18F-AV-1451-FR01 SAP Amendment 2

# A Reader Study to Assess Accuracy and Reliability of Flortaucipir F 18 PET Scan Interpretation

NCT03901092

Approval date: 26 Apr 2019

# STATISTICAL ANALYSIS PLAN

## DATE OF PLAN:

*26-April-2019*

## STUDY DRUG:

*Flortaucipir $^{18}$F ($^{18}$F-AV-1451)*

## PROTOCOL NUMBER:

*$^{18}$F-AV-1451-FR01 Amendment 1*

## STUDY TITLE:

*A READER STUDY TO ASSESS ACCURACY AND RELIABILITY OF FLORTAUCIPIR F 18 PET SCAN INTERPRETATION*

## SPONSOR:

*Avid Radiopharmaceuticals, Inc.*

[redacted]

This study is being conducted in compliance with good clinical practice, including the archiving of essential documents.

Approval Date: 30-Apr-2019 GMT

# SIGNATURE PAGE

This document has been prepared and/or reviewed by:

[redacted] [redacted]

S gnature Date

This document has been reviewed and accepted by:

[redacted] [redacted]

Signature

[redacted] [redacted]

Signature

# TABLE OF CONTENTS

## LIST OF TABLES

# 1. LIST OF ABBREVIATIONS

**Table 1: List of Abbreviations**

| | |
|---|---|
| **AD** | **Alzheimer's Disease** |
| **CN** | **Cognitively Normal** |
| **TS** | **Truth Standard** |
| **MCI** | **Mild Cognitive Impairment** |
| **SAP** | **Statistical Analysis Plan** |
| **CRO** | **Contract Research Organization** |
| **ODD** | **Other Dementia Disorder** |
| **τAD-** | **Not consistent with AD pattern** |
| **τAD** | **AD pattern** |
| **CRF/eCRF** | **Case Report Form/ Electronic Case Report Form** |
| **SAS** | **SAS® software** |
| **ADNC** | **AD Neuropathological Change** |
| **EFFA** | **Efficacy Analysis Population for Accuracy Evaluation** |
| **EFFP** | **Efficacy Analysis Population for Precision Evaluation** |
| **SAC** | **Supplementary Academic Cohort** |
| **PET** | **Positron Emission Tomography** |
| **PPV** | **Positive Predictive Value** |
| **NPV** | **Negative Predictive Value** |
| **LR+** | **Likelihood Ratio Positive** |
| **LR-** | **Likelihood Ratio Negative** |
| **FN** | **False Negative** |
| **TN** | **True Negative** |
| **FP** | **False Positive** |
| **TP** | **True Positive** |

# 2. INTRODUCTION

In prior clinical studies supporting accuracy and reliability of flortaucipir F 18 positron emission tomography (PET) scans, interpretation has been shown to have acceptable performance when compared to the autopsy truth standard (TS).

This study is designed to further evaluate the accuracy and reliability of multiple readers' interpretations of flortaucipir F 18 PET scans not only from those subjects who came to autopsy, but also in the intended population for clinical use.

The purpose of this statistical analysis plan (SAP) is to describe the statistical analyses for this study. The SAP should be read in conjunction with the protocol.

# 3. STUDY OBJECTIVES

## 3.1. Primary Objective

There are two co-primary objectives for this study:

- Test the relationship between ante-mortem flortaucipir F 18 PET imaging and tau neurofibrillary pathology associated with Alzheimer's disease (AD), as measured at autopsy.
- Assess inter-reader reliability

## 3.2. Secondary Objective

There are four secondary objectives for this study:

- Test the relationship between ante-mortem flortaucipir F 18 PET imaging of an AD pattern with uptake beyond the temporal/occipital regions (i.e., τAD++; refer to Protocol Table 1) and tau neurofibrillary pathology associated with Alzheimer's disease (AD), as measured at autopsy
- Assess inter-reader reliability for scans with an AD pattern that is beyond the temporal/occipital regions (i.e., τAD++; refer to Protocol Table 1)
- Assess agreement among readers of flortaucipir F 18 PET scans in subjects known to be from the intended population (interpretation of scans from Avid study $^{18}$F-AV-1451-A05 [Study A05])
- Assess intra-reader reliability for scans read twice by each reader

# 4. STUDY DESIGN

All training and reads will be conducted by an imaging contract research organization (CRO) as described in the imaging review charter (IRC). Five readers will be trained in-person on the flortaucipir F 18 PET scan read methodology using the previously developed read method. The training will consist of teaching the readers the steps of interpretation, followed by a practice session using a set of demonstration and practice cases. After the training phase is complete, readers will then independently read 262 scans; 83 from Study $^{18}$F-AV-1451-A16 (A16) and 159 from Study A05.

