

**CENTER FOR DRUG EVALUATION AND  
RESEARCH**

*APPLICATION NUMBER:*

**21-711**

**STATISTICAL REVIEW(S)**



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

## STATISTICAL REVIEW

### CLINICAL STUDIES

**NDA:** NDA 21711  
**Drug Name:** Vasovist Injection  
**Indication:** MRA Imaging of Aortoiliac Blood Vessels  
**Applicant:** EPIX Pharmaceuticals  
**Date(s):** Stamp: July 1 2008 PDUFA: Dec 31 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:** Barbara Stinson M.D. (Medical Reviewer)  
Alex Gorovetz, M.D. (Clinical Team Leader)  
**Project Managers:** James Moore

**Keywords:** Superiority, Non-Inferiority, Sensitivity, Specificity, Imaging

# Table of Contents

<b>U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES .....</b>	<b>1</b>
<b>FOOD AND DRUG ADMINISTRATION .....</b>	<b>1</b>
<b>STATISTICAL REVIEW.....</b>	<b>1</b>
<b>1. EXECUTIVE SUMMARY .....</b>	<b>3</b>
1.1 CONCLUSIONS AND RECOMMENDATIONS .....	4
1.2 BRIEF OVERVIEW OF CLINICAL STUDIES .....	5
<b>2. INTRODUCTION .....</b>	<b>7</b>
2.1 OVERVIEW.....	7
2.2 DATA SOURCES .....	9
<b>3. STATISTICAL EVALUATION .....</b>	<b>10</b>
3.1 EVALUATION OF EFFICACY .....	10
3.2 EVALUATION OF SAFETY .....	15
<b>4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS .....</b>	<b>16</b>
4.1 GENDER, RACE AND AGE .....	16
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS .....	17
<b>5. SUMMARY AND CONCLUSIONS .....</b>	<b>18</b>
5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE .....	18
5.2 CONCLUSIONS AND RECOMMENDATIONS .....	18
<b>SIGNATURES/DISTRIBUTION LIST (OPTIONAL).....</b>	<b>20</b>

## List of Tables

Table 1: Statistics for Worst Outcome Imputation .....	6
Table 2: Statistics for Half Credit Imputation.....	6
Table 3: Statistics on Subsets of Interpretables Only .....	6
Table 4: Reader-Averaged Summary of Original vesus New Statistics .....	8
Table 5: Demographics by Study.....	12
Table 6: Statistics On Country and Imaging Device by Study .....	12
Table 7: Original Read versus Re-Read Reader-Averaged Uninterpretable Rates .....	13
Table 8: Agreement Levels on Uninterpretables .....	13
Table 9: Comparisons of Basic Results For Worst Outcome Analyses ( Condition(1) ) .....	14
Table 10: Vasovist Sensitivities/Specificities on Uninterpretable Baseline Reads .....	14
Table 11: Vessel Level Statistics by Gender .....	16
Table 12: Vessel Level Statistics by Age Group .....	17

# 1. EXECUTIVE SUMMARY

## 1.1 Conclusions and Recommendations

EPIX originally submitted four, primary Phase III Diagnostic Imaging studies of Vasovist Enhanced Magnetic Resonance Angiography (MRA) for the evaluation of patients with known or suspected aortoiliac vascular disease (two studies), renal artery disease (one study), or pedal artery disease (one study). All four trials were multi-center, blinded read, single arm trials.

The primary efficacy objective in all four studies was the determination of the presence or absence of stenosis in the vessels under examination. Each study used Baseline MRA (unenhanced) as Comparator, Vasovist MRA as Test, and X-Ray Angiography (XRA) as the Standard of Reference, and each study provided three blinded readers. The Sponsor's original goal was to demonstrate improved Sensitivity and Specificity for Enhanced MRA over Baseline MRA. The Sponsor claimed that these objectives were met in the original submission. The Agency noted that the demonstration of Superiority was dependent on the high percentages of Baseline images judged to be uninterpretable. The Agency further noted there was no Baseline Read protocol in place in the submission to provide confidence that these uninterpretable levels were intrinsic to Baseline and not due to inadequate readings by the readers. Consequently, the Agency stipulated that a re-read of images by properly trained new readers would be necessary before Vasovist could be considered for Approval.

The current submission consists essentially of a blinded re-read of the original images from the combined aortoiliac studies by three newly trained readers. The Sponsor and the Agency reached agreement that a claim for Vasovist Efficacy would rest on the following paired conditions:

**Condition(1):** Two of the three new readers had to simultaneously achieve vessel-level Superiority for Sensitivity and Non-Inferiority for Specificity of Vasovist MRA over Unenhanced MRA on the reads of the combined aortoiliac studies where:

- (a): Superiority required that the lower limit of the 2-sided 95% CI for the Vasovist minus Unenhanced Sensitivity exceed zero.
- (b): Non-Inferiority required that the lower limit of the 2-sided 95% CI for the Vasovist minus Unenhanced Specificity exceed  $-.05$ .

In this scenario the previous scoring of uninterpretable vessels as Wrong Outcome was preserved.

**Condition(2):** For each reader for whom Condition(1) obtained, Vasovist vessel-level Sensitivity and Specificity had to statistically exceed chance ( $.50$ ) on the subset of vessels that reader had classified as uninterpretable.

Both of these criteria were met in the re-read of the combined aortoiliac studies. The recommendation is for Approval for an aortoiliac indication.

## 1.2 Brief Overview of Clinical Studies

EPIX originally submitted four, primary Phase III Diagnostic Imaging studies of Vasovist Enhanced Magnetic Resonance Angiography (MRA) for the evaluation of patients with known or suspected aortoiliac vascular disease (two studies), renal artery disease (one study), or pedal artery disease (one study). All four trials were multi-center, open-label, blinded read, single arm trials.

The primary efficacy objective in all four studies was the determination of the presence or absence of stenosis in the vessels under examination. Each study used Baseline MRA (unenhanced) as Comparator, Vasovist MRA as Test, and X-Ray Angiography (XRA) as the Standard of Reference, and each study provided three blinded readers. The Sponsor's original goal was to demonstrate improved Sensitivity and Specificity for Enhanced MRA over Baseline MRA. The Sponsor claimed that these objectives were met in the original submission. The Agency noted that the demonstration of Superiority was dependent on the high percentages of Baseline images judged to be uninterpretable. The Agency further noted there was no Baseline Read protocol in place in the submission to provide confidence that these uninterpretable levels were intrinsic to Baseline and not due to inadequate readings by the readers. Consequently, the Agency stipulated that a re-read of images by properly trained new readers would be necessary before Vasovist could be considered for Approval.

The current submission of NDA 21711 consists essentially of a blinded re-read of images from two previously submitted studies from the earlier NDA 21711 submission which the agency determined to have Approvable status. The reads for the current submission are new reads of the old images from the combined aortoiliac trials by three newly trained readers.

**Appears This Way  
On Original**

### 1.3 Statistical Issues and Findings

A long series of communications between the Agency and the Sponsor, beginning with the original submission and extending through to the final agreement on criteria for Efficacy in the re-reads, focused on the evaluations of Sensitivity and Specificity under various scenarios for the treatment of Uninterpretables. For completeness, an overview of the statistics under these various scenarios is presented below for the re-reads.

