

**CENTER FOR DRUG EVALUATION AND RESEARCH**

**APPLICATION NUMBER: NDA 20937**

**STATISTICAL REVIEW(S)**

## Statistical Review and Evaluation - NDA Review

**NDA#:** 20-937 **DATE:** NOV 6 1998

**SPONSOR:** Mallinckrodt Inc.

**DRUG:** Optimark

**INDICATION:** MRI Contrast Agent for CNS and Liver

**DOCUMENTS REVIEWED:** NDA Volumes 2.1 and 2.46 through 2.87 and electronic database

**DATE:** Date received by Medical Division (Stamp Date): March 4, 1998  
Date received by Division of Biometrics: March 9, 1998

**PDUFA DATE:** March 2, 1999

**STATISTICAL REVIEWER:** Ruthanna Davi, M.S.

**MEDICAL REVIEWERS:** Ramesh Raman, M.D. (CNS, safety)  
Robert Yaes, M.D. (Liver)

**PROJECT MANAGER/CSO:** Kim Colangelo, James Moore

### **KEY POINTS:**

- Studies 488 and 525 may contain an over-representation of post-treatment patients. The over inclusion of this type of patient is enhancing the magnitude of the mean change in the primary endpoints from pre-image to image pair for both the Optimark and Magnevist groups. (See Appendix A for details.)
- In most cases, there are statistically significant increases in the mean and median of the primary endpoints from pre-image to image pair for both Magnevist and Optimark. (See Section 3.2.1 for details.)
- In all but one case, the confidence interval for the mean difference between the change in primary endpoints using Optimark and the same such change using Magnevist was completely contained within  $-1.5$  to  $+1.5$ . These results indicate that the effects of Magnevist and Optimark (in terms of the primary endpoints) are similar. (See Section 3.2.2 for details.)
- Equivalence trials are inherently difficult to interpret due in part to the fact that poor designs, ill-defined endpoints, etc. may make it impossible to show a difference between treatment groups even if one exists. i.e., Unlike in traditional superiority trials, conducting a sloppy trial can be an asset in an "equivalence" trial.

Key words: clinical studies, equivalence (clinical)

## **1.0 Introduction**

The sponsor has submitted the results of six phase 3 trials in support of the efficacy of Optimark to be used with MRI of the central nervous system and liver. Two of these studies (#484/485, 486/487) were open-label and are not considered pivotal by the sponsor for the demonstration of efficacy. Therefore, this review will consist of an assessment of studies #488 and #525 in support of the CNS indication and studies #490 and #526 for the liver indication.

## **2.0 Study Design**

Since the study design and primary analyses for the four pivotal studies were nearly identical, an integrated discussion of these trials will be presented. However to allow for interpretation of each study on its own merits, data analyses will be presented separately for each study.

The pivotal studies were designed to evaluate the safety and efficacy of intravenously administered 0.1 mmol/kg Optimark (gadoversetamide injection) compared to 0.1 mmol/kg Magnevist (gadopentetate dimeglumine injection) as a contrast-enhanced MRI agent in patients with highly suspected CNS pathology (study 488 and 525) or known or highly suspected liver pathology (studies 490 and 526). All trials were multicenter, double-blind, parallel group, randomized studies. The objective of these studies was to show that Optimark is not inferior to Magnevist for the previously described indication.

## **2.1 Subject Enrollment**

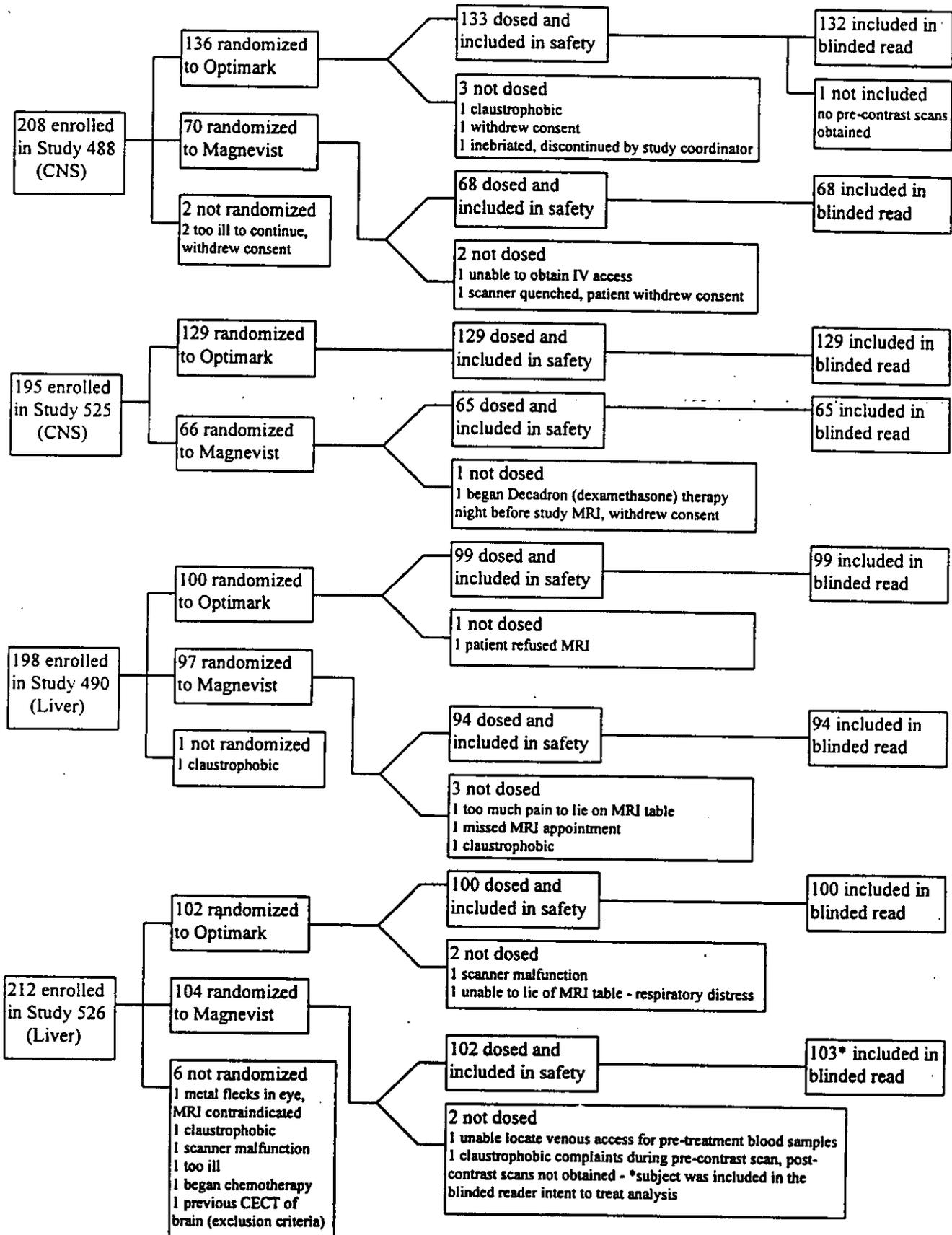
Four hundred three subjects with highly suspected CNS pathology (per an inclusion criteria requiring a "qualifying MRI evaluation" within eight weeks prior to study MR examination) were enrolled in studies 488 and 525. Four hundred ten subjects with known-or-highly suspected liver pathology (per an inclusion criteria requiring a "qualifying CECT evaluation" within three weeks prior to study MR examination) were enrolled in studies 490 and 526.

Enrolling subjects based on a qualifying exam which is identical to the procedure being used in the study (as in studies 488 and 525) may have allowed investigators to subconsciously not enroll subjects for whom MR imaging is unclear or inconclusive. This would result in the study samples containing an over-representation of patients with obvious disease. Concerns that the study group contains an over-representation of post-treatment patients (including post-surgical, post-biopsy, post-radiation, post-chemotherapy) have been expressed by the FDA Medical Reviewer. (Please see the medical review for complete details regarding the enrollment of post-treatment patients.) It is conceivable that the blinded readers' ratings of the primary endpoints for the images of post-treatment patients may be different from their ratings of other patients' images because of the post-treatment patients' medical status. If this is true, scoring of the pre-images as well as the scoring of the image pairs would be effected for these patients. However, the magnitude of this effect may be larger for the image pair than for the pre-image readings causing an inflated change in scores from pre-image to image pair for both the Optimark and Magnevist groups. This relationship is explored further in Appendix A.

In interpreting the results of these trials, the reader should keep in mind that the study samples may have been enriched with post-treatment patients and thus may present the phenomenon described above and in Appendix A.

The patient disposition for each of the studies follows that outlined in the schematics in Figure 1.

**Figure 1: Patient Disposition**



## 2.2 Imaging Methods and Blinded Reads

Each patient was imaged prior to contrast administration and after contrast administration. Patients received T1-weighted, T2-weighted, and proton density images prior to contrast administration. The imaging parameters and imaging plane were determined by the investigator at pre-injection and were held constant for post-injection imaging. Post-contrast imaging was to begin within one hour of contrast administration.

Within this review, the terms "pre-contrast image set" or "pre-image" will be used to represent the pre-contrast T1-weighted, T2-weighted, and proton density images for a patient. The term "pre and post-contrast pair" or "image pair" refers to those images in the pre-contrast image set as well as the post-contrast T1-weighted image for that patient.

