

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:
NDA 22-090

STATISTICAL REVIEW(S)



U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Pharmacoeconomics and Statistical Science
Office of Biostatistics

STATISTICAL REVIEW

CLINICAL STUDIES

NDA: NDA 22090

Drug Name: Eovist (Gd-EOB-DTPA)

Indication: MRI of the Liver

Applicant: Bayer Health Care

Date(s): Stamp: July 2 2007 PDUFA: May 2 2008

Review Priority: Standard

Biometrics Division: Division of Biometrics V (DBV)

Statistical Reviewer: A G Mucci, Ph.D.

Concurring Reviewers: Jyoti Zalkikar, Ph.D., Statistical Team Leader, DBV
Aloka Chakravarty, Ph.D., Director, DBV

Medical Division: Medical Imaging and Hematology Products

Clinical Team: Cynthia Welsh M.D. (Medical Reviewer)
Liberio Marzella, M.D. (Clinical Team Leader)

Project Managers: Tiffany Brown/James Moore

Keywords: Superiority, Non-Inferiority, Imaging

Table of Contents

Table Of Contents	2
List Of Tables	3
1. EXECUTIVE SUMMARY	4
1.1 Conclusions and Recommendations	4
1.2 Overview of the Clinical Studies	4
1.3. Statistical Issues and Findings	6
2. INTRODUCTION	10
2.1 Overview	10
2.2 Data Sources	12
3. STATISTICAL EVALUATION	13
3.1 Evaluation of Efficacy	13
3.1.1 Evaluation of the Efficacy for the Detection Studies.....	13
3.1.2 Evaluation of the Efficacy for the Characterization Studies.....	22
3.2 Evaluation of Safety.....	27
4. Findings in Special/Subgroup Populations	28
4.1 Gender, Race and Age	28
4.2 Other Special/Subgroup Populations	29
5. SUMMARY AND CONCLUSIONS	30
5.1 Statistical Issues and Collective Evidence	30
5.2. Conclusions and Recommendations	34
Appendix: Mathematics	35

List of Tables

Table 1: Demographics	16
Table 2: Patient Disposition.....	16
Table 3: Primary Results for the Detection Studies.....	17
Table 4: Percentages of Detection Study Patients with at Least One False Positive.....	17
Table 5: Detection Studies Sensitivities/Specificities at the Liver Lobe Level.....	18
Table 6: Test versus Comparator (CT) Success Levels in Lesion Detection	18
Table 7: Within Patient versus Lesion Weighted Detection Rates	19
Table 8: Lesion Detection Statistics by Classification Category	20
Table 9: Lesion Detection Statistics by Classification Category.....	20
Table 10: Lesion Detection Statistics by Cluster Status	21
Table 11: Lesion Detection Statistics by Cluster Status	22
Table 12: Demographics	24
Table 13: Patient Disposition.....	24
Table 14: Test versus Comparator Success Levels in Characterization	25
Table 15: Test versus CT Success Levels in Characterization	26
Table 16: Lesion Classification Statistics by Classification Category	27
Table 17: Lesion Detection Rates by Age	28
Table 18: Lesion Characterization Rates by Age	28
Table 19: Lesion Detection Rates by Gender	28
Table 20: Lesion Characterization Rates by Gender	29
Table 21: Detection Results for a Random Reader in the Combined Studies	31
Table 22: Random Reader Characterization/Classification Results	33

1. EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

The Bayer Health Care submission (NDA22090) provided results for Eovist from four pivotal Phase III studies to support an indication for enhanced MRI of the liver. The four pivotal trials consisted of two identically designed liver lesion detection studies and two identically designed liver lesion characterization studies. For the primary analyses in the detection studies, lesion detection rates for the Test modality of combined pre and post Eovist contrast MRI were compared to the detection rates for pre contrast MRI. The objective was to show that the pre+post contrast MRI reads provided higher rates of liver lesion detections than did the pre contrast MRI reads. The lesion characterization studies compared pre+post contrast MRI image reads to pre contrast MRI image reads for lesion characterization in the primary analyses. The objective was to show that the pre+post contrast MRI reads provided higher rates of correct liver lesion characterizations than did the pre contrast MRI reads. Both sets of studies included secondary endpoint comparisons of pre+post contrast MRI to contrast CT, which is the current standard of care in liver imaging. Both sets of studies met their success criteria:

For each of the detection studies, two of three blinded readers achieved superior lesion detection rates for pre+post contrast MRI over pre contrast MRI.

For each of the characterization studies, two of the three blinded readers achieved superior lesion characterization rates for pre+post contrast MRI over pre contrast MRI.

The success in the detection studies was marginal in the sense that although the statistical significance was achieved, clinical meaningfulness of the result may be questionable. This success was achieved on patients with several rather than single lesions. The success in the characterization studies was somewhat more substantial and may be clinically meaningful, and was achieved largely in the characterization of benign lesions.

There is adequate evidence from the submitted studies to support approval of Eovist for enhanced MRI of the liver.

1.2 Overview of the Clinical Studies

NDA22090 provided four Phase III trials in support of its proposed indication. These consisted of two identically designed lesion detection trials and two identically designed lesion characterization trials.

Lesion Detection Trials: StudyA03779 (USA) and Study A00518 (EU)

Lesion Detection Trials: Essential Design Elements

Common Design and Objectives: These were open label, cross-over, diagnostic studies of patients suspected of liver lesions who were scheduled to undergo liver surgery.

Patients were imaged under three imaging modalities:

Test = Pre + Post Eovist Contrast MRI ;
Primary Comparator = Pre-Contrast MRI ;
Secondary comparator = Contrast CT.

The various image reads were directed toward detection of lesions which were identified during the scheduled surgery and confirmed by histopathology or IOUS (Intra-Operative Ultra-Sound). Three blinded readers independently read all the images. The primary endpoint was the lesion detection rate. The statistical criterion for success for the primary comparison of Test to Comparator was that two of the three blinded readers achieve superiority for Test over Comparator in detection rates. The principal secondary comparison was the evaluation of Test lesion detection rates to Contrast CT lesion detection rates. (Contrast CT is the current standard of care.)

Lesion Characterization Trials: StudyA05742 (EU) and Study A01908 (USA)

Lesion Characterization Trials: Essential Design Elements

Common Design and Objectives: These trials were open-label, cross-over diagnostic imaging studies of patients with liver lesions verified and characterized by an appropriate SOR. Patients were once again imaged under three imaging modalities:

Test = Pre + Post Eovist Contrast MRI ;
Primary Comparator = Pre-Contrast MRI ;
Secondary Comparator = Contrast CT.

The competing reads were directed toward characterization of the SOR verified lesions, which were clearly marked on the images by the On-Site investigators. Three blinded readers independently read all the images. The primary endpoint was the lesion characterization rate. (Characterization was with respect to at least 12 categories, some of which were benign, such as adenomas, hemangiomas, and cysts, and some of which were malignant, such as hepatocellular carcinomas and metastases.) The statistical criterion for success for the primary comparison of Test to Comparator was that two of the three blinded readers achieve superiority for Test in lesion characterization rates. The principal secondary comparison was the evaluation of Test lesion characterization rates to Contrast CT lesion characterization rates. (Contrast CT is the current standard of care.)

Overall Conclusions: *The Test met its primary objectives for superiority over Comparator for both lesion detection and lesion characterization. The detection superiority was achieved only on patients with multiple lesions. The characterization superiority was achieved primarily on benign lesions.*

1.3. Statistical Issues and Findings

Lesion Detection Trials:

The combined detection trials enrolled 341 subjects; 260 of whom qualified for the per-protocol analyses. There were 56 patients excluded from the primary analyses because no SOR was available; the remaining 25 patients lost to the primary analyses had significant protocol deviations. The per-protocol patients had a total of 617 Lesions, 72% of which were malignant.

