CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

202008Orig1s000

STATISTICAL REVIEW(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

CLINICAL STUDIES

NDA/Serial Number:	202-008 resubmission
Drug Name:	Amyvid
Indication(s):	AMYViD™ is a radiopharmaceutical indicated for Positron Emission Tomography (PET) imaging of β-amyloid neuritic plaques in the brains of adult patients with cognitive impairment
Applicant:	Avid Pharmaceuticals
Date(s):	Submitted date: Oct 7, 2011
	Received date: Oct 7, 2011
	PDUFA date: April 7, 2012
Review Priority:	Priority
Biometrics Division:	Division of Biometrics V
Statistical Reviewer:	Lan Huang, Ph.D.
Concurring Reviewers:	Jyoti Zalkikar (Team Leader)
	Rajeshwari Sridhara, Ph.D. (Division director)
Medical Division:	Division of Medical Imaging Products
Clinical Team:	Qi Feng, M.D. (clinical reviewer),
	Lucie Yang, M.D. (clinical team leader),
	Dwaine Rieves, M.D. (clinical division director)
Project Manager:	Sharon Thomas

Keywords: Amyloid detection, open label, reader agreement, Fleiss kappa, sensitivity and specificity,

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. EXECUTIVE SUMMARY

The medical product candidate under development is florbetapir F18 (Amyvid). Amyvid is intended for the indication (b) (4) .

This resubmission comprised of 4 completed clinical trials (involving Studies A08, A09, A16) involving 19 readers across the four studies. All images (scans) for the 4 studies were obtained in A05 and A07 studies (for more details on these studies, please go the review of the original submission with FDUFA date as March 17, 2011).

Studies A08, A09, and A16 used "in-person" reader training program. They are exploratory studies for evaluating the reproducibility and reliability of image interpretation, and reader performance characteristics (sensitivity and specificity) compared to the autopsy truth standard (using both the original set of autopsy cases as well as using the full set of autopsies completed since the original NDA submission); and for developing the web-based, self-training program conducted in each image reader's typical clinical practice setting. Study PT01, the pivotal study, evaluated the reader reliability and reader performance for the binary reading when the readers are trained by the web-based self-training program. In this review, our focus is PT01, which includes scans of the 59 autopsy cases from Study A07 and 92 non-autopsy cases from Study A05. Study A16 is also explored because it used the scans from all subjects with autopsy (59 cases from Study A07) in the whole program. Note that the 59 autopsy subjects are the same in PT01 and A16.

From Study PT01, as for inter-reader agreement among multiple readers, the kappa value is 0.83 (95% CI: 0.78, 0.88) for the 151 subjects without repeated reads, 0.75 (95% CI: 0.67, 0.83) for the 59 autopsy subjects from study A07, 0.88 (95% CI: 0.82, 0.94) for the 92 non-autopsy scans from study A05. The kappa statistics are all greater than 0.75 with lower bounds of confidence intervals greater than 0.65. The percentages of agreement for five readers are greater than 75% in different patient populations (all 151 subjects, autopsy subjects, and non-autopsy subjects). The results on kappa statistics are consistent with percent of agreement.

Sensitivity and specificity values for the subjects with autopsy are greater than 65% (lower bounds of confidence intervals >50%) for the five readers. Specifically, the sensitivity and specificity (%), respectively, are 79 (95% CI: 63.5, 90.7) and 90 (95% CI: 68.3, 98.8) for reader 1, 92 (79.1, 98.4) and 90 (68.3, 98.8) for reader 2, 69 (52.4, 83.0) and 95 (75.1, 99.9) for reader 3, 87 (72.6, 95.7) and 95 (75.1, 99.9) for reader 4, and 82 (66.5, 92.5) and 95 (75.1, 99.9) for reader 5.

The kappa statistic values are 0.83 (95% CI: 0.69, 0.97) for the 20 healthy subjects from A05 study and 0.73 (95%CI: 0.55, 0.87) for the 12 healthy subjects from A07 study; and 0.59 (95%CI: 0.47, 0.71) for the 29 AD subjects from A05 and 0.71 (95%CI: 0.57, 0.85) for the 20 AD subjects from the A07 study. This indicates that inter-reader agreement on scans from A05 study is higher than that from A07 study. Subgroup analyses on agreement indicate that readers have higher agreement on MCI patients (Kappa as 0.91 for the 57 MCI subjects), but lower

agreement on Other Dementia Disorders (ODD) patients (kappa as 0.52 for the 13 ODD subjects) .

As for intra-reader agreement, the percent of agreement values for the five readers are all greater than 90% for the readings on the 33 subjects with repeats. The time lag for the 1st reading and the 2nd reading ranges from 1 to 8 days. However, one reader has scans read within one day.

One major issue is that only five subjects who belong to the targeted population (Mild cognitive impairment (MCI) patients) have autopsy in the whole program. Therefore, sensitivity and specificity of the image read versus Standard of Truth (SoT) from autopsy can not be obtained for the targeted patient population. Even though the reader agreement is very good for MCI subjects (kappa as 0.91), but it does not indicate good sensitivity and specificity.

Additionally, from Study A16, the sensitivity and specificity of the binary reads for the 59 autopsy subjects from A05 study are 0.92 (95%CI: 0.79, 0.98) and 0.95 (95% CI: 0.75, 1) for reader 5, 0.95 (95%CI: 0.83, 0.99) and 0.95 (0.75, 1) for reader 6, 0.87 (0.72, 0.96) and 0.95 (0.75, 1) for reader 7, 0.92 (0.79, 0.98) and 1 (0.83, 1) for reader 8, and 0.69 (0.52, 0.83) and 0.90(0.68, 0.99) for reader 9. Fleiss' kappa statistic is 0.75 (95% CI: 0.67, 0.88) for the 59 subjects among the five blinded readers for the binary read analysis. Sensitivity and specificity is slightly higher for the 59 subjects in A16 study compared with PT01 study.

This application is approvable from the statistical view point in spite of the drawbacks of Fleiss' kappa because of reasonable sensitivity and specificity in a limited population and reasonably large value of Fleiss' kappa in a varied patient population. The only cautions statistical team would like to express are that 1> Amyvid scan results are indicative of the brain amyloid content only at the time of image acquisition, 2> a negative scan result does not preclude the development of brain amyloid in the future, and 3> the extremely small (n=5) number of MCI subjects had autopsy confirmation, a limiting factor in interpreting the performance characteristics (sensitivity and specificity).

2. INTRODUCTION

2.1. Overview

2.1.1. Indication

The sponsor seeks approval for the indication, (b) (4)

2.1.2. History of Program Development

The original clinical development program of florbetapir F 18 (Amyvid) comprised 6 completed clinical trials involving 496 patients: 18F-AV-45-A01 (A01), 18F-AV-45-A02 (A02), 18F-AV-45-A03 (A03), 18F-AV-45-A04 (A04), 18F-AV-45-A05 (A05), and 18F-AV-45-A07 (A07). Study A07 was the pivotal trial comparing β-amyloid levels as evaluated by florbetapir-PET imaging to post-mortem amyloid levels on histopathology. The original product development of florbetapir F18 demonstrated the efficacy of Amyvid in reflecting beta-amyloid levels in the human brain (using mainly semi-quantitative and quantitative image interpretations). More details are included in the reviews of the original submission.

Guided by recommendations of both the FDA division of medical imaging and the January 20, 2011 Peripheral and CNS Advisory Committee, Avid developed and tested the performance of a visual binary image read methodlogy for Amyvid that is suitable for routine clinical practice. This methodology has been applied following either a centralized/core lab (in-person) training of image readers (Studies A08, A09, A16), or in a web-based self-training format conducted in each image reader's typical clinical practice setting (Study PT01). The resubmission comprised of the 4 trials involving 19 readers across the four studies.

All four studies test the binary read methodology which applies a standardized set of criteria for determining if a scan is positive or negative and is proposed for florbetapir-PET scan interpretation in the routine clinical setting.

2.1.3. Specific Studies Reviewed

The resubmission comprised of the 4 trials (Studies A08, A09, A16 and PT01) involving 19 readers across the four studies.

All four studies test the binary read methodology which applies a standardized set of criteria for determining if a scan is positive or negative and is proposed for florbetapir-PET scan interpretation in the routine clinical setting.

There are no new scans for the above 4 studies (all images are from studies A07 and A05). Image reads are new. Results on Standard of Truth from autopsy are the same as those in A07 study.

There are 151 patients from A05 and A07 selected for Study PT01, 59 patients from A07 for Study A16, 40 randomly selected subjects from A05 for Study A09, 35 patients from A07 for A08.

Studies A08, A09, and A16 used "in-person" reader training program. They are exploratory studies for evaluating the reproducibility and reliability of image interpretation, and reader performance characteristics (sensitivity and specificity) compared to the autopsy truth standard (using both the original set of autopsy cases as well as using the full set of autopsies completed sinace the NDA submission); and for developing the web-based, self-training program conducted in each image reader's typical clinical practice setting. Study PT01, the pivotal study, evaluated

the reader reliability and reader performance for the binary reading when the readers are trained by the web-based self-training program.

A summary of the 4 studies is shown in Table 1.

In the following, only the pivotal study PT01 with images from 151 subjects with or without autopsy and the exploratory study A16 including all 59 autopsy subjects are explored in this review. There is no un-interpretable image in these two studies.

Table 1: Summary of Clinical Efficacy Studies Conducted to Support Validity of Reader Training Program
(Read methodology is binary read method for all studies.)

Study	# of Readers	# of Cases Read	Training Delivery Method	Training & Read Setting	Standard of Truth (SoT)	Primary Outcome Variables
A08	9	35 A07 autopsy	In person	Core lab	Neuropathology at autopsy	• Individual Reader Sensitivity and Specificity • Inter-reader reliability (overall Fleiss kappa and reader to reader kappa)
A09	7	25 MCI, 15 AD cases (40 total) from Study A05	In person	Core lab	None	• Inter-reader reliabilities (kappa) • Overall Fleiss kappa
A16	5	59 autopsy cases • 46 < one year post-scan • 13 > one year post-scan	In person	Core lab	Neuropathology at autopsy	• Sensitivity and Specificity of majority read • Individual Reader Sensitivity and Specificity • Inter-reader reliability (overall Fleiss kappa and reader to reader kappa)
PT01	5	184 • 59 autopsy cases from study A16 • 52 MCI from study A05b • 20 HC from study A05 • 20 AD from study A05 • 33 cases, including 20 MCI were reread for intra-rater reliability	Automated computer-web-based self-study (no in person contact)	Physician office	Neuropathology (for those cases with autopsy)	• Sensitivity and specificity on A16 cases • Inter-rater reliability for 119 cases: 60 cases (20 MCI, 20 AD, 20 HC) from study A05 and 59 autopsy cases from A16 • Inter-rater reliability in MCI (n=52) • Intra-rater reliability in a mix of 20 HC, MCI and AD, and separately in 20 MCI cases

2.2 Data Sources

All materials reviewed including the applicant study reports, data sets and literature referenced are provided electronically, and the full electronic path of the documents are \\Cdsesub1\evsprod\NDA202008\0013, \\Cdsesub1\evsprod\NDA202008\0014, \\Cdsesub1\evsprod\NDA202008\0018, \\Cdsesub1\evsprod\NDA202008\0021.

The application study reports for PT01 reviewed include Clinical overview, Summary of Clinical Efficacy, Summary of Clinical Safety in \\Cdsesub1\evsprod\NDA202008\0018, M2.

The application study reports for A08 are in \\Cdsesub1\evsprod\NDA202008\0013, M5.
The application study reports for A09 are in \\Cdsesub1\evsprod\NDA202008\0014, M5.
The application study reports for A16 are in \\Cdsesub1\evsprod\NDA202008\0018, M5.

Data sets analyzed for study **A08, A09, A16, and PT01** (with data definition document) were located in \\Cdsesub1\evsprod\NDA202008\0021, M5 and \\Cdsesub1\evsprod\NDA202008\0018, M5.

The datasets analyzed include adsl.xpt and adrt.xpt for PT01, adqa.xpt, adsl.xpt, adih.xpt, and adbs.xpt for A16.

3. STATISTICAL EVALUATION

3.1. Data and Analysis Quality

It was possible to reproduce the primary analyses from tabulation. Related SAS data sets (for study A08, A09 and A16) were provided to Agency after the request.

3.2. Evaluation of Efficacy

3.2.1. Study Design and Analyses

In this section, the reviewer's comments for the design, analyses are in italics.

Study PT01

The study was designed to evaluate a web-based training program that would be used to educate imaging physicians in the interpretation of florbetapir-PET images using the previously developed binary read methodology (positive or negative for significant tracer accumulation in cortical gray matter). The study included a training phase using a web-based program where readers were trained on the binary read method including, interpretation steps (sequential review of gray-scaled axial sections from inferior to superior) and read criteria (focused on loss of gray white contrast), and a validation phase where readers were tested in a format meant to simulate routine clinical practice.

The evaluation and validation of the training program was conducted by having five readers interpret florbetapir-PET images (scans) in their usual clinical image-reading environment after completion of the above training. A total of 151 unique florbetapir-PET image sets with 33 repeat images sets (for a total of 184) were grouped into 6 batches (or tranches). Readers could only submit the results of each tranche in sequential order and only one tranche could be submitted on any given calendar day. Thus, the validation reads were spread out over a minimum of 6 days. At no time did any repeat scan appear in the same tranche as the original scan.

Note that repeat images were used from completed studies. However, the reading processes in different studies are all blinded, readers are different, and reads are all new in different studies. The Standard of Truth (SoT) is the same for all studies for the autopsy subjects.

Some scans were read in the same day by one reader. Therefore, the intra-reader reliability is questionable. More details are presented in the result sections.

Readers were blinded to all demographic and clinical data of the cases they were rating and were not aware of the presence of repeat scans. Images (scans) were interpreted on the physicians' own computer monitors.

Readers rated each case as either positive for significant tracer accumulation or negative for insignificant tracer accumulation in cortical gray matter. The readers then indicated their

confidence (low, medium or high) in their assessment of each scan. In the case of a low confidence rating, readers selected the scan feature(s) contributing to reduced confidence.

There is no missing image reads because of the forced decision process.

The primary analysis evaluated the inter-reader reliability of the binary rating among the five readers. The primary outcome variable was the overall inter-reader Fleiss' kappa statistic among the 5 readers across the 119 cases. Assuming an expected Kappa of 0.70, the study had 90% power to test the hypothesis that the observed Kappa would be ≥ 0.64 with a lower bound of confidence interval ≥ 0.58 at a two-sided type I error rate of 5%.

The secondary analysis evaluated sensitivity and specificity of each reader's assessment versus the reference standard of CERAD-derived positive or negative neuropathological assessment of more than sparse β-amyloid plaques for the 59 subjects with autopsy. The hypothesis tested was that the same 3 of 5 readers would achieve a lower bound of the 95% confidence interval of at least 0.50 for both statistical measures (i.e. the same 3 of 5 readers would show lower bound of 95% confidence interval of ≥ 0.50 for sensitivity and ≥ 0.50 for specificity).

Additional exploratory analyses evaluated reader confidence, agreement with the majority read score, inter-reader reliability in MCI subjects only, and intra-reader reliability.

Inter-reader agreement and reader performance (sensitivity and specificity) need to be evaluated for the 151 images (59 cases with autopsy and 92 cases without autopsy) and for subgroups of interest.

Study A16

The study design of A16 is similar to that of PT01. The major difference is the training method. In-person training program is used to train the readers in Study A16 and web-based training program is used in Study PT01.

Only the 59 subjects with autopsy are included in Study A16. Sensitivity and specificity are evaluated in the primary analyses, and inter-reader agreement measure (kappa) is evaluated in the secondary analyses.

Study A16 is only a supportive study because the training program is not a typical type of training method used in medical practice.

3.2.2. Patient Disposition, Demographic and Baseline Characteristics

The patient's demographic and baseline characteristics for PT01 are shown in Table 2. Patients in Study A16 belong to a subset of patients in PT01 (the 59 subjects with autopsy).

The 59 autopsy subjects (scans from A05) are older than the non-autopsy subjects. This is expected since Study A05 enrolled end-of-life patients in order to obtain more autopsy cases.

About 85% of autopsy subjects are >=65 years old, and about 73% of non-autopsy subjects are >=65 years old. The percent of males and females are close to 50% in different groups. Majority of the patients are Caucasian.

Among the 151 subjects, there are 49/151 (32%) patients with baseline clinical diagnosis as Alzheimer's Disease (AD), 57/151 (38%) patients with Mild cognitive impairment (MCI), 13/151 (9%) patients with Other Dementia Disorders (ODD), 20/151 (13%) Health controls (HC) and 12/151 (8%) Normal controls (NC). Both Health controls (HC) and Normal controls (NC) are health subjects, HC is used in subjects from study A05 without autopsy and NC is used in subjects from study A07 with autopsy. The percent of AD subjects is higher in autopsy subjects (49%) and lower in non-autopsy subjects (22%). The percent of MCI subjects is very high in non-autopsy subjects (56%), and very low in autopsy subjects (8%).

Table 2: Patients disposition, demographics and Baseline diagnosis (PT01)

	Total subjects in PT01 (151)	Subjects with autopsy (59)	Non-autopsy subjects (92)
Age: mean (std)/(min, max)	75(12) / (47, 103)	79 (13) / (47, 103)	72 (10) / (50, 90)
Age <65	34 (23%)	9 (15%)	25 (27%)
Age >=65	117 (77%)	50 (85%)	67 (73%)
sex			
Male	75 (50%)	30 (51%)	45 (49%)
Female	76 (50%)	29 (49%)	47 (51%)
Races			
Caucasian	141 (93%)	55 (93%)	86 (93%)
Black or African-American	7 (5%)	3 (5%)	4 (4%)
Other	3 (2%)	1 (2%)	2 (2%)
Baseline diagnosis			
AD (from both A05 and A07)	49(32%)	29(49%)	20(22%)
HC (from A05 without autopsy)	20(13%)	0	20(22%)
MCI (from both A05 and A07)	57(38%)	5(8%)	52(56%)
NC (from A07 with autopsy)	12(8%)	12(20%)	
ODD (from A07 with autopsy)	13(9%)	13(22%)	

3.2.3. Statistical Methodologies

In this section, the reviewer's comments and alternative methods are in italics. If there is no comment, the methods proposed by the sponsor were used by the reviewer.

Study PT01:

Fleiss' kappa statistics were used to assess the overall inter-reader reliability of binary interpretation of florbetapir PET scan images. Sensitivity and specificity were calculated for the readers' binary interpretation of florbetapir-PET scan images using neuropathologist CERAD diagnosis as a reference standard. Confidence intervals for sensitivity and specificity were calculated using Wilson score method. Confidence measures and the frequency of agreement between individual image readers and the majority of all readers were presented descriptively. An intra-class kappa statistic was used to assess intra-reader reliability.

Exact method is used to obtain Clopper-Pearson intervals in the review, which are similar to intervals obtained using Wilson score method. Percent of agreement among five readers is also used to evaluate the inter-reader reliability. Percent agreement for the repeated reads is also used to evaluate the intra-reader reliability by reader.

The primary analysis evaluated the inter-reader reliability of the binary (positive vs negative scan) rating among the 5 readers. The primary outcome variable was the overall inter-reader Fleiss' kappa statistic among the 5 readers across the 119 cases. Assuming an expected Kappa of 0.70, the study had 90% power to test the hypothesis that the observed Kappa would be ≥ 0.64 with a lower bound of confidence interval ≥ 0.58 at a two-sided type I error rate of 5%.

In this review, the inter-reader agreement among the 5 readers across the total of 151 subjects is considered more important than that across the 119 cases. Note that 32 MCI subjects (targeted patient population) from A05 study were added to the original 119 cases by the sponsor, according to FDA's request.