**First:** Uninterpretables are scored as Incorrect ( Worst Outcome)

**Next :** Uninterpretables are given half-credit

**Final :** Uninterpretables are removed

**Note:** \* = Superiority (Lower limit of CI is above zero) ;

# = Inferiority (Upper Limit of CI is below zero)

**Table 1: Statistics for Worst Outcome Imputation**

	Sensitivity				Specificity		
	Pre	Post	Diff		Pre	Post	Diff
RDR D	69%	89%	20%* ( 15%, 25%)	RDR D	71%	72%	1% (-3%, 5%)
RDR E	70%	82%	12%* (7%, 17%)	RDR E	73%	81%	8%* ( 4%, 12%)
RDR F	64%	79%	15%* (9%, 21%)	RDR F	85%	85%	0% ( -2%, 2%)

**Table 2: Statistics for Half Credit Imputation**

	Sensitivity				Specificity		
	Pre	Post	Diff		Pre	Post	Diff
RDR D	74%	90%	16%* ( 12%, 20%)	RDR D	77%	72%	-5%# (-7%, -3%)
RDR E	73%	83%	10%* (6%, 14%)	RDR E	79%	81%	2%* ( 1%, 3%)
RDR F	68%	81%	13%* (9%, 17%)	RDR F	88%	86%	-2%# ( -3%, -1%)

**Table 3: Statistics on Subsets of Interpretables Only**

	Sensitivity				Specificity		
	Pre	Post	Diff		Pre	Post	Diff
RDR D	78%	91%	13%* ( 9%, 17%)	RDR D	81%	73%	-8%# (-10%, -6%)
RDR E	75%	83%	8%* (4%, 12%)	RDR E	83%	81%	-2%# (-4%, 0%)
RDR F	70%	82%	12%* (8%, 16%)	RDR F	90%	87%	-3%# ( -4%, -2%)

## 2. INTRODUCTION

### 2.1 Overview

#### Background on Submissions

The Sponsor originally provided four, primary Phase III Diagnostic Imaging studies of Vasovist Enhanced Magnetic Resonance Angiography (MRA) for the evaluation of patients with known or suspected peripheral vascular disease (two studies), renal artery disease (one study), or pedal artery disease (one study). The primary efficacy objective in all four studies was the determination of the presence or absence of stenosis in the vessels under examination. Each study used Baseline MRA (unenhanced) as Comparator, and X-Ray Angiography (XRA) as the Standard of Reference, and each study provided three blinded readers. The Sponsor's original goal was to demonstrate improved Sensitivity and Specificity for Enhanced MRA over Baseline MRA. The Sponsor's criteria for improvement in diagnostic performance required that the lower bound of the two-sided 95% Confidence Interval for the difference in performance - Enhanced MRA minus Baseline MRA - exceed zero. The Sponsor did not state that these criteria be met by all readers. In the absence of a clear, protocol statement of minimal criteria for success, two sets of criteria were considered by the Agency:

**Strong Criteria:** At least two of the three readers must simultaneously demonstrate superiority on both statistics (Sensitivity and Specificity).

**Moderate Criteria:** At least two of the three readers must simultaneously demonstrate superiority on one statistic, non-inferiority on the other. Furthermore, if two studies investigate the same vessel group, then the statistic with diagnostic superiority should be the same in both these studies.

**Principal Results:** The Sponsor's results in three of the four studies (two peripheral, r satisfied the Strong Criteria, but only under conditions which the Agency found arguable, namely:

All four studies presented large percentages of uninterpretable vessels for Baseline image reads (ranging from 10% to 40% across studies), as contrasted with less than 2% for Vasovist Enhanced image reads. The Sponsor chose to impute incorrect diagnoses to uninterpretable images. This "Worst Outcome" imputation, coupled with the high rates of Baseline Uninterpretables, ensured successful performance for Enhanced MRA diagnoses over Baseline MRA diagnoses.

The high levels of Baseline Uninterpretables could, of course, be consistent with inherent limitations in Baseline MRA diagnostics, in which case Vasovist enhanced images would present a diagnostic advantage. But these rates could also be consistent with an underspecified Baseline imaging protocol and poor reader training on baseline images. Since none of the protocols included rigorous specifications for Baseline imaging procedures, the Agency concern was that Vasovist superiority could have been driven by less than optimal readings of Baseline images.

The Sponsor received an Approvable determination, with the recommendation that, at a minimum, new reads be carried out on the existing data by new readers operating in accordance with a rigorous reading protocol. The eventual agreement between the Agency and the Sponsor was that an Approval for Vasovist MRA could be based on successful vessel-level statistics obtained from a re-read of the combined aortoiliac studies. The agreed-upon paired conditions for successful demonstration of Efficacy were:

**Condition(1):** Two of the three new readers had to simultaneously achieve vessel-level Superiority for Sensitivity and Non-Inferiority for Specificity of Vasovist MRA over Unenhanced MRA on the reads of the combined aortoiliac studies where:

(a): Superiority required that the lower limit of the 2-sided 95% CI for the Vasovist minus Unenhanced Sensitivity exceed zero.

(b): Non-Inferiority required that the lower limit of the 2-sided 95% CI for the Vasovist minus Unenhanced Specificity exceed -.05.

In this scenario the previous scoring of uninterpretable vessels as Wrong Outcome was preserved.

**Condition(2):** For each reader for whom Condition(1) obtained, Vasovist vessel-level Sensitivity and Specificity had to statistically exceed chance (.50) on the subset of vessels that reader had classified as uninterpretable.

**Results:** The Sponsor has achieved these paired criteria for the combined aortoiliac studies. The details are presented in tables in Section 3. A broad overview, comparing the original results to the new results on a reader-averaged level, restricted to Condition(1), is given below:

**Table 4: Reader-Averaged Summary of Original versus New Statistics**

Original Reads		New Reads	
<i>Sensitivity Results</i>			
# Patients	225		212
# Vessels	383		353
# Baseline Uninterpretables	40 (10%)		28 (8%)
Vasovist Sensitivity	74%		83%
Baseline Sensitivity	56%		68%
Difference	18%		15%
<i>Specificity Results</i>			
# Patients	422		412
# Vessels	2427		2230
# Baseline Uninterpretables	265(11%)		202 (9%)
Vasovist Specificity	88%		79%
Baseline Specificity	77%		76%
Difference	11%		3%

**Comments:**

(1): There was only a marginal decrease in Unenhanced Uninterpretables from Old Reads to New Reads: Old = 10% New = 9%

(2): Both Vasovist and Unenhanced Sensitivities increased about 10% from Old Reads to New Reads

(3): Vasovist Specificities dropped about 10% from Old Reads to New Reads ; Unenhanced Specificities remained the same.

Thus: Vasovist Sensitivities preserved their Superiority to Unenhanced Sensitivities ( about 15% ) , while Vasovist versus Unenhanced Specificities dropped from Superiority ( 11% ) to Non-Inferiority (3%).

Thus, the rigorously trained new readers appear to:

(a): *Overcall stenoses*

(b): *Call Uninterpretables at about the same rate as the old readers*

**2.2 Data Sources**

\\FDSWA150\NONECTD\N21711\N\_000\2008-06-30

**Appears This Way  
On Original**

### 3. STATISTICAL EVALUATION

#### 3.1 Evaluation of Efficacy

##### Comments on the Vasovist Re-Read Results in the EPIX Pre-Meeting Package

EPIX submitted a pre-meeting package as background for the upcoming ( June 5 2008) Type C Meeting discussion of formatting issues for resubmission of NDA21711. The NDA21711 resubmission is dedicated to results from a Re-Read of image data from two previously submitted studies: MS-325-12 and MS-325-13. These two studies evaluated the diagnostic performance Vasovist enhanced MRA images versus unenhanced MRA images for the detection of vessel stenoses in patients suspected of aortoiliac disease. The original submission of NDA21711 provided results from four trials – the two aortoiliac trials , along with trials concentrated on pedal and renal vessel diagnoses. The latter two trials will not enter into consideration in the resubmission.