The study population for the A16 Study ($^{18}$F-AV-1451-A16: "A Clinico-Pathological Study of the Correspondence Between $^{18}$F-AV-1451 PET Imaging and Post-Mortem Assessment of Tau Pathology") consisted of subjects at the end of life, who were imaged with flortaucipir F 18 and came to autopsy. Subjects were enrolled in the study with the intent of capturing a range of tau neurofibrillary pathology in AD.

The study population for the A05 Study Confirmatory Cohort ($^{18}$F-AV-1451-A05 "An open label, multicenter study, evaluating the safety and imaging characteristics of $^{18}$F-AV-1451 in cognitively healthy volunteers, subjects with Mild Cognitive Impairment, and subjects with Alzheimer's disease") consisted of subjects with cognitive impairment who had Mild Cognitive Impairment (MCI) or dementia with a suspected neurodegenerative cause, and a Mini Mental State Examination (MMSE) score of 20-27, inclusive. Subjects were imaged with flortaucipir F 18 and followed longitudinally for 18 months to assess the subsequent rate of cognitive decline.

# 5. SAMPLE SIZE JUSTIFICATION

Assuming 80% sensitivity and 80% specificity, 14 truth standard (TS) positive or TS negative cases will be needed to show the lower bounds of 95% confidence interval (CI; 2-sided) greater than 50%, for either sensitivity or specificity, with a Wilson score method to calculate the 95% CI. Out of the 83 A16 cases with autopsy results, 47 were neurofibrillary tangle (NFT) stage B3 per pathologist's panel diagnosis (TS positive for primary objective #1 efficacy analysis #1), and 36 were NFT stage B2 or lower (TS negative for primary objective #1 efficacy analysis #1); 41 were high AD neuropathologic change ( ADNC) according to panel diagnosis (TS positive for primary objective #1 efficacy analysis #2), and 42 were low to intermediate ADNC (TS negative for primary objective #1 efficacy analysis #2). Therefore, this sample size provides adequate power to assess the accuracy of flortaucipir F 18 PET scan in detecting underlying pathological changes.

Fleiss' Kappa will be used to assess the inter-reader reliability in flortaucipir F 18 scan visual interpretation. Five independent qualified physicians will read 242 unique scans (83 autopsy scans from study A16 and 159 scans from study A05). Assuming the Fleiss' kappa is expected to be 0.7, this sample size provides over 90% power in detecting a kappa value greater than or equal to 0.60, under a two-sided type I error rate of 0.05.

# 6. FLORTAUCIPIR PET SCAN SELECTION

Flortaucipir F 18 PET scans were selected from subjects enrolled in previous studies $^{18}$F-AV-1451-A16 and $^{18}$F-AV-1451-A05. The following images will be used for this study:

- All 83 subjects from Study A16 (primary and supplemental) who have a valid scan and autopsy. This includes:
  - 3 front-runners
  - 64 main study autopsy cases
  - 16 autopsy cases collected under SAC
- All 159 subjects from the study $^{18}$F-AV-1451-A05 who have a valid scan, representing the intended population for clinical use.
- 20 scans will be randomly selected (by simple random selection method) from studies A16 and A05, to be read twice independently by each of the 5 readers for intra-reader reading reliability evaluation purposes.

The majority read results from A05 and A16 both showed a ~56% scan positive (τAD+/τAD++) rate. Similarly, it is expected that the 20 scans randomly selected for intra-reader reliability evaluation will have a similar scan positive rate. If by chance the randomly selected 20 cases have a skewed positive or negative rate, to avoid possible interpretation bias from a sample population of primarily positive or negative scans, the random sampling process will be repeated to ensure a relatively balanced positive/negative rate from the selected samples (scan positive rate by majority reads from corresponding study is in the range of 30% - 70%).