The secondary analysis evaluated sensitivity and specificity of each reader's assessment versus the reference standard of CERAD-derived positive or negative neuropathological assessment of more than sparse β-amyloid plaques. The hypothesis tested was that the same 3 of 5 readers would achieve a lower bound of the 95% confidence interval of at least 0.50 for both statistical measures (i.e. the same 3 of 5 readers would show lower bound of 95% confidence interval of $\geq$ 0.50 for sensitivity and ≥ 0.50 for specificity).

In this review, sensitivity and specificity of the read versus the SoT for the 59 subjects are considered important secondary analyses.

Additional exploratory analyses evaluated reader confidence, agreement with the majority read score, inter-reader reliability in MCI subjects only, and intra-reader reliability.

Study A16:

Two primary analyses were conducted:
1) The sensitivity and specificity of the qualitative image read (majority rating among five readers) were measured for the population of all subjects who came to autopsy within 24 months of the florbetapir-PET scan. Two co-hypotheses for this aim were tested:
• Hypothesis A: Observed sensitivity of florbetapir-PET scan is ≥ 80%.
• Hypothesis B: Observed specificity of florbetapir-PET scan is ≥ 80%.

2) The correlation between amyloid observed on PET scan and the true level of amyloid determined at autopsy was measured, and the primary hypothesis from the linked 18F-AV-45-A07 trial (i.e., that there would be a significant correlation between the semi-quantitative assessment of PET scan and quantitative amyloid burden at autopsy) was re-tested in this study, including all subjects imaged in trial 18F-AV-45-A07 who came to autopsy under either protocol (A07 or A16).

Secondary analyses duplicated the primary analyses described above for all subjects whose brain autopsy occurred within 12 months of their florbetapir-PET scan. The 12 month PET-scan to autopsy interval was used to define the primary efficacy population for the A07 study. This interval was set, in part, to determine if the time interval between imaging and autopsy could affect the correlation and diagnostic agreement between PET image and neuropathology diagnosis results.

In this review, only sensitivity and specificity of the read versus the SoT for the 59 subjects, and the inter-reader agreement using Fleiss' kappa and percent of agreement among five readers in Study A16 were explored. Analyses by reader instead of majority read were used in the review.

3.2.4. Results and Conclusions

Brief summary of the applicant's results and conclusion

Per FDA recommendation, a binary read methodology for interpretation of florbetapir-PET scans has been developed for routine clinical implementation. A reader training program has been prepared and used to train 19 different independent blinded readers across three trials which have an autopsy neuropathology reference standard. The reader training was carried out either in a centralized lab, in-person training or in an individual clinician's standard setting using electronic, self-study training media. Blinded reader assessments of PET scans using this binary read methodology in studies A08, A16 and PT01 have shown good to excellent inter-reader repeatability (overall Fleiss' kappa values ≥0.75) together with good sensitivity and specificity (16/19 readers with sensitivity >80% and median sensitivity of 92%, and 18/19 readers with specificity >80% and a median specificity of 94%) for the detection of neuropathologically significant levels of β-amyloid neuritic plaques (CERAD moderate/frequent).

These inter-reader reliability and sensitivity and specificity results indicate the validity of the binary read methodology for detection of moderate/frequent neuritic plaque pathology (i.e.

probable/definite AD pathology) by Amyvid-PET, as well as the acceptability of the training processes (in-person or remote web training) utilized to ensure reliable reader interpretation of Amyvid-PET scans. Avid believes that these studies address the FDAs primary concern expressed in the CRL regarding the development of a florbetapir-PET scan interpretation methodology and training process which is suitable for market implementation.

Since the previous NDA submission, new training materials have been developed, including a web-based program suitable for use in a clinical setting. The clinical development program has shown that readers can be trained to interpret florbetapir-PET images with high reliability and with high accuracy. Furthermore, based on performance, both web-based and in-person training formats are appropriate for reader training. Finally, when compared to results submitted in the original NDA, the new binary read methodology does not represent a change in overall florbetapir-PET interpretation, but a refinement allowing more reliable education of imaging physicians.

FDA Results on efficacy

The results on reader agreement and reader performance (in terms of sensitivity and specificity) obtained using the data provided by the sponsor is consistent with the results in the clinical report provided by the sponsor.

Study PT01

Primary and secondary analyses on inter-reader reliability

As shown in Table 3, the kappa is 0.81 (95% CI: 0.75, 0.87) for the pre-specified primary analysis population, 0.83 (0.78, 0.88) for the 151 subjects without repeats, 0.75 (0.67, 0.83) for the 59 autopsy subjects, and 0.88 (0.82, 0.94) for the 92 non-autopsy subjects. The kappa values and the lower bound of confidence intervals are all greater than 0.64. The kappa value for the autopsy subjects is lower than that for the non-autopsy subjects.

Among the five readers, it is possible to observe all five readers agree, four readers agree, and three readers agree with each other. The kappa statistics are consistent with the percent of agreement in this case. Higher kappa value, higher percent of agreement among five readers is observed (Table 3).

Note that for the 33 repeated scans, only the first reads were used for evaluating the inter-reader reliability.

Table 3**:** Fleiss' kappa with 95% confidence interval (CI) and reader agreement

Subjects and scans	Kappa	Percent agreement (in %) (sum in row as 100%)		
		3 readers agree	4 readers agree	5 readers agree
Primary				
119 unique scans in Batches 1-4	0.81 (0.75, 0.87)	8	11	81
Secondary				
151 subjects and scans (33 repeated with 1st scan)	0.83 (0.78, 0.88)	7	11	83
59 autopsy subjects (A07)	0.75 (0.67, 0.83)	14	10	76
92 non-autopsy scans (A05)	0.88 (0.82, 0.94)	2	11	87

Secondary analysis----sensitivity and specificity

As shown in Table 4, the point estimates of sensitivity and specificity are all greater than 65% for the five readers for the 59 subjects and all the lower bound of the confidence intervals are greater than 50%. Reader 3 has a little lower sensitivity with lower bound of 95% confidence interval as 52.4%. The sensitivity and specificity for the 46 subjects died within one year is higher than those for all the subjects died within two years. However, because of other factors, we can not state that the performance is always better in shorter duration.

Table 4: Sensitivity and specificity (in %) with 95% CI for the 59 subjects with autopsy from A05 study and 46 subjects died within one year

readers	59 subjects (39 positive and 20 negative)		46 subjects (28 positive and 18 negative)	
	Sensitivity %	Specificity %	Sensitivity %	Specificity %
1	79 (64, 91)	90 (68, 99)	86 (67, 96)	89 (65, 99)
2	92 (79, 98)	90 (68, 99)	100 (88, 100)	89 (65, 99)
3	69 (52, 83)	95 (75, 100)	75 (55, 89)	94 (73, 100)
4	87 (73, 96)	95 (75, 100)	93 (77, 99)	94 (73, 100)
5	82 (67, 93)	95 (75, 100)	89 (72, 98)	100 (82, 100)

Secondary analysis----percent of positive reading by baseline clinical diagnosis (151 subjects)

There is about 70-85% subjects with positive reading results for the 49 AD subjects, about 30% subjects with positive reading results for the 57 MCI subjects (Table 5). More healthy subjects without autopsy had positive reads compared with those with autopsy. The percentages are consistent among the five readers for AD, MCI, and healthy subjects. The percent of positive reads varied a lot among the five readers for ODD subjects (ranges from 46% to 77%).

Table 5: # Percent of positive reading by baseline clinical diagnosis (151 subjects)

Baseline diagnosis	Reader 1	Reader 2	Reader 3	Reader 4	Reader 5
AD (49)	76	86	65	80	76
MCI (57)	28	32	28	32	30
HC (20) without autopsy	20	20	15	20	15
NC (12) with autopsy	8	8	0	8	8
ODD (13)	54	69	46	77	54

Secondary analyses---evaluate intra-reader agreement

The inter-reader agreement measures among the five readers are similar for the 1st and the 2nd read (Table 6). Percent of agreement for the two repeated reads is high for all subjects (Table 7).

Table 6: Agreement of the five readers for 33 subjects with repeated readings

Subjects	Kappa	Percent agreement (%)(sum in row as 100%)		
		3 readers agree	4 readers agree	5 readers agree
33 repeats (1st)	0.89 (0.78, 0.99)	3	9	88
33 repeats (2nd)	0.86 (0.75, 0.97)	3	12	85

Table 7: Percent of agreement for the repeated reads

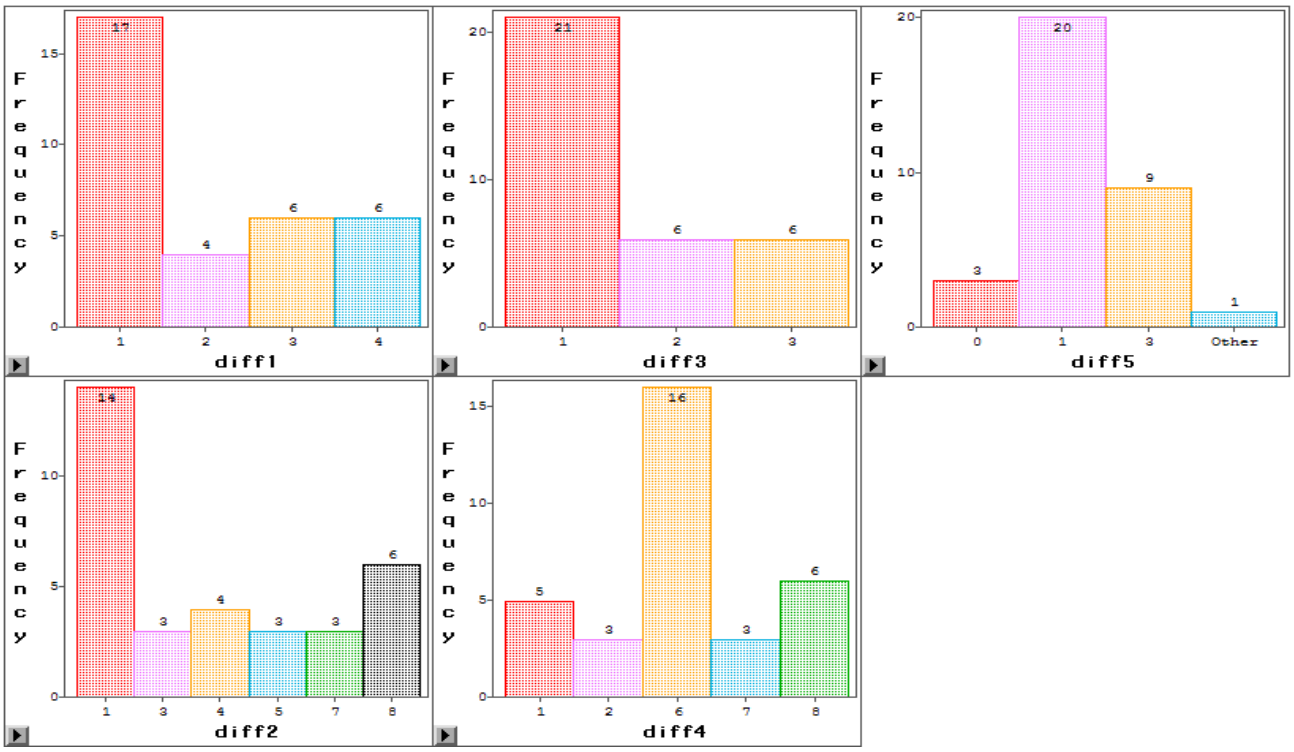
	N	Reader 1	Reader 2	Reader 3	Reader 4	Reader 5
All scans with repeats	33	94	100	91	97	97
AD	7	100	100	71	100	86
HC	6	83	100	100	83	100
MCI	20	95	100	95	100	100

Secondary analyses----read time lag for repeated reads

Figure 1 shows the frequency of the read time lag for the repeated reads (time between the 1st read and the 2nd read for the same scan). Diff1 is for reader 1, diff2 is for reader2, diff3 is for reader 3, diff4 is for reader 4 and diff5 is for reader 5. For readers 1, 2, 3, and 5, more cases had one day lag time. Maximum lag time is 8 days. Reader 4 had most of the lag time as 6 days. For reader 5, there are 3 scans with lag time 0 days and therefore are not useful

The results for the intra-reader agreement are questionable, especially for reader 5.

Figure 1: Frequency of read time lag for repeated reads (33 scans)



Secondary analyses ---- logistic regression

Logistic regression models were constructed to evaluate the effect of covariates (such as age (<65 vs. >=65), race, sex, baseline diagnosis, and other) on the percent of majority readers agree with each other.

The response is defined as 0 when 4 or 5 readers agree and 1 when 3 readers agree.

Age, sex are not significant and models with baseline diagnosis and race do not have valid fit because of small sample size in some categories.

Study A16

The sensitivity and specificity values are all greater than 65% and all the lower bound of the confidence intervals are greater than 50%. The reader performance in terms of sensitivity and specificity is similar for the 59 subjects and the 46 subjects.

The sensitivity and specificity values from Study A16 (Table 8) is slightly higher than those from Study PT01 (Table 4).

Table 8: Sensitivity and specificity (in %) with 95% CI for the 59 subjects with autopsy from A05 study and 46 subjects died within one year

Readers	59 subjects (39 positive and 20 negative)		46 subjects (28 positive and 18 negative)	
	Sensitivity %	Specificity %	Sensitivity %	Specificity %
5	92 (79, 98)	95 (75, 100)	96 (81, 100)	94 (72, 100)
6	95 (83, 99)	95 (75, 100)	97 (83, 100)	94 (72, 100)
7	87 (72, 96)	95 (75, 100)	89 (71, 98)	100 (81, 100)
8	92 (79, 98)	100 (83, 100)	96 (81, 100)	100 (81, 100)
9	69 (52, 83)	90 (68, 99)	75 (55, 89)	89 (65, 99)

The Fleiss' kappa is 0.75 (95% CI: 0.67, 0.83) for the 59 subjects and 0.76 (95% CI: 0.67, 0.85) for the subgroup of 46 subjects died within one year.

Percent that five readers agree with each other is 73% for the 59 subjects and 74% for the 46 subjects, which are consistent with results shown in kappa.

3.3. Evaluation of Safety

There is no major safety issue for this product (For more details please see clinical review).

4. FINDINGS IN SPECIFAL/SUBGROUP POPUATLIONS

Study PT01:

As shown in Table 9, reader agreement is the highest among the five readers for the 57 MCI subjects (0.91 kappa and 91% percent of five readers agree with each other). The agreement for AD subjects is moderate and low for the 13 ODD subjects.

The agreement is higher for scans from study A05 including subjects without autopsy compared with those from study A07 including all autopsy subjects (Table 9 and 10). The healthy subjects had kappa 0.83 for cases without autopsy and 0.73 for cases with autopsy. The AD subjects had kappa 0.71 for cases without autopsy and 0.56 for cases with autopsy. The MCI subjects had kappa 0.93 for cases without autopsy and 0.18 for cases with autopsy.

Table 9: Agreement by baseline clinical diagnosis (151 subjects)

Subjects and scans	Kappa	Percent agreement (%) (sum in row as 100%)		
		3 readers agree	4 readers agree	5 readers agree
49 AD in 151 subjects (scans from A05, A07)	0.67 (0.58, 0.76)	10	14	76
20 HC (only from A05 without autopsy)	0.83 (0.69, 0.97)	5	5	90
57 MCI (52 scans from A05, 5 scans from A07)	0.91 (0.83, 0.99)	2	8	91
12 NC (only from A07 with autopsy)	0.73 (0.55, 0.87)	0	8	92
13 ODD (only from A07 with autopsy)	0.52 (0.35, 0.69)	23	23	54

Note: subjects from A07 had autopsy, and subjects from A05 did not have autopsy.

Table 10: Agreement in AD and MCI subjects using scans from Study A05 and A07

Subjects and scans	Kappa	Percent agreement (%)(sum in row as 100%)		
		3 readers agree	4 readers agree	5 readers agree
29 AD from A07	0.59 (0.47, 0.71)	14	7	79
20 AD from A05	0.71 (0.57, 0.85)	5	25	70
5 MCI from A07 (sample size small)	0.18 (-0.1, 0.46)	20	0	80
52 MCI from A05	0.93 (0.84, 1)	0	8	92

As shown in Table 11, the reader agreement is higher in younger subjects (age<65) compared with those older ones (age>=65). The agreement is similar for males and females, Hispanics and non-Hispanics.

Table 11: Agreement by demographics (151 subjects)

	N	kappa	Percent agreement (%)(sum in row as 100%)		
			3 readers agree	4 readers agree	5 readers agree
Age					
<65	34	0.92 (0.81, 1)	3	3	94
>=65	117	0.81 (0.75, 0.87)	8	13	79
Sex					
Male	76	0.81 (0.76, 0.86)	5	14	80
Female	75	0.85 (0.78, 0.92)	8	7	85
Ethnic					
Hisp or Lat	20	0.84 (0.78, 0.88)	5	10	85
Non-Hisp or Lat	131	0.83 (0.78, 0.92)	7	11	82
Race					
Caucasian	141	0.83 (0.78, 0.88)	7	11	82
Black or A -A	7	1 (0.76, 1)	0	0	100
Other	3	0.72 (0.36, 1)	0	33	67

In addition to reader agreement, sensitivity and specificity of image read versus SoT were evaluated for the 59 subjects with autopsy. A shown in Table 12, the sensitivity and specificity values for the 29 AD cases and the 12 NC cases are very high. The sensitivity and specificity values for the 13 ODD cases are low. Due to the small sample size (5), it is difficult to evaluate the sensitivity and specificity for MCI subjects. Younger subjects have higher sensitivity and specificity values than older ones.

Table 12: Sensitivity/specificity by subgroups (59 subjects with autopsy)

	Reader 1	2	3	4	5
Baseline diagnosis					
AD (29)	86/100	96/100	79/100	86/100	89/100
MCI (5)	0/75	0/75	0/100	0/100	0/100
NC (12)	100/100	100/100	0/100	100/100	100/100
ODD (13)	67/75	89/75	56/75	100/75	67/75
Age					
<65 (9)	75/100	75/100	75/100	75/100	75/80
>=65 (50)	80/87	94/87	69/93	89/93	83/100

Note: Confidence intervals are not calculated because of the small sample size in subgroups.

Confidence (low, median and high) were recorded for each reader. The median percent of cases with high confidence among the five readers is about 71% for autopsy cases and 78% for non-autopsy cases (Table 13). Very low percent of cases with high confidence is observed in ODD subjects and very high percent of cases with high confidence is observed in healthy cases and MCI cases.

Younger subjects have higher percentage of cases with high confidence.

Range reflects the variation of the percent among the five readers. Higher variation in terms of confidence is observed in autopsy subjects compared with non-autopsy subjects.

Table 13: Median and range of the percent of subjects with high confidence among the five readers (in %) by subgroups

	High confidence in percent: median (range)		
N in different groups	151	59	92
Confidence in %			
All	74(23)	71(32)	78(17)
Baseline diagnosis			
AD	70(37)	72(48)	65(20)
MCI	81(19)	80(20)	83(23)
HC	95(30)		95(30)
NC	83(33)	83(33)	
ODD	54(15)	54(15)	
Age			
<65	88(26)	89(56)	88(24)
>=65	73(22)	68(30)	76(16)

Note: range = max percent – min percent among the five readers

Study A16

As shown in Table 14, the sensitivity and specificity values for all the 59 subjects and the subgroups by baseline diagnosis are all greater than 80% except the 5 MCI cases and the 13 ODD cases. The results in terms of sensitivity and specificity for Study A16 are higher than those for Study PT01.

