given in Table 1 below:

Table 1. Statistical Significances of Tests of Homogeneity and Trend in Survival

	Males		Females	
	Cox	K-W	Cox	K-W
Homogeneity over Groups 1-4	0.0091	0.0058	0.2237	0.1927
Trend over Groups 1-4	0.0037	0.0045	0.9165	0.8071
Departure from trend in 1-4	0.2125	0.1080	0.1128	0.0968

One problem with testing for carcinogenicity is adjusting for the large number of statistical tests that need to be performed (please see section 1.3.1.3 for details). Simulations by Rahman and Lin (2008) have shown that generally the Haseman-Lin-Rahman rules roughly apply to results using the poly-3 tests as is done here. That is, since this is a one species study,

for rare tumors (incidence ≤ 1%) both the tests of trend and pairwise comparisons between the high dose group and the controls andshould be tested at a 0.05 (5%) level. The corresponding tests for common tumors (incidence > 1%) should be tested at a 0.01 level. This approach is intended to balance both Type I error and Type II error (i.e., the error of concluding there is no evidence of a relation to tumorgenicity when there actually is such a relation).

Table 2. below lists tumors that are potentially statistically significant in that at least one test is statistically significant at a 0.05 level. In female rats the test of trend in granular cell tumors of the vagina would be classified as rare, and thus the result would be considered as statistically significant (p = 0.0323<0.05), while benign c-cell adenoma in the thyroid would be considered as common, and thus close to statistical significance (p=0.0107 ≈ 0.01). No other tests in either gender even achieve the 0.05 level of significance, let alone, significance after the Haseman-Lin-Rahman adjustment.

Table 2. Potentially Statistically Significant Results of Poly-k tests for Neoplasms

	Incidence:				P-values:			
	Ctrl	Low	Med	High	Trend	High vs Ctrl	Med vs Ctrl	Low vs Ctrl
Females P<=0.05								
THYROID GLAND								
C-CELL ADENOMA [B]	4	3	4	10	.0107	.0716	.3823	.5606
VAGINA								
GRANULAR CELL TUMOUR [B]	0	1	0	3	.0323	.1249	.	.5275

Detailed incidence tables with further discussion are provided in Appendix 2.

1.2. Brief Overview of the Studies

This submission had one rat study:

Report 25316: Aztreonam, 104 Week Inhalation Carcinogenicity Study of Aztreonam in Rats

According to the Sponsor: "Aztreonam, a monobactam antibiotic, is being developed by the Sponsor as an inhalation drug (b) (4) (page 9 of report) The purpose of this study was to assess the oncogenic potential of Aztreonam when administered to IGS(CD) Sprague-Dawley rats by daily inhalation administration for 104 consecutive weeks. There were four treatment groups per gender, each with 55 animals, including a vehicle control, and three treatment groups with nominal target inhalation doses of 0, 30, 60, and 120 mg Aztreonam/kg/day, and claimed achived doses of 0, 31, 56, and 120 mg Aztreonam/kg/day. The study comprised 55 animals per sex and group. These were labeled as Control, Low, Medium or Intermediate, and High dose groups.

1.3. Statistical Issues and Findings

1.3.1. Statistical Issues

In this section, several issues typical of statistical analyses of these studies, are considered. These issues include details of the survival analyses, tests on tumorigenicity, multiplicity of tests on neoplasms, and the validity of the designs.

1.3.1.1. Survival Analysis:

The Cox test is usually called the logrank test, while the K-W, i.e., Kruskal-Wallis test, is more commonly called the Wilcoxon test or the generalized Wilcoxon test. Note that the Wilcoxon test places more weight on earlier events than does the logrank test. Both the Cox logrank and Kruskal-Wallis-Wilcoxon tests were used to test homogeneity of survival among the treatment groups. Tests of dose related trend using a Cox proportional odds model were also performed. The number of such tests raises issues of multiple testing, but from the point of view of finding differences among treatment groups (i.e., reducing the probability of Type II error), this should be acceptable. Appendix 1 reviews the animal survival analyses in some detail. The Sponsor's analyses are summarized in Section 3.2.1.1.

1.3.1.2. Tests on Neoplasms:

The Sponsor presents the results from a Peto analysis. This is has been the usual primary carcinogenicity analysis utilized in submissions to CDER. However, this approach does require accurate assessment of whether a tumor is fatal or incidental, as well a data dependent allocation to time periods for the incidental tumors. Largely on the basis of the former consideration, the Society of Toxicological Pathology had a town hall meeting in June 2001 where this methodology was criticized. The alternative recommended in the commentary on this meeting (STP Peto Working Group, 2002) is the poly-k modification of the Cochran-Armitage test of trend for tumor incidence, presented in Appendix 2, and used in the current FDA analysis. This is based on the apparent fact that the overall course of tumor onset usually seems to follow a polynomial in time.

1.3.1.3. Multiplicity of Tests on Neoplasms:

Testing the various neoplasms involves a large number of statistical tests, which in turn necessitates an adjustment in experiment-wise Type I error. The usual methodology for a two species, two gender, two year study is the Peto analysis, with testing for trend over four doses and comparing the high dose group to controls with the Haseman-Lin-Rahman rules. Based on his extensive experience with such analyses, for pairwise tests between the high dose group and controls in two species, Haseman (1983) claimed that for a roughly 0.10 (10%) overall false positive error rate, rare tumors should be tested at a 0.05 (5%) level, and common tumors (with a historical control incidence greater than 1%) at a 0.01 level. For a standard chronic study in two species (i.e., mice and rats) study, based on simulations and their experience, Lin & Rahman (1998) proposed a further p-value adjustment for tests of trend. That is, for a roughly 0.10 (10%) overall false positive error rate in tests of trend, rare tumors should be tested at a 0.025 (2.5%) level and common tumors at a 0.005 (0.5%) level. This is the adjustment used by the Sponsor.

However, since this is a one species study, for rare tumors both the test of trend and the test of pairwise comparison between the high dose group and the appropriate control should be tested at a 0.05 (5%) level. The corresponding tests for common tumors should be tested at a 0.01 level. Rahman & Lin (2008) indicate that the same adjustment can be used for poly-3 analyses. In this analysis we will use the observed incidence in the control group to decide if a tumor is rare or common. This approach to the multiplicity of tests is intended to balance both Type I error and Type II error (i.e., the error of concluding there is no evidence of a relation to tumorgenicity when there actually is such a relation).

Further, note that strictly speaking, these rules only control the overall errors of the test of trend in Aztreonam and the corresponding comparison between the high dose Aztreonam group and the control group. It is not clear how the error rate would apply to other possible tests, such as the comparison between the medium dose group and the low dose group with the control group.

1.3.1.4. Validity of the Designs:

When determining the validity of designs there are two key points:
1) adequate drug exposure,
2) tumor challenge to the tested animals.

1) is related to whether or not sufficient animals survived long enough to be at risk of forming late-developing tumors and 2) is related to the Maximum Tolerated Dose (MTD), designed to achieve the greatest likelihood of tumorigenicity.

Lin and Ali (1994), quoting work by Haseman, have suggested that a survival rate of about 25 animals, out of 50 or more animals, between weeks 80-90 of a two-year study may be considered a sufficient number of survivors as well as one measure of adequate exposure. From tables 9 and 10 below, note that this criterion does seem to be satisfied.

Chu, Ceuto, and Ward (1981), citing earlier work by Sontag *et al* (1976) recommend that the MTD "is taken as 'the highest dose that causes no more than a 10% weight decrement as compared to the appropriate control groups, and does not produce mortality, clinical signs of toxicity, or pathologic lesions (other than those that may be related to a neoplastic response) that would be predicted to shorten the animal's natural life span' " The values in the following tables were taken from the Sponsor's data set weight.sas7bdat. Table 3 gives body weights at initiation of treatment and the mean of the body weights at the last two assessments in the study, as well as and the corresponding final percent weight change relative to the weight change in the control group in each gender. In males the low dose group does seem to be associated with weight decrement greater than 10%, the same is not true of higher doses, and thus is probably an artifactual result. However, in females, all treated groups had weight gains, and thus, the above criterion in MTD does not seem to have been satisfied for females.

Table 3. Relative Weight Change (compared to control)

Dose Label Nominal Dose mg/kg/day	Males				Females			
	Baseline Day 0	Final Days	Change	% from Control	Baseline Day 1	Final Days	Change	% from Control
Control - 0	212.1	608.0	359.9		165.1	362.8	197.7	
Low - 31	212.0	555.1	342.9	-13.4%	165.2	379.3	214.1	8.3%
Medium - 56	212.6	603.6	391.0	-1.2%	167.2	366.3	199.1	0.7%
High - 120	212.5	587.9	375.3	-5.2%	167.5	372.0	204.5	3.4%

The following table displays the overall mean over weeks of the mean food consumption per treatment group. The values are taken from the table of relative food consumption. Within each gender, there is no evidence of consistent changes in food consumption across treatment groups.

Table 4. Means of Mean Food Consumption (g/day)

Dose Label	Dose mg/kg/day	Males		Females	
		Mean	% from Control	Mean	% from Control
Controls	0	25.7		18.7	
Low	31	25.6	1.2%	19.0	-0.4%
Medium	56	26.0	0.9%	18.9	1.2%
High	120	25.4	1.8%	19.1	-1.2%

As discussed in 2) above, excess mortality not associated with any tumor or sacrifice in the higher dose groups might suggest that the MTD was exceeded. One way to assess this possibility is to measure mortality not associated with any identified tumor. Note this seems to be a new way to assess if the high dose is at the MTD. Table 5 below indicates the number of animals in each dose group that died of a natural death or moribund sacrifice, but did not show any tumors:

Table 5. Natural Death or Accident with No Identified Tumor

Group Label	Dose mg/kg/day	Males		Females	
		Died w/o tumor	Other	Died w/o tumor	Other
Control	0	9	46	1	54
Low	31	8	47	1	54
Medium	56	9	46	2	53
High	120	0	55	1	54

To compare the incidence of deaths without tumors we can specify the usual survival tests where animals that die with a tumor or are sacrificed are considered as censored. The remaining animals are those that die prior to developing a tumor. If the MTD is exceeded we would expect a dose related excess toxicity, resulting in a dose related trend in these deaths. In females, no increasing trend over dose is apparent. In males the high dose group has fewer early

deaths than the other dose groups. Although this is a decision for the toxicologist, this may suggest that the MTD may have not even be met, let alone exceeded.

1.3.2. Statistical Findings

Please see Section 1.1 above.

2. INTRODUCTION

2.1. Overview

Results from a study in Crl:CD®(SD) IGS BR Sprague Dawley Rats (b) (4) were submitted to assess the carcinogenic potential of Aztreonam when administered daily in an inhalation chamber.

2.2. Data Sources

Four SAS transport files were provided by the Sponsor and placed in the CDER electronic data room (edr):

Tumor.xpt, weights.xpt, mortal.xpt, and food.xpt each contained the corresponding SAS data sets tumor.sas7bdat, weights.sas7bdat, mortal.sas7bdat, and food.sas7bdat respectively.

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

NA

3.2. Evaluation of Safety

More detailed results on the study are presented below.

3.2.1. Study Report 25316: Aztreonam, 104 Week Inhalation Carcinogenicity Study of Aztreonam in Rats

STUDY DURATION: 104 Weeks.
DOSING DATES: Reported as Experimental Start and Completion Dates of October 8, 2004 – October 10, 2007 [sic].
TERMINAL SACRIFICE (NECROPSY) DATE: October 19, 2006.
STUDY ENDING DATE (Final Report dated): October 12, 2007.
RAT: Crl:CD®(SD) IGS BR Sprague Dawley Rats (from (b) (4)
ROUTE: Fixed Inhalation Chamber

The Sponsor describes the route of treatment administration as follows: "The Inhalation exposures were conducted in a room adjacent to, but separate from, the animal holding room. Exposures to the test aerosols were performed using appropriately sized modular nose only flow past systems . . . . This exposure technique allows a continuous supply of test aerosol to be delivered to each animal; the biased flow ensures that there is no re-breathing of the test atmosphere. Separate chambers were used for the Control and test aerosols. A vacuum pump system was used to continuously exhaust the test atmosphere. Each exposure chamber was located in an extract booth to prevent any cross-group contamination and for the protection of the personnel undertaking the animal inhalation exposure procedures. Each exposure chamber was operated to sustain a dynamic air flow sufficient to ensure an evenly distributed exposure atmosphere. The chamber airflow rate was ca 48 L/min. Air flow rates were monitored continuously using calibrated flow meters. Chamber air flow rates, temperature and humidity were monitored and recorded at appropriate intervals during each exposure period.

"For inhalation exposure, the rats were restrained in clear, tapered, polycarbonate tubes with an adjustable back-stop to prevent the animals from turning in the tubes. The animals' snouts protruded through the tapered end of the restraint tubes which were connected to the exposure chamber by way of a push-fit through rubber 'o' ring in the aerosol delivery port. This exposure technique was used to minimise concurrent exposure by the oral and dermal routes." (page 18 of report)

The Sponsor indicates that this method was designed to achieve targeted inhalation doses of 0, 30, 60, and 120 mg of Aztreonam per kg/day in the vehicle control, low, medium and high dose groups respectively. They claim achieved doses of 0, 31, 56, and 120 mg/kg/day, respectively. In particular, the Sponsor reports that "The Control group was exposed to 30 mM NaCl for 200-260 min (average weekly range) over the course of the 104 week exposure period. The Low, Intermediate and High dose groups were exposed to Aztreonam formulations for 104 consecutive weeks at mean aerosol concentrations of 1.07, 1.01 and 1.10 mg Aztreonam/L respectively, for 50-70 min, 100-140 min and 200-260 min (average weekly ranges) respectively. Aerosol particle size measurements indicated an overall gravimetric mass median aerodynamic diameter (MMAD) of (b) (4) for the Control group aerosol. For the Aztreonam test aerosols the overall gravimetric MMAD was in the range (b) (4) Analytically, aerosol particle size measurements indicated an overall MMAD of (b) (4) for the Aztreonam aerosol." (page 9 of report)

Dose levels were justified as follows: " The dose levels used on this study were selected by the Sponsor, and discussed with the Study Director, based on the results obtained from a 90 Day Inhalation Toxicity Study of Aztreonam in Rats ((b) (4) Study No. 664348, Report No. 23679) and review of the draft protocol by CDER's Executive Carcinogenicity Committee. Results from the 90 day study indicated that there were no adverse effects on any of the in-life study parameters investigated (body weight, food consumption, laboratory investigations, ophthalmoscopy or organ weights). Histologically there was no evidence of systemic toxicity. Adverse histological effects were noted in the nasal cavities (minimal or mild olfactory epithelial

atrophy and/or rhinitis) and larynx (minimal or mild squamous metaplasia on the medial aspect of the arytenoid cartilage or of the U-shaped cartilage) at 120 mg Aztreonam/kg/day. These same findings were noted at a lower incidence in animals dosed at 60 mg Aztreonam/kg/day. There was evidence of a partial recovery from these lesions following a 28 day recovery period. The effects observed were considered minor in nature and did not preclude dosing chronically at 120 mg Aztreonam/kg/day. In addition, the High dose level provided sufficient systemic levels to evaluate more fully the clinical safety of Aztreonam lysine at the highest dose planned for the clinic." (page 17 of report)

The Sponsor reports that 220 male and 220 female 6-7 week old rats were chosen for the study from among 235 IGS rats (Crl:CD®(SD)IGS BR) of each gender, that had been acclimatized. "The remaining animals were retained as contingency replacements and removed from the study following completion of the second week of dosing."

Animals were observed daily for signs of reaction to treatment, with more detailed weekly examinations. Food consumption and animal weights were assessed weekly up to week 14, and every four weeks thereafter. The Sponsor states that food and water were available *ad libitum*, except during inhalation treatment. Toxicokinetic blood samples were reported as being collected from 3 animals/sex/group on Day 1 and during Week 26 of dosing, at nine times at each of these weeks.

3.2.1.1. Sponsor's Results and Conclusions

This section will present a summary of the Sponsor's analysis on survivability and tumorigenicity in rats.

Survival analysis:

According to the Sponsor: "There were a total of 189 animals (91 males and 98 females) that died or were prematurely killed for welfare reasons during the 104 week dosing period. Details are presented in the table below." (page 30 of report) Simple mortality results are summarized in the following table:

Table 6. Sponsor's Mortality Summary

Dose Group/ Treatment	Number of Premature Decedents			
	Males		Females	
	Found Dead	Killed Prematurely	Found Dead	Killed Prematurely
1 Vehicle Control	3	28	0	29
2 Low Dose	4	18	0	19
3 Medium Dose	6	17	2	22
4 High Dose	1	14	0	26

Although it is not made explicit in the Sponsor's report, it does appear that the p-values for the tests of equality in survival presented below, come from the standard tests of equality of Kaplan-Meier survival curves.

Table 7. Significance Levels of Sponsor's Tests of Equality in Survival

Comparison	Males	Females
Vehicle vs. Low	0.010	0.033
Vehicle vs. Medium	0.066	0.42
Vehicle vs. High	0.002	0.77

In fact, these tests suggest mortality was statistically significantly lower in the low dose compared with the vehicle control for both male and female rats. (Males: p=0.010, Females: p=0.033). Further, the observed decrement in mortality in the high dose group versus vehicle in males was also highly statistically significant (p = 0.002), while the test comparing the medium dose to vehicle was close to the usual significance level (p=0.066).

Tumorigenicity analysis:

The Sponsor reports that "Peto analysis revealed a significant increase in C-cell adenoma in the thyroid gland in Group 4 (High dose) females compared with Control Group 1 (p=0.026). In addition there was evidence of a dose related increase in thyroid gland C-cell adenoma in females (p=0.005). However when Group 4 was excluded, Peto analysis did not reveal any significant trends or pairwise comparisons in the remaining groups. . . . The incidence of anterior lobe pituitary adenomas was increased in Group 2 (Low dose) males, although the increased incidence did not achieve statistical significance. . . . There was an increase in the incidence of granular cell tumours in the vagina of Group 4 (High dose) females. The incidence of granular cell tumours in the uterine cervix is also shown, for comparison. Peto analysis revealed weak evidence of a dose related increase in vagina granular cell tumour (p=0.026). However when the incidences of granular cell tumours in uterine cervix and vagina were combined, Peto analysis did not reveal any significant trends or pairwise comparisons. . . . The incidence of mammary fibroadenomas was increased in Group 2 (Low dose) and 3 (Intermediate [i.e., Medium] dose) females, but the increased incidences did not achieve statistical significance. . . . The number of animals with benign tumours was similar in all groups, while the number of animals with malignant tumours was decreased in treated groups when compared with Group 1 (Control)." (see pages 33-34). Note these p-values are similar to those in the FDA analysis using the poly-3 model. Tumor incidences (# out of N) reported by the Sponsor are summarized in table 8 below:

Table 8. Tumor Incidences

Incidence:		Males				Females			
		Ctr1	Low	Med	High	Ctr1	Low	Med	High
Thyroid Gland	N	55	51	54	54	55	55	55	55
C-Cell Adenoma	#	7	8	8	7	4	3	4	10
Pituitary Gland	N	54	55	55	55	55	55	55	55
Adenoma, Anterior Lobe [B]	#	11	17	15	15	29	32	32	27
Mammary Gland	N	48	51	50	47	54	55	55	54
Fibroadenoma	#	1	0	0	0	9	17	17	12
Uterus	N					55	55	55	55
Granular Cell Tumor [B]	#					0	2	1	0
Vagina	N					55	55	54	55
Granular Cell Tumor [B]	#					0	1	0	3

Note the above counts seem to agree with those presented in the FDA analysis below and in Appendix 2.

