18F-AV-1451-PX01 SAP v1.0

Evaluation of the relationship between baseline flortaucipir PET signal and cognitive change in subjects with early Alzheimer's disease participating in the I8D-MC-AZES Protocol Addendum D5010C00009 (2.1) (Tau Imaging)

NCT03901105

Approval date: 16 Apr 2019

Protocol No. 18F-AV-1451-PX01

Protocol Addendum D5010C00009 (2.1) (Tau Imaging) A 24-Month, Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Efficacy, Safety, Tolerability, Biomarker, and Pharmacokinetic Study of AZD3293 in Early Alzheimer's Disease (The AZES Study)

Statistical Analysis Plan

Prepared by:
Avid Radiopharmaceuticals, Inc

Final Version 1.0
16 April 2019

Approval Date: 24-Apr-2019 GMT

Table of Contents

1 LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Table 1: Abbreviations and Definitions of Terms

Aβ	amyloid-β
AD	Alzheimer's disease
ADAS-Cog	Alzheimer's Disease Assessment Scale-Cognitive
CDR	Clinical Dementia Rating Scale
CDR-SB	Clinical Dementia Rating Scale-Sum of Boxes
CFB	Change from Baseline
CSR	clinical study report
FAQ	Pfeffer Functional Activities Questionnaire
LSM	least squares mean
MMRM	mixed model repeated measures
MMSE	Mini-Mental State Examination
PET	positron emission tomography
SAP	statistical analysis plan

2 INTRODUCTION

Brain amyloid plaques, composed of amyloid-beta (Aβ) aggregates, and neurofibrillary tangles, composed of phosphorylated aggregated tau deposits, are defining pathological features of Alzheimer's disease (AD). Thus, an amyloid negative florbetapir positron emission tomography (PET) scan of the brain, which indicates the absence of moderate or frequent Aβ neuritic plaques, is inconsistent with a diagnosis of AD. However, because amyloid is believed to accumulate very early in the AD process (Jack et al. 2010) and may be present in clinically normal elderly subjects (Price and Morris 1999; Sperling et al. 2011), the density or distribution of amyloid in subjects with a positive scan is not thought to reflect the severity of AD.

In contrast to Aβ neuritic plaques, the density and distribution in the brain of phosphorylated tau, aggregated in neurofibrillary tangles, increases with AD-related cognitive impairment and correlates with neurodegeneration (Duyckaerts et al. 1987). Thus, a PET imaging agent that binds to phosphorylated tau has the potential to indicate disease severity and with longitudinal measures could provide important information on disease progression. The PET tracer (18F-AV-1451) possesses a high affinity to human phosphorylated tau deposits. Up to three scans will be performed over the two year study.

3 STUDY OBJECTIVES

The primary objective of this study is:

- To assess whether a visual interpretation (τAD++ vs. non-τAD++ pattern) of the baseline flortaucipir PET scan can predict the risk of subjects' clinically meaningful cognitive and functional deterioration within 18 months of scan, as measured by the Clinical Dementia Rating Scale (CDR) Sum of Boxes (CDR-SB) change from baseline (CFB).

The secondary objectives of this study are:

- To assess whether a visual interpretation (τAD++ vs. non-τAD++ pattern) of the baseline flortaucipir PET scan can predict the risk of subjects' clinically meaningful cognitive and functional deterioration within 18 months of scan, as measured by Mini-Mental State Examination (MMSE), Alzheimer's Disease Assessment Scale–Cognitive subscale (ADAS-Cog11), Pfeffer Functional Activities Questionnaire (FAQ), and CDR Global CFB.

- To assess the relationship between visual interpretation (τAD++ vs. non-τAD++ pattern) of the baseline flortaucipir PET scan and magnitude of cognitive and functional deterioration within 18 months of scan, as measured by the mean change from baseline of CDR-SB, MMSE, ADAS-Cog11, and FAQ.

- To assess inter-reader reliability of the flortaucipir visual interpretation by 5 independent, blinded readers.

The exploratory objectives of this study are:

- To assess whether a visual interpretation (τAD++ vs. non-τAD++ pattern) of the baseline flortaucipir PET scan can predict the risk of subjects' clinically meaningful cognitive and functional deterioration within 24 months of scan, as measured by the CDR, MMSE, ADAS-Cog11 and FAQ CFB.

- To assess the diagnostic performance of baseline flortaucipir in predicting clinically meaningful deterioration at 18 months.