Table 14: Sensitivity/Specificity (in %) of the five readers and the 95% confidence intervals (CI) for Study A16

	Reader 5	Reader 6	Reader 7	Reader 8	Reader 9
59 total (39 pos and 20 neg)	92 (79, 98)/ 95 (75, 100)	95 (83, 99) / 95 (75, 100)	87 (72, 96)/ 95 (75, 100)	92 (79, 98)/ 100 (83, 100)	69 (52, 83)/ 90(68, 99)
29 AD	93/ 100	96/ 100	89/ 100	93/ 100	82/ 100
5 MCI	0/ 100	0/ 100	0/ 100	0/ 100	0/ 100
12 NC	100/ 91	100/ 100	100/ 100	100/ 100	100/ 90
13 ODD	100/ 100	100/ 75	89/ 75	100/ 100	44/ 75

Note: confidence intervals are not shown for diagnosis groups because of the small sample size.

The kappa values and percent agreement are good for the 59 subjects and 46 subjects (Table 15). The agreement is low for the ODD cases (kappa as 0.43 and percent of agreement for five reader agree with each other as 38%), which is consistent with the result obtained in PT01 study. Agreement is higher for younger subjects compared with older subjects. Overall, the agreement for all autopsy subjects and for the subgroups by baseline diagnosis is lower in A16 study compared with PT01 study.

Table 15: Inter-reader agreement in terms of Fleiss' kappa and percent agreement (A16 study)

Subjects and scans	General kappa	Percent agreement (%)(sum in row as 100%)		
		3 agree	4 agree	5 agree
59 subjects	0.75 (0.67, 0.83)	5	22	73
46 subjects (death within one year)	0.76(0.67, 0.85)	7	20	74
Baseline diagnosis				
29 AD	0.62 (0.50, 0.74)	7	10	83
5 MCI	Can not calculate	0	0	100 (all read as negative)
13 NC	0.44(0.26, 0.62)	0	25	75
13 ODD	0.43 (0.26, 0.60)	8	54	38
Age				
<65	0.81 (0.60, 1)	0	22	78
>=65	0.73 (0.64, 0.82)	6	22	72

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

The sample size for MCI subjects (targeted patient population) with autopsy is very small (five cases). Therefore, for MCI subjects, we can not conduct formal statistical evaluation of the reader performance in terms of sensitivity and specificity of image reads versus SoT obtained from autopsy. Only summary statistics are provided and we can not obtain conclusive evidence for the performance in terms of sensitivity and specificity on MCI population.

Only reader agreement (inter-reader and intra-reader) is evaluated using Fleiss' kappa and percent agreement for the patients with and without autopsy (with baseline diagnosis: AD, MCI, HC, and ODD). Even though the reader agreement is high for MCI subjects, it does not always indicate high sensitivity and specificity for MCI subjects.

Training program was developed using the scans from A05 and A07 studies. The same scans were used to evaluate the reader performance of the readers using the training program. The optimal case should be using one set of images (scans) for training program development and using another set of images for evaluating reader performance.

5.2 Conclusions and Recommendations

Inter-reader agreement among multiple readers is good (kappa as 0.83 (95% CI: 0.78, 0.88) for the 151 subjects without repeated reads). And the percent of agreement for five readers agree with each other are greater than 75% in different patient populations (all 151 subjects, autopsy subjects, and non-autopsy subjects). The results on kappa statistics are consistent with percent of agreement.

Sensitivity and specificity to the subjects with autopsy are consistently good for the five readers. From study PT01, the sensitivity and specificity (%) for the 151 subjects including 59 autopsy and 92 non-autopsy subjects are 79 (95% CI: 63.5, 90.7) and 90 (95% CI: 68.3, 98.8) for reader 1, 92 (79.1, 98.4) and 90 (68.3, 98.8) for reader 2, 69 (52.4, 83.0) and 95 (75.1, 99.9) for reader 3, 87 (72.6, 95.7) and 95 (75.1, 99.9) for reader 4, and 82 (66.5, 92.5) and 95 (75.1, 99.9) for reader 5. From Study A16, the sensitivity and specificity (%) of the binary reads for the 59 autopsy subjects from A05 study are 92 (95%CI: 79, 98) and 95 (95% CI: 75, 100) for reader 5, 95 (95%CI: 83, 99) and 95 (75, 100) for reader 6, 87 (72, 96) and 95 (75, 100) for reader 7, 92 (79, 98) and 100 (83, 100) for reader 8, and 69 (52, 83) and 90(68, 99) for reader 9.

Sensitivity and specificity values are higher for the 59 subjects in A16 study compared with PT01 study suggesting in-person training may be better than web-based training.

Inter-reader agreement on A05 patients without autopsy is higher than A07 patients with autopsy. Subgroup analyses on agreement indicate that readers have high agreement on MCI patients, but lower agreement on ODD patients.

Intra-reader agreement is good for the readings on the 33 subjects with repeats. The time lag for the 1st reading and the 2nd reading ranges from 1 to 8 days. However, one reader had scans read within one day.

Primary Statistical Reviewer:	Lan Huang, Ph.D.
Date:	March 10, 2012
Concurring Reviewer(s):	
Statistical Team Leader:	Jyoti Zalkikar, Ph.D.
Biometrics Division Director:	Rajeshwari Sridhara, Ph.D.

Cc:
CDTL: **Lucie Yang**
Clinical: **Qi Feng**
Non-Clinical: **Sunday Awe**
CMC: **Ravindra Kasliwal**
Statistics: **Lan Huang**
Clin Pharm: **Christy John**
Microbiology: **Stephen Langille**
Safety Team: **Denise Baugh/ Sandra Griffith**
Jyoti Zalkikar
Rajeshwari Sridahara
Lillian Patrician

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

/s/

LAN HUANG
03/25/2012

JYOTI ZALKIKAR
03/25/2012
Concur with the reviewer.

RAJESHWARI SRIDHARA
03/26/2012

U.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/BLA Serial Number:	202008
Drug Name:	Amyvid
Indication(s):	Amyvid is indicated for Positron Emission Tomography (PET) imaging of the brain to detect β-amyloid neuritic plaques in adult patients with cognitive impairment.
Applicant:	Avid Pharmaceuticals
Date(s):	Primary Review: March 10, 2012 Secondary Review: March 14, 2012 PDUFA date : April 7, 2012
Review Priority:	Re-submission 6 months
Biometrics Division:	Division of Biometrics V
Statistical Reviewer:	Lan Huang, Ph.D.
Concurring Reviewers:	Jyoti Zalkikar, Ph.D. Rajeshwari Sridhara, Ph.D.
Medical Division:	Division of Medical Imaging Products
Clinical Team:	Qi Feng, M.D. Lucie Yang, M.D Dwaine Rieves, M.D.
Project Manager:	Sharon Thomas

Keywords: Fleiss' Kappa, Agreement, Sensitivity, Specificity, Patients with autopsy

EXECUTIVE SUMMARY

(b) (6)

Amyvid is indicated for Positron Emission Tomography (PET) imaging of the brain to detect β-amyloid neuritic plaques in adult patients with cognitive impairment.

A negative scan indicates sparse to no plaques and is inconsistent with a neuropathological diagnosis of Alzheimers Disease (AD) at the time of image acquisition. Obtaining a negative Amyvid scan in a patient with cognitive impairment may assist in the detection of non-AD causes of cognitive impairment. A positive scan indicates more than a sparse amount of plaque in the brain and has been observed in older patients with normal cognition as well as patients with cognitive disorders due to AD or other neurologic conditions.

The Amyvid assessment of brain amyloid plaque density is based upon correlation of images with brain tissue microscopic slides scored for plaque density using modification of criteria developed by the Consortium to Establish a Registry for Alzheimer's Disease (1997).

Amyvid efficacy has not been established for its use an AD diagnostic test, for prediction of the development of dementia or for monitoring the response to therapies.

Reader training is essential to minimize the risk for image misinterpretation. The training may be completed by participation in electronic media-based tutorials as well as in-person tutorials. Even among well trained readers, erroneous image interpretations have been reported. In clinical studies, 5% to 31% of patients had images that readers incorrectly interpreted when the image result was compared to a postmortem brain amyloid truth standard. An extended duration of time between the Amyvid image acquisition (> one year) and subsequent patient death may have contributed to the apparent errors although some errors were detected even when the time duration was shorter (< one year). Amyvid scan results are indicative of the brain amyloid plaque content only at the time of image acquisition and a negative scan result does not preclude the development of brain amyloid in the future.

We evaluated three clinical studies that examined Amyvid images from healthy adult subjects as well as subjects with a range of cognitive disorders, including some terminally ill patients who had agreed to participate in a postmortem brain donation program. All these studies were single arm studies in which subjects underwent an Amyvid injection and scan and then had images interpreted by multiple independent readers who were blinded to all clinical information. Image interpretations used co-registration with computed tomography (CT) scans when PET scans were performed on dual PET-CT scanners. Study A07 was submitted to FDA and reviewed in earlier submission of this NDA.

In Study A07, a semi-quantitative Amyvid image interpretation method, which is not intended for clinical use, was used by three readers to interpret images from 152 terminally ill patients of which 35 underwent autopsy within one year following PET imaging procedure. The first 6 were part of a predefined front-runner study group. The subsequent 29 subjects comprised the primary autopsy analysis population. The median patient age of this group was 85 years (range 55 to 103 years) and 14 of the patients were female. Eighteen patients had dementia, 9 had no cognitive impairment and 2 had mild cognitive impairment (MCI). The main study outcome was a comparison of premortem Amyvid images to the findings from a postmortem brain examination (truth standard). The semi-quantitative measure consisted of a five-point scale of whole brain Amyvid uptake that was compared to a global score of the percentage of the whole brain that contained amyloid, as determined by immunohistochemical microscopy. The percentage of postmortem cortical amyloid burden ranged from 0 to 7% and correlated with the median Amyvid scores (Spearman's rho=0.78; $p<0.0001$, 95% CI, 0.58 to 0.89).

Study A16 and study PT01 used a clinically-applicable binary image interpretation method (positive/negative) to evaluate images from a range of patients who had participated in earlier studies. These studies assessed performance characteristics (sensitivity and specificity) among subjects with a postmortem amyloid plaque density truth standard. Additionally, inter-reader and intra-reader image interpretation reproducibility was assessed among

all the subjects, including subjects who lacked a postmortem truth standard. Before image interpretation, all readers underwent special training; Study A16 used an in-person tutoring type of training and Study PT01 used an electronic media-based training method. Five trained readers interpreted images independently within each study. For purposes of correlating Amyvid image results to the whole brain amyloid plaque density, Amyvid results (negative/positive) were pre-specified to correlate with specific plaque density scores, based upon a modification of criteria developed by the Consortium to Establish a Registry for Alzheimer's Disease (CERAD).

Table 1: Plaque Density* Correlates to AMYViD Image Results

Neuritic Plaque Counts	CERAD Score	AMYViD Image Result
< 1	none	Negative
1 - 5	sparse	
6 - 19	moderate	Positive
20 +	frequent	

* J of Neuropathology and Experimental Neurology 1997; 56(10):1095.

Study A16 examined images only from terminally ill patients who had premortem Amyvid scans and postmortem brain examinations to determine a truth standard (including the 35 autopsy subjects in Study A07). Among the 59 patients, the median age was 83 years (range 47 to 103 years), half were females and most were Caucasian (93%). Twenty-nine patients had an AD clinical diagnosis, 13 had another type of dementing disorder, 12 had no history of cognitive impairment and 5 had MCI. The time interval between the Amyvid scan and death was less than one year for 46 patients and between one and two years for 13 patients. Among the subset of patients who died within one year of Amyvid scanning, the median sensitivity among the readers was 96% (95% CI: 80% to 100%) and specificity was 100% (95% CI:78% to 100%). At autopsy, the global brain plaque density category (CERAD score as in Table 1) was: frequent n = 30; moderate n = 9; sparse n = 5; and none n = 15. Inter-reader reproducibility analyses showed a Fleiss' kappa statistic of 0.75. Tables 2 and 3 shows the Amyvid performance characteristics among all the patients.

Table 2: Amyvid Scan Results by Reader Training Method and Reader Performance among Autopsied Patients (n = 59)

Test Performance		In-Person Training (Study A16)	Electronic Media Training (Study PT01)
Sensitivity (%)	Median	92	82
	Range among the 5 readers	(69 – 95)	(69 – 92)
Specificity (%)	Median	95	95
	Range among the 5 readers	(90 – 100)	(90 – 95)

Table 3: Amyvid False Negative/False Positive Scan Results by Reader Training Method among Autopsied Patients (n = 59)

Read result	In-Person Training (Study A16)					Electronic Media Traning (Study PT01)				
	Reader					Reader				
	1	2	3	4	5	6	7	8	9	10
Correct	55	56	53	56	45	49	54	46	53	51
False Negative	3	2	5	3	12	8	3	12	5	7
False Positive	1	1	1	0	2	2	2	1	1	1

Study PT01 included images from 92 subjects who did not have a truth standard (20 healthy volunteers, 52 patients with mild cognitive impairment, 20 patients with AD) as well as all 59 of the patients who underwent an autopsy and provided a truth standard. Duplicate images of some patients were included within the total pool of images in order to assess intra-reader image reproducibility. Among the 151 subjects, the median age was 76 years (range 47 to 103 years), half were females and most were Caucasian (93.4%). Performance characteristics for patients with a truth standard are shown above (Tables 2 and 3). The median number of Amyvid positive scans, Fleiss' kappa results and distribution of agreement for baseline diagnosis are shown in Table 4. Inter-reader reproducibility analyses showed overall Fleiss' kappa statistic of 0.83 (95% CI: 0.78 to 0.88) with a lower bound of 0.78 which exceeded the pre-specified success criterion of 0.58.

Table 4: Positive Amyvid Scans within Study PT01 Subject Groups and Reproducibility of Scans Among Readers

Subject group by cognitive and truth standard (TS= autopsy) status	Positive Scans, n (%)*	Kappa (95% CI)	Percent agreement (%)**		
			3 agree	4 agree	5 agree
All subjects with a TS, n = 59	33	0.75 (0.67, 0.83)	14	10	76
All subjects without a TS, n = 92	33	0.88 (0.82, 0.94)	2	11	87
AD, n = 49 (29 with TS; 20 no TS)	38	0.67 (0.58, 0.76)	10	14	76
MCI, n = 57 (5 with TS; 52 no TS)	17	0.91 (0.83, 0.99)	2	8	91
Cognitively normal without TS, n = 20	4	0.83 (0.69, 0.97)	5	5	90
Cognitively normal with TS, n = 12	1	0.73 (0.55, 0.87)	0	8	92
Other (non-AD) dementia with TS, n = 13	7	0.52 (0.35, 0.69)	23	23	54

*shown is the median number of scans interpreted as positive across the 5 readers for each subgroup of patients listed in the first column;
**Percent agreement among the 5 readers.

This application is approvable from the statistical view point in spite of the drawbacks of Fleiss' kappa because of reasonable sensitivity and specificity in a limited population and reasonably large value of Fleiss' kappa in a varied patient population. The only cautions statistical team would like to express are that 1> Amyvid scan results are indicative of the brain amyloid content only at the time of image acquisition, 2> a negative scan result does not preclude the development of brain amyloid in the future, and 3> the extremely small (n=5) number of MCI subjects had autopsy confirmation, a limiting factor in interpreting the performance characteristics (sensitivity and specificity).

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

/s/

JYOTI ZALKIKAR
03/14/2012

RAJESHWARI SRIDHARA
03/14/2012
Concur with the reviewer

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

CLINICAL STUDIES

BLA/Serial Number:	202-008
Drug Name:	Amyvid (Florbetapir F 18 Injection)
Indication(s):	Detection of β-amyloid in the brain
Applicant:	Avid Raiopharmaceuticals, Inc.
Date(s):	Submitted date: September 17, 2010
	Received date: September 17, 2010
	AC date: January 20, 2011
	PDUFA date: March 17, 2011
Review Priority:	Priority
Biometrics Division:	Division of Biometrics V
Statistical Reviewer:	Lan Huang, Ph.D.
Concurring Reviewers:	Anthony Mucci (Team leader)
	Rajeshwari Sridhara, Ph.D. (Division director)
Medical Division:	Division of Medical Imaging Products
Clinical Team:	Qi Feng, M.D. (Clinical reviewer),
	Lucie Yang, M.D. (Clinical team leader),
	Rafel Rieves, M.D. (Clinical division director)
Project Manager:	Sharon Thomas

Keywords: Amyloid detection, open label, autopsy, reader agreement, correlation, sensitivity and specificity

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIX

1. EXECUTIVE SUMMARY

Florbetapir F 18 (formerly known as 18F-AV-45 or florpiramine F 18) is a molecular imaging agent proposed here for PET imaging of β-amyloid aggregates in the human brain. An indication for imaging of β-amyloid pathology, rather than a diagnosis of AD, is therefore sought. While the data suggest correlation between the semi-quantitative Amyvid PET image visual read results and the quantitative immunohistochemistry, because of the observed inter-reader variability, use of different rating methods, lack of data to ascertain sensitivity and specificity in the intended patient population, the data as submitted in the application does not support the proposed indication.

The clinical development program of florbetapir F 18 comprised 6 completed clinical trials involving 496 patients: 18F-AV-45-A01 (A01), 18F-AV-45-A02 (A02), 18F-AV-45-A03 (A03), 18F-AV-45-A04 (A04), 18F-AV-45-A05 (A05), and 18F-AV-45-A07 (A07). Study A07 was the pivotal trial comparing β-amyloid levels as evaluated by florbetapir-PET imaging to post-mortem amyloid levels on histopathology.

Description of the study A07

The pivotal trial, Study A07, is an open label, single arm study. It was designed to (1) determine the relationship between measurements of brain β-amyloid using florbetapir-PET imaging and true levels of β-amyloid measured post mortem (Autopsy Cohort) and to (2) demonstrate the specificity of florbetapir- PET in a cohort of individuals unlikely to have, and therefore assumed not to have, brain amyloid plaque (Specificity Cohort). The study was conducted at 34 study centers in the United States.

The study tested two hypotheses:
Primary hypothesis #1: Correlation analysis
There is a statistically significant correlation (Spearman's rho $\rho>0$) between the semi-quantitative visual rating of amyloid burden of the florbetapir-PET scan and the cortical amyloid burden at autopsy as assessed by quantitative immunohistochemistry (IHC).

Primary hypothesis #2: Specificity analysis
The observed specificity of florbetapir-PET imaging is ≥90% in young healthy controls.

A total of 226 subjects were enrolled in the study, 152 subjects in the autopsy cohort and 74 young healthy volunteer subjects in the specificity cohort.

Autopsy Cohort Results

The 152 subjects in the Autopsy Cohort were enrolled from various end-of-life (e.g., hospice/hospital/nursing home) and late-life (longitudinal studies of aging) populations and yielded 35 autopsies within 1 year following the PET imaging procedure. The first 6 were part of a predefined front-runner study group. The subsequent 29 subjects comprised the primary autopsy analysis population. Three independent imaging physicians (reader 1, 2, and 3)

evaluated the florbetapir-PET scans in randomized blinded fashion. The neuropathology analyses were independently performed and were blinded to any clinical information, image data or reading results.

The primary read was a visual semi-quantitative rating assessment performed by the three independent readers. Each autopsy-cohort reader rated the degree of florbetapir retention in the grey matter on a scale from 0 (no amyloid) to 4 (high levels of β-amyloid deposition), and the median score of the 3 readers was the primary efficacy endpoint. The primary correlation analysis produced a statistically significant Spearman's rho of 0.78($p<0.0001$, 95% CI: 0.58 - 0.89) between the semi-quantitative visual rating of Amyid PET scan and the cortical amyloid level as assessed by quantitative IHC.

Specificity Cohort Results

For the primary Specificity analysis, an additional cohort (specificity cohort) of 47 young (age < 40), cognitively and neurologically healthy individuals who were not ApoE ε4 allele carriers (thus could be expected with high confidence to be devoid of brain amyloid) had their scans randomly mixed with 40 scans rated positive (median rating of 2, 3 or 4) from the autopsy cohort, and these were read for binary outcomes (amyloid positive, amyloid negative) by three additional independent imaging physicians (reader 4, 5, and 6). The primary specificity analysis focused on the majority qualitative read for the 47 controls. For the primary specificity analysis, 100% (47/47) of young healthy subjects were rated as amyloid negative on the florbetapir-PET scan by the median read, with readers 4 and 6 agreeing on all 47 cases as negative, and reader 5 scoring negative on 46 of 47 cases.