The pre-images and image pairs were evaluated in random order by one of three randomly assigned blinded (to dose and clinical history) readers. Independent randomizations were used to order the pre-images and the image pairs. Each blinded reader assessed images from approximately one-third of the subjects in that study. None of the readers for Studies 488, 490, 525, and 526 participated in more than one trial. Among other things, the blinded readers were asked to assess the following.

- I. Primary Endpoints:
  - A. Choose one number that best reflects the **level of conspicuity for all lesions visualized**. Possible responses include 1, 2, . . . , 10 (1=no lesions, . . . , 5=moderately visualized, . . . , 10=clearly visualized)
  - B. Choose one number that best reflects your **ability to delineate lesion borders**. Possible responses include 1, 2, . . . , 10 (1=no lesions, . . . , 5=moderately visualized, . . . , 10=clearly visualized)
  - C. Select one number that best reflects your **degree of confidence in the diagnosis(es)**. Possible responses include 1, 2, . . . , 10 (1=no lesions, . . . , 5=moderately visualized, . . . , 10=clearly visualized)
- II. Select Secondary Endpoints
  - A. Indicate the **number of lesions** for this patient as indicated by the given set of images. Possible responses include 0, 1, . . . , 10, >10  
Being examined in this review as a result of interest initiated by FDA medical division
  - B. Indicate the **diagnosis(es)** for this patient based on the MR images provided.  
A checklist of possible diagnoses was provided to the blinded reader  
Used to assess diagnostic agreement of the images with the "final diagnosis"

In addition to the blinded readers described above, a fourth independent blinded (to dose) reader was used to assess the **extent of agreement between, a patient's "final diagnosis" and the blinded reader's "image pair diagnosis"** for each patient. The "final diagnosis" was established by the site investigator based on any or all of the following: computed tomography, CECT evaluation (including the qualifying CECT), previous MR evaluations with or without contrast medium, unenhanced MR evaluation from this study, ultrasound, nuclear medicine evaluation, CTAP, patient's clinical course, physical examination, laboratory evaluations, biopsy/surgery, histology findings, autopsy report. Possible responses for the extent of agreement appraisal include:

**Not Evaluable:** Information from the blinded review image record cannot be compared to the final clinical diagnosis (e.g., the images were not technically satisfactory).

**No Agreement:** No agreement in the diagnosis(es) indicated from the blinded review image record compared to those indicated in the final clinical diagnosis record.

**Partial Agreement:** Incomplete or fractional agreement in the diagnosis(es) indicated from the blinded review image record compared to those indicated in the final clinical diagnosis record.

**Basic Agreement:** Basic agreement supported by identical diagnosis(es) yet different number of lesions detected from the blinded review image record compared to final clinical diagnosis record.

**Absolute Agreement:** Total agreement based on identical diagnosis(es) and same number of lesions detected in the blinded review image record compared to final clinical diagnosis record.

**2.3 Safety Information**

Since the safety data (submitted electronically on a laptop) was received on October 7, 1998, assessment of the safety of Optimark as compared to Magnevist will be considered in an addendum to this review.

**3.0 Efficacy**

The objective of the trials in terms of efficacy was to demonstrate that Optimark is not inferior to Magnevist as a contrast-enhanced MRI agent in patients with highly suspected CNS pathology (study 488 and 525) or known or highly suspected liver pathology (studies 490 and 526).

Although the designation of the primary and secondary efficacy endpoints was consistent from the protocol to the NDA submission, the statistical methods being used to analyze these endpoints were not. The efficacy results contained in the NDA are derived using Analysis of Variance, a *customized* method for calculating confidence intervals, and an equivalence region defined as -1.5 to +1.5. The protocol calls for the use of t-tests and *standard* confidence interval methods, but does not clearly define an equivalence region. Due to these discrepancies, the following section will describe, compare, and discuss the advantages and disadvantages of the statistical procedures which were planned for in the protocol, used in the NDA, and implemented by this reviewer.

**3.1 Efficacy Analysis Plans**

In terms of the primary endpoints, there are two important questions to be answered to demonstrate the efficacy of Optimark:

- (1.) Within the two drug groups, is the change in the primary endpoints from the pre-image to the image pair significantly different from zero;
- (2.) Given that there is such a change as is referred to in #1, is the magnitude of this change the same for Magnevist and Optimark. (This comparison is designated by the sponsor as the primary analysis to demonstrate non-inferiority. This designation is made both in the protocol and in the NDA submission.)

In addition, since not all readers read all images, it is necessary to convince ourselves that the results we are seeing for items #1 and #2 and not being driven by certain readers. In other words, we would like to confirm that overall there is not a significant reader effect.

**Table 2: Summary of Analyses (planned in protocol, used in the NDA, used in this review)**

Comparison	#1. Is there a statistically significant change between the pre-image and the image pair scores?	#2. Is the change in #1 equivalent for Magnevist and Optimark? (Primary Analysis)	#3. Assessment of Reader Effect
Per-Protocol Analyses	Paired t-test comparing the mean pre-image score to the mean image pair score within each drug group $t = \frac{\sqrt{n}(\bar{x}_{pair} - \bar{x}_{pre})}{S_d}$	Two-one sided confidence interval methodology $(\bar{d}_O - \bar{d}_M) \pm t_{(1-\alpha/2, n_O+n_M-2)} \sqrt{\frac{S_{d,O}^2}{n_O} + \frac{S_{d,M}^2}{n_M}}$ equivalence region not clearly specified	not addressed
NDA Study Report Analyses	<i>Customized</i> paired t-test comparing the mean pre-image score to the mean image pair score within each drug group $t = \frac{\sqrt{n}(\bar{x}_{pair} - \bar{x}_{pre})}{MS(R * T)}$	Two-one sided <i>customized</i> confidence interval methodology $(\bar{d}_O - \bar{d}_M) \pm t_{(1-\alpha/2, n_O+n_M-2)} \sqrt{\frac{MS(R * T)}{n_O} + \frac{MS(R * T)}{n_M}}$ contained within -1.5 and +1.5 equivalence region	ANOVA with reader term
Analyses utilized in this review	Wilcoxon sign rank test comparing median pre-image score to median image pair score within each drug group.	Per-protocol confidence interval methodology	Graphically display differences between readers (ignoring drug group)*

\*Evidence of an interaction between reader and dose group generally, was not present in the sponsor's ANOVA results.

### **3.1.1 Regarding the analyses to answer question #1**

A paired t-test comparing the mean pre-image score to the mean image pair score provides information regarding the mean change from pre-image to image pair. However, in some cases, analysis of means can be misleading. Clearly, the estimate of a mean can be over-influenced by outlying observations. The data from these trials is such that in most cases, approximately 35% of the subjects' images were rated the same on the pre-image as on the image pair. This resulted in the difference from pre-image to image pair being zero for approximately 35% of subjects. This clustering of the data around zero caused the variance for this data to be very small. A small variance coupled with the fact that an outlying observation may draw the mean in one direction, sets the stage for reaching statistical significance. Therefore, it is being proposed by this reviewer that in addition to evaluating the primary endpoints using the paired t-test, the results the Wilcoxon sign rank test comparing the median pre-image score to the median image pair score should be considered.

Note in Table 2 that the NDA study report implemented a *customized* paired t-test. Rather than using the standard deviation of the data in each drug group in the calculation of the confidence intervals, the sponsor used the overall mean square error due to the treatment-by-reader interaction term from an ANOVA model. The ANOVA model contained a term for reader, treatment, and reader-by-treatment interaction with the response variable being the change between pre-image and image pair scores. The sponsor's justification for this is that this is an attempt to discount the variability due to reader. This is not a commonly used and/or appropriate statistical method and it was not prescribed in the protocol. In addition, if Optimark were to be used in the clinical setting, the variability due to the image reader would be inevitable. For these reasons, it is the opinion of this reviewer that the results of the standard paired t-test and the previously mentioned Wilcoxon sign rank test would be most useful in this scenario. This review presents the results of these two tests in Section 3.2 Efficacy Results.

### **3.1.2 Regarding the analyses to answer question #2**

Note in Table 2 that the NDA study report implemented a *customized* method for calculating the confidence intervals around the difference between the change in scores using Optimark and the change in scores using Magnevist. As was the problem associated with question #1, rather than using the standard deviation of the data in each drug group in the calculation of the confidence intervals, the sponsor used the mean square error due to the treatment-by-reader interaction term from the ANOVA model. Again, this is not a commonly used and/or accepted statistical method and it was not prescribed in the protocol. Once again, the variability due to the image reader in a clinical setting will be unavoidable and therefore should be considered in the analysis.

In addition, the use of ANOVA in this situation may not be completely straightforward. The sponsor proposed using a completely random model (i.e., both reader and treatment designated as random effects). However, it is not completely evident that this is the most appropriate designation. In these studies, we are examining two treatments, Optimark and Magnevist, and we wish to state results about only these two treatments, not some universal population of treatments. Therefore, the purpose for designating treatment as a random effect is unclear. Considering reader as a random effect is perhaps more plausible since we would like to generalize the results of these trials to a larger population of all readers who may use Optimark in a clinical setting. This type of generalization is most appropriate however, if the number of readers is large. Since there are only three readers represented in each of these trials, one may argue that making general comments about all readers after examining only three of them is inappropriate and therefore, that reader should be a fixed effect. For further information, refer to the ICH Harmonized Tripartite Guideline, Statistical Principles for Clinical Trials where designation of center (which is analogous to reader in this trial) as a random effect is discussed. The issues associated with designation of fixed and random effects will be avoided by using the per-protocol analysis.

