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 201277
Drug Name:	Gadobutrol (Gadavist)
Indication(s):	MRA of Supra-Aortics and Renals
Applicant:	Bayer Health Care Pharmaceuticals
Date(s):	Submission Date: 6 / 30 / 2015
	PDUFA Date: 4 /30/ 2016
Review Priority:	Standard
Biometrics Division:	DBV
Statistical Reviewer:	A G Mucci, PH. D.
Concurring Reviewers:	Jyoti Zalkikar, PH. D. (Secondary Reviewer)
	Thomas Gwise PH. D. (Deputy Director)
Medical Division:	Division of Medical Imaging
Clinical Team:	Medical Reviewer: Anthony Fotenos MD
Project Manager:	

Table of Contents

1 EXECUTIVE SUMMARY

The Sponsor (Bayer) has submitted NDA201277 for Gadavist MRA for visualization and disease detection in the Supra-Aortic and Renal arteries. The proposed Indication is:

(b) (4)

Gadavist currently has the following indications:

(1):Gadavist is indicated for use with MRI in adult and pediatric patients (including term neonates) to detect and visualize areas with disrupted brain barrier (BBB) and/or abnormal vascularity of the central nervous system.

(2): Gadavist is indicated for use with MRI to assess the presence and extent of malignant breast disease.

The Sponsor has submitted analyses and results from several studies, which include two Phase III studies, in support of this new and extended claim. This review will focus exclusively on the two Phase III trials:

Study14607 (GEMSAV): A Multi-center, open-label study to evaluate the Safety and Efficacy (by blinded reading) of contrast enhanced MRA after a single intravenous injection of 0.1 mmol/kg of gadobutrol in subjects with known or suspected vascular disease of the supra-aortic vessels.

Study14607 (GRAMS): A Multi-center, open-label study to evaluate the Safety and Efficacy (by blinded reading) of contrast enhanced MRA after a single intravenous injection of 0.1 Mmol kg gadobutrol in subjects with known or suspected renal artery disease.

The two studies share identical designs in which qualifying patients undergo both unenhanced MRA (Control = TOF) and Gadavist enhanced MRA (Test) for the acquisition of images which are evaluated at a multiple segment level per patient for detection of significant stenosis (70% or more for supra-aortics ; 50% or more for renals.) For each study there were three independent blinded readers of the TOF and Gadavist images, and also three blinded Standard of Truth (CTA) readers.

The Sponsor presented five primary statistical objectives for determination of Efficacy, the first of which focused on the Visualization of segments, while the remaining four focused on the standard performance characteristics of Sensitivity and Specificity:

First Objective of Visualization: Gadavist was required to provide a greater proportion of visualized segments in comparison to TOF.

Objectives (2) and (3) for Performance Characteristics: Gadavist was required to be non-inferior to TOF for segment level Sensitivity and Specificity.

Objectives (4) and (5): for Performance Characteristics: Gadavist was required to provide segment level Sensitivities and Specificities greater than 50%.

The statistics were evaluated with respect to a "Majority Read" derived from the three blinded reads. (Majority Read applies to all three types of reads: TOF/Gadavist?CTA).

The results were:

Supra-Aortic Study: All Objectives were met, but the success levels for the four objectives dedicated to Performance Characteristics were marginal.

Renal Study: All objectives other than the minimal Sensitivity requirement (> 50%) were met. Again, all other performance characteristic levels of success were marginal

This Review will analyze and evaluate these objectives, as stated, in Section 3.1. However, the Reviewer considered both the segment level analyses and the reduction of the three blinded reads to a single "Majority Read" to have no clear connection to clinical value. Therefore, this Review includes, as its primary analyses, the material in Section 5.1 (Statistical Issues and Collective Evidence.) This material includes:

(a): Analyses that explore the limitations of segment based and Majority Read based analyses and
(b): A limited Subject-Level analysis that could carry more clinical meaning.

The Reviewer concludes, in the light of the results for the Objectives proposed by the Sponsor and the results from the additional and alternative analyses, that the established Efficacy for Gadavist MRA for the Supra-Aortics is marginal, and that Efficacy for Gadavist MRA for the Renals will have to be based on clinical consideration that minimizes the role of the poor Sensitivity performance on the Renals.

2 INTRODUCTION

2.1 Overview

The core of Bayer's NDA 201277 submission, with respect to Efficacy, consists of two prospective Phase III studies:

Study14607 (GEMSAV): A Multi-center, open-label study to evaluate the Safety and Efficacy (by blinded reading) of contrast enhanced MRA after a single intravenous injection of 0.1 mmol/kg of gadobutrol in subjects with known or suspected vascular disease of the supra-aortic vessels.

Study14607 (GRAMS): A Multi-center, open-label study to evaluate the Safety and Efficacy (by blinded reading) of contrast enhanced MRA after a single intravenous injection of 0.1 Mmol kg gadobutrol in subjects with known or suspected renal artery disease.

The intention in both studies is to establish Gadavist MRA diagnostic Efficacy for the indicated regions (Supra-Aortic, Renal) through the satisfaction of five statistical Criteria:

Criterion#1: Superiority of Gadavist MRA to TOF in Visualization

Criterion#2 & Criterion#3: Non-Inferiority of Gadavist MRA to TOF with respect to the Performance Characteristics of Sensitivity and Specificity

Criterion#4 & Criterion#5: A Minimal Performance of Gadavist MRAfor Sensitivity and Specificity(These statistics must exceed Chance = 50%.)

All of these criteria were to be satisfied at a Segment Level rather than at a Subject Level. That is, the relevant statistics involved summing segment level diagnoses across all patients. So, the Criteria take the form:

For Visualization: Gadavist MRA adequately visualized more segments than did TOF

For Non-Inferiority: Gadavist was non-inferior to TOF in classifications of segments as stenosed/non stenosed (-7.5% non-inferiority limit.)

For Minimal Performance: Gadavist MRA correctly classified segments as stenosed/non-stenosed at a rate better than chance.

Note that:
(a): Visualization requires no Truth validation
(b): Non-Inferiority requires that the Gadavist/TOF comparisons be validated through a Standard of Truth , in this case CTA.
(c): The Minimal Performance required Truth but no comparison to TOF.

The statistics in both studies were derived from Image reads from three blinded readers (three different readers per study) who read both the Gadavist images and the TOF images. There were also three blinded Standard of Truth (CTA) readers per study. All reads provided a stenosis level per segment, whenever the segment was visualized , or defaulted to a decision of Non-Assessable (NA) otherwise. Thus, the first level of classification for segments was either a stenosis level or a binary classification of the segment as NA. For purposes of disease classification, there was a next level of classification, defined as follows:

If the segment was visualized, then a binary decision followed – Positive or Negative – with Positive for Supra-Aortics if the stenosis was at least 70%, Positive for Renals if the stenosis was at least 50% .If a segment was Non-Assessable (NA) , then a random decision – Positive or Negative – was assigned. ***This was the Sponsor's Imputation Scheme.***

There was then a third level of classification (binary):
The three individual binary results (per segment) were collapsed into a "Majority" diagnosis

This Majority binary diagnosis was the diagnosis used for evaluation of Sensitivities and Specificities. Moreover, an identical procedure was in place for the CTA Standard of Truth. Thus, for the primary objectives, all performance statistics were evaluated via Majority decisions. It is noted here that the classification of a segment as Non-Assessable was also a Majority decision, and was employed for evaluation of the Visualization Objective.
The principal results for the two studies were:

(A): All five criteria were satisfied for the Supra-Aortic Study.