4 STUDY DESIGN

4.1 General Design

The D5010C00009 (AZES) study was a multicenter, randomized, double blind placebo-controlled, parallel-group, longitudinal study evaluating the efficacy of LY3314814 (lanabecestat) in early AD subjects, whose primary efficacy measure was ADAS-Cog13. AZES was terminated for futility, however over 400 subjects received a bolus injection of flortaucipir F18 and quantitative imaging as part of a tau addendum.

Because the AZES study was terminated early for futility, not all subjects had the opportunity to complete the full term follow up. To avoid bias for these analyses, it was decided to include only the subjects that had: 1) a valid baseline flortaucipir scan (no later than 91 days post randomization, considering that the flortaucipir scans were added to this study after the initiation of AZES); and 2) a CDR assessment at 18 months visit. 205 subjects met these criteria, and approximately 90 of these subjects completed a CDR assessment at the 24 month visit.

Five independent radiologists or nuclear medicine physicians visually interpreted the PET scans from the 205 qualified subjects as either τAD++ (a pattern indicating spread of aggregated tau beyond the posterolateral temporal [PLT] or occipital lobe), τAD + (a pattern indicating aggregated tau confined to posterolateral temporal/occipital lobe) or τAD- (inconsistent with an AD pattern). The primary hypothesis tested by this study is that the risk of clinically meaningful cognitive deterioration would vary as a function of flortaucipir PET scan status at baseline. Specifically, that risk would be significantly greater for subjects in the τAD++ group as compared to those in the non-τAD++ group (including τAD- and τAD+).

4.2 Discussion of Study Design

18F-AV-1451-PX01 (PX01) was designed to take advantage of the therapeutic trial cohort in AZES that had flortaucipir PET scans and up to 2 years cognitive follow-up to

further evaluate the relationship between flortaucipir PET signal and cognitive decline. Patients who had a florbetapir PET scan at screening in the main AZES study to document the presence of amyloid for study inclusion and participated in the longitudinal amyloid PET sub-study also received a flortaucipir PET scan performed at baseline, 12 months, and 24 months at participating sites. Patients who established eligibility by historical amyloid scan were not eligible to participate in the flortaucipir (tau) addendum unless they also had an optional subsequent florbetapir scan as part of the main AZES study. Flortaucipir PET scans were conducted in over 400 study participants. Once the randomization targets for Addendum 2 were reached, enrollment was considered complete and, at the discretion of the sponsor, enrollment of additional patients was stopped. Flortaucipir PET scanning was conducted under the management of a central PET vendor. The PET imaging data was acquired after injection of flortaucipir F 18. The baseline flortaucipir scan was performed only after amyloid positivity had been demonstrated. There was at least 16 hours between the follow-up florbetapir scan and the flortaucipir scan at week 104 or early discontinuation due to the half-life of 18-fluorine.

Procedures relevant to this addendum are provided in section 6.1.

PX01 was designed to evaluate the performance of flortaucipir scans in subjects with early AD to establish a relationship between baseline flortaucipir imaging results and longitudinal cognitive assessments. Longitudinal cognitive assessments, with CDR-SB being the primary assessment, will provide evidence whether greater degrees of change in cognitive/functional impairment correlate with higher levels of flortaucipir uptake at baseline. This relationship is important in order to determine the utility and predictive power of flortaucipir on the degree of change in cognitive impairment in 18 to 24 months. Since the parent study AZES was terminated prior to reaching the endpoint, not all study participants could complete study measurements as planned. Therefore, only 205 subjects with a valid baseline tau scan and a CDR assessment at week 78 (month 18) will be included in this study for efficacy analyses.

4.3 Method of Assignment of Subjects to Treatment Groups

4.4 Blinding

A blinded design was not used in PX01 for flortaucipir exposure because all subjects received flortaucipir F 18. Subjects and investigators were blind to lanabecestat treatment throughout the trial.

4.5 Determination of Sample Size

This study will include up to 205 subjects who have a valid baseline flortaucipir scan, and a change from baseline value of CDR-SB at 18 months. In Study TZAX, which had a similar design and entry criteria, approximately 75% of subjects were rated as τAD++ and 25% were rated as non-τAD++. Among non-τAD++ subjects (reference group), approximately 40% had an increase of 1 point or more on CDR-SB at 18 months follow up visit. Assuming a distribution of visual reads and CDR-SB change in this study is

similar to that for TZAX, a sample of 205 subjects will provide 90% power to detect a risk ratio of 1.65 or larger, under a two-sided type I error rate of 5%.