The Agency rendered an Approvable decision for NDA21711; the primary concern driving the Approvable was the imbalance in percentages of vessels declared uninterpretable on unenhanced reads when compared with enhanced reads. The levels of uninterpretability for unenhanced reads was often in the 20% range. Enhanced reads, by contrast, were judged uninterpretable in less than 2% of the cases. Since uninterpretable reads were defaulted to “Incorrect” for diagnoses, a potential bias favoring Vasovist reads could have been introduced into the statistics. The Sponsor claimed that the advantage of Vasovist enhanced MRA lay precisely in its capacity to provide successful diagnoses where unenhanced MRA could provide no diagnoses at all. The Agency remarked that the absence of reader consistency in declaring unenhanced images uninterpretable, along with the anomaly that unenhanced images declared uninterpretable by one reader were often successfully diagnosed by the other readers, presented the possibility that the readers were insufficiently trained in the reading of such images. It was therefore determined that reconsideration for approval of the NDA would require, at a minimum, a re-read of the aortoiliac trials by new, carefully trained readers.

The resubmission therefore consists of a new set of reads of the original images from the two aortoiliac trials, which are now combined into a single data set so as to constitute one study. No new patients were enrolled, and the angiographic results from the original Truth reads were retained. The indication attendant on success for the new reads remains open for discussion. The criteria for success, however, have been established. They are as follows:

For two of the three readers, three conditions must be simultaneously met:

On the full data set, and with uninterpretable images imputed as “incorrect”:

- (1): Vasovist MRA Sensitivity must be superior to unenhanced MRA Sensitivity
- (2): Vasovist MRA Specificity must be non-inferior to unenhanced MRA Specificity

On the reduced ( per reader) data set where unenhanced reads were uninterpretable:

(3) Vasovist Sensitivity and Specificity must statistically exceed chance ( 50% success)

In terms of the calculated statistics:

On the full data set: Let

$D_S$  = Vasovist Sensitivity minus Unenhanced Sensitivity

$D_{SP}$  = Vasovist Specificity minus Unenhanced Specificity

Then:

(1): The lower limit of the 95% CI for  $D_S$  exceeds zero

(2): The lower limit for the 95% CI for  $D_{SP}$  exceeds -5%

Next, let S and SP be Sensitivity and Specificity for Vasovist. Then:

On the reduced ( per reader) data set where unenhanced reads were uninterpretable:

(3) The lower limit of the 95% CI for both S and SP must statistically exceed 50%

### **Demographics**

The re-read was performed on the combined patient populations from Study#12 and Study#13. The demographics table below shows that the population characteristics did not differ much between these studies, other than for Race, where the difference is accounted for by the fact that Study#13 was confined to US/Canada and Germany, while Study#12 included patients from Columbia.

**Appears This Way  
On Original**

**Table 5: Demographics by Study**

	STUDY#12 ( N = 251 )		STUDY#13 ( N = 173 )	
<b>Gender</b>	<b>M</b>	178 (71%)	<b>M</b>	112 (65%)
	<b>F</b>	73 (29%)	<b>F</b>	61 (35%)
<b>Race</b>	<b>Caucasian</b>	184 (73%)	<b>Caucasian</b>	167 (97%)
	<b>Hispanic</b>	36 (14%)		
	<b>Black</b>	30 (13%)		
<b>Age Group</b>	<b>&lt; 65 yrs</b>	107 (43%)	<b>&lt; 65 yrs</b>	72 (42%)
	<b>&gt;= 65 yrs</b>	144 (57%)	<b>&gt;= 65 yrs</b>	101 (58%)
<b>Disease Level</b>	<b>No Vessels</b>	111 (44%)	<b>No Vessels</b>	88 (51%)
	<b>One Vessel</b>	72 (29%)	<b>One Vessel</b>	44 (25%)
	<b>&gt;= Two Vessels</b>	68 (27%)	<b>&gt;= Two Vessels</b>	41 (24%)

**Device and Country Distribution:** The table below provides a breakdown of population distribution for both country and imaging device. It should be noted that the device distribution was fairly consistent between the studies.

**Table 6: Statistics On Country and Imaging Device by Study**

	STUDY#12 ( N = 251 Patients )		STUDY#13 ( N = 173 Patients )	
<b>COUNTRY</b>	<b>US/Canada</b>	221 (88%)	<b>US/Canada</b>	112 (65%)
	<b>Columbia</b>	30 (12%)	<b>Germany</b>	32 (18%)
			<b>Other</b>	29 (17%)
<b>DEVICE</b>		112 (45%)		91 (53%)
		96 (38%)		75 (43%)
		35 (14%)		7 (4%)
		8 (3%)		

b(4)

**Uninterpretable Rates – Original versus Re-Read:** The table below presents the original versus new read uninterpretable rates averaged over readers.

**Table 7: Original Read versus Re-Read Reader-Averaged Uninterpretable Rates**

	Sensitivity		Specificity	
	# Vessels	% Uninterpretable	# Vessels	% Uninterpretable
Combined Original Reads	383	10%	2427	11%
Combined Re- Reads	353	8%	2230	9%

**Comment#1:** *There was a relative drop of about 2% in Uninterpretables from the original reads to the re-reads.*

**Comment#2:** *About 6% of the vessels were not read in the re-read. Moreover, in the re-read, there were some differences in numbers of vessels read, reader by reader.*

**Agreement Levels on Uninterpretables by Reader:**

One of the concerns with the original reads was the inconsistency among readers in assignments of uninterpretable status to a vessel. As the table below indicates, this inconsistency has not been removed with the new reads. The sample over which the table statistics are calculated consists, in each row, of those vessels which at least one of the three readers scored uninterpretable. The entries by column are the percentages among such vessels scored uninterpretable by one, two, or three readers respectively.

*The critical result implied by the table is that the agreement levels among readers on uninterpretables did not improve under the new reads. The reviewer infers that the category of uninterpretability is reader-dependent rather than an intrinsic feature of baseline images, and can, at best, be interpreted as an indication that some baseline images are difficult to read rather than impossible to read.*

**Table 8: Agreement Levels on Uninterpretables**

	Exactly One Reader	Two Readers	All Three Readers
Study#12 Reads	60%	26%	14%
Study#13 Reads	48%	18%	34%
New Reads	61%	25%	14%

**Principal Results (Condition(1)):** The Table below presents the basic results by Reader and Study.

**Table 9: Comparisons of Basic Results For Worst Outcome Analyses ( Condition(1) )**  
 ( \* indicates a successful statistic)

		Original Reads Study #12			Original Reads Study #13			New Combined Reads		
		A12	B12	C12	A13	B13	C13	D	E	F
Sens	#Patients	140	140	140	85	85	85	212	215	217
	# Vessels	237	237	237	146	146	146	353	360	410
	#Uninterpretables	(29) 12%	(48) 20%	(14) 6%	( 9 ) 6%	(16) 11%	( 6 ) 4%	(36) 10%	(22) 6%	(25) 6%
	Vasovist	.80	.73	.61	.83	.84	.71	.89	.82	.79
	Unenhanced	.62	.67	.42	.52	.60	.49	.69	.70	.64
	Difference	.18	.06	.19	.31	.24	.22	.20	.12	.15
	Lower 95% CI	.10*	.00*	.12*	.21*	.14*	.11*	.15*	.07*	.09*
Spec	#Patients	250	250	250	172	172	172	411	412	410
	# Vessels	1409	1409	1409	1018	1018	1018	2185	2232	2272
	#Uninterpretables	(126) 9%	(283) 20%	(72) 5%	(66) 6%	(144) 14%	(107) 11%	(239) 11%	(257) 12%	(109) 5%
	Vasovist	.84	.93	.95	.80	.83	.90	.72	.81	.85
	Unenhanced	.75	.85	.75	.71	.75	.78	.71	.73	.85
	Difference	.09	.08	.20	.09	.08	.12	.01	.08	.00
	Lower 95% CI	.05*	.05*	.16*	.03*	.04*	.07*	-.03*	.04*	-.02*

*Comment: Condition(1) was met for both measures by all readers.*

**Vasovist versus Unenhanced Results respecting Condition(2):**

Two of the three readers who demonstrated successful statistics under Condition(1) also had to provide better than chance statistics for Vasovist Sensitivity and Specificity on the images each had classified as Uninterpretable at Baseline. The results are provided in the table below. Note that in all cases the CI lower limit exceeds .50.

