

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:

204677Orig1s000

STATISTICAL REVIEW(S)

STATISTICAL REVIEW AND EVALUATION

Amendment to Review of NDA 204-677

Date:	1/31/2014
From	Lan Huang
Subject	Statistical Review - Amendment
NDA # and Supplement #	NDA 204677
Applicant	Piramal Imaging
Date of Original Submission	12/21/2012
Original PDUFA Goal Date	12/22/2013 (PDUFA V, Program)
Date of Major Amendment	11/22/2013
Proprietary Name / Established	Neuraceq (Florbetaben)
Dosage forms / Strength	Solution for Injection /50 to 5000 MBq
Proposed Indication(s)	Florbetaben is indicated for the detection of β-amyloid in the brain, thereby assisting in the differential diagnosis in adult patients who are being evaluated for Alzheimer's disease and other causes of cognitive decline.
Recommended:	Approval, see comments in summary

Table of Contents

1. EXECUTIVE SUMMARY4

2. INTRODUCTION5

2.1 OVERVIEW.....5

2.1.1. CLASS AND INDICATION5

2.1.2. HISTORY OF PROGRAM DEVELOPMENT6

2.1.3. SPECIFIC STUDIES REVIEWED7

2.2 DATA SOURCES8

3. STATISTICAL EVALUATION9

3.1 DATA AND ANALYSIS QUALITY9

3.2 EVALUATION OF EFFICACY9

3.2.1 *Study Design and Endpoints*9

3.2.2 *Statistical Methodologies*11

3.2.3 *Patient Disposition, Demographic and Baseline Characteristics*.....13

3.2.4 *Results and Conclusions*15

3.3 EVALUATION OF SAFETY17

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS18

4.1 GENDER, RACE, AND AGE18

4.2 OTHER SPECIAL/SUBGROUP POPULATIONS20

5. SUMMARY AND CONCLUSIONS23

5.1 STATISTICAL ISSUES23

5.2 COLLECTIVE EVIDENCE24

5.3 CONCLUSIONS AND RECOMMENDATIONS25

LIST OF TABLES

Table 1: Summary of Study FBB-01_01_13. 7

Table 2: Patient disposition (% in ()) for Study FBB-01_01_13 for different patient populations
..... 14

Table 3: Sensitivity and specificity with web-training PET assessment for 54 autopsy subjects
and the 10 YHVs..... 16

Table 4: Inter-reader agreement and percent of agreement 17

Table 5: Sensitivity (sen) and specificity (spe) by gender, race and age (in %) (for the 82 subjects
with autopsy)..... 19

Table 6: Reader agreement evaluation by gender, race, and age groups 20

Table 7: Sensitivity (sen) and specificity (spe) point estimates by clinical diagnosis (in %),
without the 10 YHVs 21

Table 8: Sensitivity (sen) and specificity (spe) point estimates by study part (in %)..... 22

Table 9: Reader agreement subgroup analyses by study part and clinical diagnosis 23

1. EXECUTIVE SUMMARY

The medical product under development is Florbetaben F18 (Neuraceq). Neuraceq is indicated for the detection of β -amyloid in the brain, thereby assisting in the differential diagnosis in adult patients who are being evaluated for Alzheimer's disease and other causes of cognitive decline. A negative scan indicates sparse to no neuritic plaques, and is inconsistent with a neuropathological diagnosis of AD at the time of image acquisition; a negative scan result reduces the likelihood that a patient's cognitive impairment is due to AD.

Piramal Imaging (Piramal) submitted an original NDA to the Agency on December 21, 2012, for marketing approval of Neuraceq. The original submission includes two pivotal studies: Study 14595 (the autopsy study) and Study 16034 (the pool-read study). The major issues in the original submission are the low specificity for subjects with autopsy and the varying definitions of Standard of Truth (SoT). The different SoT definitions led to inconsistent results on sensitivity and specificity. This NDA major amendment submission includes one study (FBB-01_01_13), for addressing the clinical and statistical issues identified and conveyed to the sponsor in the Late Cycle Meeting in August 2013.

The new read study [Study FBB-01_01_13] enrolled no subjects but analyzed images from Study 14595 to assess the effectiveness of an electronic program for training clinicians in the appropriate interpretation of [18F] florbetaben PET images. The total number of images is 92 plus 20 images for re-read.

The web-based training used in Study 16034 will be used in this new read study (FBB-01_01_13). The Standard of Truth was determined by histopathological consensus panel, blinded to the clinical and imaging results. The rules are a) Bielschowsky Silver Stain (BSS) in combination with immunohistochemistry (IHC) for neuritic plaque, scored according to CERAO; and b) BSS for neuritic plaque according to CERAD. The sponsor proposed to use BSS plus IHC criteria in the primary analyses, and FDA required BSS only in the analyses (see details in COR-NDAIR-10(General Advice Letter) dated on 10/18/2013).

For the analyses using SoT from BSS only and without the 10 young healthy volunteers (YHVs), the lower bounds of the exact 95% confidence intervals (CIs) of sensitivity are >0.7 for all five readers. The lower bounds of the 95% CIs (specificity) are >0.5 for three readers (<0.6 for three readers). Three out of five reader had the lower bounds of sensitivity >0.6 and the lower bounds of specificity >0.5 (for the same reader). The pre-specified criteria for the primary analyses were met, which is consistent with the sponsor's findings. However, two readers had inadequate performance in terms of specificity (the lower bounds of the CIs are 28% and 37%, less than the pre-specified threshold). With a moderate prevalence, high sensitivity indicates high negative predictive value (NPV), and high NPV indicates that the drug product is clinical useful in ruling out subjects without AD. However, one reader can achieve very good sensitivity by sacrificing the level of specificity, which is not acceptable.

In terms of inter-reader agreement among the 92 subjects (82 autopsy plus 10 YHVs), the kappa value for the 92 subjects is 0.76 with a lower bound of 0.70, and the percent of five readers agree with each other is 74%. This is consistent with the sponsor's finding on reader agreement. In terms of intra-reader agreement, the percent of agreement for all the 20 images is 100%, 85%, 95%, 95%, and 100% for reader 1, 2, 3, 4, and 5, respectively.

Note that the brains (31, 23, and 28) were collected in three periods (Study 14595 period, Study 16034 period, and Study FBB-01_01_13 period). The analyses by the collection show that the reader performance is not the same in different collections. The specificity values (point estimates) of the five readers are extremely low for the 23 brains collected in Study 16034 (56%, 33%, 67%, 56%, and 33%). The performance for the first collection (31 brains collected in Study 14595) is the best among the three collections (sensitivity values 89%-100%, specificity values as 92%, 58%, 92%, 92%, 75%).

Comparing the reader performance between FBB-01_01_13 and Study 16034 (same training, same SoT, different five readers), FBB-01_01_13 has more variation than Study 16034 (especially for specificity). For the analyses with 31 brains, the specificity values from the five readers are 58-92% for FBB-01_01_13, and 67-83% for Study 16034. For the analyses with 54 brains, the specificity values are 48-81% for FBB-01_01_13, and 57-76% for Study 16034.

The statistical results in terms of accuracy (sensitivity and specificity) and reproducibility provide evidence to support the claim for the detection of β -amyloid in the brain proposed in this NDA. There are no MCI subject data, the population most needing the diagnostic method, in Study FBB-01_01_13. Also, there are extremely low values of specificity for the two readers with poor performance, different results in the different collections of brains, and different results in terms of sensitivity and specificity between Study 16034 and FBB-01_01_13. We conclude that the data and analyses support approval, noting the limitations of the study, as listed above.

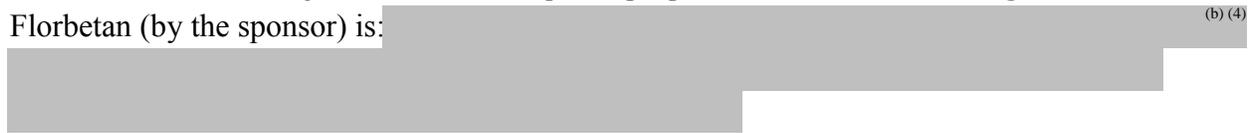
2. INTRODUCTION

2.1 Overview

2.1.1. Class and Indication

Florbetaben is a diagnostic radiopharmaceutical product developed for use with positron emission tomography (PET) imaging for the visual detection of fibrillar amyloid in the form of neuritic plaques in the brain. It is in the same pharmacologic class as Amyvid™ (florbetapir), a

product approved in the United States for the same indication. The proposed commercial name for Florbetaben F 18 Injection is Neuraceq. The proposed Indications and Usage statement for Florbetan (by the sponsor) is: (b) (4)



2.1.2. History of Program Development

Piramal Imaging (Piramal) submitted an original NDA to the Agency on December 21, 2012, for marketing approval of Neuraceq. FDA's Late Cycle Meeting Background Package (dated August 29, 2013) was provided to the Sponsor on August 30, 2013, and the late cycle review meeting was held September 10, 2013. Responses to the statistical substantive review issues were submitted in Sequence 0023, and responses to the clinical substantive review issues were submitted in Sequence 0024. In Sequence 0028 and 0030 additional material was provided to facilitate the review process.

In addition, the sponsor performed study FBB-01_01_13, a new read study of PET scans collected and histopathologically assessed by a pathology consensus panel within study 14595, as announced in Sequence 0024 and suggested by the Agency in the minutes of the Late Cycle Review meeting.

The protocol for the New Read Study FBB-01_01_13 was submitted to the florbetaben IND 78868 on 1 October 2013 (Serial 0104). The new read was performed on 8-9 October 2013; comments from the Agency concerning the protocol were received by the sponsor on 18 October 2013. In the submission to NDA 204677 on 11/6/2013 (sequence 0031, SD 34), the sponsor provided a top-level summary of the data obtained from study FBB-01_01_13, responses to the protocol comments from the Agency (10/18/2013) and the requested datasets.

The major amendment including the complete clinical study report (CSR) for the clinical trial (new read study conducted on 8-9 October 2013) was submitted to NDA 204677, on 11/22/2014 (sequence 0032, SD 35). Note that no new additional data or analyses are presented in this CSR as compared to Amendment 0031. The report in the 11/22/2013 submission is intended to be a formal presentation of the results in ICH E3 format.

In the 11/22/2013 submission, the clinical study report for FBB-01_01_13, the analysis data sets, the programs, and the data definition are included in Section 5.3.5.1.

The sponsor also submitted additional data sets (response to FDA's information request on 11/27/2013) to the NDA on 12/6/2013 (sequence 0034, SD 37).

2.1.3. Specific Studies Reviewed

The full statistical review and evaluation were conducted for Study FBB-01_01_13 (new read study). This new read study evaluated the sensitivity and specificity of the web-training method for reader training, and used pre-specified SoT definition derived from BSS or BSS+IHC, and included 82 patients with autopsy enrolled in Study 14595 plus 10 young healthy volunteers.

Title of the new read study is “A non-interventional study to assess the efficacy, reliability, and reproducibility of the florbetaben-F 18 (FBB) /J-amyloid Positron Emission Tomography (PET) scan visual assessment method as trained via an electronic training tool, using images from the histopathology study 14595”.

Note that Study 14595 (the autopsy study) and Study 16034 (the pool read study) have been reviewed in the original submission. In the amendment, only the new read study is reviewed. A summary of the new read study is shown in Table 1. There are 31 brains collected in Study 14595, additional 23 brains collected for Study 16034 (the total number of brains is 54 for this study), additional 28 brains collected for the new read study (the total number of brains is 82 for this study). All patients with autopsy were enrolled in Study 14595. No additional patients enrolled in Study 16034 and Study FBB-01_01_13.

Table 1: Summary of Study FBB-01_01_13.

	Phase and Design	Primary objective	Treatment period and follow-up	# of Subjects per Arm	Study Population
Study FBB-01_01_13	Phase 3, Single arm, open-label, new read	To assess the efficacy of the visual assessment of PET scans in the detection of beta-amyloid neuritic plaques in the brain, using an electronic training tool.	No subject was dosed in this study	92 images were read (82 EOL subjects with autopsy and 10 YHVs <38yr, and plus 20 of the 92 images for re-read) Five blinded readers with Web-based training	60 AD 10 YHV 51 MCI 4 DLB 9 DEM 9 HV

EOL = end of life; PET = positron emission tomography; SoT = Standard of truth; AD: Alzheimer’s disease. MCI: Mild cognitive impairment. YHV: young healthy volunteer. HV: healthy (non-demential) volunteer DEM: other dementia. DLB: dementia with Lewy bodies.

Major Statistical Issues

- The presence of the 10 YHVs may inflate the performance in terms of specificity. In addition, it is not clear how the 10 YHVs were selected. Excluding the YHVs, the sample size is 30 for specificity evaluation.
- The sponsor proposed to use BSS+IHC for SoT determination. Prior NDA submissions all used BSS only, to determine the SoT in the primary analyses.
- The sponsor used normal approximation to obtain the confidence intervals of sensitivity and specificity. Exact confidence intervals will be more proper in the case of the small sample sizes (i.e., for specificity evaluation with 30 brains).
- The patterns of the reader performance (the five readers in the new read study) for the brains collected in Study 14595 (31 brains), Study 16034 (23 brains), and the new read study (28 brains) are different.
- The patterns of the reader performance of the five readers in the new read study are different compared with the patterns of the five readers in Study 16034 (31 brains and 54 brains).
- Two readers performed much worse than the other three readers in the new read study (the sponsor, Piramal, tried to explain the variation by saying that the two readers didn't follow the instructions properly in spite of Web-based training).
- The re-use of the images for Study 14595, 16034, and FBB-01_01_13 may introduce bias on the reader performance.

The Agency did not approve the inclusion of 10 YHVs, the use of BSS+IHC for SoT determination in the sensitivity and specificity primary analyses (see details in COR-NDAIR-10(General Advice Letter) dated on 10/18/2013).

2.2 Data Sources

Most of the materials reviewed including the applicant study reports and the data sets are provided electronically, and the full electronic path of the documents is

<\\Cdsesub1\evsprod\NDA204677\0032\M5>.

Additional datasets with histopathology information are included in

<\\Cdsesub1\evsprod\NDA204677\0034\M5>.

The datasets analyzed include assess01.xpt, sotwNEUR.xpt, sotwSSTN.xpt, sotpSSTN.xpt, pathoSOT.xpt, and FBB_01_01_13_D_pathoSOT_20131129.xpt. The data do not follow ADAM and SDTM data format.

3. STATISTICAL EVALUATION

3.1 Data and Analysis Quality

Moderate level of effort is needed to process the submitted data.

The following statistical information requests (IRs) were sent to the sponsor on 11/27/2013:

Please provide data for the 82 patients together in one xlsx file (also in xpt file). The content should be the same as the attached file with additional column (an indicator variable) to identify the 31 patient group, the 23 (54-31) patient group, the 28 (82-54) patient group, and the 10 healthy volunteer group. A define file should be provided. The subject id (patient id) should be unique so that the data can be merged with other data sets in this new submission.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

The reviewer's comments will be in italics in this Section.