3.2.1.2. FDA Reviewer's Results

This section will present the current Agency findings on survival and tumorigenicity in male and female rats.

Survival analysis:

The following tables (Table 9 for male rats, Table 10 for female rats) summarize the mortality results for the dose groups. The data were grouped for the specified time period, and present the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage cited is the percent survived at the end of the interval.

Table 9. Summary of Male Rat Survival (Aztreonam: daily dose)

Period (Weeks)	Vehicle Control	Low 31 mg/kg	Medium 56 mg/kg	High 120 mg/kg
0-50	5/55[1] 90.9%[2]	1/55 98.2%	2/50 96.4%	0/50
51-78	8/50 76.4%	0/54	6/53 85.5%	8/55 85.5%
79-91	11/42 56.4%	9/54 81.8%	6/47 74.5%	2/47 81.8%
92-104	7/31 43.6%	12/45 60.0%	9/41 58.2%	5/45 72.7%
Terminal 105	24	33	32	40

[1] number deaths / number at risk
[2] per cent survival to end of period.

In these tables all animals that died during the terminal sacrifice period are counted as having been sacrificed, even those that died of other causes.

Table 10. Summary of Female Rat Survival (Aztreonam: dose/kg/day)

Period (Weeks)	Vehicle Control	Low 31 mg/kg	Medium 56 mg/kg	High 120 mg/kg
0-50	1/55[1] 98.2%[2]	1/55 98.2%	1/55[1] 98.2%[2]	3/55 94.5%
51-78	8/54 83.6%	3/54 92.7%	6/54 87.3%	4/52 87.3%
79-91	2/46 80.0%	5/51 83.6%	7/48 74.5%	9/48 70.9%
92-104	18/44 47.3%	10/46 65.5%	10/41 56.4%	10/39 52.7%
Terminal 105	26	36	31	29

[1] number deaths / number at risk

[2] per cent survival to end of period.

The results of the tests of trend in survival, departure from trend, and overall homogeneity over the four dose groups (including controls) are given in Table 11 below.

Table 11. Statistical Significances of Tests of Homogeneity and Trend in Survival

	Males		Females	
	Cox	K-W	Cox	K-W
Homogeneity over Groups 1-4	0.0091	0.0058	0.2237	0.1927
Trend over Groups 1-4	0.0037	0.0045	0.9165	0.8071
Departure from trend in 1-4	0.2125	0.1080	0.1128	0.0968

In male rats, the tests of homogeneity were statistically significant, i.e., there was evidence of heterogeneity over groups (both $p \leq 0.0091$), as was the test of trend (both $p \leq 0.0045$). There was no strong evidence of a departure from trend (both $p \geq 0.1080$). However, from the Kaplan-Meier plots in Appendix 1 and Table 9 in the report it seems clear that most of this lack of homogeneity in survival is due to higher mortality in the control group. The survival curves of the actual aztreonam doses are fairly closely intertwined, though there is weak evidence of a decrease in mortality over increasing dose. In female rats there is no strong evidence of heterogeneity in over doses or a dose related trend.

Tumorigenicity analysis:

Table 12 below lists tumors that have any p-value less than 0.05. The table displays the tumor incidence over the four dosing groups, including controls, as well as the p-values using the poly-k adjustment to the Cochran-Armitage test of trend in dose. The first p-value provides the results of the overall poly-k test of trend, here with k=3. The poly-k test modifies the original Cochran-Armitage test to adjust for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). The next last three columns present the results of tests between the

control group and each of the high dose group, the medium dose group, the low dose group respectively. Complete incidence tables and test results are presented in Appendix 2.

Note that in male rats no tumors reached the nominal 0.05 level of significance. Applying the Haseman-Lin-Rahman rules for a single species study (see section 1.3.1.3), rare tumors could be considered statistically significant if the observed p-value is 0.05 or less, while common tumors would be considered statistically significant if the observed p-value is 0.01 or less. Using the incidence in the control group to determine the rarity of the tumor, in female rats the trend test in granular cell tumors of the vagina would be classified as rare, and thus would be considered as statistically significant ($p = 0.0323<0.05$), while benign c-cell adenoma in the thyroid would be considered as common, and thus close to statistical significance ($p=0.0107 \approx 0.01$). No other tests in female rats even achieve the 0.05 level of significance, let alone, significance after the Haseman-Lin-Rahman adjustment.

Table 12. Potentially Statistically Significant Results of Poly-k Tests for Neoplasms

	Incidence:				P-values:			
	Ctrl	Low	Med	High	Trend	High vs Ctrl	Med vs Ctrl	Low vs Ctrl
sex=F P<=0.05								
THYROID GLAND								
C-CELL ADENOMA [B]	4	3	4	10	.0107	.0716	.3823	.5606
VAGINA								
GRANULAR CELL TUMOUR [B]	0	1	0	3	.0323	.1249	.	.5275

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

NA

5. SUMMARY AND CONCLUSIONS

5.1. Statistical Issues and Collective Evidence

Please see Section 1.3 above.

5.2. Conclusions and Recommendations

Please see section 1.1 above.

APPENDICES:

Appendix 1. Survival Analysis

The statistical significances of the tests of differences in survival across treatment groups are given below. A test for homogeneity in survival is a test that survival is equal across the controls, low, medium/intermediate, and high dose treatment groups, while the test of trend is a test of dose related trend across these groups. Note that the Cox test is usually called the logrank test, while the K-W, i.e., Kruskal-Wallis test, is more commonly called the Wilcoxon test. Note that the Wilcoxon test places more weight on earlier events than does the logrank test.

In both male rats, the tests of homogeneity were statistically significant, i.e., there was evidence of heterogeneity over groups (both $p \leq 0.0091$), as was the test of trend (both $p \leq 0.0045$). There was no strong evidence of a departure from trend (both $p \geq 0.1080$). However, from the Kaplan-Meier plots it appears and Table 9 in the report it seems clear that most of this lack of homogeneity in survival is due to higher mortality in the control group. The actual aztreonam doses are fairly closely intertwined, though there was be weak evidence of a decrease in mortality over increasing dose. In female rats there is no strong evidence of heterogeneity in over doses or a dose related trend. The actual p-values are given in A.1.1 below:

Table A.1.1 Statistical Significances of Tests of Homogeneity and Trend in Survival

	Males		Females	
	Cox	K-W	Cox	K-W
Homogeneity over Groups 1-4	0.0091	0.0058	0.2237	0.1927
Trend over Groups 1-4	0.0037	0.0045	0.9165	0.8071
Departure from trend in 1-4	0.2125	0.1080	0.1128	0.0968

Further, the figures A.1.1 and A.1.2, below, display these Kaplan-Meier estimated survival curves for the two genders.

Figure A.1.1 Kaplan-Meier Survival Curves for Male Rats

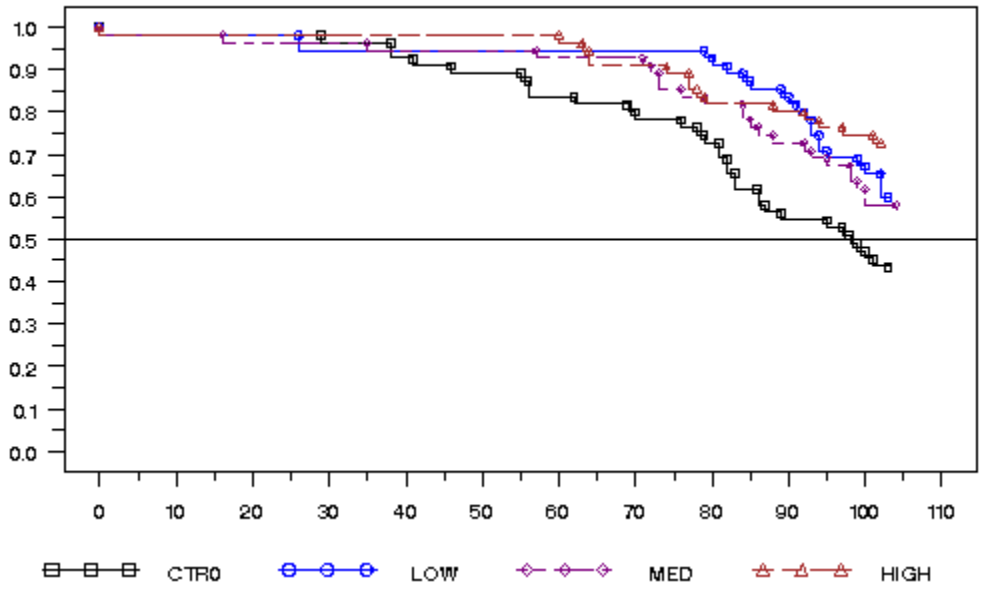
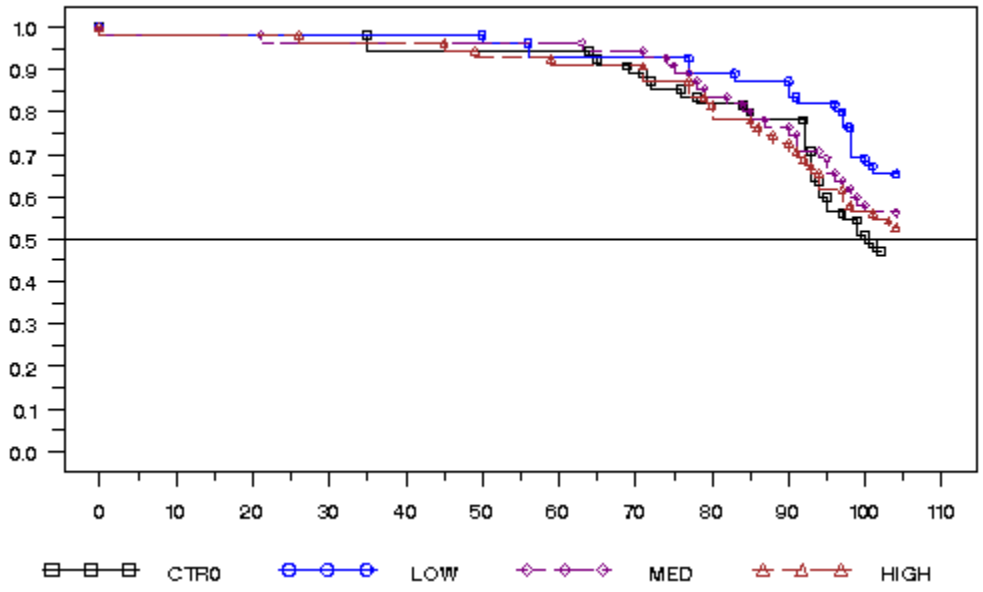


Figure A.1.2 Kaplan-Meier Survival Curves for Female Rats



Appendix 2. Poly-k Tumorigenicity Analysis

The tables below display the tumor incidence over the four dosing groups, including controls, as well as the p-values using the poly-k adjustment to the Cochran-Armitage test of trend in dose. The first p-value provides the results of the overall poly-k test of trend, here with k=3. The poly-k test modifies the original Cochran-Armitage test to adjust for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). The last three columns present the results of tests between the control group and each of the high dose group, the medium dose group, the low dose group respectively.

As noted in the report, at the Society of Toxicological Pathology "town hall" meeting in June 2001 the poly-k modification of the Cochran-Armitage test of trend seemed to have been recommended over the Peto tests. The tests used here are small sample exact tests. These do assume all marginal totals are fixed, a debatable assumption. To adjust for the multiplicity of tests, tentatively, the Haseman-Lin-Rahman rules discussed in Section 1.3.1.3. of the report seem to apply, here with the modification for a single species study.

Table A.2.1 below lists tumors that have any p-value less than 0.05. Note that in male rats no tumors reached the nominal 0.05 level of significance. Applying the Haseman-Lin-Rahman rules for a single species study, rare tumors could be considered statistically significant if the observed p-value is 0.05 or less, while common tumors would be considered statistically significant if the observed p-value is 0.01 or less. Using the incidence in the control group to determine the rarity of the tumor, in female rats the trend test in granular cell tumors of the vagina would be classified as rare, and thus would be considered as statistically significant ($p = 0.0323 < 0.05$), while benign c-cell adenoma in the thyroid would be considered as common, and thus quite close to statistical significance ($p = 0.0107 \approx 0.01$). No other tests even achieve the 0.05 level of significance, let alone, significance after the Haseman-Lin-Rahman adjustment.

Table A.2.1. Potentially Statistically Significant Results of Poly-k Tests of Neoplasms

	Incidence:				P-values:			
	Ctrl	Low	Med	High	Trend	High vs Ctrl	Med vs Ctrl	Low vs Ctrl
sex=F P<=0.05								
THYROID GLAND								
C-CELL ADENOMA [B]	4	3	4	10	.0107	.0716	.3823	.5606
VAGINA								
GRANULAR CELL TUMOUR [B]	0	1	0	3	.0323	.1249	.	.5275

Table A.2.2. Significance Levels of Poly-k Tests for Neoplasms in Male Rats

	Incidence:				P-values:	High vs	Med vs	Low vs
	Ctrl	Low	Med	High	Trend	Ctrl	Ctrl	Ctrl
ABDOMINAL CAVITY								
LIPOMA [B]	0	1	0	0	.5169	.	.	.5581
ADRENAL GLAND								
CORTICAL ADENOMA [B]	1	1	0	0	.8402	.5581	.5366	.3028
CORTICAL CARCINOMA [M]	0	1	0	0	.5169	.	.	.5581
GANGLIONEUROMA [B]	0	0	1	0	.5198	.	.5366	.
PHAEOCHROMOCYTOMA [B]	1	2	1	1	.5885	.3086	.2849	.5801
PHAEOCHROMOCYTOMA [M]	2	0	0	0	.9549	.8077	.7883	.8031
BRAIN								
ASTROCYTOMA [B]	1	0	0	0	.7853	.5581	.5366	.5529
GRANULAR CELL TUMOUR [B]	1	0	0	0	.7853	.5581	.5366	.5529
MALIGNANT ASTROCYTOMA [M]	4	1	0	0	.9968	.9630	.9552	.8725
PARAGANGLIOMA [B]	0	0	1	0	.5198	.	.5366	.
EYE								
AMELANOTIC MELANOMA [M]	0	0	1	0	.5198	.	.5366	.
FEMUR								
FIBROSARCOMA [M]	0	0	1	0	.5198	.	.5366	.
FOOT/LEG								
HISTIOCYTIC SARCOMA [M]	0	1	0	0	.5169	.	.	.5581
SARCOMA (NOT OTHERWISE SPECIFIED) [M]	1	0	0	0	.7853	.5581	.5366	.5529
HAEMOPOIETIC SYSTEM								
HAEMOPOIETIC TUMOUR (NOS) [M]	0	0	0	1	.2712	.5581	.	.
HISTIOCYTIC SARCOMA [M]	0	1	0	2	.1310	.3086	.	.5581
LEUKAEMIA, GRANULOCYTIC [M]	2	0	0	0	.9530	.8019	.7823	.7973
LEUKAEMIA, LARGE GRANULAR CELL [M]	0	1	1	0	.5277	.	.5422	.5581
LYMPHOMA, FOLLICULAR CENTRE CELL [M]	2	0	0	0	.9530	.8019	.7823	.7973
LYMPHOMA, LYMPHOCYTIC [M]	0	1	0	0	.5169	.	.	.5581
JEJUNUM								
FIBROSARCOMA [M]	1	0	0	0	.7853	.5581	.5366	.5529
KIDNEY								
LIPOSARCOMA [M]	0	0	0	1	.2712	.5581	.	.
NEPHROBLASTOMA [M]	1	0	0	0	.7809	.5517	.5301	.5465
TUBULAR CELL ADENOMA [B]	0	0	0	1	.2712	.5581	.	.
TUBULAR CELL CARCINOMA [M]	1	0	0	0	.7853	.5581	.5366	.5529
LIVER								
HEPATOCELLULAR ADENOMA [B]	0	1	0	1	.3519	.5581	.	.5581
HEPATOCELLULAR CARCINOMA [M]	0	1	0	0	.5198	.	.	.5529
LYMPH NODE (MANDIBULAR)								
HAEMANGIOSARCOMA [M]	0	0	1	0	.5225	.	.5422	.
MAMMARY GLAND								
CARCINOMA [M]	0	0	0	1	.2712	.5581	.	.
FIBROADENOMA [B]	1	0	0	0	.7853	.5581	.5366	.5529
NASAL CAVITY								
POLYP [B]	0	0	0	1	.2712	.5581	.	.
ORAL CAVITY								
SQUAMOUS-CELL CARCINOMA [M]	0	0	0	1	.2712	.5581	.	.
PANCREAS (ENDOCRINE)								
ISLET CELL ADENOMA [B]	4	1	1	1	.9301	.8776	.8555	.8725
PARATHYROID GLAND								
ADENOMA [B]	0	0	0	1	.2712	.5581	.	.
PITUITARY GLAND								
ADENOMA, ANTERIOR LOBE [B]	11	17	15	15	.4829	.4854	.4128	.3290
ADENOMA, INTERMEDIATE LOBE [B]	1	1	0	1	.4984	.3086	.5366	.3028
CARCINOMA, ANTERIOR LOBE [M]	0	0	0	1	.2712	.5581	.	.
SKELETAL MUSCLE								
FIBROSARCOMA [M]	0	1	0	0	.5169	.	.	.5581
HAEMANGIOSARCOMA [M]	0	0	1	0	.5198	.	.5366	.