Statistical issues

The data from pivotal trial A07 provide statistically significant evidence that median semi-quantitative Florbetapir F 18 image reads of amyloid burden are highly correlated with pathological read of amyloid burden. However, the data do not produce evidence of clinical usefulness of the Amyloid detection by Amyvid since its performance characteristics (sensitivity and specificity) show considerable inconsistency among the readers for the patients from various end-of-life populations. The median read proposed by the sponsor masks the individual reader performance. It is not clear if this reader-to-reader variability will increase or decrease in the intended patient population. Also with the small sample of 29 subjects in the primary efficacy population, the confidence intervals are very wide for the sensitivity and specificity values.

The inter-reader agreement is evaluated using studies A05 and A07 data, but not sensitivity or specificity, since most subjects did not have autopsy results. The exploration showed that the reader agreement varied by studies, age groups, and diagnosis at baseline. So the specificity results in A07, although consistent across readers, are obtained from the population of young healthy volunteers and it is not clear if these results will be upheld in the intended patient population.

Moreover, the sponsor has proposed using binary, qualitative read of Florbetapir F 18 images which has been applied only to 14 patients in whom pathological standard of truth was available. This sample size is too small to assess the clinical usefulness of the proposed qualitative read.

2. INTRODUCTION

2.1. Overview

2.1.1. Indication

Florbetapir F 18 (Amyvid) is a diagnostic radiopharmaceutical for use with positron emission tomography. It is first in the class of β-amyloid PET imaging agents to be submitted under a New Drug Application. The Sponsor's proposed indication is:

Florbetapir F 18 is indicated for PET imaging of β amyloid aggregates in the brain. A negative florbetapir-PET scan is clinically useful in ruling out the presence of β-amyloid, a defining pathology of Alzheimer's disease (AD).

2.1.2. History of Program Development

Florbetapir F 18 (Amyvid) is a molecular imaging agent designed for PET imaging of β-amyloid aggregates in the human brain. An indication for imaging of β-amyloid pathology, rather than a diagnosis of AD, is therefore sought. The Food and Drug Administration's (FDA's) Peripheral and Central Nervous System Drugs Advisory Committee (AC) Meeting of October 23, 2008 brought consensus for this approach. Furthermore, the AC recommended histopathology following autopsy as a reference for the standard of truth to evaluate the performance characteristics for the detection of amyloid pathology by PET imaging. This recommendation was reiterated by the Agency in a follow-up teleconference between the Division of Medical Imaging and Hematology Products and Avid Radiopharmaceuticals, Inc on November 3, 2008 and in a Type C meeting on February 11, 2009. Therefore, the focus of the florbetapir F 18 development program and this New Drug Application (NDA) was to establish the relationship between β-amyloid levels, as evidenced on the PET image, and the underlying true β-amyloid levels determined by postmortem histopathology.

The clinical development program of florbetapir F 18 comprised 6 completed clinical trials involving 496 patients: 18F-AV-45-A01 (A01), 18F-AV-45-A02 (A02), 18F-AV-45-A03 (A03), 18F-AV-45-A04 (A04), 18F-AV-45-A05 (A05), and 18F-AV-45-A07 (A07). A pooled, blinded read analysis of 2 image acquisition time points from Studies A01 and A03 was conducted and is reported as study report A06. Studies A01, A02, A03 and A04 were Phase I studies examining brain uptake and retention (β-amyloid binding) in AD patients and controls (A01, A03, A04), whole body biodistribution and dosimetry (A01, A03), pharmacokinetics and metabolism (A01, A03), dose response (A03) and test-retest reliability (A04). Study A05 was a Phase II study comparing β-amyloid binding in AD patients, MCI subjects and cognitively healthy subjects across the age range of 50 to more than 80. Study A07 was the pivotal trial comparing β-amyloid levels as evaluated by florbetapir-PET imaging to post-mortem amyloid levels on histopathology.

On 9/30/2010, the priority review designation for the NDA 202-008 was communicated to Sponsor in writing.

2.1.3. Specific Studies Reviewed

This statistical review is focused on the single pivotal study 180-F-AV-45-A07 which included two study cohorts: Autopsy cohort and Specificity cohort. However, this review includes other studies for exploration:
(1): Study A05 is explored for investigating inter-reader agreement on the binary rating method. The rating results are compared with those from A07 study.
(2): Study 18-F-AV-45-A04 is explored to evaluate test-retest reliability.
(3):Studies A01, A05 and A07 are explored for understanding the distribution of the quantitative PET scan results (SUVR) among subjects with full range of age included in the program.

A brief description of these studies is presented below:
Study A07 is a phase III study, open label, single arm study of correlation between florbetapir F 18 (Amyvid) positron emission tomography imaging and amyloid pathology. The study was conducted at 34 study centers in the United States, 25 of which enrolled at least 1 subject. A total of 226 subjects were enrolled in the study, 152 subjects in the autopsy cohort and 74 young healthy volunteer subjects in the specificity cohort. There are 35 observed death in the autopsy cohort (6 frontrunners and 29 subjects in the primary efficacy population) and 47 subjects without ApoE 4 carrier (a known risk factor for AD) in the specificity cohort.

Study A05 is a phase II, open label, parallel group, multicenter study for evaluating the safety and imaging characteristics of 18-F-AV-45 in healthy volunteers, patients with mild cognitive impairment (MCI) and patients with Alzheimer's disease (AD). The study was conducted at 24 study centers. A total of 184 subjects (45 subjects with AD, 60 subjects with MCI, and 79 cognitively normal subjects) were enrolled in the study.

Study A04 is a phase I, open label study for investigating the Test-retest reproducibility of 18-F-AV-45 (Amyvid) for brain imaging of amyloid in healthy volunteers and Alzheimer's disease patients. The study was conducted at 4 study centers. A total of 25 subjects were enrolled: 21 subjects (11 subjects with AD and 10 healthy volunteers) were enrolled in the primary test-retest phase of the study, and additional 4 AD subjects were enrolled and dosed using a revised injection protocol to evaluate the effect of slow vs. fast IV injection.

Study A01 is a phase I, open label study for a preliminary evaluation of the amyloid binding properties, pharmacokinetics and safety of 18-F-AV-45 in healthy elderly volunteers and patients with Alzheimer's disease. The study was conducted at 3 study centers. Sixteen (16) healthy volunteers and 16 subjects with AD were enrolled.

A summary of the above studies is included in the following table (AD for Alzheimer's disease, MCI for Mild cognitive impairment, ODD for other dementia disorder and HC for healthy control).

Table 1: Summary of studies explored in this review

	A07	A05	A04	A01
Phase	III	II	I	I
Objective	Image-amyloid correlation and "specificity"	Amyloid risk factors	Test-retest	PK
Subjects	152 end of life (AD, MCI, ODD, and HC) for correlation and 74 HC	45 AD, 60 MCI, and 79 HC	15 AD and 10 HC	16 AD, and 16 HC
Reader	3 independent	3 independent	1	NA
Rating method for primary analyses	Semi-quantitative for correlation and binary for "specificity"	Binary rating	Binary rating	NA
SUVR	Available	Available	Available	Available

Both A07 and A05 studies have three independent readers. Study A04 has one reader to read the images twice in different time points after one PET scan. Two methods for reading the images are used: binary qualitative method and semi-quantitative method. For binary qualitative method, readers will evaluate the whole brain and provide results as positive or negative (positive for with amyloid deposits and negative for no amyloid deposits in brain). For semi-quantitative method, the readers provide results as 0 (None), 1 (low), 2 (low to moderate), 3(moderate to high) and 4(high) in terms of the level of amyloid deposits in brain shown from the images. SUVR (standardized uptake value ratio) is the quantitative measure of amloid level in brain (more details on defining SUVR are included in the medical review of this application).

Note that A07 study enrolled 152 "end of life" subjects (life expectancy less than six month or less) without criterion on the degree of cognitive impairment for the autopsy cohort and 74 healthy young subjects for the specificity cohort. The population is not the intended patient population (with some degree of cognitive impairment and middle or old age). Note all subjects in A07 study have rating results using binary read method proposed for clinical use. However, A05 study enrolled 183 subjects with various levels of cognitive impairment levels (AD, MCI and HC) and all images are evaluated with the binary qualitative rating method for clinical practice. Therefore, A05 is another important study even though it is a phase II study.

2.1.4. Major Statistical Issues

The data from pivotal trial A07 provide statistically significant evidence that median semi-quantitaive Florbetapir F 18 image reads of amyloid burden are highly correlated with pathological read of amyloid burden. However, the data do not produce evidence of clinical

usefulness of the Amyloid detection by Amyvid since the sponsor has proposed using binary, qualitative read of Florbetapir F 18 images which has been applied only to 14 patients in whom pathological standard of truth was available. This sample size is too small to assess the clinical usefulness of the proposed qualitative read. We can not obtain the autopsy based performance characteristics (sensitivity and specificity measures) using the available data from A07.

The inter-reader agreement is evaluated using A05 and A07 data, not sensitivity or specificity, since most subjects did not have autopsy results. The exploration showed that the reader agreement on the binary, qualitative rating results varied by studies, age groups, and diagnosis at baseline. Also, the specificity results in A07, although consistent across readers, are obtained from the population of young healthy volunteers and it is not clear if these results will be upheld in the intended patient population.

For the 35 subjects with autopsy, the Pearson correlation between SUVR and IHC is 0.73 ($p<0.0001$). However, the Pearson correlation between SUVR and IHC is 0 for subjects with SUVR <=1.1 and 0.14 for subjects with SUVR > 1.1. These discrepant results render the correlation evaluation as inadequate to assess the usefulness of quantitative reads of Florbetapir F 18 images.

2.2 Data Sources

All materials reviewed including the applicant study reports, data sets and literature referenced are provided electronically, and the full electronic path of the document is \\Cdsesub1\evsprod\NDA202008\.

The application study reports reviewed include Clinical overview, Summary of Clinical Efficacy, Summary of Clinical Safety in M2.

Data sets analyzed for study A07 (with data definition document) were located in M5 \\Cdsesub1\evsprod\NDA202008\0000\m5\datasets\av45a07\analysis\datasets.

Data sets analyzed for study A04 (with data definition document) were located in M5 \Cdsesub1\evsprod\NDA202008\0000\m5\datasets\av45a04\analysis\datasets.

Data sets analyzed for study A05 (with data definition document) were located in M5 \Cdsesub1\evsprod\NDA202008\0000\m5\datasets\av45a05\analysis\datasets.

Partial data sets analyzed for study A01 (for SUVR analysis) were located in M1 \\Cdsesub1\evsprod\NDA202008\0008\m1\us\111-info-amend\1113-eff-info-amend.

The datasets analyzed include ADVS (Vital Signs), ADIH (Immunohistochemistry Results), ADEFF (Efficacy FDA requested), ADQA (Blinded Reads), ADSL (Subject Disposition), ADSV (Standardized Uptake Values), ADAE (Adverse Events).

3. STATISTICAL EVALUATION

3.1. Data and Analysis Quality

It was possible to reproduce the primary analyses from tabulation. Related SAS programs were provided to Agency after the request.

3.2. Evaluation of Efficacy

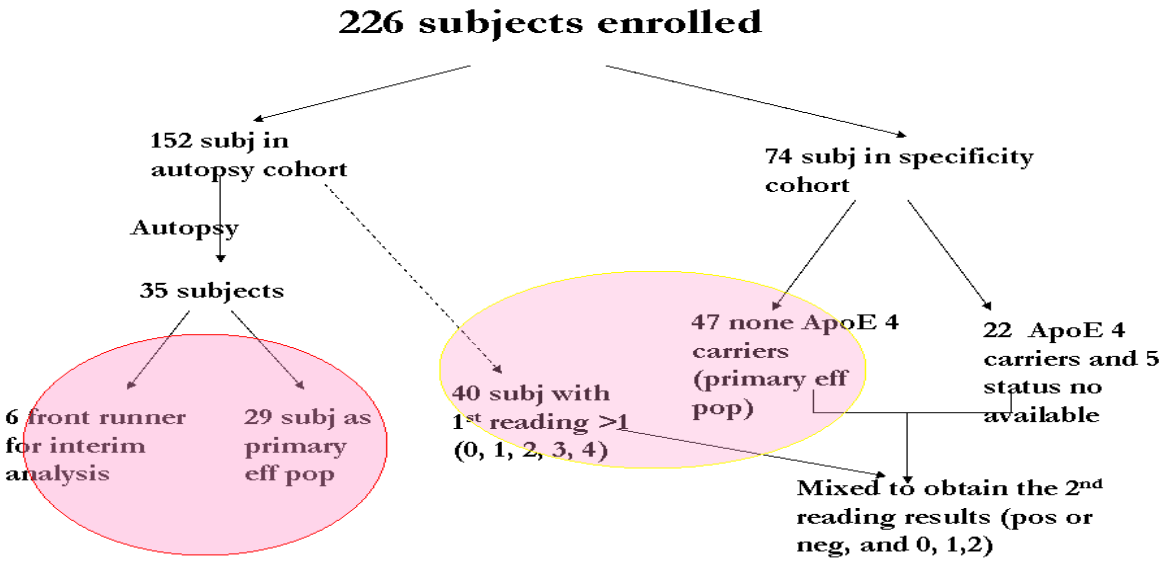
3.2.1. Study Design and Endpoints

In this and the following sections, the reviewer's comments are in italics.

Phase III study A07:

Study A07 is a phase III study, open label, single arm study of correlation between florbetapir F 18 (Amyvid) positron emission tomography imaging and amyloid pathology. The study was conducted at 34 study centers in the United States, 25 of which enrolled at least 1 subject. A simple flow chart is used to describe the study design (in Figure 1).

Figure 1: A07 study design and patients distribution (ApoE 4 is a possible risk factor for AD, therefore, in order to avoid confounding, only 47 non-ApoE 4 carriers are included as the primary efficacy population in the specificity cohort)



A total of 226 subjects were enrolled in the study, 152 subjects in the autopsy cohort and 74 young healthy volunteer subjects in the specificity cohort.

For autopsy cohort, male or female subjects, ≥18 years of age and with projected life expectancy of ≤6 months, from various end-of-life (e.g., hospice/hospital/nursing home) and late-life (longitudinal studies of aging) and could tolerate a 10-minute PET scan were eligible to enroll in the autopsy cohort. An effort was made to enroll subjects with various levels of cognitive status, ranging from cognitively normal through dementia.

In order to evaluate the specificity of florbetapir-PET scans for detecting the absence of amyloid, an additional separate cohort of 74 young cognitively and neurologically healthy individuals was enrolled for imaging only. For specificity cohort, cognitively and neurologically healthy male and female subjects, 18 to 40 years of age, who had no known risk factors for Alzheimer's disease (AD), including genetic risk factors for AD, such as apolipoprotein (Apo) E 4, performed in an age-appropriate normal range on the Wechsler Logical Memory I & II, story A, and could tolerate a 10-minute PET scan were eligible to enroll in the specificity (young control) cohort. Note that this study did not exclude ApoE 4 carriers at screening as informing the healthy subjects of this reason for exclusion could create undue anxiety. Thus, these subjects received florbetapir-PET scans, but the data collected were not used in the primary analysis.

The subjects nearing the end-of-life were enrolled in the autopsy cohort, rather than a population of patients with cognitive impairment seeking diagnosis. On the other hand, very healthy young subjects were enrolled in the specificity cohort. The study population for both autopsy cohort and the specificity cohort are not representative of the intended patient population who are in middle age group with MCI.

Subjects who qualified for the study had a catheter(s) placed for intravenous (IV) administration of florbetapir F 18. Subjects received a single IV bolus of 370 Megabecquerel (MBq) (10 millicurie [mCi]) of florbetapir F 18 followed by brain PET imaging for 10 minutes duration, beginning approximately 50 minutes post-injection. Vital signs and safety laboratory tests were obtained before the administration of florbetapir F 18, at the completion of the imaging session, and in a telephone contact made 24 to 48 hours after imaging.

The subjects in the autopsy cohort were followed for one year and anyone who died had autopsy if the family did not withdraw the consent. For subjects in the specificity cohort, only imaging sessions were conducted and results were evaluated.

For the correlation analysis, PET images from the autopsy cohort were assessed for amyloid levels semi-quantitatively (interpretation by three independent and blinded image readers (reader 1, 2, 3) using a 5-point scale (0=none, 1=possible, 2=possible to moderate, 3=moderate to high, and 4=high) for whole brain amyloid deposition) and quantitatively (regional amyloid levels and average cortical amyloid levels, using semi-automated procedures to calculate cortical to cerebellar standard uptake value ratios [SUVRs]).

For the specificity analysis, 40 subjects enrolled in the autopsy cohort with the score ≥ 2 were then mixed with the 74 healthy young subjects enrolled in the specificity cohort. PET images for

the 114 subjects were assessed for presence or absence of amyloid by three independent and blinded image readers (reader 4, 5, and 6) using a binary rating: amyloid positive (Aβ+) or amyloid negative (Aβ-), which is used for the primary analysis. For exploratory analysis, a 3-point reading scale (0=none, 1=moderate, and 2=high) is also used for the 114 subjects by reader 4, 5, and 6.

Three rating methods are defined in this study, which leads to confusion when analyzing the data. Later the sponsor proposed to use the binary rating in clinical practice. However, not all subjects with autopsy have the binary rating. It is more suitable to use only one rating system in the single study.

All of the visual blinded reads were completed at the imaging core lab in accordance with the Independent Review Charter. Separate groups of three readers were used for the autopsy and specificity analysis. All PET scan readers were board-certified in radiology and/or nuclear medicine and were blinded to any clinical or histopathology data on the subject scans being evaluated.

It is proper for the readers to be blinded to clinical information.

As shown in Figure 1, in the autopsy, there are 35 observed deaths within 1 year following the PET imaging procedure. The first 6 subjects to come to autopsy were considered front runners, and an interim analysis was completed on data from these subjects in order to finalize the study methods. 29 subjects were used as the primary efficacy population to test the primary correlation hypothesis in the main phase of the trial. For the specificity analysis, 47 subjects without ApoE 4 carrier (a known risk factor for AD) is the primary efficacy population.

Neuropathological specimens (from subjects coming to autopsy) were assessed for amyloid levels using standardized methods to measure cortical amyloid levels both quantitatively (immunohistochemical measurement of Aβ (IHC)) and semi-quantitatively using the modified Consortium to Establish a Registry for Alzheimer's Disease (CERAD) criteria.

Primary endpoints:

For the autopsy cohort, a visual semi-quantitative rating assessment was performed by three independent readers. Each autopsy-cohort reader rated the degree of florbetapir retention in the grey matter on a scale from 0 (no amyloid) to 4 (high levels of β-amyloid deposition), and the median score of the 3 readers was the primary efficacy endpoint. The global cortical amyloid burden measured by IHC is the primary outcome variable.

We recommend using reader specific analysis instead of the median score approach. The median score may obscure the individual performance of the readers.

For the specificity cohort, the majority qualitative read result of the blinded readers was the primary efficacy endpoint for the specificity evaluation.

Secondary endpoints:

- Regional semi-quantitative visual ratings of amyloid burden on the florbetapir-PET image and the regional cortical amyloid levels as assessed by quantitative IHC (for six brain regions).

- Regional standardized uptake values ratios (SUVRs) for 6 cortical target regions and the global SUVR (as the mean of the SUVRs).

- Reader specific visual semi-quantitative rating for the autopsy cohort (0-4)

- Reader specific qualitative read results for specificity cohort (positive vs. negative): Negative (Aβ-) = Little or no increased cortical gray matter tracer retention by comparison to cerebellum is seen. The defining feature of a positive scan is the localization of the tracer's activity in cortical gray matter.