Study FBB-01_01_31:

A total of 82 brains (subjects enrolled in Study 14595) were available with histopathology assessed. In study FBB-01_01_13, two consensus panel (CP) histopathology-based Standards of Truth (SoTs) were used for the primary and secondary endpoints, respectively:

Bielschowsky Silver Stain (BSS) in combination with immunohistochemistry (IHC) for neuritic plaque, scored according to CERA0; and b) BSS for neuritic plaque according to CERAD. All available PET scan images with CP histopathology from the 14595 study were included in the read, along with scans from 10 young healthy volunteers (YHVs) without histopathology but who can be assumed to be amyloid negative.

Study design:

PET images from 92 scans (plus 20 scans for re-read) obtained at 90-110 min post injection (pi) were randomly assigned for, blinded visual assessment by five independent blinded readers.

- Electronic training of the visual assessment methodology as it is intended to be used as part of the future training for users were the basis for the training/ qualification of the blinded readers. For each image data set 4 brain regions were visually assessed according to the: Regional Cortical Tracer Uptake (RCTU) and Brain Amyloid Plaque Load (BAPL) scoring algorithm. Here, the RCTU (ie, regional assessment) is collapsed to render the overall BAPL (ie, subject level) score.
- The final assessment is binary. A scan with a BAPL score of “1” were considered to be a normal image (ie, a β -amyloid negative subject and a match for specificity) and scans with BAPL scores of “2” and “3” were considered abnormal (ie, a β -amyloid-positive subject and a match for sensitivity).
- All primary and secondary variables used a subject-level visual assessment.
- Sensitivity and specificity were calculated for both SoTs and for each reader separately.
- Inter-reader agreement were assessed by calculating kappa values across all five readers and between each pair of readers
- Intra-reader agreement were assessed by calculating kappa values per reader for a re-read of 20 randomly selected cases from the image data set.

Standard of Truth (SoT)

For the ‘sensitivity and specificity to detect β -amyloid’, two SoTs based on the CP histopathology assessment for neuritic amyloid plaques will be used:

- a) BSS(CERAD)/IHC
- b) BSS(CERAD)

Inter- and intra-reader agreement were assessed via kappa statistics. For this analysis no SoT is necessary.

Primary efficacy variables:

Sensitivity and specificity with the CP histopathology SoTs for neuritic plaque using BSS(CERAD)/IHC based on the individual reads:

The sensitivity and specificity were assessed for each of the 5 readers in the 82 post mortem subjects from Study 14595 with available CP histopathology for neuritic plaque as SoT plus 10 young HVs who served as negative controls (without autopsy).

Secondary efficacy variables:

- Sensitivity and specificity with the CP histopathology SoT for neuritic plaque using BSS according to CERAD based on the individual reads: These variables were analyzed with the same techniques formulated for the primary efficacy variable.

The analysis with BSS should be considered as primary efficacy variable.

- Kappa (κ inter) value across all five blinded readers:
- Inter-individual kappa values for all 10 reader pairs.
- Intra-individual kappa values based on the re-reads separately for all 5 blinded readers.

Missing data

The Sponsor implemented a forced decision rule that required the readers required to provide a score in all cases; no imputation procedure was used.

3.2.2 Statistical Methodologies

The reviewer's comments will be in italics in this Section.

Study FBB-01_01_13

Primary Endpoint and analyses:

Sensitivity and specificity with the CP histopathology SoTs for neuritic plaque using BSS(CERAD)/IHC based on the individual reads:

The sensitivity and specificity was assessed for each of the 5 readers in the 82 post mortem subjects from Study 14595 with available CP histopathology for neuritic plaque as SoT plus 10 young HVs who served as negative controls (without autopsy). Corresponding normal approximated 95% confidence intervals will be calculated.

The following combined hypotheses will be tested:

H0, sens: sensitivity \leq 0.6 vs. H1, sens: sensitivity $>$ 0.6

H0, spec: specificity \leq 0.5 vs. H1, spec: specificity $>$ 0.5

The combined hypotheses will be rejected if the lower limits of the 95% confidence intervals for sensitivity and specificity are higher than the thresholds of 0.6 and 0.5 respectively for at least 3 out of the 5 blinded readers.

BSS only should be used for SoT determination in the primary analyses (to be consistent with the prior NDAs).

Exact confidence intervals should be used for data with small sample sizes (30 for the specificity population).

The sensitivity and specificity values should be higher than the threshold for at least the same 3 out of 5 blinded readers.

Secondary efficacy variables:

Sensitivity and specificity with the CP histopathology SoT for neuritic plaque using BSS according to CERAD based on the individual reads:

These variables were analyzed with the same techniques formulated for the primary efficacy variable.

This should be the primary analyses.

Kappa value across all readers:

The kappa value across the 5 blinded readers for the binary assessment normal / abnormal on the subject level was calculated over all PET scan images read. The confidence interval was calculated based on an asymptotic variance estimate.

The 95% confidence intervals were also calculated for the kappa values obtained for all reader pairs using the same methodology as applied for the calculation of the kappa value across all five readers.

No statistical hypotheses were formulated.

Subgroup analyses and post-hoc analyses:

Sensitivity and specificity compared to both SoTs without HVs:

The primary and secondary analyses were repeated excluding the 10 HVs without autopsy.

Sensitivity and specificity compared to both SoTs with exact confidence intervals:

The primary analyses were repeated using exact 95% Clopper-Pearson confidence intervals.

The above analyses should be treated as primary analyses.

Agreement of blinded readers in percent:

Agreement of blinded readers is presented in terms of percentage of PET scan images in which 5, 4, 3 or 2 readers agreed.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

In Study FBB-01_01_13, the effectiveness of an electronic training program for image orientation and interpretation was evaluated using PET images across subjects with different cognitive abilities who had participated in Study 14595. Inter-reader reproducibility of image interpretation was assessed using images from subjects with a truth standard (82 patients who underwent an autopsy and 10 YHVs).

Among the 82 subjects with autopsy, the median age was 81 years (range 48 to 98), 43% were females, and 66% were Caucasian (White). There are 31 brains collected in Study 14595, 23 brains collected in Study 16034, and 28 brains collected in Study FBB-01_01_12. Among the 10 YHV subjects, the median age was 25 (range 22-38), 4 were females, and 10 were White.

More information on the patient disposition is presented in Table 2.

Table 2: Patient disposition (% in ()) for Study FBB-01_01_13 for different patient populations

Group		Total subjects without YHVs (n=82)	Brains from 14595 (31 brains)	From 16034 (23 brains)	From FBB-01_01_13 (28 brains)	YHVs (n=10)
Gender	Female	35 (43)	11 (35)	9 (39)	15 (54)	4 (40)
	Male	47 (57)	20 (65)	14 (61)	13 (46)	6 (60)
Age	<=65	9 (11)	1 (3)	3 (13)	5 (18)	10 (100)
	>65	73 (89)	30 (97)	20 (87)	23 (82)	0
Age	<=75	25 (3)	6 (19)	8 (35)	11 (39)	10 (100)
	>75	57 (70)	25 (81)	15 (65)	17 (61)	0
Age	Median (range)	81 (48-98)	83 (62-97)	78 (58-98)	80 (48-98)	25 (22-38)
Race	White	54 (66)	16(52)	17 (74)	21 (75)	10 (100)
	Asian	28 (34)	15(48)	6 (26)	7 (25)	0
Clinical diagnosis	AD	60 (73)	22 (71)	19 (83)	19 (68)	0
	MCI	0 (0)	0	0	0	0
	HV excluding YHVs	9 (11)	6 (19)	2 (8)	1 (4)	0
	YHV	0 (0)	0	0 (0)	0(0)	10 (100)
	DEM	9 (11)	2 (6)	1 (4)	6 (21)	0
	DLB	4 (5)	1 (3)	1 (4)	2 (7)	0

AD: Alzheimer's disease. MCI: Mild cognitive impairment. YHV: young healthy volunteer. HV: healthy volunteer DEM: other dementia. DLB: dementia with Lewy bodies

3.2.4 Results and Conclusions

Summary of the main results and conclusions provided by the sponsor (report no. A0001) is shown below.

Study FBB-01_01_13:o

The present Histopathology Read Study FBB-01_01_13 was successfully performed. All pre-specified endpoints were met.

PET scans from 82 deceased subjects and 10 young healthy volunteers were formally evaluated by five electronically trained blinded readers using the assessment method proposed for clinical practice.

Reading results were compared with presence (moderate or frequent) or absence (none or sparse) of amyloid neuritic plaques and scored by the histopathology consensus panel:

- Median Sensitivity: 96.4% (range 91.1% - 100.0%)
- Median Specificity: 86.1% (range 52.8% - 94.4%)

The combined endpoint for sensitivity and specificity was successfully met with the same 3/5 readers exceeding the pre-specified thresholds of 0.6 (sensitivity) and 0.5 (specificity).

The results were not affected by exclusion of the 10 young healthy volunteers. No change was observed in point estimates:

- Median Sensitivity: 96.4% (range 91.1% - 100.0%)
- Median Specificity: 88.5% (range 53.9% - 92.3%)

Similar sensitivity results were obtained applying the pre-specified analysis of BSS only, demonstrating that sensitivity results are independent of the SoT used. However, with BSS alone, median specificity is 76.7% (range 46.7% - 80.0%); as BSS alone missed neuritic plaques in four subjects, which therefore were incorrectly analyzed as false-positives (thus impacting the specificity).

Observed inter-reader agreement was $\kappa = 0.71$. This may be considered a measure of the reproducibility of the results obtained with the visual assessment method and an assessment of the electronic training program.

The reviewer’s findings (based on analyses with SoT determined by BSS only) are presented below.

Study FBB-01_01_13:

Sensitivity and Specificity analyses with SoT determined by BSS only

Before looking at the data, FDA has pre-specified the analyses with SoT determined by BSS only and without the 10 YHVs as the **primary analyses**. Exact confidence intervals (CIs) were calculated, which is more proper for data with small sample size.

As shown in Table 3, without the 10 YHVs, the lower bounds of the exact 95% CIs (sensitivity) are >0.7 for all five readers. The lower bounds of the 95% CIs (specificity) are >0.5 for three readers (<0.6 for three readers). Three out of five reader had the lower bounds of sensitivity >0.6 and the lower bounds of specificity >0.5 (for the same reader). The pre-specified criteria for the primary analyses were met, **which is consistent with the sponsor’s findings. However, two readers had very low levels of specificity, not achieving pre-specified level of adequacy (the lower bounds of the CIs are <0.4).**

The reader performance in terms of specificity is better for analyses including 10 YHVs compared with the analyses without the 10 YHVs, which is expected because it is easier to read the images for the YHVs. The lower bounds of the 95% CIs (specificity) are >0.6 for three readers. **Inclusion of the 10 YHVs did affect the results, which is not the same as what the sponsor claimed.**

Two readers had low specificity values (the lower bounds of the CIs <0.4) with and without the 10 YHVs.

Table 3: Sensitivity and specificity with web-training PET assessment for 54 autopsy subjects and the 10 YHVs

Reader	Without YHVs, n=82=52 pos + 30 neg				With YHVs, n=52 pos+ 40 neg	
	Sensitivity	CI	Specificity	CI	Specificity	CI
1	94	(84, 99)	80	(61, 92)	85	(70,94)
2	98	(90, 100)	47	(28, 66)	47.5	(32, 64)
3	90	(79, 97)	80	(61, 92)	85	(70, 94)
4	96	(87, 99.5)	77	(58, 90)	77.5	(62, 89)
5	100	(93, 100)	57	(37, 75)	55	(38, 71)

Inter-reader kappa and agreement

The sponsor used a force-decision rule for the image interpretation, so there is no missing data for the evaluation. As shown in Table 4, the kappa value for the 92 subjects is 0.76 with a lower bound of 0.70, and the percent of five readers agree with each other is 74%. **This is consistent with the sponsor’s finding on reader agreement.**

There is no MCI subject in the study population for FBB-01_01_13. Reader agreement cannot be evaluated for MCI subjects in this study.

Table 4: Inter-reader agreement and percent of agreement

Subject Group by Cognitive and Standard of Truth (SoT)	Positive Scans, n ^a	Kappa (95% CI)	Percent of Scans with Inter-reader Agreement		
			3 of 5 readers agreed	4 of 5 readers agreed	5 of 5 readers agreed
92 subjects (82 autopsy + 10 YHVs)	59	0.76 (0.70-0.81)	11	15	74
82 images from subjects with autopsy collected in 14595, 16034, and FBB-01_01_13	57	0.75 (0.67, 0.83)	7	14	78

^aShown is the median number of scans interpreted as positive across the 5 readers for each subgroup of subjects listed in the first column.

Intra-reader agreement

Intra-reader reproducibility analysis showed that, between the two readings for each of the 20 duplicate patient images, two readers had discordant reads for a single image, one reader (reader 2) had discordant reads for three images. The percent of agreement for all the 20 images is 100%, 85%, 95%, 95%, and 100% for reader 1, 2, 3, 4, and 5, respectively.

3.3 Evaluation of Safety

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

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

In this Section, subgroup analyses as exploratory analyses were only presented for Study FBB-01_01_13. Only reviewer's results are presented below.

Note that no conclusions can be drawn from the subgroup analyses due to lack of representation and limited sample size. Further exploration should be conducted for some subgroups of interest.

4.1 Gender, Race, and Age

Sensitivity and specificity were explored by Race, gender, and age (>75 or not) among the autopsy subjects (without the 10 YHVs). Confidence intervals are not calculated because of the small sample sizes. The results are shown in Table 5.

For the five readers, sensitivity values are 97-100% for Male and 83-100% for Female. Specificity values are 33-78% for Male and 67-92% for Females (more variation is observed in Males). Sensitivity values are 92-100% for Whites and 88-100% for Asian. Specificity values are 50-83% for Whites and 42-75% for Asian (lower level of specificity for Asian is observed).

Sensitivity values are 100-100% for subjects with age≤75 (n=25) and 93-100% for subjects with age>75. Specificity values are 38-69% for subjects with age≤75 and 47-88% for subjects with age>75 (lower level of specificity values is observed for subjects with age ≤ 75).

The performance is different in different subgroups. More studies with large sample size should be conducted for reaching a conclusion.

Table 5: Sensitivity (sen) and specificity (spe) by gender, race and age (in %) (for the 82 subjects with autopsy).

	sen, spe	n_sen, n_spe	reader 1	Reader 2	Reader 3	Reader 4	Reader 5
sex	Male	29, 18 (T=47)	97, 72	100, 33	97, 78	97, 72	100, 44
	female	23, 12 (T=35)	91, 92	96, 67	83, 83	96, 83	100, 75
race	White	36, 18 (T=54)	94, 83	97, 50	92, 83	94, 83	100, 56
	Asian	16, 12 (T=28)	94, 75	100, 42	88, 75	100, 67	100, 58
age	<=75	12, 13 (T=25)	100, 69	100, 46	100, 69	100, 69	100, 38
	>75	40, 17 (T=57)	93, 88	93, 47	88, 88	95, 82	100, 71

Note: pos is for positive cases, neg is for negative cases based on histopathology. n_sen is the sample size for sensitivity evaluation, and n_spe is the sample size for specificity evaluation. T is the total sample size (n_sen+ n_spe).

Inter-reader agreement was evaluated by subgroups among the 82 subjects with autopsy. The kappa values for different subgroups (Table 6) are similar (0.71-0.78) for the subgroups except the age<=65 group with sample size of 5. Kappa is not a proper statistic for such small sample size (i.e. 5). The values of percent of agreement (all five readers agree with each other) for different subgroups are 71-84% except the age<=65 group. The percent of agreement (all five readers agree with each other) is 89% for the age<=65 group. We observe higher agreement among readers for White compared with Asian, for young subjects compared with older ones.