An external CRO will randomize all scans into a randomized sequence for visual reading.

# 7. DEFINITIONS AND CONVENTIONS

## 7.1. General Summary Table and Individual Patient Data Listing Considerations

Summary tables and listings will be prepared according to ICH Guideline E3 that will include a footer providing the following notes:

1. Date of output generation.
2. SAS program name that generates the output.
3. Reference to the data listings that the summary table based on.
4. Any other output specific details that require further elaboration.

## 7.2. Analysis Populations

Since there will be no flortaucipir F 18 dose in this study, there will be no safety evaluation planned thus the analysis population will be for efficacy analyses only.

Efficacy analysis population will include all valid scan reading results from 5 readers on 242 cases. Valid images will be considered unevaluable only if 3 out of 5 independent readers declare the image unevaluable for the same reason. Subjects with invalid or unevaluable PET data will be excluded from efficacy analyses. Criteria for declaring an image invalid or not evaluable will be specified in advance in the Image Review Charter.

### 7.2.1. Efficacy Analysis Population for Accuracy Evaluation (EFFA)

The EFFA will include reading results for all valid and interpretable scans from all 83 autopsy cases from study A16. The analysis set will be used to conduct the primary objective #1 efficacy analyses #1 and #2 (refer to Section 10.3.1 and 10.3.2), and secondary objective #1 analysis #1 and #2 (refer to section 10.4.1).

For scans that will be read twice for intra-reader reliability evaluation, the first read will be included in the EFFA.

Sixteen cases enrolled as supplementary academic center (SAC) autopsy cases only had neuropathologist diagnosis from one hemisphere. Therefore, the visual reads from the corresponding side will be used in the primary objective analyses where the accuracy of τAD visual reads is assessed, since the τAD visual read data is available from both sides of the brain. For secondary objective analyses where the accuracy of τAD++ visual reads is assessed, the τAD++ read for the whole brain will be used due to data availability.

### 7.2.2. Efficacy Analysis Population for Precision Evaluation (EFFP)

The EFFP are defined according to analyses, as below:

#### 7.2.2.1. EFFP1

EFFP1 will include reading results for all valid and interpretable scans from all 242 cases from studies A16 and A05. The analysis set will be used to conduct primary objective #2 efficacy analysis (refer to Section 10.3.3), and secondary objective #2 efficacy analysis (refer to section 10.4.2).

For scans that will be read twice for intra-reader reliability evaluation, the first read will be included in the EFFP1. For the 16 SAC cases, the overall visual interpretation based on whole brain will be used.

#### 7.2.2.2. EFFP2

EFFP2 will include reading results for all valid and interpretable scans from the 159 cases from study A05. The analysis set will be used to conduct analysis for secondary objective #3 (refer to Section 10.4.3).

For scans that will be read twice for intra-reader reliability evaluation, the first read will be included in the EFFP2.

#### 7.2.2.3. EFFP3

EFFP3 will include reading results for all valid and interpretable scans from the 20 cases randomly selected to evaluate intra-reader reliability. The analysis set will be used to conduct efficacy analysis for secondary objective #4 (refer to Section 10.4.4). For the 16 SAC cases, the overall visual interpretation based on whole brain will be used.

# 8. DATA PRESENTATIONS AND DATA MANAGEMENT

## 8.1. Data Presentations

The data will be presented with tables and listings in the following categories:

1. Demographic profile for the study participants
2. Analyses of the primary, secondary and exploratory efficacy data

## 8.2. Data Management

The data will be collected using an electronic case report form (eCRF) system that was developed by data vendor American College of Radiology (ACR). There will be an edit check and quality control before the database lock to ensure the highest possible quality of the study data.

Derived working datasets will be created using the Statistical Analysis System (SAS) software. Data analyses and summary tables will be generated using SAS version 9.0 or higher.

## 8.3. General Post Text Summary

### 8.3.1. Table and Individual Patient Data Listing Format Considerations

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 two digits per level (e.g., Table XX.YY.ZZ…). Tables will be presented in Appendix 14 and individual patient data listings are presented in Appendix 16.