For these reasons, it is the opinion of this reviewer that the results of the per-protocol analysis would be more appropriate. Unfortunately however, the protocol does not clearly specify an equivalence region so the confidence interval for the difference between the mean changes may need to be interpreted on its own merits (e.g., does it include zero, failing to demonstrate a statistically significant difference).

Otherwise, one may elect to use the equivalence region established (possibly post hoc) in the NDA study report which is as follows. If the previously mentioned confidence interval is completely contained within -1.5 to +1.5 then the effects of Optimark and Magnevist would be considered equivalent. This review presents the per-protocol results in Section 3.2 Efficacy Results.

### **3.1.3 Regarding the analyses to address issue #3**

A graphical presentation of the difference in the means (and confidence intervals) from pre-image to image pair for each of the readers allows an informal look at the way the readers tended to respond. When this presentation is coupled with the p-value associated with a reader term resulting from the ANOVA provided in the NDA study report, the presence or absence of a reader effect can be formally confirmed. These graphical presentations and a summary of the sponsor's results will be provided in section 3.2 Efficacy Results.

## **3.2 Efficacy Results**

### **3.2.1 Discussion of results of analyses to answer question #1**

Recall that we are interested in discovering if there is a statistically significant change between the pre-image scores and the image pair scores. Appendix B contains results that are presented for that end.

The frequency tables are provided to give a "feeling" for the way the data looks. In considering the frequency tables, one may wish to note how many subjects have pre-image scores which are the same (on diagonal), higher (below diagonal), or lower (above diagonal) than the image pair scores. In addition, a visual comparison of the changes from pre-image to image pair for Optimark relative to Magnevist can be made in considering the Optimark and Magnevist tables simultaneously.

The paired t-test and the Wilcoxon sign rank test are provided in order to quantify what is seen in the frequency tables. Within each drug group, the paired t-test is being used to test whether the mean score for the image pair is different from that of the pre-image. The Wilcoxon sign rank test is being used to test if the median score for the image pair is different from that of the pre-image..

*Reviewer's Conclusion #1: In most cases, a statistically significant change from the pre-image scores to image pair scores was observed for both the Optimark and Magnevist groups. The paired t-test shows a statistically significant increase in the mean score from pre-image to image pair in nearly all cases (e.g., a mean change of approximately 1 unit on a scale of -10 to +10 was observed in many cases). The results of the Wilcoxon sign rank test are consistent with those of the paired t-test. According to the Wilcoxon sign rank test, the median image pair score was higher than the pre-image score in the majority of cases.*

### **3.2.2 Discussion of results of analyses to answer question #2**

Recall that we are interested in discovering if the change in scores from pre-image to image pair is equivalent for Magnevist and Optimark. Table 3 and Figure 2 are presented for that end.

Table 3 and Figure 2 show the 95% confidence intervals for the mean difference in each of the changes in the primary endpoints between Optimark and Magnevist. If these confidence intervals fall completely within the "equivalence region" (defined possibly post-hoc by the sponsor in the NDA study report as -1.5 to +1.5) then Optimark and Magnevist were considered equivalent in terms of that endpoint.

Defining the equivalence region is not a statistical argument. The clinical relevance and validity of an equivalence region from -1.5 to +1.5 should be carefully considered since redefining this region would completely alter the interpretation of the results in Table 3 and Figure 2. In considering this issue, it may be important to note that due to the fact that approximately 35% of the subjects had pre-image scores equal to their image pair scores (i.e., approximately 35% of the changes in scores were zero), the variability of the pre to pair differences was very small (i.e., achieving narrow confidence intervals was almost certain).

**Table 3: 95% Confidence Intervals for the Difference in Mean Change Between Optimark and Magnevist in each of the Primary Endpoints\*\***

Endpoint	Study	Analysis Method	Lower 95% Confidence Limit	Mean	Upper 95% Confidence Limit
Conspicuity	488 (CNS)	NDA Submission	-0.015	0.665	1.345
		Per-Protocol	-0.103	0.665	1.433
	525 (CNS)	NDA Submission	0.084	0.234	0.383
		Per-Protocol	-0.900	0.234	1.362
	490 (Liver)	NDA Submission	*	*	*
		Per-Protocol	*	*	*
	526 (Liver)	NDA Submission	-0.371	0.027	0.425
		Per-Protocol	-0.653	0.027	0.707
Border Delineation	488 (CNS)	NDA Submission	-0.174	0.207	0.588
		Per-Protocol	-0.577	0.207	0.991
	525 (CNS)	NDA Submission	0.109	0.424	0.740
		Per-Protocol	-0.810	0.424	1.658
	490 (Liver)	NDA Submission	-1.347	-0.491	0.365
		Per-Protocol	-1.264	-0.491	0.281
	526 (Liver)	NDA Submission	-0.095	0.164	0.423
		Per-Protocol	-0.513	0.164	0.842
Diagnostic Confidence	488 (CNS)	NDA Submission	-0.230	-0.780	0.074
		Per-Protocol	-0.662	-0.780	0.506
	525 (CNS)	NDA Submission	0.017	0.218	0.418
		Per-Protocol	-0.521	0.218	0.956
	490 (Liver)	NDA Submission	-0.214	-0.050	0.115
		Per-Protocol	-0.842	-0.050	0.742
	526 (Liver)	NDA Submission	-0.584	-0.241	0.102
		Per-Protocol	-0.988	-0.241	0.507

\*A significant interaction between reader and treatment was observed for conspicuity in Study 490. For both Readers 2 and 3, neither of the mean difference scores for Optimark and Magnevist were significantly different from zero. However, for Reader 1, the mean difference score for Optimark was significantly greater than zero and the mean difference score for Magnevist was not different from zero. Calculation of confidence intervals ignoring reader may be misleading.

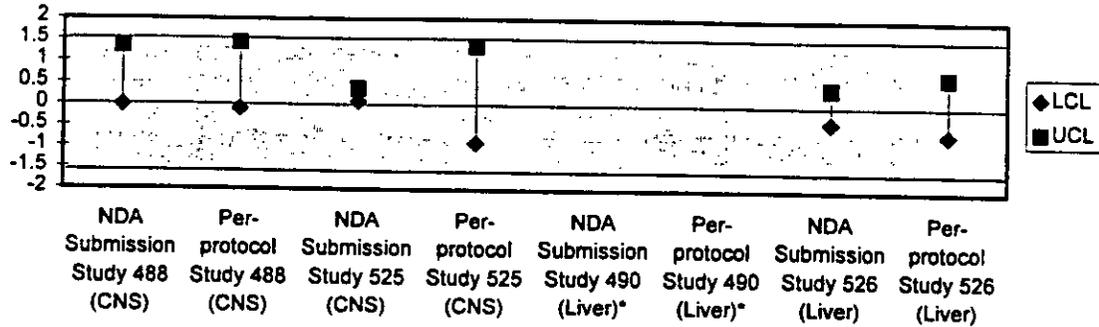
\*\*Difference = Magnevist change minus Optimark change.

**Reviewer's Conclusion #2:** *If an equivalence region from -1.5 to +1.5 is clinically acceptable then it is the opinion of this reviewer that equivalence between Optimark and Magnevist in terms of the mean change in scores from pre-image to image pair has been demonstrated. Although using the per-protocol analysis methods resulted in confidence intervals which are wider than those presented in the NDA submission, it is still true that in all but one case (per-protocol analysis of border delineation for study 490) the confidence intervals are completely contained within -1.5 and +1.5.*

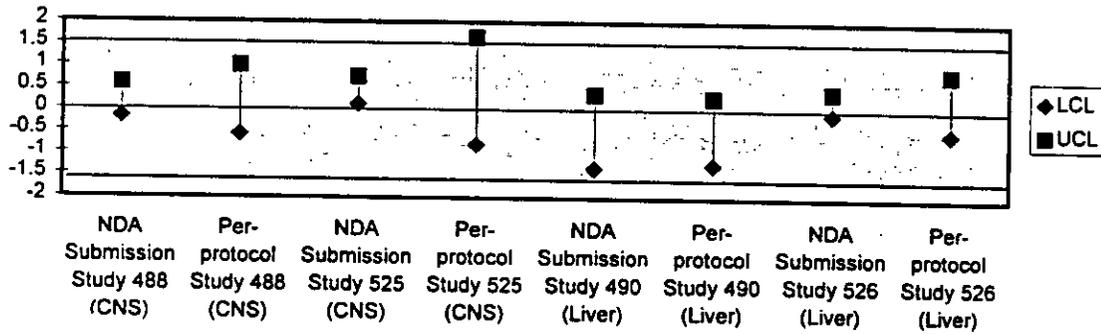
**APPEARS THIS WAY  
ON ORIGINAL**

Figure 2: Graphical Display of Primary Analysis\*\*

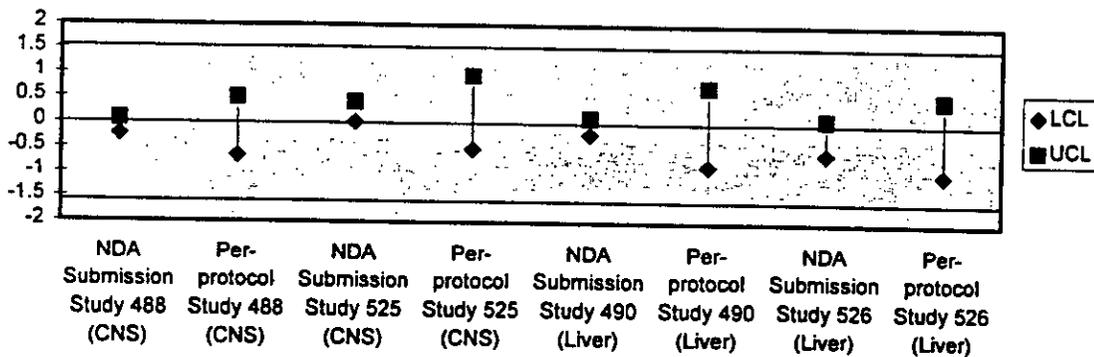
95% C.I. for Mean Difference in Conspicuity Between Optimark and Magnevist



95% C.I. for Mean Difference in Border Delineation Between Optimark and Magnevist



95% C.I. for Mean Difference in Diagnostic Confidence Between Optimark and Magnevist



\*A significant interaction between reader and treatment was observed for Conspicuity in Study 490. For both Readers 2 and 3, neither of the mean difference scores for Optimark and Magnevist were significantly different from zero. However, for Reader 1, the mean difference score for Optimark was significantly greater than zero and the mean difference score for Magnevist was not different from zero. Calculation of confidence intervals ignoring reader may be misleading.

