CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

203137Orig1s000

STATISTICAL REVIEW(S)

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 203-137
Supplement #:	
Drug Name:	VIZAMYL (Flutemetamol F18 injection) for Intravenous Use
Indication(s):	A radioactive diagnostic agent for Positron Emission Tomography (PET) imaging of the brain to estimate β amyloid neuritic plaque density in adult patients with cognitive impairment who are being evaluated for Alzheimer's disease (AD) or other causes of cognitive decline.
Applicant:	GE Healthcare
Date(s):	date submitted 10/26/2012
	PDUFA due date 6/28/2013
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:	Philip Davis, M.D. (clinical reviewer)
	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 6

2.1 OVERVIEW 6
2.1.1. CLASS AND INDICATION 6
2.1.2. HISTORY OF PROGRAM DEVELOPMENT 7
2.1.3. SPECIFIC STUDIES REVIEWED 8
2.1.4. MAJOR STATISTICAL ISSUES 8
2.2 DATA SOURCES 10

3. STATISTICAL EVALUATION 10

3.3 DATA AND ANALYSIS QUALITY 10
3.4 EVALUATION OF EFFICACY 10
3.4.1 Study Design and Endpoints 10
3.4.2 Statistical Methodologies 14
3.4.3 Patient Disposition, Demographic and Baseline Characteristics 16
3.4.4 Results and Conclusions 17
3.5 EVALUATION OF SAFETY 22

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS 22

4.1 GENDER, RACE, AND AGE 23
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS 24

5. SUMMARY AND CONCLUSIONS 26

5.1 STATISTICAL ISSUES 26
5.2 COLLECTIVE EVIDENCE 27
5.3 CONCLUSIONS AND RECOMMENDATIONS 28
5.4 LABELING RECOMMENDATIONS 28

Reference ID: 3334092

LIST OF TABLES

Table 1: Summary of pivotal clinical efficacy studies GE067-007, GE067-015, and GE067-021 9
Table 2: Patient disposition (% in ()) for GE067-021 for different patient populations..... 17
Table 3: Sensitivity and specificity (%) with and without anatomic images by reader..... 19
Table 4: Specificity for the 181 Young Healthy Volunteers (YHVs) 20
Table 5: By reader sensitivity and specificity (point estimate and 95% CIs) without anatomic images for different analysis populations 21
Table 6: Inter-Reader agreement for different analysis populations 22
Table 7: Sensitivity (sen) and specificity(spe) by gender (in %). pos for positive cases, neg for negative cases based on histopathology, or biopsy, or negative assumption for YHVs) 23
Table 8: Reader agreement evaluation by gender and age 23
Table 9: Sensitivity (sen) and specificity (spe) point estimates by clinical diagnosis (in %)..... 25
Table 10: Reader agreement subgroup analyses by baseline clinical diagnosis..... 26

Reference ID: 3334092

1. EXECUTIVE SUMMARY

The medical product under development is Flutemetamol F18 (VIZAMYL). VIZAMYL is intended for estimating β amyloid neuritic plaque density in adult patients with cognitive impairment who are being evaluated for Alzheimer's disease (AD) or other causes of cognitive decline. A negative Vizamyl 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 11 studies in the clinical development program. Data were collected from 761 subjects enrolled in 10 clinical studies, including 2 Phase 1 studies (ALZ103 and GE067-014 for determination of biodistribution and dosimetry); 1 Phase 2 study (ALZ201 for differentiating Healthy volunteer (HV) and probable AD (pAD) subjects); and 7 Phase 3 studies (GE067-005 for prediction of conversion from Mild Cognitive Impairment (MCI) to pAD, GE067-007 (autopsy study for evaluating sensitivity with Standard of Truth based on histopathology), GE067-008, GE067-009, GE067-010, and GE067-011 for evaluating correlation between SUVR and brain amyloid level, GE067-015 for evaluating specificity in young healthy volunteers (YHVs) assuming they are all negative cases). GE067-021 is a pool read study including images obtained from the previous studies for evaluating sensitivity and specificity with Standard of Truth (SoT) based on histopathology or biopsy, and reader agreement for a broad range of subjects. The YHVs are assumed to be negative cases.

Studies GE067-007, GE067-015, and GE067-021 were selected for full statistical review and evaluation. Note that only GE067-021 used the web-training approach in the whole clinical program. The in-person training approach used in the other studies will not be used in clinical practice. In GE067-007, only analysis of sensitivity is included in the proposed primary analysis. Sensitivity may be biased without evaluating specificity together. In addition, only end-of-life subjects were enrolled, who are not from the intended patient population (MCI subjects). In GE067-015**,** the images from Young Healthy Volunteers (YHVs) were mixed with the ones from MCI subjects, and were read by blinded readers. It is assumed that the YHVs had negative amyloid level and there is no SoT for MCI subjects. It is impossible to evaluate the sensitivity in this study. For any given diagnostic test there is a trade-off between sensitivity and specificity. Therefore, to properly interpret diagnostic test performance, sensitivity and specificity must be evaluated together using the same study sample and blinded readers.

Study GE067-021 includes both sensitivity and specificity analyses in the primary analyses. However, the 135 subjects in the primary analyses population include 68 autopsy subjects, 36 subjects with biopsy, and 31 YHVs (54 positive + 84 negative cases by SoT). Including the 31 YHVs (assumed to be negative cases) may lead to inflated specificity. Without the YHVs, the sample size is 53 for specificity evaluation and 51 for sensitivity evaluation. This sample size is large enough for the evaluation without the YHVs.

For GE067-007, without anatomic images and with in-person training, the lower bounds of 95% CIs for sensitivity and specificity are both >70% only for **the same 2 out of the 5 blinded readers** (Table 3)**.** The results with anatomic images are similar.

Reference ID: 3334092

Readers with web-based training read the images of the 68 autopsy subjects (from GE067-007) in the pool read study GE067-021. For this sub-population with SoT based on histopathology, the lower bounds of 95% CIs for sensitivity and specificity are both >70% only for the **same 1 out of the 5 blinded readers**, which is slightly worse than the performance with in-person training (Table 5).

In the pool-read study GE067-021, for the primary analysis population (the subjects with SoT based on histopathology, or biopsy, or assumed negative YHV cases), the sensitivity of the blinded visual interpretation of PET images ranged from 84% to 94% with a median value of 92%. The specificity ranged from 77% to 96% by reader, with a median value of 81%. The study met the primary success criterion, because the lower bounds of the 2-sided 95% CI for both sensitivity and specificity **exceeded 70% for each of 3 out of 5 readers** (Table 5).

Excluding the 31 YHVs, for the 104 subjects with SoT from histopathology or biopsy, the lower bounds of 95% CI for sensitivity and specificity are both >70% for **2 out of the 5** blinded readers (Table 5)**.**

The availability of anatomic images (CT) for consultation during the blinded visual interpretation of PET images did not significantly affect the results.

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 blinded visual interpretations of PET images with web-trained readers, the percent of agreement (all five readers agreed with each other) is 81% (Analysis Population 4 including all subjects in the study; n = 276), and the kappa value for inter-reader agreement across all 5 readers is 0.83 with 95% CI as (0.79, 0.86). Inter-reader agreement met the pre-specified secondary success criterion, which required the lower bound of the 2-sided CI for kappa to exceed 0.6 (Table 6). As shown in Table 10, the percent of agreement (all five readers agreed with each other) is 89% for MCI subjects (n = 80), and the kappa value for inter-reader agreement across all 5 readers is 0.89 with 95% CI as (0.82, 0.96). In addition, intra-reader agreement ranged from 93% to 100% among the five blinded readers.

The statistical results in terms of sensitivity and specificity and reproducibility (in terms of reader agreement) provide some evidence to support the claims proposed in this NDA. However, the issue of relatively low specificity for subjects with autopsy and biopsy should be noted. There are no MCI subjects with autopsy. Therefore, we cannot obtain conclusive evidence for the performance in terms of sensitivity and specificity on MCI population.