# 9. BASELINE PATIENT DATA AND EFFICACY ENDPOINT

## 9.1. Demographic, PET Scan, and Autopsy Data

The baseline data will be carried forward from the previous studies $^{18}$F-AV-1451-A05 and $^{18}$F-AV-1451-A16, which include the following:

- Demographics: age, gender, race, and ethnicity
- PET scan and autopsy information:
  Date of flortaucipir PET scan, date of autopsy, and Interval between image and autopsy.

## 9.2. Neuropathological Evaluation

All neuropathological measurements on brain tissue will be obtained from study A16, which were evaluated in a standardized fashion in a qualified laboratory.

Neuropathological assessment was derived from the NIA-AA guidelines (Hyman et al., 2012). The truth standard (TS) for primary objective #1 efficacy analysis #1 was constructed from NFT scores at autopsy as shown in Table 2.

**Table 2. Autopsy NFT Score Truth Standard**

| Braak Stage | NFT Score[A] | Truth Standard |
|---|---|---|
| 0 (no NFTs) | B0 | Negative |
| I-II | B1 | |
| III-IV | B2 | |
| V-VI | B3 | Positive |

[A]Adapted from Hyman et al., 2012.

The TS for primary objective #1 efficacy analysis #2 was constructed from levels of AD neuropathological change as shown in Table 3.

**Table 3. NIA-AA Autopsy Diagnosis Level of AD Neuropathological Change Truth Standard**

| Level of AD Neuropathological Change Present[A] | Truth Standard |
|---|---|
| None | Negative |
| Low | |
| Intermediate | |
| High | Positive |

[A]Adapted from Hyman et al., 2012.

## 9.3. Flortaucipir F 18 PET Imaging Evaluation

Criteria for declaring an image invalid or not evaluable were specified in advance in the Image Review Charter. In this SAP, valid images will be considered as unevaluable only if at least 3 out of 5 independent readers declare the image unevaluable for the same reason(s).

Flortaucipir F 18 PET images will be interpreted by visual examination using the following criteria:

| Read Outcome | | Objective Image Features |
|---|---|---|
| Not consistent with AD pattern (τAD-) | | No increased neocortical activity, or increased neocortical activity isolated to the mesial temporal, anterolateral temporal, and/or frontal regions. |
| AD pattern (τAD) | 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 purpose of testing primary objective #1, a scan interpretation will be considered positive (hypothesized to correspond to a B3 NFT score for primary objective #1 efficacy analysis 1 and an NIA-AA score of high AD pathology for primary objective #1 efficacy analysis #2) if the scan was interpreted as at least consistent with an AD pattern (τAD+ or τAD++).

For the purpose of testing secondary objective #1, a scan interpretation will be considered positive (hypothesized to correspond to a B3 NFT score for secondary objective #1 efficacy analysis #1 and an NIA-AA score of high AD pathology for secondary objective #1 efficacy analysis #2) if the scan was interpreted as τAD++.

# 10. EFFICACY ANALYSIS

## 10.1. General Considerations

Categorical variables will be summarized as frequencies and percentages by enrolling diagnostic group (Dementia, MCI, or CN). Continuous variables will be summarized using descriptive statistics (n, mean, median, standard deviation, minimum and maximum) by each diagnostic group.
All inferential statistics will be performed at the 2-sided, 0.05 level of significance.

SAS Version 9.0 or above will be used to perform all analyses.

## 10.2. Multiple comparisons/multiplicity of hypotheses testing

Three hypotheses will be tested for the two co-primary objectives, as detailed in section 10.3. This trial will be considered successful if all three primary efficacy analyses achieve their success criteria.

## 10.3. Analyses of the Primary Objectives

### 10.3.1. Primary Objective #1 Efficacy Analysis 1: Accuracy of Flortaucipir F 18 PET Scan Interpreted as AD Pattern in Detecting NFT Tau Stage

The first primary analysis will evaluate the accuracy of flortaucipir F 18 PET scan in detecting neurofibrillary tangle as measured by NFT stage (truth standard), as detailed in Table 1 of section 9.2. The hypothesis to be tested is that for at least the same 3 out of 5 independent readers, the lower bound of the two-sided 95% CI for both sensitivity and specificity of flortaucipir PET reading interpretations will be $\geq 50\%$.