Table 6: Reader agreement evaluation by gender, race, and age groups

Subject Groups	Positive Scans, n ^a	Kappa (95% CI)	Percent of Scans with Inter-reader Agreement		
			3 of 5 readers agreed	4 of 5 readers agreed	5 of 5 readers agreed
Gender					
Male (n=47)	33	0.71 (0.62, 0.80)	11	11	79
Female (n=35)	24	0.78 (0.67, 0.88)	3	20	77
Race					
White (n=54)	37	0.76 (0.68, 0.85)	9	9	81
Asian (n=28)	20	0.71 (0.59, 0.82)	4	25	71
Age group					
<=65 (n=9)	5	0.91 (0.70, 1.00)	0	11	89
>65 (n=73)	52	0.71 (0.64, 0.79)	8	15	77
<=75 (n=25)	16	0.79 (0.67, 0.92)	12	4	84
>75 (n=57)	41	0.72 (0.64, 0.80)	5	19	75

^aShown is the median number of scans interpreted as positive across the 5 readers for each subgroup of subjects listed in the first column.

4.2 Other Special/Subgroup Populations

The accuracy in term of sensitivity and specificity were evaluated by baseline clinical diagnosis.

As shown in Table 7, most subjects with autopsy are ADs. The specificity values for AD subjects are 34-69% for the five readers (Table 7). The confidence intervals were not calculated because of the small sample sizes.

Table 7: Sensitivity (sen) and specificity (spe) point estimates by clinical diagnosis (in %), without the 10 YHVs

sen, spe	N	reader 1	2	3	4	5
AD	44, 16 (T=60)	95, 63	98, 44	91, 69	98, 63	100, 34
DLB	2, 2 (T=4)	100, 100	100, 00	100, 100	100, 100	100, 50
DEM	2, 7 (T=9)	100, 100	100, 43	100, 86	100, 86	100, 71
HV	4, 5 (T=9)	75, 100	100, 80	75, 100	75, 100	100, 100

AD: Alzheimer's disease. MCI: Mild cognitive impairment. YHV: young healthy volunteer. HV: healthy volunteer DEM: other dementia. DLB: dementia with Lewy bodies.

As shown in Table 8, the accuracy in term of sensitivity and specificity was evaluated by study parts (14595, 16034, FBB-01_01_13) with 31, 23, 28 brains, 54 brains, and 82 brains, respectively. The accuracy for Study 16034 is also presented for comparison purpose (only images of 54 subjects were read by five readers in Study 16034). Note that the five readers in Study 16034 and FBB-01_01_13 are different. Specificity values are in bold.

The reader performance (FBB-01_01_13) is not consistent for the brains collected in different study period (14595, 16034, and FBB-01_01_13). The specificity values (point estimates) of the five readers are extremely low for the 23 brains collected in Study 16034 (56%, 33%, 67%, 56%, and 33%). The performance for the first collection (31 brains collected in Study 14595) is the best among the three collections (sensitivity values 89%-100%, specificity values as 92%, 58%, 92%, 92%, 75%).

Comparing the reader performance between FBB-01_01_13 and Study 16034 (same training, same SoT, different five readers), FBB-01_01_13 has more variation than Study 16034 (especially for specificity). For the analyses with 31 brains, the sensitivity values are 90-100% for FBB-01_01_13, and 84-100% for Study 16034; the specificity values are 58-92% for FBB-01_01_13, and 67-83% for Study 16034. For the analyses with 54 brains, the sensitivity values are 94-100% for FBB-01_01_13, and 85-100% for Study 16034; the specificity values are 48-81% for FBB-01_01_13, and 57-76% for Study 16034.

Table 8: Sensitivity (sen) and specificity (spe) point estimates by study part (in %)

Study FBB-01_01_13

sensitivity, specificity	N	reader 1	2	3	4	5
31 brains	19, 12	89.5, 91.7	100, 58.3	89.5, 91.7	94.7, 91.7	100, 75.0
54-31=23 brains	14, 9	100, 55.6	100, 33.3	100, 66.7	100, 55.6	100, 33.3
82=54=28 brains	19, 9	94.7, 88.9	94.7, 44.4	84.2, 77.8	94.7, 77.8	100, 55.6
54 brains	33, 21	93.9, 76.2	100, 47.6	93.9, 81.0	97, 76.2	100, 57.1
82 brains	52, 30	94.2, 80.0	98.1, 46.7	90.4, 80.0	96.2, 76.7	100, 56.7

Study 16034

	N	1a	2a	3a	4a	5a
31 brains	19, 12	100, 75.0	100, 66.7	94.7, 75.0	89.5, 66.7	84.2, 83.3
54 brains	33, 21	100, 66.7	100, 61.9	97.0, 66.7	93.9, 57.1	84.9, 76.2

The reader agreement (Table 9) was evaluated by clinical diagnosis and study parts (14595, 16034, FBB-01_01_13) with 31, 23, and 28 brains, respectively. Note that kappa is not proper for subgroups with small sample sizes.

The percent of agreement (all five readers agree with each other) is 83%, 78%, 50%, and 56% for ADs, HVs, DLBs, and DEMs. The AD group has higher level of agreement.

The agreement for the first collection (31 brains) and the second collection (23 brains) are similar (kappa values 0.76-0.78). The second collection (23 brains) has 87%, the highest level of percent of agreement (all five readers agree with each other). Recall that the accuracy (Sp) is low for this subset of 23 brains, the good agreement on incorrect interpretations for the 23 brains suggests possible training issues or inadequate interpretation method. The reader agreement is worst in the third collection (28 brains), with kappa value of 0.69, and only 71% of cases the five readers agree with each other.

Table 9: Reader agreement subgroup analyses by study part and clinical diagnosis

Subject Group by Cognitive and Standard of Truth (SoT)	Positive Scans, n ^a	Kappa (95% CI)	Percent of Scans with Inter-reader Agreement		
			3 of 5 readers agreed	4 of 5 readers agreed	5 of 5 readers agreed
31 brains collected in 14595	19	0.76 (0.65, 0.87)	10	13	77
23 brains in 16034	18	0.78 (0.65, 0.91)	9	4	87
28 brains in FBB-01_01_13	20	0.69 (0.57, 0.81)	4	25	71
54 brains collected in 14595 and 16034	37	0.77 (0.69, 0.85)	9	9	81
Baseline clinical diagnosis					
AD (n=60)	49	0.73 (0.65, 0.81)	5	12	83
DLB (n=4)	2	0.45 (0.14, 0.76)	25	25	50
DEM (n=9)	3	0.58 (0.38m 0.79)	11	33	56
HV (n=9)	3	0.77 (0.56, 0.98)	11	11	78

^aShown is the median number of scans interpreted as positive across the 5 readers for each subgroup of subjects listed in the first column. *AD*: Alzheimer's disease. *MCI*: Mild cognitive impairment. *YHV*: young healthy volunteer. *HV*: healthy volunteer *DEM*: other dementia. *DLB*: dementia with Lewy bodies.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

- The presence of the 10 YHVs may inflate the performance in terms of specificity. In addition, it is not clear how the 10 YHVs were selected. Excluding the YHVs, the sample size is 30 for specificity evaluation.
- The sponsor proposed to use BSS+IHC for SoT determination. Prior NDA submissions all used BSS only, to determine the SoT in the primary analyses.

- The sponsor used normal approximation to obtain the confidence intervals of sensitivity and specificity. Exact confidence intervals will be more proper in the case of the small sample sizes (i.e., for specificity evaluation with 30 brains).
- The patterns of the reader performance (the five readers in the new read study) for the brains collected in Study 14595 (31 brains), Study 16034 (23 brains), and the new read study (28 brains) are different.
- The patterns of the reader performance of the five readers in the new read study are different compared with the patterns of the five readers in Study 16034 (31 brains and 54 brains).
- Two readers performed much worse than the other three readers in the new read study (the sponsor, Piramal, tried to explain the variation by saying that the two readers didn't follow the instructions properly in spite of Web-based training).
- The re-use of the images for Study 14595, 16034, and FBB-01_01_13 may introduce bias on the reader performance.

The Agency did not approve the inclusion of 10 YHVs, the use of BSS+IHC for SoT determination in the sensitivity and specificity primary analyses (see details in COR-NDAIR-10(General Advice Letter) dated on 10/18/2013).

5.2 Collective Evidence

For the analyses using SoT from BSS only and without the 10 young healthy volunteers (YHVs), the lower bounds of the exact 95% confidence intervals (CIs) of sensitivity are >0.7 for all five readers. The lower bounds of the 95% CIs (specificity) are >0.5 for three readers (<0.6 for three readers). Three out of five reader had the lower bounds of sensitivity >0.6 and the lower bounds of specificity >0.5 (for the same reader). The pre-specified criteria for the primary analyses were met, which is consistent with the sponsor's findings. However, two readers had inadequate performance in terms of specificity (the lower bounds of the CIs are 28% and 37%, less than the pre-specified threshold). With a moderate prevalence, high sensitivity indicates high negative predictive value (NPV), and high NPV indicates that the drug product is clinical useful in ruling out subjects without AD. However, one reader can achieve very good sensitivity by sacrificing the level of specificity, which is not acceptable.

In terms of inter-reader agreement among the 92 subjects (82 autopsy plus 10 YHVs), the kappa value for the 92 subjects is 0.76 with a lower bound of 0.70, and the percent of five readers agree with each other is 74%. This is consistent with the sponsor's finding on reader agreement. In terms of intra-reader agreement, the percent of agreement for all the 20 images is 100%, 85%, 95%, 95%, and 100% for reader 1, 2, 3, 4, and 5, respectively.

Note that the brains (31, 23, and 28) were collected in three periods (Study 14595 period, Study 16034 period, and Study FBB-01_01_13 period). The analyses by the collection show that the reader performance is not the same in different collections. The specificity values (point estimates) of the five readers are extremely low for the 23 brains collected in Study 16034 (56%, 33%, 67%, 56%, and 33%). The performance for the first collection (31 brains collected in Study 14595) is the best among the three collections (sensitivity values 89%-100%, specificity values as 92%, 58%, 92%, 92%, 75%).

Comparing the reader performance between FBB-01_01_13 and Study 16034 (same training, same SoT, different five readers), FBB-01_01_13 has more variation than Study 16034 (especially for specificity). For the analyses with 31 brains, the specificity values from the five readers are 58-92% for FBB-01_01_13, and 67-83% for Study 16034. For the analyses with 54 brains, the specificity values are 48-81% for FBB-01_01_13, and 57-76% for Study 16034.

5.3 Conclusions and Recommendations

The statistical results in terms of accuracy (sensitivity and specificity) and reproducibility provide evidence to support the claim for the detection of β -amyloid in the brain proposed in this NDA. There are no MCI subject data, the population most needing the diagnostic method, in Study FBB-01_01_13. Also, there are extremely low values of specificity for the two readers with poor performance, different results in the different collections of brains, and different results in terms of sensitivity and specificity between Study 16034 and FBB-01_01_13. We conclude that the data and analyses support approval, noting the limitations of the studies as described above.

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
01/31/2014

JYOTI ZALKIKAR
01/31/2014

I concur with the primary reviewer and support approval of Florbetaben for amyloid detection. The only limitation of this NDA is that there no patients with Mild Cognitive Impairment with autopsy confirmation, a extremely restraining factor in interpreting the performance characteristics fully (sensitivity and specificity) in the most applicable patient population.

THOMAS E GWISE
01/31/2014



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 #: NDA 204-677

Supplement #:

Drug Name: Neuraceq (Florbetaben)

Indication(s): *for the detection of β -amyloid in the brain, thereby assisting in the differential diagnosis in adult patients who are being evaluated for Alzheimer's disease and other causes of cognitive decline.*

Applicant: Piramal Imaging

Date(s): date submitted 12/21/2012
PDUFA due date 8/23/2013

Review Priority: Standard

Biometrics Division: Division of Biometrics V

Statistical Reviewer: Lan Huang, Ph.D.

Concurring Reviewers: Jyoti Zalkikar (Team Leader)

Medical Division: Division of Medical Imaging Products

Clinical Team: Brenda Ye, M.D. (clinical reviewer)
Alex, Gorovets (clinical team leader)
Libero Marzella (clinical division director)

Project Manager: Sharon Thomas

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

Table of Contents

1. EXECUTIVE SUMMARY	4
2. INTRODUCTION	7
2.1 OVERVIEW.....	7
2.1.1. CLASS AND INDICATION	7
2.1.2. HISTORY OF PROGRAM DEVELOPMENT	7
2.1.3. SPECIFIC STUDIES REVIEWED.....	8
2.1.4. MAJOR STATISTICAL ISSUES.....	10
2.2 DATA SOURCES	11
3. STATISTICAL EVALUATION	12
3.1 DATA AND ANALYSIS QUALITY	12
3.2 EVALUATION OF EFFICACY	13
3.2.1 <i>Study Design and Endpoints</i>	13
3.2.2 <i>Statistical Methodologies</i>	20
3.2.3 <i>Patient Disposition, Demographic and Baseline Characteristics</i>	23
3.2.4 <i>Results and Conclusions</i>	25
3.3 EVALUATION OF SAFETY	31
4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	31
4.1 GENDER, RACE, AND AGE	32
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS	33
5. SUMMARY AND CONCLUSIONS	34
5.1 STATISTICAL ISSUES.....	34
5.2 COLLECTIVE EVIDENCE.....	36
5.3 CONCLUSIONS AND RECOMMENDATIONS	37

LIST OF TABLES

Table 1: Summary of pivotal clinical efficacy studies 14595 and 16034.....	9
Table 2: Rules for the assessment of BAPL by the independent blinded reader.....	16
Table 3: PET assessment and SoT determination in Study 14595 and 16034.....	19
Table 4: Patient disposition (% in ()) for Study 16034 for different patient populations.....	24
Table 5: Regional-level sensitivity and specificity (%) for all regions by reader and majority read	27
Table 6: Regional-level sensitivity and specificity (%) by region and by reader.....	28
Table 7: Subject-level analyses with two approaches proposed by the sponsor.....	30
Table 8: Inter-reader agreement and percent of agreement.....	30
Table 9: Sensitivity and specificity with web-training PET assessment for 54 autopsy subjects and the 10 YHVs.....	31
Table 10: Sensitivity (sen) and specificity (spe) by gender, race and age (in %)......	32
Table 11: Reader agreement evaluation by gender, race, and age group (<=65 and >65).....	33
Table 12: Sensitivity (sen) and specificity (spe) point estimates by clinical diagnosis (in %).	33
Table 13: Reader agreement subgroup analyses by baseline clinical diagnosis.....	34

1. EXECUTIVE SUMMARY

The medical product under development is Florbetaben F18 (Neuraceq). Neuraceq is indicated for the detection of β -amyloid in the brain, thereby assisting in the differential diagnosis in adult patients who are being evaluated for Alzheimer's disease and other causes of cognitive decline. A negative scan indicates sparse to no neuritic plaques, and is inconsistent with a neuropathological diagnosis of AD at the time of image acquisition; a negative scan result reduces the likelihood that a patient's cognitive impairment is due to AD.