(B): Four of the five criteria were satisfied for the Renal Study (this Study failed to achieve Criterion#5: better than chance Sensitivity.) The details with respect to these results are presented in Section 3.2. A brief presentation of the results is provided directly below.

Note:
(a): *Parentheticals are Lower Limits of a One-Sided 96% CI.*
(b):An asterisk indicates Failure.

Visualization Percentages:
Supra-Aortics: Test(Gadavist) = 95% ; Control (TOF) = 73% ; Difference = 22%
Renals: Test (Gadavist) = 96% ; Control (TOF) = 78% ; Difference = 18%

Non-Inferiority(Lower Limit must be greater than -7.5%)
Supra Aortics: Test Sensitivity = 60% ; Control Sensitivity = 54% ; Difference = 6% (- 4%)
Test Specificity = 96% ; Control Specificity = 87% ; Difference = 9% (8%)
Renals: Test Sensitivity = 53% ; Control Sensitivity = 47% ; Difference = 6% (-2%)
Test Specificity = 95% ; Control Specificity = 86% ; Difference = 9% (7%)

Minimal Performance Statistics
Supra-Aortics: Gadavist Sensitivity = 62% (55%)
Gadavist Specificity = 98% (97%)
Renals: ***** *Gadavist Sensitivity = 55% (46%) (Failure)*
Gadavist Specificity = 96% (95%)

Problems:
Several serious problems attach to the statistical design and objectives:

(1): Problem with Majority Diagnosis:
The Majority Diagnosis obscures the clinically meaningful results that attach to individual reader performances. Typically, the Medical Imaging Division recommends that there be three independent blinded readers and that at least two of these readers simultaneously achieve the various Efficacy Objectives (usually confined to Sensitivity and Specificity requirements.) The Reviewer has therefore provided the individual statistics by reader for all five objectives. As it turns out, for both studies, the Majority Read statistic succeeded if and only if the majority of the individual reader statistics also succeeded (success for two out of three readers.)

(2): First Problem with Segment Level Statistics
The segment level results do not readily translate into clinically meaningful results at the patient level. The Reviewer investigated a subject Level statistic in order to determine if the Success/Failure profile obtained at the segment level also obtained at this subject level. The only statistic of sufficient interest was Sensitivity. The critical outcome was that, for the renals, the segment level failure to achieve 50% Sensitivity carried over to a Subject Level failure, but, additionally, there was a further failure for Gadavist to achieve non-inferior Sensitivity to TOF.

(3): Second Problem with Segment Level Statistics
The segment level performances, as achieved in these studies, can be closely approximated through Default diagnoses that do not require image reads. This peculiarity is due to an imbalance in disease prevalences across segments, a fact known to the readers. In particular, in the Supra-Aortic Study, only two among 21 segments were likely to be Positive for the SOT, while in the Renal Study, only two among six segments were likely to be Positive for the SOT. This circumstance allows for the implementation of an Imputation Scheme which simply classifies the two High Prevalence segments as Positive, all others as Negative. This Default diagnosis provided Performance Characteristics consistent with the Gadavist results. The Reviewer believes that any chosen segment level success criteria should be sufficiently robust so as to produce statistics superior to those produced by this Default Imputation; this wasn't the case here,

(4): Problem with the Visualization Objective:
The positioning of Superiority in Visualization as a primary objective only makes clinical sense if the Test outperforms Chance on segments that are Non-Assessable for the Control. This result did not obtain in these studies.

(5): Problem with the Standard of Truth (SOT)
CTA is a weak Standard of Truth, attested to in these studies by the "noise" in the three CTA stenosis measurements (large variations from CTA reader to CTA reader.)

The analyses relevant to problems (1) through (4) are provided in Section 5.1, which the Reviewer considers to be the critical section in this review.

The overall conclusions the Reviewer draws from the results regarding the Sponsor's stated objectives and the results obtained through his alternative analyses are:

(1): The Supra-Aortic Study was successful, but the level of success in performance for Gadavist MRA in comparison to TOF or in comparison to minimal Performance Characteristic requirements was marginal.

(2): The Renal Study did not meet all of its objectives – Sensitivity failed to beat the minimal performance level of 50%. Therefore, if Study success is to be accorded here, it will have to be with respect to some "combined" (Sensitivity plus Specificity) performance, such as Se + Sp > 1. However:

(3): The overall clinical meaning that attaches to Segment Level objectives is unclear.

2.2 Data Sources

\\CDSESUB1\evsprod\NDA201277\0115

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

Tables and data sets were adequate for purposes of the review.

3.2 Evaluation of Efficacy

Bayer submitted results from two primary Phase III trials:

Study14607: A Multi-center, Open-label Study to Evaluate the Safety and Efficacy (by blinded reading) of Gadobutrol-Enhanced MRA after a Single Injection 0f 0.1 mmol/kg of Gadobutrol in Subjects with Known or Suspected Vascular Disease of the Supra-Aortic Vessels.

Study91759: A Multi-Center, Open-label Study to Evaluate the Safety and Efficacy (by blinded reading) of Gadobutrol–Enhanced MRA after a Single Injection of 0.1 mmol/kg of Gadobutrol in Subjects with known or Suspected Renal Artery Disease.

The two studies shared an identical Statistical Design and identical Statistical Objectives.
The only differences between the studies were:

The Supra-Aortic Study involved assignment of a binary result – presence/absence of significant stenosis – to each of 21 segments, by means of stenosis levels, where the cut-off for significant stenosis was set at 70%.

The Renal Study involved assignment of a binary result – presence/absence of significant stenosis – to each of 6 segments, by means of stenosis levels, where the cut-off for significant stenosis was set at 50%.

Note: The presentation below focuses exclusively on the evaluation of Success/Failure of the studies with respect to the Sponsor's five objectives, as stated. However, the Reviewer believes that the framework within which these objectives are placed (for example: Majority Reads; Segment Level statistics) do not allow for inferences to clear clinical meaning. Consequently, the Reviewer will provide additional and alternative analyses afterwards, in Section 5.1. The Reviewer believes these latter analyses to be more meaningful.

Common Study Design and Study Objectives

Standard of Reference Read: CTA was the Standard of Reference (SOR). There were three CTA blinded readers (three different readers for each Study) who independently evaluated each segment and recorded a stenosis level (when possible). If a read registered a stenosis of at least X%, (X = 70% for Supra-Aortics ; X = 50% for Renals) the segment was scored as Positive; otherwise as Negative. For primary statistical analyses and objectives, the Majority Diagnosis (Positive/Negative) derived from the three binary decisions then became the SOR Diagnosis. This requires that at least 2 of the 3 CTA readers render a diagnosis and also agree on the diagnosis. There was also an "overall" stenosis level derived from the individual registered stenoses, whose value was determined by a complicate algorithm, but this stenosis level played no role in the Primary Analyses. (It appears to coincide with the median read in many cases.)

Blinded Reads of Test /Control:There were three blinded readers who read both the Test Images (Gadavist-Enhanced MRA) and the Control Images (Unenhanced MRA), with, again, three different readers for each Study. Each reader registered a stenosis level for each segment, whenever possible, and then collapsed this continuous variable into a binary decision in the same manner as did each of the CTA readers. Once more, for Test and Control, respectively, a "Majority Diagnosis" , derived from the individual binary decisions was used for primary statistical analyses and objectives. This Majority Diagnosis often required imputations, especially for the TOF reads. The Imputation Scheme will be defined later on in this review.