Statistical Endpoints and Hypotheses: The primary statistical endpoint was the Test versus Comparator difference in per- patient numbers of lesions detected. In detail:

For patient K: If the SOR determined L_K true lesions, from among which the Test matched T_K , and the Comparator C_K , then the patient's primary statistic was

$$D_K = (T_K - C_K) / L_K \quad K = 1, 2, \dots, N \quad (N = \text{Number of patients})$$

The Sponsor collected all these differences D_K and transformed them into rankings R_K in order to apply the Wilcoxon Signed Rank Test to the sequence R_1, \dots, R_N . This statistic was then used to test the hypotheses:

H_0 : The Sensitivity of lesion detection of combined pre+post image reads is the same as the Sensitivity of lesion detection of pre image reads.

H_1 : The Sensitivity of lesion detection of combined pre+post image reads differs from the Sensitivity of lesion detection of pre image reads.

Reviewer's Comments:

1> The Null Hypothesis was to be tested at the two-sided .05 level. The protocol (Section 7.8.1.5.1 – p30 of 38) states that the hypotheses were to be tested separately for each reader, with p values generated under the Null Hypothesis, and rejection of the Null obtaining when $p < .05$. However, the reviewer could find no statement regarding an overall success criterion, for example, that two of the three readers achieve p values $< .05$. The Agency typically requires that overall success be equated with two of the three readers achieving statistical significance, unless the Sponsor proposes an acceptable alternative.

2> The statistical reviewer remains unconvinced that the Sponsor's proposed Wilcoxon Signed Rank Test, applied at the "Within-Patient" level, is appropriate. A more intuitive and direct approach would use the mean values for the per patient differences in detection rates, D_K , described above. This alternative approach requires that the lower limit of the 95% two-sided CI for the mean per patient differences exceed zero for two of the three readers. As the tables in the sections below will reveal, the Wilcoxon Signed Rank Test

and the reviewer's CI Test succeed for the same readers, so there is good reason to believe that the Sponsor's approach is formally equivalent to the more natural approach.

3> Both the Sponsor's Wilcoxon statistic and the reviewer's alternative "Within Patient" lesion detection statistic are inappropriate for several secondary subset analyses that compare Test to Comparator lesion detection rates by lesion classification, lesion size, etc. In all these lesion base cases, the appropriate statistic is:

Detection Rate = # Lesions Detected in the Subset/ # Lesions in the Subset

Parenthetically, as will be shown in the results section below, the Sponsor would have done better in choosing this lesion based statistic for all comparisons.

Statistical Findings:

In both trials the success criterion that two of three readers achieve superiority in lesion detection rates for Test over Comparator was met. In both trials, however, the success level was marginal, as attested by the average reader performances given below.

A detailed investigation of the detection rates in both trials revealed that the marginal superiority of Test over Comparator could be traced to improved detection rates in patients with multiple lesions. (Single lesions were typically large, malignant, and equally well detected by all image modalities.) The average performances stratified by isolated versus multiple lesions are listed below. Note that statistical significance is achieved only in the multiple lesion case, and that, despite this significance, the rates there are not especially good.

Trial A03779: Stratified Detection Rates and CI's for Differences:

All Lesions: Test = 79%; Comparator = 74%; CI for Difference = (0%, 10%)

Isolated Lesions: Test = 93% ; Comparator = 89% ; CI for Difference = (-1% , 7%)

Clustered Lesions: Test = 64% ; Comparator = 55% ; CI for Differences = (4% ,9%)

Trial A00518: Stratified Detection Rates and CI's for Differences:

All Lesions: Test 78%; Compaator = 74%; CI for Difference = (0%, 8%)

Isolated Lesions: Test = 88% ; Comparator = 88% ; CI for Difference = (-3% , 3%)

Clustered Lesions: Test = 67% ; Comparator = 62% ; CI for Differences = (3% ,9%)

The Sponsor's Wilcoxon statistic in the Detection studies was, for purposes of hypothesis testing, indistinguishable from a "Within-Patient" lesion detection statistic. This "Within-Patient" statistic focuses on detection rates achieved per-patient rather than at the lesion level (sometimes described as "Lesion-Weighted").

Whenever the Sponsor supplemented Wilcoxon analyses with point estimates for detection rates, these estimates were always computed at the lesion level. The derivation of necessary mathematical results for distinguishing between these two statistics – "Within-Patient" and "Lesion-Weighted" - tend to be complex, and are relegated to the

Appendix. The overall conclusion drawn from these results is that the “Within-Patient” statistic will be larger than the “Lesion-Weighted” statistic in a fairly predictable manner whenever the detection rates per lesion decrease with increasing numbers of lesions per patient, which is the case in these studies.

These comments should not be taken to imply anything, in general, about the statistics on differences, such as Within-Patient” Test minus Comparator “statistics versus “Lesion-Weighted” Test minus Comparator statistics. This issue is not addressed in this review. Consequently, there is as yet no general argument for preferring one of the statistics over the other, in detection studies.

Lesion Characterization Trials:

A total of 475 patients were enrolled in the combined trials, with 359 of these qualifying for the per-protocol analyses. A total of 67 patients were excluded from the primary analyses because no SOR was available; the remaining 49 patients had significant protocol violations.

Primary Statistical Endpoints and Hypotheses: The primary statistical endpoint was the proportion of SOR detected and characterized lesions which were correctly characterized by the individual blinded readers.

H₀: The proportion of correctly characterized lesions under pre+post contrast MRI reads is the same as the proportion of lesions with correct characterization for pre contrast MRI reads.

H₁: The proportion of correctly characterized lesions under pre+post contrast MRI reads differs from the proportion of lesions with correct characterization for pre contrast MRI reads.

Since some subjects might have several lesions, a McNemar Test adjusted for clustering would be used to evaluate these hypotheses.

Reviewer’s Comments:

1>: The Sponsor used the McNemar Test, applied at the lesion level for the test statistic. There was adjustment for clustering, since there were several lesions per patient. The reviewer agrees that this is an appropriate statistic for these studies. The test for statistical significance using this statistic is equivalent to satisfaction of the condition that the lower limit of the 95% two-sided CI for the Test versus Comparator mean difference in lesion classification rates exceed zero.

2>: In these characterization studies the discordance rates were again moderately high and could be interpreted as evidence that the competing modalities are sensitive to different lesion types.

Statistical Findings

In both trials the success criterion that two of three readers achieve superiority in lesion characterization rates for Test over Comparator was met. The success levels here were somewhat better than the levels achieved in the detection studies:

There was a concern in these trials with the possibility that mischaracterizations could be serious, in that they could also be misclassifications. For instance, the characterization of a cyst as metastases is not only a mischaracterization, but also a misclassification in that it classifies a benign lesion as a malignant lesion. Consequently, secondary analyses concentrated on comparisons of correct lesion classification levels: benign as benign and malignant as malignant. The statistics showed that there was a slight improvement in correct lesion classification rates for Test over Comparator.

Trial 01908: Average Correct Classification Rates

All Lesions: Test = 68%; Comparator = 58%; CI for Difference = (6%, 14%)

Benign Lesions: Test = 79% ; Comparator = 71%; CI of Difference = (3% , 8%)

Malignant Lesions: Test = 91% ; Comparator = 91% ; CI = (-5% , 5%)

Trial 05742: Average Correct Classification Rates

All Lesions: Test = 67%; Comparator = 54%; CI for Difference = (9%, 17%)

Benign Lesions: Test = 86% ; Comparator = 77%; CI of Difference = (4% , 14%)

Malignant Lesions: Test = 85% ; Comparator = 77% ; CI = (3% , 13%)

**APPEARS THIS WAY
ON ORIGINAL**

2. INTRODUCTION

2.1 Overview

NDA22090 provided four Phase III trials in support of its proposed indication. These consisted of two identically designed lesion detection trials and two identically designed lesion characterization trials:

Lesion Detection Trials: StudyA03779 and Study A00518

Lesion Characterization Trials: StudyA05742 and Study A01908

Lesion Detection Trials: Essential Design Elements and Statistical Results

Common Design and Objectives: These were open label, cross-over, diagnostic studies of patients suspected of liver lesions who were scheduled to undergo liver surgery. Patients were imaged under three imaging modalities:

Test = Pre + Post Eovist Contrast MRI ;
Primary Comparator = Pre-Contrast MRI ;
Secondary comparator = Contrast CT.