This NDA submission included 10 studies in the clinical development program. Data were collected from 944 subjects enrolled in the studies, including **6 Phase 1 studies** (n=197) for PK/PD evaluation (A42404, Study 310863 (A35694), Study 311722 (A40922), Study 91790 (A42441), Study 312161 (A41147), and Study 312043(A50622)), **2 Phase 2 studies** (n=531, Study 311741 (A45264) for sensitivity (sen) and specificity (spe); Study 14311 (A51672) for subjects with Down Syndrome), and **2 Phase 3 studies** (Study 14595 (A47592): autopsy study (n=216), Study 16034 (A45264): pool read study). Note that the A numbers were assigned by the previous developer Bayer and the Study numbers were assigned by the current sponsor Priamal.

The pool read study [Study 16034] enrolled no subjects but analyzed images from other six studies to assess the effectiveness of an electronic program for training clinicians in the appropriate interpretation of [18F] florbetaben PET images. The total number of images is 461 plus 46 images for re-read. Studies 14595 and 16034 were selected for full statistical review and evaluation. Note that the in-person training approach used in the all studies except Study 16034 will not be used in clinical practice.

In Study 14595, regional-level comparison (sensitivity and specificity) between PET assessment and Standard of Truth (SoT) using majority read were conducted in the primary analyses (*by sponsor*). However, regional-level analyses and majority read will not be used in clinical practice. The sponsor claimed that the confidence intervals are calculated taking within brain dependencies into account using normal approximation (by Rao and Scott 1992 method). However, the method is not proper because of the small number of clusters (brains) for sensitivity and specificity evaluation. The primary analyses on regional-level images includes 60 (6 regions \times 10 Young Healthy Volunteers (YHVs)) negative regions assuming the regions in the YHVs are negative (44% of the total negative regions), which will lead to inflated specificity values for regional-level analyses. Two approaches for obtaining subject-level results were used by the sponsor, but none of them will be used in future clinical practice. In addition, only end-of-life subjects were enrolled, who do not belong the intended patient population.

In Study 16034, the images in the pool-read study were selected from some earlier studies (six studies), but not all earlier studies. There may be a selection bias. The majority of the subjects in the study are subjects with Alzheimer's dementia (182/461=40%), HVs (186/461=41%). The number of the intended patient population (Mild cognitive impairment (MCI) subjects) is only 51 (11%) out of the total 461 subjects.

For both **Studies 14595 and 16034**, subject-level specificity evaluation includes 10 YHVs, which is 10/24 =42% of the total negative cases with SoT. The presence of large YHVs may

inflate the performance in terms of specificity. Excluding the YHVs, the sample size is 14 for specificity evaluation (small), which will lead to wide confidence intervals. The number of negative cases (by SoT) among the 31 brains (obtained in Study 14595) is 8 by histopathology pathology consensus panel (CP) in Study 14595, 14 by onsite neuropathological diagnosis in Study 14595, and 10 by CP in Study 16034. According to the protocol, the rules are the same for the CP approach in Studies 14595 and 16034. The discrepancy on SoT between Study 14595 and 16034 cannot be explained. For both studies, the sponsor used normal approximation to obtain the confidence intervals of sensitivity and specificity at subject-level. Exact confidence intervals will be more proper in the case of the small sample sizes.

There is no agreement on the proposed primary analyses on regional-level (Study 14595) between the sponsor and the Agency (see details in REV-BIOMETRICS-01(General Review) for IND 78868 with submit date as 10/3/2011 and COR-MEET-03 with submit date as 6/16/2011). For the pool-read study, the Agency did not approve the inclusion of 10 YHVs in the sensitivity and specificity analyses and the use of normal approximation for the calculation of the confidence intervals (see details in IND 78868 information request or advice, communication ID COR-INDAD-02).

In the NDA submission, the sponsor claimed that the combined hypothesis for Study 14595 that sensitivity is $\leq 60\%$ or specificity is $\leq 80\%$ could be rejected (co-primary analyses). However, the statistical method used in the primary analyses is not proper. The sponsor also claimed that the primary endpoint exceeded the pre-specified kappa value threshold of 0.6 (for the lower bound of the two sided 95% CI) showing a value of 0.787 (CI: 0.750 – 0.824) across all five readers for Study 16034. However, the sponsor also claimed that the combined hypothesis that sensitivity is smaller or equal to 0.6 and specificity is smaller or equal to 0.7 (specified as secondary analyses in the protocol) could only be rejected for one out of the 5 readers. Thus the goal to reject this hypothesis for at least 3 out of the 5 readers was not met.

At regional-level (Study 14595), with the 10 YHVs, assuming independent regions within each brain (best scenario), two out of three lower bounds of 95% CI are more than 0.7 for sensitivity, and all three readers had specificity 95% CI lower bound >0.8 . The specificity point estimates without YHVs are all lower than the ones with YHVs (about 6% less) and the lower bound of the 95% CIs are between 0.7 and 0.8 for the three readers. **However, at subject-level** (by Approach 1: subject-level PET assessment and SoT were obtained using regional level PET assessment and regional histological findings from the pathology consensus), the lower bounds of 95% CIs for specificity with the 10 YHVs were 47%, 52%, and 65% for the three readers respectively. The rule for collapsing from regional assessment and SoT to subject assessment and SoT is that if one region in a brain is positive, the whole brain is positive. This rule applies to both PET assessment and SoT.

The sponsor used another approach to obtain the **subject-level sensitivity and specificity** for the 31 brains and 10YHVs. The in-person training procedure (PET assessment for the whole brain) used the rules for future clinical practice. SoT for the whole brain was obtained by on-site neuropathological diagnosis. The point estimate of the sensitivity is 100% for all three readers with 95% CI as (80%, 100%), and two out of three readers had lower bound of 95% CIs for

specificity >0.7. The better performance using Approach 2 vs. 1 may be due to the use of the new rule for reader training, or the change of SoT determination.

The **pool-read study** evaluated the Approach 2 PET assessment with web-training. The SoT values were obtained using Approach 1 regional histological findings from CP. From the definitions, the SoT for the pool-read study should be the same as the one used in Approach 1 (in Study 14595). However, data exploration shows that the SoT for Approach 1, Approach 2 in Study 14595, and the SoT in Study 16034 are all different. The number of positive cases is 23 by SoT in Approach 1, 17 by SoT in Approach 2, and 21 by SoT in the pool-read study, respectively; the number of negative cases is 8 by SoT in Approach 1, 14 by SoT in Approach 2, and 10 by SoT in the pool-read study, respectively. The rules for determining the SoT by the sponsor in the pool-read study are not clear.

In the pool-read study (using the web-training and PET assessment rules same as Approach 2 and SoT using Approach 1 described in Study 14595), with the 10 YHVs, the lower bound of the 95% CIs (sensitivity) are >0.7 for four readers, and between 0.6 and 0.7 for one reader. The lower bound of the 95% CIs (specificity) are <0.6 for three readers. Without the 10 YHVs, there are only 14 negative cases for specificity evaluation. The point estimates of the specificity without 10 YHVs are less than those with the 10YHVs. Because of the small sample size, the CIs are very wide. In addition, the specificity values for subgroup AD (0.38-0.75 for the five readers) are much lower than the specificity values of the HVs (0.62-1.00 for the five readers with and without 10 YHVs).

Since the subjects with SoT do not include the intended patient population (MCI subjects), the reader agreement was evaluated among a broad range of subjects in the pool read study. **For inter-reader agreement**, the kappa value for the primary analysis in **Study 16034** is 0.799 (7 subjects with missing reads were removed), and the percent of five readers agree with each other is 78%. This is consistent with the sponsor's results on the inter-reader agreement (primary analysis). The inter-reader agreement for MCI subjects was evaluated with kappa =0.75. In addition, intra-reader agreement ranged from 91% to 98% among the five blinded readers.

The statistical results in terms of accuracy (sensitivity and specificity) and reproducibility provide very limited evidence to support the claim for the detection of β -amyloid in the brain proposed in this NDA. The reader agreement is good in terms of intra-reader and inter-reader agreement, which indicates that the results are reproducible. However, the low specificity (indicating high false positive images) for subjects with autopsy in Study 16034 using web-training process should be noted. Agreement on incorrect interpretations suggests an inadequate interpretation method. In addition, the rules for obtaining the subject-level PET assessment and SoT are not the same in the two pivotal studies. Particularly, it is not clear how the sponsor determined the SoT using the CP histopathology. The performance varies in different studies with different training processes and different approaches for obtaining the subject-level results. There are no MCI subject data, the population most needing the diagnostic method, with autopsy information. Conclusive evidence for the performance in terms of sensitivity and specificity cannot be obtained on MCI population.

2. INTRODUCTION

2.1 Overview

2.1.1. Class and Indication

Florbetaben is a diagnostic radiopharmaceutical product developed for use with positron emission tomography (PET) imaging for the visual detection of fibrillar amyloid in the form of neuritic plaques in the brain. It is in the same pharmacologic class as Amyvid™ (florbetapir), a product approved in the United States for the same indication. The proposed commercial name for Florbetaben F 18 Injection is Neuraceq. The proposed Indications and Usage statement for Florbetan (by the sponsor) is:

[REDACTED] (b) (4)

2.1.2. History of Program Development

The first clinical trial authorization for florbetaben to be used in a company-sponsored clinical trial was granted in Germany on 22 MAR 2007. An investigational new drug application (IND 78,868) for florbetaben was submitted to the US FDA and has been active since 28 JUL 2008.

Additional meetings with the FDA during 2009-2011 were aimed at further discussions on the appropriate SoT for the verification of efficacy, the statistical analysis plan (SAP) for the Phase 3 pivotal histopathology study, and the impact of the FDA Advisory Committee Meeting (held 20 JAN 2011) on the submission strategy for florbetaben. The sponsor obtained the FDA's advice on the Phase 3 "pool-read" study in a Type B meeting (12 Dec 2011).

A pre-NDA meeting was held on [August 24, 2012] at which Primal presented the clinical documentation for supporting an NDA submission for Florbetan F18 Injection.

Note that before the submission of this NDA, the sponsor was Bayer (for IND 78868). The current sponsor for this NDA 204677 is Priamal, who purchased the product from Bayer.

The 10 studies in the clinical development program for Florbetaben were designed to provide safety and efficacy data for worldwide regulatory approval for the target indication presented above. Data were collected from 944 subjects enrolled in 9 clinical studies, including

- 6 Phase 1 studies (n=197) for PK/PD evaluation (Study 123456(A42404), Study **310863** (A35694), Study 311722 (A40922), Study 91790 (A42441), Study **312161** (A41147), and Study 312043(A50622)),
- 2 Phase 2 studies (n=531, Study 311741 (A45264) for sensitivity and specificity among 150 subjects with ≤ 5 microgram iv **in part A** and about 260 subjects with ≤ 50 microgram iv in part B; **Study 14311** (A51672) for subjects with Down Syndrome),

- 2 Phase 3 studies
 - Study 14595 (A47592): autopsy study (n=216),
 - Study 16034 (A45264): pool read study (n=461).

Note that the study numbers were assigned by the current sponsor Piramal. The A numbers were assigned by Bayer.

All phase II and III studies are open-label studies. Among the phase I studies, A42404, A35694, A50622 are open-label studies; A40922 and A42441 are placebo-controlled randomized studies; Study 312161 (A41147) is a randomized, cross-over study.

Populations studied in the clinical development program included healthy volunteers (HVs), subjects with Mild Cognitive Impairment (MCI), Alzheimer's disease (AD), other dementia, subjects with a life expectancy of 1 year or less (regardless of primary diagnosis), and others.

The pool read study (Phase 3 electronic reader training study [Study 16034]) enrolled no subjects but analyzed images from other 6 studies (all subjects from Studies 311741 part B, 312043, 14595; all subjects from Study 91790 and Study 311722 receiving florbetaben except one subject with major protocol deviation; randomly selected subjects from Study A42404) to assess the effectiveness of an electronic program for training clinicians in the appropriate interpretation of [18F]florbetaben PET images. The total number of images is 461 plus 46 images for re-read.

2.1.3. Specific Studies Reviewed

Studies 14595 and 16034 were selected for full statistical review and evaluation. Study 14595 is the autopsy study including the information of SoT from autopsy for end-of-life subjects. Sensitivity and specificity can be evaluated using the data obtained from Study 14595. Study 16034 evaluated the sensitivity and specificity of the web-training method for reader training, and included a broad range of subjects including the intended patient population (MCI subjects) for reader agreement evaluation. Note that only Study 16034 used the web-training approach in the whole clinical program.

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

Table 1: Summary of pivotal clinical efficacy studies 14595 and 16034.

	Phase and Design	Primary objective	Treatment period and follow-up	# of Subjects per Arm	Study Population
Study 14595	Phase 3, multi-center, open label single arm	To determine the sensitivity and specificity of the visual assessment of regional tracer compared to histological verification of the presence or absence of cerebral beta-amyloid in the respective postmortem specimens as the standard of truth (SoT).	each subject received a single IV injection of the study drug and scanning was performed from 90 to 110 minutes post-injection (pi). Each subject was asked to return to the site for a follow-up visit 20 to 28 hours after study drug administration and a telephone contact occurred 7 days thereafter.	216 administered florbetaben and scanned 32 underwent brain autopsy 31 evaluable in-person training	Subjects with short life expectancy 216 subjects (137 AD subjects, 31 other dementia subjects, 5 DLB subjects, 32 NDVs, and 11 HVs).
Study 16034	Phase 3, Single arm, open-label, pool read, no new enrollment	To assess the reproducibility of the visual assessment of PET scans from a patient population that closely represents the “future use” population via assessing the inter-reader agreement of the visual assessment results of 461 florbetaben PET scans pooled from various florbetaben clinical studies.	No subject was dosed in this study	507 images were read (461 other trials and 10% re-read) Five blinded readers Web-based training	182 (21) AD 188 (15) HV 51 (5) MCI 10 (2) DLB 12 FTLD 5PD 4VaD 3DEM 3(2) Other 3(1)CP diagnosis not available

Numbers in parenthesis indicate image re-reads for assessment of intra-reader agreement

EOL= end of life; IHC = Immunohistochemical; PET = positron emission tomography; SoT = Standard of truth;

AD: Alzheimer’s disease. MCI: Mild cognitive impairment. YHV: young healthy volunteer. HV: healthy volunteer DEM: other dementia. DLB: dementia with Lewy bodies. FTLD: fronto-temporal lobe degeneration (dementia). PD: Parkinson’s disease.

VaD: vascular dementia. NA: not available.