5 CHANGES IN THE CONDUCT OF THE STUDY OR PLANNED ANALYSES

5.1 Changes in the Conduct of the Study

There were no changes in the conduct of the study at the time of preparing this SAP.

6 BASELINE, EFFICACY AND SAFETY EVALUATIONS

6.1 Schedule of Evaluations

Study Procedure	Screening	Randomization/ Baseline									
Visit Number	V1	V2-V4	V6	V8	V10	V12	V14	V16	V20	V30	ED
Study Month (m)			3m	6m	9m	12m	15m	18m	24m	156w	
Tau Addendum											
Interval for Visit (d)					±7	±10		±7	±10	±7	
Flortaucipir PET Imaging		X			X	X		X	X	X	
Efficacy Measures		X									
CDR-SB	X			X		X		X	X		X
MMSE	X		X	X		X		X	X		X
ADAS			X	X		X	X	X	X		X
FAQ		X		X		X		X	X		X

6.2 Time Point Algorithms

6.2.1 Windows

For all analyses, results will be summarized at the planned study visit they were obtained. Data was collected in weeks in the AZES study, therefore the following conversion will be used to analyze in months:

Weeks	Months
13	3
26	6
52	12
65	15
78	18
104	24

6.3 Screening and Baseline Assessments

Baseline cognitive and functional assessments will include ADAS-Cog11, MMSE, CDR-SB, and FAQ. Baseline will be defined in the same manner as in the AZES study.

6.4 Efficacy Variables

6.4.1 Primary Efficacy Variable(s)

6.4.1.1 Tau Status Based on Flortaucipir

Flortaucipir images were visually assessed by five independent readers blinded to all demographic and clinical data information of the subjects. The majority read (i.e., the tau status for which 3 or more readers have agreed upon) was used as the tau status.

For the visual reads, the images were classified as either τAD++ (a pattern indicating spread of aggregated tau beyond the PLT lobe or occipital lobe), τAD + (a pattern indicating aggregated tau confined to PLT and/or occipital lobe) or τAD- (inconsistent with an AD pattern). See table below for details.

Read Outcome		Objective Imaging Features
Not Consistent with AD pattern (τAD-)		No increased neocortical activity, or increased neocortical activity isolated to mesial temporal, anterolateral temporal, and/or frontal regions.
AD pattern	τAD +	In either hemisphere, increased neocortical activity in the posterolateral temporal (PLT) or occipital region(s).
	τAD ++	In either hemisphere, increased neocortical activity in the parietal/precuneus region(s) or frontal region(s) with increased

		uptake in the PLT, parietal, or occipital region(s).

For the purposes of these efficacy analyses, image reads will be contrasted in the following ways: τAD++ vs. non-τAD++ (τAD+ and τAD-); τAD (τAD++ and τAD+) vs. τAD-; τAD++ vs. τAD-.

6.4.1.2 Clinical Dementia Rating (CDR) Scale (Berg, 1988)

The CDR examines 6 categories of cognitive functioning domains. Each domain is scored on a scale ranging from 0 to 3 (including 0.5). A CDR-SB will be generated as the sum of the values in each of the six domains. The CDR-SB sum scores ranges from 0-18, where higher scores indicate greater cognitive impairment. CDR was scheduled to be performed at screening, 6 months, 12 months, 18 months, and 24 months post-baseline. CDR-SB will be used as a continuous scale in exploratory objective analysis, when assessing the average CFB by tau status.

For the primary analysis, the CDR-SB score CFB will be dichotomized in two groups: CDR-SB score change 1 point or more at 18 months, or otherwise.

The CDR Global score (a global rating of dementia) was derived based on CDR item/domain values. CDR global will be dichotomized in two groups for exploratory analysis: any change in CDR global or otherwise.

Any CDR measurement collected outside of this intended visit schedule (e.g. Early Discontinuation) will not be included in the efficacy analysis.

6.4.2 Additional Efficacy Variables

6.4.2.1 Cognitive Assessments

These additional cognitive assessments from AZES will be analyzed:

Mini-Mental State Examination (MMSE)

The MMSE (Folstein et al., 1975) is a brief instrument used to assess cognitive function in elderly patients. The range for the total MMSE score is 0 to 30, the sum of each correct answer, with higher scores indicating better cognition. MMSE was scheduled to be collected at screening, 3 months, 6 months, 12 months, 18 months, and 24 months post-baseline. The score on a continuous scale will be used for secondary and exploratory analyses when assessing the average CFB by tau status. Any measurement collected outside of this intended visit schedule (e.g. Early Discontinuation) will not be included in the efficacy analysis.