- Reader specific and median visual semi-quantitative rating for specificity cohort (0-2): 0=Low/None (regions with no difference or slight difference in uptake between cortex and cerebellum), 1=Moderate (regions with modestly higher uptake versus cerebellum (but still lower than white matter), 2=High (regions with significantly higher uptake versus the cerebellum and versus the white matter).

- Regional quantitative analysis of β-amyloid burden, as determined by IHC

- Semi-quantitative evaluation of neuritic plaque burden, as recorded as CERAD score (0=none, 1=sparse, 2=moderate, 3=severe).

Phase I and II studies

Studies A01, A04, A05 are not the major focus of this review. Here, we simply describe the study designs.

Study A01 is a phase I, open label study for providing a preliminary evaluation of the amyloid binding properties, pharmacokinetics and safety of 18-F-AV-45 in healthy elderly volunteers and patients with Alzheimer's disease. Sixteen (16) healthy volunteers and 16 subjects with AD (>50 years old) were enrolled. All subjects then received a single IV bolus of approximately 10 mCi (370 MBq) 18F-AV-45 and positron emission tomography (PET) imaging began. This study assessed the brain uptake (SUV and SUVR) and distribution, as well as the PK, metabolism, and safety of 18F-AV-45.

Study A04 is a phase I, open label study for investigating Test-retest reproducibility of 18F-AV-45 for brain imaging of amyloid in healthy volunteers and Alzheimer's disease patients. A total of 25 subjects were enrolled: 21 subjects (11 subjects with AD (>50 years old) and 10 healthy volunteers (35-55 years old)), and an additional 4 AD subjects were enrolled and dosed using a

revised injection protocol to evaluate the effect of slow vs. fast IV injection were enrolled in the primary test-retest phase of the study. Subjects who qualified for the study returned to the clinic for a second imaging session within 4 weeks of the initial imaging session. At each imaging session, subjects were injected with a single IV bolus of 370 MBq (10 mCi) of 18F-AV-45. Approximately 50 minutes after the injection of 18F-AV-45, the subject received a 20 minute continuous dynamic PET scan. Images were evaluated both qualitatively and quantitatively. For the qualitative image evaluation, the kappa statistic was calculated for the agreement of the blinder reader's interpretation of the test and retest scans. The intraclass correlation between the global SUV ratios (SUVR) for the 2 scans was determined for the quantitative image evaluation.

For A04, a blinded design was not used for this study because the same investigational agent was taken in both treatment arms. However, the quantitative image analysis was fully automated and the qualitative / semi-quantitative visual read of the PET images was performed by an expert reader who was blinded to the subject's clinical and experimental condition.

Study A05 is a phase II, open label, parallel group, multicenter study, evaluating the safety and imaging characteristics of 18-F-AV-45 in healthy volunteers, patients with mild cognitive impairment (MCI) and patients with Alzheimer's disease (AD). A total of 184 subjects (45 subjects with AD, 60 subjects with MCI, and 79 cognitively normal subjects) were enrolled in the study. All subjects should be ≥50 years old. Subjects received a single IV bolus injection of 18F-AV-45, followed by brain PET imaging of 10-minutes duration approximately 50 minutes after dose injection. The PET images were evaluated qualitatively (blinded readers), semi-quantitatively (blinded readers), and quantitatively (computerized analysis). For semi-quantitative and qualitative evaluations (visual reads), the images were visually examined by 3 readers who were blinded to all clinical information. For the semi-quantitative evaluation, each reader rated the amyloid burden level on a scale from 0 to 4, and the median score of the 3 readers was the primary efficacy endpoint. For the qualitative evaluation, the readers classified images as either positive for amyloid-beta (Aβ+, AD-like) or negative for amyloid-beta (Aβ-, not AD-like). The majority read was the primary efficacy endpoint for the qualitative evaluation. If the image was only read by 2 readers and they differed on their classification, the majority read was classified as Aβ+. For the quantitative evaluation (computerized analysis), tracer uptake levels were measured for the following 6 target cortical brain regions: frontal cortex, temporal cortex, precuneus, parietal cortex, anterior cingulate, and posterior cingulate. The cerebellum was measured for use as a reference region. The SUVR was calculated for each cortical target region relative to the cerebellum. The primary efficacy endpoint for quantitative evaluation of each subject was the mean of the SUVRs for the 6 cortical target regions.

3.2.2. Patient Disposition, Demographic and Baseline Characteristics

Only information on the pivotal study (Study A07) is summarized here.

As shown in Table 2 and 3, a total of 226 subjects (152+74) were enrolled in the study, 152 in the autopsy cohort (56 subjects with AD, 25 subjects with MCI, 21 with other dementing disorders, and 50 cognitively normal subjects) and 74 (all cognitively normal) in the specificity cohort. At the end of the study, 110 subjects in the autopsy cohort were alive and had valid

images, and 37 subjects had died. Of the 37 subjects who had died, consent to perform the autopsy for 2 subjects was withdrawn by their families. Thus, there were 35 subjects in the autopsy cohort with evaluable scans and with complete autopsy procedures, and had data available for the correlation efficacy analyses.

The first six subjects to come to autopsy were used in the front-runner analysis, and the remaining 29 subjects comprised the primary efficacy population for the autopsy cohort. Of the 74 subjects in the specificity cohort, 47 were identified as non-ApoE 4 carriers and were included in the primary specificity efficacy analyses.

All 226 subjects injected with florbetapir F 18 were included in the safety analyses.

Table 2: Subject disposition in autopsy and specificity cohort

	Autopsy cohort (N=152)
Injected (safety population)	152 (100%)
Imaged with valid images	147 (97%)
imaging acquisition failure	2 (1%)
injected with invalid image or not imaged	3 (2%)
Valid image but subject still living at study completion	110 (72%)
Deceased within time of study	37 (24%)
Deceased with autopsy	35 (23%)
Deceased with no autopsy	2 (1%)
Subjects with an autopsy and an evaluable image	35 (23%)
Front-runners	6 (4%)
Efficacy population	29 (19%)

	Specificity cohort (N=74)
Injected (safety population)	74 (100%)
Imaged with valid images	74 (100%)
Subjects with an evaluable image	74 (100%)
Valid image and an ApoE 4 carrier	22 (30%)
Valid image and ApoE 4 status not available	5 (7%)
Efficacy population	47 (63%)

Table 3**:** Study populations

	Autopsy cohort	Specificity cohort
Safety population	152	74
Primary efficacy population	29	47
Efficacy population including front runners/ ApoE 4 carriers	35	74

The selected variables on demographic and baseline characteristics are the following:

- Age (years)
- Gender: male, female
- Race: Caucasian, Black or African-American, Asian, American Indian or Alaska native, other
- Ethnicity: Non-Hispanic or Latino, Hispanic or Latino
- Height
- Weight
- Diagnosis (AD, Mild cognitive impairment, Other dementing disorder, No cognitive impairment)
- Mini-Mental State Examination (MMSE) score (0-30)
- Wechsler memory scale –immediate recall
- Wechsler memory scale – delayed recall

The distribution of the three populations by the above variables is shown in Table 4 and 5.

Table 4: Demographic characteristics by cohort

	Autopsy Cohort			Specificity Cohort	
Characteristic	Subjects imaged n=152	Subjects with autopsy n=35	Subjects in primary efficacy population n=29	Subjects Imaged n=74	Non-ApoE 4 carriers n=47
Age (years)					
mean (sd)	78 (13)	79 (13)	80 (13)	27 (6)	26 (7)
Median	81.5	84	85	25.5	24
Range	(38, 103)	(47, 103)	(55, 103)	(18, 50)	(18, 50)
Gender					
Male	71 (46.7%)	18 (51.4%)	15 (51.7%)	48 (64.9%)	32 (68.1%)
Female	81 (53.3%)	17 (48.6%)	14 (48.3%)	26 (35.1%)	15 (31.9%)
Race					
Caucasian	134 (88.2%)	32 (91.4%)	26 (89.7%)	57 (77.0%)	36 (76.6%)
Black or African-American	10 (6.6%)	2 (5.7%)	2 (6.9%)	6 (8.1%)	4 (8.5%)
Asian	2 (1.3%)	0	0	4 (5.4%)	3 (6.4%)
Native American/Alaskan Native	2 (1.3%)	0	0	7 (9.5%)	0
Other	4 (2.6%)	2 (5.7%)	1 (3.4%)	0	4 (8.5%)
Ethnicity					
Non-Hispanic or Latino	139 (91.5%)	33 (94.3%)	28 (96.6%)	69 (93.2%)	44 (93.6%)
Hispanic or Latino	13 (8.5%)	2 (5.7%)	1 (3.4%)	5 (6.8%)	3 (6.4%)
Weight (kg)					
mean (sd)	68 (19)	69 (21)	67 (19)	81 (16)	80 (16)
Median	64	62	62	79	78
Range	(33, 133)	(42, 133)	(42, 133)	(52, 130)	(53, 130)
Height (cm)					
mean (sd)	166 (11)	169 (11)	168(12)	175 (10)	175 (11)
Median	165	166	165	175	178
Range	(127, 200)	(152, 200)	(152, 200)	(145, 201)	(145, 201)

Table 5: Baseline characteristics by cohort

		Autopsy Cohort		Specificity Cohort	
Characteristic	Subjects imaged n=152	Subjects with autopsy n=35	Subjects in primary efficacy population n=29	Subjects Imaged n=74	Non-ApoE 4 carriers n=47
Diagnosis					
Alzheimer's disease	56 (36.8%)	17 (48.6%)	13 (44.8%)	0	0
Mild cognitive impairment	25 (16.5%)	3 (8.6%)	2 (6.9%)	0	0
Other dementing disorder	21 (13.8%)	6 (17.1%)	5 (17.2%)	0	0
No cognitive impairment	50 (32.9%)	9 (25.7%)	9 (31.0%)	74 (100%)	47 (100%)
MMSE					
N	115	26	21	74	47
mean (sd)	21.2 (9.3)	18.1 (10.2)	19.9 (10)	29.7 (0.6)	29.8 (0.4)
Median	25	20	23	30	30
Range	(0, 30)	(0, 30)	(0, 30)	(27, 30)	(29, 30)
Wechsler memory scale-immediate recall					
N	107	22	19	74	47
mean (sd)	6.3 (5.0)	5.5 (5.0)	5.9 (5.2)	16.1 (3.3)	16.3(2.6)
Median	6	5.5	6	16	16
Range	(0, 19)	(0, 17)	(0 , 17)	(6, 23)	(12, 22)
Wechsler memory scale-delayed recall					
N	107	22	19	74	47
mean(sd)	5.1 (4.8)	3.8 (4.3)	3.8 (4.5)	15.4(3.5)	15.4 (2.8)
Median	5	2.5	1	15	15
Range	(0 , 23)	(0, 13)	(0, 13)	(5, 22)	(8, 21)

As shown in Table 3a, the average age (SD) for the 29 subjects included for the primary efficacy analysis for the autopsy cohort is 80 (SD=13). The median age is 85, maximum is 103 and minimum is 50. On the other hand, very healthy young subjects were enrolled in the specificity cohort. The average age (SD) of the 47 subjects for the primary analysis is 26 (SD=7). The median age is 24, maximum is 50, and minimum age is 18. There are more Males in the specificity cohort. The subjects in the specificity cohort are also higher and heavier compared with those in the autopsy cohort. The subjects in the two cohorts are very different in terms of baseline demographics.

As shown in Table 5, the subjects in the autopsy cohort had diagnosis with AD, MCI, other dementia disorder, and No cognitive impairment; the subjects in the specificity cohort are all healthy subjects without cognitive impairment.

3.2.3. Statistical Methodologies

In this section, the statistical methods used by the sponsor are in black, and the reviewer's comments and alternative methods are in italics. If there is no comment, the methods proposed by the sponsor were used by the reviewer.

Study A07:

All correlation analyses were one-sided while all other statistical tests were two-sided with a significance level of α=0.05 and were performed using statistical analysis system (SAS®) version 9.0 or higher. Data were summarized using descriptive statistics.

For the primary efficacy correlation analysis and the secondary efficacy analysis in the autopsy cohort study, Spearman's rank order correlation was determined as well as the asymptotic standard error (ASE) and 95% CI using Fisher z-transformation. The primary hypothesis was that there was a significant positive correlation between the visual semi-quantitative rating of the florbetapir-PET images (median of three readers) and the quantitative measurements of cortical amyloid burden (IHC).

For some secondary efficacy analyses, if both of the two variables for the correlation study are continuous, Pearson correlation was used instead of the Spearman's rank order correlation.

For the specificity cohort, the primary hypothesis was that the observed specificity of the florbetapir-PET imaging in the specificity cohort would be ≥90% (i.e., ≥90% of the florbetapir-PET scans from subjects in the specificity cohort would be rated as Aβ- on an independent read, using the majority view of three readers). For the majority view of the three readers, the number and percent (specificity) of Aβ- using the florbetapir-PET scan was determined as well as the 95% CI.

The method for computing the 95% CI is not clearly stated by sponsor. In this review, Wilson score interval for proportion will be used because of the small sample size.

Exploratory efficacy analyses were also conducted and some are listed in the following:

1. Correlation Analysis between Visual Semi-quantitative Rating of Florbetapir-PET images and Measurement of Cortical Amyloid Burden (IHC) for Individual Brain Regions.

The hypothesis for these analyses was that in each region there would be a significant correlation between the visual semi-quantitative rating of the florbetapir-PET images (median of three readers) and the quantitative measurements of cortical amyloid burden (IHC). For each of the six brain regions (anterior cingulate, frontal cortex, parietal cortex, posterior cingulate, precuneus, and temporal cortex),

3. Correlation Analysis between Quantitative Assessment of Florbetapir-PET SUVR and Measurement of Cortical Amyloid Burden (IHC)

The hypothesis for this analysis was that there was a significant positive correlation between the quantitative assessment (SUVR) of the florbetapir-PET signal and the quantitative measurements of cortical amyloid burden (IHC).

4. Correlation Analysis Including Front-runner Subjects (29+6=35)

5. Specificity Determination Including ApoE 4 carriers (47+27=74)

6. Inter-reader Agreement for Visual Semi-quantitative Rating of Global Amyloid Burden on the PET Images using Kappa Statistic

The degree of agreement between two readers for the visual semi-quantitative rating of global amyloid burden on the PET images was assessed using the weighted kappa statistic in a pair-wise manner. A weighted kappa statistic for multiple categories was determined resulting in three inter-reader agreement analyses. This analysis was performed for the Efficacy Population.

In this review, the reader agreement will be evaluated by simple kappa statistic and weighted kappa statistics with different weight functions for the primary efficacy population, and all subjects with valid images.

7. Correlation Analysis between Quantitative Assessment of Florbetapir-PET SUVR and Visual Semi-quantitative Rating of Florbetapir-PET Images

The hypothesis for this analysis was that there was a significant positive correlation between the quantitative assessment (SUVR) of the florbetapir-PET signal and visual semi-quantitative rating of the florbetapir-PET images (median of three readers).

There are also some secondary analyses related to CERAD. In this review, *CERAD will not be included because IHC is a more objective and reliable measure of Amyloid level from autopsy.*

In addition to the above secondary analyses, diagnostic agreement will be explored by dichotomizing the IHC measure, the semi-quantitative read, and SUVR. If IHC >1%, then autopsy outcome is positive; otherwise outcome is negative. If semi-quantitative read=0 or 1, then PET scan outcome is positive; otherwise outcome is negative. If SUVR>1.1, then PET scan outcome is positive; otherwise it is negative. Sensitivity and specificity will be evaluated using the binary outcomes.

All of the secondary analyses with median read from the three readers will also be explored by reader.

All adverse event summaries were prepared using the set of treatment-emergent adverse events only. Treatment-emergent adverse events were summarized by cohort (autopsy, non-autopsy

specificity) as well as all subjects (both cohorts). The change from baseline in clinical laboratory values and vital sign measurements were analyzed within treatment group.

Study A01, A04, and A05:

In this review, we do not redo the statistical analyses conducted for phase I and II studies A01, A04, and A05 by the sponsor.

For A04, only the correlation between the test and retest PET scan results (qualitative, semi-quantitative, and quantitative) are explored in this review.

For A01 and A05, the SUVR distribution is explored with the pooled data A01, A04, A05, and A07.

For A05, the binary rating is explored together with A07 data, using the observed agreement and kappa statistics. The 95% confidence intervals for the observed agreement are computed by Wilson score interval for proportion, and the 95% confidence intervals for kappa statistics are obtained using normal approximation.

3.2.4. Results and Conclusions

Brief summary of the applicant's results and conclusion from A07

The results of this study demonstrate that florbetapir-PET accurately detects the presence and density of β-amyloid aggregates. Strong statistically significant positive correlations were observed between florbetapir-PET (both visual blinded reader assessment and computerized SUVR measurement) and histopathologic measurements of β-amyloid. The specificity of the florbetapir-PET scan blinded read was 100% in this study for the young healthy subjects.

The study met both of the primary endpoints: 1.) a significant correlation was observed between the visual reader semi-quantitative assessment of florbetapir-PET scans and the amyloid burden measured at autopsy by immunohistochemistry, and 2.) the measured specificity in the specificity cohort was > 90%.

In addition, this study demonstrated that the previously defined quantitative SUVR measurement threshold of ≥1.10 for a positive (i.e., abnormal) scan was highly accurate for the detection of pathologically significant β-amyloid levels.

Florbetapir F 18 Injection was well tolerated in this study. There was one serious adverse event/death occurring in the autopsy cohort during the safety monitoring period and was considered unrelated to drug treatment. The overall adverse event rate was not higher in the end-of-life autopsy population as compared to the healthy volunteers enrolled in this study.

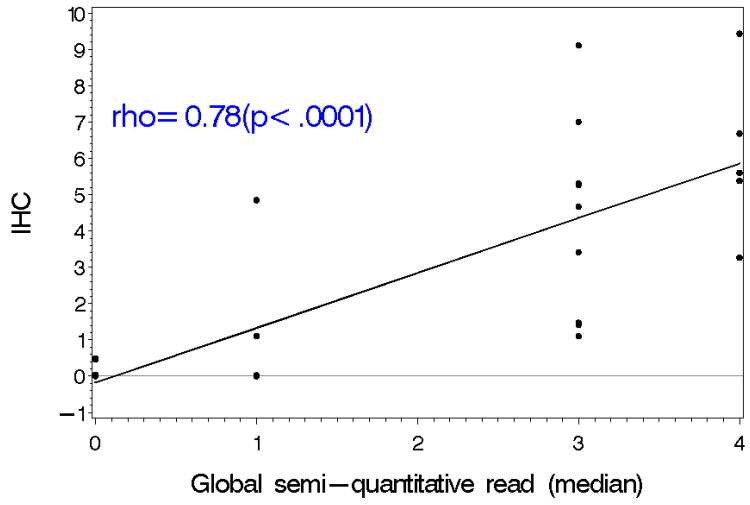
Results on efficacy from the reviewer

In summary, the results on the primary analyses are consistent with the point from the applicant's results and conclusion. The good correlation results, however, do not in themselves provide a clear indication of clinical utility. More details are discussed in the following sections. If not stated, the results are for A07 study.

Primary efficacy analyses

For the primary efficacy analyses, the results are consistent with those provided by the sponsor. For the correlation between the median blinded visual read of the PET images and the quantitative IHC measures assessed post-mortem, a highly significant Spearman correlation (rho=0.77 with pvalue<.0001, 95% CI: 0.58 to 0.88) is obtained for the 29 subjects in primary efficacy population (also shown in Figure 2). For the specificity cohort, 100% (47/47) of young healthy subjects were rated as amyloid negative on the florbetapir-PET scan, with 95% CI as (92%, 100%).

Figure 2: IHC vs. global semi-quantitative read (median read) from readers 1, 2, and 3 for the 29 primary efficacy population.