**Table 10: Vasovist Sensitivities/Specificities on Uninterpretable Baseline Reads**

Reader		#Patients	# Vessels	Point Estimate	95% CI
RDR D	Sensitivity	32	36	.97	(.93, 1.00)
	Specificity	118	239	.72	(.67, .76)
RDR E	Sensitivity	21	22	.91	(.79, 1.00)
	Specificity	97	257	.84	(.81, .88)
RDR F	Sensitivity	23	25	.72	(.54, .90)
	Specificity	65	109	.82	(.76, .88)

**Comment#1:** *The Sponsor calculated the CI's under the assumption that the vessel diagnoses were independent. This assumption is most likely non-controversial for sensitivity calculations, since there was, typically, about one uninterpretable stenosed vessel per patient. However, for specificity, there were, on the average, about two uninterpretable non-stenosed vessels per patient, so that the statistics should account for possible clustering effects in diagnoses. The reviewer therefore modified the Sponsor's CI's to reflect a worst case scenario for dependence in within-patient diagnoses in the specificity calculations. This approach increased the CI's, but not to an extent that affected the positive profile for Vasovist diagnoses.*

**Conclusions:** *The Sponsor has met the proposed criteria.*

### **3.2 Evaluation of Safety**

This section is not application to the material reviewed in this report.

**Appears This Way  
On Original**

#### 4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

##### 4.1 Gender, Race and Age

The vessel level performance statistics by gender are given in the following table. There is no treatment (vasovist) by gender interaction for any reader.

**Table 11: Vessel Level Statistics by Gender**

Male	Female
<b>RDR D</b>	<b>RDR D</b>
( N = 241 )	( N = 112 )
Baseline Sensitivity = 70%	Baseline Sensitivity = 65%
Vasovist Sensitivity = 90%	Vasovist Sensitivity = 86%
( N = 1502 )	( N = 683 )
Baseline Specificity = 72%	Baseline Specificity = 69%
Vasovist Specificity = 73%	Vasovist Specificity = 69%
<b>RDR E</b>	<b>RDR E</b>
( N = 244 )	( N = 116 )
Baseline Sensitivity = 68%	Baseline Sensitivity = 74%
Vasovist Sensitivity = 83%	Vasovist Sensitivity = 81%
( N = 1532 )	( N = 700 )
Baseline Specificity = 73%	Baseline Specificity = 73%
Vasovist Specificity = 83%	Vasovist Specificity = 77%
<b>RDR F</b>	<b>RDR F</b>
( N = 243 )	( N = 120 )
Baseline Sensitivity = 65%	Baseline Sensitivity = 64%
Vasovist Sensitivity = 83%	Vasovist Sensitivity = 73%
( N = 1575 )	( N = 697 )
Baseline Specificity = 85%	Baseline Specificity = 87%
Vasovist Specificity = 85%	Vasovist Specificity = 85%

Appears This Way  
On Original

The vessel level performance statistics for each of the two age-groups are given in the following table. There is no treatment (vasovist) by age-group interaction for any reader.

**Table 12: Vessel Level Statistics by Age Group**

<b>Age &lt; 65 Years</b>	<b>Age &gt;= 65 Years</b>
<b>RDR D</b>	<b>RDR D</b>
( N = 150 ) Baseline Sensitivity = 69% Vasovist Sensitivity = 88%	( N = 203 ) Baseline Sensitivity = 68% Vasovist Sensitivity = 90%
( N = 918 ) Baseline Specificity = 76% Vasovist Specificity = 71%	( N = 1267 ) Baseline Specificity = 68% Vasovist Specificity = 72%
<b>RDR E</b>	<b>RDR E</b>
( N = 155 ) Baseline Sensitivity = 70% Vasovist Sensitivity = 88%	( N = 205 ) Baseline Sensitivity = 71% Vasovist Sensitivity = 78%
( N = 920 ) Baseline Specificity = 76% Vasovist Specificity = 79%	( N = 1312 ) Baseline Specificity = 70% Vasovist Specificity = 82%
<b>RDR F</b>	<b>RDR F</b>
( N = 160 ) Baseline Sensitivity = 68% Vasovist Sensitivity = 82%	( N = 203 ) Baseline Sensitivity = 62% Vasovist Sensitivity = 77%
( N = 947 ) Baseline Specificity = 86% Vasovist Specificity = 84%	( N = 1325 ) Baseline Specificity = 84% Vasovist Specificity = 86%

#### 4.2 Other Special/Subgroup Populations

This section is not applicable to the materials reviewed in this report.

Appears This Way  
On Original

## 5. SUMMARY AND CONCLUSIONS

### 5.1 Statistical Issues and Collective Evidence

This section is not applicable to the materials reviewed in this report.

### 5.2 Conclusions and Recommendations

EPIX originally submitted four, primary Phase III Diagnostic Imaging studies of Vasovist Enhanced Magnetic Resonance Angiography (MRA) for the evaluation of patients with known or suspected aortoiliac vascular disease (two studies), renal artery disease (one study), or pedal artery disease (one study). All four trials were multi-center, open-label, blinded read, single arm trials.

The primary efficacy objective in all four studies was the determination of the presence or absence of stenosis in the vessels under examination. Each study used Baseline MRA (unenhanced) as Comparator, Vasovist MRA as Test, and X-Ray Angiography (XRA) as the Standard of Reference, and each study provided three blinded readers. The Sponsor's original goal was to demonstrate improved Sensitivity and Specificity for Enhanced MRA over Baseline MRA. The Sponsor claimed that these objectives were met in the original submission. The Agency noted that the demonstration of Superiority was dependent on the high percentages of Baseline images judged to be uninterpretable. The Agency further noted there was no Baseline Read protocol in place in the submission to provide confidence that these uninterpretable levels were intrinsic to Baseline and not due to inadequate readings by the readers. Consequently, the Agency stipulated that a re-read of images by properly trained new readers would be necessary before Vasovist could be considered for Approval.

The current submission consists essentially of a blinded re-read of the original images from the combined aortoiliac studies by three newly trained readers. The Sponsor and the Agency reached agreement that a claim for Vasovist Efficacy would rest on following paired conditions:

**Condition(1):** Two of the three new readers had to simultaneously achieve vessel-level Superiority for Sensitivity and Non-Inferiority for Specificity of Vasovist MRA over Unenhanced MRA on the reads of the combined aortoiliac studies where:

- (a): Superiority required that the lower limit of the 2-sided 95% CI for the Vasovist minus Unenhanced Sensitivity exceed zero.
- (b): Non-Inferiority required that the lower limit of the 2-sided 95% CI for the Vasovist minus Unenhanced Specificity exceed  $-0.05$ .

In this scenario the previous scoring of uninterpretable vessels as Wrong Outcome was preserved.

**Condition(2):** For each reader for whom Condition(1) obtained, Vasovist vessel-level Sensitivity and Specificity had to statistically exceed chance (.50) on the subset of vessels that reader had classified as uninterpretable.

Both of these criteria were met in the re-read of the combined aortoiliac studies. The recommendation is for Approval for an aortoiliac indication.

Appears This Way  
On Original

Appears This Way  
On Original

## **SIGNATURES/DISTRIBUTION LIST (Optional)**

An example of this optional documentation is as follows:

Primary Statistical Reviewer: A G Mucci Ph.D.  
Date: 12 -19 - 08

Concurring Reviewer(s):

Statistical Team Leader: Jyoti Zalkikar Ph. D.

Biometrics Deputy Division Director: Aloka Chakravarty Ph. D.

cc:

HFD-Number/Project Manager James Moore

HFD-Number/Medical Officer Barbara Stinson M.D.