Alzheimer's Disease Assessment Scale—Cognitive subscale

The ADAS (ADAS-Cog; Rosen et al. 1984) was designed to assess the severity of the dysfunction in the cognitive and non-cognitive behaviors characteristic of persons with AD. Although AZES reported ADAS-Cog13 as its primary endpoint, ADAS-Cog11 (a subset version of ADAS-Cog13 comprising of 11 questions) will be chosen for the PX01 analyses to remain consistent with other historical Avid studies. The cognitive subscale ADAS-Cog11, consists of 11 items assessing areas of function most typically impaired in AD: orientation, verbal memory, language, and praxis. The overall score for ADAS-Cog11 ranges from 0 to 70, with higher scores indicating greater disease severity, and is calculated as the sum of all 11 individual component scores. ADAS total score will be used as a continuous variable for the secondary and exploratory analyses to assess the average CFB by tau status. ADAS was scheduled to be collected at randomization (baseline), 3 months, 6 months, 12 months, 15 months, 18 months, and 24 months post-baseline. Any measurement collected outside of this intended visit schedule (e.g. Early Discontinuation) will not be included in the efficacy analysis.

Pfeffer Functional Activities Questionnaire (FAQ), (Pfeffer et al. 1982)

Functional status is conceptualized as the "ability to perform self-care, self-maintenance and physical activities." The FAQ was developed to assess instrumental activities of daily living involving higher level functional skills such as shopping alone, writing checks, remembering appointments, etc. The FAQ asks the informant to rate the patient's ability using the following scoring system: Dependent = 3; Requires assistance = 2; Has difficulty but does by self = 1; Normal = 0; Never did [the activity] but could do now = 0; Never did and would have difficulty now = 1. The sum scores ranges from 0-30, where higher scores indicate greater functional impairment. FAQ sum score will be used as a continuous variable in secondary objective analysis, when assessing the average CFB by tau status. FAQ was scheduled to be collected at randomization (baseline), 6 months, 12 months, 18 months, and 24 months post-baseline. Any measurement collected outside of this intended visit schedule (e.g. Early Discontinuation) will not be included in the efficacy analysis.

6.4.2.2 Clinically Meaningful Deterioration

The cognitive/functional assessments from 6.4.2.1 will be dichotomized to allow for the assessment of risk of clinically significant deterioration. Subjects will be categorized to these groups by the criteria accordingly.

CDR-SB: clinically significant deterioration is defined as a 1 point or more increase from baseline.

MMSE: clinically significant deterioration is defined as 3 points or more decrease from baseline.

FAQ: clinically significant deterioration is defined as 3 points or more increase from baseline.

ADAS-Cog11: clinically significant deterioration is defined as 4 points or more increase from baseline.

CDR Global: CDR global CFB will also be categorized as any increase (change >0) versus no change or decrease (change <= 0).

7 STATISTICAL METHODS

7.1 Definitions and Conventions

All analysis will be performed using SAS version 9.2 or higher.

Data will be summarized using descriptive statistics (number of subjects [n], mean, standard deviation [SD], median, minimum [min], and maximum [max]) for continuous variables and frequency counts and percentages for discrete variables. Percentages will be rounded to one decimal place, except 100% will be displayed without any decimal places and percentages will not be displayed for zero counts.

The tables and listings will be numbered using a decimal system to reflect main levels of unique tables and listings and sub-levels of replicate tables and the listings with maximum two digits per level (e.g., Table XX.YY.ZZ…). Tables will be numbered as 14.YY.ZZ. Baseline analysis will be reported in table series 14.1, and efficacy analysis in series 14.2. Listings will be numbered as 16.YY.ZZ.

Unless otherwise specified, hypothesis testing will be two-sided with type I error rate of 0.05.

7.2 Adjustments for Covariates

Multivariate models using cognitive or function assessments as the dependent variables will be adjusted for treatment arm [lanabecestat (20mg, 50 mg) or placebo], baseline cognitive score, years of education (categorical), and age.

7.3 Handling of Dropouts or Missing Data

Dropout subjects will not be replaced in this study. For situations with no rules for handling missing data the default will be no imputation.

Likelihood-based mixed effects models for repeated measures will be used to handle missing data for the cognitive/functional assessment mean CFB analyses. The model parameters are simultaneously estimated using restricted likelihood estimation incorporating all of the observed data. Estimates have been shown to be unbiased when the missing data are missing at random and when there is ignorable non-random missing data.