Statistical Criteria for Success for Efficacy:
There were five Statistical Criteria that had to be met for Study Success. Each criterion involved statistics based on the aforementioned binary "Majority" Reads derived from the individual reads. The five criteria, relative to this Majority Read, will be presented below. However, we mention here that the Medical Imaging Division typically requires that all Success Criteria, whatever form they take, be achieved by at least two out of three blinded readers rather than by some collapsed read. Therefore, although the Majority Read outcomes were accepted by the Division as determinants of success for the two studies under review, the outcomes reader-by-reader will receive equal attention in this Review in order to determine if the Division's standard "two out of three" criterion was met for each of the five statistics. (The Division's acceptance of the Majority outcome was most likely predicated on the circumstance that there were five conditions to be met, rather than the more common set of two conditions involving only Sensitivity and Specificity.)

The five criteria are:

(1): Superiority of Test over Control for Visualization: For each reader, each segment was registered as Assessable or Non-Assessable. (We will often substitute the term "Visualized" for Assessable.) An Assessable segment was a segment for which a stenosis level was registered. Segments too opaque for such a measurement were denoted as Non-Assessable (NA).
The Test was deemed ***Superior*** to the Control if the proportion of Test Assessables statistically exceeded the proportion of Control Assessables. It is noted here that the determination that a segment was Assessable was also a Majority Decision outcome, but in the obvious sense that at least two of the three readers provided a stenosis level for the segment.

With A(Test), A(Control) denoting the "true" proportions of Majority Read Assessables for Test/Control respectively, this first Success Criterion is:

Visualization Success Criterion: A(Test)>A(Control) .

In terms of point estimates derived from the data, the operational Criterion for Success is as follows: Let:
$\overline{A}$ (Test) = Point Estimate for A(Test); $\overline{A}$ (Control) = Point Estimate for A(Control)
D = A(Test) - A(Control) ; $\overline{D}$ = Point Estimate for D
Sigma = Point Estimate of Standard Error for D
Then:

"Operational" Visualization Success Criterion: $\overline{D}$ – 1.64*Sigma > 0

Important Note#1: The calculation for the Success Criterion utilized a 95% One-Sided Confidence Interval (CI) rather than a Two-Sided 95% CI. Equivalently, the Criterion utilizes a One-Sided 95% CI rather than a One-Sided 97.5% CI. The Agency accepted this more liberal Criterion, along with the substitution of the Majority Read for the individual reads, presumably because there were five criteria that had to be met.

Next, were two Performance Characteristic Success Criteria:

(2): Non-Inferiority of Test with respect to Control for segment-level Sensitivity.

(3): Non-Inferiority of the Test with respect to the Control for segment-level Specificity.

It is with respect to these two criteria that an Imputation Scheme was put in place, since there were significant levels of NAs (primarily for the Control). All these NAs were randomly assigned Positive or Negative status; that is, each NA segment was assigned Positive or Negative status with probability 1/ 2. Note that this Imputation Scheme was in place for each Reader. Thus, if the three reads were, respectively, Positive/Negative/NA, and if the NA was randomly assigned as Positive, then the Majority result was Positive.

For both of these objectives the Non-Inferiority limit was set at -7.5% (.075.) So, with Se(Test)/ Se(Control) and Sp(Test)/ Sp(Control) denoting the True Sensitivities /Specificities for Test/Control, respectively, we have the Success Criteria:

(*): Se(Test)> Se(Control) - .075 ; Sp(Test) > Sp(Control) - .075

As with the Visualization Criterion, and with respect to the Majority Rule, we need notation:

$\overline{Se}$ (Test) = Point Estimate for Se(Test); $\overline{Se}$ (Control) = Point Estimate for Se(Control)
$\overline{Sp}$ (Test) = Point Estimate for Sp(Test); $\overline{Sp}$ (Control) = Point Estimate for Sp(Control)
De = Se(Test) – Se(Control) ; Dp = Sp(Test) – Sp(Control)
$\overline{De}$ = Point estimate of De ; $\overline{Dp}$ = Point Estimate of Dp
Sigma(e) = Point Estimate of Standard Error for De
Sigma(p) = Point Estimate for Standard Error for Dp

Note a further refinement: These point estimates are dependent on the results of the randomization; that is, each implementation of randomization across the set of images and segments produces its own estimates. So let's define, for the k-th run-through of the data (across readers and segments), the point estimates : $\overline{De}$ (k) , Sigma(k, e) and $\overline{Dp}$ (k), Sigma(k, p).
We then define:

LL(k,e) = $\overline{De}$ (k)- 1.64* Sigma(k, e)
LL(k,p) = $\overline{Dp}$ (k) - 1.64* Sigma(k, p)
Then let:

LL(e) = Mean of the LL(k,e) across 1000 bootstrappings (k-th producing LL(k,e))
LL(p) = Mean of the LL(k,p) across 1000 bootstrappings (k-th producing LL(k,p))

Then, (*) above is established if:

(**): Operational Performance Comparisons Criteria: LL(e > -.075 and LL(p) > - .075

Important Note#2: The Reviewer believes:

(a):That the "half-credit" Imputation Scheme has dubious value, largely because the disease prevalence differs radically from segment to segment, and readers have knowledge of this fact, and, if forced to make a binary decision, would not flip a coin, but would decide in light of this knowledge and thereby produce more representative diagnoses.

(b): "Across Segment" Level, as contrasted with Patient Level, Sensitivity and Specificity, have no clear clinical meaning, and also little statistical meaning.

Consequently, although the Sponsor's results will be presented and assessed with respect to the agreed-upon segment-level and chance imputation scenario, the Reviewer will also provide alternative Subject Level statistical analyses that are likely to be better indicators of clinical performance for Sensitivity and Specificity.

We now turn to the final two Success Criteria:

(4): Segment level minimum Sensitivity must statistically exceed 50%.

(5): Segment level minimum Specificity of Test must statistically exceed 50%.

As with Criteria (2) and (3), the Operational conditions that have to be met are:

Lower Limit(LL) of one-Sided 95% CI determined by the corresponding point estimates must exceed 50% (.50)

However, in contrast to the framework employed for evaluations under Criteria (2) and (3), only assessable segments were included in the evaluation of these last two objectives, and therefore no bootstrapping is required.

Comments
(1):The Reviewer's assumption regarding the restriction of to Assessables is that: if the point estimates are just above .50 for visualized segments, then the Imputation Scheme will reduce these estimates slightly, to the disadvantage of the Test. Thus, the restriction to Assessables can bias the statistic upwards.

(2):The Reviewer has a problem with the concept of a Majority Read Assessable. Since each reader has his own set of Assessables, the "Majority" Read proper to conditions (4) and (5) is complicated. (What does it mean for a segment to be assessable when, for instance, two readers find it assessable and the third reader doesn't, and the two readers do not agree on the binary diagnosis?) The Reviewer will therefore exercise the option of evaluating each reader's outcomes in turn for criteria (4) and (5), although the Sponsor's "Majority" outcome will still be reported.

Before turning to the statistical evaluations for Criteria (1) through (5), we state:

Overall Results: All statistical objectives ((1) through (5)) were met in both Studies.

Caveat: The success levels for Criteria (2) through (5) were, at best, marginal.

Presentation and Critique of the Results for the Primary Objectives

Patient Dispositions

Study 14067 (Supra-Aortic) enrolled 487 subjects, 457 of whom qualified for the Primary Efficacy evaluations (Full Analysis Set = FAS). Subjects were included in the FAS if all three Image types were collected and read: Unenhanced MRA , Enhanced MRA, CTA.