The various image reads were directed toward detection of lesions which were identified during the scheduled surgery and confirmed by histopathology or IOUS (Intra-Operative Ultra-Sound). Three blinded readers independently read all the images. The primary endpoint was the lesion detection rate. The statistical criterion for success for the primary comparison of Test to Comparator was that two of the three blinded readers achieve superiority for Test in detection rates. The principal secondary comparison was the evaluation of Test lesion detection rates to Contrast CT lesion detection rates. (Contrast CT is the current standard of care.)

Principal Results: The combined detection trials enrolled 341 subjects; 260 of whom qualified for the per-protocol analyses. There were 56 patients excluded from the primary analyses because no SOR was available; the remaining 25 patients lost to the primary analyses had significant protocol deviations. These protocol patients had a total of 617 Lesions, 72% of which were malignant. In both trials the success criterion that two of three readers achieve superiority in lesion detection rates for Test over Comparator was met. In both trials, however, the success level was marginal; this minimal success performance can be appreciated by looking at average reader performances:

The average reader lesion detection rates in the US trial were 79% for Test (pre+post contrast MRI) and 74% for Comparator (pre contrast MRI), resulting in a point estimate difference of 5% and a 95% CI lower limit for this difference just above the success limit of 0%.

The average reader lesion detection rates in the EU trial were 78% for Test (pre+post contrast MRI) and 74% for Comparator (pre contrast MRI), resulting in a point estimate difference of 4% and a 95% CI lower limit for this difference once more just above the success limit of 0%.

A detailed investigation of the detection rates in both trials revealed that the marginal superiority of Test over Comparator could be traced to improved detection rates in patients with multiple lesions. (Single lesions were typically large, malignant, and equally well detected by all image modalities.) The average performances stratified by isolated versus multiple lesions are listed below. Note that statistical significance is achieved only in the multiple lesion case, and that, despite this significance, the rates there are not especially good.

Trial A03779: Stratified Detection Rates and CI's for Differences:

Isolated Lesions: Test = 93% ; Comparator = 89% ; CI for Difference = (-1% , 7%)
Clustered Lesions: Test = 64% ; Comparator = 55% ; CI for Differences = (4% ,9%)

Trial A00518: Stratified Detection Rates and CI's for Differences:

Isolated Lesions: Test = 88% ; Comparator = 88% ; CI for Difference = (-3% , 3%)
Clustered Lesions: Test = 67% ; Comparator = 62% ; CI for Differences = (3% ,9%)

Lesion Characterization Trials: Essential Design Elements and Statistical Results

Common Design and Objectives: These trials were open-label, cross-over diagnostic imaging studies of patients with liver lesions verified and characterized by an appropriate SOR. Patients were once again imaged under three imaging modalities:

Test = Pre + Post Eovist Contrast MRI ;
Primary Comparator = Pre-Contrast MRI ;
Secondary Comparator = Contrast CT.

The competing reads were directed toward characterization of the SOR verified lesions, which were clearly marked on the images by the On-Site investigators. Three blinded readers independently read all the images. The primary endpoint was the lesion characterization rate. (Characterization was with respect to at least 12 categories, some of which were benign, such as adenomas, hemangiomas, and cysts, and some of which were malignant, such as hepatocellular carcinomas and metastases.) The statistical criterion for success for the primary comparison of Test to Comparator was that two of the three blinded readers achieve superiority for Test in lesion characterization rates. The principal secondary comparison was the evaluation of Test lesion characterization rates to Contrast CT lesion characterization rates. (Contrast CT is the current standard of care.)

Principal Results: A total of 475 patients were enrolled in the combined trials, with 359 of these qualifying for the per-protocol analyses. A total of 67 patients were excluded from the primary analyses because no SOR was available; the remaining 49 patients had significant protocol violations. In both trials the success criterion that two of three readers achieve superiority in lesion characterization rates for Test over Comparator was met. The success levels here were somewhat better than the levels achieved in the detection studies:

The average reader lesion characterization rates in the US trial were 68% for Test (pre+post contrast MRI) and 58% for Comparator (pre contrast MRI), resulting in a point estimate difference of 10% and a 95% CI lower limit for this difference of 6%.

The average reader lesion characterization rates in the EU trial were 67% for Test (pre+post contrast MRI) and 54% for Comparator (pre contrast MRI), resulting in a point estimate difference of 13% and a 95% CI lower limit for this difference of 9%.

There was a concern in these trials with the possibility that mischaracterizations could be serious, in that they could also be misclassifications.. For instance, the characterization of a cyst as a metastases is not only a mischaracterization, but also a misclassification in that it classifies a benign lesion as a malignant lesion.. Consequently, secondary analyses concentrated on comparisons of correct lesion classification levels: benign as benign and malignant as malignant. The statistics showed that there was a slight improvement in correct lesion classification rates for Test over Comparator:

US Trial: Average Correct Classification Rates

Benign Lesions: Test = 79% ; Comparator = 71%; CI of Difference = (3% , 8%)

Malignant Lesions: Test = 91% ; Comparator = 91% ; CI = (-5% , 5%)

EU Trial: Average Correct Classification Rates

Benign Lesions: Test = 86% ; Comparator = 77%; CI of Difference = (4% , 14%)

Malignant Lesions: Test = 85% ; Comparator = 77% ; CI = (3% , 13%)

Overall Conclusions: *The Test met its primary objectives for superiority over Comparator for both lesion detection and lesion characterization. The detection superiority was achieved only on patients with multiple lesions. The characterization superiority was achieved primarily on benign lesions. (See Section 5.1: Statistical Issues and Collective Evidence.)*

2.2 Data Sources

The data source for this review is: \\CDSESUB1\EVXPROD\NDA022090\022090.ENX

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Evaluation of the Efficacy for the Detection Studies

There were two identical Phase III Lesion Detection studies. These will be treated side-by-side below.

Overview of the Design for the Detection Studies

StudyA00518: This study enrolled 169 patients, with 131 patients included in the preferred efficacy analysis. These 131 patients provided 302 SOR verified lesions for the primary efficacy analyses.

StudyA03779: This study enrolled 172 patients, with 131 patients included in the preferred efficacy analysis. These 131 patients provided 316 SOR verified lesions for the primary efficacy analyses.

These two studies had the following identical Title and Design:

Study Title: A multi-center, open-label study with corresponding blinded read to evaluate SH L 569 B (Eovist) in adult patients with known/suspected focal liver lesions who are scheduled for liver surgery.

Primary Study Objective: To demonstrate Superiority of pre+post Eovist contrast MRI to pre contrast MRI in detection of liver lesions.

Principal Secondary Efficacy Objectives:

(a):To compare pre + post Eovist MRI to contrast enhanced spiral CT for liver lesion detection. (Contrast CT is the current Standard of Care.)

(b): To compare pre + post Eovist MRI to both pre contrast MRI and to contrast enhanced spiral CT for Sensitivity and Specificity in liver lesion detection at a liver lobe level. Also, to compare Test versus Comparator for false positive rates in lesion detections. (The primary comparisons only allow for true positives and false negatives. Determinations of presence/absence of lesions in liver lobes allow for Specificity calculations. These specificity statistics are reinforced with direct statistics on false positive rates, and, taken together, are intended to ensure that the objective of improved lesion detection is not achieved at the expense of increased false positives.

(c): To compare pre + post MRI to pre MRI and to CT for lesion classification (malignant; benign; non-assessable). This comparison intersects with secondary comparisons performed in the Characterization studies.

Primary Inclusion Criterion: Patients are known or suspected for liver lesions and are scheduled for liver surgery.