Repeated measures analyses will only use data from visits where the data was scheduled to be collected. When subjects discontinue from the study early, there may be efficacy data measurements at visits where the variables were not scheduled to be collected. These data will appear in listings only.

7.4 Interim Analyses and Data Monitoring

No interim analysis or data monitoring is planned for this study.

7.5 Multi-center Studies and Pooling of Centers

The data from all centers will be pooled. The pooled data will be analyzed and presented.

7.6 Multiple Comparisons/Multiplicity

No multiple comparisons/multiplicity adjustment is planned.

7.7 Examination of Subgroups

Subgroups of interest will be age, sex, and CDR Global group (see section 8.4.4).

8 STATISTICAL ANALYSIS

8.1 Analysis Population

Five readers independently interpreted the 205 flortaucipir PET scans collected from the AZES PET sub-study. No new subjects were enrolled for the purpose of this study.

Valid images were considered unevaluable in the present study only if 3 out of 5 readers independently declare the image unevaluable for the same reason. Subjects with invalid or unevaluable PET data were excluded from analyses. Criteria for declaring an image invalid or not evaluable is specified in the Image Review Charter. The majority read on each flortaucipir scan (baseline) from these 5 readers was used to determine the tau status of the corresponding study subject.

Subject disposition and baseline characteristics will be summarized using the entire analysis population (regardless if the baseline scan was evaluable or not). Analyses outlined in section 8.4 will be summarized using subjects in the analysis population who have a valid visual tau interpretation as described in section 6.4.1.1.

8.2 Summary of the Analysis Population

The analysis population will be summarized in the disposition table by total available scans, scans read, and breakdown of majority read (detailed in section 6.4.1.1).

8.3 Demographic and Other Baseline Characteristics

Demographics and baseline summaries will be reported on the analysis population. Age (years), sex, education (categorical: 1-5, 6-9, 10-12, 13+ years), race, ethnicity, treatment (lanabecestat – -, 2-0 mg or 50 mg –or placebo), florbetapir SUVr (referenced to the cerebellum), ApoE4 (E4+/E4-), ADAS-Cog11, CDR-SB, CDR Global (categorical: 0, 0.5, 1, 2, 3), FAQ, and MMSE will be reported in a table by enrolling diagnosis in the AZES study (AD or MCI).

8.4 Analysis of Efficacy Parameters

8.4.1 Analysis of Primary Efficacy Variable

8.4.1.1 Risk Ratio of τAD++ vs. Non-τAD++ Subjects in Progression to Clinically Meaningful Deterioration Evaluated by CDR-SB at 18 Months

This analysis will test the hypothesis that the risk of progressing to a clinically meaningful event (≥1 change) as determined by CDR-SB value change at 18 months will be significantly greater for subjects in the τAD++ group as compared to those in the non-τAD++ group (τAD- and τAD+).

Since the study goal is to evaluate the risk ratio of τAD++ group vs non-τAD++ group at month 18 instead of marginal risk ratio by tau status, only month 18 measurements will be included for this analysis.

As described in section 6.4.1.2, the primary efficacy variable will be the dichotomized CDR-SB score CFB (1 point or more increase vs. otherwise). Incidence of this clinically meaningful event by tau visual read groups will be compared using a log-linear model adjusted for treatment arm (lanabecestat – 20 mg or 50 mg, or placebo), baseline age, years of education (categorical), and baseline CDR-SB score. The Poisson distribution will be chosen to describe the distribution of the dependent variable and a log link function will be used to model the risk ratio. To improve the model's stability and reliability, a modified Poisson regression [Zou, 2004] model will be used using a robust error variance estimation, although there is only one observation per subject.

The risk ratio of τAD++ rated subjects progressing to the event over non-τAD++ rated subjects along with the 95% confidence interval will be provided.

8.4.2 Analysis of Secondary Efficacy Variable

8.4.2.1 Risk Ratio of τAD++ vs. Non-τAD++ Subjects in Progression to Clinically Meaningful Deterioration Evaluated by MMSE at 18 Months

The same analysis as described in section 8.4.1.1 will be performed for MMSE, however a CFB in MMSE <=-3 will be considered clinically meaningful.

8.4.2.2 Risk Ratio of τAD++ vs. Non-τAD++ Subjects in Progression to Clinically Meaningful Deterioration Evaluated by ADAS-Cog11 at 18 Months

The same analysis as described in section 8.4.1.1 will be performed for ADAS-Cog11, however a CFB in ADAS >=4 will be considered clinically meaningful.