Reference ID: 3334092

2. INTRODUCTION

2.1 Overview

2.1.1. Class and Indication

Flutemetamol F 18 Injection 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 Flutemetamol F 18 Injection is *VIZAMYL*. The proposed Indications and Usage statement for Flutemetamol F 18 Injection (by the sponsor) is:

(b) (4)

The revised indication provided by the Agency is

Vizamyl is a radioactive diagnostic agent for Positron Emission Tomography (PET) imaging of the brain to estimate β amyloid neuritic plaque density in adult patients with cognitive impairment who are being evaluated for Alzheimer's disease (AD) or other causes of cognitive decline. A negative Vizamyl 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. A positive Vizamyl scan indicates moderate to frequent amyloid neuritic plaques; neuropathological examination has shown this amount of neuritic plaque is present in patients with AD, but may also be present in patients with other types of neurologic conditions as well as older people with normal cognition. Vizamyl is an adjunct to other diagnostic evaluations.

A positive Vizamyl scan does not establish a diagnosis of AD or other cognitive disorder. Safety and effectiveness of Vizamyl have not been established for:

Reference ID: 3334092

- *Predicting the development of dementia or other neurological condition;*
- *Monitoring responses to therapies.*

2.1.2. History of Program Development

A pre-IND meeting was held with the US FDA on [May 19, 2008]. At this meeting FDA stated that the standard of truth and measures of clinical utility would be further discussed at an Advisory Committee meeting (AC). In the AC [October 23, 2008], the Committee agreed that "positive" amyloid test would have very limited utility since cerebral amyloid is known to be present in multiple conditions, including normal aging. A "negative" amyloid test could have clinical utility in ruling out a diagnosis of AD. The committee overwhelmingly agreed that histopathological correlation should be the Standard of Truth (SoT) in phase 3 clinical studies for these agents.

An end-of-Phase-2 meeting was held on [September 7, 2010]. FDA confirmed that the primary study to support the approval of Flutemetamol F 18 Injection should be an autopsy study [GE067-007] and a study in presumed amyloid negative subjects (subjects ≤40 years of age, [GE067-015]). At a type-C meeting (October 25, 2011), it was agreed that an electronic training program would be developed and a clinical trial conducted to validate that the program was successful at training physicians to read and accurately interpret [18F]flutemetamol PET images.

A pre-NDA meeting was held on [April 12, 2012] at which GE present the clinical documentation for supporting an NDA submission for Flutemetamol F18 Injection.

The 11 studies in the clinical development program for Flutemetamol F 18 Injection were designed to provide safety and efficacy data for worldwide regulatory approval for the target indication presented above. Data were collected from 761 subjects enrolled in 10 clinical studies, including

- 2 Phase 1 studies (ALZ103 and GE067-014 for determination of biodistribution and dosimetry),
- 1 Phase 2 study (ALZ201 for differentiating Healthy volunteer (HV) and probable AD (pAD) subjects),
- 7 Phase 3 studies
    - GE067-005 for prediction of conversion from MCI to pAD,
    - GE067-007 (autopsy study),
    - GE067-008, GE067-009, GE067-010, and GE067-011 for evaluating correlation between SUVR and brain amyloid level,
    - GE067-015 for specificity in young healthy volunteers (YHVs).

All studies are open-label studies. GE067-005 is still on-going and the other nine studies are completed. Populations studied in the clinical development program included European, Japanese and US healthy volunteers (HVs), subjects with amnestic Mild Cognitive Impairment (aMCI), probable Alzheimer's disease (pAD), known or suspected normal pressure

Reference ID: 3334092

hydrocephalus (NPH), and subjects with a life expectancy of 1 year or less (regardless of primary diagnosis).
The 11th study (Phase 3 electronic reader training study [GE067-021]) enrolled no subjects but analyzed images from the other 10 studies (not including the Japanese studies) to assess the effectiveness of an electronic program for training clinicians in the appropriate interpretation of [18F]flutemetamol PET images. This study was requested by the US FDA in an [advice letter of June 20, 2011] and discussed with the FDA during a [Type C Meeting held on October 25, 2011].

2.1.3. Specific Studies Reviewed

Studies GE067-007, GE067-015, and GE067-021 were selected for full statistical review and evaluation. GE067-007 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 GE067-007. GE067-015 included young healthy volunteers, who are assumed to be negative cases for evaluating specificity. GE067-021 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 GE067-021 used the web-training approach in the whole clinical program.

A summary of the 3 studies is shown in Table 1.

2.1.4. Major Statistical Issues

GE067-007:
In-person training was used, which will not be used in clinical practice. Only end-of-life subjects were enrolled, who are not in the intended patient population. Only sensitivity analysis is included in the proposed primary analysis. The sensitivity may be biased without evaluating specificity together.

GE067-015:
In-person training was used, which is not used in clinical practice. Only specificity analysis is included in the proposed primary analysis. The images from Young Healthy Volunteers (YHVs) were mixed with the ones from MCI subjects, and were read by blinded readers is assumed that the YHVs had negative amyloid level and there is no SoT for MCI subjects. It is impossible to evaluate the sensitivity in this study. The specificity may be biased without evaluating sensitivity together. For any given diagnostic test there is a trade-off between sensitivity and specificity. Therefore, to properly interpret diagnostic test performance, sensitivity and specificity must be evaluated together using the same study sample and blinded readers.

GE067-021:
This study includes both sensitivity and specificity in the primary analyses. However, the 135 subjects in the primary analysis population include 68 autopsy subjects, 36 subjects with biopsy, and 31 YHVs (SoT: 54 positive + 84 negative cases). The primary population does not include any MCI subjects, which is the intended population.

Reference ID: 3334092

Including the 31 YHVs (assumed to be negative cases) may lead to inflated specificity. Without the YHVs, the sample size is 53 for specificity evaluation and 51 for sensitivity evaluation. This sample size is large enough for the evaluation without the YHVs.

Table 1: Summary of pivotal clinical efficacy studies GE067-007, GE067-015, and GE067-021

	Phase and Design	Primary objective	Treatment period and follow-up	# of Subjects per Arm	Study Population
GE067-007	Phase 3, Multi-center open label single arm	sensitivity of blinded visual interpretations of flutemetamol F18 PET images *without* anatomic brain images for detecting brain amyloid.	Injection within 40 seconds; 24-hr safety follow-up post injection	180 scanned 69 underwent brain autopsy (68 evaluable)	≥55 yr with short life expectancy
GE067-015	Phase 3, Multi-center, open-label Single arm	specificity of F 18 Injection for excluding the presence of brain amyloid based on the visual assessment of a PET scan by independent blinded readers	Bolus injection within 40 seconds; 24-hr safety followup post injection	181	healthy young adult subjects aged 18 to 40.
GE067-021	Phase 3, Single arm, open-label, pool read, no new enrollment	Sensitivity and specificity of image interpretation of 5 blinded readers (trained by an electronic program) compared with SoT from autopsy. Reader agreement as important secondary endpoint	No subject was dosed in this study	305 images from 276 subjects (including 29 randomly interspersed duplicate images to assess within reader reproducibility)	68 EOL, 36 NPH, 31 YHV, 80 aMCI, 28 EHV, 33 pAD

aMCI = amnestic mild cognitive impairment; EHV= elderly healthy volunteer; EOL= end of life; HV = healthy volunteers; IHC = Immunohistochemical; NPH = normal pressure hydrocephalus; pAD = probable Alzheimer's disease; PET = positron emission tomography; SoT = Standard of truth; YHV= young healthy volunteer

Reference ID: 3334092

2.2 Data Sources

All materials reviewed including the applicant study reports, data sets and literature referenced are provided electronically, and the full electronic path of the documents are \\Cdsesub1\evsprod\NDA203137\0000

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

Data sets analyzed for study **GE067-007, GE067-015, and GE0-67-021** (with data definition document) were located in M5.