Table A.2.2. (cont.) Significance Levels of Poly-k Tests for Neoplasms in Male Rats

	Incidence:				P-values:	High vs Ctrl	Med vs Ctrl	Low vs Ctrl
	Ctrl	Low	Med	High	Trend			
SKIN AND SUBCUTIS								
BASAL CELL ADENOMA [B]	0	0	0	1	.2712	.5581	.	.
BASAL CELL CARCINOMA [M]	0	0	0	1	.2712	.5581	.	.
BASOSQUAMOUS CARCINOMA [M]	1	0	1	0	.7283	.5517	.2840	.5465
FIBROMA [B]	2	4	2	4	.3422	.4554	.3745	.4445
FIBROSARCOMA [M]	0	0	0	1	.2712	.5581	.	.
KERATOACANTHOMA [B]	3	5	2	2	.8328	.6106	.5807	.4819
MALIGNANT FIBROUS HISTIOCYTOMA [M]	0	0	1	0	.5198	.	.5366	.
RHABDOMYOSARCOMA [M]	1	0	0	0	.7853	.5581	.5366	.5529
SQUAMOUS-CELL PAPILLOMA [B]	1	1	0	0	.8402	.5581	.5366	.3028
SPINAL CORD								
MALIGNANT ASTROCYTOMA [M]	1	0	0	1	.4677	.3015	.5301	.5465
STERNUM								
CHONDROMA [B]	0	0	1	0	.5225	.	.5422	.
TESTIS								
INTERSTITIAL CELL ADENOMA [B]	2	2	2	3	.3786	.6106	.3654	.4011
THYMUS								
MALIGNANT THYMOMA [M]	1	0	0	0	.7853	.5581	.5366	.5529
THYMOMA [B]	1	0	1	0	.7329	.5581	.2849	.5529
THYROID GLAND								
C-CELL ADENOMA [B]	7	8	8	7	.6708	.5759	.4190	.4739
C-CELL CARCINOMA [M]	0	1	1	0	.5307	.	.5366	.5529
FOLLICULAR CELL ADENOMA [B]	2	0	0	4	.0724	.4554	.7883	.8031

Table A.2.3. Significance Levels of Poly-k Tests for Neoplasms in Female Rats

	Incidence:				P-values:	High vs Ctrl	Med vs Ctrl	Low vs Ctrl
	Ctrl	Low	Med	High	Trend			
ADRENAL GLAND								
CORTICAL ADENOMA [B]	1	1	1	0	.7471	.5000	.2586	.2755
CORTICAL CARCINOMA [M]	0	0	1	0	.4916	.	.5114	.
PHAEOCHROMOCYTOMA [B]	4	0	1	1	.8873	.8198	.8267	.9514
BRAIN								
MALIGNANT ASTROCYTOMA [M]	4	0	1	0	.9902	.9361	.8196	.9490
MIXED GLIOMA [M]	1	0	0	0	.7556	.4943	.5056	.5217
HAEMOPOIETIC SYSTEM								
HISTIOCYTIC SARCOMA [M]	2	0	0	3	.1340	.5110	.7641	.7795
LEUKAEMIA, LARGE GRANULAR CELL [M]	0	0	1	0	.4916	.	.5114	.
LYMPHOMA, FOLLICULAR CENTRE CELL [M]	1	0	1	0	.6923	.5000	.2586	.5275
HEART								
MALIGNANT SCHWANNOMA [M]	1	0	0	0	.7598	.5000	.5114	.5275
KIDNEY								
LIPOMA [B]	0	0	0	1	.2402	.5000	.	.
NEPHROBLASTOMA [M]	1	0	0	0	.7556	.4943	.5056	.5217
LIVER								
CHOLANGIOMA [B]	0	1	0	0	.4889	.	.	.5326
HEPATOCELLULAR CARCINOMA [M]	0	0	1	0	.4916	.	.5114	.
RHABDOMYOSARCOMA [M]	0	0	1	0	.4916	.	.5114	.
LUNG								
BRONCHIOLO-ALVEOLAR ADENOMA [B]	0	0	0	1	.2402	.5000	.	.

Table A.2.3. (cont.) Significance Levels of Poly-k Tests for Neoplasms in Female Rats

	Incidence:				P-values:	High vs	Med vs	Low vs
	Ctrl	Low	Med	High	Trend	Ctrl	Ctrl	Ctrl
MAMMARY GLAND								
ADENOCARCINOMA [M]	8	8	9	5	.7967	.7249	.5590	.5016
ADENOMA [B]	5	5	7	7	.2235	.3789	.4106	.4258
FIBROADENOMA [B]	9	17	17	12	.3848	.3304	.0767	.0967
NASAL CAVITY								
SARCOMA (NOT OTHERWISE SPECIFIED) [M]	1	0	0	0	.7598	.5000	.5114	.5275
ORAL CAVITY								
SQUAMOUS-CELL CARCINOMA [M]	1	1	1	0	.7418	.4943	.2528	.2749
OVARY								
GRANULOSA CELL TUMOUR [B]	0	1	0	1	.3077	.5000	.	.5275
TUBULAR CELL ADENOMA [B]	2	1	0	0	.9386	.7471	.7584	.5330
TUBULOSTROMAL ADENOMA [B]	0	0	0	1	.2402	.5000	.	.
PANCREAS (ENDOCRINE)								
ISLET CELL ADENOMA [B]	0	3	0	1	.4320	.5057	.	.1467
PITUITARY GLAND								
ADENOMA, ANTERIOR LOBE [B]	29	32	32	27	.7050	.5798	.4355	.4355
ADENOMA, INTERMEDIATE LOBE [B]	1	1	0	0	.8132	.5000	.5114	.2809
CARCINOMA, ANTERIOR LOBE [M]	0	0	0	1	.2402	.5000	.	.
CRANIOPHARYNGIOMA [B]	1	0	0	0	.7598	.5000	.5114	.5275
SKIN AND SUBCUTIS								
BASAL CELL ADENOMA [B]	0	0	1	0	.4916	.	.5114	.
FIBROMA [B]	0	0	1	1	.1816	.5057	.5114	.
FIBROSARCOMA [M]	0	1	0	0	.4916	.	.	.5275
KERATOACANTHOMA [B]	1	2	1	1	.5177	.7529	.2586	.5416
LIPOMA [B]	0	1	1	0	.4859	.	.5114	.5275
OSTEOSARCOMA [M]	0	0	0	1	.2444	.5057	.	.
SQUAMOUS-CELL CARCINOMA [M]	1	0	0	1	.4281	.7529	.5056	.5217
THYMUS								
MALIGNANT THYMOMA [M]	0	0	0	1	.2444	.5057	.	.
THYMOMA [B]	1	1	1	2	.2635	.5000	.2528	.2695
THYROID GLAND								
C-CELL ADENOMA [B]	4	3	4	10	.0107	.0716	.3823	.5606
FOLLICULAR CELL ADENOMA [B]	0	1	0	0	.4916	.	.	.5275
TONGUE								
GRANULAR CELL TUMOUR [B]	1	0	0	0	.7598	.5000	.5114	.5275
UTERUS								
ENDOMETRIAL ADENOMA [B]	0	0	1	0	.4916	.	.5114	.
GRANULAR CELL TUMOUR [B]	0	2	1	0	.6294	.	.5114	.2755
LEIOMYOMA [B]	0	1	0	0	.4916	.	.	.5275
STROMAL POLYP [B]	9	7	9	3	.9453	.9304	.5831	.6602
VAGINA								
GRANULAR CELL TUMOUR [B]	0	1	0	3	.0323	.1249	.	.5275
HAEMANGIOMA [B]	1	0	0	0	.7556	.4943	.5056	.5217

Appendix 3. References

Bailer, A. and Portier, C. (1988), "Effects of Treatment-Induced Mortality on Tests for Carcinogenicity in Small Samples", *Biometrics*, **44**, 4, 417-431.

Bieler, G.S., and Williams, R.L. (1993), "Ratio Estimates, the Delta Method, and Quantal Response Tests for Increased Carcinogenicity", *Biometrics*, **49**, 4, 793-801.

Chu, K.C., Ceuto, C., and Ward, J.M. (1981), Factors in the Evaluation of 200 National Cancer Institute Carcinogen Bioassays, *Journal of Toxicology and Environmental Health,* 8, 251-280.

Haseman, J. K. (1983), A Reexamination of False-positive Rates for Carcinogenicity Studies, *Fundamental and Applied Toxicology,* 3, 334-339.

Lin, K. K. and Ali, M.W. (2006), Statistical Review and Evaluation of Animal Tumorigenicity Studies, *Statistics in the Pharmaceutical Industry,Third Edition,* edited by C.R. Buncher and J.Y. Tsay, Marcel Dekker, Inc. New York.

Lin, K. K. and Rahman, M.A. (1998), Overall False Positive Rates in Tests for Linear Trend in Tumor Incidence in Animal Carcinogenicity Studies of New Drugs, *Journal of Biopharmaceutical Statistics*. 8(1), 1-15.

McConnell, E.E., Solleveld, H.A., Swnberg, J.A., and Boorman, G.A. (1986), Guidelines for Combining Neoplasms for Evaluation of Rodent Carcinogenesis Studies, *Journal of the National Cancer Institute*. 76, 283-289.

Peto, R., Pike, M.C., Day, N.E., Gray, R.G., Lee, P.N., Parrish, S., Peto, J., Richards, S., and Wahrendorf, J. (1980). Guidelines for sample sensitive significance tests for carcinogenic effects in long-term animal experiments, *IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans, supplement 2: Long term and Short term Screening Assays for Carcinogens: A Critical Appraisal*, International Agency for Research Against Cancer, 311-426.

Rahman, M.A. and Lin, K.K. (2008), A Comparison of False Positive Rates of Peto and Poly-3 Methods for Long Term Carcinogenicity Data Analysis Using Multiple Comparison Adjustment Method Suggested by Lin and Rahman, *Journal of Biopharmaceutical Statistics*. 18, 949-958.

STP Peto Working Group (2002), Statistical Methods for Carcinogenicity Studies, *Toxicologic Pathology*. 30 (3), 403-414.

U.S. Department of Health and Human Services (2001), *Guidance for Industry Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals (DRAFT GUIDANCE),* Center for Drug Evaluation and Research, Food and Drug Administration.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

Steven Thomson
10/14/2008 02:57:14 PM
BIOMETRICS

Karl Lin
10/14/2008 03:25:52 PM
BIOMETRICS
Concur with review

Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA # :	50814
Drug Name:	Aztreonam lysine for inhalation (75mg/mL TID)
Indication(s):	To improve respiratory symptoms and pulmonary function in cystic fibrosis patients with *P.aeruginosa*
Applicant:	Gilead Sciences, Inc.
Stamp Date:	November 16, 2007
PDUFA Goal Date:	September 16, 2008
Reviewer Completion Date:	July 22, 2008
Biometrics Division:	Division of Biometrics IV (HFD-725)
Medical Division:	Division of Anti-Infective and Ophthalmology Drug Products (HFD-520)
Documents Reviewed:	\\CDSESUB1\evsprod\Nda050814\0000
Statistical Reviewer:	Christopher Kadoorie, PhD
Concurring Reviewers:	Thamban Valappil, PhD
Clinical Reviewer:	Menfo Imoisili, MD, MPH
Project Manager:	Kyong Hyon, BSN, MA

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. EXECUTIVE SUMMARY

1.1 Introduction

Gilead Sciences, Inc. has submitted NDA 50814 to support approval of CAYSTON™ (aztreonam inhalation (AI)) as a three times daily (TID) treatment to improve respiratory symptoms and pulmonary function in patients with cystic fibrosis due to *P.aeruginosa*. This submission includes results from six clinical studies including two phase I pharmacokinetic studies, one phase II placebo-controlled study assessing safety and efficacy (Study CP-AI-003), two phase III placebo-controlled studies assessing safety and efficacy (Studies CP-AI-005 and CP-AI-007) and one open-label follow-on phase III study (Study CP-AI-006) assessing safety only. This statistical review primarily focuses on the efficacy results presented from Studies CP-AI-005 and CP-AI-007, hereafter referred to as Study 005 and Study 007.

1.2 Conclusions and Recommendations

Overall evidence presented in NDA 50814 failed to demonstrate a substantial treatment benefit for patients with cystic fibrosis due to *Psuedomonas aeruginosa* (*PA*) using 75mg aztreonam inhalation three times daily (AI TID). Study 005 failed to provide meaningful evidence regarding a treatment benefit from AI TID therapy and Study 007 provided only marginal evidence. Demonstration of treatment efficacy requires substantial evidence essentially replicated from two or more adequate and well controlled trials.

In Study 005, ITT patients in the AI TID treatment arm failed to achieve significant improvement compared to Placebo for several key endpoints. This included the primary endpoint, 'Time to Need for Inhaled or IV Antibiotics,' a key secondary endpoint 'Actual Change in CFQ-R Respiratory Domain Scores at Day 28,' and several other secondary endpoints. In Study 005, primary analyses were also unclear due to a "regimen effect" (i.e. BID vs. TID dosing) which was observed to be stronger than the treatment effect (i.e. AI vs. Placebo) such that patients on a placebo BID regimen had actually fared better than patients on the AI TID regimen. Due to the strength of this "regimen effect" in the Placebo BID and Placebo TID arms, FDA primary analysis comparisons of AI TID vs. Placebo BID and AI TID vs. Placebo TID were entirely inconsistent. Similarly, comparisons of AI BID vs. Placebo BID and AI BID vs. Placebo TID were also inconsistent (Table 6).

In Study 007, although the primary endpoint of 'Actual Change in Cystic Fibrosis Questionnaire- Revised (CFQ-R) respiratory domain scores at Day 28' was met and supported by findings from several secondary endpoints, there were still limitations with the evidence presented. First, there are recent Agency concerns regarding the validity of the CFQ-R instrument due to the potential for recall bias. This may limit the strength and interpretability of study findings. Second, primary analysis results were not considered robust since patients 18 years or older and patients with greater disease severity demonstrated only a marginal benefit from AI TID therapy. Finally, Study 007 patients

in the AI TID regimen failed to achieve a significant finding in the key secondary endpoint of 'Time to Need for Inhaled or IV Antibiotics due to Pre-defined Symptoms." Since this endpoint was considered as primary in Study 005, significant findings in both studies could have provided meaningful evidence regarding a treatment benefit for AI patients. This endpoint, however, was not found to be significant in either of the 005 and 007 studies.

In summary, due to the lack of substantial evidence presented in this submission, we recommend that an additional adequate and well controlled Phase III study be conducted to demonstrate; (1) a reduction in the 'Time to Need for IV or Inhaled Antibiotics due to Pre-defined Symptoms' and (2) 'Actual Change in CFQ-R Respiratory Domain Scores at Day 28' as co-primary endpoints. A sequential testing procedure or other methodologies may be considered to control for the overall type-I error rate due to multiple testing. It is also recommended that the additional study utilizes an updated version of the CFQ-R (respiratory domain) questionnaire that is appropriately revised and validated to address recent Agency concerns with the instrument.

As a note, the Sponsor may consider whether treatment regimens involving a higher dose (e.g. AI 150mg) and less frequent dosing regimen (e.g. BID) would provide a more optimal dosing regimen. Possible study options could include a two arm study with AI 75mg TID vs. Placebo TID; or a four-arm trial comparing the AI 75mg TID, AI 150mg BID and the corresponding placebo treatment regimens. Including a higher dose of AI may provide an alternative to patients if AI 75mg is not found to be the optimal dose.

1.3 Brief Overview of Clinical Studies

Both Studies 005 and 007 were Phase 3 randomized, double-blind studies designed to assess the safety and efficacy of a 28 day course of 75mg aztreonam inhalation (AI) versus placebo in cystic fibrosis (CF) patients aged 6 years or older with lung disease due to *PA*. Included patients also had FEV_1% predicted measurements between 25% and 75%. Study 005 (N=211) assessed both the BID and TID regimens of AI against placebo (BID and TID) with 2:2:1:1 randomization whereas Study 007 (N=164) assessed only the TID regimen of AI against placebo (1:1). The primary endpoint of Study 005 was 'Time to Need for Inhaled or IV Antibiotics After Initial AI/Placebo Dosing (Day 0)' whereas the primary endpoint of Study 007 was 'Actual Change in CFQ-R Respiratory Domain Scores at End of Treatment (Day 28)'. Patients in Study 005 were randomized prior to receiving one 28 day course of Tobramycin solution for inhalation (TSI) and one subsequent 28 day course of AI/Placebo and then followed for 56 days post-therapy. Patients in Study 007 were randomized prior to receiving one 28 day course of AI/Placebo and followed for 14 days post-therapy.

1.3 Statistical Issues and Findings

The following are statistical issues and findings identified in this submission. Issues identified were categorized according to the study or studies most relevant (e.g. 'Studies

005 & 007 (Combined)' vs. 'Study 005' vs. 'Study 007)') and by the type of issue ('Study Design' vs. 'Study Analysis' vs. 'Study Evidence' related).

Studies 005 & 007 (Combined)

Study 005 & 007 (Combined) Design Issues

- Studies 005 and 007 analyzed different primary endpoints.

  Statistical Reviewer Comments: *Overall strength of evidence would be stronger with two studies analyzing the same primary endpoint. However, each of the studies did consider the primary endpoint of the other study as a key secondary endpoint. Therefore, significant findings for the Study 005 and Study 007 primary endpoints which are replicated in the Study 007 and Study 005 secondary analyses would be critical in demonstrating an AI TID treatment benefit.*

- Studies 005 and 007 do not aim to address whether efficacy benefits achieved using AI therapy can be considered to be non-inferior to benefits achieved using the standard therapy of Tobramycin Solution for Inhalation (TSI) therapy.

  Statistical Reviewer Comments: *Studies 005 and 007 were designed to show a benefit of AI therapy over placebo rather than compare the efficacy of AI therapy versus TSI therapy. Although a non-inferiority (NI) study of AI vs. TSI had been previously suggested by the Agency, the Sponsor had opted away from such a NI study due primarily to difficulties with enrolling an adequate number of CF patients.*

Study 005 & 007 (Combined) Analysis Issues

- Studies 005 and 007 involved secondary analyses which tested a large number of secondary endpoints often without controlling for multiplicity for most or all of these endpoints.

  Statistical Reviewer Comments: *Inferences regarding secondary endpoints in Studies 005 and 007 were highly limited since there was often no control for the inflation of the overall type I error rate that resulted from multiple testing. Study 007 did pre-specify statistical control of four secondary endpoints tested whereas Study 005 failed to pre-specify statistical control for any of the secondary endpoints tested.*

Study 005 & 007 (Combined) Evidence Issues

- Overall evidence presented for Studies 005 and 007 failed to demonstrate a substantial treatment benefit for patients with cystic fibrosis due to *Psuedomonas aeruginosa* (*PA*) using 75mg AI TID.

  Statistical Reviewer Comments: *Demonstration of treatment efficacy requires replicative evidence from two or more adequate and well controlled trials. Such*

replicative evidence was not provided in this submission from Studies 005 and 007. This was primarily due to Study 005 primary analysis results which failed to show significance in the AI TID regimen.

- Patients in the AI TID arm failed to achieve a significant improvement versus placebo in the ‘‘Time to Need for Inhaled or IV Antibiotics due to Pre-defined Symptoms’ endpoint in either of the 005 and 007 studies.

  Statistical Reviewer Comments: *The above endpoint was defined as ‘primary’ in Study 005 and as ‘secondary’ in Study 007 but was not found to be significant in either Study 005 or Study 007. This finding appears inconsistent with a hypothesis of a substantial ‘AI TID’ treatment benefit over placebo. However, it should also be noted that Study 007 was not powered appropriately for showing significance in this endpoint since the follow-up period involved a 14 day follow-up period that was considerably shorter than the 84 day follow-up period used in Study 005.*

- Patients in the AI TID arm failed to achieve a significant improvement versus placebo in the ‘Actual Change in CFQ-R Respiratory Domain Scores at Day 28’ endpoint that was observed to be consistent across both the 005 and 007 studies.

  Statistical Reviewer Comments: *The above endpoint was defined as ‘primary’ in Study 007 and as ‘secondary’ in Study 005. It was found to be significant in Study 007 but not in Study 005. This further suggests a lack of consistency in Study 005 and 007 efficacy findings.*

Study 005

Study 005 Design Issues

- The Sponsor’s point of randomization (Day -28) led to increases in variation among study groups due to group differences occurring during the 28 day course of TSI preceding initial AI/placebo dosing (Day 0).

  Statistical Reviewer Comments: *Patients in the Placebo TID arm experienced worse respiratory symptoms as evidenced by a lower CFQ-R respiratory domain score, a lower mean FEV_1 % predicted and a lower mean FEV_1 (L) at Day 0 compared to other treatment arms. Such an imbalance in respiratory symptoms could potentially lead to lower estimates for Placebo TID patients who may be more likely to require treatment with IV or inhaled antibiotics. Note that randomization at Day 0 after the 28 day course of TSI instead of Day -28 would offer the advantage of minimizing differences among study groups at baseline.*

- Under the Sponsor’s 005 study design, patients in the AI regimen received 56 days of continuous medication (i.e. 28 days TSI followed by 28 days AI). This design feature may be problematic due to potential carryover effects of TSI which may confound the

treatment drug effect (AI vs. Placebo) as well as the regimen effect (BID vs. TID). Many patients would also have to endure a longer period of therapy (i.e. 56 days) than would be typical (e.g. 28 days).