8.4.2.3 Risk Ratio of τAD++ vs. Non-τAD++ Subjects in Progression to Clinically Meaningful Deterioration Evaluated by FAQ at 18 Months

The same analysis as described in section 8.4.1.1 will be performed for FAQ, however a CFB in FAQ >=3 will be considered clinically meaningful.

8.4.2.4 Risk Ratio of τAD++ vs. Non-τAD++ Subjects in Progression to Clinically Meaningful Deterioration Evaluated by CDR Global at 18 Months

The same analysis as described in section 8.4.1.1 will be performed for CDR Global, however a CFB in CDR Global >0 will be considered clinically meaningful

8.4.2.5 Mixed Model Repeated Measures for Comparison of τAD++ vs. Non-τAD++ CDR-SB Change at 18 Months

The objective of this analysis is to test the hypothesis that baseline tau status as determined by the flortaucipir scan will predict subjects' cognitive deterioration.

This analysis will test the hypothesis that the CDR-SB CFB at 18 months will be significantly larger (worse) for the τAD++ group compared to the non-τAD++ group (τAD- and τAD+). The null hypothesis is that the least squares mean (LSM) difference between τAD++ and non-τAD++ groups at the month 18 visit equals zero.

This hypothesis will be tested using a mixed model repeated measures (MMRM). MMSE CFB will be the dependent variable. The model will include tau status, visit, and the tau status-visit interaction as fixed effects, and be adjusted for treatment arm (lanabecestat – 20 mg or 50 mg, or placebo), baseline CDR-SB score, age, and years of education (categorical) as covariates.

An unstructured covariance matrix will be used to model the within-subject variance-covariance. If the unstructured covariance structure matrix results in a lack of model convergence, the following tests will be used in sequence: heterogeneous Toeplitz covariance structure, heterogeneous autoregressive covariance structure, and compound symmetry covariance structure. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom.

Descriptive statistics for the CDR-SB at baseline and all scheduled follow-up visits will be summarized by tau read status. The CFB to all follow-ups will be summarized in the same corresponding table.

CDR-SB LSM change will be graphed across time by τAD++ and non-τAD++ groups.

8.4.2.6 MMRM for Comparison of τAD++ vs. Non-τAD++ MMSE Change at 18 Months

The same analyses will be performed as described in Section 8.4.2.5 with the substitution of MMSE for CDR-SB.

8.4.2.7 MMRM for Comparison of τAD++ vs. Non-τAD++ ADAS-Cog11 Change at 18 Months

The same analyses will be performed as described in Section 8.4.2.5 with the substitution of ADAS-Cog11 for CDR-SB.

8.4.2.8 MMRM for Comparison of τAD++ vs. Non-τAD++ FAQ Change at 18 Months

The same analyses will be performed as described in Section 8.4.2.5 with the substitution of FAQ for CDR-SB.

8.4.2.9 Inter-reader Consistency

The inter-reader consistency of flortaucipir scan interpretation (τAD++ vs. non-τAD++) across the 5 independent readers will be assessed using a Fleiss' Kappa statistics. The overall percent of agreement, Fleiss' Kappa, and 95% CI around the Kappa will be summarized in a table. Pairwise comparisons between readers will also be presented with simple kappa statistics evaluating agreement.

8.4.3 Exploratory Analysis

8.4.3.1 Risk Ratio of τAD++ vs. Non-τAD++ Subjects in Progression to Clinically Meaningful Deterioration Evaluated by CDR-SB at 24 Months

The same analysis as described in section 8.4.1.1 will be performed for CDR-SB at 24 months.

8.4.3.2 Risk Ratio of τAD++ vs. Non-τAD++ Subjects in Progression to Clinically Meaningful Deterioration Evaluated by MMSE at 24 Months

The same analysis as described in section 8.4.1.1 will be performed for clinically meaningful deterioration in MMSE (≤ -3 point change) at 24 months.

8.4.3.3 Risk Ratio of τAD++ vs. Non-τAD++ Subjects in Progression to Clinically Meaningful Deterioration Evaluated by ADAS-Cog11 at 24 Months

The same analysis as described in section 8.4.1.1 will be performed for clinically meaningful deterioration in ADAS-Cog11 (≥4 point change) at 24 months.

8.4.3.4 Risk Ratio of τAD++ vs. Non-τAD++ Subjects in Progression to Clinically Meaningful Deterioration Evaluated by FAQ at 24 Months

The same analysis as described in section 8.4.1.1 will be performed for clinically meaningful deterioration in FAQ (≥3 point change) at 24 months.