Study 91759 (Renal) enrolled 317 subjects, 292 of whom qualified for the Primary Efficacy evaluations (Full Analysis Set = FAS). As with the Supra-Aortic Study, subjects were included in the FAS if all three Image types were collected and read: Unenhanced MRA , Enhanced MRA, CTA.

Table(1):Demographics

	STUDY14607 (Supra-Aortic) (457 Subjects)	**Study 91759 (Renal) (292 Subjects)**
Sex		
Male	**294 (64%)**	**158 (54%)**
Female	**163 (36%)**	**134 (46%)**
Age		
< 65	**160 (35%)**	**189 (65%)**
≥ 65	**297 (65%)**	**103 (35%)**
Race		
White	**366 (80%)**	**199 (68%)**
Black	**13 (3%)**	**21 (7%)**
Asian	**71 (16%)**	**64 (22%)**
Not Reported	**7 (1%)**	**7 (2%)**

Presentation of Results for the Five Statistical Objectives

Objective (1): Superiority of Test over Control for Assessable Segments

Table(2): Percentages of Assessable Segments by Study

STUDY14607 Supra-Aortic Subjects =457 ; Segments = 9597				
	Test	**Control**	**Difference**	**95% CI**
Majority Read	**95%**	**73%**	**22%**	**(20% , 24%)**
RDR#1	**88%**	**24%**	**64%**	**(61% , 67%)**
RDR#2	**95%**	**75%**	**20%**	**(18% , 22%)**
RDR#3	**97%**	**82%**	**15%**	**(13% , 17%)**
On-Site	**97%**	**79%**	**19%**	**(17% , 21%)**
STUDY 91759 Renals Subjects =292 ; Segments = 1734 to 1752				
	Test	**Control**	**Difference**	**95% CI**
Majority Read	**96%**	**78%**	**18%**	**(15% , 21%)**
RDR#1	**98%**	**82%**	**16%**	**(13% , 19%)**
RDR#2	**96%**	**72%**	**24%**	**(21% , 27%)**
RDR#3	**96%**	**78%**	**17%**	**(14% , 20%)**
On-Site	**94%**	**69%**	**25%**	**(22% , 28%)**

Comments on the Table:

(1):The method used to generate the 95% CIs was **McNemar** adjusted for Clustering.) Clustering accounts for the fact that the Sponsor's CIs are slightly wider than CIs calculated under the assumption that assessability occurs independently, segment by segment.

(2): A rough description regarding the Table above (one exception) is:

Supra-Aortics:
About 16 in 20 segments were visualized in Control Images
About 19 in 20 segments were visualized in the Test Images
Renals:
About 15 in 20 segments were visualized in Control Images
About 19 in 20 segments were visualized in the Test Images

Exception: The shaded cells in the Supra-Aortic table highlight that Reader#1 called most segments Non-Assessable for the Control.

Conclusion: ***The Assessability Objective was met by all Readers and by the Majority Read.***

Objectives (2)/(3): Non-Inferiority of Test to Control

Preliminary Comments
Both studies present with segment level prevalences that partition into two classes: High Prevalence and Low Prevalence, as determined by the SOR of CTA. More will be made of this later on, but an overview of subject level stenosed segment dispositions is presented directly below as a backdrop for the presentation of the results for Objectives (2) and (3), and also as a rational for alternative analyses.

For Supra-Aortics
There are 2 High Prevalence segments, with individual Prevalences $\approx .12$
There are 19 Low Prevalence segments, with individual Prevalences $\approx .006$

The two High Prevalence segments are the Right Internal Carotid and the Left Internal Carotid. The Prevalence P = .12 represents the likelihood that any given high prevalence segment is stenosed (Positive). For the 19 Low Prevalence segments, the likelihood that any one of them is stenosed is .006.

For Renals:
There are 2 High Prevalence segments, with individual Prevalences $\approx .167$
There are 4 Low Prevalence segments, with combined Prevalence $\approx .023$

The two High Prevalence segments are the Left Proximal Third and the Right Proximal Third. The Prevalence P = .167 represents the likelihood that any given high prevalence segment is stenosed (Positive).

Profile of the Distribution of Significant Stenoses
The table below provides a profile of locations for stenosed segments; the intention is to give a broad picture of where stenosess are found when they are found. The information provided is Subject Level. The Table is read as follows:

High & Low means that there were stenosed segments both in the High Prevalence class and in the Low Prevalence class , and # Subjects indicates the number of subjects with this Profile

No High, Low means there were no stenosed segments in the High Prevalence class , but there were stenosed segments in the Low Prevalence class. Again, # Subjects indicates the number of subjects with this Profile. (etc)

Table(3) Profile of Locations for Stenosed Segments

SUPRA-AORTICS 457 Subjects)			RENALS (274 Subjects)	
Stenosis Profile	**# Subjects**		**Stenosis Profile**	**# Subjects**
High & Low	**11 (2%)**		**High & Low**	**15 (5%)**
High, No Low	**94 (21%)**		**High, No Low**	**57 (21%)**
No High, Low	**36 (8%)**		**No High, Low**	**8 (3%)**
No High , No Low	**316 (69%)**		**No High , No Low**	**194 (71%)**

Note that the dominant category is::
No High, No Low (69% /71% respectively ; about 7 in every 10 subjects , both studies.)

Observations per Supra-Aortics: If the likelihood of being positive is independent across these segments, then the likelihood that any given subject has at least one stenosed high prevalence segment is 2*.12*(.88) + (.12)*(.12) ≈ .23 (about one in every four patients.) . This is exactly what Table (3) presents.
Again, if we assume independence of positivity from segment to segment, then the likelihood that any given subject has at least one stenosed low prevalence segment is approximately 1-(Negative)**19 = 1 – (.994)**19 ≈.11. Table (3) shows that the actual probability is ≈.10. Thus, the evidence suggests that Positivity occurs independently across these segments also.

Observations per Renals: If the likelihood of stenosis (.167) was independent across high prevalence renals, then the probability that any renal subject had at least one stenosed renal segment would be ≈ 2*(.167)*(.833)+ (.167)**2 = .31. Table(3) reveals this probability to be .26, which suggests a weak dependence.
As for the likelihood that a subject has at least one stenosis among the four low prevalence segments, this would be 1 – (.977)**4 ≈.09. Table(3) reveals the probability to be .08, which suggests independence for Positivity from segment to segment for the Low Prevalence class.

Additional Information: (Not all of the following dertivable from the Table):

For Supra-Aortics
(a): 23% of Subjects (about 1 in every 4) had stenosed segments in High Prevalence class
Expected number of stenosed segments in High Prevalence class, given at least 1 is: 1.0
(b): 10% of Subjects (about 1 in every 10) had stenosed segments in Low Prevalence class
Expected number of stenosed segments in Low Prevalence class, given at least 1 is: 1.1