Statistical Reviewer Comments: *An alternative design would be to allow a 28 day off cycle between the 28 day TSI therapy and the start of the 28 day AI/placebo therapy. This would limit the potential for confounding due to carryover effects and allow most patients to keep their usual routine of alternating cycles involving 28 days on therapy and 28 days off therapy. Additionally, more direct comparisons regarding patient improvement during the TSI cycle vs. patient improvement during the AI/placebo cycle would be possible. Note that the point of randomization should be immediately prior to the start of the AI/Placebo cycle.*

Study 005 Analysis Issues

- The Sponsor's primary and secondary analyses included comparisons based on 'Pooled AI'.

  Statistical Reviewer Comments: *Such comparisons would be problematic in drawing inferences regarding the treatment regimen of interest, AI TID, since patients in the AI TID group generally performed worse than those in the AI BID group with respect to most efficacy endpoints.*

- The Sponsor's primary analysis considers each of the AI regimens against 'Pooled Placebo'. Pooling of the Placebo BID and Placebo TID regimens is not justified in the primary analysis.

  Statistical Reviewer Comments: *Based on the primary endpoint, Placebo BID patients fared significantly better than Placebo TID patients (p=.0043) suggesting separate rather than common distributions for Placebo BID and Placebo TID patients. Under the scenario that the true underlying placebo rate for patients follows a Placebo BID rather than a Pooled Placebo distribution, substantial inflation of the overall type I error can occur when testing AI vs. Pooled Placebo. Note that FDA analyses assume separate distributions of Placebo BID and Placebo TID in the primary analysis.*

- The Sponsor's use of the log rank test for pair-wise comparisons of regimens of the primary endpoint, such as (Survival of) AI TID vs. (Survival of) Placebo BID, inappropriately tests for a two-sided alternative (below):

  - H_{A1}: Survival: AI TID > Placebo BID, if $Z_{AI\ TID} = -Z_{Placebo} > 0$

    OR

  - H_{A2}: Survival: Placebo BID > AI TID, if $Z_{Placebo\ BID} = -Z_{AI\ TID} > 0$

  Since only the survival benefit of AI TID therapy over Placebo BID is of interest and not the survival benefit of Placebo BID therapy over AI TID therapy as shown by

H_{A2}, a one-sided alternative H_A (using a two-sided test) would be more appropriate as shown below:

- H_A: Survival: AI TID > Placebo BID, for all $Z_{AI\ TID}$ values

Statistical Reviewer Comments: *Survival in the AI TID regimen ($Z_{AI\ TID}$= -0.61636) was less favorable than in the Placebo BID regimen ($Z_{Placebo\ BID}$ = 0.61636). However, the Sponsor's computed p-value was based on a two-sided alternative using* 'H$_{A2}$: Placebo BID > AI TID' which was *not considered relevant in testing a survival benefit from AI TID therapy (as mentioned above). Consequently, the Sponsor's p-value (two-sided) = 2 *Pr ($Z > Z_{Pl\ BID}$= 0.61636) = 0.5377. The FDA computed p-value of 0.9999 used a one-sided alternative* 'H$_A$: AI TID > Placebo BID' (*with a two sided test) which estimated the p-value (two-sided) = 2 *Pr ($Z > Z_{AI\ TID}$ = -0.61636)) > 0.9999 ≈1.000 (upper bound). Note that FDA and Sponsor computations were identical for all other pairwise comparisons shown in Table 6. In these cases, Z values computed for AI were greater than 0 and use of the two-sided alternative was operationally equivalent to the one-sided alternative.*

Study 005 Evidence Issues

- ITT patients in the AI TID regimen failed to achieve significant improvement versus 'Pooled Placebo' in the Sponsor's analysis of the primary endpoint, 'Time to Need for Inhaled or IV Antibiotics.'

  Statistical Reviewer Comments: *Although 'Pooled Placebo' is not recommended as an appropriate comparison group, failure to achieve significance against 'Pooled Placebo' provides evidence which is inconsistent with the hypothesis of a treatment benefit from the use of AI TID.*

- Primary analysis results and interpretations were unclear due to an unexpected regimen effect (i.e. effect from BID vs. TID dosing). Additionally, this regimen effect was observed to be stronger than the AI treatment effect.

  Statistical Reviewer Comments: *We would not expect the benefit from BID vs. TID dosing to be significantly different from placebo (p=.0012) and especially not more significant than the treatment effect (AI vs. Placebo therapy) such that patients taking Placebo BID would actually fare better than patients on an AI TID regimen (Figure 4).* Additionally, s*ince the benefit from BID vs. TID dosing was strong in the placebo arms (p=.0043), treatment comparisons against Placebo BID, Placebo TID and Pooled Placebo could not be clearly interpreted as they failed to provide consistent findings of treatment efficacy in the Sponsor's primary analysis for any of the AI treatment arms tested (AI BID, AI TID or Pooled AI).*

- The majority of the secondary endpoints failed to show significance for the AI TID arm in comparison to Pooled Placebo. Endpoints found to be generally supportive of an AI benefit over placebo were not found to be robust.

Statistical Reviewer Comments: *Changes in pulmonary function and changes in log $_{10}$ PA CFU density in sputum were generally supportive of a treatment benefit of AI TID over Pooled Placebo. These endpoints were not considered robust, however. For example, change from Day 0 of FEV_1 (L) at Day 28 (using observed case data rather than LOCF imputation) failed to show significance in AI BID and AI TID arms. Change from Day 0 at Day 28 in log_{10} CFUs failed to show significance in either of the AI TID vs. Placebo TID or AI TID vs. Placebo BID comparisons. Additionally, all other secondary endpoints failed to show significance and appeared inconsistent with a treatment benefit of AI TID over Pooled Placebo. Note that there is no control of the overall type I error rate involved with testing multiple secondary endpoints.*

Study 007

Study 007 Design Issues

- Study 007 had performed an interim analysis involving sample size re-estimation performed by an unblinded independent third party based on variability of CFQ-R results. The actual sample size used in Study 007 (n=140) differed from the sample size recommended by the third party (n=150).

  Statistical Reviewer Comments: Sample size re-estimation (SSR) conducted by an unblinded third party creates concerns of potential operational bias. There are additional concerns when the actual sample size used (e.g. 140 subjects) differs from what is dictated by the SSR conducted by the third party (e.g. 150 subjects). According to the Sponsor, this difference in sample size resulted from operational difficulties with meeting the recommended patient enrollment of 150 subjects. Based on the review of Study 007 data, the Sponsor's sample size selection of 140 subjects did not appear to affect the statistical validity for findings reported in Study 007.

- Issues with the CFQ-R instrument used in measuring the primary endpoint have been identified by the FDA Study Endpoints and Label Development (SEALD) team.

  Statistical Reviewer Comments: *Based on language used in the CFQ-R instrument, there are concerns of potential recall bias. The Agency does not recommend that the current version of the CFQ-R instrument is used in future clinical studies. Also refer to the SEALD team review of Dr. Elektra Papadopoulos.*

  Study 007 Analysis Issues

  The Sponsor's primary endpoint ''Actual Change in CFQ-R Respiratory Domain Scores at End of Treatment (Day 28)' was analyzed based on analysis of covariance (ANCOVA) with Day 0 values and Day -14 disease severity as covariates.

  Statistical Reviewer Comments: *We generally recommend that primary analyses are conducted using unadjusted analyses. Covariates such as disease severity at Day -14*

should be controlled for by the randomization and should not be used in the primary analysis. Unadjusted analyses were also performed by the Sponsor and were found to be generally consistent with primary analysis results.

Study 007 Evidence Issues

- Patients in the AI TID regimen failed to achieve a significant improvement over patients in the placebo TID arm in the ''Time to Need for Inhaled or IV Antibiotics due to Pre-defined Symptoms' endpoint.

  Statistical Reviewer Comments: *Study 007 may not have been powered appropriately for showing significance in this endpoint due to short study follow-up period of only 14 days.*

- Subgroup analyses show a strong influence from the 'disease severity' and 'age' variables on primary analysis results. This raises concerns regarding the robustness of the AI TID treatment effect in older patients or patients with greater disease severity (e.g. FEV_1% predicted less than 50%) since primary analysis results were only marginally significant or non-significant in these populations.

  Statistical Reviewer Comments: For 'age', primary analysis results were significant in patients <18 years of age (p=.0006) and marginally significant in patients ≥ 18 yrs (p=.0495). For disease severity, the AI benefit was smaller in patients with FEV_1% predicted less than 50% (p=.0839) and larger in patients with FEV_1% predicted greater than 50% (p=.0018).

- Primary analysis results reported in Study 007 may be considered less reliable in light of issues (e.g. potential recall bias) raised by the FDA SEALD team concerning the validity of the CFQ-R instrument.

  Statistical Reviewer Comments: *Currently there is not clear evidence available to determine the effect that issues with CFQ-R instrument may have had on primary analysis results. However, given the strength and robustness of CFQ-R findings observed in at Days 28 and Day 42. It does not appear likely under a well-controlled study design that issues identified by the SEALD team would significantly affect overall Study 007 findings.*

2. INTRODUCTION

2.1 Class and Indication

According to the Sponsor, AI is a novel formulation of a synthetic monobactam antibiotic aztreonam, which has been used extensively as parenteral therapy for infections caused by a wide range of gram-negative bacteria. The parenteral form of aztreonam, aztreonam for infection (Azactam®) has been approved for use in the US since the mid 1980s. Azactam is indicated for the following infections caused by gram-negative organisms: urinary tract infections, lower respiratory tract infections, septicemia, skin and

skin structure infections, intra-abdominal infections, and gynecological infections. It is also indicated for adjunctive therapy to surgery caused by susceptible organisms.

2.2 Sponsor's Rationale

Tobramycin Solution for Inhalation (TSI) is the only FDA approved antibiotic solution as an aerosol and it remains the most widely used aerosolized antibiotic for treatment of CF patients. The Sponsor argues that AI will give patients another option with its activity against gram-negative bacteria, especially *P. aeruginosa,* and its safety profile. According to the Sponsor, many physicians prescribe parenteral formulations of various antibiotics for aerosol administration to CF patients. However, such compounded therapies increase the risk of airway toxicity associated with components of the parenteral formulations. For example, the parenteral formulation of tobramycin (Nebcin®) contains phenol, a known respiratory irritant. Likewise, Azactam contains approximately 780 mg arginine per gram. The Sponsor argues that since inhalation will concentrate high levels in the lung as compared to IV administration, it would be unlikely that a product containing arginine would be safe for inhalation. Therefore, the lysine salt was selected for development as the inhaled aztreonam formulation. Lysine is listed as generally recognized as safe by the FDA.

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy (Study 005)

3.1.1 Study Design and Endpoints

Study 005 Key Inclusion Criteria

Patients had to meet all of the following inclusion criteria to be eligible for participation in this trial:

- Male or female patients aged ≥ 6 years diagnosed with CF.
- Patients had to have received ≥ 3 courses of TSI within the previous 12 months.
- *PA* present in expectorated sputum or throat swab culture at Visit 1.
- Patient had to be able to provide written informed consent/assent prior to any trial related procedures. A parent or legally authorized representative/guardian had to be able to give written informed consent as necessary prior to any trial related procedure.
- Patients on chronic azithromycin had to have no change in regimen in the previous months and had to have had need for TSI and/or additional antipseudomonal therapy since initiation of azithromycin.
- FEV1 ≥ 25% and ≤ 75% predicted at Visit 1.
- Ability to perform reproducible pulmonary function tests.
- Arterial oxygen saturation (SaO_2) ≥ 90% on room air at Visit 1.

Study 005 Analysis Populations

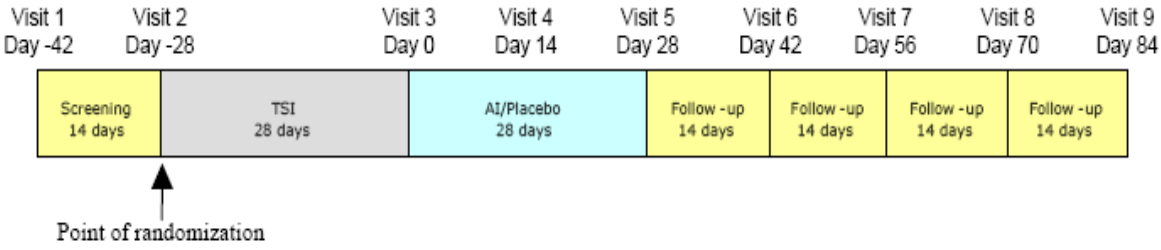
The ITT population was considered the primary analysis population. All primary efficacy and safety analyses were performed on the ITT population. The Safety population was the same as the ITT population.
The PP population was defined as all patients randomized to treatment who received at least one dose of trial drug (AI or placebo), except those identified with at least one of the protocol deviations. Analyses of the primary endpoint and selected secondary endpoints, including the following, were conducted on the PP population:

- Changes in FEV_1 (L)
- Log change in *PA* CFU in sputum
- Clinical symptoms as assessed by the CFQ-R respiratory domain

Study 005 Design

Study 005 randomized 211 patients in a 2:2:1:1 ratio to the following treatment arms for comparison: AI 75mg BID, AI 75mg TID, placebo BID and placebo TID placebo regimens. After initial screening, patients received a 28 day course of TSI therapy immediately followed by a 28 day course of AI/placebo therapy. Patients were followed for 84 days after initial AI/placebo dosing.

Figure 1: Design (Study 005)



Study 005 Efficacy Analyses

All efficacy analyses conducted by the Sponsor were on the pooled AI vs. pooled placebo treatment groups, followed by pair wise comparisons between AI TID and AI BID vs. pooled placebo if the null hypotheses based on pooled data were rejected. If there was a strong suggestion that the two placebo groups were different, they could have been analyzed separately as sensitivity analyses.

Study 005 Primary Endpoint

In Study 005, the primary endpoint was defined as "The Time to Need After Start of AI/placebo Therapy for Inhaled or IV Antibiotics".

The primary analysis of Study 005 first tested $H_{0(Pooled)}$ and if rejected then tested $H_{0(BID)}$ and $H_{0(TID)}$ at the $\alpha = .05$ level.

$H_{0(Pooled)}$: No difference between pooled 75mg AI & pooled placebo
$H_{0(BID)}$: No difference between BID 75mg AI & pooled placebo
$H_{0(TID)}$: No difference between TID 75mg AI & pooled placebo

Study 005 Secondary Endpoints

Secondary endpoints included those related to the following:

- Change in CFQ-R respiratory symptoms domain
- Change in CFQ-R non-respiratory domains
- Change in assessment of symptoms using GRCQ
- Change in pulmonary function
- Change in log 10 *PA* CFU density in sputum
- Change in patient's ability to produce sputum
- Hospitalization
- School/work missed
- Use of other antipseudomonal antibiotics
- Change in CF Symptoms and Severity
- % Change in Weight
- Change in BMI

3.1.2 Subject Disposition, Demographic and Baseline Characteristics

Table 1: Patient Disposition (Study 005)

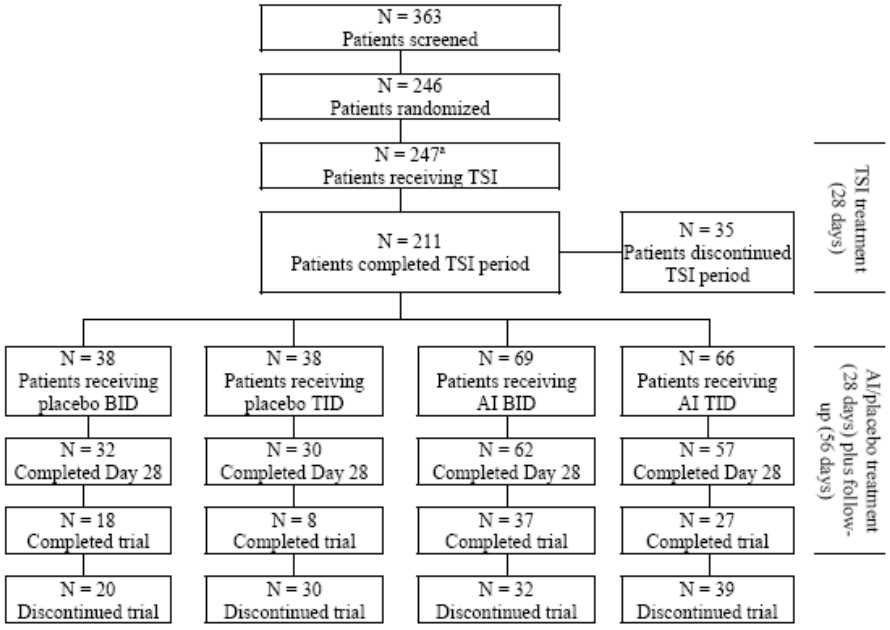
	PLACEBO			AI			TOTAL
	BID n (%)	TID n (%)	Pooled n (%)	BID n (%)	TID n (%)	Pooled n (%)	
No of patients Randomized	41	41	82	82	82	164	246
No of Patients Treated with TSI	41	41	82	82	82	164	247[a]
No of patients: ITT/Safety pop. (Treated with AI/Placebo)	38 (92.7)	38 (92.7)	76 (92.7)	69 (84.1)	66 (80.5)	135 (82.3)	211 (85.8)
No of patients excluded from ITT/Safety population[b]	3 (7.3)	3 (7.3)	6 (7.3)	13 (15.9)	16 (19.5)	29 (17.7)	36 (14.6)
Reason: Unrelated Adverse Event	3 (7.3)	3 (7.3)	6 (7.3)	9 (11.0)	12 (14.6)	21 (12.8)	27 (10.9)
Reason: Other	0	0	0	4 (4.9)	4 (4.9)	8 (4.9)	9 (3.6)

[a] Patient 21208 received TSI but was not randomized and is included in the total summary only.
[b] These patients did not receive at least one dose of AI/Placebo.
Source: Partially Adapted from Sponsor Table 10 of Study Report

Of the 246 randomized patients receiving TSI, 82 were randomized to placebo and 164 were randomized to AI. Of the 164 patients randomized to AI, 135 (82.3%) received at least one dose of AI and were included in the ITT population. Of the 82 patients randomized to placebo, 76 (92.7%) received at least one dose of placebo and were included in the ITT population.

Statistical Reviewer Comments: *The percentage of patients excluded from the AI arms was significantly higher than the percentage of patients excluded from the placebo arms, 29/164 (17.7%) vs. 6/82 (7.3%) with p-value of 0.028.*

Figure 2: Flow Chart of Patient Disposition (Study 005)



Source: Sponsor Figure 2 of Study Report

In the figure above, of the 246 patients randomized and treated with TSI in the trial, 90 (37%) completed the trial and 156 (63%) discontinued. One additional patient who was not randomized discontinued during the TSI period. A higher proportion of patients discontinued in the placebo TID group (81%) in comparison with the other groups (placebo BID group, 56%, AI BID group, 55%, AI TID group, 61%).