HFD-Number/Medical Team Leader Alex Gorovetz M. D.

HFD-Number/Office of Biostatistics Associate Director Ram Tiwari Ph. D.

HFD-Number/Office of Biostatistics Deputy Director Edward Nevius Ph. D.

c:\NDA\statreview.doc

-----  
**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  
12/19/2008 02:34:38 PM  
BIOMETRICS

Jyoti Zalkikar  
12/19/2008 02:43:24 PM  
BIOMETRICS

Aloka Chakravarty  
12/19/2008 04:04:54 PM  
BIOMETRICS

P



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

## STATISTICAL REVIEW AND EVALUATION CLINICAL STUDIES

**NDA:** NDA 21711 ( Complete Response to Approvable Letter)  
**Drug Name:** Vasovist Injection (gadofosveset trisodium)  
**Indication:** MRA Imaging of Blood Vessels  
**Applicant:** EPIX Pharmaceuticals  
**Date(s):** Letter Date: May 23, 2005 ; PDUFA Date Nov 23, 2005  
**Review Priority:** Standard  
**Biometrics Division:** Division of Biometrics II  
**Statistical Reviewer:** A G Mucci, Ph.D.  
**Concurring Reviewer:** Mike Welch, Ph.D.  
**Medical Division:** Medical Imaging and Hematology Products  
**Clinical Team:** Melanie Blank, M.D. (Medical Reviewer)  
Libero Marzella, M.D. (Team Leader)  
**Project Manager:** James Moore

**Keywords:** Clinical Studies, NDA Review, Sensitivity, Specificity

**1.0 EXECUTIVE SUMMARY .....3**

    1.1 Conclusions and Recommendations .....3

    1.2 Overview .....3

**2.0 INTRODUCTION.....4**

    2.1 Summary of First Cycle Review .....4

    2.2 Summary of Current Submission ..... 5

    2.3 Data Sources.....6

**3.0 STATISTICAL EVALUATION.....6**

    3.1 Summary of FDA Approvable Letter.....6

    3.2 Summary of Complete Response to the Approvable Letter.....7

    3.3 Reviewer’s Comments on Issues Raised in the Complete Response.....9

    3.4 Reviewer’s Evaluation of the Resubmitted Analyses.....12

**4.0 CONCLUSIONS AND RECOMMENDATIONS.....18**

## 1.0 EXECUTIVE SUMMARY

### 1.1 Conclusions and Recommendations

The exploratory evidence that the Sponsor's Complete Response to the Agency Approvable Letter for NDA21711 presents for Superiority of Vasovist Enhanced Image Diagnostics over Baseline Image Diagnostics continues to be partially driven by the Imputation of Incorrect Diagnosis to Uninterpretable Images. Since Baseline Images present relatively high percentages of Uninterpretables, when compared to Enhanced Images, the Incorrect Diagnosis default could introduce a bias against Baseline if some non-negligible percentage of Baseline Uninterpretables are contingently rather than intrinsically Uninterpretable. Such a contingency could, for instance, arise through insufficient reader training or an underspecified uniform calibration of Baseline imaging. The Reviewer's exploratory analyses of the original data, partially developed from several of the Sponsor's new analyses of these data, suggest both that more expert readings could both reduce the percentages of Uninterpretables and also provide acceptable evidence for Vasovist Diagnostic Superiority over Baseline. Therefore the Agency's original recommendation for new reads and new studies should stand.

### 1.2 Overview

The Agency determined an Approvable status for the original submission of NDA#21711 and requested both new blinded reads of existing data and new studies. The critical problem informing the determination of the Approvable status was the asymmetry in percentages of uninterpretable reads – typically 10% to 20% for Baseline images versus 1% to 2% for Vasovist Images. Since the Sponsor's original analyses used a "Worst Outcome" imputation scheme, in which Uninterpretables were classified as misdiagnoses, this asymmetry, if due to correctable practices in machine calibration and Baseline image reading, and thereby not *intrinsic* to Baseline Imaging, could have introduced an element of bias that was critical for the favorable results for Vasovist over Baseline in the Sensitivity and Specificity statistics. This possibility was investigated by the statistical reviewer in the original review through analyses of the subsample of "Interpretables Only" images, where it was found that no clear evidence for improved Sensitivity and/or Specificity for Vasovist was shown in these trials. In lieu of superiority for both measures, a review criterion used by the reviewer was that for at least one of the two statistics – Sensitivity or Specificity - two of the three readers provide superiority for Enhanced over Baseline reads, and, simultaneously, *are* not statistically inferior in the other statistic. Moreover, in identically designed trials, the statistic providing superiority must be the same in both trials. The failure of the Sponsor's statistics on the subsample of Interpretables led, in part, to the request for the re-reads and new studies.

The Sponsor's "Complete Response to the Agency Approvable Letter" is a new submission which provides none of the Agency's requested new studies, but, rather, consists largely of new exploratory analyses of the original database. The Sponsor asserts that these new analyses of old data are, in themselves, confirmatory of Vasovist efficacy. These new analyses, though typically successful with respect to the Agency standard presented above, are exclusively exploratory and retrospective, and therefore unacceptable as evidence for an Efficacy indication; they are, however, suggestive of the possibility that Vasovist diagnostics would improve sufficiently upon Baseline diagnostics in the presence of reduced percentages of Baseline Uninterpretables, provided the population of vessels was suitably large. These exploratory analyses typically penalized as misdiagnoses only those images which most readers classified as Uninterpretable; on all other images for which some readers provided diagnoses the individual diagnoses were weighted and then applied to the Uninterpretable. The effect of these new scorings was to extend the previous "Interpretables Only" subsample to the full sample size, with default misdiagnoses now restricted only to those images where the majority of readers provided a classification of Uninterpretable. These procedures were extended and modified by the Agency statistical reviewer to produce an exploratory "Modified Majority" diagnosis, and the statistics on Sensitivity and Specificity for this rule were favorable for Vasovist over Baseline. The Agency therefore has reason to believe, from these exploratory analyses, that readers properly trained in the reading of Baseline images, (which are acquired according to a standard protocol), will classify fewer images as Uninterpretable and still present superior statistics for Vasovist over Baseline. However, as stated in the Approvable Letter, the evidence favorable to Vasovist must be acquired prospectively, and the appropriate prospective studies have not been provided.

## **2.0 INTRODUCTION**

### **2.1 Summary of First Cycle Review**

In the previous submission of NDA 21711 the Sponsor provided four, primary Phase III Diagnostic Imaging studies of Vasovist MRA for the evaluation of patients with known or suspected peripheral vascular disease: two aortoiliac studies, one renal artery study, and one pedal artery study. The *Primary Efficacy Objective* in all four studies was the determination of the presence or absence of vessel level stenosis. Each study used Baseline MRA as Comparator, and X-Ray Angiography (XRA) as the Standard of Reference, and each study provided three blinded readers. The Sponsor's presumed goal was to demonstrate improved Sensitivity and Specificity for Enhanced MRA over Baseline MRA, although no specific success criteria derived from combinations of individual blinded reader performance were provided. An acceptable criterion for the Agency is that the lower bound of the two-sided 95% Confidence Interval for the difference in both Sensitivity and Specificity performance for Enhanced MRA minus Baseline MRA should exceed zero for at least two readers. A reduced, albeit exploratory criterion is:

(\*): At least two of the three readers must demonstrate Superiority on one statistic, while the performance of the other statistic not be significantly inferior. Furthermore, if two studies investigate the same vessel group, then the statistic with Diagnostic Superiority should be the same in both these studies. This criterion was used by the reviewer to suggest alternative pathways for new study recommendations.