Criteria for Inclusion in the Preferred Efficacy Analysis: Patients will have undergone unenhanced and Eovist enhanced MRI, CT imaging, and liver surgery within a six week period.

Primary Statistical Efficacy Variables: Per- patient sensitivity of pre + post MRI and pre MRI in lesion detection. Sensitivity is defined as the proportion of MRI detected lesions that match the SOR validated lesions.

Primary Reads for the Statistical Analyses: Three blinded readers were to independently read all images from all image modalities. The readers were to list, locate, classify (benign, malignant, non-assessable), and characterize (by type) detected lesions

Standards of Reference (SOR): There were three possible sources for the SOR for Presence or absence of lesions: Histopathology , Intra-Operative Ultrasound (IOUS) , or three month Follow-Up. Histopathology of resected liver obtained during the mandated surgery was primary as the SOR, with IOUS substituting in liver areas where resection was not indicated. In the rare cases where surgery was not performed, the default SOR was a three month follow-Up evaluation as the SOR. This evaluation could include all imaging results exclusive of the Test modality.

Blinded Read Protocol: The three blinded readers independently read all MRI and CT images. The image types pertinent to the primary and principal secondary analyses were read in two blocks, each consisting of three sessions conducted sequentially as follows:

Session#1: Only CT scans.

Session#2: Both pre and post MRI images, presented separately and randomly.

Session#3: Combined pre + post MRI images.

Two to four weeks separated each session, in order to minimize recall. Presumably, the two distinct blocks of three separate sessions each included approximately half the patients, and the three sessions for block#1 preceded the three sessions for block#2.

Procedure for Lesion Identification and Tracking: The blinded readers numbered and located all lesions they individually detected on their respective CRF's , according to a pre-specified segmental mapping procedure. This same procedure was employed for SOR readings. Subsequent to completion of these several lesion mappings, an independent radiologist compared all these CRF's and established a common numbering for the various lesions so that direct comparisons of reader detections to SOR detections could be carried out for the statistical analyses. Only those lesions identified by the SOR were used for the primary sensitivity analyses; lesions recorded by blinded readers but not verified by the SOR constituted false positives.

Statistical Endpoints and Hypotheses: The primary statistical endpoint was the Test versus Comparator difference in per- patient numbers of lesions detected. In detail:

For patient K: If the SOR determined L_K true lesions, from among which the Test matched T_K , and the Comparator C_K , then the patient's primary statistic was

$$D_K = (T_K - C_K) / L_K \quad K = 1, 2, \dots, N \quad (N = \text{Number of patients})$$

The Sponsor collected all these differences D_K and transformed them into rankings R_K in order to apply the Wilcoxon Signed Rank Test to the sequence R_1, \dots, R_N . This statistic was then used to test the hypotheses:

H₀ : The Sensitivity of lesion detection of combined pre+post image reads is the same as the Sensitivity of lesion detection of pre image reads.

H₁: The Sensitivity of lesion detection of combined pre+post image reads differs from the Sensitivity of lesion detection of pre image reads.

Comment#1: The Null Hypothesis was to be tested at the two-sided .05 level. The protocol (Section 7.8.1.5.1 – p30 of 38) states that the hypotheses were to be tested separately for each reader, with p values generated under the Null Hypothesis, and rejection of the Null obtaining when $p < .05$. However, the reviewer could find no statement regarding an overall success criterion, for example, that two of the three readers achieve p values $< .05$. The Agency typically requires that overall success be equated with two of the three readers achieving statistical significance, unless the Sponsor proposes an acceptable alternative.

Comment#2: The statistical reviewer remains unconvinced that the Sponsor's proposed Wilcoxon Signed Rank Test, applied at the "Within-Patient" level, is appropriate. A more intuitive and direct approach would use the mean values for the per patient differences in detection rates, D_K , described above. This alternative approach requires that the lower limit of the 95% two-sided CI for the mean per patient differences exceed zero for two of the three readers. As the tables in the sections below will reveal, the Wilcoxon Signed Rank Test and the reviewer's CI Test succeed for the same readers, so there is good reason to believe that the Sponsor's approach is formally equivalent to the more natural approach.

Comment#3: Both the Sponsor's Wilcoxon statistic and the reviewer's alternative "Within Patient" lesion detection statistic are inappropriate for several secondary subset analyses that compare Test to Comparator lesion detection rates by lesion classification, lesion size, etc. In all these lesion base cases, the appropriate statistic is:

$$\text{Detection Rate} = \# \text{ Lesions Detected in the Subset} / \# \text{ Lesions in the Subset}$$

Parenthetically, as will be shown in the results section below, the Sponsor would have done better in choosing this lesion based statistic for all comparisons.

Overview of the Efficacy Results for the Detection Studies

Table 1: Demographics

	Study A03779	Study A00518
# Dosed Patients/ #Per Protocol Patients	169/ 131	162 ; 131
Mean Age (+- Sigma)	59 +-13 yrs	58 +- 12 yrs
Per-Protocol Patients		
Age Range	19 - 84	21 - 83
Gender	94 Male (56%) ; 75 (44%)Female	97 Male (60%); 65 Female (40%)
Race:		
Caucasian	142(84%)	159 (98%)
Black	12 (7%)	
Hispanic	12 (7%)	
Other	3 (2%)	

Comment#4: The ethnic distribution was very predominantly Caucasian. Statistics stratified by ethnicity will therefore be uninformative.

Table 2: Patient Disposition

	Study A03779	Study A00518
# Patients Enrolled	172	169
# Patients Dosed	169	162
#Patients excluded per No SOR evaluation	31	25
# Patients excluded per Protocol Deviations	7	6
# Remaining Preferred Population	131	131

Comment#5: The majority of patients excluded from the primary analyses were excluded because an SOR was unavailable.

Reviewer's Statistical Analyses Results: The principal findings are paraphrased in the numbered comments below. The supportive statistical analyses are presented in the subsequent tables.

(1): Primary Efficacy Result: *The Sponsor met the primary success criterion: Two out of three blinded readers achieved statistically significant increases in detection rates for Test (Pre+Post Contrast MRI) over Comparator (Pre Contrast MRI) in both studies. The Level of Superiority, however, was marginal: On the average, across the two studies, the point estimate difference in lesion detection rates was 5%, and the lower bound of the 95% two-sided CI was just slightly above zero. It is noted here that these analyses employed a "Within-Patient" detection statistic whose results were entirely consistent with the Sponsor's Wilcoxon statistic.*

**Table 3: Primary Results for the Detection Studies
Test (Pre+Post) versus Comparator (Pre) Success Levels in Lesion Detection**

Study A03779 (N = # Lesions = 316 ; # Patients = 131)					
	Pre+Post	Pre	Difference	95% CI	Wilcoxon Result
RDR#1	82%	77%	5%	(1% , 9)*	*
RDR#2	76%	73%	3%	(-1% , 7)	
RDR#3	78%	72%	6%	(0% , 10%)*	*
Average	79%	74%	5%	(0% , 10%)	
Study A00518 (N = # Lesions = 302 ; # Patients = 131)					
	Pre+Post	Pre	Difference	95% CI	Wilcoxon Result
RDR#1	81%	76%	5%	(1% , 9%)*	*
RDR#2	78%	76%	2%	(-1% , 5%)	
RDR#3	74%	71%	3%	(0 ⁺ % , 6%)*	*
Average	78%	74%	4%	(0% , 8%)	

* = Success under Wilcoxon Test and Within Patient T-Test

(2): Specificities and False Positive Rates: *A natural concern in lesion detection studies is the possibility that improvements in detection for a Test over a Comparator are reflective only of “overcalling” of lesions, in which case the gains are achieved at the expense of increased false positive rates. Consequently, false positive rates are required as a check on “overcalling”. Specificities are also a means of checking for overcalling. Specificity, however, has no meaning in these studies. In order to evaluate specificities, the anatomical area of interest needs to be partitioned into regions whose diagnoses of normality/abnormality can be verified by the SOR. The Agency requested that the Sponsor provide detection analyses of on a liver lobe level so that specificity could be evaluated. The analyses of both the false positive rates and the liver lobe level specificities reveal that overcalling was not an issue in these studies.*

Table 4: Percentages of Detection Study Patients with at Least One False Positive

Study A03779				Study A00518			
# MRI Patients =131 ; # CT Patients =128				# MRI Patients =131 ; # CT Patients =127			
	Pre	Pre+Post	CT		Pre	Pre+Post	CT
RDR#1	34%	37%	45%	RDR#1	44%	39%	53%
RDR#2	29%	31%	36%	RDR#2	24%	28%	32%
RDR#3	37%	34%	43%	RDR#3	28%	35%	26%
Average	33%	34%	41%	Average	32%	34%	37%

The Pre+Post reads do not present a statistically significant increase in false positives at the patient level. Note that CT has slightly higher false positive rates.