Table 2: Demographic Characteristics at Day -42: ITT Population (Study 005)

Variable	Treatment						
	Placebo			AI			
	BID (N = 38)	TID (N = 38)	Pooled (N = 76)	BID (N = 69)	TID (N = 66)	Pooled (N = 135)	Total (N=211)
Gender(n,%)							
Male	26 (68.4)	19 (50.0)	45 (59.2)	38 (55.1)	38 (57.6)	76 (56.3)	121 (57.3)
Female	12 (31.6)	19 (50.0)	31 (40.8)	31 (44.9)	28 (42.4)	59 (43.7)	90 (42.7)
Race, n(%)							
Caucasian	34 (89.5)	35 (92.1)	69 (90.8)	61 (88.4)	63 (95.5)	124 (91.9)	193 (91.5)
African American	0	0	0	3 (4.3)	0	3 (2.2)	3 (1.4)
Hispanic	3 (7.9)	3 (7.9)	6 (7.9)	5 (7.2)	3 (4.5)	8 (5.9)	14 (6.6)

Other	1 (2.6)	0	1 (1.3)	0	0	0	1 (0.5)
Genotype, n(%)							
Homozygous	16 (50.0)	18 (56.3)	34 (53.1)	25 (49.0)	31 (60.8)	56 (54.9)	90 (54.2)
Heterozygous	10 (31.3)	10 (31.3)	20 (31.3)	15 (29.4)	12 (23.5)	27 (26.5)	47 (28.3)
Unidentified	2 (6.3)	3 (9.4)	5 (7.8)	5 (9.8)	7 (13.7)	12 (11.8)	17 (10.2)
Other	4 (12.5)	1 (3.1)	5 (7.8)	6 (11.8)	1 (2.0)	7 (6.9)	12 (7.2)
Age group; n (%)							
≥ 6 years to ≤ 12 years	1 (2.6)	0	1 (1.3)	4 (5.8)	5 (7.6)	9 (6.7)	10 (4.7)
> 12 years to < 18 years	7 (18.4)	4 (10.5)	11 (14.5)	13 (18.8)	12 (18.2)	25 (18.5)	36 (17.1)
≥ 18 years	30 (78.9)	34 (89.5)	64 (84.2)	52 (75.4)	49 (74.2)	101 (74.8)	165 (78.2)
Age (years) Mean (SD)	27.8 (12.0)	28.1 (8.8)	27.9 (10.4)	26.5 (10.7)	24.1 (9.7)	25.3 (10.2)	26.2 (10.4)
Sweat chloride test (mEQ/L)							
Mean (SD)	106.2 (17.5)	103.2 (19.3)	104.8 (18.3)	100.4 (17.8)	103.5 (20.3)	101.9 (19.0)	102.9 (18.7)
n	35	29	64	60	54	114	178
SaO2							
Mean (SD)	96.5 (1.5)	96.8 (2.3)	96.6 (2.0)	96.5 (1.7)	96.2 (1.8)	96.3 (1.7)	96.5 (1.8)
n	38	38	76	69	66	135	211

Source: Sponsor Table 7 of Study Report

In the table above, there was a slightly greater percentage of males in the placebo BID group in comparison with the other groups, however, there was no statistically significant difference among the pooled placebo, AI BID, and AI TID groups (p = 0.8807). The majority of patients were Caucasian. There were no major differences for other variables. Most of the patients in the trial were adults, with 78% being at least 18 years old. There was a higher percentage of children (patients in the ≥ 6 years to ≤ 12 years category) in the pooled AI group (9 [7%] patients) than in the pooled placebo group (1 [1.3%] patient). However, the treatment comparison for age was not statistically significant (p = 0.3470). The demographic characteristics of patients in the PP and Enrolled populations were generally similar to those for patients in the ITT population.

Table 3: Baseline Characteristics at Day 0: ITT Population Study 005

Variable	Treatment					
	Placebo			AI		
	BID (N = 38)	TID (N = 38)	Pooled (N = 76)	BID (N = 69)	TID (N = 66)	Pooled (N = 135)
Mean (SD) FEV_1 (L)	2.005 (0.695)	1.842 (0.678)	1.923 (0.687)	1.899 (0.654)	1.929 (0.725)	1.914 (0.687)
Mean (SD) FEV_1 % predicted	55.655 (17.299)	52.333 (16.258)	53.994 (16.758)	56.192 (15.628)	57.064 (16.608)	56.618 (16.060)
Mean (SD) CFQ-R respiratory domain	65.66 (17.39)	58.71 (21.46)	62.14 (19.74)	63.14 (16.74)	64.23 (18.13)	63.68 (17.38)
MIC of aztreonam all PA isolates (µg/mL)						
MIC_{50}	4	≤1	≤1	2	2	2
MIC_{90}	64	64	64	64	32	32
Minimum MIC	≤1	≤1	≤1	≤1	≤1	≤1
Maximum MIC	512	1024	1024	>2048	1024	>2048
Number of isolates	61	64	125	105	111	216

Source: Sponsor Table 9 of Study Report

Statistical Reviewer Comments: *Although treatment regimens were generally similar with regard to variables at baseline, patients in the Placebo TID arm experienced worse respiratory symptoms as evidenced by a lower CFQ-R respiratory domain score, a lower mean FEV_1 % predicted and a lower mean FEV_1 (L) compared to other treatment arms. Such an imbalance in respiratory symptoms could potentially lead to lower estimates for Placebo TID patients who may be more likely to require treatment with IV or inhaled antibiotics. Note that randomization at Day 0 after the 28 day course of TSI instead of Day -28 would offer the advantage of minimizing such differences among study groups at baseline.*

3.1.3 Statistical Methodologies

Sample Size Considerations

The primary efficacy variable in Study 005 is time to need of a course of inhaled or IV antipseudomonal antibiotics. The null hypothesis is that there is no difference between 75 mg AI and placebo in terms of the distributions of time to need of a course of inhaled or IV antipseudomonal antibiotics. There are four treatment arms: 75 mg AI BID, 75 mg AI TID, Placebo BID, and Placebo TID. The two Al arms were pooled against the pooled placebo to calculate sample size with a 2:1 ratio. Assuming a two-sided significance level of 0.05, approximately 210 patients (70 in each Al group, 35 in each placebo group) will be required to provide > 90% power to reject the null hypothesis. Therefore, 250 patients will be randomized at Visit 2 (Day -28) to ensure that at least 210 patients enter the double-blind treatment period at Visit 3 (Day 0). This estimate is based upon the Log-rank test. Exponential rates of loss to follow-up are assumed as 10% for both treatment groups. The time points and assumed survival rates where survival is defined as no need for antibiotics are listed below and are based on previous experience:

Table 4: Sponsor's Assumed Times without Need for Antibiotics (Study 005)

Time	Survival Rate	
	Pooled Placebo	Pooled 75 mg AI
Day 0	1.00	1.00
Day 14	0.98	1.00
Day 28	0.83	0.97
Day 42	0.73	0.91
Day 56	0.63	0.85
Day 70	0.47	0.79
Day 84	0.45	0.68

Source: Sponsor's Table 6 of Study Report

Interim Analyses
Two interim summaries of safety will be provided to the DSMB. These summaries will include only safety-related data and will be prepared by an unblinded independent party not be involved in the final analysis of the study.

Premature Discontinuation and Missing Data
Missing baseline data will not be imputed. Any patient with missing baseline data will not be included in the analysis. Data collected at the Early Termination (ET) Visit will be included in the analyses and assigned to visits. Missing individual CFQ-R values will be imputed according to the method described in the protocol. CFU data values where PA was not isolated from a valid culture will be set to zero (0). Missing spirometry, CFQ-R (domain scores), and CFU data will be imputed from Day 14 to Day 42 according to the following rules. If a patient withdraws from the study due to an adverse event or study drug intolerance and an assessment value is not available on the withdrawal date, then the worst case value (from the baseline value up to the latest available value) will be used to impute subsequent missing data values, up through Day 42. For all other reasons that lead to missing data, last observation carried forward (LOCF) will be used to impute missing data values, up through Day 42.

Multiple Comparisons Adjustment
Multiplicity for multiple treatment group comparisons will be handled using a gatekeeper strategy for each efficacy endpoint as described below. The following three hypotheses are the key hypotheses of interest:
$H_{O(Pooled)}$: There is no difference between pooled *75* mg AI and pooled placebo.
$H_{O(BID)}$: There is no difference between 75 mg AI BID and pooled placebo.
$H_{O(TID)}$: There is no difference between 75 mg AI TID and pooled placebo.
The $H_{O(Pooled)}$ hypothesis will serve as the gatekeeper hypothesis.
If $H_{O(Pooled)}$ is rejected, $H_{O(BID)}$ and $H_{O(TID)}$ will both be tested at the two-sided 0.05 significance level, in parallel (i.e. without further adjustments).

Covariates
All parametric efficacy analyses of continuous endpoints will use the respective baseline

characteristics as covariates unless otherwise stated.

Examination of Subgroups
Subgroup analyses on the primary efficacy variable will be performed using the following subsets on the ITT population:

- Age group at Visit 1 ($\geq$ 6 to $\leq$ 12 years, > 12 to $\leq$ 18 years, or $\geq$ 18 years).
- Gender at Visit 1 (male or female).
- Disease severity at Visit 2 (> 50% or < 50% FEV_1 of predicted).
- Highest aztreonam MIC for PA at Visit 3 (> 8 or < 8 $ug*ml^{-1}$).
- Highest tobramycin MIC for PA at Visit 2 (> 8 or < 8 $ug*ml^{-1}$).

3.1.4 Results

Sponsor's Primary Analysis
The table below provides results of the Sponsor's primary analysis of Study 005 of "Time to Need (Days) for Inhaled or IV Antibiotics Due to Pre-Defined Symptoms (ITT Population). Based on the Study 005 protocol, this primary endpoint was measured from initiation of study drug (Day 0) to the end of study (Day 84). However, cases in which patients required the need for IV or inhaled antibiotics outside the nominal 84 day study period were included in the Sponsor's analysis.

Statistical Reviewer Comments: *Consideration of these cases outside the nominal 84 day study period provided a more conservative analysis and was therefore considered acceptable.*

Table 5: Sponsor's Primary Analysis: Time to Need (Days) for Inhaled or IV Antibiotics Due to Pre-Defined Symptoms: ITT Population (Study 005)

Statistic	Treatment					
	Placebo			AI		
	BID (N = 38)	TID (N = 38)	Pooled (N = 76)	BID (N = 69)	TID (N = 66)	Pooled (N = 135)
Min	10	11	10	3	2	2
25th percentile	59	43	45	77	56	59
Median	-	54	71	-	87[b]	92[b]
95% CI for median	(71, -)	(46, 66)	(57, 97)	(89, -)	(71, -)	(89, -)
75th percentile	-	97	97	-	-	-
Max	71	97	97	92	87	92
Number of censored values	26	12	38	50	42	92
Number of events	12	26	38	19	24	43
p-value[a]				0.0019	0.1816	0.007

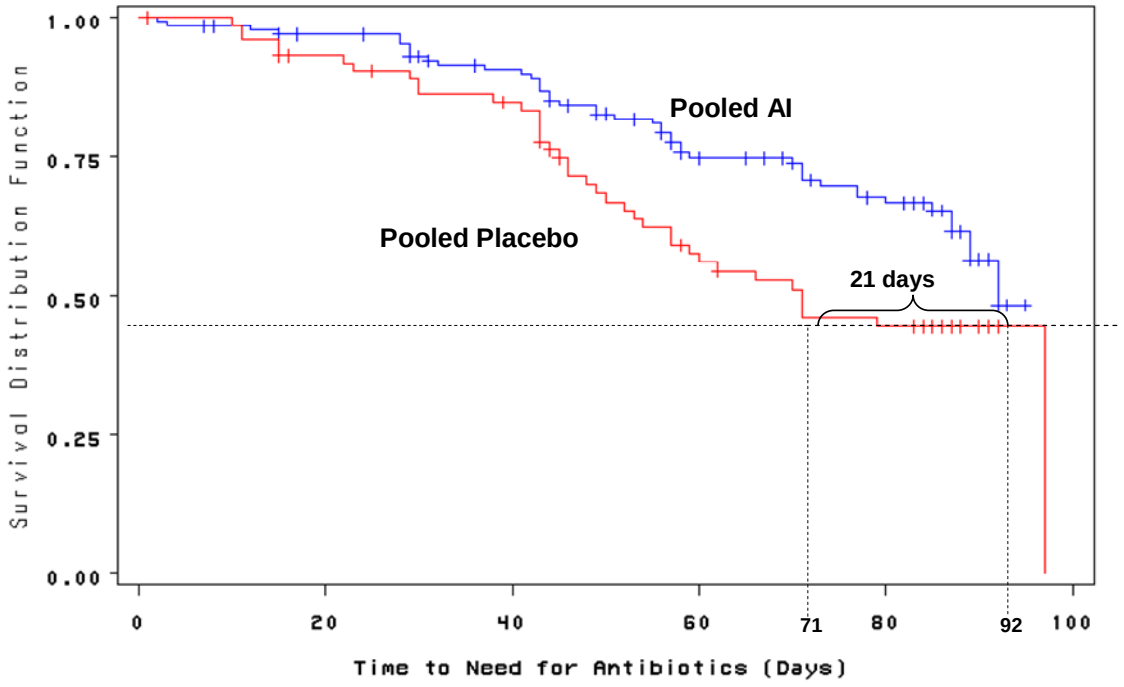
- = not estimable.

[a] All comparisons are made against pooled placebo.
[b] Note that estimate is outside the nominal 84-day study period.
Source: Sponsor Table 16 of Study Report

Statistical Reviewer Comments: *In the Sponsor's analysis of Study 005, there was no clear evidence of a significant benefit in the primary endpoint, 'Time to Need for Inhaled or IV Antibiotics Due to Pre-Defined Symptoms' using AI TID therapy. Comparisons of this primary endpoint in AI TID patients versus Pooled Placebo patients were not significant (p=.1816). Although patients in the Pooled AI arms fared significantly better than patients in the Pooled Placebo arms (p=0.0070), these results could not be reliably interpreted due a significant overall regimen effect (p=0.0012) and placebo regimen effect (p=0.0043). Additionally, results based on the pooling of AI BID and AI TID patient arms may not provide valid inferences regarding the efficacy of AI TID. This is because efficacy of AI TID therapy cannot be assumed to be equal to or greater than efficacy with AI BID therapy. In Study 005, findings from primary and secondary endpoints showed patients on AI BID therapy as generally performing better patients on AI TID therapy.*

The figure below depicts the Sponsor's Study 005 primary analysis of 'Time to Need (Days) for Inhaled or IV Antibiotics Due to Pre-Defined Symptoms (ITT Population)'. This figure depicts only the pooled placebo and AI treatment groups. According to the Sponsor's primary analysis, the proportion of patients not requiring IV or inhaled antibiotics over time after starting AI/Placebo regimens (or survival rate) was significantly higher in the pooled AI treatment arms versus the pooled placebo arms. The figure above shows an increase in the median time to need for IV or inhaled antibiotics of approximately 21 days for patients taking AI therapy versus placebo therapy.

Figure 3: Sponsor's Primary Analysis: Time to Need (Days) for Inhaled or IV Antibiotics Due to Pre-Defined Symptoms: ITT Population (Study 005)



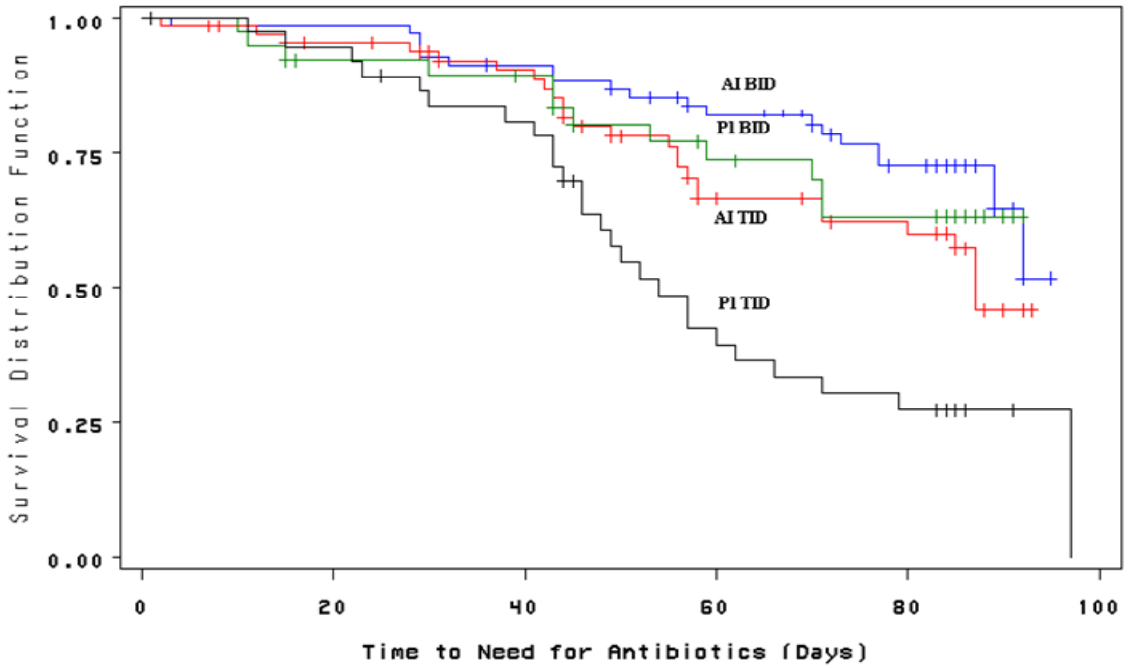
Source: FDA Figure

Statistical Reviewer Comments: *Since comparisons involving pooling across either the placebo or AI regimens may be problematic, the above figure cannot be meaningfully interpreted. Consequently, FDA primary analyses were also conducted to assess survival rates in each of the four treatment arms (without pooling) and are included below.*

FDA Primary Analysis

Due to overall evidence of a "regimen effect" in the primary analysis which was stronger than the effect due to treatment, the FDA primary analysis considered the survival distributions of each of the four treatment arms without considering pooling of treatment arms. These survival distributions are depicted in the figure below.