For autopsy cohort, 35 subjects were also evaluated for the correlation between the semi-quantitative reads and the quantitative IHC measures. The Spearman correlation rho is 0.77 with pvalue <.0001, which is the same as that obtained with 29 subjects. Therefore, in the following, secondary analyses will be conducted for the 35 subjects with autopsy in the autopsy cohort.

Secondary analysis – reader specific correlation

Table 6: Correlation between the semi-quantitative visual blinded read vs. quantitative IHC measures for 35 subjects with autopsy by reader

Reader	Spearman's rho	95% CI
1	0.74	0.57, 0.87
2	0.74	0.46, 0.83
3	0.66	0.48, 0.84
Median read	0.78	0.62, 0.89

As shown in table 6, the correlation values are all high for the three readers. Reader 1 and 2 have similar correlation, reader 3 has slightly lower correlation. However, the confidence interval for reader 2 is wider than the other two readers.

Secondary analysis – Specificity by reader

- For the specificity cohort, 100% (47/47) of young healthy subjects were rated as amyloid negative on the florbetapir-PET scan (reader 4 and 6). The "specificity" is 100% with 95% CI (92%, 100%).
- 98% (46/47) of young healthy subjects were rated as amyloid negative on the florbetapir-PET scan (reader 5). The "specificity" is 98% with 95% CI (89%, 96%).

Secondary analysis – regional correlation

Table 7: Spearman correlation between the semi-quantitative visual blinded read vs. quantitative IHC measures by region (35 subjects with autopsy)

Region	Spearman's rho	95% CI
Frontal cortex	0.71	0.48,0.84
Temporal cortex	0.68	0.44,0.82
Precuneus	0.76	0.56, 0.87
Parietal cortex	0.72	0.50, 0.84
Anterior Cingulate	0.75	0.54,0.84
Posterior Cingulate	0.68	0.43, 0.82

As shown in Table 7, regional correlation values are all high, which is consistent with the global correlation result.

Secondary analysis – Diagnosis agreement

Dichotomizing the semi-quantitative reads and the quantitative IHC measures, we construct the following 2×2 table for the autopsy cohort. Note that IHC ranges from 0% to 10%. And the classification shown in Table 8 is proposed by the sponsor.

Table 8: Diagnosis agreement between the visual read and IHC groups

		IHC	
		Positive (IHC>1%)	Negative (IHC ≤1%)
Semi-quantitative reading	Positive (visual read=2, 3, 4)		
	Negative (visual read=0, 1)		

From Table 8 formulation, sensitivity and specificity (35 autopsy subjects) were obtained for the median reads and also for each reader (1, 2, and 3).

Table 9: Reader specific sensitivity and specificity for the 35 autopsy subjects

Reader	Sensitivity (%)(CI)	Specificity (%)
1	**90 (69, 97)**	**100 (82, 100)**
2	**55 (28, 79)**	**100 (86,100)**
3	**85 (64, 95)**	**80 (55, 93)**
Median read	**85 (62, 95)**	**100 (82, 100)**

Reader specific sensitivity and specificity were evaluated and shown in Table 9. Reader 2 has the worst sensitivity and reader 3 has the worst specificity among the three readers. Because of the small sample size, the confidence intervals are wide. Note that even though the correlation for reader 2 is similar to that for reader 1 (shown in Table 6), the sensitivity and specificity for reader 2 are worse than reader 3. Therefore, a good correlation does not indicate clinical usefulness in terms of ability to classify the disease and non-disease cases with good sensitivity and specificity. The confidence intervals are wide because of the sample size and some lower limits are lower than 70%.

Secondary analysis – IHC values by reader and reading results

For investigating the performance of the readers in identifying the level of IHC level, we also summarized the IHC values by the semi-quantitative reads (Table 10). The reader performance is different. Reader 1 splits the subjects into two big groups (0 and 1, 3 and 4) with low IHC values vs. high IHC values. Reader 2 classified more subjects into low amyloid level groups (0 and 1).

Table 10: Mean (std) IHC values by reader and reading values (35 autopsy subjects)

	Reader1	Reader2	Reader3	Median read
0	**0.28** (0.24)	**0.10** (0.18)	**0.23** (0.36)	**0.13** (0.20)
1	**0.47** (1.35)	**3.46** (2.81)	**1.71** (2.35)	**1.07** (1.86)
2	NA	**6.02** (3.68)	**2.70** (3.81)	**5.39** (.)
3	**4.58** (2.83)	**5.65** (2.80)	**3.68** (2.91)	**4.31** (2.74)
4	**5.40** (2.60)	**5.38** (.)	**6.92** (2.25)	**6.70** (2.14)

Secondary analysis – reader agreement for A07 (two cohorts separately)

The reader agreements among the readers 1, 2, and 3 for the autopsy cohort and readers 4, 5, and 6 for the specificity cohort are investigated using kappa statistic (-1, 1) for the pairwise comparison. Large kappa values are considered indicative of Chance corrected good agreement.

In the specificity study, there are 40 subjects from the autopsy cohort and 74 subjects from the specificity cohort (a total of 114 subjects), who have been evaluated by reader 4, 5, and 6.

As shown in Tables 11 and 12, overall, the reader agreement is better in the specificity study than that in the autopsy cohort. It may be due to the different rating scale, or the design, or the reader. Because of the better reader agreement from the binary rating, the sponsor proposed to use binary rating in clinical practice. However, it has to be investigated in future studies.

Table 11: Reader agreement evaluation for autopsy cohort (rating 0-4)

Reader comp	n	Obs agreement	Kappa (pvalue)	weighted kappa (CA)	weighted kappa (FC)
1 vs. 2	29	7/29=0.24	0.07(0.36)	0.47	0.72
1 vs. 3	29	13/29=0.45	0.33	0.65	0.84
2 vs. 3	29	13/29=0.45	0.24 (0.02)	0.51	0.68

Reader comp	n	Obs agreement	Kappa (pvalue)	weighted kappa (CA)	weighted kappa (FC)
1 vs. 2	35	9/35=0.25	0.08(0.25)	0.4	0.61
1 vs. 3	35	17/35=0.49	0.37	0.63	0.8
2 vs. 3	35	16/39=0.46	0.28	0.49	0.64
1 vs. 2	147	47/147=0.32	0.14(0.0003)	0.38	0.54
1 vs. 3	147	69/147=0.47	0.33	0.56	0.7
2 vs. 3	147	72/147=0.49	0.32	0.52	0.68

Note: if pvalue is not included, pvalue is <.00001. Pvalue is obtained from the test of kappa statistic away from 0. P-values are not adjusted for multiple analyses.

Table 12: Reader agreement evaluation for specificity cohort (rating 0-2 and binary rating)

Reader comp	N	obs agreement	Kappa (pvalue)	weighted kappa (Ciccheti-Alison-CA)	weighted kappa (Fleiss-Cohen-FC)
114 subject for the specificity study (rating scale 0, 1, 2)					
4 vs. 5	114	93/114=0.82	0.63	0.72	0.78
4 vs. 6	114	107/114=0.94	0.87	0.93	0.96
5 vs. 6	114	90/114=0.79	0.57	0.68	0.76
114 subject for the specificity study (binary rating)					
4 vs. 5	114	107/114=0.94	0.86		
4 vs. 6	114	113/114=0.99	0.98		
5 vs. 6	114	106/114=0.93	0.84		

Note: if pvalue is not included, pvalue is <.00001. Pvalue is obtained from the test of kappa statistic away from 0. P-values are not adjusted for multiple analayses.

We also compared the rating for the 40 subjects who have both semi-quantitative rating results and binary qualitative rating results (**Appendix 1**). The results show the rating performance is not consistent when different read method were used.

Secondary analysis – inter-reader agreement comparison (A05 and A07)

Here, we compared the reader performance in different studies and subgroups. Focus is on the binary rating proposed for clinical practice by the sponsor after discussion with the Agency recently. As shown in Table 13, a phase II study A05 with 183 subjects and a phase III study A07 with 114 subjects (40 + 74) have direct binary rating results and all have three readers. Most of the study subjects with binary rating do not have autopsy, so we only can evaluate the reader agreement using autopsy based sensitivity and specificity measures. To assess if the binary

image reads are reproducible, we evaluate the reader agreement by using observed agreement and kappa statistic with 95% confidence intervals.

Table 13: Summary of Study A05 and A07 with rating information

Study	Patients	reader	rating
A05 (II)	183 (age 50-92, AD, MCI, ODD, HC	1, 2, 3	Direct binary
A07 (III) (auto)	152 (age 38-103, AD, MCI, HC)	a, b, c	0-4
A07 (autopsy cohort)	40 out of 152 with score >2, age 55-94, AD, MCI, ODD	4, 5, 6	Direct binary
A07 (specificity cohort)	74 (age 18-50, YHC)	4, 5, 6	Direct binary

As shown in table 14, both measures for reader agreement evaluation show that study A05 has worse reader agreement than A07. There is a drop in agreement levels from A07 to A05. For the agreement results for just healthy patients in both studies A07 and A05, the drop of agreement level from A07 to A05 is even larger. This discrepancy implies possibility of age by binary image read interaction and makes the general applicability of specificity results of A07 study questionable.

Table 14: reader agreement comparison (A05 and A07)

Study	Age	Reader comp	N	obs agreement (95% CI)	Kappa (95% CI)
A07	18-94	4 vs. 5	114	0.94(0.88, 0.97)	0.86(0.76, 0.96)
		4 vs. 6	114	0.99 (0.95, 1.00)	0.98(0.94, 1)
		5 vs. 6	114	0.93 (0.87, 0.96)	0.84(0.73, 0.95)
A05	50-92	1 vs. 2	182	0.71 (0.64, 0.77)	0.46(0.34, 0.57)
		1 vs. 3	183	0.93 (0.88, 0.96)	0.86(0.78, 0.94)
		2 vs. 3	183	0.72 (0.65, 0.78)	0.48(0.38, 0.59)
A07 YHC	<=50	4 vs. 5	74	0.96 (0.89, 0.99)	Not estimable
		4 vs. 6	74	1 (0.95, 1.00)	Not estimable
		5 vs. 6	74	0.96 (0.89, 0.99)	Not estimable
A05 HC	50-92	1 vs. 2	78	0.63 (0.52, 0.73)	0.20(0.03, 0.37)
		1 vs. 3	78	0.91 (0.83, 0.96)	0.62(0.36, 0.87)
		2 vs. 3	78	0.67 (0.56, 0.76)	0.28(0.12, 0.43)

Note: all subjects in A07 specificity cohort (YHC group) have age <=50, but only one with 50 and all others are <=40.

Study A05 enrolled three major groups of subjects, presumably healthy controls over 50 years old, subjects with a clinical diagnosis of mild cognitive impairment and subjects with a clinical diagnosis of Alzheimer's Disease.

Table 15 shows that the reader agreement is not consistently strong across all three subgroups of baseline clinical diagnosis. From Table 16, there is a substantial variation in the reader agreement by age groups. Clearly, these data raise questions about the reproducibility of image interpretation for patients who may have a broad range of ages and various degrees of amyloid deposition.

Table 15: reader agreement by baseline diagnosis (A05)

Diag	Readers	Obs agreement	kappa
AD	1 vs. 2	0.89 (0.76, 0.95)	0.64 (0.35, 0.92)
N=44	1 vs. 3	0.91 (0.79, 0.96)	0.77 (0.57, 0.98)
	2 vs. 3	0.84 (0.71, 0.92)	0.54 (0.28, 0.80)
MCI	1 vs. 2	0.70 (0.57, 0.80)	0.44 (0.25, 0.62)
N=60	1 vs. 3	0.98 (0.91, 1.00)	0.97 (0.90, 1.00)
	2 vs. 3	0.72 (0.60, 0.82)	0.47 (0.29, 0.65)
HC	1 vs. 2	0.63 (0.52, 0.73)	0.20 (0.03, 0.37)
N=78	1 vs. 3	0.91 (0.83, 0.96)	0.62 (0.36, 0.87)
	2 vs. 3	0.67 (0.56, 0.76)	0.28 (0.12, 0.43)

Table 16: reader agreement for healthy subjects by age groups (A05)

age	Reader comparison	Observed agreement (95% CI)	Kappa (95% CI)
50-60	1 vs. 2	0.65 (0.45, 0.81)	0.13 (-0.10, 0.37)
N=23	1 vs.3	1.00 (0.86, 1.00)	1.00 (1, 1)
	2 vs.3	0.65 (0.45, 0.81)	0.13 (-0.10, 0.37)
60-70	1 vs. 2	0.67 (0.44, 0.84)	0.27 (-0.06, 0.60)
N=18	1 vs.3	0.94 (0.74, 0.99)	0.64 (0, 1)
	2 vs.3	0.67 (0.44, 0.84)	0.14 (-0.12, 0.39)
70-80	1 vs. 2	0.53 (0.31, 0.74)	0.03 (-0.39, 0.44)
N=17	1 vs.3	0.82 (0.59, 0.94)	0.46 (-0.05, 0.97)
	2 vs.3	0.71 (0.47, 0.87)	0.39 (0.03, 0.74)
80+	1 vs. 2	0.65 (0.43, 0.82)	0.30 (-0.06,0.66)
N=20	1 vs.3	0.85 (0.64, 0.95)	0.57 (0.14, 1)
	2 vs.3	0.70 (0.48, 0.85)	0.40 (0.08, 0.72)

In addition, exploration of PET scan test-retest correlation in study A04 is included in **Appendix 2**. Exploration of SUVR (including correlation between SUVR vs. IHC, Diagnosis agreement on SUVR, and SUVR distribution in subjects from A01, 04, 05 and 07) are included in **Appendix 3**.

3.3. Evaluation of Safety

The safety population consisted of 496 subjects who received an Amyvid injection. There were no withdrawals due to treatment emergent adverse events or TEAEs. As shown in table 17, the most commonly reported adverse event was headache which was reported in approximately 2% of the population.

There were 4 Serious adverse events reported in the study, including 2 deaths although we agree with the sponsor that none of the serious adverse events appeared related to Amyvid. There were also lab value, vital signs, and ECG changes that were not clinically meaningful.

The total radiation dose from an Amyvid PET/CT scan is about 9 mSv. This dose appears reasonable given that there are other approved nuclear medicine drugs that result in a higher radiation dose. Also, one Amyvid PET/CT exposes an individual to a radiation dose that is much lower than the annual limit set by the nuclear radiation commission for radiation exposure to subjects which is 50 mSv /yr limit.

Table 17: Most frequent treatment emergent adverse events (TEAE) for all subjects from A01, A02, A03, A04, A05, and A07 (496 subjects)

TEAE	Subjects (%)
Headache	9 (1.8)
Musculoskeletal pain	4 (0.8)
Fatigue	3 (0.6)
Nausea	3 (0.6)

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

Because of the small sample size (35 subjects with autopsy), we simply summarize the PET scan and IHC results by age groups, baseline diagnosis groups and gender for the subjects in autopsy cohort. We do not explore by race because majority of the subjects are Caucasian, and do not explore by Mini-Mental State Examination (MMSE), Wechsler memory scale-immediate recall, and Wechsler memory scale-delayed recall because of many missing values for the three variables.

We explore the spearman correlation by baseline diagnosis (AD, Mild cognitive impairment (MCI), other dementia disorder (ODD), No cognitive impairment (NC)), age (60-, 60-70, 70-85, 85+) and gender (Male and Female). Some of the confidence intervals can not be estimated and most of the confidence intervals are very wide. This exploration only provides limited information on the patterns for the subgroups.

From Table 18 and Figure 3, 4 and 5, the positive correlation pattern is similar for subjects in different age and gender groups. The subjects diagnosed with AD and other dementia disorders have positive correlation, but those with MCI and without cognitive impairment have negative correlation (IHC vs. the visual reads). More work is needed for investigating the correlation pattern for subjects with different diagnosis at baseline if data with larger sample size is available.

Table 18: Spearmen correlation of IHC and semi-quantitative visual read (reader 1, 2, 3, and median read) by age group, diagnosis at baseline and gender

	Age	Corr	95% CI	Diag	corr	95% CI
	<=60			AD		
IHC vs. median read	(n=4)	0.8	(-0.76, 0.99)	(n=17)	0.48	(-0.02, 0.77)
IHC vs. read from reader 1		0.89	(-0.58, 0.997)		0.38	(-0.13, 0.72)
IHC vs. read from reader 2		0.74	(0-0.81, 0.99)		0.19	(-0.32, 0.61)
IHC vs. read from reader 3		0.8	(-0.76, 0.99)		0.48	(-0.02, 0.77)
	60-70			MCI		
IHC vs. median read	(n=5)	0.87	(-0.18, 0.99)	(n=3)	-0.87	.
IHC vs. read from reader 1		0.63	(-0.62, 0.97)		.	.
IHC vs. read from reader 2		0.87	(-0.18, 0.99)		-0.87	.
IHC vs. read from reader 3		0.87	(-0.18, 0.99)		-0.87	.
	70-85			ODD		
IHC vs. median read	(n=12)	0.74	(0.25, 0.92)	(n=6)	0.58	(-0.49, 0.94)
IHC vs. read from reader 1		0.74	(0.26, 0.92)		0.74	(-0.25, 0.96)
IHC vs. read from reader 2		0.75	(0.27, 0.92)		0.62	(-0.44, 0.95)
IHC vs. read from reader 3		0.54	(-0.08, 0.84)		0.32	(-0.68, 0.89)
	85+			NC		
IHC vs. median read	(n=14)	0.61	(0.15, 0.87)	(n=9)	-0.17	(-0.75, 0.56)
IHC vs. read from reader 1		0.65	(0.27, 0.89)		-0.31	(--0.80, 0.46)
IHC vs. read from reader 2		0.72	(0.01, 0.83)		0.55	(-0.22, 0.88)
IHC vs. read from reader 3		0.55	(0.09, 0.85)		-0.63	(-0.91, 0.09)

	Gender	Corr	95% CI
IHC vs. median read	Male	0.78	(0.47, 0.91)
IHC vs. read from reader 1		0.63	(0.21, 0.84)
IHC vs. read from reader 2		0.82	(0.55, 0.93)
IHC vs. read from reader 3		0.70	(0.32, 0.87)
IHC vs. median read	Female	0.77	(0.44, 0.91)
IHC vs. read from reader 1		0.80	(0.50, 0.92)
IHC vs. read from reader 2		0.71	(0.32, 0.88)
IHC vs. read from reader 3		0.60	(0.15, 0.83)

Figure 3: Scatter plots of ICH vs. semi-quantitative visual read (median read) by age group