8.4.3.5 Risk Ratio of τAD++ vs. Non-τAD++ Subjects in Progression to Clinically Meaningful Deterioration Evaluated by CDR Global at 24 Months

The same analysis as described in section 8.4.1.1 will be performed for clinically meaningful deterioration in CDR Global (>0 point change) at 24.

8.4.3.6 Risk Ratio of τAD (τAD+/τAD++) vs. τAD- Subjects in Progression to Clinically Meaningful Deterioration Evaluated by CDR-SB at 18 Months and 24 Months

The same analysis as described in section 8.4.1.1 will be performed for CDR-SB, comparing τAD (τAD+/τAD++) and τAD- groups at 18 months and 24 months.

8.4.3.7 Risk Ratio of τAD vs. τAD- Subjects in Progression to Clinically Meaningful Deterioration Evaluated by MMSE at 18 Months and 24 Months

The same analysis as described in section 8.4.1.1 will be performed for clinically meaningful deterioration in MMSE (≤-3 point change), comparing τAD (τAD+/τAD++) and τAD- groups at 18 months and 24 months.

8.4.3.8 Risk Ratio of τAD vs. τAD- Subjects in Progression to Clinically Meaningful Deterioration Evaluated by ADAS-Cog11 at 18 Months and 24 Months

The same analysis as described in section 8.4.1.1 will be performed for clinically meaningful deterioration in ADAS-Cog11 (≥4 point change), comparing τAD (τAD+/τAD++) and τAD- groups at 18 months and 24 months.

8.4.3.9 Risk Ratio of τAD vs. τAD- Subjects in Progression to Clinically Meaningful Deterioration Evaluated by FAQ at 18 Months and 24 Months

The same analysis as described in section 8.4.1.1 will be performed for clinically meaningful deterioration in FAQ (≥3 point change) comparing τAD (τAD+/τAD++) and τAD- groups at 18 months and 24 months.

8.4.3.10 Risk Ratio of τAD vs. τAD- Subjects in Progression to Clinically Meaningful Deterioration Evaluated by CDR Global at 18 Months and 24 Months

The same analysis as described in section 8.4.1.1 will be performed for clinically meaningful deterioration in CDR Global (>0) comparing τAD (τAD+/τAD++) and τAD- groups at 18 months and 24 months.

8.4.3.11 MMRM for Additional Comparisons of CDR-SB Change across Tau Groups

Differences between τAD++ vs non-τAD++LSM changes in CDR-SB at 24 months will also be tested and presented with the secondary analysis outlined in section 8.4.2.5.

Separate models similar to that outlined in 8.4.2.5 will be used to evaluate CDR-SB changes over time between τAD (τAD+ and τAD++) vs τAD-, and τAD++ vs τAD-.

8.4.3.12 MMRM for Additional Comparisons of MMSE Change across Tau Groups

Differences between τAD++ vs non-τAD++ LSM changes in MMSE at 24 months will also be tested and presented with the secondary analysis outlined in section 8.4.2.6.

Separate models similar to that outlined in 8.4.2.6 will be used to evaluate MMSE changes over time between τAD (τAD+ and τAD++) vs τAD-, and τAD++ vs τAD- .

8.4.3.13 MMRM for Additional Comparisons of ADAS-Cog11 Change across Tau Groups

Differences between τAD++ and non-τAD++ LSM changes in ADAS-Cog11 at 24 months will also be tested and presented with the secondary analysis outlined in section 8.4.2.7.

Separate models similar to that outlined in 8.4.2.7 will be used to evaluate ADAS Cog11 changes over time between τAD (τAD+ and τAD++) and τAD-, and τAD++ and τAD-.

8.4.3.14 MMRM for Additional Comparisons of FAQ Change across Tau Groups

Differences between τAD++ and non-τAD++ LSM changes in FAQ at 24 months will also be tested and presented with the secondary analysis outlined in section 8.4.2.8.

Separate models similar to that outlined in 8.4.2.8 will be used to evaluate FAQ changes over time between τAD (τAD+ and τAD++) and τAD-, and τAD++ and τAD-.