Table 5: Detection Studies Sensitivities/Specificities at the Liver Lobe Level

Study A03779					Study A00518				
		Pre	Pre+Post	Diff			Pre	Pre+Post	Diff
RDR#1	Sens	88%	89%	1%	RDR#1	Sens	86%	89%	3%
	Spec	73%	76%	3%		Spec	65%	73%	8%
RDR#2	Sens	85%	87%	2%	RDR#2	Sens	88%	90%	2%
	Spec	73%	76%	3%		Spec	74%	74%	0%
RDR#3	Sens	85%	90%	5%	RDR#3	Sens	82%	85%	3%
	Spec	70%	76%	6%		Spec	85%	71%	-14%*
Average	Sens	86%	89%	3%	Average	Sens	85%	88%	3%
	Spec	72%	76%	4%		Spec	75%	73%	-2%*

(3): Principal Secondary Analysis Result: *The lesion detection rates for Pre+Post Contrast MRI were , on the average, essentially equivalent to the lesion detection rates for Contrast CT. There is therefore no evidence that the new modality is either superior or inferior to the standard of practice modality for lesion detection. Again, the statistics used in these analyses were the reviewer’s “Within-Patient” statistic and the Sponsor’s Wilcoxon statistic, and the results were consistent across statistics.*

Table 6: Test versus Comparator (CT) Success Levels in Lesion Detection

Study A01908 (N = # Lesions =299 ; # Patients = 126)						
	Pre+Post	CT	Diff	Discordance	95% CI	Wilcoxon Result
RDR#1	82%	76%	7%	34%	(0% , 10%)	*
RDR#2	76%	72%	4%	39%	(-4% , 8%)	
RDR#3	78%	69%	9%	37%	(1% , 17%)	*
Average	79%	72%	7%	37%		
Study A01908 (N = # Lesions = 297 ; # Patients = 129)						
	Pre+Post	CT	Diff	Discordance	95% CI	Wilcoxon Result
RDR#1	81%	77%	3%	37%	(-3% , 9%)	
RDR#2	78%	73%	4%	42%	(-4% , 8%)	
RDR#3	74%	66%	8%	44%	(0% , 16%)	*
Average	78%	72%	6%	41%		

* = Rejection of Equality under both Wilcoxon and T-Test

Comment#6: The Discordance rates for these MRI versus CT comparisons are quite large approximately 35% to 40% . This could be reasonably strong evidence that the two modalities do not detect the same lesion types.

(4): Lesion Level Analyses: For several secondary analyses, “Lesion Level” statistics were more appropriate than the “Within-Patient” statistic. In fact, the lesion detection rates recorded by the Sponsor were always lesion level, although the Sponsor’s primary Wilcoxon analyses were not lesion level. This choice of lesion level over Within-Patient statistics worked to the Sponsor’s disadvantage since the lesion level results are more favorable for Eovist than were the “within-Patient statistics. A comparison of Lesion level to Within-Patient statistics for the primary endpoint is presented in the table below.

Table 7: Within Patient versus Lesion Weighted Detection Rates

STUDY A03779 (# Patients = 131 ; # Lesions = 316)			
	PRE	PRE+POST	DIFF
RDR#1	Within Patient = 77%	Within Patient = 82%	Within Patient = 5% (1%, 9%)
	Lesion Weighted = 63%	Lesion Weighted = 72%	Lesion Weighted = 9% (4%, 13%)
RDR#2	Within Patient = 73%	Within Patient = 76%	Within Patient = 3% (-1%, 7%)
	Lesion Weighted = 62%	Lesion Weighted = 68%	Lesion Weighted = 6% (1%, 11%)
RDR#3	Within Patient = 72%	Within Patient = 78%	Within Patient = 6% (0%, 10%)
	Lesion Weighted = 59%	Lesion Weighted = 68%	Lesion Weighted = 9% (4%, 14%)
STUDY A00518 (# Patients = 131 ; # Lesions = 302)			
	PRE	PRE+POST	DIFF
RDR#1	Within Patient = 76%	Within Patient = 81%	Within Patient = 5% (1%, 9%)
	Lesion Weighted = 71%	Lesion Weighted = 76%	Lesion Weighted = 5% (1%, 9%)
RDR#2	Within Patient = 76%	Within Patient = 78%	Within Patient = 2% (-1%, 5%)
	Lesion Weighted = 65%	Lesion Weighted = 70%	Lesion Weighted = 5% (1%, 8%)
RDR#3	Within Patient = 71%	Within Patient = 74%	Within Patient = 3% (0%, 6%)
	Lesion Weighted = 63%	Lesion Weighted = 68%	Lesion Weighted = 5% (1%, 9%)

Comment#7: The “Within-Patient” detection rates are always larger than the “Lesion Weighted” detection rates, by an average of 11% in the US Study and 7% in the EU Study.

Comment#8: The “Lesion Weighted” values, although smaller than the “Within-Patient” values, provide 95% two-sided CI’s for Test versus Comparator differences where all readers, rather than two out of three readers, present success for Test. It would therefore have been to the Sponsor’s advantage to use lesion level rather than Wilcoxon statistics in the primary analyses. In any event, all secondary analyses in this review will be carried out on the lesion level.

(5): Detection Rates Stratified by Classification: *It is important to know if the marginal improvement in detection rates for Pre+Post Contrast MRI over Pre Contrast MRI was consistent across the significant classification strata: Benign versus Malignant. A secondary analysis of detection rates stratified to benign and malignant lesions revealed that the superiority of Eovist over Pre Contrast MRI was achieved on both strata. The statistics also revealed that, averaged over readers and studies, CT was slightly better at detecting benign lesions.*

**Table 8: Lesion Detection Statistics by Classification Category
(Study A03779)**

Benign: 67 Lesions for CT; 79 Lesions for MRI Malignant: 227 Lesions for CT; 232 Lesions for MRI						
RDR	Truth	Pre	CT	Pre+Post	Pre+post - CT	Pre+Post - Pre
RDR#1	Benign	57%	73%	65%	- 8%	8%
	Malignant	66%	70%	74%	4%	8%
RDR#2	Benign	53%	60%	63%	3%	10%
	Malignant	66%	70%	70%	0%	4%
RDR#3	Benign	54%	54%	59%	5%	5%
	Malignant	61%	66%	72%	6%	11%
Average	Benign	55%	62%	62%	0%	8% (4%, 12%)
	Malignant	64%	69%	72%	3%	8% (6% , 10%)

**Table 9: Lesion Detection Statistics by Classification Category
(Study A00518)**

Benign: 79 Lesions for CT; 80 Lesions for MRI Malignant: 211 Lesions for CT; 215 Lesions for MRI						
RDR	Truth	Pre	CT	Pre+Post	Pre+post - CT	Pre+Post - Pre
RDR#1	Benign	64%	75%	71%	- 4%	7%
	Malignant	75%	77%	79%	2%	4%
RDR#2	Benign	58%	76%	60%	-16%	2%
	Malignant	70%	69%	74%	5%	4%
RDR#3	Benign	53%	62%	58%	-4%	5%
	Malignant	69%	66%	73%	7%	4%
Average	Benign	58%	71%	63%	-8%	5%* (0%, 10%)
	Malignant	71%	71%	75%	4%	4%* (1%, 7%)

Comment#9: The improved performance of Pre+Post over Pre (approximately 8%) in the US Study, 5% in the EU Study, distributes equally over benign and malignant strata.