This analysis will be based on the EFFA, as defined in 7.2.1.

The following table contains terms used to define sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), likelihood ratio positive (LR+) and likelihood ratio negative (LR-):

| **Diagnostic Test: Flortaucipir F 18 PET Scan Read Result** | **Truth Standard** | |
|---|---|---|
| | **NFT B3** | **NFT B0-B2** |
| Positive: Flortaucipir F 18 neocortical uptake; AD pattern (τAD+, τAD++) | True Positive (TP) | False Positive (FP) |
| Negative: Flortaucipir F 18 neocortical uptake; not consistent with AD pattern (τAD-) | False Negative (FN) | True Negative (TN) |

The sensitivity, which evaluates how good a test is at detecting a positive disease, will be calculated as:

$$100 \times \frac{TP}{TP + FN}$$

The specificity, which estimates how likely subjects without the disease can be correctly ruled out, will be calculated as:

$$100 \times \frac{TN}{FP + TN}$$

The accuracy, which evaluates the overall proportion of observed agreement, will be calculated as:

$$100 \times \frac{TP + TN}{TP + FN + FP + TN}$$

The PPV, which is the proportion of positive test results that are true positives, will be calculated as:

$$100 \times \frac{TP}{TP + FP}$$

The NPV, which is the proportion of subjects with a negative test result who are correctly diagnosed, will be calculated as:

$$100 \times \frac{TN}{FN + TN}$$

The 95% CIs provided for specificity, sensitivity, accuracy, PPV and NPV will be based on the Wilson score method (Wilson 1927; Newcombe 1998).

The LR+, which is true positive rate divided by the false positive rate, will be calculated as:

$$\frac{sensitivity}{1 - specificity}$$

The LR-, which is the false negative rate divided by the true negative rate, will be calculated as:

$$\frac{1 - sensitivity}{specificity}$$

### 10.3.2. Primary Objective #1 Efficacy Analysis #2: Accuracy of Flortaucipir F 18 PET Scan Interpreted as AD Pattern in Detecting AD Neuropathological Change

This analysis will evaluate the accuracy of flortaucipir F 18 PET scan in detecting High AD neuropathological change (ADNC), as detailed in Table 2 of section 9.2. Like primary analysis #1, the hypothesis to be tested is that for at least the same 3 out of 5 independent readers, the lower bound of the two-sided 95% CI for both sensitivity and specificity of flortaucipir PET reading interpretations will be $\geq$ 50%.

This analysis will be based on the EFFA, as defined in 7.2.1.

The calculation of sensitivity, specificity, NPV, PPV, LR+ and LR- analyses, as well as the confidence interval calculation will be identical to primary analysis #1, with ADNC as TS.

### 10.3.3. Primary Objective #2 Efficacy Analysis: Inter-Reader Agreement of Flortaucipir F 18 PET Scan Interpreted as AD Pattern

The overall reader to reader reliability across 5 readers for interpreting the flortaucipir PET scan will be assessed using Fleiss' Kappa. The primary hypothesis to be tested is that the lower bound of two-sided 95% confidence interval of Fleiss' Kappa will be greater than or equal to 0.6. This analysis will be based on the EFFP1, as defined in 7.2.2.1.

The degree of agreement between two readers for the interpretation of flortaucipir PET scan will be assessed in a pair-wise manner using Cohen's Kappa statistics.

The percent of agreement between the two readers will be calculated for each reader pair. The over percent of agreement across 5 readers will be calculated too. All these results will be summarized in tables.

## 10.4. Analysis of Secondary Objectives

### 10.4.1. Test the Relationship Between Ante-mortem Flortaucipir F 18 PET Imaging of an AD Pattern with Uptake Beyond the Temporal/Occipital Regions (τAD++) and Tau Neurofibrillary Pathology Associated with AD, as Measured at Autopsy

To further evaluation the diagnostic performance of flortaucipir scans, the same scan interpretation from the readers will be reclassified as an AD pattern with uptake beyond the temporal/occipital regions (τAD++: positive) vs. otherwise (τAD+/ τAD-: negative) and compared to the TS as detailed in section 9.2. That is, calculating the diagnostic performance statistics following the tables as below:

Secondary Objective #1 Efficacy Analysis #1

| Diagnostic Test: Flortaucipir F 18 PET Scan Read Result | **Truth Standard** | |
|---|---|---|
| | **NFT Score B3 [Truth Positive]** | **NFT Score B0–B2 [Truth Negative]** |
| Positive: Flortaucipir F 18 neocortical uptake; τAD++ pattern | True Positive (TP) | False Positive (FP) |
| Negative: Flortaucipir F 18 neocortical uptake; non- τAD++ pattern (τAD-/ τAD+) | False Negative (FN) | True Negative (TN) |

And:

Secondary Objective #1 Efficacy Analysis #2

| Diagnostic Test: Flortaucipir F 18 PET Scan Read Result | **Truth Standard** | |
|---|---|---|
| | **High AD Neuropathologic Change [Truth Positive]** | **Not, Low, or Intermediate AD Neuropathologic Change [Truth Negative]** |
| Positive: Flortaucipir F 18 neocortical uptake; τAD++ pattern | True Positive (TP) | False Positive (FP) |
| Negative: Flortaucipir F 18 neocortical uptake; non-τAD++ pattern (τAD-/ τAD+) | False Negative (FN) | True Negative (TN) |

The calculation of sensitivity, specificity, NPV, PPV, LR+, LR-, as well as the associate 95% CI will be the same as described in section 10.3.1.

### 10.4.2. Inter-Reader Reliability of Flortaucipir F 18 PET Scan Interpreted as τAD++ Pattern

This analysis will assess the reliability of flortaucipir F 18 scan interpreted as τAD++ pattern across 5 readers. Same as analysis described in section 10.3.3, Fleiss' Kappa and the associated two-sided 95% CI will be calculated. The hypothesis to be tested is that the lower bound of two-sided 95% confidence interval of Fleiss' Kappa will be greater than or equal to 0.6. This analysis will be based on the EFFP1, as defined in 7.2.2.1.

### 10.4.3. Inter-Reader Agreement of Flortaucipir F 18 PET Scan Interpreted as AD Pattern for Intended Clinical Use Population

The subjects enrolled in study A05 were representative cases of the intended clinical use population for flortaucipir F 18 scan. This analysis will evaluate the inter-reader reliability of flortaucipir scan interpreted as AD pattern with this intended clinical practice population. Analyses similar to 10.3.3 will be conducted using EFFP2, which will be based on 159 A05 cases as defined in 7.2.2.2.

### 10.4.4. Intra-Reader Agreement of Flortaucipir F 18 PET Scan Visual Interpretation Interpreted as AD Pattern

Intra-reader reliability will be assessed using randomly selected 20 cases (EFFP3) as defined in section 7.2.2.3. These randomly selected cases will be read twice by every reader. A Cohen's Kappa statistics will be used to assess the agreement of the two reading results (scans that are interpreted as an AD pattern, i.e., τAD (τAD+/ τAD++) vs τAD-) by every reader.

The percent of agreement between the two readings from same reader will be calculated for each reader. All these results will be summarized in tables.

## 10.5. Other Efficacy Analyses

### 10.5.1. Accuracy of Flortaucipir F 18 Scan in Detecting Tau Neurofibrillary Pathology Associated with Alzheimer's Disease (AD), Based on Majority Scan Interpretations

For each image, the majority interpretation of 5 independent readers will be derived using individual readers' interpretation results. Diagnostic performance of this majority interpretation will be assessed relative to the autopsy NFT score TS as per primary objective #1 efficacy analysis #1 (refer to Section **Error! Reference source not found.**0.3.1) and again relative to the NIA-AA autopsy diagnosis TS as per primary objective #1 efficacy analysis #2 (refer to Section **Error! Reference source not found.**0.3.2).

The sensitivity and specificity along with their two-sided 95% CIs (based on the Wilson score method) will be calculated for the majority interpretation of 5 independent readers relative to the appropriate TS.

The accuracy, PPV, NPV, LR+ and LR- of the flortaucipir F 18 PET imaging classification for the majority interpretation of 5 independent readers relative to the appropriate TS will be also calculated. Two-sided 95% CIs (based on the Wilson score method) will be provided for the accuracy, the PPV and NPV.