\*\*Difference = Magnevist score minus Optimark score.

**3.2.3 Discussion of results to assess reader effect:**

Figure 3 gives a graphical display of the 95% confidence intervals for the mean difference between pre-image and image pair scores for each of the primary endpoints. The intervals are shown for each reader and ignore drug group assignment (in order to assess reader as a main effect).

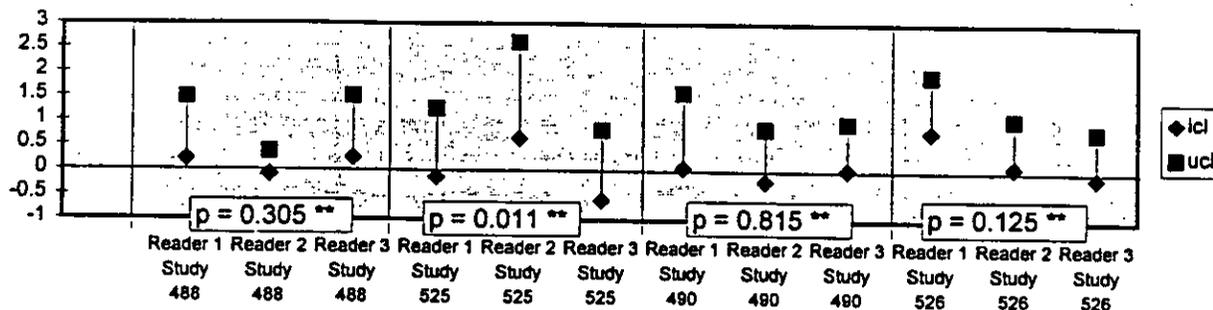
According to the sponsor's ANOVA, the only statistically significant reader effect occurred in Study 525 in the analysis of the change from pre-image to image pair Conspicuity scores and Diagnostic Confidence scores. For the conspicuity endpoint, Figure 3 shows that this significance is a result of Reader 2 having a change in scores that was significantly different from zero while these changes for Readers 1 and 3 were not statistically significantly different from zero. For the diagnostic confidence endpoint, Readers 1 and 2 had changes in scores that were significantly different from zero while Reader 3 did not.

In addition to differences in means, Figure 3 can be used to qualitatively assess differences in variability. For example, in Study 488 Reader 2 results have lower variability (as evidenced by the narrow confidence intervals) than those of Readers 1 and 3 for that study.

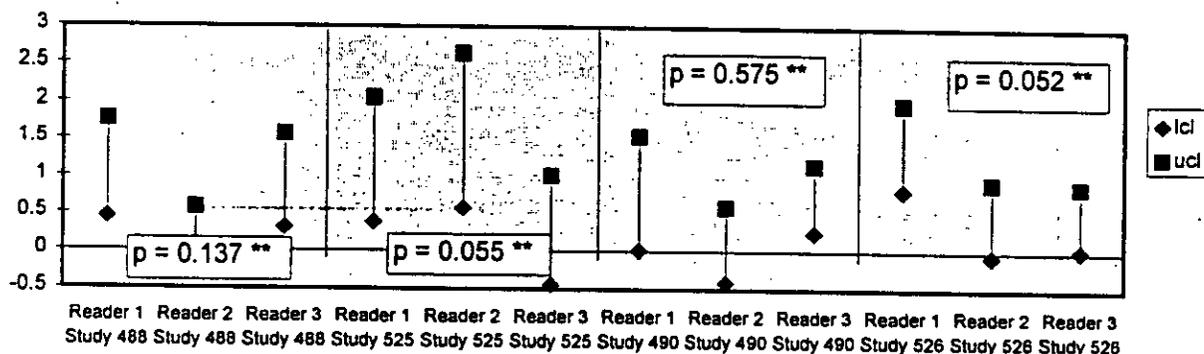
**APPEARS THIS WAY  
ON ORIGINAL**

Figure 3: Graphical Display of Reader Effect

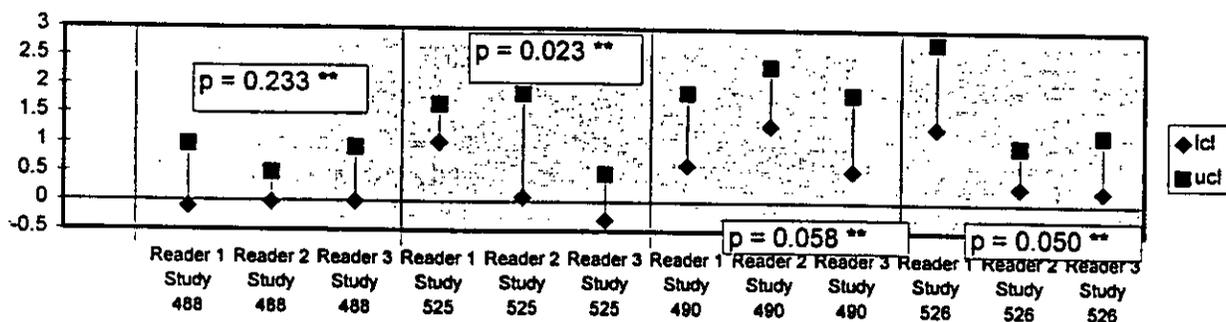
95% C.I. for Mean Difference Between Pre-Image and Image Pair Conspicuity Scores (by reader, ignoring drug group)\*



95% C.I. for Mean Difference Between Pre-Image and Image Pair Border Delineation Scores (by reader, ignoring drug group)\*



95% C.I. for Mean Difference Between Pre-Image and Image Pair Diagnostic Confidence Scores (by reader, ignoring drug group)\*



\*Difference = Image Pair Score - Pre-Image Score

\*\*P-value corresponds to reader effect in ANOVA model with reader, treatment, and reader\*treatment terms. (per NDA submission)

### 3.4 MRI Diagnoses Compared to the Final Diagnosis

The results for agreement with the final diagnosis are presented below. This endpoint was designated as a secondary endpoint by the sponsor (in the protocol as well as in the NDA study report). It is being examined in this review as a result of interest initiated from the FDA medical division. The following tables are taken directly from the NDA study report.

Since the trials enrolled only subjects with highly suspected disease, very few truly normal subjects were entered. This limits the ability of these trials in proving that Optimark (or Magnevist) can be used to distinguish normal subjects from a mixed patient group. This may not be a substantial concern however, since this agent would typically be used in a patient population which is highly suspected of disease, not as a screening tool.

**Table 4: Agreement Between MRI Diagnoses and the Final Diagnoses - Study 488 (CNS)**

Optimark	Pre-Contrast Diagnosis N(%)				
	Not Evaluable	No Agreement	Partial Agreement	Basic Agreement	Absolute Agreement
Disease	1 (0.8)	65 (53.3)	14 (11.5)	11 (9.0)	31 (25.4)
No Disease		5 (71.4)			2 (28.6)
	Pre- plus Post-Contrast Diagnosis N(%)				
Disease	1 (0.8)	57 (46.7)	18 (14.8)	13 (10.7)	33 (27.1)
No Disease		5 (71.4)			2 (28.6)
Magnevist	Pre-Contrast Diagnosis N(%)				
Disease		25 (39.7)	9 (14.3)	9 (14.3)	20 (31.8)
No Disease		1 (50.0)			1 (50.0)
	Pre- plus Post-Contrast Diagnosis N(%)				
Disease		25 (39.7)	7 (11.1)	9 (14.3)	22 (34.9)
No Disease		1 (50.0)			1 (50.0)

**Table 5: Agreement Between MRI Diagnoses and the Final Diagnoses - Study 525 (CNS)**

Optimark	Pre-Contrast Diagnosis N(%)				
	Not Evaluable	No Agreement	Partial Agreement	Basic Agreement	Absolute Agreement
Disease	5 (4.5)	42 (37.5)	18 (16.1)	17 (15.2)	30 (26.8)
No Disease	1 (5.9)	5 (29.4)			11 (64.7)
	Pre- plus Post-Contrast Diagnosis N(%)				
Disease	3 (2.7)	42 (37.5)	22 (19.6)	17 (15.2)	28 (25.0)
No Disease	1 (5.9)	7 (41.2)			9 (52.9)
Magnevist	Pre-Contrast Diagnosis N(%)				
Disease	1 (1.9)	22 (40.7)	10 (18.5)	11 (20.4)	10 (18.5)
No Disease		6 (54.5)			5 (45.5)
	Pre- plus Post-Contrast Diagnosis N(%)				
Disease	1 (1.9)	22 (40.7)	12 (22.2)	8 (14.8)	11 (20.4)
No Disease		6 (54.5)			5 (45.5)