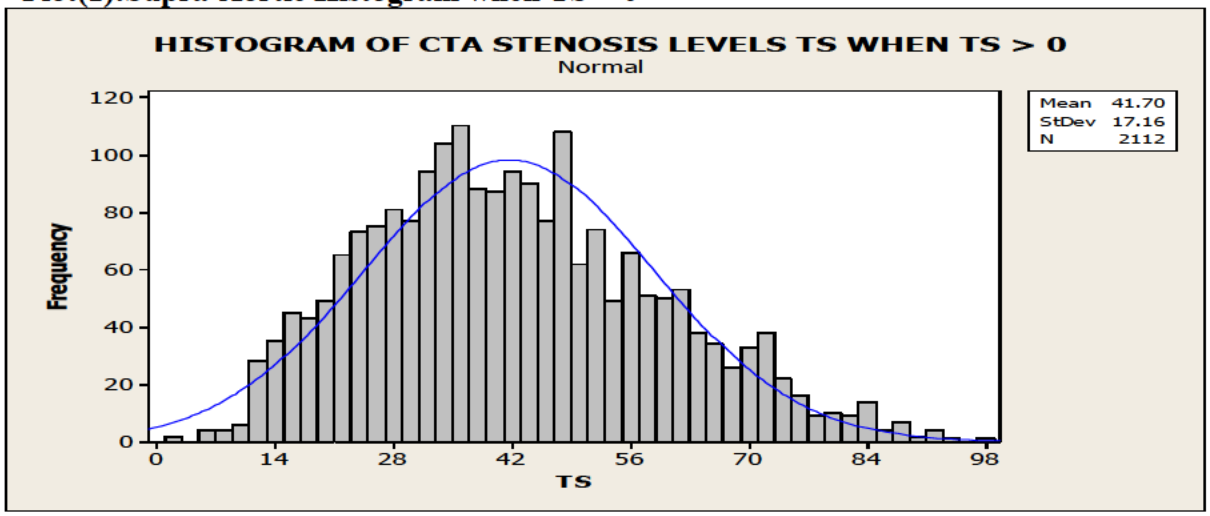
For Renals
(a): 26% of Subjects (about 1 in every 4) had stenosed segments in High Prevalence class
Expected number of stenosed segments in High Prevalence class, given at least 1 is: 1.3
(b): 8% of Subjects (about 1 in every 12) had stenosed segments in Low Prevalence class
Expected number of stenosed segments in Low Prevalence class, given at least 1 is: 1.1

Overall: For the 3 in 10 subjects with at least one significant stenosis, about 1 in 12 had A significant stenosis in both the High and Low Prevalence class; about 2 in 3 had the significant stenosis confined to the High Prevalence class; about 1 in 4 had significant stenosis confined to the Low Prevalence class.

Finally, before the presentation of the Performance results, the following CTA stenosis level profile is provided:

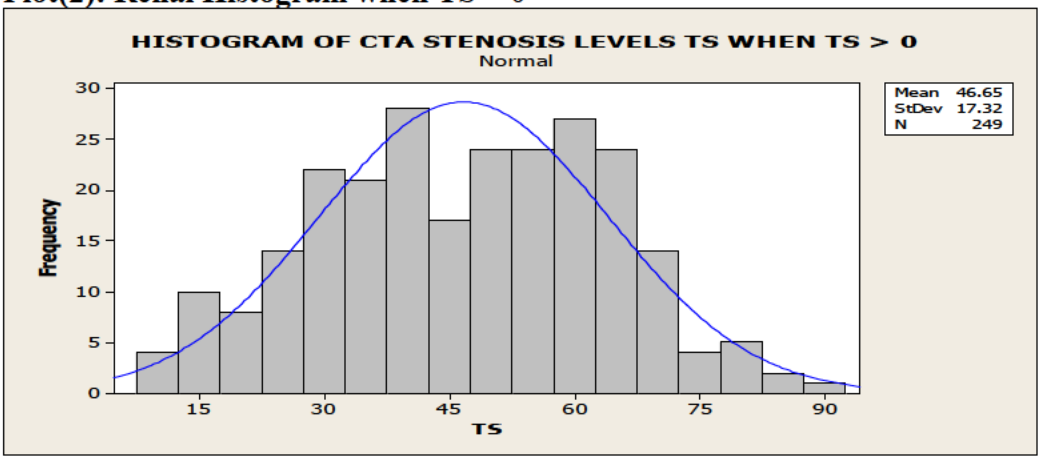
The Supra-Aortic Distribution for CTA Stenosis Levels is a Mixture:
TS = 0 in 78% of Segments
TS is approximately Norma(42, 17) in remaining 22% of Segments

Plot(1):Supra-Aortic Histogram when TS > 0



Renal Distribution for CTA Stenosis Levels is a Mixture:
TS = 0 in 85% of Segments
TS is approximately Norma(47, 17) in remaining 15% of Segments

Plot(2): Renal Histogram when TS > 0



Non-Inferiority Results (Objectives (2) and (3)

Supra-Aortics:

Table(4): Sensaitivity/Specificity for Supra-Aortic Study

STUDY14607 Supra-Aortic Sensitivity Stenosed Subjects =141 ; Stenosed Segments = 158 Average Stenosed Segments per Stenosed Subject = 1.1				
	Test	**Control**	**Difference**	**95% 1-Sided LL**
Majority Read	**60%**	**54%**	**6%**	**-3.6% > -7.5%**
RDR#1	**60%**	**54%**	**5%**	**-3.8% > -7.5%**
RDR#2	**60%**	**54%**	**6%**	**-2.5% > -7.5%**
RDR#3	**58%**	**55%**	**3%**	**-4.4% > -7.5%**
STUDY14607 Supra-Aortic Specificity Subjects =457 ; Non-Stenosed Segments = 9321				
	Test	**Control**	**Difference**	**95% 1-Sided LL**
Majority Read	**96%**	**87%**	**9%**	**8%**
RDR#1	**92%**	**62%**	**30%**	**29%**
RDR#2	**95%**	**85%**	**10%**	**9%**
RDR#3	**97%**	**89%**	**8%**	**7%**

Comments on Table:

The statistics, in particular, the CIs, were generated by assigning values of Stenosed or Non-Stenosed randomly to each non-assessable segment, and repeating this process 1000 times (bootstrapping) over the entire data set of segments. The Sponsor calculated the 95% 1-Sided Lower Limit (LL) by adjusting for "clustering", but it is not clear that this was necessary, given the bootstrapping procedure.

Conclusion: ***The Supra-Aortic Study met its Performance Characteristic Objectives both with respect to the Majority Read and with respect to the Agency's preferred criterion that at least 2 of the 3 Readers succeed on both Sensitivity and Specificity. In fact, all Readers were successful.***

Renals

Table(5):Sensitivity / Specificity for Renal Study

STUDY 91759 Renal Sensitivity Stenosed Subjects =93 ; Stenosed Segments = 133 Average Stenosed Segments per Stenosed Subject = 1.4				
	Test	**Control**	**Difference**	**95% 1-Sided LL**
Majority Read	**53%**	**47%**	**7%**	**-2.2% > -7.5%**
RDR#1	**52%**	**51%**	**1%**	**-9%**
RDR#2	**54%**	**39%**	**14%**	**6% > -7.5%**
RDR#3	**53%**	**50%**	**3%**	**-6.6% > -7.5%**
On-Site	**69%**	**50%**	**20%**	**11% > -7.5%**
STUDY91759 Renal Specificity Subjects =292 ; Non-Stenosed Segments = 1605				
	Test	**Control**	**Difference**	**95% 1-Sided LL**
Majority Read	**95%**	**86%**	**9%**	**7%**
RDR#1	**94%**	**83%**	**11%**	**9%**
RDR#2	**95%**	**85%**	**10%**	**8%**
RDR#3	**94%**	**81%**	**13%**	**11%**
On-Site	**97%**	**84%**	**13%**	**11%**

Comments on Table

The statistics were generated by assigning values stenosed/non-stenosed randomly to each non-assessable segment, and repeating this process 1000 times (bootstrapping) over the entire data set of segments. The Sponsor calculated the 95% 1-Sided Lower Limit (LL) by adjusting for "clustering", but, again, it is not clear that this was necessary, given the bootstrapping procedure.