Note also that these differences are point estimates derived from lesion level statistics rather than “Within-Patient” statistics. (The latter difference was 5%, as documented in Table(D).)

Comment#10: Both the Pre reads and the Pre+Post reads registered improved detection rates for malignant over benign lesions – a “study averaged” improvement of about 11% for both reads. The CT reads revealed little or no improvement in lesion detection from benign to malignant categories.

(6): Detection Rates Stratified by Isolated versus Clustered Status: *The principal exploratory finding in this review was the determination that Pre+Post Contrast MRI achieved its margin of superiority in lesion detection strictly on patients with “clustered” lesions, that is, in patients with more than one lesion. When a patient presented with only one lesion, (and here the lesion was predominantly malignant), all three modalities provided equivalent detection rates.*

**Table 10: Lesion Detection Statistics by Cluster Status
(Study A03779)**

Isolated Lesions: 57 Clustered Lesions: 258						
RDR	Status	Pre	CT	Pre+Post	Pre+post - CT	Pre+Post - Pre
RDR#1	Isolated	93%	98%	96%	-2%	3%
	Clustered	57%	64%	66%	2%	10%
RDR#2	Isolated	89%	96%	91%	-5%	2%
	Clustered	56%	61%	63%	2%	7%
RDR#3	Isolated	86%	89%	91%	2%	5%
	Clustered	53%	57%	63%	5%	10%
Average	Isolated	89%	94%	93%	-1% (-5%, 3%)	3% (-1, 7%)
	Clustered	55%	61%	64%	3% (-2%, 8%)	9%* (4% , 14%)

**APPEARS THIS WAY
ON ORIGINAL**

**Table 11: Lesion Detection Statistics by Cluster Status
(Study A00518)**

Isolated Lesions:55 Clustered Lesions: 247						
RDR	Status	Pre	CT	Pre+Post	Pre+post - CT	Pre+Post - Pre
RDR#1	Isolated	.85	.96	.87	-8%	2%
	Clustered	.68	.66	.74	8%	6%
RDR#2	Isolated	.96	.96	.93	-4%	-3%
	Clustered	.58	.65	.64	-1%	6%
RDR#3	Isolated	.84	.94	.85	-2%	2%
	Clustered	.59	.59	.64	7%	5%
Average	Isolated	88%	96%	88%	-7% (-12%, -2%)	0% (-3%, 3%)
	Clustered	62%	63%	67%	4% (0%, 8%)	6% (3%, 9%)

Comment#11: The important conclusion to draw from this table is that the margin of improvement in lesion detection for Pre+Post Images over Pre Images is achieved in the subpopulation of patients with multiple lesions.

3.1.2 Evaluation of the Efficacy for the Characterization Studies

There were two identical Phase III Lesion Detection studies. These will be treated side-by-side below.

Overview of the Design for the Detection Studies

StudyA05742 (Europe): This study enrolled 235 patients, with 182 patients included in the per-protocol analysis.

StudyA01908 (USA): This study enrolled 240 patients , with 177 patients included in the per-protocol analysis.

These two studies had the following identical Title and Design:

Study Title: A Multicenter open-label study of Gd-EOB-DTPA (Eovist) with a single i.v. injection (25 umol/kg BW) in patients with known or suspected focal liver lesions.

Study Efficacy Objective: To evaluate the ability of Eovist to provide additional information for characterization of liver lesions. (Lesion types range from malignancies such as hepatocellular carcinomas and metastases, through to benign types such as adenomas, hemangiomas, and cysts.)

Primary Inclusion Criterion: Patients are known or suspected for focal liver lesions.

Criteria for the Per-Protocol Population: Patients will have undergone pre and post Eovist MRI and a contrast CT within a six week period and will have received an appropriate SOR examination within a time frame that can be as long as one year prior to inclusion or three months after inclusion. The allowable pre-inclusion time period can depend on the SOR findings.

SOR and Lesion Identification: The On-Site investigator is the final judge for the SOR characterization. The investigator will receive all SOR reports and will determine the acceptability of the SOR characterization for each lesion. Malignant lesions must have histopathology as their SOR. Non-Study related MRI and sonography can serve as the SOR for benign lesion types. Lesions with SOR verification and characterization will be circled by the On-Site investigator, and these lesions will be the only lesions that will contribute to the Efficacy analyses.

Primary Reads for the Statistical Analyses: Three blinded readers will independently read all images from all image modalities, and will characterize all the lesions circled by the On-Site investigators.

Blinded Read Protocol:

Reading Sessions: There were three blinded readers who independently read all MRI and CT images. The image types pertinent to the primary and principal secondary analyses were read in two blocks of readings, each consisting of three separate sessions conducted sequentially as follows:

Session#1: Only CT scans.

Session#2: Both pre and post MRI images, presented separately and randomly.

Session#3: Combined pre + post MRI images.

Two to four weeks separated each session, in order to minimize recall. Presumably, the two distinct blocks of three separate sessions each included approximately half the patients, and the three sessions for Block#1 preceded the three sessions for Block#2.

Primary Efficacy Variable: The proportion of SOR detected and characterized lesions which were correctly characterized by the individual blinded readers.

Primary Statistical Hypotheses:

H₀: The proportion of correctly characterized lesions under pre+post contrast MRI reads is the same as the proportion of lesions with correct characterization for pre contrast MRI reads.

H₁: The proportion of correctly characterized lesions under pre+post contrast MRI reads differs from the proportion of lesions with correct characterization for pre contrast MRI reads.

Since some subjects will have several lesions, a McNemar Test adjusted for clustering will be used to evaluate these hypotheses.

Principal Secondary Efficacy Objectives:

(a): To compare pre + post Eovist MRI to contrast enhanced spiral CT for liver lesion characterization.

(b): To compare pre + post MRI to pre MRI and to CT for lesion classification (malignant ; benign ; non-assessable).

Overview of the Efficacy Results for the Characterization Studies

Table 12: Demographics

	Study A01908	Study A05742
# Dosed Patients/ #Per Protocol Patients	235/177	231/ 182
Mean Age (+- Sigma)	54 +-12 yrs	57 +- 15 yrs
Age Range	22 - 81	19 - 83
Gender	124 Male (53%) ; 111 (47%)Female	119Male (52%); 112 Female (48%)
Race:		
Caucasian	173(74%)	226 (98%)
Black	12 (5%)	
Hispanic	24 (10%)	
Asian	20 (9%)	
Other	6 (2%)	

Table 13: Patient Disposition

	Study A01908	Study A05742
# Patients Enrolled	240	235
# Patients Dosed	235	231
#Patients excluded per No SOR evaluation	38	29
# Patients excluded per Protocol Deviations	20	20
# Remaining Preferred Population	177	182

Reviewer’s Statistical Analyses Results: The principal findings are paraphrased in the numbered comments below. The data analyses are presented in the subsequent tables.