Figure 4: FDA Primary Analysis: Time to Need (Days) for Inhaled or IV Antibiotics Due to Pre-Defined Symptoms: ITT Population (Study 005)

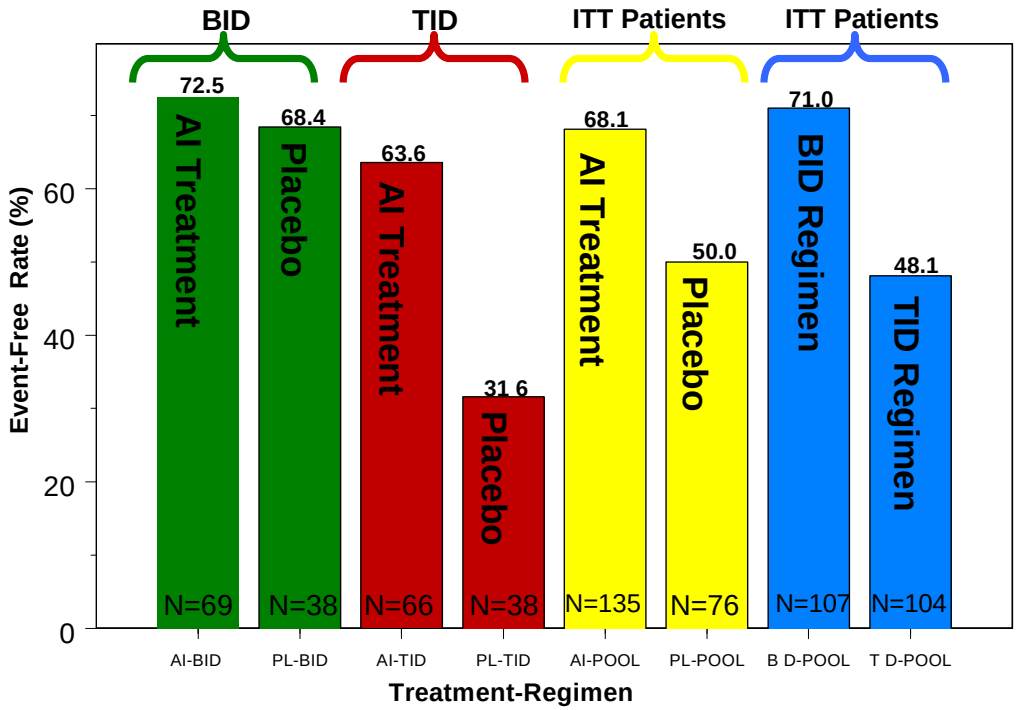


Source: FDA Figure

Statistical Reviewer Comments: *In the figure above, ITT patients on a BID regimen achieved a greater time to need for inhaled or IV antibiotics due to pre-defined symptoms than patients on a TID regimen, regardless of whether an AI or placebo regimen was received. Note that interpretations of the effect due to AI are not clear from the figure above. Although patients on a TID regimen benefited substantially from receiving AI rather than placebo, a substantial benefit was not observed for patients on a BID regimen receiving AI rather than placebo. Moreover, patients on an AI TID regimen did not fare as well as patients on a placebo BID regimen with respect to the primary endpoint.*

A sensitivity analysis was conducted which re-classified patients in the primary analysis as either having an event (i.e. requiring the need for inhaled or IV antibiotics) versus not having an event. Differences in the AI BID, placebo BID, AI TID and placebo regimens were similar to those differences observed in the primary analysis described earlier. Similar to the primary analysis, there is strong evidence of a "regimen effect" which makes inferences regarding an effect due to AI TID therapy unclear.

Figure 5: FDA Comparison of Event-Free Rates (%) by Treatment Regimens (Study 005)



Source: FDA Figure

Statistical Reviewer Comments: *Interpretations of the effect due to AI are not clear from the figure above which shows event-free rates for treatment regimens. Although patients on a TID regimen benefited substantially from receiving AI rather than placebo (31.6% vs. 63.6%), similar benefits were not observed for patients on a BID regimen receiving AI rather than placebo (68.4% vs. 72.5%). Moreover, patients on an AI TID regimen did not fare as well as patients on a placebo BID regimen with respect to event-free rates (63.6% vs. 68.4%). Note that event-free rates were driven primarily by the regimen used (BID (71.0%) vs. TID (48.1%) regardless of the treatment (AI or placebo) received.*

The table below shows how different assumptions regarding the placebo rate would affect significance testing in the primary analysis. The Sponsor's primary analysis had pre-specified three comparisons, each of the AI BID, AI TID and Pooled AI regimens compared to Pooled Placebo with respect to the primary endpoint of "Time to Need for IV or Inhaled Antibiotics". However, due to significant differences observed in patients on the placebo BID with patients on the placebo TID regimen, a common distribution for patients receiving a placebo regimen cannot be assumed and would fail to justify pooling of the placebo BID and TID regimens. Therefore, FDA analyses considered the following comparisons as most relevant for assessing for the primary endpoint:

- AI BID versus Placebo BID
- AI BID versus Placebo TID
- AI TID versus Placebo BID
- AI TID versus Placebo TID

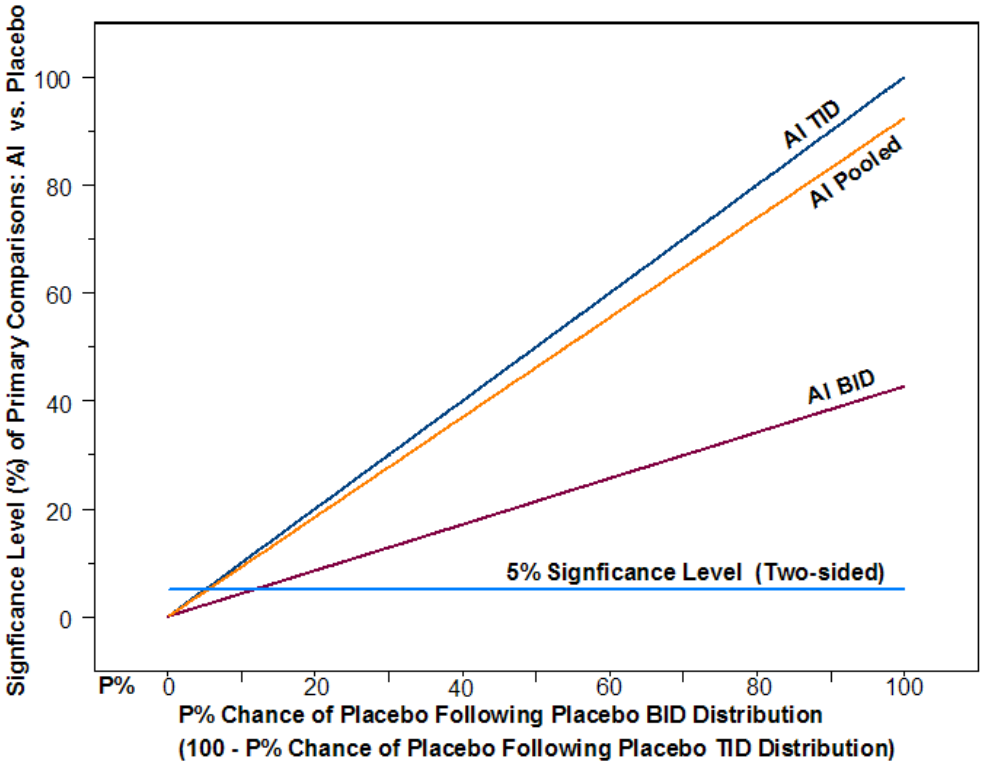
Table 6: FDA Primary Analysis: Significance Testing of AI Regimens Against Possible Placebo Regimens Based on Two-Sided P-values (Study 005)

AI → vs. Placebo ↓	AI BID	AI TID	AI Pooled
Placebo BID	**0.4269**	**> 0.9999***	**0.9240**
Placebo TID	**< .0001**	**0.0043**	**<.0001**
Placebo Pooled	**0.0019**	**0.1816**	**0.0070**

*This p-value was computed using a one-sided alternative (H_A: Survival $_{AI\ TID}$ > Survival Placebo $_{BID}$) based on a two-sided test.
Source: FDA Figure

Statistical Reviewer Comments: *Since pooling of placebo and/or AI regimens may be problematic, the FDA analysis did not give primary consideration to pooled analysis results of the outer row/column in the figure above. Instead, the FDA analysis focused on comparisons which did not consider pooling. This included AI BID versus Placebo BID (p=0.4269), AI TID versus Placebo BID (p=0.9999), AI BID versus Placebo TID (p <0.0001) and AI TID versus Placebo TID (p=0.0043). In the table above, significance of either AI BID or AI TID largely depends upon whether we assume Placebo BID or Placebo TID as the appropriate comparison group. Since it is unclear as to which placebo estimate (i.e. Placebo TID or Placebo BID) most reliably estimates the underlying placebo rate, any conservative analysis should adequately control against the possibility that the underlying (true) placebo rate is best estimated by Placebo BID. Even under the assumption that the true placebo rate is no more than 50% likely to follow a Placebo BID vs. Placebo TID estimate, the expected type I error rate would still be substantially above 5% in testing the primary hypothesis for the AI TID regimen (as shown in the figure below). Based on the FDA analysis of the primary endpoint of Study 005, there is no clear evidence to suggest a treatment benefit with the use of AI therapy.*

Figure 6: Significance Level (%) of Primary Analysis Comparisons: Pooled AI, AI BID & AI TID vs. Placebo Assuming a P% Chance that Placebo Follows a Placebo BID Distribution (or 100 – P% Chance Placebo Follows a Placebo TID Distribution), Study 005



Statistical Reviewer Comments: *From the figure above, significance of AI vs. Placebo would depend upon assumptions of Placebo being much more likely to follow a Placebo TID distribution than a Placebo BID distribution. Such an assumption would not be appropriate in any conservative analysis. While the assumption that Placebo may follow a Placebo BID or Placebo TID distribution with similar likelihood is not necessarily conservative, we can see that under this assumption all AI comparisons fail to approach significance at the 5% level. This suggests a lack of adequate evidence regarding any treatment benefit from AI therapy over placebo therapy. For adequate and substantial evidence of a treatment benefit from AI using a conservative analysis, we would expect all AI regimens to show significance when compared to the Placebo BID regimen. In the above figure, this corresponds to AI BID, AI TID and Pooled AI having observed significance levels (two-sided) below 5% at P=100%.*

Table 7: Sponsor's Summary of Efficacy Endpoints (Study 005)

Efficacy Endpoint Key Test or Evaluation	Result			
	Pooled Placebo (N = 76)	Pooled AI (N = 135)	Treat. Diff.	p-value
Median time to need for IV or inhaled antipseudomonal antibiotics (days) (Primary Endpoint)	71	92	21	0.007
Clinical symptoms as assessed by CFQ-R respiratory domain Mean change in CFQ-R respiratory domain score at Day 28	-0.66	4.34	5.01	0.0196
Categorical result: % of patients who improved at Day 28	37.0	51.5	-	0.0289
% of patients who worsened at Day 28	38.4	28.0	-	
Change in pulmonary function Mean percent change in FEV_1 at Day 28	-2.4	3.9	6.3	0.0012
Hospitalization Number (%) of patients hospitalized at least once during trial	3 (3.9)	14 (10.4)	-	NS
Weight Median percent change in weight at Day 28	-0.17	0.38	0.66	0.0377
BMI Mean change in BMI (kg/m2) at Day 28	-0.108	0.066	0.174	0.0362
Change in CF symptoms and severity % of patients showing improvement in respiratory, thoracic and mediastinal disorders SOC CF symptoms at Day 28	1	12	-	-
Change in $\log_{10}$ PA CFUs in sputum Mean change in sputum $\log_{10}$ PA CFUs at Day 28	0.225	-0.434	-0.659	0.0059

Notes: - not applicable; NS = not significant; Treat. Diff. = Treatment Difference.
Source: Sponsor Table 54 of Study Report

Statistical Reviewer Comments: *In the Sponsor's summary of clinical efficacy endpoints above, the primary endpoint computation is problematic for two reasons: (1) Pooled Placebo was used as a comparison group rather than individual Placebo BID and Placebo TID arms and (2) Pooled AI was used as a comparison group rather than individual AI BID and AI TID arms.*

Secondary endpoint computations are also problematic but mainly due to reason (2). Since AI BID effects were generally more favorable than AI TID effects across all endpoints considered, Pooled AI effects would tend to over-estimate the AI TID effect. Finally, it should be noted that the Sponsor has tested numerous secondary endpoints without pre-defined measures to control against inflation of the overall type I error rate. Due to the potential for such type I error inflation, statistical inferences regarding secondary endpoints listed in the above table are limited.

Table 8: Observed P-values of AI regimens versus Placebo for Key Endpoints (Studies 005 and 007)

Study:	**Study 005**		Study 007
Comparison:	**AI BID vs. Pooled Placebo**	**AI TID vs. Pooled Placebo**	AI TID vs. Placebo TID
Endpoint:	**Observed P-values:**		

Median Time to Need for IV/Inhaled Antibiotics Due to Pre-Defined Symptoms (Study 005 Primary Endpoint)	**.0019**[a]	**.1816**	.0949
Actual Change in CFQ-R Respiratory Domain Scores from Day 0 (ITT) at Day 28 (Study 007 Primary Endpoint)	**.0207**	**.0920**	.0005
Actual Change in CFQ-R Respiratory Domain Scores From Day 0 (ITT) at Day 42	**--** [b]	**--** [b]	.0154
Categorized Change in CFQ-R Respiratory Domain Scores From Day 0 (ITT) (Day 28)	**.0299**	**.1405**	.0055
FEV_1% Predicted Actual Change from Day 0 (Day 28)	**.0106**	**.0215**	.0001
FEV_1 (L) Change from Day 0 (Day 28)	**.0060**	**.0052**	.0001
Change in sputum log10 PA CFUs (Day 28)	**.0106**	**.0313**	.0001
Mean (adj) change in Weight (Day 28)	**.0225**	**.3046**	.0039
Mean (adj) change in BMI (Day 28)	**.0181**	**.2367**	.0054
Number of school/work days missed during Study (Day 28)	**.9033**	**.7437**	.1860
Number of Hospitalizations [c]	**.2540**	**.0829**	.0487
Proportion of patients hospitalized [c]	**.3096**	**.1130**	.0640
Proportion of patients with school/work missed during Study (Day 28)	**>.9999**	**.8371**	.2201

[a] Signficance of this result is not clear due to the presence of a regimen effect. Comparison of AI BID treatment with placebo BID was non-signficant (p=.4269) whereas comparisons with placebo TID were signfiicant (p=.0001). [b]Endpoint was not observed at Day 42 in Study 005. [c] Endpoint measured from Day 0 to Day 28 in Study 005 and from Day 0 to Day 42 in Study 007

Source: FDA Table

Statistical Reviewer Comments: In *Study 005, the majority of secondary endpoints failed to show significance at the α =.05 level for the AI TID regimen and appeared inconsistent with results observed in Study 007 (provided above for comparative purposes only). From the endpoints included in the above table for Study 005, the AI TID regimen was significant for only 'FEV$_1$% Predicted Actual Change from Day 0 (Day 28),' 'FEV$_1$ (L) Change from Day 0 (Day 28),' and 'Change in Sputum log$_{10}$ PA CFUs (Day 28)'. Note also that 'Median Time to Need for IV/Inhaled Antibiotics Due to Pre-Defined Symptoms' and 'Actual Change in CFQ-R Respiratory Domain Scores from Day 0 (ITT) at Day 28' were defined as the primary endpoints of Studies 005 and 007, respectively. However, neither of these endpoints was shown to be significant in Study 005 for AI TID compared to Pooled Placebo.*

Finally, it should be noted that in Study 005 the Sponsor has tested numerous secondary endpoints without pre-defined measures to control against inflation of the overall type I error rate. Due to the potential for such type I error inflation, statistical inferences regarding Study 005 secondary endpoints listed in the above table are limited. (Refer to the statistical comments for Table 13 regarding analysis of Study 007 secondary endpoints.)

3.1.5 Conclusions

Evidence presented for Study 005 failed to demonstrate a substantial treatment benefit for patients with cystic fibrosis due to *Psuedomonas aeruginosa* (*PA*) using 75mg aztreonam inhalation (TID). ITT patients in the AI TID arm failed to achieve significant improvement versus Pooled Placebo in 'Time to Need for Inhaled or IV Antibiotics,' the primary endpoint, 'Actual Change in CFQ-R respiratory domain scores at Day 28,' a key secondary endpoint, and several other secondary endpoints. Primary analyses were also unclear due to a "regimen effect" (i.e. BID vs. TID dosing) that was stronger than the treatment effect (AI vs. Placebo). Since this "regimen effect" was especially strong in the placebo arms, treatment comparisons against Placebo BID, Placebo TID and Pooled Placebo failed to provide consistent findings of treatment efficacy in either of the AI BID, AI TID or Pooled AI comparison groups.

3.2 Evaluation of Efficacy (Study 007)

3.2.1 Study Design and Endpoints

Study 007 Inclusion Criteria

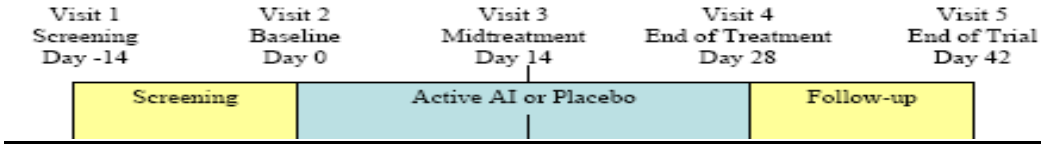
Patients had to meet all of the following inclusion criteria to be eligible for participation in this trial.

- Patients ≥ 6 years of age with documentation of CF diagnosis
- *PA* present in expectorated sputum or throat swab culture at Visit 1, for patients enrolled under protocol versions 1, 2, and 3.
- *PA* present in expectorated sputum or throat swab culture at Visit 1 or documented *PA*
- $FEV_1 \geq 25\%$ and $\leq 75\%$ predicted at Visit 1.
- Patients (and parent/guardian as required) had to be able to provide written informed consent/assent prior to any trial related procedures.
- Females of childbearing potential had to have a negative serum pregnancy test at Visit 1.
- Ability to perform reproducible pulmonary function tests.
- Arterial oxygen saturation (SaO_2) ≥ 90% on room air at Visit 1.

Study 007 Design

Study 007 randomized 164 patients in a 1:1 ratio to either the AI 75mg TID arm or the placebo TID arm. After initial screening, patients underwent a 28 day course of AI or placebo. Patients were followed up to 42 days following initial AI/placebo dosing.

Figure 7: Design (Study 007)



Study 007 Primary Endpoint

In Study 007, the primary endpoint was defined as 'Change in respiratory symptoms as measured by CFQ-R respiratory domain (Day 0 to Day 28)'. The primary analysis tested '$H_{0(TID)}$: No difference between TID 75mg AI and TID placebo.'