Conclusion: ***The Renal Study met its Performance Characteristic Objectives with respect to the Pre-Specified Majority Read and also with respect to the Agency's preferred criterion that at least 2 of the 3 Readers succeed on both Sensitivity and Specificity.***

Note: ***RDR#1 did not succeed for Sensitivity.***

Objectives (4) and (5): Performance for Test versus 50% Threshold

Table(6): Performance Characteristics for Test versus Threshold of 50%

SENSITIVITY						
SUPRA-AORTIC		**95% LL**		**RENAL**		**95% LL**
Majority Read	**62%**	**55%**		**Majority Read**	**55%**	**46%**
Reader#1	**60%**	**54%**		**Reader#1**	**52%**	**44%**
Reader#2	**60%**	**53%**		**Reader#2**	**54%**	**46%**
Reader#3	**59%**	**52%**		**Reader#3**	**53%**	**45%**
SPECIFICITY						
SUPRA-AORTIC		**95% LL**		**RENAL**		**95% LL**
Majority Read	**98%**	**97.7%**		**Majority Read**	**96%**	**95%**
Reader#1	**98%**	**97%**		**Reader#1**	**95%**	**94%**
Reader#2	**97%**	**96.9%**		**Reader#2**	**96%**	**95%**
Reader#3	**98%**	**97.7%**		**Reader#3**	**96%**	**95%**

Comments on Table

The number of segments that enter into the statistics vary from Reader to Reader since Test non-assessables are removed from the calculations. (The Reviewer is uncertain as to how the Majority Read non-assessables were determined since this read required some kind of "pooling" of the individual reader non-assessables. However, the details on this pooling are irrelevant since the table shows that all readers individually either succeeded or failed.)

Conclusions

(1): The Performance Objectives were met in the Supra-Aortic Study

(2): The Sensitivity Objective was not met in the Renal Study.

Overall Conclusions regarding the five Statistical Objectives:

(A): All objectives were met except for the minimal Sensitivity Objective in the Renals

However:

(B):Sensitivities, in both the Supra-Aortics and the Renals, were only marginally better than Chance, both with Imputation for Non-Assessables for Objective (3), and on the Assessables for Objective (5). Thus, the concentration below under Alternative analyses will generally be on Sensitivity.

Overall Conclusions

The Sponsor defined Study Success as the achievement of five statistical objectives, involving segment level statistics which can be loosely stated as follows:

(1): The Test (Gadavist) registered more segments as Visualized than did the Control (TOF)

(2)&(3):The Test was Non-Inferior to the Control for Segment Level Sensitivity/Specificity

(4)&(5): The Test had Segment Level Sensitivity and Specificity > 50% for the Supra-Aortics, but failed to achieve a Sensitivity > 50% for the Renals.

The Reviewer concludes that a marginal level of success was achieved for the Supra-Aortics, but Success for the Renals is questionable.

3.3 EVALUATION OF SAFETY

There were no significant Safety issues.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender/Race/Age

Preliminary Comments:

(1): Presentations of all results will be confined to Sensitivity at the segment level since Specificity here, or, in general in these studies, presents no problems.

(2): No definitive conclusions will be drawn from the results in this section. However, these results are sometimes suggestive of performance asymmetries – for example, male versus female sensitivities.

GENDER

Table(7): Supra-Aortic Sensitivities by Gender

SUPRA-AORTIC SEGMENT LEVEL SENSITIVITIES BY GENDER (Males =64% ; Male Positives = 64% Females = 46% ; Female Positives = 46%) (V = Visualized Positive Segments ; A = All Positive Segments)								
		RDR#1			RDR#2		RDR#3	
		M	F		M	F	M	F
Visualized	Test	91%	95%		98%	100%	97%	100%
	Control	21%	19%		79%	82%	80%	81%
Sensitivity on V	Test	65%	54%		65%	51%	65%	49%
	Control	81%	55%		62%	45%	61%	48%
Sensitivity on A	Test	64%	54%		65%	51%	65%	49%
	Control	56%	51%		60%	46%	50%	48%

Table(8): Renal Sensitivities by Gender

RENAL SEGMENT LEVEL SENSITIVITIES BY GENDER (Males =54% ; Male Positives = 65% Females = 46% ; Female Positives = 35%) (V = Visualized Positive Segments ; A = All Positive Segments)								
		RDR#1			RDR#2		RDR#3	
		M	F		M	F	M	F
Visualized	Test	99%	95%		95%	73%	95%	80%
	Control	83%	66%		68%	41%	80%	44%
Sensitivity on V	Test	52%	51%		57%	53%	56%	45%
	Control	59%	44%		37%	24%	57%	28%
Sensitivity on A	Test	52%	51%		57%	52%	56%	46%
	Control	58%	46%		41%	40%	56%	40%

Comments:

These tables suggest that Sensitivities are higher for males than for females, for both Gadavist and TOF.

RACE

Note: W = White ; A = Asian (No other groups with non-negligible representation)

Table(9): Supra-Aortic Sensitivities by Race

SUPRA-AORTIC SEGMENT LEVEL SENSITIVITIES BY RACE (White =80% ; White Positives = 80% Asian= 15% ; Asian Positives = 14%) (V = Visualized Positive Segments ; A = All Positive Segments)								
		RDR#1			RDR#2		RDR#3	
		W	A		W	A	W	A
Visualized	Test	96%	91%		99%	95%	99%	100%
	Control	20%	23%		81%	82%	82%	82%
Sensitivity on V	Test	60%	70%		57%	81%	56%	77%
	Control	69%	100%		51%	78%	55%	61%
Sensitivity on A	Test	60%	68%		57%	81%	56%	77%
	Control	54%	62%		50%	73%	54%	58%

Table(10): Renal Sensitivities by Race

RENAL SEGMENT LEVEL SENSITIVITIES BY RACE (White =69% ; Male Positives = 76% Asian = 22% ; Asian Positives = 17%) (V = Visualized Positive Segments ; A = All Positive Segments)								
		RDR#1			RDR#2		RDR#3	
		W	A		W	A	W	A
Visualized	Test	98%	95%		84%	95%	88%	95%
	Control	73%	85%		58%	55%	66%	70%
Sensitivity on V	Test	54%	58%		57%	58%	56%	53%
	Control	52%	59%		33%	27%	53%	50%
Sensitivity on A	Test	54%	58%		56%	58%	55%	53%
	Control	51%	59%		40%	37%	52%	50%

Comments: The is a suggestion here that Sensitivities are higher on Asians for the Supra-Aortics, but this trend is not continued in the Renals**.**

AGE

Table(11): Supra-Aortic Sensitivities by Age

SUPRA-AORTIC SEGMENT LEVEL SENSITIVITIES BY AGE ((≥ 65)= 65% ; (≥ 65) Positives = 62% (<65) = 35% ; (<65)Positive = 38%) (V = Visualized Positive Segments ; A = All Positive Segments)								
		RDR#1			RDR#2		RDR#3	
		≥ 65	<65		≥ 65	<65	≥ 65	<65
Visualized	Test	95%	88%		100%	97%	100%	95%
	Control	17%	25%		84%	75%	86%	71%
Sensitivity on V	Test	66%	52%		63%	54%	64%	50%
	Control	77%	67%		58%	50%	60%	50%
Sensitivity on A	Test	65%	52%		63%	54%	64%	50%
	Control	55%	54%		57%	50%	59%	50%