The datasets analyzed include adsl.xpt, adzi.xpt, and adzp.xpt for GE067-007, adzi.xpt for GE067-015, and adeff.xpt for GE067-021.

The data follow ADAM and SDTM data format.

3. STATISTICAL EVALUATION

3.3 Data and Analysis Quality

It is possible to reproduce the primary analysis dataset, and in particular the primary endpoint, from the original data source.

It is easy to find the related data. Moderate level of effort is needed to process the data.
One statistical Information Request (IR) was sent to the sponsor for the data with baseline clinical diagnosis. A data, diagsot.xpt, were submitted and delivered to the reviewers by the project manager.

3.4 Evaluation of Efficacy

3.4.1 Study Design and Endpoints

GE067-007:

Study design:
This was a multicenter, open-label PET study to evaluate the efficacy and safety of a single intravenous (i.v.) dose of Flutemetamol F 18 Injection for the detection of cortical fibrillar amyloid β. Subjects underwent anatomic brain imaging (usually CT), followed by open-label i.v.

Reference ID: 3334092

administration of approximately 185 to 370 megabecquerels (MBq) (5 to 10 millicuries [mCi]) of study drug, followed by PET imaging of the brain.

Subjects were followed until death or study termination, whichever occurred first. The brains of subjects who died during the study were analyzed for neuritic (amyloid) plaque density. Subjects had up to 3 visits: 1) screening within 35 days before the administration of study drug, 2) a second screening visit if the anatomic imaging and clinical assessments could not be performed during a single visit, and 3) the PET imaging visit.

On the day of PET imaging, each subject received a single i.v. dose of Flutemetamol F 18 Injection, given within approximately 40 seconds. PET imaging for 30 minutes was to start approximately 90 minutes after the injection. Subjects and their families were blinded to imaging results during the study.

PET images without corresponding anatomic images were interpreted visually by 5 independent readers, trained in the interpretation of such images (in-person training) and blinded to the standard of truth (SoT) and all other information about the subject.

Images were classified as normal (negative for fibrillar amyloid β) or abnormal (positive for fibrillar amyloid β). Each interpretation of a PET image set without the anatomic images had to be recorded and the read session concluded before interpretation of another image set could be conducted.

After all of the PET-only image sets had been interpreted, the order of PET image presentation was re-randomized and each PET scan was presented together with corresponding anatomic images to the reader.

Personnel conducting image preparation and those facilitating the visual image interpretations were blinded to autopsy results as well as all clinical information about the subjects.

Primary endpoints: Blinded visual interpretation of each subject's Flutemetamol F 18 Injection brain PET images as normal or abnormal, *without* anatomic brain images for reference
Secondary endpoints:

- Blinded visual interpretation, *with* anatomic CT brain images for reference
- Global and region-specific estimates of cortical tracer uptake (SUVR value) determined from quantitative analysis of the [18F]flutemetamol PET image, normalized for the mean uptake in reference regions known to have low or no amyloid (cerebellum and pons)

Standard of Truth (SoT): SoT results were available for those subjects who died during the study and underwent brain autopsy. The SoT used to classify the accuracy of image interpretations was the categorization of each autopsied subject's brain as either normal or abnormal for neuritic (amyloid) plaque density, based on a pre-defined threshold (1.5) of neuritic plaque density score; brain tissue assessments were based on Bielschowsky silver stain.

Specimens were taken from 8 cortical regions in the left hemisphere: the precuneus, mid-frontal, superior temporal, middle temporal, inferior parietal, anterior cingulate gyrus, posterior cingulate

Reference ID: 3334092

gyrus, and primary visual cortex regions of the brain. There are a total of 48 slides for the regions.

For each slide, each of five 100x-magnification grey-matter fields were scored from 0 to 3 for neuritic plaque density using a modified Consortium to Establish a Registry for Alzheimer's Disease (CERAD) scale: 0 = none (0 plaques); 1 = sparse (1 to 5 plaques); 2 = (6 to 19 plaques); 3 = (≥20 plaques).

The five scores per slide were averaged to give a mean slide score, and the 6 mean slide scores per region were averaged to give a mean region score. Each region score was classified *normal* if < 1.5, and *abnormal* if > 1.5. A brain was classified abnormal if at least one region was classified as abnormal.

GE067-015:

Study design:

This was a multi-center, open-label PET study to determine the specificity of Flutemetamol F 18 Injection for excluding the presence of brain amyloid in healthy young adult subjects aged 18 to 40 based on visual PET image assessment.

Each subject attended a screening visit 45 days before their Flutemetamol F 18 Injection. A second screening visit occurred if the MRI and clinical assessments could not be performed in a single visit (in the first visit). On the day of PET scanning, each subject was administered Flutemetamol F 18 injection and underwent a PET scan of 30 minutes duration. Adverse events (AEs) were monitored during the course of the study and at a 24-hour post-injection follow-up telephone call. Subjects were instructed to report serious adverse events (SAEs) occurring within 30 days of administration of Flutemetamol F 18 Injection for which a causal relationship could not be ruled out.

The visual assessment of [18F]flutemetamol PET images was performed by 5 independent blinded readers trained (in-person training) in the evaluation of PET brain amyloid imaging.

Primary Efficacy Endpoint: The primary endpoint was visual assessment of the individual [18F]flutemetamol PET images classified as normal or abnormal for the presence of brain amyloid based on independent blinded reads conducted by 5 blinded independent readers. All scans from this study were blindly and randomly mixed with approximately equal numbers of [18F]flutemetamol scans from the GE-067-005 mild cognitive impairment study (which was expected to contain some abnormal images) to avoid potential bias if readers saw only (or predominantly) normal scans.

Secondary Efficacy Endpoint: The secondary efficacy endpoint was the composite SUVR defined as an average of frontal, anterior cingulate, parietal, lateral-temporal and posterior cingulate/precuneous uptake following administration of Flutemetamol F 18 Injection.

Reference ID: 3334092

Standard of Truth: The subjects evaluated in this study, healthy young adults aged 18 to 40, are presumed to be amyloid negative and this assumption constituted the standard of truth in this study.

GE067-021:

Study design:

This study evaluated the effectiveness of an electronic program for training technologists and image readers *via* blinded visual interpretations of [18F]flutemetamol PET image sets selected from previous flutemetamol clinical trials.

There are 305 [18F]flutemetamol images including 276 unique images from subjects with histopathological confirmation of their brain amyloid status (n = 104), young healthy volunteers (n = 31), elderly healthy volunteers (n = 28), subjects with clinically probable Alzheimer's disease (pAD; n = 33), and subjects with amnestic mild cognitive impairment (aMCI; n = 80).

The selected subjects represent the entire spectrum of cognitive health, from cognitively normal, to aMCI, to dementia (including AD). Duplicates of 29 images (approximately 10%) randomly selected from the 276 unique images were inserted in random order into the set of 276 images (to give a total of 305 images) to determine intra-reader reproducibility (IRR).

The complete, oriented PET-only image set was randomized.

Five physicians (3 nuclear medicine physicians and 2 radiologists) ("reader trainees") who were untrained and inexperienced in the orientation and interpretation of amyloid images were asked to complete an electronic reader training program.

To qualify as a reader for this study, each reader-candidate had to watch the training program, correctly answer the review questions provided by the program, and correctly interpret the test images at the end of the program, without assistance from anyone else. The criteria for successful reader training completion and qualification was at least 90% agreement between the reader-candidate's interpretations of the test images and the previously documented interpretations of the same images by a consultant physician with extensive experience in flutemetamol imaging.

After completing the training, the readers read the images according to the random schedule. None of the images used in the electronic training program were included in the blinded image interpretation.

The blinded visual interpretation of each image as *normal* or *abnormal* was recorded on a case report form (CRF), along with the reader's self-rating of confidence in his/her image interpretation; the read was a forced choice between *normal* and *abnormal*, and readers were not allowed to exclude any images from interpretation.

Reference ID: 3334092

During the blinded visual image interpretation, if any reader decided that re-orientation of an image was needed, he/she could either re-orient the image him/herself or request that re-orientation be done by an on-call NMT who was available during the read.