The Sponsor's results in three of the four studies (two peripheral, ———, met the requirement (\*), but only under the condition that Uninterpretables be imputed "Worst Outcome", that is, that they be classified as incorrectly diagnosed. The problem here was that all four studies presented asymmetrical distributions of Uninterpretables, with relatively large percentages of Uninterpretables at Baseline ( 15%-20%) and much smaller percentages of Uninterpretables under Vasovist enhancement (1%- 2%). The Agency was not convinced that baseline Uninterpretables were inherently so, and not due to either inadequate reader training in baseline image analysis or to the absence of a pre-specified uniform protocol for baseline image acquisition. The Agency review team therefore analyzed the Sensitivity and Specificity statistics restricted to Interpretables only. The results of this analysis did not meet the restricted two out of three requirement (\*) for any of the vessel groups under investigation. Consequently, in the Approvable Action Letter, the Agency stated:

b(4)

*"The Sponsor should conduct adequate and well-controlled studies that demonstrate superior efficacy for Vasovist in those vascular regions for which an indication is sought. In particular, for the Aortoiliac region, the two studies under review could be combined for a re-read using pre-specified criteria for selecting and interpreting Baseline images; if the results are positive, then one additional confirmatory study might suffice for an Aortoiliac indication. In order to address the problem of Uninterpretables, the new studies should specify a standardization of Baseline imaging procedures (including a provision for re-reads of Baseline Uninterpretables), and separate evaluations of dynamic and steady-state images."*

## **2.2 Summary of the Current Submission**

The current submission consisted of the Sponsor's Complete Response to the FDA Approvable Letter. In this Complete Response the Sponsor provides no new studies, no new data, and no re-reads of the existing data, as requested by the Agency. Instead, the Sponsor supplemented the original analyses with several new analyses of the original data, in particular Sensitivity and Specificity analyses using three new Imputation Schemes for Uninterpretables. These new schemes score Uninterpretables somewhere between 0 and 1, where correct diagnoses are given a score=1, incorrect diagnoses a score=0, and yield statistics which lie between the Sponsor's successful original "Worst Outcome" high end statistics, and the unsuccessful low end statistics achieved on the restricted subsample of Interpretables. The Sponsor claims that the Sensitivities and Specificities determined under all these schemes – the original Worst Outcome scheme, the Interpretables only scheme, and the three new Imputation schemes -demonstrate improved efficacy for Vasovist images over Baseline images. The statistical reviewer concurs that, using criterion (\*), improved efficacy obtained under the original Worst

Outcome scheme and under the three new schemes, <sup>1</sup> However, success under analyses employing new, retrospective, exploratory imputation schemes is not deemed acceptable as an alternative to the requested new reads and new studies. In fact, in all four Imputation schemes the asymmetry of the distribution of Uninterpretables continued to play a role that favored Vasovist, a role which would be legitimate only if the Uninterpretables were intrinsically so. The statistical reviewer therefore undertook an exploratory investigation of Uninterpretables in order to determine if there was evidence that uninterpretability was intrinsic to an image. The analyses consisted of two investigations: first, a determination of the agreement among readers on uninterpretability; next, a determination of each reader's performance on Baseline images other readers classified as Uninterpretable, but which he classified as Interpretable. The first analysis revealed that there was nothing near unanimity among readers in the classification of images as Uninterpretable; the second analysis revealed that each reader's diagnostic performance on images other readers classified as Uninterpretable was similar to his overall performance on Interpretables. These exploratory results suggested that the classification of images as Uninterpretable is too reader-specific and might possibly result from inadequate training of readers in the reading of Baseline images, and, therefore, that the Agency had a rationale for its recommendations for new reads and new studies.

b(4)

### 2.3 Data Sources

\\Cdsub1\N21711\N\_000\2003-12-12  
\\Cdsub1\N21711\N\_000\2005-05-23

## 3.0 STATISTICAL EVALUATION

### 3.1 Summary of FDA Approvable Letter

The FDA sent an Approvable Letter to EPIX for their NDA#21711 on January 12 2005. The critical concern which determined the Approvable status was:

The large percentages of Baseline images classified as Uninterpretable, and the possibility that these percentages were correlated with the absence of a standardized procedure for Baseline Imaging and inadequate training for the readers. Since the study protocols specified that Uninterpretables were to be classified as diagnostic failures, these percentages could have biased the diagnostic success rates toward Vasovist enhanced images. In fact, under this Imputation Scheme the Sensitivities and Specificities of Vasovist enhanced images were significantly superior to Baseline Sensitivities and Specificities in most of the studies.

#### Agency Review Cycle Response to the "Worst Outcome" Imputation Scheme

The Agency, during the review cycle, suggested that the Sponsor recalculate Sensitivity and Specificity for the subset of Interpretable images alone, that is, for the subset

consisting of images for which both Baseline and Vasovist images were classified as Uninterpretable. *Diagnostic Superiority for Vasovist over Baseline on this subset would be sufficient grounds for approval. As noted in the Approvable Letter, this data reduction procedure is commonly performed in Imaging studies.* The Sponsor complied with this request. The Sponsor's results for this reduced data set were verified by the FDA statistician; however, the conclusions the FDA drew from these results differed from the conclusions drawn by the Sponsor. EPIX determined that the results were consistent with a diagnostic advantage for Vasovist images over Baseline images; the FDA found no clear evidence of such success. Since the FDA reviewer and the Sponsor produced the same numbers, this disagreement rests not on the statistics themselves, but on differing criteria for determining success for these statistics. For instance, in the two identically designed Aortoiliac studies, the reviewer proposed a minimal requirement for success:

(a): In each study, and for at least two of the three blinded readers, Vasovist outperform Baseline for one of the measures – Sensitivity or Specificity – and not be statistically inferior for the other measure.

(b): The measure providing Superiority be the same in similarly designed studies.

The Sponsor, on the other hand, appears to have considered it sufficient that the two out of three criterion be met in each study without the additional constraint that the successful measure be the same for identically designed studies. Moreover, the Sponsor considered positive performance for Vasovist to have been established in terms of several other (secondary endpoint) measures.

The impasse described above was addressed through the following synopsis of FDA requests in the Approvable Letter:

The Sponsor should conduct adequate and well-controlled studies that demonstrate superior efficacy for Vasovist in those vascular regions for which an indication is sought. *In particular, for the Aortoiliac region, the two studies under review could be combined for a re-read using pre-specified criteria for selecting and interpreting Baseline images; if the results are positive, then one additional confirmatory study might suffice for an Aortoiliac indication.* In order to address the problem of Uninterpretables, the new studies should specify a standardization of Baseline imaging procedures (including a provision for re-reads of Baseline Uninterpretables), and separate evaluations of dynamic and steady-state images. Moreover, a Superiority margin, suggested at 10%, for Vasovist statistics over Baseline statistics should be set.

### **3.2 Summary of Complete Response to the Approvable Letter**

This section provides an overview of the major issues raised by the Sponsor in response to the Agency Approvable Letter, along with comments and responses by the FDA Statistical Reviewer. It should be understood that the pro and con statements below are not standard for reviews and are not intended as formal scientific evidence for or against

Approval of a product. The intention of the overview of this material here is to provide a context for the understanding of the more formal analyses provided and suggested in Section 3.4. Several of the comments in this overview reflect material covered above, but in the context of responses to issues raised by the Sponsor in the Complete Response.

The Sponsor's Complete Response to the FDA Approvable Letter addresses the concerns and requests listed above in the following manner:

(1): The Uninterpretability rates for Baseline images are consistent with expectations and provide a major rationale for utilizing Vasovist enhanced Imagings. That is, *high percentages of Uninterpretables are intrinsic to Baseline imagings, (references were provided), and the conversion of uninterpretable images into diagnosable images under Vasovist enhancement is an indicator of Efficacy.*

(2): Moreover, no uniform Baseline standardization protocol exists, so none was set; it was deemed more appropriate to allow the centers and investigators to determine their own Baseline settings, and these settings were consistent with the manufacturers' suggested settings for the various machines employed. Consequently, the pre-specified imputation scheme which assigned incorrect diagnoses to Uninterpretables was justified.