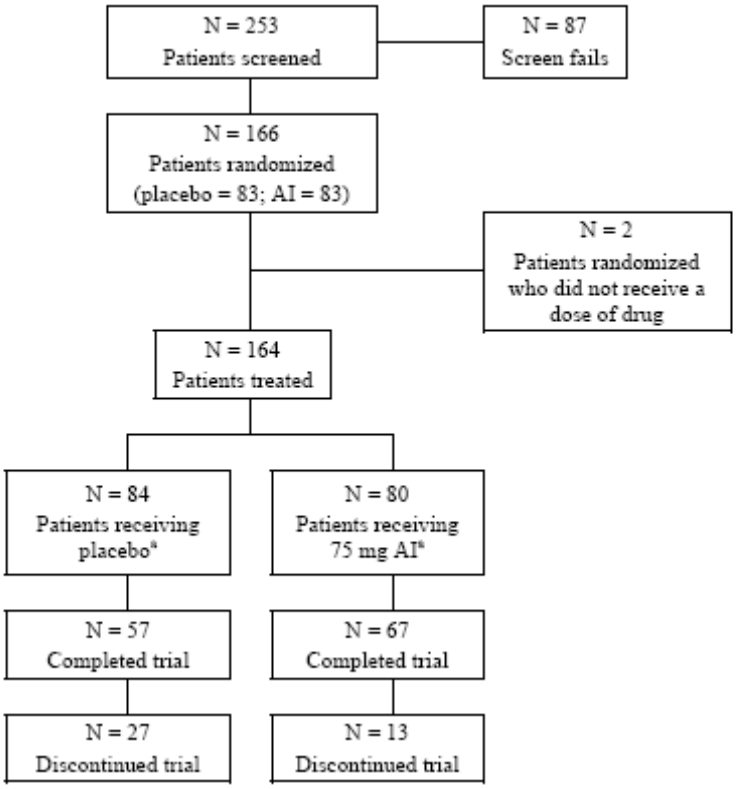
Study 007 Secondary Endpoints

Secondary endpoints included the following:

- Change in CFQ-R respiratory symptoms domain
- Change in CFQ-R non-respiratory domains
- Change in assessment of symptoms using GRCQ
- Change in pulmonary function
- Change in log 10 *PA* CFU density in sputum
- Change in patient's ability to produce sputum
- Hospitalization
- School/work missed
- Use of other antipseudomonal antibiotics
- Change in CF Symptoms and Severity
- % Change in Weight.
- Change in BMI

3.2.2 Subject Disposition, Demographic and Baseline Characteristics

Figure 8: Flow Chart of Patient Disposition (Study 007)



Source: Sponsor Figure 2 of Study Report

Of the 253 patients screened in Study 007, 34% failed to meet screening criteria. The most common reasons for screen failure were no *PA* in sputum sample and $FEV_1 > 75\%$ or < 25%. The disposition of all patients by treatment received and the details of the reasons for withdrawal are presented in the table below. Two patients randomized to receive AI were withdrawn before receiving trial drug. Of the 164 patients who were randomized and received trial drug, 124 (76%) patients completed the trial and 40 (24%) patients discontinued. One patient (Patient 40954) received placebo rather than AI as randomized. For the Safety population, the proportion of patients who discontinued in the AI group was 16% vs. 32% in the placebo group.

Table 9: Disposition of Patients: All Patients by Treatment Received (Study 007)

	Treatment as Received		
	Placebo (N = 84) n (%)	75 mg AI (N = 80) n (%)	Total (N = 164) n (%)
No. of patients screened	-	-	253

No. of patients randomized	83	83	166
No. of patients treated[a]	84 (100.0)	80 (100.0)	164 (100.0)
No. of patients completing the trial	57 (67.9)	67 (83.8)	124 (75.6)
Number of patients discontinued	27 (32.1)	13 (16.3)	40 (24.4)
Reasons for early withdrawal			
Unrelated AE	16 (19.0)	8 (10.0)	24 (14.6)
Trial drug intolerance (AE)	2 (2.4)	0	2 (1.2)
Related AE	5 (6.0)	3 (3.8)	8 (4.9)
Death	0	0	0
Lost to follow-up	0	1 (1.3)	1 (0.6)
Noncompliance	0	0	0
Personal/administrative	1 (1.2)	1 (1.3)	2 (1.2)
Other	3 (3.6)	0	3 (1.8)

a Patients treated are those receiving at least part of one dose of trial drug.
- not applicable.
One patient was randomized to receive AI but received placebo.
Source: Sponsor Table 10 of Study Report

Overall, 11 patients receiving AI discontinued due to adverse events vs. 23 patients in the placebo group. In the placebo group 16 (19%) patients withdrew because of unrelated adverse events compared to 8 (10%) patients in the AI group. The majority of adverse events that led to discontinuation were pulmonary exacerbations.

Statistical Reviewer Comments: *Among patients receiving treatment, a higher rate of early withdrawal was observed in the Placebo arm (32.1%) than in the AI arm (16.3%).*

Table 10: Demographic Characteristics at Day -14: ITT Population (Study 007)

Variable	**Treatment**		**Total (N = 164)**
	Placebo (N=84)	**75 mg AI (N = 80)**	
Gender; n (%)			
Male	45 (53.6)	48 (60.0)	93 (56.7)
Female	39 (46.4)	32 (40.0)	71 (43.3)
Disease severity; n (%)			
FEV_1 % pred ≤ 50%	30 (35.7)	30 (37.5)	60 (36.6)
FEV_1 % pred > 50%	54 (64.3)	50 (62.5)	104 (63.4)
Race; n (%)			
Caucasian	82 (97.6)	76 (95.0)	158 (96.3)
Hispanic	2 (2.4)	4 (5.0)	6 (3.7)
Genotype; n (%)			
Homozygous (ΔF508)	30 (42.9)	38 (54.3)	68 (48.6)
Heterozygous (ΔF508)	22 (31.4)	21 (30.0)	43 (30.7)
Unidentified	18 (25.7)	9 (12.9)	27 (19.3)
Other	0	2 (2.9)	2 (1.4)
n	70	70	140
Age group; n (%)			

≥ 6 years to ≤ 12 years	4 (4.8)	11 (13.8)	15 (9.1)
> 12 years to < 18 years	12 (14.3)	10 (12.5)	22 (13.4)
≥ 18 years	68 (81.0)	59 (73.8)	127 (77.4)
< 18 years	16 (19.0)	21 (26.3)	37 (22.6)
Age (years)			
Mean (± SD)	31.7 (14.8)	27.4 (12.8)	29.6 (14.0)

Source: Sponsor Table 5 of Study Report

The mean age (± SD) of patients in the trial was 29.6 (± 14.0) years with 72% of patients at least 18 years old. The majority of patients in each treatment group were in the milder of the two disease severity categories, with 63% of patients in the AI group and 64% of patients in the placebo group having an FEV_1 greater than 50% of their predicted values at Day -14.

Statistical Reviewer Comments: *There were no significant differences between the AI and placebo groups in any demographic characteristics tested.*

3.2.3 Statistical Methodologies

Primary Analysis

The primary efficacy analyses were conducted on the ITT (imputed data) population. In addition, the analysis of the primary efficacy endpoint was also conducted on the ITT (observed case data) and PP (imputed and observed case data) populations. The actual change at Day 28 from Day 0 in the CFQ-R respiratory domain (Child/Teen/Adult combined) was analyzed using analysis of covariance (ANCOVA), with Day 0 value and disease severity at Day -14 as covariates. ANCOVA was planned as the primary method used to analyze the CFQ-R endpoint. In the case that the ANCOVA analysis was conducted, a sensitivity analysis was to be conducted using nonparametric ANCOVA.

Interim Analyses

A planned sample size re-estimation (SSR) analysis was conducted by an independent third party to monitor assumptions of variability that affect study power. The SSR used estimates of the actual change from baseline CFQ-R variances for each of the two treatment groups. Other than using the variances from the first 40-50 patients, all other protocol assumptions did not change.

Sample Size Adjustment

The variance of the primary endpoint observed in the ongoing Study 007 was larger than the variance assumed in the original sample size calculation, this lead to a sample size increase from approximately 58 to 70 subjects per treatment group.

Covariates

All parametric efficacy analyses of continuous endpoints will use disease severity from Visit 1 (Day -14) for stratification and the respective baseline characteristics as covariates unless stated otherwise.

Analysis of subgroups

Subgroup analyses were performed to investigate the effects of age , gender, country, baseline for each respective endpoint, disease severity (FEV_1 % Predicted-- ≤ 50% or > 50%) , highest aztreonam MIC for PA (> 8, ≤ 8 µg/ml)

Missing Data

Missing baseline data will not be imputed and not included in the analysis. Missing spirometry and CFQ-R data for non-baseline visits will be imputed based on the worst case value (from the baseline value up to the latest available value). For all other reasons, LOCF will be used to impute missing data values.

Multi-center Study

Due to the rarity of CF disease, the majority of centers are expected to enroll small numbers of patients. Since large within-center variance is expected, the investigative center effect will not be analyzed in any statistical models.

Multiple-Endpoint Adjustment

To preserve the Type I error rate, a gatekeeper approach will be used to test multiple endpoints. For secondary endpoints, type I error is preserved by testing the following:
1) the percent change in FEV_1 (from Day 0 to Day 28) at the α =.025 level. If this is significant then test change in *PA* CFUs in sputum at the α =.025 level.
2) the proportion of patients receiving IV or inhaled antipsueomonal antibiotics other than study drug through Visit 5 (nominal Day 42) at the α =.025 level. If significant the proportions of patients hospitalized through Visit 5 will be compared at the α =.025 level.

3.2.4 Results

The Sponsor's primary analysis of Study 007 compared the 75mg AI TID regimen to the Placebo TID regimen according to the 'Actual Change from Visit 2 (Day 0) in CFQ-R Respiratory Scores at Visit 4 (Day 28) on Imputed Data- Child and Teen/Adult Versions Combined.' Results of the primary analysis are shown in the table below.

Table 11: Primary Analysis: Actual Change from Visit 2 (Day 0) in CFQ-R Respiratory Scores at Visit 4 (Day 28) on Imputed Data (Study 007)

Primary Analysis	**ITT Patients**	
Treatment	**Placebo N=84**	**75 mg AI N=80**
Mean (± SD)	-1.91±18.64	7.88±18.88
n	83	80
Adjusted mean	-2.63	7.08
Treatment difference: AI – placebo	9.71	
95% CI (p-value)	4.31,15.11 (p-value = 0.0005)	

Source: FDA Table

Statistical Reviewer Comments: *ITT patients in the 75mg AI TID treatment regimen achieved significantly higher CFQ-R respiratory scores at Day 28 compared with*

patients in the placebo TID regimen. The treatment difference of 9.71 (p =.0005) at Day 28 reflects a clinically significant improvement in respiratory symptoms for AI-treated patients based on the MCID estimate of 5 established in Study 005. Similar findings were also observed when analyzing the primary endpoint based on per protocol rather than ITT patients (p=.0009), observed case data rather than imputed data (p=.0005), and non-parametric analyses rather than parametric analyses (p=.0013).

Table 12: Actual Change from Visit 2 (Day 0) in CFQ-R Respiratory Scores at Visit 5 (Day 42) on Imputed Data (Study 007)

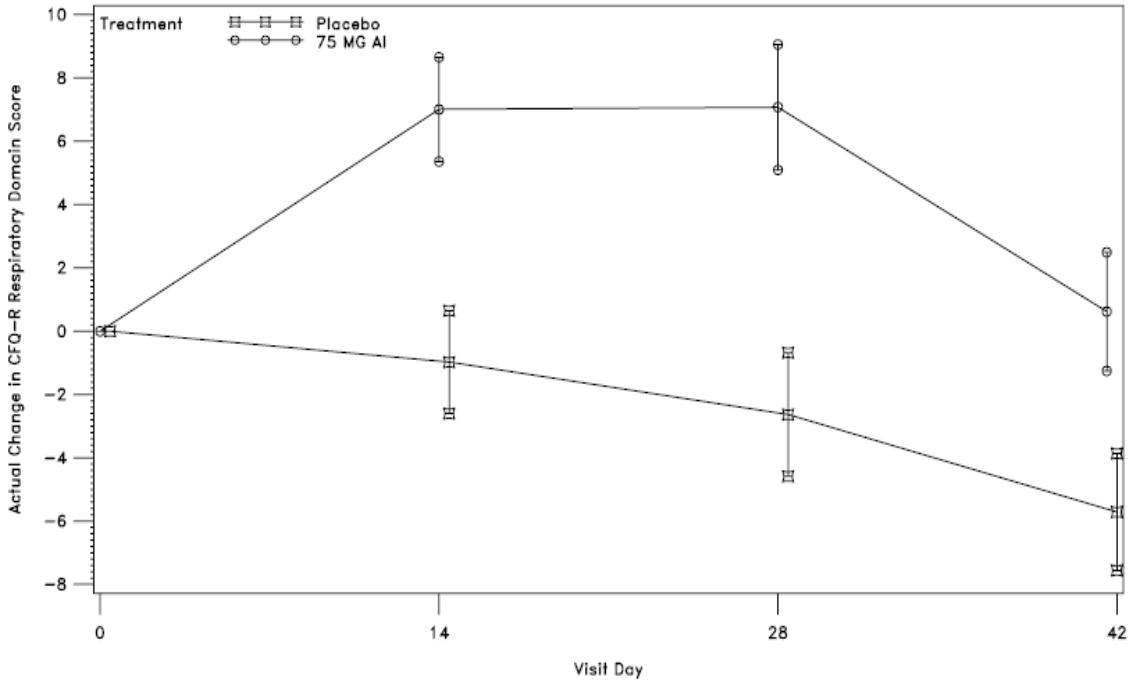
Primary Analysis	**ITT Patients**	
Treatment	**Placebo N = 84**	**75 mg AI N = 80**
Mean (± SD)	-5.09 ± 17.02	1.32 ± 18.31
n	83	80
Adjusted mean	-5.71	0.62
Treatment difference: AI – placebo	6.33	
95% CI (p-value)	1.22,11.43 (p-value = 0.0154)	

Source: FDA Table

Statistical Reviewer Comments: *The primary endpoint was also significant at Day 42 suggesting a sustained treatment benefit over placebo for an additional 14 days after Day 28. However, the treatment difference (AI – placebo) decreased from 9.71 to 6.33 points during these 14 days suggesting some possible loss of treatment efficacy over placebo. Note that there is a substantial absolute loss of efficacy ≈ 6.5 points in AI patients from Day 28 to Day 42 as evidenced by the (adjusted) change in CFQ-R from 7.08 (Day 28) to 0.62 Day (42). This is a larger absolute loss than observed in placebo arms of -2.63 – (-5.71) ≈ 3.1 points.*

The figure below graphically depicts the actual change from Visit 2 (Day 0) in CFQ-R respiratory domain score to Visits 3, 4 and 5 on imputed data (Child/Teen/Adult Combined) according to adjusted mean values +/- standard error.

Figure 9: Actual Change from Visit 2 (Day 0) in CFQ-R Respiratory Domain Score to Visits 3, 4 & 5 Using Imputed Data (Study 007)



Source: Sponsor Figure 14.7.2 from Study Report

Statistical Reviewer Comments: *Actual change from Visit 2 (Day 0) in CFQ-R respiratory domain scores increased substantially in the AI regimen from Day 0 to Day 14, remained fairly constant from Day 14 to Day 28 and decreased substantially from Day 28 to Day 42. The actual change in the placebo regimen decreased from Day 0 to Day 14, Day 14 to Day 28 and Day 28 to Day 42. These decreases tended to accelerate slightly in later time intervals.*

Table 13: Observed P-values Testing AI TID Subjects versus Placebo TID Subjects For Key Endpoints in the ITT Population (Study 007)

Study 007 Endpoints (AI TID vs. Placebo TID)	**Observed P-values**
Actual Change in CFQ-R Respiratory Domain Scores, Day 0 to Day 28	.0005
Actual Change in CFQ-R Respiratory Domain Scores, Day 0 to Day 42	.0154
Categorized Change in CFQ-R Respiratory Domain Scores, Day 0 to Day 28	.0055
Median Time to Need for IV/Inhaled Antibiotics Due to Pre-Defined Symptoms	.0949
Proportion of Patients Using IV/Inhaled Antibiotics, Day 0 to Day 42	.2364[a]
FEV_1% Predicted Actual Change from Day 0 to Day 28	.0001
FEV_1 (L) Percent Change from Day 0 to Day 28	.0001[a]
Change in sputum $\log_{10}$ *PA* CFUs, Day 0 to Day 28	.0001[b]
Mean (adj) change in Weight, Day 0 to Day 28	.0039
Mean (adj) change in BMI, Day 0 to Day 28	.0054
Number of school/work days missed, Day 0 to Day 28	.1860

Number of Hospitalizations, Day 0 to Day 42	.0487
Proportion of patients hospitalized, Day 0 to Day 42	.0640[c]
Proportion of patients with school/work missed, Day 0 to Day 28	.2201

[a] The Sponsor pre-specified control for testing these endpoints at the α=.025 level if the primary hypothesis was rejected.
[b] Sponsor pre-specified control for testing 'Change in sputum $\log_{10}$ *PA* CFUs, Day 0 to Day 28' at the α=.025 level if significance was found in 'FEV_1 (L) Percent Change from Day 0 to Day 28' was found at the α=.025 level.
[c] Sponsor pre-specified control for testing 'Proportion of patients hospitalized, Day 0 to Day 42' at the α=.025 level if significance in 'Proportion of Patients Using IV/Inhaled Antibiotics, Day 0 to Day 42' was found significant at the α=.025 level.

Source: FDA Table

Statistical Reviewer Comments: *Study 007 achieved its primary endpoint of 'Actual Change in CFQ-Respiratory Domain Scores from Day 0 at Day 28' (p=.0005). This result was supported by secondary analysis results which included an analysis of the CFQ-R respiratory domain at Day 42 (p=.0154), analysis of the categorized change in CFQ-R respiratory domain (p=.0055). Other secondary endpoints supporting Study 007 primary analysis findings included 'FEV$_1$ (L) Actual Change' (p=.0001), 'FEV$_1$% Predicted Actual Change' (p=.0001), 'Change in sputum log$_{10}$ PA CFUs' (p=.0001), 'Change in BMI' (p=.0054) and 'Change in Weight' (p=.0039).*

However, it should be noted that Study 007 patients in the AI TID regimen failed to achieve a significant finding in the key secondary endpoint of 'Time to Need for Inhaled or IV Antibiotics due to Pre-defined Symptoms' (p=.0949). Since this endpoint was considered as primary in Study 005, significant findings in both studies could have provided meaningful evidence regarding a treatment benefit for AI patients. This endpoint, however, was not found to be significant in either of the 005 and 007 studies.

It should also be noted that statistical inferences based on secondary analyses may be limited due to potential inflation of the overall type I error rate due to multiple testing. However, the Sponsor's secondary analysis did show significance in the 'FEV$_1$ (L) percent change from Day 0 to Day 28' and the 'Change in Sputum log$_{10}$ PA *CFUs, Day 0 to Day 28' which were both pre-specified for testing at the α=.025 level thus controlling against inflation of the type I error rate.*

3.2.5 Conclusions

Study 007 provided some evidence to suggest that 75mg aztreonam inhalation (TID) may provide a substantial treatment benefit for patients with cystic fibrosis due to *Psuedomonas aeruginosa* (*PA*). Study 007 met its primary endpoint of 'Actual Change in CFQ-R respiratory domain scores at Day 28' and provided some evidence towards a treatment benefit of AI TID. There were also several secondary endpoints which supported primary analysis findings. The strength of study 007 findings was however limited by the short follow-up period of 14 days following end of AI/placebo treatment in which statistical significance could not be demonstrated for a key secondary endpoint in Study 007 'Time to Need for IV/inhaled Antibiotics Due to Pre-defined Criteria.' (p=.0949). Primary analysis results were not considered robust since patients 18 years or

older and patients with greater disease severity only demonstrated a marginal benefit. It should also be noted that there are recent Agency concerns regarding the validity of the CFQ-R instrument due to the potential for recall bias. This may further limit the strength and interpretability of study findings.

3.3 Evaluation of Safety (Study 005)

The treatment-emergent adverse event summary for the AI/placebo period in the table below includes data from Day 0 to Day 84.