**Table 6: Agreement Between MRI Diagnoses and the Final Diagnoses - Study 490 (Liver)**

Optimark	Pre-Contrast Diagnosis N(%)				
	Not Evaluable	No Agreement	Partial Agreement	Basic Agreement	Absolute Agreement
Disease	1 (1.1)	15 (16.9)	26 (29.2)	28 (31.5)	19 (21.4)
No Disease		3 (75.0)	1 (25.0)		
	Pre- plus Post-Contrast Diagnosis N(%)				
Disease		15 (16.9)	27 (30.3)	27 (30.3)	20 (22.5)
No Disease		2 (50.0)	2 (50.0)		
<b>Magnevist</b>	Pre-Contrast Diagnosis N(%)				
Disease	1 (1.2)	13 (15.5)	21 (25.0)	30 (35.7)	19 (22.6)
No Disease			2 (33.3)	1 (16.7)	3 (50.0)
	Pre- plus Post-Contrast Diagnosis N(%)				
Disease		17 (20.2)	17 (20.2)	27 (32.1)	23 (27.4)
No Disease		1 (16.7)	2 (33.3)	1 (16.7)	2 (33.3)

**Table 7: Agreement Between MRI Diagnoses and the Final Diagnoses - Study 526 (Liver)**

Optimark	Pre-Contrast Diagnosis N(%)				
	Not Evaluable	No Agreement	Partial Agreement	Basic Agreement	Absolute Agreement
Disease		11 (12.2)	13 (14.4)	37 (41.1)	29 (32.2)
No Disease		2 (22.2)		3 (33.3)	4 (44.4)
	Pre- plus Post-Contrast Diagnosis N(%)				
Disease	1 (1.1)	8 (8.9)	14 (15.6)	33 (36.7)	34 (37.8)
No Disease		3 (33.3)	1 (11.1)	2 (22.2)	3 (33.3)
<b>Magnevist</b>	Pre-Contrast Diagnosis N(%)				
Disease		17 (17.4)	24 (24.5)	29 (29.6)	28 (28.6)
No Disease		2 (50.0)	1 (25.0)		1 (25.0)
	Pre- plus Post-Contrast Diagnosis N(%)				
Disease		13 (13.3)	15 (15.3)	37 (37.8)	33 (33.7)
No Disease		1 (25.0)	2 (50.0)		1 (25.0)

#### 4.0 Conclusions

The following points summarize this reviewer's concerns.

- Studies 488 and 525 may contain an over-representation of post-treatment patients. The over inclusion of this type of patient is enhancing the magnitude of the mean change in the primary endpoints from pre-image to image pair for both the Optimark and Magnevist groups. (See Appendix A for details.)
- In most cases, there are statistically significant increases in the mean and median of the primary endpoints from pre-image to image pair for both Magnevist and Optimark. (See Section 3.2.1 for details.)
- In all but one case, the confidence interval for the mean difference between the change in primary endpoints using Optimark and the same such change using Magnevist was completely contained within -1.5 to +1.5. These results indicate that the effects of Magnevist and Optimark (in terms of the primary endpoints) are similar. (See Section 3.2.2 for details.)
- Equivalence trials are inherently difficult to interpret due in part to the fact that poor designs, ill-defined endpoints, etc. may make it impossible to show a difference between treatment groups even if one exists. i.e., Unlike in traditional superiority trials, conducting a sloppy trial can be an asset in an "equivalence" trial.

The following excerpt is taken from the sponsor's proposed label and pertains to the efficacy of Optimark as the sponsor felt was demonstrated in the pivotal trials.

If the caveats discussed in this review (possible over-enrollment of post-treatment patients and the sponsor's definition of the equivalence region) are acceptable to the FDA medical division then it is the opinion of this reviewer that the efficacy of Optimark and its non-inferiority to Magnevist has been demonstrated in the sense described by the above excerpt. In addition, it is the opinion of this reviewer that if Optimark is approved the following points should be conveyed in the label:

- The post-contrast image was not evaluated alone. The label should reflect the need for adjunctive use.
- For a small group of patients, the pre-contrast images yielded primary endpoint scores higher than those of the pre- and post-contrast pair images. The label should reveal this fact in an attempt to convince users of Optimark not to automatically ignore the results of the pre-image when it appears to be different from the post-contrast image.
- Using the term *equivalent* in describing the relationship between Optimark and Magnevist as is done in the sponsor's proposed label is not recommended. Since only certain endpoints were studied, under certain conditions, it is possible that Optimark is not absolutely equivalent to Magnevist in all circumstances. Using a phrase such as, *for the endpoints studied, Optimark Injection was shown to be similar to Magnevist*, would be preferable.

*/S/* 11/6/98  
Ruthanna C. Davi  
Mathematical Statistician, HFD-720

Concur:

*/S/* 11/6/98  
Michael Welch, Ph.D.  
Acting Division Director

*/S/* 11/4/98  
Mahboob Sobhan, Ph.D.  
Acting Team Leader

cc:

Archival NDA#20-937  
HFD-701/C. Anello  
HFD-160/P. Love/E. Jones/R. Raman/B. Yaes  
HFD-160/R. Leedham/J. Moore  
HFD-160/File Copy  
HFD-344/B. Barton  
HFD-715/E. Nevius/M. Welch/M. Sobhan/R. Davi  
HFD-715/File Copy/Chron. Copy

R. Davi/x73122/Word/9/21/98

This review contains 33 pages of text, tables, and figures.

**6.0 Appendix A**

This appendix explores the possibility that the over-representation of post-treatment patients (including post-surgical, post-biopsy, post-radiation, post-chemotherapy) has over-influenced the efficacy results seen in studies 488 and 525.

Patients were divided into two groups, post-treatment patients and all other patients, according to the judgement of the FDA medical reviewer. The following tables show the mean change (and t-test p-value testing whether the mean change is different from zero) in each of the primary endpoints from pre-image to image pair within the post-treatment group as well as within the non-post-treatment group. P-values <0.05 have been shaded.

**Table 8: Study 488 – Mean Change in Primary Endpoints by Post-treatment Patient Grouping**

Endpoint	Patient Status	Treatment Group	Mean Difference (pair-pre)	Standard Deviation of Differences	t-test p-value
Conspicuity	Post-Treatment	Optimark n=55	0.5273	2.1418	0.0734
		Magnevist n=28	1.7857	1.9693	0.0001
	Non-Post-Treatment	Optimark n=77	0.2987	2.1464	0.2258
		Magnevist n=40	0.5500	2.4802	0.1687
Border Delineation	Post-Treatment	Optimark n=55	1.1636	2.2752	0.0004
		Magnevist n=28	1.0714	2.3401	0.0224
	Non-Post-Treatment	Optimark n=77	0.3766	2.2422	0.1446
		Magnevist n=40	0.8000	2.3772	0.0397
Diagnostic Confidence	Post-Treatment	Optimark n=55	0.6545	2.1707	0.0295
		Magnevist n=28	0.8571	1.8402	0.0204
	Non-Post-Treatment	Optimark n=77	0.2208	1.6593	0.2466
		Magnevist n=40	-0.0500	1.3950	0.8219

**Table 9: Study 525 – Mean Change in Primary Endpoints by Post-treatment Patient Grouping**

Endpoint	Patient Status	Treatment Group	Mean Difference (pair-pre)	Standard Deviation of Differences	t-test p-value
Conspicuity	Post-Treatment	Optimark n=33	0.9091	2.8103	0.0723
		Magnevist n=16	1.6250	2.5528	0.0224
	Non-Post-Treatment	Optimark n=33	0.5729	3.3865	0.1007
		Magnevist n=16	0.6531	3.5092	0.1989
Border Delineation	Post-Treatment	Optimark n=33	1.1818	3.4951	0.0609
		Magnevist n=16	2.1875	0.7372	0.0096
	Non-Post-Treatment	Optimark n=33	0.7604	3.4205	0.0319
		Magnevist n=16	1.0000	3.8568	0.0758
Diagnostic Confidence	Post-Treatment	Optimark n=33	0.9394	2.8167	0.0644
		Magnevist n=16	1.3750	0.5836	0.0325
	Non-Post-Treatment	Optimark n=33	0.6250	2.4804	0.0153
		Magnevist n=16	0.7755	1.7472	0.0032

From the mean changes displayed in Tables 8 and 9, it appears that although in most cases, there is an increase in scores on the image pair compared the pre-image, the magnitude of this increase is larger for post-treatment patients than for non-post-treatment patients. For that reason when considering the overall results of these studies one should keep in mind that the magnitude of the changes may be enhanced by over-inclusion of post-treatment patients. It should be noted that this analysis is a post-hoc exploratory analysis and that reaching statistically significant changes within each of these groups was not the intent of the studies.