(1): Primary Endpoint Results: *The Sponsor met the primary success criterion: Two out of three blinded readers had statistically significant increases in characterization rates for Pre+Post Contrast MRI over Comparator Pre Contrast MRI.*

Table 14: Test versus Comparator Success Levels in Characterization

Study A01908 (N = # Lesions = 302 ; # Patients = 182)					
	Pre+Post	Pre	Diff	Discordance	95% CI
RDR#1	61%	60%	1%	24%	(-7% , 10%)
RDR#2	76%	65%	11%	22%	(5% , 18%)*
RDR#3	67%	48%	19%*	30%	(11% , 27%)*
Average	68%	58%	10%*	25%	(6% , 14%)*
Study A05742 (N = # Lesions = 259 ; # Patients = 177)					
	Pre+Post	Pre	Diff	Discordance	95% CI
RDR#1	67%	51%	16%*	31%	(7% , 25%)*
RDR#2	76%	59%	17%*	29%	(9% , 25%)*
RDR#3	58%	53%	5%	27%	(-2% , 12%)
Average	67%	54%	13%	29%	(9% , 17%)

* = Success under Adjusted McNemar Test

Comment#12: The Sponsor used the McNemar Test, applied at the lesion level for the test statistic. There was adjustment for clustering, since there were several lesions per patient. The reviewer agrees that this is an appropriate statistic for these studies. The test for statistical significance using this statistic is equivalent to satisfaction of the condition that the lower limit of the 95% two-sided CI for the Test versus Comparator mean difference in lesion classification rates exceed zero. As the table reveals, this criterion was met for two of the three readers in each characterization study.

Comment#13: The reviewer has included discordance rates in the tables. In these characterization studies the discordance rates are again moderately high and could be interpreted as evidence that the competing modalities are sensitive to different lesion types.

Comment#14: Unlike the results in the detection studies , the mean differences here vary considerably from reader to reader: 1% to 19%.

(2):Principal Secondary Comparison: *There was approximate equivalence in performance levels for lesion characterization between Pre+Post Contrast MRI and Contrast CT, with a marginal level of superiority for Pre+Post Contrast MRI in the European study.*

Table 15: Test versus CT Success Levels in Characterization

Study A01908 (N = # Lesions = 252 for Pre+Post versus CT Comparisons)					
	Pre+Post	CT	Diff	Discordance	95% CI
RDR#1	60%	57%	3%	49%	(-6% , 12%)
RDR#2	76%	75%	1%	35%	(-7% , 9%)
RDR#3	65%	65%	0%	41%	(-8% , 8%)
Average	68%	66%	2%		(-5% , 9%)
Study A05742 (N = # Lesions = 251 for Pre+Post versus CT Comparisons)					
	Pre+Post	CT	Diff	Discordance	95% CI
RDR#1	68%	58%	10%	27%	(3% , 17%)*
RDR#2	77%	64%	13%	26%	(7% , 19%)*
RDR#3	58%	55%	3%	28%	(-4% , 10%)
Average	68%	59%	8%		(4% , 12%)*

* = Rejection of Equality under Adjusted McNemar Test

Comment#15: As with the Detection Study, the discordance rates were large for secondary endpoint comparisons, (at least for the US Study) and most likely indicative that the modalities are sensitive to different lesion types.

Comment#16: The results present evidence that the characterizations by MRI were equal to CT in the US Study but better than CT in the EU Study. Since actual MRI characterization rates were the same in both studies, this discordance in the statistics between the two studies resides in the slightly poorer performance of CT in the EU Study. But the results are not sufficiently dramatic as to require a deeper investigation.

(3): Classification Results: *Characterizations – carcinomas, adenomas, cysts, etc - are finer distinctions than Classifications – benign ; malignant. A critical concern in the characterization studies was the subsidiary classification problem of correct lesion classification into benign and malignant: mischaracterizations that place a true malignant lesion into a benign class, or a true benign lesion into a malignant class, are more serious than mischaracterizations that preserve classifications. Consequently, secondary analyses on lesion classifications were carried out. These secondary analyses showed that , averaged over readers and studies, Pre+Post Contrast MRI was slightly superior to both Contrast CT and Pre Contrast MRI in lesion classifications.*

**APPEARS THIS WAY
ON ORIGINAL**

Table 16: Lesion Classification Statistics by Classification Category

Study A01908 (174 Benign Lesions ; 94 Malignant Lesions)					
	Pre	CT	Pre+Post	Pre+Post – CT	Pre+Post – Pre
Benign	71%	78%	79%	0% (-6% , 6%)	8%* (3% , 13%)
Malignant	91%	80%	91%	12% (5% , 19%)	0% (-5% , 5%)
Study A05742 (154 Benign Lesions ; 105 Malignant Lesions)					
	Pre	CT	Pre+Post	Pre+Post – CT	Pre+Post – Pre
Benign	77%	79%	86%	7% (3% , 11%)	9%* (4% , 14%)
Malignant	77%	84%	85%	2% (-4% , 8%)	8%* (3% , 13%)
Combined Studies (328 Benign Lesions ; 199 Malignant Lesions)					
	Pre	CT	Pre+Post	Pre+Post – CT	Pre+Post – Pre
Benign	74%	78%	82%	4%* (0% , 8%)	8%* (4% , 12%)
Malignant	84%	82%	88%	6%* (2% , 10%)	4%* (0% , 8%)

Comment#17: The overall average statistics reveal that the statistical superiority of Pre+Post Contrast MRI over both Pre Contrast MRI and Contrast CT in lesion classification is marginal.

3.2 Evaluation of Safety

The medical reviewer reported that there are no significant safety issues.

**APPEARS THIS WAY
ON ORIGINAL**

4. Findings in Special/Subgroup Populations

4.1 Gender, Race and Age

The two tables directly below list reader-averaged detection and characterization rates by age stratum. There are no outstanding differences from stratum to stratum for either of these rates.

Table 17: Lesion Detection Rates by Age

	Detection Study A03779					Detection Study A00518				
	#Patients	#Lesions	Pre	CT	Pre+Post	#Patients	#Lesions	Pre	CT	Pre+Post
Age < 45	23	52	60%	67%	60%	15	32	66%	69%	72%
45 ≤ Age < 65	51	152	59%	63%	69%	75	200	66%	68%	71%
65 ≤ Age	52	112	66%	72%	74%	39	70	70%	80%	74%
Overall	126	316	62%	67%	69%	129	302	67%	71%	72%

Table 18: Lesion Characterization Rates by Age

	Characterization Study A01908					Characterization Study A05742				
	#Patients	#Lesions	Pre	CT	Pre+Post	#Patients	#Lesions	Pre	CT	Pre+Post
Age < 45	57	68	60%	65%	71%	55	83	49%	57%	67%
45 ≤ Age < 65	123	145	57%	69%	68%	85	86	57%	62%	71%
65 ≤ Age	55	56	55%	58%	64%	91	90	56%	58%	63%
Overall	235	269	57%	66%	68%	231	259	54%	59%	67%

Comment #18: The number of patients in the Age stratified tables above, as provided by the Sponsor, are not for the Preferred Population; they are for the (Intent to Diagnose) ITD population. However, the entries under #Lesions are the correct numbers for the Preferred Population, and the statistics depend only on the number of lesions.

The next two tables, directly below, list reader-averaged detection and characterization rates by gender. There are no outstanding differences by gender for either of these rates.

Table 19: Lesion Detection Rates by Gender

	Detection Study A03779					Detection Study A00518				
	#Patients	#Lesions	Pre	CT	Pre+Post	#Patients	#Lesions	Pre	CT	Pre+Post
Male	71	189	60%	67%	68%	77	171	65%	68%	71%
Female	55	127	64%	67%	71%	52	131	69%	75%	73%
Overall	126	316	62%	67%	69%	129	302	67%	71%	72%

Table 20: Lesion Characterization Rates by Gender

	Characterization Study A01908					Characterization Study A05742				
	#Patients	#Lesions	Pre	CT	Pre+Post	#Patients	#Lesions	Pre	CT	Pre+Post
Male	124	125	55%	64%	65%	119	112	55%	58%	63%
Female	111	144	60%	67%	70%	112	147	54%	60%	70%
Overall	235	269	57%	66%	68%	231	259	54%	59%	67%

Comment #19: Once more, The number of patients in the Age stratified tables above, as provided by the Sponsor, are not for the Preferred Population; they are for the (Intent to Diagnose) ITD population. However, the entries under #Llesions are the correct numbers for the Preferred Population, and the statistics depend only on the number of lesions.