8.4.3.15 Diagnostic Performance of Baseline Tau Status in Predicting Clinically Meaningful Deterioration Evaluated by CDR-SB

This analysis will use dichotomized CDR-SB Change (1 point or more increase vs. otherwise at 18 months) as a truth standard to assess the diagnostic performance of baseline tau status as determined by flortaucipir scan. Sensitivity, specificity, accuracy,

PPV, and NPV will be presented in a table, along with their respective 95% Wilson score confidence intervals, with details as below:

	Clinically Meaningful Deterioration Change +	Change-
τAD++	A	B
Non-τAD++	C	D

- Sensitivity = A*/(A+C)*
- Specificity = *D/(B+D)*
- Accuracy = *(A+D)/(A+B+C+D)*
- *PPV*
- *NPV*

8.4.3.16 Diagnostic Performance of Baseline Tau Status in Predicting Clinically Meaningful Deterioration Evaluated by MMSE

The same analysis as described in section 8.4.3.15 will be performed for clinically meaningful deterioration in MMSE (≤-3 point change).

8.4.3.17 Diagnostic Performance of Baseline Tau Status in Predicting Clinically Meaningful Deterioration Evaluated by ADAS-Cog11

The same analysis as described in section 8.4.3.15 will be performed for clinically meaningful deterioration in ADAS-Cog11 (≥4 point change).

8.4.3.18 Diagnostic Performance of Baseline Tau Status in Predicting Clinically Meaningful Deterioration Evaluated by FAQ

The same analysis as described in section 8.4.3.15 will be performed for clinically meaningful deterioration in FAQ (≥3 point change).

8.4.3.19 Diagnostic Performance of Baseline Tau Status in Predicting Clinically Meaningful Deterioration Evaluated by CDR Global

The same analysis as described in section 8.4.3.15 will be performed for clinically meaningful deterioration in CDR Global (>0 point change).

8.4.4 Subgroup Analyses

The primary and secondary analyses described in sections 8.4.1 and 8.4.2, with the exception of inter-reader consistency, will be repeated by age group (≤75 years; > 75 years), sex (female; male), and baseline CDR Global score (≥1; < 1). The analysis in 8.4.2.4 will not be run by the CDR Global subgroups as this model is already adjusted for baseline CDR Global score.

8.4.5 Sensitivity Analysis

To avoid misinterpretation of potentially biased estimates caused by missing data, the MMRM analyses described in sections 8.4.2.5 through 8.4.2.8 will be run on the set of 90 subjects who completed the 24 month visit.

9 REFERENCES

- Jack, C R, Knopman, D S, Jagust, W J, Shaw, LM, Aisen, P S, Weiner, M W, Petersen, R C, Trojanowski, J Q. Hypothetical model of dynamic biomarkers of the Alzheimer;s pathological cascade. Lancet Neruol 9(1): 119-128, 2010
- Price, J.L., Morris, J.C. "Tangles and plaques in nondemented aging and "preclinical" Alzheimer's disease." Ann Neurol 45:358–368, 1999.
- Sperling, R.A., Aisen, P.S., Beckett, L.A., Bennett, D.A., Craft, S., Fagan, A.M., Iwatsubo, T., Jack, C.R., Jr., Kaye, J., Montine, T.J., Park, D.C., Reiman, E.M., Rowe, C.C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M.C., Thies, B., Morrison-Bogorad, M., Wagster, M.V., Phelps, C.H. 2011. "Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease". Alzheimers Dement 7(3), 280-92.
- Berg L. Clinical Dementia Rating (CDR). Psychopharmacol Bull 1988; 24:637-639.
- Folstein, M.R., Folstein, S.E., and McHugh, P.R. "Mini Mental State": A practical method for grading the cognitive status of patients for the clinician. Journal of Psychiatric Research, 12: 189-198, 1975.
- Rosen, W.G., Mohs, R.C., and David, K.L. A new rating scale for Alzheimer's disease. American Journal of Psychiatry, 141: 1356-1364, 1984.
- Pfeffer, R.I., Kurosaki, T.T., Harrah, C.H., Chance, J.M., Filos, S. Measurement of functional activities in older adults in the community. J Gerontol, 37(3): 323-329, 1982.
- Zou G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. American Journal of Epidemiology. 2004; 159 (7):702-706.

Leo Document ID = 52cdf45d-6802-4c95-baca-2fb78bd0183e

Approver: [redacted]
Approval Date & Time: 16-Apr-2019 17:25:59 GMT
Signature meaning: Approved

Approver: [redacted]
Approval Date & Time: 17-Apr-2019 00:28:06 GMT
Signature meaning: Approved

Approver: [redacted]
Approval Date & Time: 22-Apr-2019 15:53:15 GMT
Signature meaning: Approved

Approver: [redacted]
Approval Date & Time: 24-Apr-2019 15:02:57 GMT
Signature meaning: Approved