The results of the blinded visual image interpretation were used as measures of the effectiveness of the reader training program, and provided measures of validity (sensitivity and specificity) and reproducibility (Intra-reader agreement and Intra-reader reproducibility).

Analysis Population 1 (135 subjects with SoT) was defined to include 68 autopsy subjects from GE067-007; 36 subjects with biopsy from GE067-009, -010, and 011; and 31 YHVs from GE067-015.
The positive cases determined by SoT for sensitivity evaluation is 51 (43 from autopsy, 8 from biopsy). The negative cases determined by SoT for sensitivity evaluation is 84 (25 from autopsy, 28 from biopsy, and 31 from YHVs).

Analysis Population 4 (276 subjects) was defined to include population 1 and subjects without SoT (60 MCI subjects from GE067-005, and 55 subjects from ALZ201, 10 subjects from ALZ 103, and 16 from ALZ 104).

3.4.2 Statistical Methodologies

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

GE067-007

Primary Efficacy Evaluation: Sensitivity of the blinded visual interpretation of images *without* the aid of anatomic images was determined for each of the 5 independent readers; point estimates and their exact 95% binomial confidence interval limits were calculated. The study was to be deemed successful if the lower bound of the two-sided 95% exact confidence limit for sensitivity was greater than the *a priori* threshold of 70% for at least 3 of the 5 readers.

Both sensitivity and specificity should be evaluated in the primary analyses by reader, in order to avoid bias. The lower bound of 95% confidence limit for both sensitivity and specificity should be > pre-specified thresholds for the same 3 out of 5 readers.

Secondary Efficacy Evaluations:
(1) The specificity of the blinded visual image interpretations *without* anatomic images was determined as point estimates and their exact 95% binomial confidence interval limits for each of the 5 readers.
(2) Sensitivity and specificity of blinded visual image interpretations *with* the aid of anatomic images were determined as point estimates and their exact 95% binomial confidence interval limits for each of the 5 readers.
(3) The levels of association between quantitative estimates of [18F]flutemetamol brain uptake (SUVR values) and *post-mortem IHC* estimates of amyloid were determined on a regional and global basis, for the first 30 brains to undergo autopsy.

Reference ID: 3334092

GE067-015

Primary efficacy analysis: Specificity with 95% confidence intervals were provided for the efficacy population for each of 5 blinded readers. The study was to be deemed successful if at least 3 of 5 blinded readers demonstrated specificity with a lower bound of the 95% confidence interval that exceeded 80%.

Secondary efficacy analyses: The exact 95% binomial confidence interval for specificity was calculated and presented for individual sub-populations based on demographic factors, clinical site, and the commonly used PET scanners used in the study:
· Gender
· Age [18 to ≤ 30 and > 30 to ≤ 40]
· Ethnicity
· Race
· Clinical site
· Geographic location
· Scanners type

Sensitivity should be evaluated together with specificity by readers. Since the MCI subjects from GE67-005 do not have the SoT, and all young healthy volunteers were assumed to be negative cases. Sensitivity cannot be evaluated in this study. Specificity may be biased due to lack of sensitivity evaluation.

GE067-021

Primary Endpoint and analyses:
The primary endpoint was the blinded visual interpretation of the PET images. Each of the blinded readers interpreted the images as either *normal* or *abnormal* for brain uptake of [18F]flutemetamol. For the 29 subjects whose image sets were read twice, the first image interpretation was used for the primary efficacy analysis.

Sensitivity and specificity of the blinded visual interpretation without anatomic images were determined from subjects with any SoT type (i.e., brain amyloid status either confirmed histopathologically (from specimens obtained at biopsy or autopsy) or assumed clinically based on age (40 years or less) and healthy status) (**Analysis Population 1**; n = 135).

Sensitivity and specificity were determined for each reader and by majority result (i.e., the image interpretation made by each of 3 or more of the 5 readers). The study was to be considered successful if the null hypothesis for sensitivity and specificity was rejected by each of 3 (or more) of the 5 blinded readers. The lower bounds of the two-sided 95% CIs for sensitivity and specificity should be each **>70% for each of 3 (or more) of the 5** blinded readers.

Secondary Analyses
1) Sensitivity, specificity, PPV, and NPV of the blinded visual interpretation *with* anatomic images for Analysis Population 1

Reference ID: 3334092

2) Sensitivity, specificity, PPV, and NPV without and with anatomic images for the 104 subjects with autopsy and biopsy; for the 68 subjects with autopsy

3) Inter-Reader Agreement of blinded visual image interpretations *without* anatomic images using Fleiss' kappa (κ) score and associated p-value (**Analysis Population 4**; n = 276)

The kappa value was reported according to the following categorical classification: >0.6 = good; >0.7 = very good; >0.8 = excellent. The study would have met its secondary objective regarding inter-reader agreement if the lower bound of the **two-sided 95% CI for kappa was >0.6.**

4) Fleiss' Kappa for the subset of MCI subjects (n = 80).

5) Intra-Reader Reproducibility of blinded visual image interpretation *without* anatomic images using percentage agreement of the first and second image interpretations for the 29 images (approximately 10% of 276) that were re-read

NPV and PPV will not be evaluated in this review.

3.4.3 Patient Disposition, Demographic and Baseline Characteristics

Only patients in GE067-021 were summarized in this section because the patients with autopsy from GE067-007 and selected subjects from GE067-015 are sub-population in the GE067-021 study.

In GE067-021, the effectiveness of an electronic training program for Vizamyl image orientation and interpretation was evaluated using Vizamyl 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 (68 patients who underwent an autopsy and 36 known or suspected normal pressure hydrocephalus patients with in vivo brain biopsy) and without a truth standard (28 cognitively normal volunteers 55 years or above, 80 patients with amnestic mild cognitive impairment (aMCI), 33 subjects with probable AD (pAD)), and 31 young healthy volunteers. Additionally, intra-reader reproducibility was assessed from 29 images (10%). Among the 276 subjects, the median age was 72 years (range 20 to 95), 136 were females, and 251 were Caucasian (WHITE).

For the autopsy subjects, the time interval between the Vizamyl scan and death ranged from 0 to 13 months, with a median of 2.6 months, and was less than one year for 66 patients and between 12 to 13 months for 2 patients. At autopsy, the global brain neuritic plaque density category (CERAD classification) was (N=68): frequent n= 19; moderate n=23; sparse n=14; and none n=12.

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

Reference ID: 3334092

Table 2: Patient disposition (% in ()) for GE067-021 for different patient populations

Group		Population 4 (n=276)	Population 1 (n=135)	autopsy subjects from GE067-007 (n=68)	Biopsy subjects (n=36)	YHVs from GE067-015 (n=31)	Subjects without SoT (n=141)
Gender	F	136 (49)	65 (48)	35 (51)	12 (33)	18 (58)	71 (50)
	M	140 (51)	70 (52)	33 (49)	24 (67)	13 (42)	70 (50)
Age	>=65	198 (72)	94 (70)	65 (96)	29 (81)	0	104 (74)
	<65	78 (28)	41 (30)	3 (4)	7 (19)	31 (100)	37 (26)
Age	Median (range)	72 (20-95)	73 (20-95)	82 (60-95)	70 (53-77)	25 (20-40)	71 (54-91)
Race	White	251 (91)	127 (94)	64 (94)	36 (100)	27 (87)	124 (88)
	Other	25 (9)	8 (6)	4 (6)	0	4 (13)	17 (12)
Clinical diagnosis	AD	63 (23)	30 (22)	30 (44)	0	0	33 (23)
	MCI	80 (29)	0	0	0	0	80 (57)
	Normal	49 (18)	21 (16)	21 (31)	0	0	28 (20)
	OD	53 (19)	53 (39)	17 (25)	36 (100)	0	0
	YHV	31 (11)	31 (23)	0	0	31 (100)	0

AD: Alzheimer's disease. MCI: Mild cognitive impairment. Normal: Cognitive normal. OD: other dementia. YHV: young healthy volunteer

3.4.4 Results and Conclusions

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

GE067-007:

Sensitivity: The point estimates (95% confidence interval limits) for the sensitivity *without* anatomic brain images were 81% (67%, 92%), 88% (74%, 96%), 93% (81%, 99%), 93% (81%, 99%), and 88% (75%, 96%) for Readers 1 through 5 respectively. Because the lower bound of the 95% confidence limit for sensitivity exceeded 70% for at least 3 of the 5 readers, the primary efficacy objective was achieved.