2.1.4. Major Statistical Issues

Study 14595

- In-person training for the readers was used, which will likely cause the study results to be biased in comparison to what would be expected in clinical practice using a web-training method.
- Regional-level comparison (for sensitivity and specificity) between PET assessment and SoT were conducted in the **primary analyses** (*by sponsor*). However, subject-level analyses are more appropriate with respect to clinical practice instead of regional-level analyses.
- The sponsor claimed that the confidence intervals are calculated taking within brain dependencies into account using normal approximation (by Rao and Scott 1992 method) However, the normal approximation theory discussed in Rao and Scott 1992 used the condition that the number of clusters should be large for normal approximation. But the total number of clusters (brains) in this study is only $31+10YHVs=41$. All the regions in the 10YHVs are assumed to be negative. Among the 31 autopsy subjects, some will have all six regions as positive or negative regions. The numbers of clusters for sensitivity and specificity evaluation are small (23 for sensitivity and 27 for specificity without the 10 YHVs and 37 with the 10 YHVs). The method proposed by the sponsor is not proper.
- Majority read were used in the primary analyses by the sponsor. However, in clinical practice, usually one reader interprets the images. By reader analyses will be more proper because the estimation using majority read data will not be representative of expected performance of single readers.
- The primary analyses on regional-level images included 60 (6 regions \times 10 YHVs) negative regions assuming all six regions in the YHVs are negative, which is $60/138=44\%$ of the total negative regions. This will lead to inflated specificity values.
- Two approaches for evaluating subject-level sensitivity and specificity were conducted by the sponsor. None of them will be used in future clinical practice.
- In addition, only end-of-life subjects were enrolled, who are not the intended patient population.

There is no agreement on the proposed primary analyses on regional-level (Study 14595) between the sponsor and the Agency (see details in REV-BIOMETRICS-01(General Review) for IND 78868 with submit date as 10/3/2011 and COR-MEET-03 with submit date as 6/16/2013).

Study 16034

- The images in the pool-read study were selected from some earlier studies, but not all earlier studies. There may be a selection bias.
- The majority of the subjects in the study are subjects with Alzheimer's dementia (182/461=40%), HVs (186/461=41%). The number of the intended patient population (MCI subjects) is only 51 (11%) out of the total 461 subjects.

For both **Studies 14595 and 16034**, subject-level specificity evaluation includes 10 YHVs, which is $10/24 = 42\%$ of the total negative cases with SoT (either from autopsy or the negative assumption for the YHVs). The presence of the 10 YHVs may inflate the performance in terms of specificity. Excluding the YHVs, the sample size is 14 for specificity evaluation (small), which will lead to wide confidence intervals.

The number of negative cases (by SoT) among the 31 brains (obtained in Study 14595) is 8 by histopathology pathology consensus panel (CP) in Study 14595, 14 by onsite neuropathological diagnosis in Study 14595, and 10 by CP in Study 16034. According to the protocol, the rules are the same for the CP approach in Studies 14595 and 16034. The discrepancy on SoT between Study 14595 and 16034 cannot be explained.

For both studies, the sponsor used normal approximation to obtain the confidence intervals of sensitivity and specificity at subject-level. Exact confidence intervals will be more proper in the case of the small sample sizes.

None of the six regions used in the regional-level analyses for PET assessment are the same as the four regions used in the subject-level analyses.

For the pool-read study, the Agency recommended against the inclusion of 10 YHVs in the sensitivity and specificity analyses and the use of normal approximation for the calculation of the confidence intervals (see details in IND 78868 information request or advice, communication ID COR-INDAD-02).

2.2 Data Sources

Most of the materials reviewed including the applicant study reports, data sets and literature referenced are provided electronically, and the full electronic path of the documents are <\\Cdsub1\evsprod\NDA204677\0000> and <\\Cdsub1\evsprod\NDA204677\0003>

The application study reports reviewed include Clinical overview, Summary of Clinical Efficacy, Summary of Clinical Safety in <\\Cdsub1\evsprod\NDA204677\0000\M2>.

Some reports, protocols and statistical analysis plans for 14595 and 16034 are in <\\Cdsub1\evsprod\NDA204677\0000\M5>,

Data sets analyzed for study **14595 and 16034** (with data definition document) were located in [\\Cdsesub1\evsprod\NDA204677\0000\M5](#). The format files are located in [\\Cdsesub1\evsprod\NDA204677\0003\M5](#).

The datasets analyzed include assess01.xpt, basic.xpt, effslv.xpt, effwbv.xpt, effregv.xpt, effslq.xpt for Study 14595; poolrd.xpt, assess01.xpt, basic.xpt for Study 16034; demo.xpt in the other individual studies.

The data do not follow ADAM and SDTM data format.

3. STATISTICAL EVALUATION

3.1 Data and Analysis Quality

It was not possible to reproduce the primary analysis datasets (particularly, the primary endpoint) from the original data source for Study 14595 and 16034. The raw data sets for SoT determination are not included in the study folders (in NDA submission).

A lot of effort is needed to process the data. There is no unique subject id in many data sets.

The following **statistical information requests (IRs)** were sent to the sponsor on 1/29/2013:

1. When we load your xpt files into sas, many variables do not have the right format. The error message is "Format was not found or could not be loaded". Please provide the format for all the variables in all the data submitted. You may need to resubmit all the xpt files to correct this problem. Failure to promptly resolve this problem may preclude our ability to review your application in a timely manner.
2. Provide the names of the data sets and the related sas programs used to generate the tables in the submission, especially the tables for the efficacy evaluation in studies 14595 and 16034.
3. The images to be assessed during the "pooled read" study (Study 16034) were chosen from various Phase 1 studies, the Phase 2 study (Part B) and the Phase 3 study. Provide a description of the criteria you used to select images/subjects for inclusion in the pooled read study. Were these criteria pre-specified in a manner that clearly identified which images/subjects would be included/excluded from the pooled read? If so, provide the documentation that verifies these details of the image/subject selection process. Also provide a table (or figure) that describes the Study 16034 subject distribution (by the study that originally enrolled the subject).
4. Regarding Study 14595, we have been unable to locate the pre-specified statistical analytical plan (SAP). Please identify the location of the SAP and/or submit this plan. We are particularly interested in the details of the interim analysis.

The sponsor answered the questions 3 and 4 in the applicant orientation meeting on 2/4/2013.

For question number 2, the sponsor submitted the sas programs for studies 14595 and 16034 on 2/12/2013 ([\\CDSESUB1\EVSPROD\NDA204677\0003](#)).

For question number 1, the sponsor submitted formats.xpt files for studies 14311, 14595, 16034, 311741, and 312043 ([\\CDSESUB1\EVSPROD\NDA204677\0003](#)).

The data do not follow ADAM and SDTM data format. Only the data sets in Study 16034 had unique patient id (UPID). The data sets in other studies only have patient id (PID), not unique patient id. The data sets in ISS do not have unique patient id.

One statistical Information Request (IR) was sent to the sponsor for the data with baseline clinical diagnosis and definitions of SoT in May, 2013. A data (diagsot.xpt) and the definitions of SoT were submitted and delivered to the reviewers by the project manager in May 2013.

The data from Study A42404 (Study 123456) is not included in the NDA submission (no folder for this study). The data sets in ISS also do not have the information for subjects from Study A42404. Some of the A42404 subjects were selected to be included in Study 16034 for pool-read study, but there is not baseline information for those subjects in Study 16034 data sets.

For Study 311722, the patient ids in the folder of Study 311722 are not matched with the ids in the pool-read study data sets (some patients from 311722 were selected for pool-read evaluation).

Subjects in Study 123456 (A42404) are not included in the ISS data sets submitted.

One statistical Information Request (IR) was sent to the sponsor for the data with basic demographical information for subjects enrolled from Study 123456 and 311722, and included in Study 16034, on July 30, 2013. A data (demodiag.sas7bdat) was submitted and delivered to the reviewers by the project manager in August 1, 2013.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

The reviewer's comments will be in italics in this Section.

Study 14595:

Study design:

Study 14595 was a pivotal open-label, non-randomized histopathology study to evaluate the efficacy and safety of florbetaben PET imaging for the detection/exclusion of cerebral β -amyloid when compared to post-mortem histopathology.

The radioactive dose used in this study was $300 \text{ MBq} \pm 20\%$, the specification of the tracer mass was $\leq 50 \mu\text{g}$. Due to the nature of the study, only male and female subjects with a short life expectancy who were willing to donate their brain after death and undergo **both a PET and an MRI scan** were eligible.

Subjects with clinically diagnosed or suspected AD or other non-AD dementia (DEM), as well as NDV were included in order to have representation of subjects with low and with high probability of cerebral β -amyloid deposition in the trial. The NDV cohort consisted mainly of subjects with end-stage carcinoma or cardio-vascular disease. To further enrich the population contributing to the specificity, 10 young cognitively normal HVs between 21 and 40 years of age, who served as negative controls were enrolled as well; for the SoT they were considered β -amyloid negative by default, but their PET scans were read independently in the same manner as the rest of the cohort.

The primary efficacy analysis was based on data from a pre-planned interim analysis after approximately 30 autopsy cases were available. A total of 218 subjects were assigned to treatment: 139 AD subjects, 31 DEM subjects, 5 DLB subjects, 32 NDVs, and 11 HVs. Two AD subjects who had enrolled did not receive study drug administration (total 137 AD subjects). Of the 216 subjects who were administered the study drug, a total of 32 brain specimens were collected of which 31 brains were evaluable for analysis (plus the 10 HVs with a PET scan image were included), and these data were reported in the interim report.

The study is still on-going and additional brains are being collected. These additional brain specimens were collected after the cut-off for the interim report in SEP 2011 up to 19 MAY 2012. The images from these 23 patients were included in the non-interventional pooled read analysis (Study 16034).

The sensitivity and specificity was determined with two different analyses: A regional (primary efficacy) analysis to determine the presence or absence of β -amyloid in the brain on a region-by-region basis and a subject-based analysis as part of the secondary efficacy parameters.

Co-primary endpoints: sensitivity and specificity of the visual assessment of *regional* tracer uptake in the florbetaben PET brain scans compared to the presence or absence of cerebral β -amyloid in the respective post-mortem histopathological brain specimens as the SoT.

The six regions used in the primary analyses are Middle frontal gyrus, Striate and parastriate areas of occipital cortex, Anterior cingulate cortex, Posterior cingulate / Precuneus, Hippocampus, and Cerebellar Cortex.

For the 10 young HVs without autopsy, the same regions were randomly mixed to be part of the scans to be visually assessed. There are a total of 246 scans with 244 evaluable scans (184 from the brains of the 31 subjects with autopsy and 60 scans from 10 young HVs).

Secondary endpoints:

- Quantitative assessment with generation of the *regional* and *subject level* SUVRs using the MR-segmented template.
- Sensitivity and specificity in detecting/excluding cerebral β -amyloid plaques on the basis of a “per-subject-analysis”.

Two different approaches (Standard of References [SORs]) were used to assess the sensitivity and specificity for the “per-subject-analysis”. These SoRs are briefly described below. In the following SoT will be used instead of SoR.

Two different approaches of PET assessment were also used and described below.

- The first analysis (Approach 1: composite whole brain) did not require additional PET reading, but was derived from the regional visual PET assessment results (collapsed to “yes” / “no” per subject) compared to the histopathology pathology consensus (CP) results collapsed to amyloid “present” / “not present” as the SoR. If one of the six regions is positive, the whole brain is positive.

The SoT in Approach 1 was the result of the histopathological analysis of the 6 pre-defined regions (see above) by a pathology CP of 3 experts in neuropathology blinded to all clinical information and to the PET scan. A brain region was considered to have ‘**relevant β -amyloid present**’, if the Pathology-CP judged it as having a final rating of “**moderate**” or “**frequent**” for **neuritic/cored or diffuse amyloid plaques** based on the Bielschowsky silver staining and immunohistochemistry for β -amyloid. The 'highest' score from the CP histopathological evaluation of the 6 pre-defined brain regions determined the composite “whole brain” regional histology result for this subject.

- The second subject-based analysis (Approach 2) determined the sensitivity and specificity of the visual assessment of florbetaben PET images on the subject level according to the visual assessment rules to be used in the future clinical application compared to the **on-site neuropathological diagnosis** as the SoR (beta-amyloid present or not) according to the classification by the CERAD. The PET read was performed by the same readers who read the regional PET scans.

The readers were trained by a tutor and validated on the basis of 20 scans that were not used as part of the study. Thereby, the PET decision of no tracer uptake (equals “normal”) or tracer uptake (equals “abnormal”) derived from the assessment of **four cortical brain regions, ie, the lateral temporal, the frontal, the parietal as well as the posterior cingulate cortex.**

The rule for future clinical application is shown in Table 2.

Table 2: Rules for the assessment of BAPL by the independent blinded reader

Brain beta-amyloid deposition score	Rules for brain beta-amyloid deposition assessment
1: without	Regional cortical tracer uptake score 1 in each of the 4 brain regions
2: moderate	Regional cortical tracer uptake score 2 in any or all of the 4 brain regions and no score 3 in these 4 regions
3: pronounced	Regional cortical tracer uptake score 3 at least in one of the four brain regions

BAPL=brain beta-amyloid plaque load

Standard of Truth (SoT) for exploration:

The following modification of the SOT will be used for exploration. A brain region will be considered to have ‘β-amyloid present’, if the CP judged it as having a final rating of “moderate” or higher for

1. any of the three β-amyloid pathologies (i.e. neuritic, diffuse or vascular) regardless of detection method used (Bielschowsky silver staining / immunohistochemistry) → SoT_all
2. for neuritic plaques only, regardless of the detection method, and not regarding the presence of diffuse or vascular amyloid → SoT_neur
3. diffuse plaques only, regardless of detection method, and not regarding presence of neuritic plaques or vascular amyloid → SoT_diff
4. vascular β-amyloid as per immunohistochemistry, and not regarding the presence of neuritic or diffuse amyloid plaques → SoT_vasc

Study 16034:

Study design:

By the very nature of the pivotal histopathology Study 14595 and the required co-primary endpoints, subjects recruited into the trial do not represent the “future use” population for a β-amyloid-targeted PET tracer. Likewise, the visual assessment algorithm for the primary regional assessment did not reflect the visual assessment recommended in clinical practice. Thus, the main purpose of the second pivotal trial, the “pooled” read Study 16034, was to test the reliability and reproducibility of the visual assessment method proposed for florbetaben PET in a cohort that was as close to the “future use” population as possible, and based on the reader training and visual assessment algorithm recommended for the future usage.

The “future use” population was expected to include subjects with suspected AD or other dementia, subjects with MCI and non-demented individuals with conditions that mimic dementia (eg, severe depression). Thus images to be assessed during this pooled read study were chosen from various Phase 1 studies, the Phase 2 study (Part B) and Phase 3 study and included a broad range of cohorts including subjects with probable/possible (mild to moderate) AD, fronto-temporal lobe degeneration (FTLD), vascular dementia (VaD), and DLB, subjects with MCI as well as both young HVs (< 40 years) and older cognitively normal HVs (> 55 years).

Primary endpoint: The primary efficacy variable for this study was the inter-reader agreement as determined by the kappa coefficient.

Secondary endpoint (related to primary): The secondary efficacy variable was the intra-reader agreement (10% re-read) also determined by the kappa coefficient.

Other secondary endpoints:

Re-determination of sensitivity and specificity of 272 images from Phase 2 (311741, Part B) to assess reliability of the visual assessment of PET scans after the introduction of a refined visual assessment algorithm and a computer (web)-based training tool). Standard of Reference: The CP confirmed clinical diagnosis.

Sensitivity and specificity in subjects from the autopsy study (Study 14595) for whom a post-mortem diagnosis was available by the time of study closure. Standard of Reference: Pathology CP diagnosis for Bielschowsky silver stain.

Readers:

For this pooled read study, the five blinded readers, trained on the visual assessment method by a computer (web)-based training tool and validated on the basis of 30 images (not otherwise included in the read), randomly assessed 461 images (and re-read 10%, or 46 images).