(3): *Moreover, even when, as under FDA suggestion, the analyses were confined to Interpretables alone, the results were indicative of Vasovist Efficacy, albeit not by the rigid standards considered appropriate in this instance by the FDA reviewers.* (The Sponsor cites several FDA approvals of diagnostic imaging agents that did not outperform their comparators but provided acceptable Benefit vs Risk ratios.) The particular type of success derived by the Sponsor for the "Interpretables Alone" analyses, in the case of the two Aortoiliac studies, for instance, consisted of the demonstration of Superiority for Specificity alone in one study, and Superiority of Sensitivity alone in the other.

(4): *The Sponsor provided no new studies, no new data, and no re-reads of the existing data.* Instead, the Sponsor supplemented the original analyses with several new analyses, and summaries of old analyses, including:

(4a): An ROC analysis in which Vasovist Sensitivities were compared to Baseline Sensitivities for various thresholds for stenosis.

(4b): An analysis of Diagnostic Accuracy rates (percent agreement) for Vasovist over the subset of Baseline Uninterpretables.

(4c): Sensitivity and Specificity analyses for five Imputation Schemes for Uninterpretables.

The FDA statistical reviewer has not examined (4a) in detail since ROC analyses were never specified as primary analyses in the original protocols. *However, (4b) and (4c), although never specified as measures originally, will be examined in this review in Section 3.4, since they shed light on the role of Uninterpretables and could point toward fruitful efficacy investigations in the FDA's proposed re-reads.*

Further highlights of the Sponsor's Complete Response, along with several comments by the Agency Statistical Reviewer, are presented below. As already indicated above, these comments should not be interpreted as intrinsic to the current review detailed in Section 3.4 below, but are, rather, intended to help contextualize the formal review within the framework of the Approvable letter and the Sponsor's Complete Response.

### **3.3 Reviewer's Comments on Issues Raised in the Complete Response**

#### **Sponsor's Comments on Criteria for Success on Sensitivity and Specificity**

(References to page numbers refer to the Complete Response document, Part I.)

The Sponsor states, on p3:

"The Agency ... has apparently concluded that improvements in both Sensitivity and Specificity must be shown, although a clinically meaningful improvement in either one can appropriately be the basis for approval."

The Sponsor then lists various cases ( p19) where approval was given for Imaging agents when there was improvement in only one measure, and calls attention to:

"FDA's recognition of trade-offs in performance between Sensitivity and Specificity."

These passages can serve as the Sponsor's rationale for their statement on p4:

"...our analyses show that MS-325 improved either Sensitivity or Specificity to a clinically meaningful degree after removing patients having Uninterpretable vessel segments, and on this basis alone warrants approval."

#### **Statistician's Response:**

Let's first note that there is no current formal FDA minimal requirement for Superiority in both measures; when prospectively established, non-inferiority designs may be acceptable. This reviewer established, a *post hoc* review criterion that could suggest further pathways for new studies; namely that *replicable Superiority* in one measure along with *replicable Non-Inferiority* in the other be shown. The difficulty here, in the case of the identically designed peripheral vessel studies ( MS-325-12 & MS-325-13, and with statistics restricted to Interpretables), is that the direction of the trade-off was not replicated: It was Sensitivity that improved in MS-325-13, while it was Specificity that improved in MS-325-12. It's difficult to gauge the meaning of such mixed results since

the studies then appear to be telling different stories. How is the product to be positioned if there isn't a replicable direction in successful diagnoses? The most one could conclude in such circumstances is that the Diagnostic was never worse for either measure, and, in an unpredictable fashion, better for one of them. This circumstance is consistent with demonstrating similarity in Diagnostic Accuracy, which would provide, in itself, no rationale for choosing Test over Comparator. This issue will be re-addressed in Section 3.4.

**Sponsor's Comments concerning Performance Levels:**

On p16 of the Complete Response, EPIX states:

"..FDA's approvable letter applied a different and unduly rigid framework to the evaluation of MS-325, one requiring 80% agreement with XRA and 10% improvements in both Sensitivity and Specificity over non-contrast MRA."

**Statistician's Response:**

(a): In their Statistical Analysis Plan for Study MS-325-13, Section 10.2 (p20), EPIX states: "based on the assumptions ... that expected sensitivities (specificities) for pre-contrast and post-contrast MRA are 70% and 85% respectively." Further, in their Statistical Analysis Plan For Study MS-325-12, Section 10.2 (p6), EPIX states: "based on the assumptions ... that expected sensitivities (specificities) for pre-contrast and post-contrast MRA are 70% and 80% respectively."

(b): In Volume#1, p2 of NDA#21711, EPIX states: "VASOVIST satisfies a significant unmet medical need for an *accurate* non-invasive diagnostic procedure for angiographic assessment ..... without the risks ... of catheter-based XRA."

(c): In a letter to Dr Houn at the FDA, dated Dec 12 2003, EPIX states on p4: "In summary, *Vasovist MRA has been shown to provide similar diagnostic results to XRA ...*"

(d): It would appear from (a), (b), and (c) that EPIX intended Vasovist to perform (sensitivity and specificity) at least at 80% to 85% levels. In addition, and in general, one would expect an *intended substitute* for XRA ( (b) and (c)) to perform at least at such levels. Moreover, the expected differences between pre-and-post statistics implied in the Sponsor's figures are 10% to 15%. Given these Sponsor expectations, both in levels of performance, differences in levels of performance, and the implied positioning for Vasovist MRA as a substitute for XRA, there doesn't appear to be anything extreme in the FDA's criteria for successful numbers for Vasovist when compared to the Sponsor's expected numbers. Thus, the Sponsor's objections to FDA expected levels of performance would appear to be objections to their own expectations.

**Further General Comments:**

There remains a very serious concern, and this concern involves the nature of Phase III submissions. Studies provided in NDA's submitted for review are not exploratory studies where one peruses the results and finds this or that favorable analysis. The Studies must contain hypotheses, the satisfaction of whose requirements are determinative of whether the studies rise or fall. The Sponsor states on p21 of the Complete Response that

“the FDA has not tried to specify in advance how much benefit a contrast agent must add compared to use of a device alone..”

*However, this does not imply that measures of benefit are not to be specified prospectively in each particular case. Failure to do so is failure to adequately define a prospective study.*

The statistical reviewer could find no clearly stated hypotheses in the submitted studies other than the following( see, for instance p9, Statistical Analysis Plan, MS-325-12, IND51,172, or Clinical Study Report for MS-325-13, p45):

Let  $p_1$ = Proportion of vessels correctly diagnosed post-contrast

Let  $p_2$ = Proportion of vessels correctly diagnosed pre-contrast

Then:

Null  $H_0$ :  $p_1=p_2= p$  vs Alternative  $H_A$ :  $p_1 \neq p_2$

First, these hypotheses are with respect to Accuracy, not Sensitivity and/or Specificity. Next, no statements are made, for instance, providing conditions on the performance of the three readers taken together which would serve as criteria that the Null Hypothesis had been rejected. Finally, the expected rejection of the Null Hypothesis would presumably point in the direction of Superiority, and it is such a direction that is captured in the powering calculations submitted by the Sponsor. It therefore does not appear to be exceptional that the Agency expect:

(1): Explicit Conditions - both on Reader combinations and on Sensitivity and Specificity combinations - for Rejection of Null hypotheses on Equality between the Test Diagnostic and the Comparator Diagnostic.

(2): Explicitly stated minimal performance levels consistent with the intended positioning of the Diagnostic.