Specificity: The point estimates (95% confidence interval limits) for specificity *without* v images were 88% (69%, 98%), 92% (74%, 99%), 44% (24%, 65%), 80% (59%, 93%), and 92% (74%, 99%) for Readers 1 through 5, respectively.

Reference ID: 3334092

The point estimates (95% confidence interval limits) for sensitivity *with* anatomic images were 91% (78%, 97%), 95% (84%, 99%), 98% (88%, 100%), 91% (78%, 97%), and 91% (77%, 97%) for the same 5 readers, respectively.

The point estimates (95% confidence interval limits) for specificity *with* anatomic images was 92% (74%, 99%), 88% (69%, 98%), 56% (35%, 76%), 88% (69%, 98%), and 92% (74%, 99%) for the same 5 readers, respectively.

The results for sensitivity and specificity *with* and *without* anatomic images are not very different.

GE067-015:

The results of blinded reads by 5 independent readers showed specificity of 100%, 68%, 99%, 99% and 99%. The lower bounds of the 95% confidence intervals for specificity for the 5 readers were 98%, 61%, 97%, 97% and 96%, exceeding 95% for 4 of the 5 readers. Since the primary efficacy objective was for the lower bound of the confidence interval to exceed 80% for at least 3 of 5 readers, the primary objective was achieved. There were no notable differences in specificity observed across any of the demographic categories (gender, age category, ethnicity, and race), clinical site, geographic location, and scanner type.

GE067-021:

By reader, the sensitivity of the blinded visual interpretation of PET images alone ranged from 84% to 94% with a median value of 92%. The specificity ranged from 77% to 96% by reader, with a median value of 81%. The PPV ranged from 73% to 94% by reader, with a median value of 75%. The NPV ranged from 91% to 96% by reader, with a median value of 94%. The study met the primary success criterion, because the lower bounds of the 2-sided 95% CI for both sensitivity and specificity exceeded 70% for each of 3 readers. The high NPVs indicate that a negative [18F]flutemetamol PET scan can reliably rule out brain amyloid and thus AD.

The availability of anatomic images (CT) for consultation during the blinded visual interpretation of PET images did not significantly affect the results.

For blinded visual interpretations of PET images alone, Inter-reader agreement (Analysis Population 4; n = 276) across all 5 readers was 81%, with a kappa value of 0.83 and 95% CI as (0.79, 0.86). Inter-reader agreement met the secondary success criterion, which required the lower bound of the 2-sided CI for kappa to exceed 0.6.

Intra-reader agreement ranged (IRR) from 93% to 100% among the five blinded readers, with most readers having an IRR of 97% or higher.

Reference ID: 3334092

Subset analysis showed no meaningful difference in effectiveness between the two methods of reader training (electronic vs. in-person), although there was some improvement in specificity with the electronic program.

The reviewer's results are consistent with the sponsor's results for the three studies. Note that for GE067-007, the specificity without anatomic images is not as good as the corresponding sensitivity results. 3 out of 5 readers had lower bound of 95% CI for specificity <70% (2 readers having the lower bound <60%). For study GE067-021, the specificity is not as good for patients excluding YHVs (point estimate ranged from 60% to 92%, and lower bound of 95% CI ranged from 49% to 74%).

The reviewer's findings are presented below. "Negative" will be used instead of "Normal". "Positive" will be used instead of "Abnormal".

GE067-007:

Sensitivity and specificity of the blinded reads versus SoT (from autopsy) are presented by reader in Table 3. There are a total of 43 positive cases and 25 negative cases based on histopathology. Reader B and G had one un-interpretable image (treated as missing value). The sensitivity values for reader B and G will be lower if mismatch approach is used to impute the one missing read value.

Without anatomic images, the lower bounds of 95% CI for sensitivity and specificity are both >70% only for the same 2 out of the 5 blinded readers. The results with and without anatomic images are similar.

Table 3: Sensitivity and specificity (%) with and without anatomic images by reader

		Without anatomic images				*With anatomic images*			
Reader	Sample size	Sen	95% CI	spe	95% CI	sen	95% CI	spe	95% CI
A	68	81.4	(66.6, 91.6)	88	(68.8, 97.5)	90.7	(77.8, 97.4)	92	(74.0, 99.0)
B	67	88.1	(74.4, 96.0)	92	(74.0, 99.1)	95.2	(83.8, 99.4)	88	(68.8, 97.4)
C	68	93.0	(80.9, 98.5)	44	(24.4, 65.1)	97.7	(87.7, 99.9)	56	(34.9, 75.6)
G	67	92.9	(80.5, 98.5)	80	(59.3, 93.2)	90.7	(77.7, 97.5)	88	(68.8, 97.5)
H	68	88.4	(74.9, 96.1)	92	(74.0, 99.0)	90.5	(77.6, 97.3)	92	(74.0, 99.0)

Reference ID: 3334092

GE067-015:

Sensitivity and specificity cannot be evaluated for the MCI subjects from GE067-005 without SoT. Assuming all the 181 YHVs are negative cases, as shown in Table 4, the specificity is very good for four readers (lower bound of 95% CIs >96% by reader), not very good for one reader with specificity as 68% (95% CI: 61%-75%).

Table 4: Specificity for the 181 Young Healthy Volunteers (YHVs)

Reader	Specificity (%)	95% CI (%)
A	100	(98, 100)
B	68	(60.6, 74.7)
C	99	(97, 100)
G	99	(97, 100)
H	99	(96.1, 100)

GE067-021:

Vizamyl performance characteristics for patients with different analysis populations are shown in Table 5. For the primary analysis population (n=135), without anatomic images and with web-based training, the lower bounds of 95% CI for sensitivity and specificity are both >70% for **3 out of the 5** blinded readers (met the pre-specified criteria for success).

Note that this primary analysis population included 68 autopsy subjects, 36 biopsy subjects and 31 YHVs. There are 51 positive case and 84 negative cases based on histopathology, or biopsy, or assumption for young healthy volunteers. The inclusion of the 31 YHVs may lead to inflated specificity estimation.

Excluding the 31 YHVs, for the 104 subjects with SoT from histopathology or biopsy, the lower bounds of 95% CI for sensitivity and specificity are both >70% for **2 out of the 5** blinded readers.

For the autopsy subjects only, without anatomic images and with web-based training, the lower bounds of 95% CI for sensitivity and specificity are both >70% only for **1 out of the 5** blinded readers. The results using in-person training is slightly better than those using the web-based training (Table 3: in-person training versus Table 5: web-based training).

For the 36 biopsy subjects, the population for sensitivity is only 8, which is too small to make conclusion on the sensitivity. The specificity is slightly better than that obtained from the autopsy only population.

By reader sensitivity and specificity (point estimate and 95% CIs) with anatomic images and web-based training are similar to the results without anatomic images (more details are included in Table A2 in Appendix).

Image reproducibility for various subject groups in GE067-021 is presented in Table 6. For the analysis population 4, inter-reader reproducibility analysis showed an overall Fleiss' kappa

Reference ID: 3334092

statistic of 0.83 (95% CI 0.79 to 0.86) which met the pre-specified success criterion (95% CI lower bound >0.60), and which is consistent with the percent of agreement results (5 out of 5 readers agree on 81% of the images). For the other populations, the kappa values ranged from 0.74 to 0.90 except the 36 biopsy subjects. The results on percent of agreement are consistent with the results from kappa statistics.