The visual assessment and scoring procedure

Four pre-specified brain regions were assessed. These regions included the lateral temporal and frontal lobes, the posterior cingulate cortex, and the parietal lobe.

After performing the systematic regional assessment and scoring the designated regions for regional tracer binding as described below, the reader was to provide an overall rating of the presence or absence of tracer uptake consistent with subject level BAPL as 1, 2, or 3 (rules in Table 2) as described below:

1= Scan without β -amyloid deposition

2= Scan with moderate β -amyloid deposition

3 = Scan with pronounced β -amyloid deposition

If the score is 2 or 3, the PET assessment for the brain is positive, otherwise negative.

Definition of subject based SoT

The SoT is obtained by using the respective histopathological findings (as established by the CP of pathologists): the subject level CP histopathological examination as determined in the histopathology Study 14595. It may occur that the CP histopathological evaluation of some

regions was not possible. A subject was included in this analysis if a CP diagnosis was available for at least 5 regions.

A summary of the PET assessment and SoT approaches is presented in Table 3. Note the different regions and rules used in different analyses and studies.

From the definitions, the SoT for the pool-read study should be the same as the one used in Approach 1 (in Study 14595). However, data exploration shows that the SoT for Approach 1, Approach 2 in Study 14595, and the SoT in Study 16034 are all different. The number of positive cases is 23 by SoT in Approach 1, 17 by SoT in Approach 2, and 21 by SoT in the pool-read study, respectively; the number of negative cases is 8 by SoT in Approach 1, 14 by SoT in Approach 2, and 10 by SoT in the pool-read study, respectively.

Including the additional 23 brains to the original 31 brains, the total number of positive cases is 40 (21+19), and the negative cases is 14 (10+4) by the SoT in pool-read study.

SoT should be consistent in Study 14595 and 16034 for confirmation. It will be difficult to evaluate the performance of the product with the mixture use of different SoTs and PET assessment rules.

Table 3: PET assessment and SoT determination in Study 14595 and 16034

		PET assessment		SoT	
14595	Regional-level analysis (primary)	6 regions: frontal, occipital cortex, hippocampus, anterior, posterior, cerebellar		6 regions	A brain region was considered to have 'relevant Beta-amyloid present', if the CP of neuropathology experts judged it as having a final rating of "moderate" or higher for neuritic or diffuse β -amyloid plaques based on the Bielschowsky silver staining.
	Subject-level approach 1 (composite whole brain)	6 regions	6 small brain regions were collapsed into a subject-based SoT. If one of the six regions is positive, the whole brain is positive.	6 regions	6 small brain regions was collapsed into a subject-based SoT. The 'highest' score from the CP histopathological evaluation of the 6 pre-defined brain regions determined the composite "whole brain" regional histology result for this subject. If the score is moderate or frequent, it is positive.
	Subject-level approach 2	4 regions: frontal cortex, posterior cingulate, lateral temporal, parietal cortex	Go from 4 regional scores to 1 brain score (see Table 2). 1 for negative, 2 and 3 for positive	No details in the submitted protocol	On-site neuropathological diagnosis
16034	Subject-level (primary)	4 regions: frontal cortex, posterior cingulate, lateral temporal, parietal cortex	Go from 4 regional scores to 1 brain score. Rules are in Table 2. 1 for negative, 2 and 3 for positive	6 regions	6 small brain regions was collapsed into a subject-based SoT. The 'highest' score from the CP histopathological evaluation of the 6 pre-defined brain regions determined the composite "whole brain" regional histology result for this subject. If the score is moderate or frequent, it is positive.

3.2.2 Statistical Methodologies

The reviewer's comments will be in italics in this Section.

Study 14595

Primary Efficacy Evaluation:

The co-primary efficacy variables were evaluated using majority results of the three independent Blinded Readers. This majority read value was determined based on the match to the SoT. If at least two readers matched the SoT, the majority reader response was considered a match. This majority read response was not a consensus read.

There are a total of 246 regional scans with 244 evaluable regional scans (184 from the brains of the 31 subjects with autopsy and 60 scans from 10 young HVs). For the 10 young HVs without autopsy, the same regions were randomly mixed to be part of the scans to be visually assessed.

The 95% CI were calculated for the majority read and for each blinded reader separately. The 95% CI were calculated taking brain dependencies into account using normal approximation (Rao and Scott 1992).

For sensitivity and specificity, the following hypotheses were formulated:

H0_sens: sensitivity \leq 0.6 vs H1_sens: sensitivity $>$ 0.6

H0_spec: specificity \leq 0.8 vs H1_spec: specificity $>$ 0.8

H0_sens was to be rejected if the lower bound of the two-sided 95% CI is larger than 0.6.

H0_spec was to be rejected if the lower bound of the two-sided 95% CI is larger than 0.8.

The study was considered a success, when both H0_sens and H0_spec could be rejected.

Regional-level evaluation is not clinical useful, which will only provide limited information on the relationship of the PET assessment and the SoT for regions. By reader analyses should be used instead of majority read in the primary analyses. Analysis assuming independent regional scans will be conducted in this review to obtain the results of the best scenario instead of models for correlation using normal approximation. By region analyses will also be explored, which reflect the worst case scenario (all six regions are highly correlated within brains and six regions are reduced to one region). There is no agreement between the sponsor and the Agency on this primary analyses proposed by the sponsor in the earlier interactions (REV-BIOMETRICS-01(General Review) for IND 78868 with submit date as 10/3/2011 and COR-MEET-03 with submit date as 6/16/2013).

The inclusion of the 10 YHVs in the analyses will lead to inflated specificity values.

The sponsor claimed that the confidence intervals are calculated taking within brain dependencies into account using normal approximation (by Rao and Scott 1992 method)

However, the normal approximation theory discussed in Rao and Scott 1992 used the condition that the number of clusters should be large for normal approximation. But the total number of clusters (brains) in this study is only 31+10YHVs=41. All the regions in the 10YHVs are assumed to be negative. Among the 31 autopsy subjects, some will have all six regions as positive or negative regions. The numbers of clusters for sensitivity and specificity evaluation are small (23 for sensitivity and 27 for specificity without the 10 YHVs and 37 with the 10 YHVs). The method proposed by the sponsor is not proper.

Secondary Efficacy Evaluations:

Sensitivity and specificity at subject-level are considered as secondary endpoints.

The Agency recommended these as primary endpoints during the communication process with the sponsor (see details in REV-BIOMETRICS-01 (general review) for IND 78868 with submit date as 10/3/2011 and COR-MEET-03 with date as 6/16/2013). However, there are two approaches used to determine the subject-level PET assessments and SoTs (described in the study design section), and the subject-level analyses are secondary analyses proposed by the sponsor. Both approaches will be explored in the review.

Important post hoc analyses

After the originally planned analyses became available and the results were reviewed, the following additional post hoc sensitivity analyses were conducted for the FAS.

- Sensitivity and specificity, analogous to the primary efficacy variable, by blinded read
- session
- Sensitivity and Specificity, analogous to the primary efficacy variable
 - excluding Region Hippocampus
 - excluding Region Cerebellar
 - excluding both above regions
- Sensitivity and specificity, analogous to the primary efficacy variable, based on
- different SoTs.

Post Hoc analyses excluding both Hippocampus and Cerebellar cortex, and with different SoTs will be explored in this review.

Study 16034

Primary Endpoint and analyses:

Inter-individual kappa (κ_{inter}) value across all readers will be analyzed. The kappa value κ_{inter} across the 5 blinded readers for the binary assessment normal/abnormal on the subject level was calculated over all images read. The CI was calculated based on an asymptotic variance estimate.

The hypothesis to be tested was: H_0 , inter: $\kappa_{\text{inter}} \leq 0.6$ vs H_1 , inter: $\kappa_{\text{inter}} > 0.6$. The hypothesis was rejected, when the lower bound of the confidence interval for κ_{inter} was larger than 0.6.

Normal/Abnormal should be replaced by negative/positive.

Secondary Analyses

- Inter-individual kappa value for all 10 reader pairs and intra-individual kappa values based on the re-reads separately for all 5 BRs:
- Sensitivity and specificity with histopathology as SoT based on the majority read:
 - Sensitivity and specificity were assessed for each of the 5 readers in the 54 postmortem subjects from the Study 14595 with available histopathology as SoT, enriched by the results from 10 HVs without autopsy for whom amyloid pathology in the brain is assumed negative by default. Corresponding 95% CI were calculated. The following combined hypotheses were tested:
 H_0 , sens: sensitivity ≤ 0.6 vs H_1 , sens: sensitivity > 0.6
 H_0 , spec: specificity ≤ 0.7 vs H_1 , spec: specificity > 0.7
The combined hypotheses were rejected if the lower limits of the 95% CI for sensitivity and specificity were higher than the thresholds of 0.6 and 0.7 respectively for at least 3 out of the 5 BRs.
 - A descriptive sub-analysis was performed excluding the 10 HVs.
 - Sensitivity and specificity with clinical diagnosis as SoT based on the majority read:
 - Sensitivity and specificity were assessed with the CP clinical diagnosis as the SoR in images from Part B of the 311741 study. In this analysis, subjects with a CP confirmed clinical diagnosis of AD served as amyloid-positive, and CP confirmed cognitively normal HVs as negative controls.

By reader analyses will be conducted instead of majority read analyses in this review. Analyses using SoT or SoR from clinical diagnosis will not be explored in this review.

Missing data/not assessable scans

“Not assessable” scans were treated using a worst case imputation for the purpose of the calculation of kappa.

The images that were “not assessable” were classified as mismatches to the SoT (ie, EN for an image with a positive SoT and FP for an image with a negative SoT) for the respective reader.

The sponsor did not conduct forced decision when the images cannot be interpreted. Therefore, a small portion of assessment is missing. In those cases, there is a mismatch imputation for sensitivity and specificity evaluation in this review. For reader agreement evaluation, the

missing cases were removed in this review and the results will be slightly better than those with worse case imputation (proposed by the sponsor). Since the proportion of missing assessment is very small, removing the missing cases for the reader agreement analyses will not have big impact.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

Only patients in Study 16034 were summarized in this section because the patients with autopsy are sub-population in the pool-read study.

In Study 16034, the effectiveness of an electronic training program for image orientation and interpretation was evaluated using PET images across subjects with different cognitive abilities who had participated in earlier studies. Inter-reader reproducibility of image interpretation was assessed using images from subjects with a truth standard (54 patients who underwent an autopsy and 10 YHVs) and without a truth standard (178 cognitively normal volunteers 55 years or above: median 70, range 55-98), 51 patients with mild cognitive impairment (MCI), 182 subjects with AD, and others.

Among the 461 subjects, the median age was 72 years (range 22 to 98), 43% were females, and 78% were Caucasian (White). Among the 54 autopsy subjects, the median age was 81 (range 58-98), 20 were females, and 33 were WHITE.

More information on the patient disposition is presented in Table 4.

Table 4: Patient disposition (% in ()) for Study 16034 for different patient populations

Group		Total population (n=461)	autopsy subjects from 14595 (n=54)	YHVs (n=10)	Autopsy and YHVs (n=64)
Gender	Female	197 (43)	20 (37)	4 (40)	24 (38)
	Male	264 (57)	34 (63)	6 (60)	40 (63)
Age	<65	111 (24)	4 (7)	10 (100)	14 (22)
	>=65	350 (76)	50 (93)	0	50 (78)
Age	Median (range)	72 (22-98)	81 (58-98)	25 (22-38)	79 (22-98)
Race	White	359 (78)	33(61)	10 (100)	43 (67)
	Asian	99 (21)	21(39)	0	21 (33)
	Other	3 (1)	0	0	0
Clinical diagnosis	AD	182 (39)	41 (76)	0	41 (64)
	MCI	51 (11)	0	0	0
	HV excluding YHVs	178 (39)	8 (15)	0	8 (13)
	YHV	10 (2)	0	10 (100)	10 (16)
	DEM	3 (1)	3 (6)	0	3 (5)
	DLB	10 (2)	2 (4)	0	2 (3)
	FTLD	12 (3)	0	0	0
	PD	5 (1)	0	0	0
	VaD	4 (1)	0	0	0
	Other	3 (1)	0	0	0
	NA	3 (1)	0	0	0

AD: Alzheimer's disease. MCI: Mild cognitive impairment. YHV: young healthy volunteer. HV: healthy volunteer DEM: other dementia. DLB: dementia with Lewy bodies. FTLD: fronto-temporal lobe degeneration (dementia). PD: Parkinson's disease. VaD: vascular dementia. NA: not available.

3.2.4 Results and Conclusions

Summary of the main results and conclusions provided by the sponsor is shown below.

Study 14595:

- For the 6 brain regions of interest, the sensitivity for the majority read was 77.36% (95% CI: 65.35% – 89.37%) and the specificity was 94.20% (95% CI: 88.57% – 99.84%).
Therefore, the combined hypothesis for this study that sensitivity is $\leq 60\%$ or specificity is $\leq 80\%$ could be rejected (co-primary analyses). Study demonstrates the ability to detect and locate a pathological state. This is the first study demonstrating a direct correlation of amyloid deposition and tracer uptake in the identical anatomic region
- Analyses using two different SoRs were performed: the composite “whole brain” histopathological assessment and the “onsite neuropathological” assessment.
 - The sensitivity and specificity of the “whole brain” regional assessment compared to the SoR were 86.96% (95% CI: 73.19 – 100.00%) and 88.89% (95% CI: 74.37% - 100%), respectively, for the majority read.---Approach 1
 - The subject level sensitivity and specificity of the visual assessment of florbetaben PET images according to the RCTU/RCTB and BAPL scoring as compared to the non-blinded “onsite neuropathological” assessments (as the SoR) revealed a sensitivity of 100% (95% CI: 80.49 – 100.00%) and specificity of 91.67% (95% CI: 80.31% – 100.00%), for the majority read.-----Approach 2

Study 16034:

- **Inter-reader agreement of the visual assessment based on the derived BAPL score (web-training): The primary endpoint exceeded the pre-specified kappa value threshold of 0.6 (for the lower bound of the two sided 95% CI) showing a value of 0.787 (CI: 0.750 – 0.824) across all five readers.**
- Based on a 10% re-read of the image data set, the intra-individual kappa values were convincing with the highest kappa value of 0.957 (95% CI: 0.872 – 1.041) for Reader 4 and the lowest being 0.823 (95% CI: 0.657 – 0.989) for Reader 1.
- Sensitivity and specificity of the visual assessment with histopathology as the Standard of Truth (SoT) using 55 (or 54 evaluable) autopsy cases enriched with 10% healthy volunteers: The sensitivity for the Blinded Readers 1 and 2 was 90% (95% CI: 80.70 – 99.30) and that for blinded readers 3 and 4 was 87.5% (95% CI: 77.25 – 97.75). Blinded Reader 5 provided a value of 77.50% (95% CI: 64.56 – 90.44). The specificity ranged from 62.5% (95% CI: 43.13 – 81.87) for Blinded Reader 3 to 91.67% (95% CI: 80.61 – 100.00) for Blinded Reader 5. The combined hypothesis that sensitivity is smaller or equal to 0.6 and specificity is smaller or equal to 0.7 could only be rejected for one out of

the 5 readers. **Thus the goal to reject this hypothesis for at least 3 out of the 5 readers was not met.**

- **Sensitivity and specificity of the visual assessment with histopathology as the Standard of Truth (SoT) using 55 (or 54 evaluable) autopsy cases *excluding* the healthy volunteers:** The sensitivity data was identical to that when the HVs were included in the analysis. The specificity values were lower than above for 4 out of the 5 readers, and higher for one reader. However, it should be noted that for 3 out of 5 readers (Readers 2, 3, and 5), the difference in the specificity for the two analyses (including and excluding the HVs) was less than 10 percentage points. **This suggests that the inclusion of the 10 HVs in the calculation of the specificity does not introduce a bias.**

In conclusion, on the basis of 461 images stemming from subjects with a broad range of diagnoses, it can be concluded that the visual assessment developed for florbetaben PET scans and as trained by a computer-(Web)-based tool, is reliable and reproducible with excellent ability to distinguish between subjects with and without β -amyloid deposition in the brain during life.