Table 14: Overall Summary of Treatment-emergent Adverse Events for the AI/Placebo Period (Study 005)

	Placebo	AI		
	Pooled (N = 76) n (%)	BID (N = 69) n (%)	TID (N = 66) n (%)	Pooled (N = 135) n (%)
Patients reporting at least one AE	67 (88.2)	62 (89.9)	61 (92.4)	123 (91.1)
Patients reporting at least one drug-related AE	23 (30.3)	18 (26.1)	20 (30.3)	38 (28.1)
Patients reporting at least one SAE	3 (3.9)	6 (8.7)	7 (10.6)	13 (9.6)
Patients reporting at least one severe AE	16 (21.1)	18 (26.1)	15 (22.7)	33 (24.4)
Patients who died	0	0	0	0
Number of AEs	377	358	368	726
Number of drug-related AEs	67	28	32	60
Number of SAEs	8	15	21	36

* Patient 40005 (Placebo TID) had AEs that started before the first dose of study drug and led to withdrawal of study drug; these AEs are not summarized in this table.
Source: Sponsor Table 39 of Study Report

Statistical Reviewer Comments: *The majority of patients in each treatment group (82% to 95%) reported at least one AE during the AI/placebo period. There were no differences between the pooled placebo and pooled AI groups in the percentages of patients reporting at least one AE, or at least one drug-related AE. A greater percentage of patients in the pooled AI group reported at least one SAE (13 [10%] patients) than in the pooled placebo group (3 [4%] patients). This difference did not reach statistical significance.*

Table 15: Treatment-emergent Adverse Events with Incidence Rate of 10% or Higher for the AI/Placebo Period: Safety Population (Study 005)

	Placebo	AI		
	Pooled (N = 76) n (%)	**BID (N = 69) n (%)**	**TID (N = 66) n (%)**	**Pooled (N = 135) n (%)**
Cough	52 (68.4)	45 (65.2)	41 (62.1)	86 (63.7)
Productive cough	29 (38.2)	25 (36.2)	30 (45.5)	55 (40.7)
Nasal congestion	11 (14.5)	16 (23.2)	12 (18.2)	28 (20.7)
Respiratory tract congestion	15 (19.7)	10 (14.5)	12 (18.2)	22 (16.3)
Wheezing	9 (11.8)	9 (13.0)	15 (22.7)	24 (17.8)
Haemoptysis	14 (18.4)	8 (11.6)	10 (15.2)	18 (13.3)
Fatigue	13 (17.1)	7 (10.1)	7 (10.6)	14 (10.4)
Pharyngolaryngeal pain	9 (11.8)	10 (14.5)	8 (12.1)	18 (13.3)

Dyspnoea	8 (10.5)	4 (5.8)	11 (16.7)	15 (11.1)
Headache	9 (11.8)	6 (8.7)	8 (12.1)	14 (10.4)
Crackles lung	10 (13.2)	7 (10.1)	5 (7.6)	12 (8.9)
Decreased appetite	11 (14.5)	9 (13.0)	2 (3.0)	11 (8.1)
Exercise tolerance decreased	11 (14.5)	6 (8.7)	5 (7.6)	11 (8.1)
Chest discomfort	6 (7.9)	9 (13.0)	6 (9.1)	15 (11.1)
Pulmonary function test decreased	8 (10.5)	4 (5.8)	9 (13.6)	13 (9.6)
Pyrexia	2 (2.6)	6 (8.7)	10 (15.2)	16 (11.9)
Rhinorrhoea	3 (3.9)	7 (10.1)	7 (10.6)	14 (10.4)
Sinus congestion	8 (10.5)	2 (2.9)	6 (9.1)	8 (5.9)

Source: Sponsor Table 40 of Study Report

Statistically Reviewer Comments: *Treatment-emergent adverse events were generally similar between treatment groups. However, for some safety variables, there were significant differences favoring placebo patients. 'Decreased appetite' (p=.0400) and 'pyrexia' (p=.0262) were both significantly lower in the pooled placebo regimen using Fisher Exact test for overall treatment effect across pooled placebo, AI BID and AI TID regimens. It should also be noted that statistical inferences based on these safety findings are limited since inflation of the type I error rate from multiple testing has not been controlled for. Please also refer to the safety review of Dr. Menfo Imoisili for further information.*

3.4 Evaluation of Safety (Study 007)

The treatment-emergent adverse event summary for the AI/placebo period in the table below includes data from Day 0 to Day 42.

Table 16: Overall Summary of Treatment-emergent Adverse Events: Safety (Study 007)

	Placebo (N = 84) n (%)	**75 mg AI (N = 80) n (%)**
Patients reporting at least one AE	69 (82.1)	63 (78.8)
Patients reporting at least one drug-related AE	19 (22.6)	28 (35.0)
Patients reporting at least one SAE	12 (14.3)	5 (6.3)
Patients reporting at least one severe AE	15 (17.9)	11 (13.8)
Patients with trial drug withdrawn as a result of an AE	17 (20.2)	7 (8.8)
Patients who died	0	0
Number of AEs	329	318
Number of drug-related AEs	38	59
Number of SAEs	41	21
Number of severe AEs	39	25

Source: Sponsor Table 35 of Study Report

Statistical Reviewer Comments: *Approximately 80% of patients reported at least one AE (79% AI and 82% placebo). A greater percentage of patients in the AI group reported at least one drug-related AE (28 [35%] patients) than in the placebo group (19 [23%]*

patients), and the difference approached statistical significance (p = 0.0869). There was no difference between the AI and placebo groups in the percentages of patients reporting at least one AE or at least SAE.

Table 17: Treatment-emergent Adverse Events with Incidence Rate of 5% or Higher (Study 007)

	Placebo (N=84) (n%)	**75 mg AI (N = 80) n (%)**
Cough	30 (35.7)	38 (47.5)
Productive cough	27 (32.1)	13 (16.3)
Nasal congestion	8 (9.5)	11 (13.8)
Crackles lung	8 (9.5)	10 (12.5)
Pharyngolaryngeal pain	8 (9.5)	10 (12.5)
Dyspnoea	10 (11.9)	7 (8.8)
Headache	10 (11.9)	6 (7.5)
Pyrexia	7 (8.3)	9 (11.3)
Wheezing	7 (8.3)	8 (10.0)
Dyspnoea exacerbated	8 (9.5)	5 (6.3)
Fatigue	8 (9.5)	5 (6.3)
Abdominal pain	6 (7.1)	6 (7.5)
Pulmonary function test decreased	7 (8.3)	4 (5.0)
Rhinorrhoea	7 (8.3)	3 (3.8)
Chest discomfort	4 (4.8)	5 (6.3)
Haemoptysis	6 (7.1)	3 (3.8)
Sputum discoloured	5 (6.0)	3 (3.8)
Sinus congestion	3 (3.6)	4 (5.0)
Throat irritation	2 (2.4)	5 (6.3)
Nausea	2 (2.4)	4 (5.0)

Source: Table 37 of Study Report

Statistical Reviewer Comments*: The most commonly reported treatment-emergent AE was cough with 38 (48%) and 30 (36%) of patients in the AI and placebo groups. There was a significant difference with respect to productive cough (p=.0192) favoring the AI patients. It should also be noted that statistical inferences based on these safety findings are limited since inflation of the type I error rate from multiple testing has not been controlled for.*

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Special/Subgroup Populations (Study 005)

Subgroup analyses of the primary endpoint in ITT subjects were conducted for the following variables: age (<18 yrs, ≥ 18 yrs), gender (male, female), disease severity at visit 2 (≤ 50% FEV_1% predicted, > 50% FEV_1% predicted) highest aztreonam MIC at day 0 (≤ 8 µg/mL, > 8 µg/mL), and highest tobramycin MIC at day -28 (< 8 µg/mL, ≥ 8 µg/mL).

Table 18: Time to Need for Inhaled or IV Antibiotics, ITT Subjects- Subgroup Analysis (Study 005)

Age						
Subgroups:	**< 18 yrs**			**≥ 18 yrs**		
Treatment	Placebo (Pooled)	AI (BID)	AI (TID)	Placebo (Pooled)	AI (BID)	AI (TID)
n	12	17	17	64	52	49
Number (%) of Events	5 (41.7)	4 (23.5)	5 (29.4)	33 (51.6)	15 (28.8)	19 (38.8)
p-value	--	NS	NS	--	.0066*	.2942
Gender						
Subgroups:	**Male**			**Female**		
Treatment	Placebo (Pooled)	AI (BID)	AI (TID)	Placebo (Pooled)	AI (BID)	AI (TID)
n	45	38	38	31	31	28
Number (%) of Events	22 (48.9)	10 (26.3)	14 (36.8)	16 (51.6)	9 (29.0)	10 (35.7)
p-value	--	.0043*	.1211	--	.1529	.7602
Disease Severity (at Visit 2)						
Subgroups:	**FEV_1% ≤ 50**			**FEV_1% > 50**		
Treatment	Placebo (Pooled)	AI (BID)	AI (TID)	Placebo (Pooled)	AI (BID)	AI (TID)
n	30	24	22	46	44	44
Number (%) of Events	18 (60.0)	9 (37.5)	10 (45.5)	20 (43.5)	10 (22.7)	14 (31.8)
p-value	--	.0795	.1464	--	.0220*	.5961
Highest Aztreonam MIC at Day 0						
	≤ 8 µg/mL			**> 8 µg/mL**		
Treatment	Placebo (Pooled)	AI (BID)	AI (TID)	Placebo (Pooled)	AI (BID)	AI (TID)
n	44	43	42	27	16	21
Number (%) of Events	19(43.2)	11 (25.6)	16(38.1)	17(63.0)	6(37.5)	8(38.1)
p-value	--	.0295*	.4305	--	.1770	.5768
Highest Tobramycin MIC at Day -28						
	< 8 µg/mL			**≥ 8 µg/mL**		
Treatment:	Placebo (Pooled)	AI (BID)	AI (TID)	Placebo (Pooled)	AI (BID)	AI (TID)
n	36	45	33	36	13	24
Number (%) of Events	18 (50)	14 (31.1)	14(42.4)	18(50)	3(23.1)	10(41.7)
p-value	--	0.0238*	0.4132	--	0.1138	0.6942

* In this subgroup analysis of the primary endpoint of Study 005, significant findings against pooled placebo may be unclear due to failure to control for a significant placebo regimen effect. There are also limitations involved with post-hoc subgroup testing including lack of randomization protection and lack of control against inflation of the overall type I error rate.
Source: FDA Table

Statistical Reviewer Comments: *Significant findings at the α=.05 level (two-sided) were not observed for comparisons of AI TID regimen against Pooled Placebo for any of the included subgroup categories. Significant findings were also not observed in the AI BID regimen for several subgroup categories. Therefore, results fail to provide supportive evidence of either the AI BID or AI TID regimen as a more effective therapy than placebo in a general study population. Note that although significant findings for the BID regimen were observed in specific subgroups, this evidence is unclear due to lack of randomization protection and the post-hoc nature of statistical testing which fails to control against inflation of the overall type I error rate from multiple testing. Additionally, comparisons in the above table 'Pooled Placebo' may be unclear. As referred to earlier, pooling placebo regimens is not justified when analyzing the primary endpoint of Study 005 due to evidence of a regimen effect in the placebo BID and placebo TID arms (p=.0043).*

4.2 Special/Subgroup Populations (Study 007)

Subgroup analyses were conducted on the primary endpoint of Study 007, "Change from Visit 2 (Day 0) to Visit 4 (Day 28) in CFQ-R Respiratory Domain of ITT Subjects". Subgroup analyses included the following variables and categories: age (<18 yrs, ≥ 18 yrs), gender (male, female), disease severity at visit 2 (≤ 50% FEV_1% predicted, >50% FEV_1% predicted) and Geographic Region (US & Canada, Australia & New Zealand).

Table 19: Change from Visit 2 (Day 0) to Visit 4 (Day 28) in CFQ-R Respiratory Domain Scores on Imputed Data, ITT Subjects- Subgroup Analysis (Study 007):

Age				
Subgroups:	**< 18 years**		**≥ 18 years**	
Treatment	**Placebo**	**75 mg AI**	**Placebo**	**75 mg AI**
n	16	21	67	59
Mean (± SD)	-6.42 (± 17.29)	16.01(± 17.50)	-0.83 (± 18.91)	5.56 (± 18.65)
Treatment difference*: AI – placebo	18.91		6.35	
95% CI (p-value)	8.78, 29.05 (p=.0006)		0.02,12.69 (p = 0.0495)	
Gender				
Subgroups:	**Male**		**Female**	
Treatment	**Placebo**	**75 mg AI**	**Placebo**	**75 mg AI**
n	44	48	39	32
Mean (± SD)	-3.22 (± 14.99)	5.96 (± 16.23)	-0.43 (± 22.16)	10.76 (± 22.25)
Treatment difference*: AI – placebo (p-value)	11.3 (p=.0003)		8.5 (p=.0035)	

Disease Severity				
Subgroups:	**FEV_1% predicted ≤ 50%**		**FEV_1% predicted > 50%**	
Treatment	**Placebo**	**75 mg AI**	**Placebo**	**75 mg AI**
n	30	30	53	50
Mean (± SD)	-4.72 (± 15.63)	4.91 (± 21.01)	-0.31 (± 20.11)	9.67 (± 17.46)
Treatment difference*: AI – placebo	8.25		10.90	
95% CI (p-value)	-1.14,17.64 (p=0.0839)		4.16,17.64 (p = 0.0018)	
Region				
Subgroups:	**US & Canada**		**Australia & New Zealand**	
Treatment	**Placebo**	**75 mg AI**	**Placebo**	**75 mg AI**
n	62	62	21	18
Mean (± SD)	-0.58 (± 18.79)	6.41 (± 18.45)	-5.82 (± 18.06)	12.96 (± 20.01)
Treatment difference*: AI – placebo	7.36		17.29	
95% CI (p-value)	1.07,13.65 (p=.0223)		6.00,28.58 (p=.0037)	

* Treatment difference is based on adjusted means for relevant covariates including CFQ-R baseline score and disease severity

Source: FDA Table

Statistical Reviewer Comments: *Based on subgroup analyses above, improvement in CFQ-R respiratory domain scores was influenced by 'disease severity' and 'age' variables. For 'age', primary analysis results were significant in patients <18 years of age (p=.0006) and marginally significant in patients ≥ 18 yrs (p=.0495). The treatment difference (AI-Placebo) was substantially larger in patients less than 18 years of age versus patients 18 years and older (18.91 vs. 6.35). For 'disease severity', subgroup analyses also showed differences in primary analysis results. The treatment benefit of AI TID therapy over placebo was smaller in patients with FEV_1% predicted less than 50% (p=.0839) than in patients with FEV_1% predicted greater than 50% (p=.0018).*

These subgroup analyses raise some concerns regarding the robustness of the AI TID treatment effect in older patients or patients with greater severity (e.g. FEV_1% predicted less than 50%) since primary analysis results were only marginally or non-significant in these populations. It should also be noted that statistical inferences based on post-hoc subgroup analyses may be severely limited due to lack of randomization protection and lack of control against inflation of the type I error rate from multiple testing.

5. SUMMARY AND CONCLUSIONS

Overall evidence presented in NDA 50814 failed to demonstrate a substantial treatment benefit for patients with cystic fibrosis due to *Psuedomonas aeruginosa* (*PA*) using 75mg aztreonam inhalation three times daily (AI TID). Study 005 failed to provide meaningful evidence regarding a treatment benefit from AI TID therapy and Study 007 provided only marginal evidence. Demonstration of treatment efficacy requires substantial evidence essentially replicated from two or more adequate and well controlled trials.

In Study 005, ITT patients in the AI TID treatment arm failed to achieve significant improvement compared to Placebo for several key endpoints. This included the primary endpoint, 'Time to Need for Inhaled or IV Antibiotics,' a key secondary endpoint 'Actual Change in CFQ-R Respiratory Domain Scores at Day 28,' and several other secondary endpoints. In Study 005, primary analyses were also unclear due to a "regimen effect" (i.e. BID vs. TID dosing) which was observed to be stronger than the treatment effect (i.e. AI vs. Placebo) such that patients on a placebo BID regimen had actually fared better than patients on the AI TID regimen. Due to the strength of this "regimen effect" in the Placebo BID and Placebo TID arms, FDA primary analysis comparisons of AI TID vs. Placebo BID and AI TID vs. Placebo TID were entirely inconsistent. Similarly, comparisons of AI BID vs. Placebo BID and AI BID vs. Placebo TID were also inconsistent (Table 6).

In Study 007, although the primary endpoint of 'Actual Change in Cystic Fibrosis Questionnaire- Revised (CFQ-R) respiratory domain scores at Day 28' was met and supported by findings from several secondary endpoints, there were still limitations with the evidence presented. First, there are recent Agency concerns regarding the validity of the CFQ-R instrument due to the potential for recall bias. This may limit the strength and interpretability of study findings. Second, primary analysis results were not considered robust since patients 18 years or older and patients with greater disease severity demonstrated only a marginal benefit from AI TID therapy. Finally, Study 007 patients in the AI TID regimen failed to achieve a significant finding in the key secondary endpoint of 'Time to Need for Inhaled or IV Antibiotics due to Pre-defined Symptoms." Since this endpoint was considered as primary in Study 005, significant findings in both studies could have provided meaningful evidence regarding a treatment benefit for AI patients. This endpoint, however, was not found to be significant in either of the 005 and 007 studies.

In summary, due to the lack of substantial evidence presented in this submission, we recommend that an additional adequate and well controlled Phase III study be conducted to demonstrate; (1) a reduction in the 'Time to Need for IV or Inhaled Antibiotics due to Pre-defined Symptoms' and (2) 'Actual Change in CFQ-R Respiratory Domain Scores at Day 28' as co-primary endpoints. A sequential testing procedure or other methodologies may be considered to control for the overall type-I error rate due to multiple testing. It is also recommended that the additional study utilizes an updated version of the CFQ-R (respiratory domain) questionnaire that is appropriately revised and validated to address recent Agency concerns with the instrument.

As a note, the Sponsor may consider whether treatment regimens involving a higher dose (e.g. AI 150mg) and less frequent dosing regimen (e.g. BID) would provide a more optimal dosing regimen. Possible study options could include a two arm study with AI 75mg TID vs. Placebo TID; or a four-arm trial comparing the AI 75mg TID, AI 150mg BID and the corresponding placebo treatment regimens. Including a higher dose of AI may provide an alternative to patients if AI 75mg is not found to be the optimal dose.

SIGNATURES/DISTRIBUTION LIST

Primary Statistical Reviewer: Christopher Kadoorie, Ph.D.

Date: 7/22/2008

Concurring Reviewer: Thamban Valappil, Ph.D, Statistical Team Leader

cc:
HFD-520/Project Manager/ Kyong Hyon, B.S.N, M.A
HFD-520/Medical Officer/Menfo Imoisili, M.D. M.P.H
HFD-520/Medical Team Leader/John Alexander, M.D. M.P.H.
HFD-520/Deputy Division Dir/ Katherine Laessig, M.D.
HFD-520/Acting Division Dir/ Wiley Chambers, M.D.
HFD-725/Primary Statistical Reviewer/ Christopher Kadoorie, Ph.D.
HFD-725/Statistical Team Leader/ Thamban Valappil, Ph.D.
HFD-725/Biometrics Deputy Division Director/ Daphne Lin, Ph.D.
HFD-725/Biometrics Division Director/ Mohammad Huque, Ph.D.
HFD-700/Office of Biostatistics Deputy Director/ Edward Nevius, Ph.D.
HFD-700/Office of Biostatistics Associate Director/ Ram Tiwari, Ph.D.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

Christopher Kadoorie
7/22/2008 06:03:14 PM
BIOMETRICS

Thamban Valappil
7/22/2008 06:07:16 PM
BIOMETRICS