Table 5: By reader sensitivity and specificity (point estimate and 95% CIs) without anatomic images for different analysis populations

primary					
Analysis population 1 (n=135=51 pos + 84 neg)					
	reader	sensitivity %	sen CI	specificity %	spe CI
	1	94.12	(83.76, 98.77)	78.57	(68.26, 86.78)
	2	92.16	(81.12, 97.82)	80.95	(70.92, 88.70)
	3	90.2	(78.59, 96.74)	92.86	(85.10, 97.33)
	4	94.12	(83.76, 98.77)	77.38	(66.95, 85.80)
	5	84.31	(71.41, 92.97)	96.43	(89.92, 99.26)
Secondary					
68 autopsy subjects (43 pos and 25 neg)					
	Reader	Sen	sen CI	Spe	spe CI
	1	93.02	(80.94, 98.54)	72	(50.61, 87.93)
	2	93.02	(80.94, 98.54)	84	(63.92, 95.46)
	3	90.7	(77.87, 97.41)	88	(68.78, 97.45)
	4	93.02	(80.94, 98.54)	60	(38.67, 78.88)
	5	86.05	(72.07, 94.71)	92	(73.97, 99.02)
104 subjects with SoT (68 autopsy + 36 biopsy)=51 pos and 53 neg					
	reader	Sen	sen CI	Spe	spe CI
	1	94.12	(83.76, 98.77)	67.92	(53.68, 80.08)
	2	92.16	(81.12, 97.82)	79.25	(65.90, 89.16)
	3	90.2	(78.59, 96.74)	88.68	(76.97, 95.73)
	4	94.12	(83.76, 98.77)	71.7	(57.65, 83.22)
	5	84.31	(71.41, 92.97)	94.34	(84.34, 98.82)
36 biopsy subjects (8 pos and 28 neg)					
	reader	Sen	sen CI	Spe	spe CI
	1	100	(63.06, 100)	64	(44.07, 81.36)
	2	87.5	(47.35, 99.68)	75	(55.13, 89.31)
	3	87.5	(47.35, 99.68)	89	(71.78, 97.74)
	4	100	(63.06, 100)	82	(63.10, 93.93)
	5	75	(34.91, 96.82)	96	(81.65, 99.91)

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

Reference ID: 3334092

Table 6: Inter-Reader agreement for different analysis populations

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
All 276 subjects (population 4)	139	0.83 (0.79, 0.86)	5	14	81
All subjects with a SoT, n=104 (68 autopsy; 36 biopsy)	58	0.74 (0.68, 0.80)	6	24	70
All subjects without a SoT including 31 YHVs, n = 172	76	0.88 (0.83, 0.92)	5	8	87
All subjects without a SoT excluding the 31 YHVs, n = 141	75	0.90 (0.85, 0.95)	4	7	89
Autopsy subjects n=68	44	0.78(0.70,0.85)	3	21	76
Biopsy subjects n=36	13	0.58(0.48,0.69)	11	31	58

[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.

Intra-reader agreement

Intra-reader reproducibility analysis showed that, between the two readings for each of the 29 duplicate patient images, one of the five readers had complete agreement for all 29 images, two readers had discordant reads for a single image, and three readers had discordant reads for two images. The percent of agreement for all the 29 images is 100%, 97%, 93%, 97%, and 93% for reader 1, 2, 3, 4, and 5, respectively.

Intra-reader reproducibility for a sub-group of 8 images from aMCI patients showed that all five readers had complete agreement for all duplicate images.

3.5 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 GE067-021 (autopsy population as one stratum) because web-based training (used in clinical practice) was only used in GE067-021. Only reviewer's results are presented below.

For GE067-015, subgroup analyses (by gender, race, model, country, etc) showed no difference by subgroups (results are shown in Table A in Appendix).

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.

Reference ID: 3334092

4.1 Gender, Race, and Age

Since majority of the subjects is WHITE (more than 90%) for the 276 subjects (population 1) and in the 68 autopsy subjects in study GE-67-021, subgroup analyses by race will not be conducted. More than 90% subjects are older than 65 in the autopsy subjects and more than 87% subjects are older than 65 in the autopsy plus biopsy subjects, sensitivity and specificity by age group will not be conducted.

As shown in Table 7, sensitivity is slightly lower, and specificity is slightly higher in Male compared with Female for most of the readers, especially for the autopsy population (68 subjects). As shown in Table 8, there is no difference in reader agreement by gender and age groups.

Table 7: Sensitivity (sen) and specificity(spe) by gender (in %). pos for positive cases, neg for negative cases based on histopathology, or biopsy, or negative assumption for YHVs)

104 subjects (68 autopsy and 36 biopsy)					68 autopsy subjects			
reader	n=pos+neg	gender	Sen	spe	n=pos+neg	gender	sen	spe
1	57=28+29	M	93	69	33=21+12	M	90	75
	47=23+24	F	96	67	35=22+13	F	95	69
2	57=28+29	M	86	83	33=21+12	M	86	92
	47=23+24	F	100	75	35=22+13	F	100	77
3	57=28+29	M	93	86	33=21+12	M	90	92
	47=23+24	F	87	92	35=22+13	F	91	85
4	57=28+29	M	93	72	33=21+12	M	90	58
	47=23+24	F	96	71	35=22+13	F	95	62
5	57=28+29	M	82	97	33=21+12	M	81	100
	47=23+24	F	87	92	35=22+13	F	91	85

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

Table 8: Reader agreement evaluation by gender and age

Subject Group by gender and age group	**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=140)	57	0.81 (0.76, 0.86)	4	18	79
Gender: Female (n=136)	80	0.84 (0.78, 0.89)	7	10	83
age>=65 (n=198)	116	0.83 (0.79, 0.87)	5	13	84
Age<65 (n=78)	21	0.74 (0.67, 0.81)	5	17	78

[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.

Reference ID: 3334092

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.

The sensitivity and specificity values are not very different in the different clinical diagnosis groups. Only one group (Other dementia) has low specificity values (point estimates ranged from 33% to 81%) in Table 9.

AD and MCI subjects had good reader agreement (kappa values>0.8 and percent of 5 out of 5 reader agreement >85%). The reader agreement is lower in the OD and cognitively normal subjects (with kappa values between 60-70%). For the 28 elderly Cognitively normal subjects without SoT, the number of positive scans is only 2. The imbalance in the positive and negative scans led to the low kappa values. Percent of agreement (more than 80% 5 out of 5 reader agreement) should be used instead of kappa in this case.

Reference ID: 3334092

Table 9: Sensitivity (sen) and specificity (spe) point estimates by clinical diagnosis (in %)

104 subjects (autopsy and 36 biopsy)					autopsy subjects				31 YHVs
Reader	n=pos+neg	Basediag	Sen	Spe	n=pos+neg	basediag	sen	spe	spe
1	30=22+8	AD	95	100	30=22+8	AD	95	100	none
	21=10+11	Normal	80	73	21=10+11	normal	80	73	97
	48=16+32	OD	100	56	17=11+6	OD	100	33	none
	5=3+2	OD/AD	100	100	None	none	none	none	none
2	30=22+8	AD	91	100	30=22+8	AD	91	100	none
	21=10+11	Normal	90	91	21=10+11	Normal	90	91	84
	48=16+32	OD	94	69	17=11+6	OD	100	50	none
	5=3+2	OD/AD	100	100	none	none	none	none	none
3	30=22+8	AD	91	100	30=22+8	AD	91	100	none
	21=10+11	Normal	80	100	21=10+11	Normal	80	100	100
	48=16+32	OD	94	81	17=11+6	OD	100	50	none
	5=3+2	OD/AD	100	100	none	none	none	none	none
4	30=22+8	AD	95	63	30=22+8	AD	95	63	none
	21=10+11	Normal	80	64	21=10+11	Normal	80	64	100
	48=16+32	OD	100	75	17=11+6	OD	100	50	none
	5=3+2	OD/AD	100	100	none	none	none	none	none
5	30=22+8	AD	91	100	30=22+8	AD	91	100	none
	21=10+11	Normal	60	100	21=10+11	Normal	60	100	100
	48=16+32	OD	88	91	17=11+6	OD	100	67	none
	5=3+2	OD/AD	100	100	none	none	none	none	none

Note: basediag is for baseline clinical diagnosis. pos is for positive cases, neg is for negative cases based on histopathology, or biopsy, or negative assumption for YHVs

AD: Alzheimer's disease. OD: other dementia. Normal: Cognitive normal

Reference ID: 3334092

Table 10: 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 diagnosis, n=63	47	0.88 (0.80, 0.96)	3	6	90
aMCI, n=80 (without SoT)	45	0.89 (0.82, 0.96)	4	7	89
Elderly cognitively normal without TS, n=28	2	0.46 (0.34, 0.57)	4	14	82
Cognitively normal with SoT[b], n=21	10	0.64 (0.5, 0.77)	5	38	57
Cognitive normal with SoT including 31 YHVs, n=52	12	0.60 (0.51, 0.68)	8	23	69
Other (non-AD) dementia with SoT, n=53[c]	24	0.68 (0.59, 0.77)	8	27	65

[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.