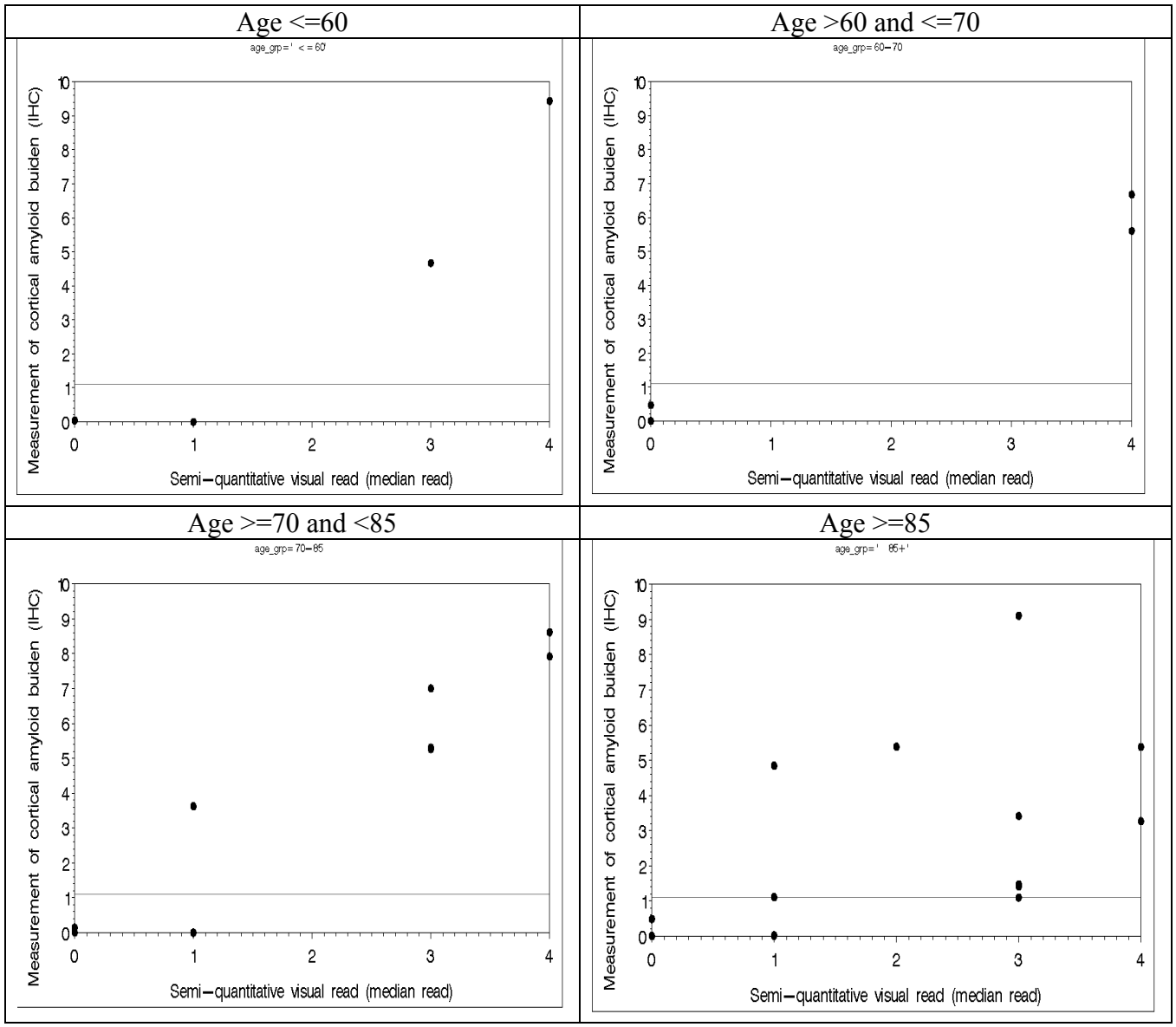


Figure 4: Scatter plots of IHC vs. semi-quantitative read (median read) by diagnosis at baseline

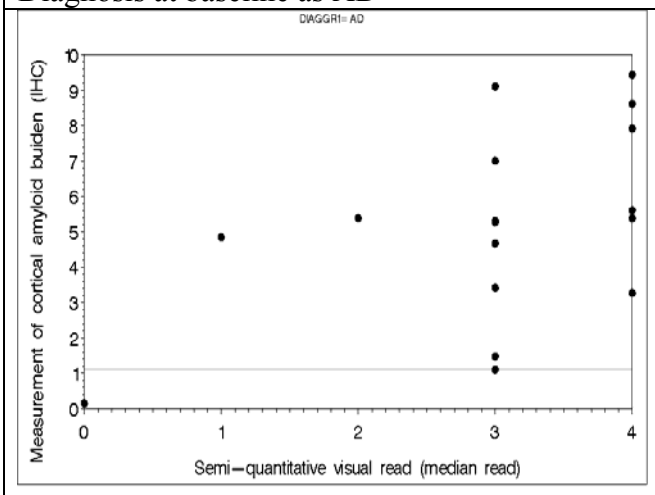
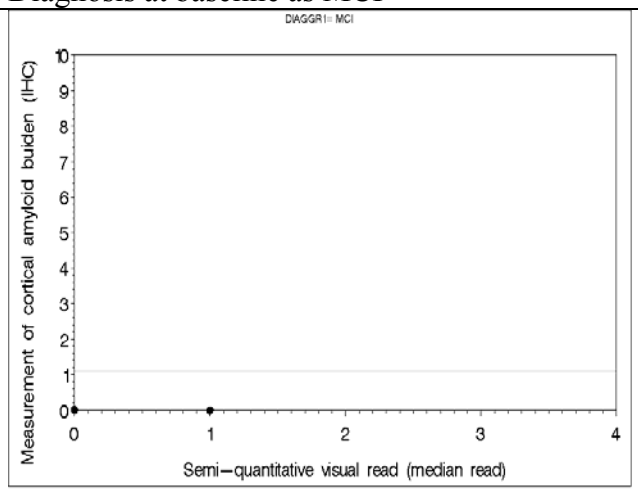
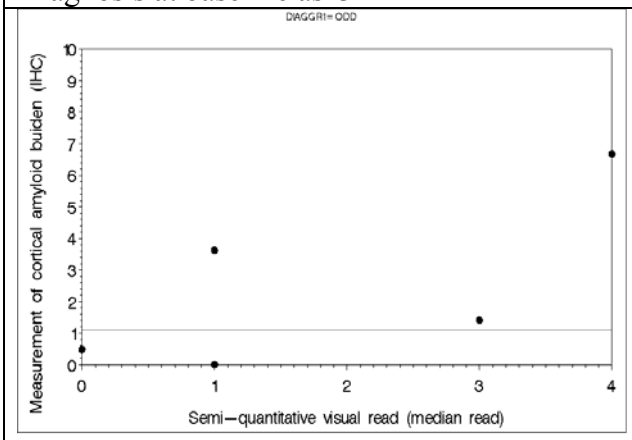
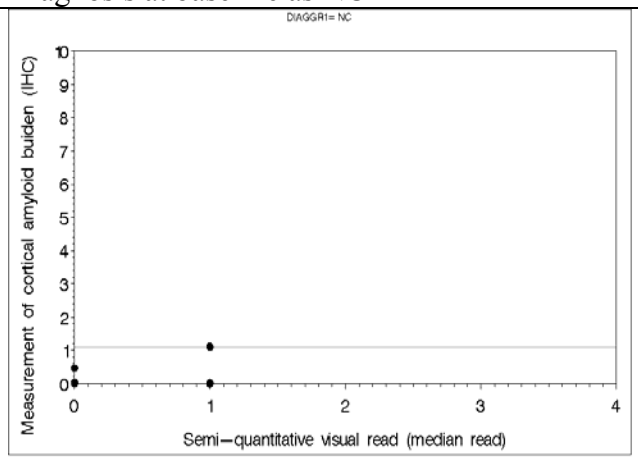
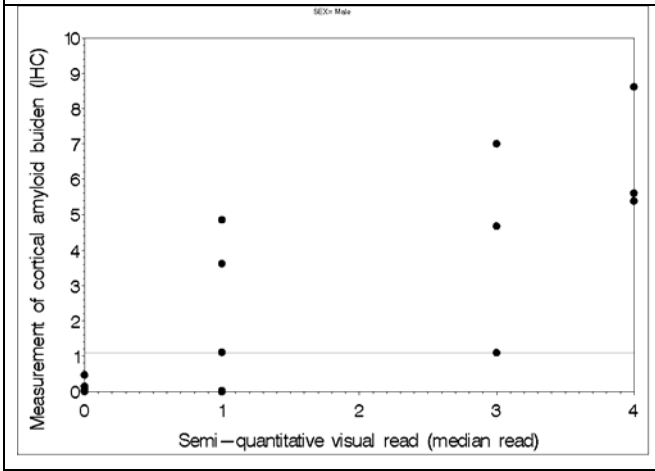
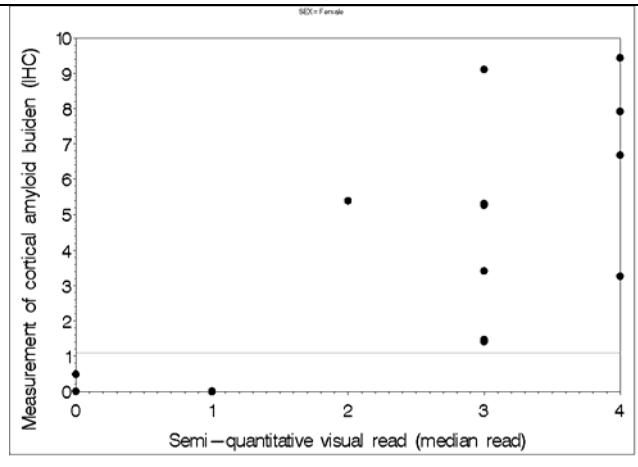
Diagnosis at baseline as AD	Diagnosis at baseline as MCI
	
Diagnosis at baseline as ODD	**Diagnosis at baseline as NC**
	

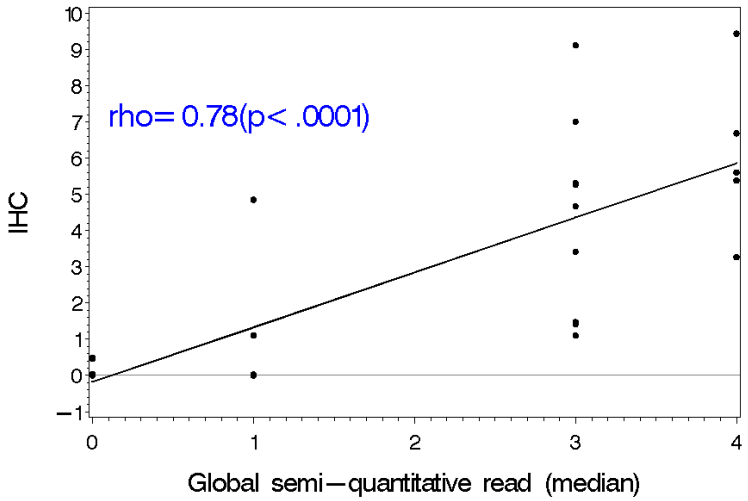
Figure 5: Scatter plots of IHC vs. semi-quantitative read (median read) by gender

Male	Female
	

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

1. For the primary correlation analysis, a statistically significant Spearman's rho of 0.78(p<0.0001, 95% CI: 0.58 - 0.89) was observed between the median of the independent reader semi-quantitative visual ratings of amyloid detected on the florbetapir-PET image and the cortical amyloid level as assessed by quantitative IHC (average percent cortical grey matter area of β-amyloid on the IHC slides).



The results are very similar from the analysis on the 29 subjects in the efficacy population and the 35 subjects with autopsy (6 frontrunners plus the efficacy population). Therefore, sensitivity and specificity were evaluated for the 35 subjects for the two by two table with a post-hoc categorization of median read (semi-quantitative median read 0 and 1 as negative, and semi-quantitative median read 2, 3, 4 as positive) as the row variable and IHC category (IHC>1% as positive and IHC <1% as negative) as column variable. The sensitivity is then 85% (95% CI: 62% to 95%) and specificity is 100% (95% CI: 82% to100%).

2. To assess the variations in the individual reader performance, the correlation, and sensitivity and specificity by reader are evaluated. The Spearman's rho is 0.74 (p<0.0001), 0.74 (p<0.0001) and 0.66 (p<0.0001) for reader 1, 2, and 3 separately. For reader 1, the sensitivity is 90% (95% CI: 69%, 97%) and specificity is 100% (95% CI: 82%, 100%); for reader 2, the sensitivity is 55% (95% CI: 28%, 79%) and specificity is 100% (95% CI: 86%, 100%); and for reader 3, the sensitivity is 85% (95% CI: 64%, 95%) and specificity is 80% (95% CI: 55%, 03%). This indicates that high correlation does not imply good sensitivity and specificity.

3. For the primary specificity analysis, 100% (47/47) of young healthy subjects were rated as amyloid negative on the florbetapir-PET scan by the median read, which reader 4 and 6 agreeing on all 47 cases as negative, and reader 5 scoring negative on 46 of 47 cases. The observed specificity for the majority read is 100% (95% CI: 91%, 100%).

4. The subjects nearing the end-of-life were enrolled in the autopsy cohort, rather than a population of patients with cognitive impairment seeking diagnosis. The average age (SD) for the 29 subjects included for the primary efficacy analysis is 80 (SD=13). The median age is 85, maximum is 103 and minimum is 50. On the other hand, very healthy young subjects were enrolled in the specificity cohort. The average age (SD) of the 47 subjects for the primary analysis is 26 (SD=7). The median age is 24, maximum is 50, and minimum age is 18. The study population for both autopsy cohort and the specificity cohort are not representative of the intended patient population who are in middle age group with MCI.

5. The kappa statistic for the reader agreement evaluation (for the total 147 subjects with 0-4 read) is low to moderate for the autopsy cohort study (0.14 for reader 1 and 2, 0.33 for reader 1 and 3 and 0.32 for reader 2 and 3 using simple kappa; 0.54 for reader 1 and 2, 0.7 for reader 1 and 3, 0.68 for reader 2 and 3 using weighted kappa). The kappa statistic for the reader agreement evaluation (for the total 114 subjects with binary read) is high for the specificity cohort study: 0.86 for reader 4 and 5, 0.98 for reader 4 and 6 and 0.84 for reader 5 and 6. Therefore, the reader performance is more consistent in the specificity cohort study than that in the autopsy cohort study. It may be due to the difference in the reading scale, the training process, or the subjects studied. Therefore, the 100% specificity from the specificity cohort study may not be generalized to the intended patient population.

6. The primary efficacy analysis on correlation for the autopsy cohort was conducted using the 29 subjects, which is a very small sample. The correlation is statistically significantly away from 0. However, the lower bounds of the 95% CI for the sensitivity and specificity by reader are not always high. The confidence intervals are also wide. If we consider the qualitative (binary) read, only 14 autopsy subjects have binary read from reader 4, 5, and 6 in A07. And all of them have autopsy and visual reading results as positive. It is impossible to evaluate the sensitivity and specificity for the (binary) qualitative read using the autopsy data.

7.For the binary reading method proposed by the company, the reader agreement in terms of observed agreement and kappa statistic shows that, overall, A05 has worse agreement than A07. For the subgroup of young healthy subjects in the two studies, the agreement is much higher in A07 than that in A05. This discrepancy implies age by binary image read interaction and makes the general applicability of specificity results of A07 study questionable. In A05, the reader agreement also varied across the subgroups of baseline clinical diagnosis (AD, MCI, HC), and across age groups (50-60, 60-70, 70-80, 80+).

8. Subgroup analysis for A07 data indicates that the positive correlation pattern (IHC vs. semi-quantitative read) is similar for subjects in different age groups. The subjects diagnosed with AD and other dementia disorders have positive correlation, but those with MCI and without cognitive impairment have negative correlation (IHC vs. the visual reads). Since the sample size is very small for each subgroup, more work is needed for investigating the correlation pattern for subjects with different diagnosis at baseline.

9. From study A04, between the test and retest results, the kappa statistic for the binary reads is 0.89 (95% CI: 0.69 to 1), the spearman correlation for the semi-quantitative reads is 0.82 (95% CI: 0.57 to 0.92), and the Pearson correlation for SUVRs is about 0.995 (95% CI: 0.98 to 0.99). The test and retest results (qualitative/semi-quantitative and quantitative) from PET scan are consistent and reproducible

10. From all SUVRs in A01, 04, 05 and 07 studies, there are slightly higher SUVRs in the AD group compared with subjects in other groups (MCI, ODD, and HC). SUVR is higher for older subjects (age >50-60), but the variation in SUVR is high when age is bigger than 50-60. There is not clear pattern for SUVR vs. Mini-Mental State Examination (MMSE). It will be difficult to use SUVR to classify a subject into a positive or negative group because of the big variation.

5.2 Conclusions and Recommendations

Florbetapir F 18 (formerly known as 18F-AV-45 or florpiramine F 18) is a molecular imaging agent proposed here for PET imaging of β-amyloid aggregates in the human brain. In the efficacy supplement reviewed in this report, the sponsor intended to show the correlation between measurements of brain β-amyloid using florbetapir-PET imaging and true levels of β-amyloid measured post mortem (Autopsy Cohort) and and to (2) demonstrate the specificity of florbetapir- PET in a cohort of individuals unlikely to have, and therefore assumed not to have, brain amyloid plaque (Specificity Cohort).

The clinical development program of florbetapir F 18 comprised 6 completed clinical trials. However, there is only one pivotal trial, study A07, which is the major focus of the review in terms of efficacy.

The data from pivotal trial A07 provide statistically significant evidence that median semi-quantitaive Florbetapir F 18 image reads of amyloid burden are highly correlated with pathological read of amyloid burden. This correlation demonstrates that Florbetapir F 18 images detect amyloid deposites in the brain.

However, these results do not provide sufficient evidence for the clinical usefulness of this detection since the performance characteristics (sensitivity and specificity) show considerable inconsistency among the readers for the patients from the end-of-life population. The specificity results, although consistent across readers, are obtained from the population of young healthy volunteers instead of the intended patient population. Moreover, the sponsor has proposed using

binary, qualitative read of Florbetapir F 18 images which has been applied only to 14 patients in whom pathological standard of truth was available. This sample size is too small to assess the clinical usefulness of the proposed qualitative read. Also, reader performance on the binary read method is not consistent among studies (A05 and A07). The perfect specificity observed in Study A07 may not be obtained in the clinical practice because of the reader variation by age and by diagnosis at baseline or because of the study variation. The statistical review team recommends that the indication statement be modified to just state the amyloid detection claim. Clinical utility of such detection can not be inferred from the data generated in the sponsor's Florbetapir F 18 development program.

Appendix 1: Comparison of the rating for the 40 subjects with all rating results

There are 40 subjects from the autopsy cohort with rating from both readers 1, 2, and 3 and readers 4, 5, and 6. According to the study design, the subjects with rating 2, 3, 4 by median read from reader 1, 2, and 3 were mixed with the 74 subjects in the specificity cohort and rated by reader 4, 5, and 6 (with binary rating or semi-quantitative rating 0-2. We compared the rating on the same subjects from the two groups of readers for investigating if the reading performance is consistent in the two cohorts.

According to the protocol provided by the sponsor, in the 0-4 scale, 0 is for none, 1 is for low, 2 is for low to moderate, 3 is for moderate to high and 4 is for high. In the 0-2 scale, 0 is for none, 1 is for moderate, and 2 is for high. In the binary scale, only positive and negative are provided by the reader 4, 5, and 6.

According to the definition, we define 0 and 1 from 0-4 scale = 0 from 0-2 scale, 2 and 3 from 0-4 scale = 1 from 0-2 scale and 4 from 0-4 scale =2 from 0-2 scale. Also if 0-4 scale is 0 and 1 or 0-2 scale is 0, we have the binary scale result as negative, otherwise the binary result as positive.

For the 0-4 scale vs. binary rating comparison, 38/40 of the subjects from autopsy cohort with 0-4 scale reading >=2 (median read from readers 1, 2, and 3) were rated as positive by reader4, 34/40 were rated as positive by reader 5, 37/40 by reader6, and 38/40 by median read.

For 0-4 scale and 0-2 scale comparison, we summarized the frequency by region and the rating score for the two methods. Only median reads are provided in Table 19. 0-4 scale scores are also transferred to 0-2 scale scores for comparison purpose.

Table 19: Number of subjects (frequency) for global cortex and a particular score (0-4 scale, 0-2 scale, and 0-4 scale transferred to 0-2 scale) by median read.