Table(12): Renal Sensitivities by Age

RENAL SEGMENT LEVEL SENSITIVITIES BY AGE ((≥ 65)= 35% ; (≥ 65) Positives = 67% (<65) = 65% ; (<65)Positive = 33%) (V = Visualized Positive Segments ; A = All Positive Segments)								
		RDR#1			RDR#2		RDR#3	
		≥ 65	<65		≥ 65	<65	≥ 65	<65
Visualized	Test	97%	97%		87%	87%	91%	87%
	Control	81%	69%		62%	54%	71%	62%
Sensitivity on V	Test	55%	45%		53%	62%	51%	56%
	Control	48%	70%		35%	29%	45%	63%
Sensitivity on A	Test	55%	45%		53%	60%	50%	55%
	Control	48%	64%		41%	39%	38%	58%

Comments
As with RACE, a suggested diagnostic improvement in Sensitivity for the >=65 yrs group in the Supra-Aortics is not sustained in the Renals.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

This section presents analyses that the Reviewer considers relevant to assessments of the clinical value of the data that were not captured within the Sponsor's framing of the five primary objectives. The Reviewer's principal concerns regarding the formulations of the Primary Objectives is:

The Segment Level statistics do not translate into any apparent clinical meaning. In particular, the fact that significant stenosis is confined to only two segments among 21 segments in the Supra-Aortic, and again, to only two segments among six in the Renal case, can introduces a bias towards high Specificities through Default readings of low prevalence segments as Negative.

Consequently:
There is good reason to consider performance statistics that arise from alternative Segment Level imputation schemes. The surprising results regarding these statistics, in turn, provide a rationale for a "Patient Level" analysis confined to the high prevalence segments in order to obtain a perspective more closely tied to clinical value.

Moreover:
Even within the framework of Segment Level statistics, the value attached to the first Primary Endpoint of segment Visualization was not explored by the Sponsor. The Reviewer contends that the fact that Gadavist images are better visualized than TOF Images has value only if Gadavist Images provide better than chance diagnoses where TOF defaults to such diagnoses (namely, on its Non-Assessables.)

These concerns and observations inform the three alternative analyses provided below.

Alternative Analyses

The Reviewer provides three alternative analyses:

The first analysis addresses segment level sensitivity in the following sense: Since TOF is provided with a chance Imputation on its non-assessables, we'd expect Gadavist to outperform chance here. This result is not achieved for the Renals; barely for the Suprs-Aortics.

The second analysis focuses on a Segment Level Alternative Imputation Scheme in order to highlight the problems inherent in segment level statistics . The results show that "good" performance at this level can be obtained without recourse to actual image reads.

The third analysis (a natural investigation, given the results from the first two analyses) is a Subject Level analysis, which the Reviewer believes to be more closely attached to clinical value than do the segment level statistics.

ANALYSIS#1: Sensitivities of Test when Control is Non-Assessable

We return to the segment level scenario in order to address question apparently not Considered by the Sponsor:What value attaches to Gadavist as a consequence of its having fewer non-assessables than TOF? Here is the relevant table:

Table(13): Test Sensitivities on Control Non-Assessables

SUPRA-AORTICS			
	# Control NAs	**# Test Positives**	**Sensitivity**
Majority	**64**	**40**	**.63 (LL= .53)**
RDR1	**125**	**73**	**.58 (LL=.51)**
RDR2	**31**	**19**	**.61 LL=(.53)**
RDR3	**31**	**19**	**.61 (LL=.53)**
RENALS			
	# Control NAs	**# Test Positives**	**Sensitivity**
Majority	**37**	**20**	**.54 (LL=.41)**
RDR1	**27**	**11**	**.41 (LL=.26)**
RDR2	**48**	**23**	**.48 (LL=.37)**
RDR3	**38**	**19**	**.50 (LL=.35)**

Comments on Table

The Superiority of Test to Control for Segment visualization would seem to be irrelevant if the Test turned in poorer than Chance Performance on Control Non-Assessables. This is the case as presented in the Table above, certainly for the Renals, and almost for the Supra-Aortics.

ANALYSIS#2: Alternative Imputation Statistics

The Sponsor employed the following Imputation to Non-Assessable segments:

Imputation Scheme Type A: Randomly impute Positive or Negative to the segment
The Reviewer believes this Imputation Scheme to be misleading since, as will be shown, it provides for both Test and Control performances that at the segment level can be fairly well matched, or surpassed, by an alternative Default Imputation Scheme that does not even require evaluation of Images. In effect, the alternative scheme simply exploits the fact that segments are known to partition significantly into a High Prevalence set and a Low Prevalence set.

The following alternative Imputation scheme was investigated by the Reviewer:

Default Imputation Scheme Type B:
Impute Positive to all High Prevalence Segments
Impute Negative to all Low Prevalence Segments

Note: ***Type B involves no Image reads*.** Thus, Type B can be implemented by simply recording Negatives for Low Prevalence segments, Positives for High Prevalence segments.

Recall:
For Supra-Aortics:
There are 2 High Prevalence segments, with individual Prevalences ≈ .12
There are 19 Low Prevalence segments, with individual Prevalences ≈ .006
For Renals:
There are 2 High Prevalence segments, with combined Prevalence ≈ .167
There are 4 Low Prevalence segments, with combined Prevalence ≈ .023

The Table below presents the segment level statistics derived under schemes (A) and (B) respectively.

Table(14): Performances under Half-Credit and Alternative Imputation Schemes

SUPRA-AORTICS					
SENSITIVITY			SPECIFICITY		
IMPUTATION (A)		IMPUTATION (B)	IMPUTATION (A)		IMPUTATION (B)
TEST	CONTROL		TEST	CONTROL	
.60	.54	.68	.96	.87	.91
RENALS					
SENSITIVITY			SPECIFICITY		
IMPUTATION (A)		IMPUTATION (B)	IMPUTATION (A)		IMPUTATION (B)
TEST	CONTROL		TEST	CONTROL	
.53	.47	.78	.95	.86	.70

Comments on Table

Note that Imputation (B) works well for the Supra-Aortics, and it presents a problem there:

For the Supra-Aortics, Overall Performance under the Type B Imputation is consistent with Overall Performance under the Type A Imputation, but it doesn't require any readers. It is even possible that this Imputation yields statistics superior to the Test statistics.

For the Renals, there is a significant drop in Specificity under scheme (B), but a large gain in Sensitivity. When Sensitivity and Specificity are averaged, the results are:

Test Average = (1/ 2) (53 + 95) = 74%
Control Average = (1/ 2) (47 + 86) = 67%
Alternative Average = (1/ 2) (78 + 70) = 74%

If an "Overall" Performance, such as is captured by averaging Sensitivity and Specificity, is used as an indicator of Efficacy, then Imputation Type (B) outperforms the Control and matches the Test. This should cause concern about the viability of a segment level scenario for Efficacy assessments.

The Reviewer infers from this that a more accurate picture of Efficacy for the Test, or for Test versus Control, requires that attention focus on the Subject Level, and that this level is itself restricted to statistics derived from the High Prevalence segments. This will be the focus below.

ANALYSIS#3: "Subject Level" Statistics

The analyses here are restricted to the two ”High Prevalence” segments., but these will default to “Subject Level” analyses in that the binary decision that will enter into the evaluations takes the following form:

For Truth Status of the Subjects:
Subjects are classified as Positive for Truth if at least 1 of the 2 segments is Positive for Truth.
Subjects are classified as Negative for Truth if both segments are Negative for Truth.