The reviewer does not agree with some of the above statements (by sponsor). The reviewer's findings are presented below. "Negative" will be used instead of "Normal". "Positive" will be used instead of "Abnormal".

Study 14595:

Regional-level sensitivity and specificity with and without 10 YHVs

The reviewer did not conduct the proposed primary analyses proposed by the sponsor because the statistical method is not proper.

Regional-level sensitivity and specificity of the blinded reads versus SoT are presented by reader in Table 5. Majority read results are also presented as the pre-specified primary analyses, but not recommended by the Agency. All regions are assumed to be independent to represent the best case scenario.

The point estimates are the same using independent assumption (conducted by the reviewer) and using Rao and Scott 1992 method (conducted by the sponsor). The confidence intervals assuming independent regions are similar to the results provided by the sponsor (Rao and Scott 1992). In addition, the number of brains (clusters) is 23 for sensitivity evaluation and 27 for specificity evaluation without 10 YHVs and (37 with 10 YHVs). The number of regions for evaluation within each brain is from 1 to 6. The number of brains (clusters) is small and the number of regions in each cluster for estimating proportions within brains (clusters) is also small. The method for incorporating the correlation information within brains proposed by the sponsor is not proper in this case.

There are a total of 106 positive regional scans and 138 negative regional scans based on histopathology and assumption for YHVs.

Without the 10 YHVs (60 negative regional scans), the total number of negative scans is 78. Sensitivity is the same with and without YHVs because YHVs only provide negative scans.

With the 10 YHVs, two out of three lower bounds of 95% CIs is more than 0.6 for sensitivity, and all three readers had specificity 95% CI lower bound more than 0.8.

It is more proper to evaluate the specificity excluding the 60 regional scans from the 10 YHVs. The specificity point estimates without YHVs are all lower than the ones with YHVs (about 6% less) and the lower bounds of the 95% CIs are between 0.7 and 0.8 for the three readers. The point estimates of specification for the 10 YHVs are ranged from 98% to 100%.

Table 5: Regional-level sensitivity and specificity (%) for all regions by reader and majority read

Reader	With YHVs				Without YHVs		10 YHVs
	Sensitivity	CI	Specificity	CI	specificity	CI	specificity
1	80(85/106)	(71, 87)	91 (125/138)	(85, 95)	85 (66/78)	(75, 92)	98 (59/60)
2	81(86/106)	(72, 88)	91 (126/138)	(85, 95)	85 (66/78)	(75, 92)	100 (60/60)
3	59 (63/106)	(49, 69)	92 (127/138)	(86, 96)	86 (67/78)	(76, 93)	100 (60/60)
Majority	77 (82/106)	(68, 85)	94 (129/138)	(88, 97)	89 (69/78)	(79, 95)	100 (60/60)

By region sensitivity and specificity (secondary analyses)

By region and by reader results are presented in Table 6, which reflect the worst case scenario for sensitivity and specificity at regional-level. Confidence intervals are not provided for the cases without YHVs since the sample size (around 10) is very small, which may lead to very wide confidence intervals. The total number of cases is 41 with YHVs, and 31 without YHVs. The number of positive scans and negative scans are different by region.

The sensitivity values for region 3 and 6 (hippocampus and cerebellar cortex) are very low (point estimates less than 0.55 for two readers), and the specificity values are very high (point estimates as 100% for two readers). The sample size for the specificity evaluation is very small (n=4) for region 6 (cerebellar cortex). The exclusion of YHVs led to reduced point estimates of the specificity (about 10% less without 10 YHVs for the by reader analyses).

Table 6: Regional-level sensitivity and specificity (%) by region and by reader

	With YHVs				Without YHVs specificity
	Sensitivity	CI	Specificity	CI	
Reader 1					
Region 1: middle frontal gyrus	85.7 (18/21)	(64, 97)	85 (17/20)	(62, 97)	70 (7/10)
Region 2: occipital cortex	88.9 (16/18)	(65, 99)	86.4 (19/22)	(65, 97)	83.3 (10/12)
Region 3: hippocampus	76.2 (16/21)	(53, 92)	90 (18/20)	(68, 99)	80 (8/10)
Region 4: Anterior cingulate cortex	90 (18/20)	(68, 99)	85.7 (18/21)	(64, 97)	72.7 (8/11)
Region 5: posterior singulate/precuneus	77.3 (17/22)	(55, 92)	88.9 (16/18)	(65, 99)	75 (6/8)
Region 6: cerebellar cortex	0 (0/4)	(0, 60)	100 (37/37)	(91, 100)	100 (27/27)
Reader 2					
1: middle frontal gyrus	95.2 (20/21)	(76, 100)	90 (18/20)	(68, 99)	80 (8/10)
2: occipital cortex	100(18/18)	(82, 100)	77.3 (17/22)	(55, 92)	58.3 (7/12)
3: hippocampus	52.4 (11/21)	(30, 74)	100 (20/20)	(83, 100)	100 (10/10)
4: Anterior cingulate cortex	95 (19/20)	(75, 100)	85.7 (18/21)	(64, 97)	72.7 (8/11)
5: posterior singulate/precuneus	81.8 (18/22)	(60, 95)	88.9 (16/18)	(65, 99)	75 (6/8)
6: cerebellar cortex	0 (0/4)	(0, 60)	100 (37/37)	(91, 100)	100 (27/27)
Reader 3					
1: middle frontal gyrus	66.7 (14/21)	(43, 85)	90 (18/20)	(68, 99)	80 (8/10)
2: occipital cortex	61.1 (11/18)	(36, 83)	86.4 (19/22)	(65, 97)	75 (9/12)
3: hippocampus	33.3 (7/21)	(15, 57)	100 (20/20)	(83, 100)	100 (10/10)
4: Anterior cingulate cortex	75 (15/20)	(51, 91)	85.7 (18/21)	(64, 97)	72.7 (8/11)
5: posterior singulate/precuneus	72.7 (16/22)	(50, 89)	94.4 (17/18)	73, 100)	87.5 (7/8)
6: cerebellar cortex	0 (0/4)	(0, 60)	94.6 (35/37)	(82, 99)	92.6 (25/27)

Subject-level analyses with SoT from two approaches (secondary analyses)

- Approach 1: reader1 PET assessment, reader2 PET assessment, reader3 PET assessment, and SoT were obtained by collapsing the regional results. For both PET assessment and SoT, if any region within a brain is positive, the whole brain is positive.
- Approach 2: PET assessments for the three readers are based rules for future use, and SoT was obtained from on-site neuropathological diagnosis.

Confidence intervals were not calculated for the cases without YHVs because of the small sample size. The results are shown in Table 7.

Note that for PET assessment, six regions were used in Approach 1 and four regions were used in Approach 2. None of the six regions (App 1) are the same as the four regions (App 2). The results on subject-level analyses using approach 2 are better than those using approach 1 in terms of both sensitivity and specificity. By Approach 2, the point estimate of the sensitivity is 100% for all three readers with 95% CI as (80%, 100%), and two out of three readers had lower bound of 95% CIs for specificity >0.7. However, by Approach 1, only one reader had the point estimate of sensitivity more than 90%, and the lower bounds of 95% CIs are 47%, 52%, and 65% for the three readers respectively. None of the two approaches will be used in future clinical practice and also in the pool-read study 16034.

For the above results, the 95% CIs are wider than those provide by the sponsor because exact confidence intervals were calculated by the reviewer and sponsor used normal approximation. Exact confidence intervals are more proper in the above cases with the small sample sizes (n=18 and n=24).

Using Approach 2, the SoT decided by the on-site histopathology led to 17 positive and 24 negative cases including the 10 negative YHVs, which are not the same as the ones for Approach 1 (23 positive and 18 negative cases).

The Excluding the 10 YHVs, the point estimate of the specificity is 50% for two out of the tree readers by Approach 1, and >0.7 for all three readers by Approach 2. There is no conclusion because of the small sample size without YHVs (n=8 Approach 1, n=14 Approach 2, for specificity evaluation).

For Approach 1, SoT was decided using neuritic and diffused plaques. Other SoTs (modification of SoT) were also explored (SoT_all, SoT_neur, SoT_diff, SoT_vasc) and the results are not shown. Different definitions yield different number of subjects for sensitivity and specificity evaluation. Including the 10YHVs (all negative cases), SoT_all and SoT_diff led to 23 positive and 18 negative cases, SoT_neur led to 19 positive and 22 negative cases, SoT_vasc led to 17 positive and 24 negative cases. Performance characteristics for different SoTs were also different.

Table 7: Subject-level analyses with two approaches proposed by the sponsor

	With YHVs		Without YHVs
Approach 1 (composite whole brain)	n=41=23 pos + 18 neg		n=31=23 pos + 8 neg
Reader	Sen (95% CI) in %	Spe (95% CI) in %	Spe in %
1	87 (66, 97)	72 (47, 90)	50
2	96 (78, 100)	78 (52, 94)	50
3	87 (66, 97)	89 (65, 99)	75
Majority read	87 (66, 97)	89 (65, 99)	75
Approach 2	n=41=17 pos + 24 neg		n=31=17 pos + 14 neg
Reader	Sen (95% CI) in %	Spe (95% CI) in %	Spe in %
1	100 (80, 100)	92 (73, 99)	86
2	100 (80, 100)	92 (73, 99)	86
3	100 (80, 100)	83 (63, 95)	79
Majority read	100 (80, 100)	92 (73, 99)	86

Study 16034:

Inter-reader kappa and agreement

The sponsor used a worse case imputation approach to impute the missing PET assessment for 7 subjects with missing reads out of the 461 subjects. In this review, the images with missing PET assessment were removed, with slightly better results compared with worse case imputation in the primary analysis (missing percentage is 7/461=1.5%). As shown in Table 8, the kappa value for the primary analysis is 0.799, and the percent of five readers agree with each other is 78%. The lower bound of the 95% CI is 0.77, which is bigger than the pre-specified threshold value of 0.6 for inter-reader agreement in the primary analysis.

The inter-reader agreement for MCI subjects is also good with kappa =0.75.

Table 8: Inter-reader agreement and percent of agreement

Subject Group by Cognitive and Standard of Truth (SoT)	Positive Scans, n ^a	Kappa (95% CI)	Percent of Scans with Inter-reader Agreement		
			3 of 5 readers agreed	4 of 5 readers agreed	5 of 5 readers agreed
Primary (n=454 excluding 7 with missing reads)	212	0.799 (0.77, 0.83)	6	15	78
Autopsy (n=60 excluding 4 with missing reads)	37	0.747 (0.67, 0.83)	10	15	75

^aShown is the median number of scans interpreted as positive across the 5 readers for each subgroup of subjects listed in the first column.

As shown in Table 9, with the 10 YHVs, the lower bounds of the 95% CIs (sensitivity) are >0.7 for four readers, and >0.6 for one reader. However, the lower bounds of the 95% CIs (specificity) are <0.6 for three readers. Without the 10 YHVs, there are only 14 negative cases for specificity evaluation. The point estimates of the specificity without 10 YHVs are less than those with the 10YHVs. The confidence intervals had very low lower bounds (0.35 to 0.57 for the five readers). Because of the small sample size, the CIs are very wide. More negative brains should be collected to evaluate the specificity with web-training PET assessment.

Table 9: Sensitivity and specificity with web-training PET assessment for 54 autopsy subjects and the 10 YHVs

Reader	With YHVs, n=64=40 pos + 24 neg				Without YHVs, n=40 pos+ 14 neg	
	sensitivity	CI	Specificity	CI	Specificity	CI
1	90	(76, 97)	83	(63, 95)	71	(42,92)
2	90	(76, 97)	63	(41, 81)	64	(35, 87)
3	87.5	(73, 96)	75	(53, 90)	71	(42, 92)
4	87.5	(73, 96)	79	(58, 93)	64	(35, 87)
5	77.5	(62, 89)	92	(73, 99)	86	(57, 98)

Intra-reader agreement

Intra-reader reproducibility analysis showed that, between the two readings for each of the 46 duplicate patient images, one reader had discordant reads for a single image, two readers had discordant reads for two images, one reader had discordant reads for three images, and one reader had discordant reads for four images. The percent of agreement for all the 46 images is 91%, 96%, 96%, 98%, and 93% for reader 1, 2, 3, 4, and 5, respectively.

Intra-reader reproducibility for a sub-group of 5 images from MCI patients showed that all five readers had complete agreement for all duplicate images.

3.3 Evaluation of Safety

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

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

In this Section, subgroup analyses as exploratory analyses were only presented for Study 16034, which is the only study with web-based training (used in clinical practice). Only reviewer's results are presented below.

Note that no conclusions can be drawn from the subgroup analyses due to lack of representation and limited sample size. Further exploration should be conducted for some subgroups of interest.

4.1 Gender, Race, and Age

Sensitivity and specificity were explored by Race, gender, and age (>65 or not) among the autopsy subjects plus the 10 YHVs. Confidence intervals are not calculated because of the small sample sizes. The results are shown in Table 10.

For the five readers, sensitivity values are 84-92% for Male and 67-87% for Female. Specificity values are 53-93% for Male and 78-100% for Females. Sensitivity values are 74-89% for Whites and 85-92% for Asian. Specificity values are 63-94% for Whites and 63-88% for Asian. Sensitivity values are 100-100% for subjects with age<=65 (n=2) and 76-89% for subjects with age>65. Specificity values are 58-100% for subjects with age<=65 and 58-83% for subjects with age>65.

The performance is different in different subgroups. More studies with large sample size should be conducted for reaching a conclusion.

Table 10: Sensitivity (sen) and specificity (spe) by gender, race and age (in %)

	gender		Race		Age	
	Male N=40=25 pos+15 neg	Female N=24=15 pos+9 neg	White N=43=27pos+16 neg	Asian N=21=13 pos+ 8 neg	<=65 N=14=2 pos+12 neg	>65 N=50=38 pos+12 neg
Reader 1						
Sen	92	87	89	92	100	89
spe	73	100	94	63	100	67
Reader 2						
sen	92	87	89	92	100	89
spe	53	78	63	63	58	67
Reader 3						
sen	88	87	85	92	100	87
spe	73	78	81	63	83	67
Reader 4						
sen	92	80	89	85	100	87
spe	73	89	88	63	100	58
Reader 5						
sen	84	67	74	85	100	76
spe	93	89	94	88	100	83

Note: pos is for positive cases, neg is for negative cases based on histopathology or negative assumption for YHVs

Inter-reader agreement was evaluated by subgroups among the 461 subjects. The performance in terms of kappa values and percent of reader agreement is very similar in different subgroups (Table 11).