### 10.5.2. Inter-Reader Agreement of Flortaucipir F 18 PET Scan Interpreted as AD Pattern for Autopsy Population

The inter-reader agreement analysis across 5 readers of flortaucipir F 18 PET scan interpretation (interpreted as AD pattern) will be repeated based on the EFFA. Fleiss' Kappa, and other statistics, will be calculated the same as described in section 10.3.3.

### 10.5.3. Additional Visual Interpretation Reliability Assessment of Flortaucipir F 18 Scan Interpreted as τAD++ Pattern

These additional visual interpretation reliability assessments will be conducted for flortaucipir F 18 scans based on the τAD++ pattern:

1) Inter-reader reliability for scans including clinical use population only (A05 cases):

> Fleiss' Kappa, and other statistics will be calculated the same as described in section 10.3.3., based on EFFP2.

2) Inter-reader reliability for scans including autopsy cases only (A16 cases):

> Fleiss' Kappa, and other statistics will be calculated the same as described in section 10.3.3., based on EFFA.

3) Intra-reader reliability for scans included in EFFP3:

Cohen's kappa and other statistics will be calculated the same as described in section 10.4.4, based on EFFP3.

### 10.5.4. Diagnostic Performance Calculation for Flortaucipir F 18 PET Scan Reader Interpretation vs. Modified Autopsy NFT Score TS (B0/1 vs. B2/3)

Flortaucipir F 18 scan visual interpretation, both as described in primary objective #1 analyses and secondary objective #1 analyses will be compared to a modified autopsy NFT score TS, which classifies NFT score B2 and B3 as truth standard positive, and NFT score B0 and B1 as truth standard negative.

The diagnostic performance statistics calculation will be based on the tables as below:

Flortaucipir scan interpretation τAD pattern (τAD+/ τAD++) as scan reads positive:

| Diagnostic Test: Flortaucipir F 18 PET Scan Read Result | **Modified Truth Standard** | |
|---|---|---|
| | **NFT Score B2/B3 [Truth Positive]** | **NFT Score B0/B1 [Truth Negative]** |
| Positive: Flortaucipir F 18 neocortical uptake; AD pattern (τAD+, τAD++) | True Positive (TP) | False Positive (FP) |
| Negative: Flortaucipir F 18 neocortical uptake; not consistent with AD pattern (τAD-) | False Negative (FN) | True Negative (TN) |

And τAD++ pattern as scan reads positive:

| Diagnostic Test: Flortaucipir F 18 PET Scan Read Result | **Modified Truth Standard** | |
|---|---|---|
| | **NFT Score B2/B3 [Truth Positive]** | **NFT Score B0/B1 [Truth Negative]** |
| Positive: Flortaucipir F 18 neocortical uptake; τAD++ pattern | True Positive (TP) | False Positive (FP) |
| Negative: Flortaucipir F 18 neocortical uptake; non-τAD++ pattern (τAD-/ τAD+) | False Negative (FN) | True Negative (TN) |

The calculation of sensitivity, specificity, NPV, PPV, LR+, LR-, as well as the associate 95% CI will be the same as described in section 10.3.1.

# 11. SENSITIVITY ANALYSIS

The primary objective analyses will be repeated by using the whole brain visual reads (for τAD pattern) for SAC cases. Identical analyses will be conducted as described in section 10.3.

# 12. SAFETY ASSESSMENT

No flortaucipir F 18 scan/dosing will be given to patient in this study, therefore the safety evaluation will not be conducted.

# 13. REFERENCES

1. Hyman BT, Phelps CH, Beach TG at al. (2012). National Institute on Aging-Alzheimer's Association Guidelines for the Neuropathologic Assessment of Alzheimer's Disease. Alzheimers Dement.; 8(1): 1-13.
2. Montine TJ et al. (2012). National Institute on Aging-Alzheimer's Association Guidelines for the Neuropathologic Assessment of Alzheimer's Disease: A Practical Approach. Acta Neuropathologica; 123.1: 1-11
3. Fleiss, J.L., Levin, B., Fleiss, J.L., Levin, B., & Paik, M.C. (2003). Statistical Methods for Rates and Proportions, 3rd
4. Newcombe RG. (1998). Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods. Statist. Med.; 17: 857-872.
5. Wilson EB. (1927). Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association; 22: 209-212