Thus, Truth Positives consist of CTA binary classifications of:

Left = Positive ; Right = Positive
Left = Positive ; Right = Negative
Left = Negative ; Right = Positive

Read Concordance with Truth
A Match with Truth for a Read is a subject-level match that requires that the Read agrees with Truth on both segments.

The first two tables below present the subject level statistics relative to the definitions provided above for the Supra-Aortics/Renals, respectively, and with the focus on Test versus Control performance relative to the Success Criteria (2) and (3) (Test versus Control Sens/Spec.)

Note:
LL = Lower Limit of 95% One-Sided CI for the Difference D
NA = % of Non-Assessable Subjects where a subject is classified as NA if at least one of the two High Prevalence segments is Non-Assessable.
E/U = Majority Gadavist/Majority TOF Reads, respectively
EJ/ UJ = Reader J Gadavist / Reader J TOF Reads, respectively
D = Gadavist minus TOF statistic

The table provides point estimates along with the Lower Limit (LL) of the One-Sided 95% CIs.

Table(15): Subject Level Performance (Test versus Control) on Supra-Aortics

SUBJECT LEVEL SENSITIVITY HIGH PREVSALENCE SEGMENTS

SUPRA-AORTIC									
	SE			SE			SE		SE
E (NA=5%)	.63		E1 (NA=7%)	.63		E2 (NA=2%)	.63	E3 (NA=1%)	.62
U (NA=33%)	.53		U1 (NA=79%)	.56		U2 (NA=10%)	.53	U3 (NA=6%)	.54
D (LL)	.10 (0.01)		D1 (LL)	.07 (-.01)		D2 (LL)	.10 (0.02)	D3 (LL)	.08 (0.0)
SUBJECT LEVEL SPECIFICITY HIGH PREVALENCE SEGMENTS SUPRA-AORTIC									
	SP			SP			SP		SP
E (NA=3%)	.82		E1 (NA=7%)	.82		E2 (NA=2%)	.82	E3 (NA=1%)	.83
U (NA=21%)	.73		U1 (NA=70%)	.75		U2 (NA=10%)	.73	U3 (NA=10%)	.75
D (LL)	.09 (.05)		D1 (LL)	.07 (.02)		D2 (LL)	.09 (.04)	D3 (LL)	.08 (.03)

Comments per the Supra-Aortic Table

For Positive Patients:

(1a): Gadavist reads agree with Truth on both segments in about 3 of every 5 patients who have at least one stenosed segment

(1b): TOF reads agree with Truth on both segments in slightly more than half the patients who have at least one stenosed segment

For Negative Patients

(2a): Gadavist reads agree with Truth on both segments in about 4 out of every 5 patients who have both segments non-stenosed

(2b):TOF reads agree with Truth on both segments in about 3 out at every 4 patients who have both segments non-stenosed

For Success Criteria (2) and (3):

All statistics are consistent with Success for Non-Inferiority when the Non-Inferiority limit is set at -7.5% . This is consistent with the Sponsor's segment level Success Criteria.

Table(16): Subject Level Performance (Test versus Control) on Renals

SUBJECT LEVEL SENSITIVITY HIGH PREVSALENCE SEGMENTS

RENALS									
	SE			SE			SE		SE
E (NA=14%)	.43		E1 (NA=5%)	.45		E2 (NA=16%)	.42	E3 (NA=15%)	.45
U (NA=40%)	.47		U1 (NA=28%)	.47		U2 (NA=50%)	.45	U3 (NA=37%)	.46
D (LL)	-.04 (-.16)		D1 (LL)	-.02 (-.15)		D2 (LL)	-.03 (-.15)	D3 (LL)	-.01 (-.14)

SUBJECT LEVEL SPECIFICITY HIGH PREVALENCE SEGMENTS RENALS									
	SP			SP			SP		SP
E (NA=3%)	.87		E1 (NA=1%)	.88		E2 (NA=4%)	.87	E3 (NA=4%)	.87
U (NA=13%)	.82		U1 (NA=11%)	.81		U2 (NA=21%)	.77	U3 (NA=14%)	.78
D (LL)	.05 (0.0)		D1 (LL)	.07 (.02)		D2 (LL)	.10 (.05)	D3 (LL)	.09 (.04)

Comments per the Renal Table

(1): Gadavist reads agree with Truth on both segments in less than half the patients who have at least one stenosed segment; likewise for TOF.

(2): Gadavist reads agree with Truth on both segments in slightly more than 4 out of every 5 Patients who have both segments non-stenosed; likewise for TOF.

Importantly: For Success Criteria (2) and (3):
For each Reader (including Majority Read): The Sensitivity statistics are not consistent with Success for Non-Inferiority when the Non-Inferiority limit is set at -7.5% .

We now next to the Minimal Sensitivity/Specificity Success Criterion (5) as applied to these "Subject Level" statistics.

Subject Level (High Prevalence) Minimal Sensitivity Criterion Results

The table below presents the Subject-Level Sensitivities for "Visualized Subjects" for both the Supra-Aortics and the Renals. This Table is based, for each Reader, on that Reader's Subject Level "Visualized" profile ; that is, a patient enters into the Reader's statistics only if both segments were visualized by the Reader. Moreover, only Sensitivity is presented since the figures for Specificity are high and do not present problems for the Success Criterion. As before, a subject Level Read matches Truth if it matches Truth on both High Prevalence segments

Note: LL = Lower Limit of 1-Sided 95% CI

Table(17): Subject Level Sensitivities for Both Studies

Supra-Aortics	**E**	**E1**	**E2**	**E3**
	(N = 100)	**(N = 98)**	**(N = 103)**	**(N = 104)**
	.64	**.62**	**.64**	**.63**
	(LL= .56)	**(LL= .54)**	**(LL = .56)**	**(LL = .55)**
Renals	**E**	**E1**	**E2**	**E3**
	(N = 64)	**(N = 71)**	**(N = 62)**	**(N = 63)**
	.50	**.46**	**.47**	**.49**
	(LL = .40)	**(LL = .36)**	**(LL = .37)**	**LL = (.39)**

Comments

(1): The minimum success level of .50 is achieved (barely) for all Supra-Aortic Reads

(2): The minimum success level of .50 is not achieved for any of the Renal Reads (Point Estimates also fail)

(3): Thus, for Sensitivity evaluated at a "Subject Level", for the Renals, Gadavist achieves Neither non-inferiority to TOF nor the minimal success performance of 50% .

Overall Conclusions derived from the Alternative Analyses

(A): Segment Level statistics with chance imputations on Non-Assessables do not provide a profile of clinically relevant diagnostics for Gadavist PET; moreover, these statistics can be approximated closely by statistics derived without readings of the images.

(B): Subject Level Sensitivities for the Renal Study do not meet minimal success requirement for Gadavist, either with respect to non-inferiority to TOF or with respect to better than chance performance.

5.2 Conclusions and Recommendations

The Supra-Aortic Study met its Objectives and the recommendation here is Approval. The Renal Study met all Objectives other than the Minimal Performance for Sensitivity. The Reviewer defers to the Medical Team for the final word on this Study; there could be some overall picture consistent with Approval here, but it doesn't reside in the results for the Objectives taken individually. Moreover, and with consideration given to the Supra-Aortic Study also:

The results ,with respect to the Sponsor's Criteria (2) through (5), and with respect to the Reviewer's additional analyses, suggest that the Gadavist Performance appears at best to be only marginally better than the TOF Performance.

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

/s/

ANTHONY G MUCCI
03/28/2016

JYOTI ZALKIKAR
03/28/2016
I concur with the primary statistical reviewer.

THOMAS E GWISE
03/28/2016