[b] 21 with SoT from autopsy, 0 with SoT from biopsy

[c] 17 from autopsy, 5 of 36 with SoT from biopsy were not definitively classified as pAD based on clinical diagnosis

pAD: probable AD; aMCI: amnestic MCI; Elderly: 55 years or above

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

In GE067-007, in-person training for the readers was used, which will not be used in clinical practice. Only analysis of sensitivity is included in the proposed primary analysis. The sensitivity may be biased without evaluating specificity together. In addition, only end-of-life subjects were enrolled, who are not the intended patient population.

In GE067-015**,** in-person training was used, which is not used in clinical practice. Only specificity analysis is included in the proposed primary analysis. The images from Young Healthy Volunteers (YHVs) were mixed with the ones from MCI subjects, and were read by blinded readers is assumed that the YHVs had negative amyloid level and there is no SoT for MCI subjects. It is impossible to evaluate the sensitivity in this study. The specificity may be biased without evaluating sensitivity together.

For any given diagnostic test there is a trade-off between sensitivity and specificity. Therefore, to properly interpret diagnostic test performance, sensitivity and specificity must be evaluated together using the same study sample and blinded readers.

Study GE067-021 includes both sensitivity and specificity in the primary analyses. However, the 135 subjects in the primary analysis population include 68 autopsy subjects, 36 subjects with biopsy, and 31 YHVs (SoT: 54 positive + 84 negative cases). Including the 31 YHVs (assumed to be negative cases) may lead to inflated specificity. Without the YHVs, the sample size is 53 for specificity evaluation and 51 for sensitivity evaluation. The sample size is large enough for the evaluation without the YHVs. In addition, none of the subjects with SoT included in the

Reference ID: 3334092

primary population are MCI subjects (the intended population). Excluding the YHVs, the accuracy in terms of specificity is lower, which should be reflected in the label.

5.2 Collective Evidence

For GE067-007, without anatomic images and with in-person training, the lower bounds of 95% CIs for sensitivity and specificity are both >70% only for **the same 2 out of the 5 blinded readers.** The results with anatomic images are similar.

Readers with web-based training read the images of the 68 autopsy subjects (from GE067-007) in the pool read study GE067-021. For this sub-population with SoT based on histopathology, the lower bounds of 95% CIs for sensitivity and specificity are both >70% only for the **same 1 out of the 5 blinded readers,** which is slightly worse than the performance with in-person training.

In the pool-read study GE067-021, for the primary analysis population (the subjects with SoT based on histopathology, or biopsy, or assumed negative YHV cases), the sensitivity of the blinded visual interpretation of PET images ranged from 84% to 94% with a median value of 92%. The specificity ranged from 77% to 96% by reader, with a median value of 81%. The study met the primary success criterion, because the lower bounds of the 2-sided 95% CI for both sensitivity and specificity **exceeded 70% for each of 3 out of 5 readers.**

Excluding the 31 YHVs, for the 104 subjects with SoT from histopathology or biopsy, the lower bounds of 95% CI for sensitivity and specificity are both >70% for **2 out of the 5** blinded readers.

The availability of anatomic images (CT) for consultation during the blinded visual interpretation of PET images did not significantly affect the results.

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 blinded visual interpretations of PET images with web-trained readers, the percent of agreement (all five readers agreed with each other) is 81% (Analysis Population 4 including all subjects in the study; n = 276), and the kappa value for inter-reader agreement across all 5 readers is 0.83 with 95% CI as (0.79, 0.86). Inter-reader agreement met the pre-specified secondary success criterion, which required the lower bound of the 2-sided CI for kappa to exceed 0.6 (Table 6). As shown in Table 10, the percent of agreement (all five readers agreed with each other) is 89% for MCI subjects (n = 80), and the kappa value for inter-reader agreement across all 5 readers is 0.89 with 95% CI as (0.82, 0.96). In addition, intra-reader agreement ranged from 93% to 100% among the five blinded readers.

Reference ID: 3334092

5.3 Conclusions and Recommendations

The statistical results in terms of accuracy and reproducibility provide some evidence to support the claims proposed in this NDA. However, the issue of relatively low specificity for subjects with autopsy and biopsy should be noted. There are no MCI subjects with autopsy. Therefore, we cannot obtain conclusive evidence for the performance in terms of sensitivity and specificity on MCI population.

5.4 Labeling Recommendations

1) Recommend using the following indication in the label:

Vizamyl is a radioactive diagnostic agent for Positron Emission Tomography (PET) imaging of the brain to estimate β amyloid neuritic plaque density in adult patients with cognitive impairment who are being evaluated for Alzheimer's disease (AD) or other causes of cognitive decline. A negative Vizamyl 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. A positive Vizamyl scan indicates moderate to frequent amyloid neuritic plaques; neuropathological examination has shown this amount of neuritic plaque is present in patients with AD, but may also be present in patients with other types of neurologic conditions as well as older people with normal cognition. Vizamyl is an adjunct to other diagnostic evaluations.

A positive Vizamyl scan does not establish a diagnosis of AD or other cognitive disorder.
Safety and effectiveness of Vizamyl have not been established for:

- *Predicting the development of dementia or other neurological condition;*
- *Monitoring responses to therapies.*

2) Images should be called negative or positive instead of [redacted] (b) (4) in the label.

3) Recommend removing the description on GE067-015, and keeping only results from GE067-007 (the autopsy study) and GE067-021 (the pool read study). The proposed paragraphs for the two studies in the clinical study section are shown below.

In Study One, one hundred eighty patients were dosed with Vizamyl and 176 were imaged. The median age was 82 years (range 47 to 98 years) and 57% of the patients were female. Forty-four patients had no cognitive impairment, 135 had dementia, no patients had mild cognitive impairment (MCI), and one patient had memory loss of unspecified nature. Sixty-nine patients died during the study; 68 had cerebral cortical amyloid status determined (43 positive and 25 negative) and were included in the primary analysis. The time interval between the Vizamyl scan and death ranged from 0 to 13 months, with a median of 2.6 months, and was less than one year for 66 patients and between 12 to 13 months for 2 patients. At autopsy, the global brain neuritic plaque density category (CERAD classification as in Table 5) was (N=67): frequent n= 19; moderate n=23; sparse n=14; and none n=12.

In Study Two, the effectiveness of an electronic training program for Vizamyl image orientation and interpretation was evaluated using Vizamyl 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 (68 patients who underwent an autopsy and 36 known or suspected normal pressure hydrocephalus patients with in vivo brain biopsy) and without a truth standard (28 cognitively normal volunteers 55 years or above, 80 patients with amnestic mild cognitive impairment (aMCI), 33 subjects with probable AD (pAD)), and 31 young healthy volunteers. Additionally, intra-reader reproducibility was assessed from 29 images (10%). Among the 276 subjects, the median age was 72 years (range 20 to 95), 136 were females, and 251 were Caucasian.