**7.0 Appendix B**

**Study 488 - CNS Study**

**Conspicuity Score Recorded by Blinded Readers - Optimark (n=132)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre-Image	6.705	
Image Pair	7.098	
Difference*	0.3939	p=0.0363 (Paired t-test**) p=0.0204 (Wilcoxon test***)

\*Difference = "Image Pair" score minus "Pre-Image" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
24/132=18.2%	69/132=52.4%	39/132=29.5%

**Conspicuity Score Recorded by Blinded Readers - Magnevist (n=68)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre	6.1912	
Pair	7.2500	
Difference*	1.0588	p=0.0004 (Paired t-test**) p=0.0001 (Wilcoxon test***)

\*Difference = "Pair" score minus "Pre" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
5/68=7.4%	37/68=54.4%	26/68=38.2%

**Cross Tabulation of Conspicuity Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Optimark (n = 132)**

Conspicuity Score - "Pre"	Conspicuity Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	22					2		1	1	
2										
3				1						1
4	2					1	1			
5					1	1			1	1
6			1	2	2	2	1	1	2	4
7					1	2	7		1	3
8	1				1		4	2	5	6
9						1		2	3	5
10			1		1	1		1	3	32

\*Empty cells are cells where the frequency is zero.

**Cross Tabulation of Conspicuity Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Magnevist (n = 68)**

Conspicuity Score - "Pre"	Conspicuity Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	13						1	1		1
2									1	
3										
4						1			1	
5					1					
6			1			1	1	2	2	2
7	1		1						1	2
8							2	2	1	2
9								2	3	
10										18

\*Empty cells are cells where the frequency is zero.

**Study 525 - CNS Study**

**Conspicuity Score Recorded by Blinded Readers - Optimark (n=129)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre-Image	6.0620	
Image Pair	6.7209	
Difference*	0.6589	p=0.0226 (Paired t-test**) p=0.0044 (Wilcoxon test***)

\*Difference = "Image Pair" score minus "Pre-Image" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
24/129=18.6%	52/129=40.3%	53/129=41.1%

**Conspicuity Score Recorded by Blinded Readers - Magnevist (n=65)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	Paired t-test p-value**
Pre	5.4923	
Pair	6.3846	
Difference*	0.8923	p=0.0333 (Paired t-test**) p=0.0086 (Wilcoxon test***)

\*Difference = "Pair" score minus "Pre" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
9/65=13.8%	27/65=41.5%	29/65=44.6%

**Cross Tabulation of Conspicuity Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Optimark (n=129)**

		Conspicuity Score - "Pair"									
		1	2	3	4	5	6	7	8	9	10
Conspicuity Score - "Pre"	1	28					1		1	3	4
	2										1
	3								1		1
	4							1			
	5						1		1		1
	6								1	2	
	7	2	1			2		1	1	6	5
	8	2				1	1	2	4	9	7
	9	1						1	2	4	6
	10	2				1			2	4	15

\*Empty cells are cells where the frequency is zero.

**Cross Tabulation of Conspicuity Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Magnevist (n=65)**

		Conspicuity Score - "Pair"										
		1	2	3	4	5	6	7	8	9	10	
Conspicuity Score - "Pre"	1	18								1	1	2
	2											1
	3			1								
	4											2
	5											
	6	1									1	
	7										2	1
	8										3	2
	9						1			1	4	3
	10	2						2	1	1	2	5
11	1									1	6	

\*Empty cells are cells where the frequency is zero.

Study 490 - Liver Study

**Conspicuity Score Recorded by Blinded Readers - Optimark (n=99)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre-Image	5.7273	
Image Pair	6.4950	
Difference*	0.7677	p=0.0027 (Paired t-test**) p=0.0038 (Wilcoxon test***)

- \*Difference = "Image Pair" score minus "Pre-Image" score
- \*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".
- \*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
21/99=21.2%	37/99=37.4%	41/99=41.4%

**Cross Tabulation of Conspicuity Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Optimark (n=99)**

Conspicuity Score - "Pre"	Conspicuity Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	11	2	1	2	1	2	2	1	3	1
2	1			1						
3	2	1		1			1			
4			1							
5	1						1			
6					1	2	1	2	1	
7				1			2	6	2	
8	1					1	2	10	5	1
9							2	2	8	3
10								1	4	4

\*Empty cells are cells where the frequency is zero.

**Conspicuity Score Recorded by Blinded Readers - Magnevist (n=94)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre	5.6489	
Pair	5.9681	
Difference*	0.3191	p=0.1837 (Paired t-test**) p=0.1478 (Wilcoxon test***)

- \*Difference = "Pair" score minus "Pre" score
- \*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".
- \*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
22/94=23.4%	41/94=43.6%	31/94=33.0%

**Cross Tabulation of Conspicuity Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Magnevist (n=94)**

Conspicuity Score - "Pre"	Conspicuity Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	21	1				1	2			1
2				1				2		
3	1	1		1				1		
4								1		
5						1		1		
6	1						1	2	1	1
7	1							4	5	
8							2	5	4	3
9	1							2	4	4
10					1				1	2

\*Empty cells are cells where the frequency is zero.

**Study 526 - Liver Study**

**Conspicuity Score Recorded by Blinded Readers - Optimark (n=100)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre-Image	7.4200	
Image Pair	8.1700	
Difference*	0.7500	p=0.0003 (Paired t-test**) p=0.0001 (Wilcoxon test***)

- \*Difference = "Image Pair" score minus "Pre-Image" score
- \*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".
- \*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
14/100=14.0%	50/100=50.0%	36/100=36.0%

**Cross Tabulation of Conspicuity Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Optimark (n=100)**

Conspicuity Score - "Pre"	Conspicuity Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	10						2		1	1
2										
3										
4	1						1	1	1	
5							1	1	1	
6							1	1	2	
7			1	2	1	1	1	1	1	
8							2	3	6	4
9								5	9	9
10									2	27

\*Empty cells are cells where the frequency is zero.

**Conspicuity Score Recorded by Blinded Readers - Magnevist (n=103)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre	7.5243	
Pair	8.3010	
Difference*	0.7767	p=0.0008 (Paired t-test**) p=0.0001 (Wilcoxon test***)

- \*Difference = "Pair" score minus "Pre" score
- \*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".
- \*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
13/103=12.6%	51/103=49.5%	39/103=37.9%

**Cross Tabulation of Conspicuity Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Magnevist (n=103)**

Conspicuity Score - "Pre"	Conspicuity Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	7			1	1		1	1	2	1
2									1	
3									1	
4	1									
5								2	1	1
6	1						1			1
7							1	1	6	2
8	1							4	1	5
9								1	2	9
10									2	33

\*Empty cells are cells where the frequency is zero.

Study 488 - CNS Study

**Border Delineation Score Recorded by Blinded Readers - Optimark (n=132)**

**Border Delineation Score Recorded by Blinded Readers - Magnevist (n=68)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre-Image	5.9015	
Image Pair	6.6061	
Difference*	0.7045	p=0.0005 (paired t-test**) p=0.0004 (Wilcoxon test***)

Image Set	Mean Observation	P-value
Pre	5.6765	
Pair	6.5882	
Difference*	0.9118	p=0.0021 (paired t-test**) p=0.0018 (Wilcoxon test***)

\*Difference = "Image Pair" score minus "Pre-Image" score

\*Difference = "Pair" score minus "Pre" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
23/132=17.4%	55/132=41.7%	54/132=40.9%

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
11/68=16.2%	28/68=41.2%	29/68=42.6%

**Cross Tabulation of Border Delineation Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Optimark (n=132)**

**Cross Tabulation of Border Delineation Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Magnevist (n=68)**

Border Delineation Score - "Pre"	Border Delineation Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	22				1			3		
2		1				1	1			
3			1	1			1	1		1
4	1			1		2	2	1	1	1
5					1	2	1			2
6	1		2	2	4		4	4	2	1
7	1					2	1	4	3	3
8		1			1	3		5	1	4
9							2	1	3	6
10								1	1	20

Border Delineation Score - "Pre"	Border Delineation Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	13						1	1		1
2		1						1		
3			1					1		
4				1					1	2
5			1	2	1	1				1
6				1			1	2	2	
7								3		
8						1	1	4	3	4
9							1	1	1	2
10								1	1	8

\*Empty cells are cells where the frequency is zero.

\*Empty cells are cells where the frequency is zero.

Study 525 - CNS Study

**Border Delineation Score Recorded by Blinded Readers - Optimark (n=129)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre-Image	5.6589	
Image Pair	6.5271	
Difference*	0.8682	p=0.0047 (Paired t-test**) p=0.0011 (Wilcoxon test***)

\*Difference = "Image Pair" score minus "Pre-Image" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
25/129=19.4%	51/129=39.5%	53/129=41.1%

**Cross Tabulation of Border Delineation Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Optimark (n=129)**

Border Delineation Score - "Pre"	Border Delineation Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	28					1		1	3	4
2									2	
3							1	2		
4								1		
5		1			2	1		1	1	4
6					1	1			4	2
7	2	1			2		1	2	5	3
8	1				1		2	7	5	6
9	1						1		4	4
10	3							3	6	8

\*Empty cells are cells where the frequency is zero.

**Border Delineation Score Recorded by Blinded Readers - Magnevist (n=65)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre	5.0615	
Pair	6.3538	
Difference*	1.2923	p=0.0060 (Paired t-test**) p=0.0017 (Wilcoxon test***)

\*Difference = "Pair" score minus "Pre" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
7/65=10.8%	29/65=44.6%	29/65=44.6%

**Cross Tabulation of Border Delineation Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Magnevist (n=65)**

Border Delineation Score - "Pre"	Border Delineation Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	18									4
2			1						1	2
3										2
4										
5									2	
6	1				1			1	4	1
7							1		2	1
8	1							3	1	4
9							2			3
10	2									7

\*Empty cells are cells where the frequency is zero.