The racial composition in both sets of trials was 74% to 98% Caucasian, with an overall average of 88% Caucasian. The numbers for other groups were too small for meaningful statistical analyses.

4.2 Other Special/Subgroup Populations

No findings.

**APPEARS THIS WAY
ON ORIGINAL**

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

The Table below was constructed from the combined detection studies by randomly choosing a reader for each lesion. The table illustrates the general conclusions that were drawn from these trials:

(1): Pre+Post contrast MRI detection rates were equivalent to contrast CT detection rates and marginally superior to Pre contrast MRI detection rates.

(2): The equivalence of Pre+Post contrast MRI detection rates to contrast CT detection rates, and the marginal superiority to Pre+Post contrast MRI detection rates to Pre contrast MRI detection rates, were preserved across the strata of Benign and Malignant lesions. However, for all modalities, all detection rates were about 10% higher for Malignant over Benign strata.

(3): Pre+Post contrast MRI detection rates were equivalent to contrast CT detection rates over the two strata of patients with Isolated or Clustered lesions. Moreover, Pre Contrast MRI detection rates were equivalent to Pre+Post Contrast MRI detection rates over Isolated lesions. Consequently, the marginal superiority of Pre+Post Contrast MRI over Pre Contrast MRI was achieved exclusively on the stratum of patients with clustered lesions.

Overall Conclusion: *The moderate Superiority of Test to Comparator in the Detection studies is traceable to superiority in detection on clustered lesions.*

**APPEARS THIS WAY
ON ORIGINAL**

Table 21: Detection Results for a Random Reader in the Combined Studies

Overall (N = 617 Lesions)				
Modality		Detection Rate	Differences	
			Pre+Post - CT	Pre+Post - Pre
Pre		65%		
CT		68%		
Pre+Post		71%	2% (-2% , 6%)	6%* (3% , 10%)
Disease Status (N = 158 Benign Lesions; 447 Malignant Lesions)				
			Differences	
			Pre+Post - CT	Pre+Post - Pre
Pre	Benign	57%		
	Malignant	69%		
CT	Benign	64%		
	Malignant	72%		
Pre+Post	Benign	63%	-1% (-8% , 6%)	6%* (1% , 11%)
	Malignant	74%	2% (-3% , 7%)	5%* (2% , 8%)
Cluster Status (N = 112 Isolated Lesions ; 505 Clustered Lesions)				
			Differences	
			Pre+Post - CT	Pre+Post - Pre
Pre	Isolated	88%		
	Clustered	59%		
CT	Isolated	96%		
	Clustered	62%		
Pre+Post	Isolated	90%	-6% (-11% , -1%)	2% (-1% , 5%)
	Clustered	66%	4% (-1% , 9%)	7%* (3% , 11%)

**APPEARS THIS WAY
ON ORIGINAL**

The Table below was constructed from the combined characterization studies by randomly choosing a reader for each lesion. The table illustrates the general conclusions that were drawn from these trials:

(1): Pre+Post contrast MRI characterization and classification rates were equivalent to contrast CT rates and moderately superior to Pre contrast MRI rates.

(2): The equivalences of Pre+Post contrast MRI characterization and classification rates to contrast CT detection rates were preserved across the strata of Benign and Malignant lesions. The moderate superiority of Pre+Post contrast MRI characterization and classification rates to Pre contrast MRI were achieved only on Benign lesions – Test reads were less likely to mischaracterize or misclassify a benign lesion than were the Comparator reads. Characterization and classification rates improved from Benign to Malignant strata.

(3): Pre+Post contrast MRI characterization and classification rates were largely equivalent to contrast CT rates over the two strata of patients with Isolated or Clustered lesions. Pre+Post Contrast MRI characterization and classification rates were generally moderately superior to Pre Contrast MRI rates over both Isolated and Clustered lesions.

Overall Conclusion: *The moderate Superiority of Test to Comparator in the Characterization studies is traceable to superiority in characterization and classification on benign lesions.*

**APPEARS THIS WAY
ON ORIGINAL**

Table 22: Random Reader Characterization/Classification Results

Combined Characterization Studies							
(N = 553)							
Characterization Rates				Classification Rates			
		Rates	Differences		Rates	Differences	
			Pre+Post - CT	Pe+Post -Pre		Pre+Post - CT	Pe+Post -Pre
Pre		55%			77%		
CT		62%			80%		
Pre+Post		66%	4% (-1% , 9%)	10%* (6% , 14%)	83%	3% (-1% , 7%)	6% (3% , 9%)
Disease Status							
(#Benign = 322 ; # Malignant = 197)							
Characterization Rates				Classification Rates			
		Rates	Differences		Rates	Differences	
			Pre+Post - CT	Pe+Post -Pre		Pre+Post - CT	Pe+Post -Pre
Pre	Benign	53%			74%		
	Malignant	60%			83%		
CT	Benign	62%			79%		
	Malignant	62%			81%		
Pre+Post	Benign	69%	6% (0% , 12%)	16%* (11% , 21%)	82%	2% (-3% , 7%)	8%* (3% , 13%)
	Malignant	61%	1% (-8% , 10%)	1% (-6% , 8%)	86%	6% (-1% , 13%)	3% (-2% , 8%)
Cluster Status							
(# Isolated = 279 ; # Clustered = 274)							
Characterization Rates				Classification Rates			
		Rates	Differences		Rates	Differences	
			Pre+Post - CT	Pe+Post -Pre		Pre+Post - CT	Pe+Post -Pre
Pre	Isolated	50%			77%		
	Clustered	61%			77%		
CT	Isolated	60%			81%		
	Clustered	65%			80%		
Pre+Post	Isolated	59%	0% (-8% , 8%)	9%* (3% , 15%)	82%	2% (-4% , 8%)	5% (0% , 10%)
	Clustered	73%	8%* (1% , 15%)	12%* (6% , 18%)	84%	5% (-1% , 11%)	7%* (2% , 12%)

5.2. Conclusions and Recommendations

The Bayer Health Care submission (NDA22090) provided results for Eovist from four pivotal Phase III studies to support an indication for enhanced MRI of the liver. The four pivotal trials consisted of two identically designed liver lesion detection studies and two identically designed liver lesion characterization studies. For the primary analyses in the detection studies, lesion detection rates for the Test modality of combined pre and post Eovist contrast MRI were compared to the detection rates for pre contrast MRI. The objective was to show that the pre+post contrast MRI reads provided higher rates of liver lesion detections than did the pre contrast MRI reads. The lesion characterization studies compared pre+post contrast MRI image reads to pre contrast MRI image reads for lesion characterization in the primary analyses. The objective was to show that the pre+post contrast MRI reads provided higher rates of correct liver lesion characterizations than did the pre contrast MRI reads. Both sets of studies included secondary endpoint comparisons of pre+post contrast MRI to contrast CT, which is the current standard of care in liver imaging. Both sets of studies met their success criteria:

For each of the detection studies, two of three blinded readers achieved superior lesion detection rates for pre+post contrast MRI over pre contrast MRI.

For each of the characterization studies, two of the three blinded readers achieved superior lesion characterization rates for pre+post contrast MRI over pre contrast MRI.

The success in the detection studies was marginal in the sense that although the statistical significance was achieved, clinical meaningfulness of the result may be questionable. This success was achieved on patients with several rather than single lesions. The success in the characterization studies was somewhat more substantial and may be clinically meaningful, and was achieved largely in the characterization of benign lesions.

There is adequate evidence from the submitted studies to support approval of Eovist for enhanced MRI of the liver.

**APPEARS THIS WAY
ON ORIGINAL**

4 Page(s) Withheld

Trade Secret / Confidential

Draft Labeling

Deliberative Process

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

/s/

Tony Mucci
6/12/2008 05:48:26 PM
BIOMETRICS

Jyoti Zalkikar
6/12/2008 05:50:13 PM
BIOMETRICS

Rajeshwari Sridhara
6/12/2008 05:52:02 PM
BIOMETRICS