Table 11: Reader agreement evaluation by gender, race, and age group (≤ 65 and > 65)

Subject Groups	Positive Scans, n ^a	Kappa (95% CI)	Percent of Scans with Inter-reader Agreement		
			3 of 5 readers agreed	4 of 5 readers agreed	5 of 5 readers agreed
Gender					
Male (n=261 excluding 3 with missing reads)	123	0.80 (0.74, 0.82)	6	19	75
Female (n=193 excluding 4 with missing reads)	89	0.83 (0.78, 0.87)	7	10	82
Race					
White (n=353 excluding 6 with missing reads)	175	0.79 (0.76, 0.83)	7	16	78
Asian (n=98 excluding 1 with missing reads)	35	0.81 (0.75, 0.87)	5	14	81
Age group					
<65 (n=109 excluding 2 with missing reads)	32	0.74 (0.68, 0.80)	6	18	76
≥ 65 (n=345 excluding 5 with missing reads)	180	0.80 (0.77, 0.84)	7	14	79

^a Shown is the median number of scans interpreted as positive across the 5 readers for each subgroup of subjects listed in the first column.

4.2 Other Special/Subgroup Populations

Clinical diagnosis is an important factor since the intended patient population is the subjects with clinical diagnosis as MCI. Therefore, the accuracy in term of sensitivity and specificity, and the reader agreement were evaluated by baseline clinical diagnosis.

As shown in Table 12, most subjects with autopsy are ADs and HVs. The specificity values for AD subjects are 0.38-0.75 for the five readers and the sensitivity values for HVs are 0.4-0.6 for the five readers (Table 13). The confidence intervals were not calculated because of the small sample sizes.

Table 12: Sensitivity (sen) and specificity (spe) point estimates by clinical diagnosis (in %)

Reader	AD (41=33 pos+8 neg)		HV (18=5 pos+ 13neg)		HV excluding 10 YHVs (8 =5 pos+ 3 neg)	
	Sen	Spe	Sen	Spe	Sen	Spe
1	94	63	60	92	60	67
2	94	63	60	62	60	67
3	94	50	40	85	40	100
4	94	38	40	100	40	100
5	82	75	40	100	40	100

As shown in Table 13, AD and MCI subjects had good reader agreement (kappa values >0.75 and percent of 5 out of 5 readers agreement with each other >80%). For the 188 Cognitive normal subjects, the number of positive scans is only 26. The imbalance in the positive and negative scans led to the low kappa values. Percent of agreement (77% 5 out of 5 reader agreement) should be used instead of kappa in this case.

Table 13: Reader agreement subgroup analyses by baseline clinical diagnosis

Subject Group by Cognitive and Standard of Truth (SoT)	Positive Scans, n ^a	Kappa (95% CI)	Percent of Scans with Inter-reader Agreement		
			3 of 5 readers agreed	4 of 5 readers agreed	5 of 5 readers agreed
AD (n=176 excluding 6 with missing reads)	139	0.76 (0.72, 0.81)	7	10	83
HV (n=188)	26	0.53 (0.49, 0.58)	7	15	77
MCI (n=50 excluding 1 with missing reads)	28	0.84 (0.75, 0.92)	0	20	80

^a Shown is the median number of scans interpreted as positive across the 5 readers for each subgroup of subjects listed in the first column.

AD: Alzheimer's disease. MCI: Mild cognitive impairment. YHV: young healthy volunteer. HV: healthy volunteer. DEM: other dementia. DLB: dementia with Lewy bodies. FTLD: fronto-temporal lobe degeneration (dementia). PD: Parkinson's disease. VaD: vascular dementia. NA: not available.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

Study 14595

- In-person training for the readers was used, which will likely cause the study results to be biased in comparison to what would be expected in clinical practice using a web-training method.
- Regional-level comparison (for sensitivity and specificity) between PET assessment and SoT were conducted in the **primary analyses** (by sponsor). However, subject-level analyses are more appropriate with respect to clinical practice instead of regional-level analyses.
- The sponsor claimed that the confidence intervals are calculated taking within brain dependencies into account using normal approximation (by Rao and Scott 1992 method) However, the normal approximation theory discussed in Rao and Scott 1992 used the condition that the number of clusters should be large for normal approximation. But the total number of clusters (brains) in this study is only 31+10YHVs=41. All the regions in

the 10YHVs are assumed to be negative. Among the 31 autopsy subjects, some will have all six regions as positive or negative regions. The numbers of clusters for sensitivity and specificity evaluation are small (23 for sensitivity and 27 for specificity without the 10 YHVs and 37 with the 10 YHVs). The method proposed by the sponsor is not proper.

- Majority read were used in the primary analyses by the sponsor. However, in clinical practice, usually one reader interprets the images. By reader analyses will be more proper because the estimation using majority read data will not be representative of expected performance of single readers.
- The primary analyses on regional-level images included 60 (6 regions × 10 YHVs) negative regions assuming all six regions in the YHVs are negative, which is $60/138=44\%$ of the total negative regions. This will lead to inflated specificity values.
- Two approaches for evaluating subject-level sensitivity and specificity were conducted by the sponsor. None of them will be used in future clinical practice.
- In addition, only end-of-life subjects were enrolled, who are not the intended patient population.

There is no agreement on the proposed primary analyses on regional-level (Study 14595) between the sponsor and the Agency (REV-BIOMETRICS-01(General Review) for IND 78868 with submit date as 10/3/2011 and COR-MEET-03 with submit date as 6/16/2011).

Study 16034

- The images in the pool-read study were selected from some earlier studies, but not all earlier studies. There may be a selection bias.
- The majority of the subjects in the study are subjects with Alzheimer's dementia (182/461=40%), HVs (186/461=41%). The number of the intended patient population (MCI subjects) is only 51 (11%) out of the total 461 subjects.

For both **Studies 14595 and 16034**, subject-level specificity evaluation includes 10 YHVs, which is $10/24=42\%$ of the total negative cases with SoT (either from autopsy or the negative assumption for the YHVs). The presence of the 10 YHVs may inflate the performance in terms of specificity. Excluding the YHVs, the sample size is 14 for specificity evaluation (small), which will lead to wide confidence intervals.

The number of negative cases (by SoT) among the 31 brains (obtained in Study 14595) is 8 by histopathology pathology consensus panel (CP) in Study 14595, 14 by onsite neuropathological diagnosis in Study 14595, and 10 by CP in Study 16034. According to the protocol, the rules are the same for the CP approach in Studies 14595 and 16034. The discrepancy on SoT between Study 14595 and 16034 cannot be explained.

For both studies, the sponsor used normal approximation to obtain the confidence intervals of sensitivity and specificity at subject-level. Exact confidence intervals will be more proper in the case of the small sample sizes.

For the pool-read study, the Agency did not approve the inclusion of 10 YHVs in the sensitivity and specificity analyses and the use of normal approximation for the calculation of the confidence intervals (see details in IND 78868 information request or advice, communication ID COR-INDAD-02).

5.2 Collective Evidence

The sponsor claimed that the combined hypothesis for Study 14595 that sensitivity is $\leq 60\%$ or specificity is $\leq 80\%$ could be rejected (co-primary analyses). However, the statistical method used in the primary analyses is not proper.

The sponsor also claimed that the primary endpoint exceeded the pre-specified kappa value threshold of 0.6 (for the lower bound of the two sided 95% CI) showing a value of 0.787 (CI: 0.750 – 0.824) across all five readers for Study 16034. However, the sponsor also claimed that the combined hypothesis that sensitivity is smaller or equal to 0.6 and specificity is smaller or equal to 0.7 (specified as secondary analyses in the protocol) could only be rejected for one out of the 5 readers. Thus the goal to reject this hypothesis for at least 3 out of the 5 readers was not met.

In the following, the reviewer's points are presented.

At regional-level (Study 14595), with the 10 YHVs, assuming independent regions within each brain (best scenario), two out of three lower bounds of 95% CI are more than 0.6 for sensitivity, and all three readers had specificity 95% CI lower bound > 0.8 . The specificity point estimates without YHVs are all lower than the ones with YHVs (about 6% less) and the lower bound of the 95% CIs are between 0.7 and 0.8 for the three readers. However, at subject-level (by Approach 1: subject-level PET assessment and SoT were obtained using regional level PET assessment and regional histological findings from the Pathology Consensus Panel), only one reader had the point estimate of sensitivity more than 90%, and the lower bounds of 95% CIs of specificity were 47%, 52%, and 65% for the three readers respectively. The rule for collapsing from regional assessment and SoT to subject assessment and SoT is that if one region in a brain is positive, the whole brain is positive. This rule applies to both PET assessment and SoT.

The sponsor used another approach to obtain the subject-level sensitivity and specificity for the 31 brains and 10YHVs. The in-person training PET assessment used the rules for future use. The SoT was determined by onsite neuropathological diagnosis. The point estimate of the sensitivity is 100% for all three readers with 95% CI as (80%, 100%), and two out of three readers had lower bound of 95% CIs for specificity >0.7 . The better performance using Approach 2 vs. 1 may be due to the use of the new rule for reader training, or the change of SoT.

None of the rules in the two approaches (approach 1 and 2) will be used in clinical practice. In addition, in-person training was used in Study 14595.

The pool-read study evaluated the Approach 2 PET assessment with web-training. The SoT values were obtained using Approach 1 regional histological findings from CP. From the definitions, the SoT for the pool-read study should be the same as the one used in Approach 1 (in Study 14595). However, data exploration shows that the SoT for Approach 1, Approach 2 in Study 14595, and the SoT in Study 16034 are all different. The number of positive cases is 23 by SoT in Approach 1, 17 by SoT in Approach 2, and 21 by SoT in the pool-read study, respectively; the number of negative cases is 8 by SoT in Approach 1, 14 by SoT in Approach 2, and 10 by SoT in the pool-read study, respectively. The rules for determining the SoT by the sponsor in the pool-read study are not clear.

In Study 16034, with the 10 YHVs, the lower bound of the 95% CIs (sensitivity) are >0.7 for four readers, and >0.6 for one reader. The lower bound of the 95% CIs (specificity) are <0.6 for three readers. Without the 10 YHVs, there are only 14 negative cases for specificity evaluation. The point estimates of the specificity without 10 YHVs are less than those with the 10YHVs. The confidence intervals had very low lower bounds (0.35 to 0.57 for the five readers). Because of the small sample size, the CIs are very wide. More negative brains should be collected to evaluate the specificity with web-training PET assessment. In addition, the specificity values for subgroup AD (0.38-0.75 for the five readers) are much lower than the specificity values of the HVs (0.62-1.00 for the five readers with and without 10 YHVs).

For inter-reader agreement, the kappa value for the primary analysis in **Study 16034** is 0.799 (7 subjects with missing reads were removed), and the percent of five readers agree with each other is 78%. The inter-reader agreement for MCI subjects is also good with kappa =0.75.

Intra-reader reproducibility analysis in Study 16034 showed that, between the two readings for each of the 46 duplicate patient images, one reader had discordant reads for a single image, two readers had discordant reads for two images, one reader had discordant reads for three images, and one reader had discordant reads for four images. The percent of agreement for all the 46 images is 91%, 96%, 96%, 98%, and 93% for reader 1, 2, 3, 4, and 5, respectively. Intra-reader reproducibility for a sub-group of 5 images from MCI patients showed that all five readers had complete agreement for all duplicate images.

5.3 Conclusions and Recommendations

The statistical results in terms of accuracy (sensitivity and specificity) and reproducibility provide very limited evidence to support the claim for the detection of β -amyloid in the brain proposed in this NDA. The reader agreement is good in terms of intra-reader and inter-reader agreement, which indicates that the results are reproducible. However, the low specificity (indicating high false positive images) for subjects with autopsy in Study 16034 using web-training process should be noted. Agreement on incorrect interpretations suggests an inadequate interpretation method. In addition, the rules for obtaining the subject-level PET assessment and SoT are not the same in the two pivotal studies. Particularly, it is not clear how the sponsor

determined the SoT using the CP histopathology. The performance varies in different studies with different training processes and different approaches for obtaining the subject-level results. There are no MCI subject data, the population most needing the diagnostic method, with autopsy information. Conclusive evidence for the performance in terms of sensitivity and specificity cannot be obtained on MCI population.

Reference:

Rao JNK, Scott AJ. A simple method for the analysis of clustered binary data. *Biometrics* 1992;48:577-85.

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
08/23/2013

JYOTI ZALKIKAR
08/23/2013

I concur with the primary reviewer who worked very diligently through this review and wrote a extensive report.

THOMAS E GWISE
08/23/2013

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

NDA Number: 204677

Applicant: Piramal Imaging

Stamp Date: 12/21/2012

Drug Name: Florbetaben

NDA/BLA Type: NME

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.	Yes			
2	ISS, ISE, and complete study reports are available (including original protocols, subsequent amendments, etc.)	Yes			
3	Safety and efficacy were investigated for gender, racial, and geriatric subgroups investigated (if applicable).	Yes			
4	Data sets in EDR are accessible and do they conform to applicable guidances (e.g., existence of define.pdf file for data sets).	Yes			

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE? Yes

We send the following information request on 1/29/2013:

1. When we load your xpt files into sas, many variables do not have the right format. The error message is "Format was not found or could not be loaded". Please provide the format for all the variables in all the data submitted. You may need to resubmit all the xpt files to correct this problem. Failure to promptly resolve this problem may preclude our ability to review your application in a timely manner.
2. Provide the names of the data sets and the related sas programs used to generate the tables in the submission, especially the tables for the efficacy evaluation in studies 14595 and 16034.
3. The images to be assessed during the "pooled read" study (Study 16034) were chosen from various Phase 1 studies, the Phase 2 study (Part B) and the Phase 3 study. Provide a description of the criteria you used to select images/subjects for inclusion in the pooled read study. Were these criteria pre-specified in a manner that clearly identified which images/subjects would be included/excluded from the pooled read? If so, provide the documentation that verifies these details of the image/subject selection process. Also provide a table (or figure) that describes the Study 16034 subject distribution (by the study that originally enrolled the subject).
4. Regarding Study 14595, we have been unable to locate the pre-specified statistical analytical plan (SAP). Please identify the location of the SAP and/or submit this plan. We are particularly interested in the details of the interim analysis.

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

The sponsor answered the questions 3 and 4 in the applicant orientation meeting on 2/4/2013.

For question number 2, the sponsor submitted the sas programs for studies 14595 and 16034 on 2/12/2013 ([\\CDSESUB1\EVSPROD\NDA204677\0003](#)).

For question number 1, the sponsor submitted formats.xpt files for studies 14311, 14595, 16034, 311741, and 312043 ([\\CDSESUB1\EVSPROD\NDA204677\0003](#)).

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.		No		Subject level endpoints should be used as primary endpoints (instead of regional endpoints)
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.	yes			
Interim analyses (if present) were pre-specified in the protocol and appropriate adjustments in significance level made. DSMB meeting minutes and data are available.			NA	
Appropriate references for novel statistical methodology (if present) are included.			NA	
Safety data organized to permit analyses across clinical trials in the NDA/BLA.	yes			
Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.			NA	

Lan Huang

02/16/2013

Reviewing Statistician

Date

Jyoti Zalkikar

02/18/2013

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
02/19/2013

JYOTI ZALKIKAR
02/19/2013