Study 490 - Liver Study

**Border Delineation Score Recorded by Blinded Readers - Optimark (n=99)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre-Image	5.3838	
Image Pair	6.1515	
Difference*	0.7677	p=0.0033 (Paired t-test**) p=0.0019 (Wilcoxon test***)

- \*Difference = "Image Pair" score minus "Pre-Image" score
- \*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".
- \*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
21/99=21.2%	38/99=38.4%	40/99=40.4%

**Cross Tabulation of Border Delineation Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Optimark (n=99)**

Border Delineation Score - "Pre"	Border Delineation Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	13	2	3	1		2	1		3	2
2	1	1		1						
3			2			1		1		
4		1	1		1			1		
5	1		1				1	1		
6					1	2	2		1	
7						2	10	7	1	
8					2		2	3	3	2
9	1							2	4	3
10									6	3

\*Empty cells are cells where the frequency is zero.

**Border Delineation Score Recorded by Blinded Readers - Magnevist (n=94)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre	5.3617	
Pair	5.6383	
Difference*	0.2766	p=0.2283 (Paired t-test**) p=0.1242 (Wilcoxon test***)

- \*Difference = "Pair" score minus "Pre" score
- \*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".
- \*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
20/94=21.3%	43/94=45.7%	31/94=33.0%

**Cross Tabulation of Border Delineation Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Magnevist (n=94)**

Border Delineation Score - "Pre"	Border Delineation Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	21	2								1
2	1	1	1	1			1	1		
3			1					1		
4	1		1				1	1	1	
5							1			
6				1			3	4	1	
7					1	2	6	5	2	
8							3	4	1	2
9	1						1	2	4	3
10									3	4

\*Empty cells are cells where the frequency is zero.

Study 526 - Liver Study

**Border Delineation Score Recorded by Blinded Readers - Optimark (n=100)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre-Image	7.2100	
Image Pair	7.9000	
Difference*	0.6900	p=0.0005 (Paired t-test**) ✓ p=0.0002 (Wilcoxon test***)

\*Difference = "Image Pair" score minus "Pre-Image" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
15/100=15.0%	45/100=45.0%	40/100=40.0%

**Border Delineation Score Recorded by Blinded Readers - Magnevist (n=103)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre	7.0874	
Pair	7.9417	
Difference*	0.8544	p=0.0003 (Paired t-test**) ✓ p=0.0002 (Wilcoxon test***)

\*Difference = "Pair" score minus "Pre" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
14/103=13.6%	50/103=48.5%	39/103=37.9%

**Cross Tabulation of Border Delineation Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Optimark (n =100)**

Border Delineation Score - "Pre"	Conspicuity Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	10					1	1		1	1
2	1									
3							1			
4					1				1	
5					1	2		1	1	1
6							1	1		
7		1			2			4	2	2
8						1	5	4	7	4
9								3	3	7
10									2	27

\*Empty cells are cells where the frequency is zero.

**Cross Tabulation of Border Delineation Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Magnevist (n=103)**

Border Delineation Score - "Pre"	Border Delineation Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	7					1	1		1	2
2										
3										
4										
5										
6										
7										
8										
9										
10										

\*Empty cells are cells where the frequency is zero.

Study 488 - CNS Study

Diagnostic Confidence Score Recorded by Blinded Readers - Optimark (n=132)

Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)

Image Set	Mean Observation	P-value
Pre-Image	7.7955	
Image Pair	8.1970	
Difference*	0.4015	p=0.0162 (Paired t-test**) p=0.0047 (Wilcoxon test***)

\*Difference = "Image Pair" score minus "Pre-Image" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

Summary of Proportions (FDA Reviewer's Analysis)

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
28/132=21.2%	51/132=38.6%	53/132=40.2%

Cross Tabulation of Diagnostic Confidence Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Optimark (n =132)

Diagnostic Confidence Score - "Pre"	Diagnostic Confidence Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	1									
2										
3										
4				1		1	1			
5					1					
6		1			2	1	4	1	3	4
7			1			3	6	11	2	4
8			1		1	1	5	10	6	7
9						1	0	3	7	2
10			1			1	2	1	4	25

\*Empty cells are cells where the frequency is zero.

Diagnostic Confidence Score Recorded by Blinded Readers - Magnevist (n=68)

Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)

Image Set	Mean Observation	P-value
Pre	8.0441	
Pair	8.3676	
Difference*	0.3235	p=0.1091 (Paired t-test**) p=0.1459 (Wilcoxon test***)

\*Difference = "Pair" score minus "Pre" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

Summary of Proportions (FDA Reviewer's Analysis)

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
18/68=26.5%	26/68=38.2%	24/68=35.3%

Cross Tabulation of Diagnostic Confidence Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Magnevist (n=68)

Diagnostic Confidence Score - "Pre"	Diagnostic Confidence Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

\*Empty cells are cells where the frequency is zero.

**Study 525 - CNS Study**

**Diagnostic Confidence Score Recorded by Blinded Readers - Optimark (n=129)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre-Image	7.6201	
Image Pair	8.3256	
Difference*	0.7054	p=0.0022 (Paired t-test**) p=0.0006 (Wilcoxon test***)

- \*Difference = "Image Pair" score minus "Pre-Image" score
- \*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".
- \*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
25/129=19.4%	40/129=31.0%	64/129=49.6%

**Cross Tabulation of Diagnostic Confidence Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Optimark (n=129)**

		Diagnostic Confidence Score - "Pair"									
		1	2	3	4	5	6	7	8	9	10
Diagnostic Confidence Score - "Pre"	1	3						1		5	
	2										
	3							1			
	4									1	
	5	3				3	1	2		3	1
	6			1			1	2			
	7				1		1	2	6	5	
	8				1		3	4	12	11	
	9	1						1	13	10	
	10	1					1	1	7	4	16

\*Empty cells are cells where the frequency is zero.

**Diagnostic Confidence Score Recorded by Blinded Readers - Magnevist (n=65)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre	7.6923	
Pair	8.6154	
Difference*	0.9231	p=0.0002 (Paired t-test**) p=0.0001 (Wilcoxon test***)

- \*Difference = "Pair" score minus "Pre" score
- \*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".
- \*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
8/65=12.3%	20/65=30.8%	37/65=56.9%

**Cross Tabulation of Diagnostic Confidence Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Magnevist (n=65)**

		Diagnostic Confidence Score - "Pair"									
		1	2	3	4	5	6	7	8	9	10
Diagnostic Confidence Score - "Pre"	1	1				1					
	2										
	3									2	
	4										
	5					1				1	1
	6							2	1	2	
	7							1	2	2	3
	8							1	6	4	10
	9								1	3	5
	10				1	1	1		2	1	8

\*Empty cells are cells where the frequency is zero.

Study 490 - Liver Study

Diagnostic Confidence Score Recorded by Blinded Readers - Optimark (n=99)

Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)

Image Set	Mean Observation	P-value
Pre-Image	6.1010	
Image Pair	7.5657	
Difference*	1.4646	p=0.0001 (Paired t-test**) p=0.0001 (Wilcoxon test***)

\*Difference = "Image Pair" score minus "Pre-Image" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

Summary of Proportions (FDA Reviewer's Analysis)

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
16/99=16.2%	14/99=14.1%	69/99=69.7%

Cross Tabulation of Diagnostic Confidence Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Optimark (n=99)

Diagnostic Confidence Score - "Pre"	Diagnostic Confidence Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	1		1				2			1
2				1	1					
3			1		1			1		
4	2									
5	1	1	1	3	2	2	3	2	8	1
6	1		1			2	1	2	4	2
7							2	4	9	8
8						1	1	5	9	5
9										
10						1		1	1	1

\*Empty cells are cells where the frequency is zero.

Diagnostic Confidence Score Recorded by Blinded Readers - Magnevist (n=94)

Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)

Image Set	Mean Observation	P-value
Pre	6.5000	
Pair	7.9149	
Difference*	1.4149	p=0.0001 (Paired t-test**) p=0.0001 (Wilcoxon test***)

\*Difference = "Pair" score minus "Pre" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

Summary of Proportions (FDA Reviewer's Analysis)

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
12/94=12.8%	19/94=20.2%	63/94=67.0%

Cross Tabulation of Diagnostic Confidence Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Magnevist (n=94)

Diagnostic Confidence Score - "Pre"	Diagnostic Confidence Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	1							2	1	
2					1			1		
3								2	1	
4				1	1	1				1
5	1			1	1	1	2	2	1	1
6					2	1			7	2
7	1						2	5	7	6
8			1		1			4	5	7
9										5
10										

\*Empty cells are cells where the frequency is zero.

**Study 526 - Liver Study**

**Diagnostic Confidence Score Recorded by Blinded Readers - Optimark (n=100)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre-Image	6.9300	
Image Pair	8.1900	
Difference*	1.2600	p=0.0001 (Paired t-test**) p=0.0001 (Wilcoxon test***)

- \*Difference = "Image Pair" score minus "Pre-Image" score
- \*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".
- \*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
17/100=17.0%	19/100=19.0%	64/100=64.0%

**Cross Tabulation of Diagnostic Confidence Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Optimark (n=100)**

Diagnostic Confidence Score - "Pre"	Diagnostic Confidence Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1	1				1			2	1	
2						1				
3		1	1				1			
4				1	2	1	1	2	2	
5				1	1	1		1	1	5
6				1			2		4	
7	1				1	1	1	1	4	4
8				1		1		3	9	7
9						2	1	2	4	11
10								2	2	7

\*Empty cells are cells where the frequency is zero.