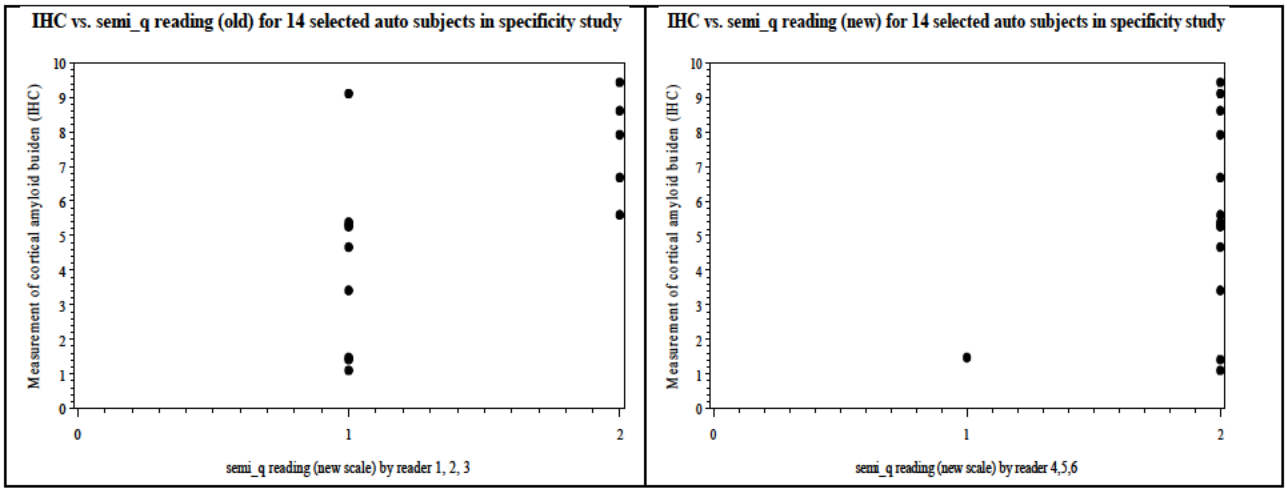
0-4 scale				0-2 scale	frequency
Original 0-4 scale	Frequency	Transferred to 0-2 scale	Frequency		
0	0	0	0	0	2
1	0				
2	9	1	32	1	6
3	23				
4	8	2	8	2	32

It is very clear that the reading results are not the same from the two groups of readers on the same sets of subjects. There are more subjects in the moderate class from the 0-4 scale, but more in the high class in the 0-2 scale. The results obtained from the study may not be generalized to other subjects and readers.

There are 14 subjects out of 40 subjects with reads from two reader groups. As shown in Figure 6, it turns out that almost all subjects are identified with HIGH amyloid level by the median read

from readers 4, 5, and 6 with full range of IHC values. Readers 1, 2, and 3 identified the low and high level of IHC more properly.

Figure 6: IHC vs. semi-quantitative read (median read) by two groups of readers (readers 1, 2, and 3, and readers 4, 5, and 6)



Appendix 2: PET scan test retest correlation (study A04)

21 subjects (11 with AD and 10 healthy volunteers) were included in the test-retest analysis. 10 subjects with AD and 10 healthy volunteers completed the study. Images (for cortical cortex) were evaluated both qualitatively and quantitatively. For the qualitative image evaluation (positive vs. negative), the kappa statistic was calculated for the agreement of the blinder reader's interpretation of the test and retest scans. For the semi-qualitative image evaluation (0-4), the Spearman correlation is evaluated. The correlation (Pearson correlation) between the global (cortical) SUVRs (cortical SUV/cerebellum SUV) for the 2 scans was determined for the quantitative image evaluation. 10 subjects with AD and 10 healthy volunteers completed the study. The exploration is based on the 20 subjects.

Due to poor subject positioning that resulted in an incomplete brain image on the retest image day, accurate quantitative analysis for healthy control Subject 031-004 was not possible, and this control subject was excluded from the SUVR-based analyses. SUVRs from 50-70 minutes and 50-60 minutes after the administration of 18-F-AV-45 are explored.

The Pearson correlation is 0.996 (95% CI: 0.985 to 0.999) for the test SUVRs and retest SUVRs from 50-60 minutes after drug administration, and 0.994 (95% CI: 0.984 to 0.998) for those after 50-70 minutes after drug administration.

The Spearman correlation between the semi-quantitative visual blinded read for test and retest is 0.82 (95% CI: 0.57, 0.92).

Table 20: Qualitative visual read from test and retest

		Visual read from Retest		
		Positive	Negative	
Visual read from Test	Positive	7	1	8
	Negative	0	12	12
		7	13	20

The kappa statistic on agreement between the test and retest binary reads is 0.89 (95% CI: 0.69 to 1). The sensitivity is 100% and specificity is 92% from Table 20.

In summary, the test and retest results (qualitative/semi-quantitative and quantitative) from PET scan are consistent and reproducible.

Appendix 3: Exploration of SUVR

Correlation between SUVR vs. IHC

We also explored the correlation between the SUVR values for the cortical (global) cortex obtained by PET scan and the IHC values from autopsy. The Pearson correlation is 0.73 with pvalue <.0001. Also as shown in Figure 7, there is positive relationship between the two variables. However, the Pearson correlation between SUVR and IHC is 0 for subjects with SUVR <=1.1 and 0.14 for subjects with SUVR > 1.1 (Figure 8). Therefore, in this case, these discrepant results render the correlation evaluation as inadequate to assess the usefulness of quantitative reads of Florbetapir F 18 images in quantifying the level of amyloid properly.

Figure 7: Scatter plot of IHC vs. global (cortical) SUVR for 35 autopsy subjects

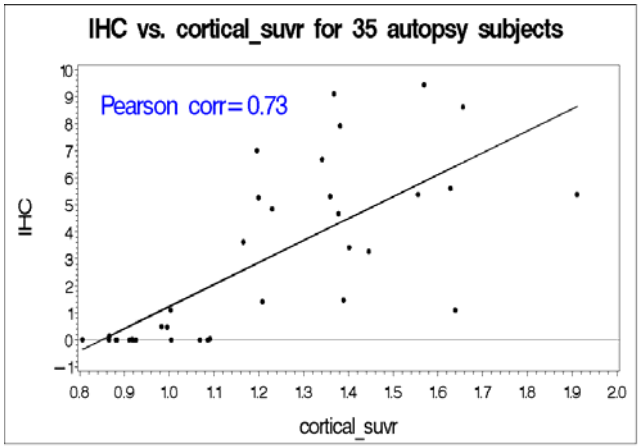
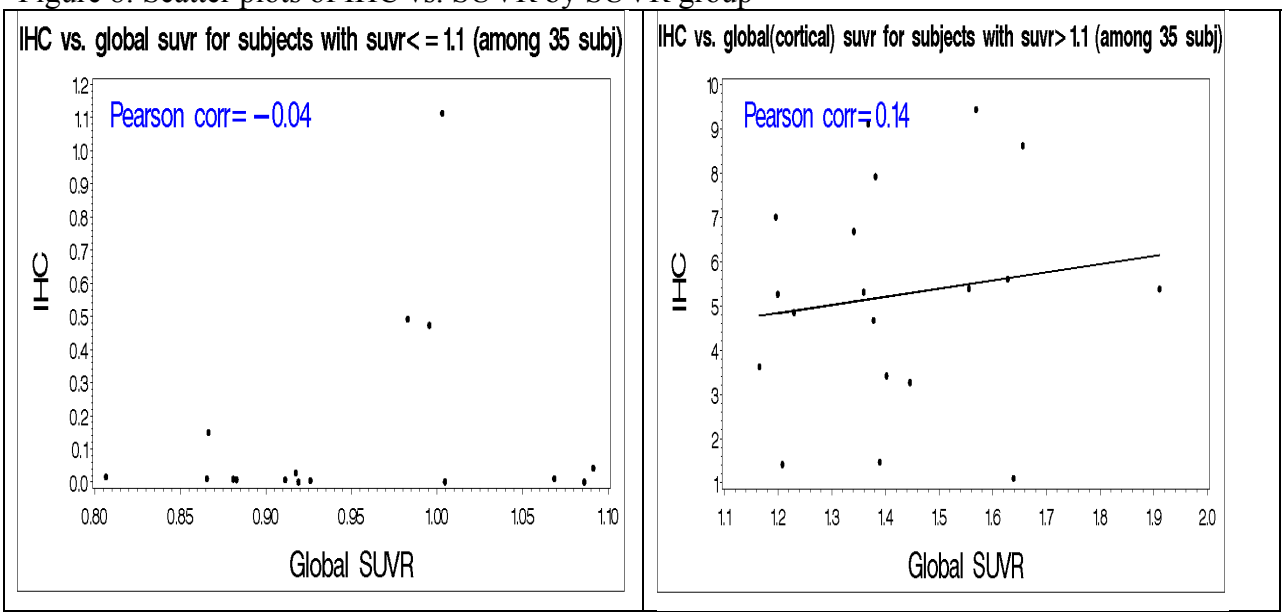


Figure 8: Scatter plots of IHC vs. SUVR by SUVR group



Diagnosis agreement on SUVR

If we define a positive outcome when IHC is bigger than 1%, and negative when IHC is ≤ 1%; and a positive diagnosis when SUVR is bigger than 1.1, and negative when SUVR ≤1.1. We also construct the following table (Table 21).

Table 21: Sensitivity and specificity by dichotomizing the IHC and SUVR variables

		IHC	
		Positive (IHC>1%)	Negative (IHC ≤1%)
SUVR	Positive (SUVR>1.1)	19	0
	Negative (SUVR≤1.1)	1	15
		Sensitivity=95%	Specificity=100%

In this case, a threshold of 1.1 on SUVR has good sensitivity and specificity for identifying the low and high level of amyloid in brain (IHC).

SUVR distribution in subjects from A01, 04, 05 and 07

There are a total of 454 subjects with PET scan results (quantitative SUVR) from A01, 04, 05 and 07 studies: 152 in the autopsy cohort in A07, 74 in the specificity cohort in A07, 26 in A01, 24 in A04 and 183 in A05.

As shown in Tables 22 and 23, and Figure 9, there are slightly higher SUVRs in the AD group compared with subjects in other groups. The SUVRs medians are similar for subject with MCI and healthy subjects. From Table 23, the SUVRs are slightly lower for subjects in A07 than those in the other studies, but do not vary too much.

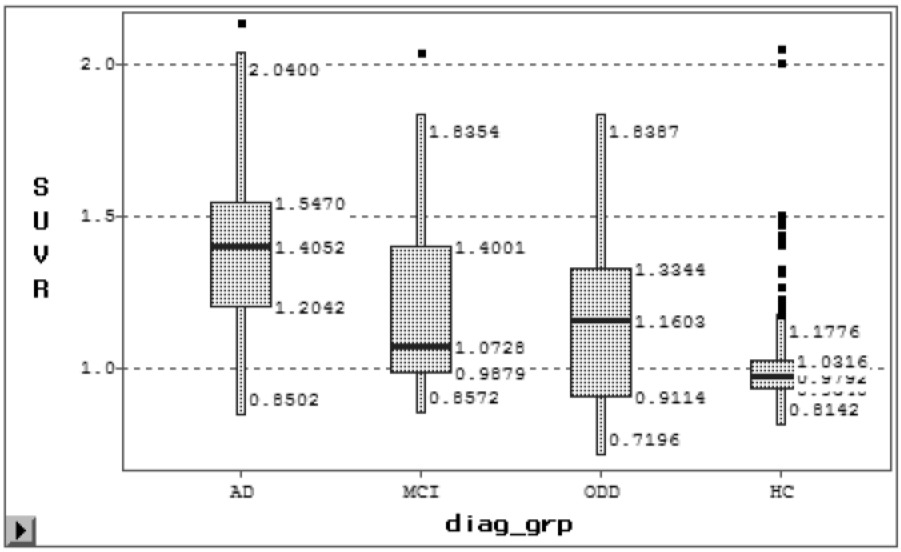
Table 22: SUVR by diagnosis at baseline

Diagnosis at baseline	N	Mean (std)	median	Min	max
AD	126	1.39(0.26)	1.41	0.85	2.14
MCI	85	1.19(0.28)	1.07	0.81	2.04
ODD (other dementing disorder)	21	1.18(0.33)	1.16	0.86	1.84
HC (healthy)	227	1.01(0.16)	0.98	0.72	2.06

Table 23: SUVR by study and diagnosis at baseline

diagnosis	Study	N	mean	std	median	min	max
AD	A01	11	1.40	0.15	1.47	1.05	1.53
AD	A04	14	1.46	0.27	1.48	0.96	2.04
AD	A05	45	1.40	0.27	1.46	0.88	1.93
AD	A07	56	1.35	0.28	1.36	0.85	2.14
HC	A01	15	1.02	0.15	0.96	0.90	1.42
HC	A04	10	1.00	0.05	1.03	0.90	1.05
HC	A05	78	1.05	0.16	1.00	0.86	1.51
HC	A07	124	0.99	0.16	0.96	0.81	2.06
MCI	A05	60	1.20	0.28	1.07	0.86	1.84
MCI	A07	25	1.18	0.29	1.07	0.87	2.04
ODD	A07	21	1.18	0.33	1.16	0.72	1.84

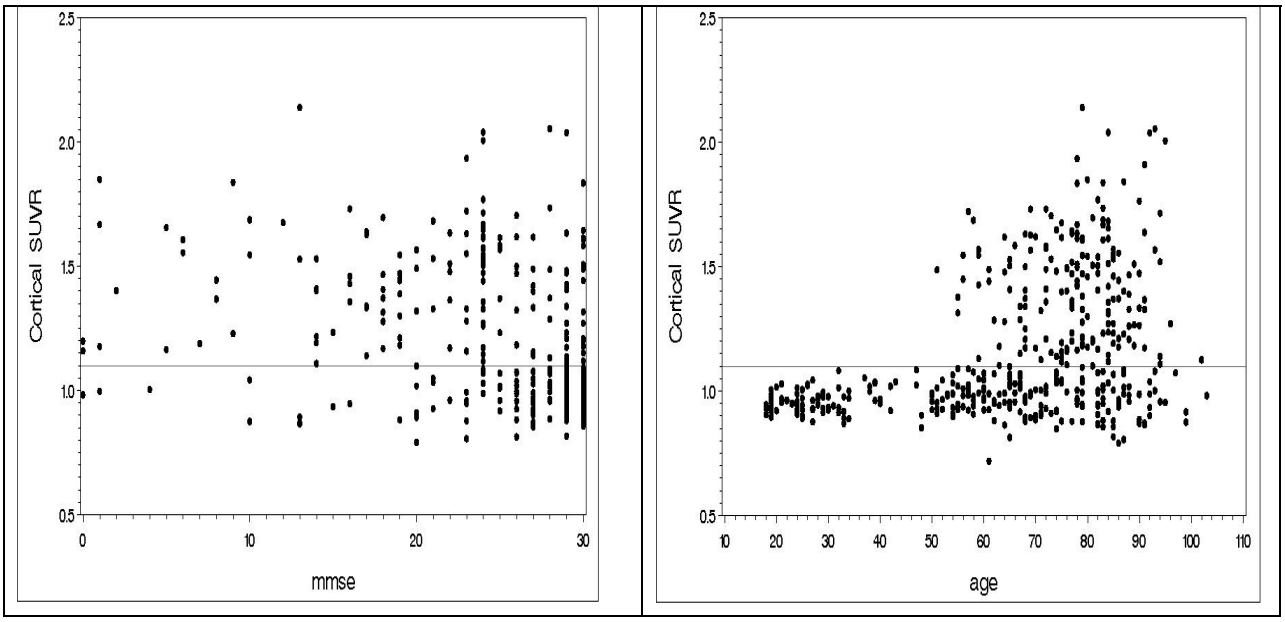
Figure 9: Boxplots of SUVRs by diagnosis at baseline



The Pearson correlation is 0.388 (95% CI: 0.32, 0.48) between SUVR and age, and -0.46 (95% CI: -0.48, -0.31) between SUVR and MMSE. The correlation values are not very high.

Also from Figure 10, SUVR is higher for older subjects (age >50-60), but the variation in SUVR is high when age is bigger than 50-60. There is not clear pattern for SUVR vs. MMSE. It will be difficult to use SUVR to classify a subject into a positive or negative group because of the big variation.

Figure 10: Scatter plots of SUVR vs. MMSE and age



Primary Statistical Reviewer:	Lan Huang, Ph.D.
Date:	Feb 16, 2010
Concurring Reviewer(s):	
Statistical Team Leader:	Anthony Mucci, Ph.D.
Biometrics Division Director:	Rajeshwari Sridhara, Ph.D.

Cc:
HFD-160/Sharon Thomas
HFD-160/Dr. Qi Feng
HFD-160/Dr. Lucie Yang
HFD-160/Dr. Rafel Rieves
HFD-745/Dr. Jyoti Zalkikar
HFD-745/Dr. Alex Gorovets

/s/

LAN HUANG
02/16/2011
ATHONY MUCCI
02/16/2011
RAJESHWARI SRIDHARA
02/16/2011

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

/s/

LAN HUANG
02/16/2011

ANTHONY G MUCCI
02/16/2011

RAJESHWARI SRIDHARA
02/16/2011
Concur with the reviewer's conclusions

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

NDA Number: 202-008

Applicant: Avid radiopharmaceuticals

Stamp Date: 9/17/2010

Drug Name: Amyvid (Florbetapir F 18 Injection)

NDA/BLA Type:
New drug application

On **initial** overview of the NDA/BLA application for RTF:

	Content Parameter	Yes	No	NA	Comments
1	Index is sufficient to locate necessary reports, tables, data, etc.	X			Reports and tables are in M2 and M5, data sets are in M5. Can not find the sas programs
2	ISS, ISE, and complete study reports are available (including original protocols, subsequent amendments, etc.)	X			M5, folder 535-rep-effic-safety-stud
3	Safety and efficacy were investigated for gender, racial, and geriatric subgroups investigated.	X			
4	Data sets in EDR are accessible and conform to applicable guidances (e.g., existence of define.pdf file for data sets).	X			All data sets are in xpt format

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE? __yes_____

If the NDA/BLA is not fileable from the statistical perspective, state the reasons and provide comments to be sent to the Applicant.

Please identify and list any potential review issues to be forwarded to the Applicant for the 74-day letter.

Content Parameter (possible review concerns for 74-day letter)	Yes	No	NA	Comment
Designs utilized are appropriate for the indications requested.	X			
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.	X			
Interim analyses (if present) were pre-specified in the protocol and appropriate adjustments in significance level made. DSMB meeting minutes and data are available.	X			In the A07 study, the 6 front runners (for the interim

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

				analysis) are not included in the primary or secondary efficacy analyses but are included in exploratory analyses.
Appropriate references for novel statistical methodology (if present) are included.			NA	
Safety data organized to permit analyses across clinical trials in the NDA/BLA.	X			
Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.			NA	No dropouts described in the studies

Brief summary of controlled clinical trials
The following table contains information on the relevant trials contained in the submission.

Study number	Design	Treatment arms/Sample size	Primary endpoint/Analysis	Sponsor's findings
F18-AV-45-A07 (one pivotal study)	Open-label, multicenter study Independent randomized blinded PET image read Independent blinded neuro-pathology analysis	Florbetapir F18 injection 10 min PET scan 226 subjects enrolled (152 for autopsy cohort, 74 health young subjects for specificity cohort) 35 subjects have autopsy (6 frontrunners +	Correlation study for the autopsy cohort (PET results vs. histochemistry results from autopsy) Specificity analysis for the specificity cohort	Primary hypothesis #1: Correlation analysis There is a statistically significant correlation ($\rho>0$) between the semi-quantitative visual rating of amyloid burden of the florbetapir-PET scan and the cortical amyloid burden at autopsy as assessed by quantitative immunohistochemistry

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

		29 final)		(IHC). Spearman's Rank Order Correlation, one-sided, $p < 0.05$, $\rho>0$, was used to assess a significant correlation. Primary hypothesis #2: Specificity analysis The observed specificity of florbetapir-PET imaging is ≥90% in young healthy controls (i.e., ≥90% of the florbetapir-PET scans from subjects in the specificity cohort would be rated as negative, which yields 95% CI bounds of 80% to 98% for n=40).
There are also other trials A01 to A06, but A07 is the only pivotal study in this submission.				

Lan Huang 10/14/2010

Reviewing Statistician Date

Jyoti Zalkikar

Supervisor/Team Leader Date

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

/s/

LAN HUANG
11/01/2010

JYOTI ZALKIKAR
11/03/2010