Reference ID: 3334092

Studies One and Two Vizamyl performance characteristics for patients with a with an autopsy-based truth standard are shown in Tables 6 and 7. Among patients who underwent autopsy (n=68; 43 positive and 25 negative based on histopathology), the median (and range) of correct read results, false negatives, and false positives were 61 (56, 62), 4 (1, 4), 3 (2, 11), respectively, for in-person training (Study One); and were 60 (55 to 61), 3 (3 to 6), 4 (2 to 10), respectively, for electronic media training (Study Two). Image reproducibility for various subject groups in Study Two is presented in Table 8. Inter-reader reproducibility analysis showed an overall Fleiss' kappa statistic of 0.83 (95% CI 0.79 to 0.86) which met the pre-specified success criterion (95% CI lower bound >0.60). Intra-reader reproducibility analysis showed that, between the two readings for each of the 29 duplicate patient images, one of the five readers had complete agreement for all 29 images, two readers had discordant reads for a single image, and three readers had discordant reads for two images. Intra-reader reproducibility for a sub-group of 8 images from aMCI patients showed that all five readers had complete agreement for all duplicate images.

4) Recommend new tables (Tables 6, 7 and 8) in the clinical study section. The tables are shown below.

Table 6. Vizamyl Scan Results by Reader Training Method among Patients with Autopsy (n = 68)

Test Performance		In-Person Training (Study One)	Electronic Media Training (Study Two)
Sensitivity (%)	Median	88	93
	Range among the 5 readers	(81 – 93)	(86 – 93)
Specificity (%)	Median	88	84
	Range among the 5 readers	(44 – 92)	(60 – 92)

Table 7: Vizamyl Scan Interpretations by Reader Training Method among Autopsied Patients (n = 68)

		In-Person Training (Study One)					Electronic Media Training (Study Two)				
		Reader					Reader				
		1	2	3	4	5	6	7	8	9	10
All scans with autopsies (n = 68[a])	Correct	62	62	56	61	61	58	61	61	55	60
	False Negative	4	2	1	4	4	3	3	4	3	6
	False Positive	2	3	11	3	2	7	4	3	10	2

[a] 43 positive and 25 negative based on histopathology

Table 8 (Study 2): Median Number of Positive Vizamyl Scans and Reproducibility of Scan Results

Subject Group by Cognitive and Truth Standard (TS)	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
All 276 subjects	139	0.83 (0.79, 0.86)	5	14	81
All subjects with a TS, n=104 (68 autopsy; 36 biopsy)	58	0.74 (0.68, 0.80)	6	24	70
All subjects without a TS, n = 172	76	0.88 (0.83, 0.92)	5	8	87
pAD, n=63 (30 with TS[b]; 33 no TS)	47	0.88 (0.80, 0.96)	3	6	90
aMCI, n=80 (without TS)	45	0.89 (0.82, 0.96)	4	7	89
Elderly cognitively normal without TS, n=28	2	0.46 (0.34, 0.57)	4	14	82
Cognitively normal with TS[c], n=21	10	0.64 (0.5, 0.77)	5	38	57
Other (non-AD) dementia with TS, n=53[d]	24	0.68 (0.59, 0.77)	8	27	65

Reference ID: 3334092

[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.
[b] 30 with TS from autopsy, [c] 21 with TS from autopsy, 0 with TS from biopsy
[d] 17 from autopsy, 5 of 36 with TS from biopsy were not definitively classified as pAD based on clinical diagnosis
pAD: probable AD; aMCI: amnestic MCI; Elderly: 55 years or above

APPENDICES

Table A1: Specificity (%) in different groups for GE067-015

	Group	Sample size	reader 1	2	3	4	5
Gender	F	103	100	74.8	100	100	100
	M	78	100	59.0	98.7	98.7	97.4
Age	18 to <=30	116	100	64.7	99.1	99.1	98.3
	>30 to <=40	65	100	73.9	100	100	100
Ethnicity	Hisp	74	100	71.6	100	100	100
	not Hisp	107	100	65.4	99.1	99.1	98.1
Race	White	156	100	66.7	99.4	99.36	98.7
	Black	19	100	79.0	100	100	100
	others (small size)	6					
Model	GE disc 690	27	100	70.4	100	100	100
	GE disc RX	20	100	65	100	100	95
	GE disc ST	35	100	42.9	97.2	97.1	97.1
	GE disc STE	7	100	71.4	100	100	100
	simens biograph	87	100	72.4	100	100	100
	simens HR+	5	100	60	100	100	100
Country	Bel	10	100	50	100	100	100
	FIN	19	100	79.0	100	100	100
	GBR	25	100	64	100	100	96
	USA	127	100	68.5	99.2	99.2	99.2

Reference ID: 3334092

Table A2: sensitivity and specificity , and corresponding 95% CIs (in %) with anatomic images and web-based training (Study GE067-021)

Analysis population 1 (n=135=51 pos + 84 neg)					
	reader	sensitivity	95% CI	specificity	95% CI
	1	96.08	(86.54, 99.52)	83.33	(73.62, 90.58)
	2	92.16	(81.12, 97.82)	84.52	(74.99, 91.49)
	3	90.2	(78.59, 96.74)	94.05	(86.66, 98.04)
	4	98.04	(89.55, 99.95)	79.76	(69.58, 87.74)
	5	86.27	(73.74, 94.30)	92.86	(85.10, 97.33)
68 autopsy subjects (43 pos and 25 neg)					
	Reader	sensitivity	95% CI	specificity	95% CI
	1	95.35	(84.19, 99.43)	68	(46.50, 85.05)
	2	90.7	(77.87, 97.41)	76	(54.87, 90.64)
	3	88.37	(74.91, 96.11)	88	(68.78, 97.45)
	4	97.67	(87.70, 99.94)	60	(38.67, 78.88)
	5	88.37	(74.91, 96.11)	84	(63.92, 95.46)
104 subjects with SoT (68 autopsy + 36 biopsy)=51 pos and 53 neg					
	Reader	sensitivity	95% CI	specificity	95% CI
	1	96.08	(86.54, 99.52)	75.47	(61.72, 86.24)
	2	92.16	(81.12, 97.82)	77.36	(63.79, 87.72)
	3	90.2	(78.59, 96.74)	90.57	(79.45, 96.87)
	4	98.04	(89.55, 99.95)	71.7	(57.65, 83.22)
	5	86.27	(73.74, 94.30)	90.57	(79.35, 96.87)
36 subjects with biopsy (8 pos and 28 neg)					
	Reader	sensitivity	95% CI	specificity	95% CI
	1	100	(63.06, 100)	82.14	(63.10, 93.93)
	2	100	(63.06, 100)	78.57	(59.05, 91.70)
	3	100	(63.06, 100)	92.86	(76.50, 99.12)
	4	100	(63.06, 100)	82.14	(63.10, 93.93)
	5	75	(34.91, 96.82)	96.43	(81.65, 99.91)

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

Reference ID: 3334092

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
06/28/2013

JYOTI ZALKIKAR
06/28/2013
I concur with the primary reviewer.

THOMAS E GWISE
06/29/2013

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

NDA Number: 203137 **Applicant: GE healthcare** **Stamp Date: 10/26/2012**

Drug Name: VIZAMYL™ (Flutemetamol F 18 Injection) **NDA/BLA Type: standard**

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_____

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

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

Content Parameter (possible review concerns for 74-day letter)	Yes	No	NA	Comment
Designs utilized are appropriate for the indications requested.	Y			
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.	Y			
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.	Y			
Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.			NA	

File name: 5_Statistics Filing Checklist for a New NDA_BLA110207

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

Lan Huang	11/30/2012
Reviewing Statistician	Date
Supervisor/Team Leader	Date

File name: 5_Statistics Filing Checklist for a New NDA_BLA110207

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

/s/

LAN HUANG
11/30/2012

JYOTI ZALKIKAR
11/30/2012