**Diagnostic Confidence Score Recorded by Blinded Readers - Magnevist (n=103)**

**Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre	6.9029	
Pair	7.9223	
Difference*	1.0194	p=0.0001 (Paired t-test**) p=0.0001 (Wilcoxon test***)

- \*Difference = "Pair" score minus "Pre" score
- \*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".
- \*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
18/103=17.5%	25/103=24.3%	60/103=58.3%

**Cross Tabulation of Diagnostic Confidence Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Magnevist (n=103)**

Diagnostic Confidence Score - "Pre"	Diagnostic Confidence Score - "Pair"									
	1	2	3	4	5	6	7	8	9	10
1								1	3	
2						1		1		
3			1		1					
4				1	2	2	1	2	2	
5	2				1	1	2	1	1	
6			1		1		2	1	3	1
7	1				2	1	3	2	4	3
8						3	1	8	7	7
9								2	5	9
10				1				1	2	6

\*Empty cells are cells where the frequency is zero.

Study 488 - CNS Study

Number of Lesions Score Recorded by Blinded Readers - Optimark (n=134)

Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)

Image Set	Mean Observation	P-value
Pre-Image	1.970	
Image Pair	2.097	
Difference*	0.127	p=0.4049 (Paired t-test**) p=0.0294 (Wilcoxon test***)

\*Difference = "Image Pair" score minus "Pre-Image" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

Summary of Proportions (FDA Reviewer's Analysis)

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
9/134=6.7%	103/134=76.9%	22/134=16.4%

Number of Lesions Score Recorded by Blinded Readers - Magnevist (n=68)

Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)

Image Set	Mean Observation	P-value
Pre	1.618	
Pair	1.544	
Difference*	-0.074	p=0.6631 (Paired t-test**) p=0.9546 (Wilcoxon test***)

\*Difference = "Pair" score minus "Pre" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

Summary of Proportions (FDA Reviewer's Analysis)

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
6/68=8.8%	56/68=82.4%	6/68=8.8%

Cross Tabulation of Number of Lesions Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Optimark (n=134)

Number of Lesions Score - "Pre"	Number of Lesions Score - "Pair"											
	0	1	2	3	4	5	6	7	8	9	10	>10
0	21	5		1								
1	2	53	8	1								
2		4	17	3		1						
3			1	1								2
4					1							
5												
6							1					
7												
8												
9												1
10												
>10	1		1									9

\*Empty cells are cells where the frequency is zero.

Cross Tabulation of Number of Lesions Score Recorded\* by Blinded Readers on "Pre" and "Pair" - Magnevist (n=68)

Number of Lesions Score - "Pre"	Number of Lesions Score - "Pair"											
	0	1	2	3	4	5	6	7	8	9	10	>10
0	13	3										
1	1	35		1					1			
2		3	2									
3			1	2								
4					1							
5												
6									1			
7												
8												
9												
10												
>10		1										3

\*Empty cells are cells where the frequency is zero.

Study 525 - CNS Study

Number of Lesions Score Recorded by Blinded Readers - Optimark (n=121)  
Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)

Image Set	Mean Observation	P-value
Pre-Image	3.165	
Image Pair	3.400	
Difference*	0.240	p=0.0925 (Paired t-test**) p=0.0226 (Wilcoxon test***)

\*Difference = "Image Pair" score minus "Pre-Image" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

Summary of Proportions (FDA Reviewer's Analysis)

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
14/121=11.6%	83/121=68.6%	24/121=19.8%

Cross Tabulation of Number of Lesions Score Recorded\*  
by Blinded Readers on "Pre" and "Pair" - Optimark (n =121)

Number of Lesions Score - "Pre"	Number of Lesions Score - "Pair"											
	0	1	2	3	4	5	6	7	8	9	10	>10
0	26	5	1									
1	3	26	4		1		1					
2		7	4	2	2		1					
3			4									
4	1			1	2		1					1
5					1		1					
6												1
7												2
8												1
9												
10												
>10			1									21

\*Empty cells are cells where the frequency is zero.

Number of Lesions Score Recorded by Blinded Readers - Magnevist (n=64)  
Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)

Image Set	Mean Observation	P-value
Pre	2.094	
Pair	2.109	
Difference*	0.016	p=0.9484 (Paired t-test**) p=0.5302 (Wilcoxon test***)

\*Difference = "Pair" score minus "Pre" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

Summary of Proportions (FDA Reviewer's Analysis)

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
7/64=10.9%	46/64=71.9%	11/64=17.2%

Cross Tabulation of Number of Lesions Score Recorded\*  
by Blinded Readers on "Pre" and "Pair" - Magnevist (n=64)

Number of Lesions Score - "Pre"	Number of Lesions Score - "Pair"											
	0	1	2	3	4	5	6	7	8	9	10	>10
0	17	2	2									
1	2	15	1				1					
2		1	3		2							
3				4								1
4					2	2						
5					1	2						
6	1											
7						1						
8												
9												
10												
>10												3

\*Empty cells are cells where the frequency is zero.

**Study 490 - Liver Study**

**Number of Lesions Score Recorded by Blinded Readers - Optimark (n=98)  
Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre-Image	2.459	
Image Pair	3.061	
Difference*	0.602	p=0.0041 (Paired t-test**) p=0.0015 (Wilcoxon test***)

- \*Difference = "Image Pair" score minus "Pre-Image" score
- \*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".
- \*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
13/98=13.3%	50/98=51.0%	35/98=35.7%

**Cross Tabulation of Number of Lesions Score Recorded\*  
by Blinded Readers on "Pre" and "Pair" - Optimark (n=98)**

Number of Lesions Score - "Pre"	Number of Lesions Score - "Pair"											
	0	1	2	3	4	5	6	7	8	9	10	>10
0	10	8	5									2
1	2	19	1	2								
2	1	3	12	3								
3			2	2	1			1				1
4		2		1	1	2	1					
5			1			3						
6				1								
7					2							
8						1						
9								1				
10										1		
>10												6

\*Empty cells are cells where the frequency is zero.

**Number of Lesions Score Recorded by Blinded Readers - Magnevist (n=93)  
Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre	2.538	
Pair	3.011	
Difference*	0.505	p=0.0144 (Paired t-test**) p=0.0377 (Wilcoxon test***)

- \*Difference = "Pair" score minus "Pre" score
- \*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".
- \*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
18/93=19.4%	46/93=49.5%	29/93=31.2%

**Cross Tabulation of Number of Lesions Score Recorded\*  
by Blinded Readers on "Pre" and "Pair" - Magnevist (n=93)**

Number of Lesions Score - "Pre"	Number of Lesions Score - "Pair"											
	0	1	2	3	4	5	6	7	8	9	10	>10
0	20	3										
1	4	7	4	2	1							
2	1	2	10	2	1		1					2
3		4	1		2			1				
4		1		3	2	2	2					
5					1			1	1	1		
6							2					2
7												
8												
9										1		1
10									1			
>10											1	3

\*Empty cells are cells where the frequency is zero.

**Study 526 - Liver Study**

**Number of Lesions Score Recorded by Blinded Readers - Optimark (n=99)  
Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre-Image	3.495	
Image Pair	3.869	
Difference*	0.373	p=0.0439 (Paired t-test**) p=0.0688 (Wilcoxon test***)

\*Difference = "Image Pair" score minus "Pre-Image" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
16/99=16.2%	57/99=57.6%	26/99=26.3%

**Cross Tabulation of Number of Lesions Score Recorded\*  
by Blinded Readers on "Pre" and "Pair" - Optimark (n=99)**

Number of Lesions Score - "Pre"	Number of Lesions Score - "Pair"											
	0	1	2	3	4	5	6	7	8	9	10	>10
0	10	2	1									
1	1	23	8	1								1
2		4	3	2	3		1					1
3			2	4	1	1			1			
4			1									
5				1	2							
6					3							
7				1								
8					1							1
9									1			
10												
>10												15

\*Empty cells are cells where the frequency is zero.

**Number of Lesions Score Recorded by Blinded Readers - Magnevist (n=103)  
Paired t-test and Wilcoxon Sign Rank Test (FDA Reviewer's Analysis)**

Image Set	Mean Observation	P-value
Pre	3.140	
Pair	3.631	
Difference*	0.495	p=0.0391 (Paired t-test**) p=0.0439 (Wilcoxon test***)

\*Difference = "Pair" score minus "Pre" score

\*\*Testing the hypothesis that the mean score for the "Image Pair" is significantly different from the mean score for the "Pre-Image".

\*\*\*Testing the hypothesis that the median score for the "Image Pair" is significantly different from the median score for the "Pre-Image".

**Summary of Proportions (FDA Reviewer's Analysis)**

Image Pair Score < Pre-Image Score	Image Pair Score = Pre-Image Score	Image Pair Score > Pre-Image Score
15/103=14.6%	65/103=63.1%	23/103=22.3%

**Cross Tabulation of Number of Lesions Score Recorded\*  
by Blinded Readers on "Pre" and "Pair" - Magnevist (n=103)**

Number of Lesions Score - "Pre"	Number of Lesions Score - "Pair"											
	0	1	2	3	4	5	6	7	8	9	10	>10
0	7	4	1	1								1
1	2	25	2	1								
2	1	6	14	1	2							
3			3	2					1			3
4			1	2	1	1						2
5					1							
6							1					
7												
8									1			
9												
10												
>10		1								1		13

\*Empty cells are cells where the frequency is zero.