### 3.4 Reviewer's Evaluation of the Resubmitted Analyses

The intention in this section is to provide exploratory evidence that:

(A): Uninterpretables are not a "Reader-Independent" class of Images; that is, the determination that an Image is Uninterpretable by Reader X provides limited implications about how Reader Y will classify the Image. In fact, when Reader Y classifies the Image as Interpretable, he is likely to diagnose the Image about as well as he would were Reader X to classify the image as Interpretable. This circumstance suggests the possibility that, to some extent, Uninterpretability is a "default" classification occasioned by poor training of readers.

(B): There are reasonable Imputation schemes and methods for combining reads that strongly suggest that a re-read of images in the Aortoiliac studies would result in a demonstration of Superiority for Vasovist images over Baseline Images. (It is to be understood that the criteria for success remain as stated earlier – the two out of three rule; the reasonable Imputation schemes explored below are intended as suggestive of results that a re-read could provide.)

To begin, the table below presents the diversity among readers when designating an Image as Uninterpretable.

**Table(3.4.1)**  
**Table of Conditional Probabilities for Uninterpretables**  
**( Probabilities are Conditional on the Occurrence of an Uninterpretable)**

STUDY	Reader			Number of Readers		
	A	B	C	One	Two	Three
#12	.61	.16	.77	.60	.26	.14
#13	.75	.48	.63	.48	.18	.34

b(4)

Note:

The Table lists statistics conditioned on the sample of all those Images for which at least one of the three Readers classified the Image as Uninterpretable. Keep in mind that different readers are used in each study.

A, B, C list the probabilities that respective Readers A,B,C classified an Image as Uninterpretable, given that at least one of them did.

"One" Means exactly One of Three Readers classified the Image as Uninterpretable, given that at least one of them did.

"Two" Means exactly Two of Three Readers classified the Image as Uninterpretable, given that at least one of them did.

“Three” Means exactly Three of Three Readers classified the Image as Uninterpretable, given that at least one of them did.

**Principal Comment Concerning Table 3.4.1:** *The lack of uniformity among readers in classifying Images as Uninterpretable suggests that Uninterpretability possesses a strong reader specific component that could have resulted from inadequate training of readers on Baseline Images.*

**Preliminary Comments Concerning Tables 3.4.1/3.4.2:** In order to further investigate if there is anything “intrinsic” to an Uninterpretable Image, the standard statistics (Sensitivity/Specificity/Accuracy) have been calculated with respect to subsamples of Images F for which at least one of the readers entered a classification of Uninterpretable. Thus:

Sensitivity for Reader A: Sensitivity was calculated for the subsample of F for which Reader A classified the image as Interpretable, while either Reader B or C ( or both) classified the Image as Uninterpretable. Likewise for Specificity and Accuracy.

The resulting statistics were then compared to the corresponding statistics calculated over the larger sample of all Interpretables for Reader A. Likewise for Reader B and Reader C.

Note:

NR = Sample size for the subsample for which the Reader registered the image as Interpretable while at least one other reader registered the Image as Uninterpretable.

N = Total sample size of Uninterpretables for the Reader

SR; SPR; AR = Sensitivity; Specificity; Accuracy over the subsample

S; SP; A = Sensitivity; Specificity; Accuracy over the larger sample of Interpretables

**Appears This Way  
On Original**

**Table(3.4.2)  
Conditional vs Unconditional Baseline Statistics**

STUDY#12														
	Sensitivity					Specificity					Accuracy			
	NR	SR	N	S		NR	SPR	N	SP		NR	AR	N	A
RDR A	29	.62	211	.70		126	.90	1196	.88		155	.85	1407	.86
RDR B	48	.81	230	.69		283	.78	1353	.88		331	.78	1583	.85
RDR C	14	.64	196	.51		72	.88	1142	.93		86	.84	1338	.87
STUDY#13														
	Sensitivity					Specificity					Accuracy			
	NR	SR	N	S		NR	SPR	N	S		NR	AR	N	S
RDR A	9	.78	126	.60		66	.80	811	.89		75	.80	937	.85
RDR B	16	.75	133	.66		144	.81	889	.85		160	.80	1022	.83
RDR C	6	.83	123	.58		107	.86	852	.93		113	.86	975	.90

b(4)

**Principal Comment Concerning the Table 3.4.2:** *If there were something intrinsic to Uninterpretables, then Images so classified by at least one Reader should at least provide diminished statistics for those Readers who, instead, entered diagnoses for these images. As the table indicates, this apparently was not the case.*

The two tables above are intended as exploratory evidence that at least some percentage of the reads classified as Uninterpretable were not intrinsically so. Consequently, it would be useful to "override" the effects of these classifications by extending the benefit of the doubt to Uninterpretables in some fashion that optimizes Baseline diagnoses in the presence of Uninterpretability. A combined diagnosis dedicated to this attempt is defined and explored below. It should be understood that this combined diagnosis was developed from the Reviewer's examination of the five Imputation schemes provided by the Sponsor in the Resubmission. The Reviewer believes:

(A): This combined diagnostic captures relatively well the Information provided by the Sponsor's several Imputation schemes in a more unified fashion.

(B): It provides exploratory evidence that a re-read could generate successful statistics for Vasovist.

**Preliminary Comments for Tables 3.4.3/3.4.4:** The standard diagnostic statistics using a classification rule based on the Reviewer's measure M (defined below) are presented for two distinct weighting schemes:

Patient-Weighted Statistics (Table 3.4.3): Sensitivity and Specificity are calculated on a vessel level for each patient individually and the results are then averaged over all patients.

Vessel-Weighted Statistics (Table 3.4.4): Sensitivity and Specificity are calculated at the vessel level.

**Note:** The Sponsor has employed both schemes: vessel-weighted statistics in the original submission and Patient-weighted statistics in the resubmission.

#### **Definition of the Reviewer's Measure M for Statistical Comparisons**

(Note: A score of 1 indicates success in diagnosis; a score of 0 indicates failure.)

Then the combined measure M is determined as follows:

If all three readers call the vessel Interpretable, then  $M = \text{Majority Score}$

If two readers score the vessel Interpretable, then  $M = 1$  only if both readers score 1

If exactly one reader scores the vessel Interpretable, then  $M = 1$  if that reader scores a 1

If all three readers score the vessel Uninterpretable, then  $M = 0$

Appears This Way  
On Original

**Table(3.4.3)  
Patient Weighted Statistics Using Reviewer's Majority Rule (M)**

Study	Sensitivity				Specificity				Accuracy			
	N	Pre	Post	Diff (95%CI)	N	Pre	Post	Diff (95%CI)	N	Pre	Post	Diff (95%CI)
12	138	.61	.73	.12 (.05, .19)	248	.87	.94	.07 (.04, .10)	249	.83	.91	.08 (.05, .11)
13	85	.57	.82	.25 (.15, .35)	171	.81	.85	.05* (0, .09)	172	.78	.86	.08 (.04, .12)

b(4)

**Table(3.4.4)  
Vessel Weighted Statistics Using Reviewer's Majority Rule (M)**

Study	Sensitivity				Specificity				Accuracy			
	N	Pre	Post	Diff (95%CI)	N	Pre	Post	Diff (95%CI)	N	Pre	Post	Diff (95%CI)
12	138	.62	.73	.11 (.04, .18)	248	.87	.94	.07 (.04, .10)	249	.83	.92	.09 (.06, .12)
13	85	.57	.82	.25 (.15, .35)	171	.81	.87	.06 (.01, .11)	172	.78	.86	.08 (.04, .12)

b(4)

**Principal Comments on Tables 3.4.3 and 3.4.4:**

(a): The results are essentially equivalent for the two sets of statistics. An asterisk (\*) denotes failure of the 95% confidence interval to exclude zero.

(b): The aortoiliac studies present Superiority for both Sensitivity and Specificity. The other studies present Superiority for Specificity.

(c): The tables suggest that a re-read which reduces the percentages of Uninterpretables could provide success for Vasovist Images.

**A Final Comparison:**

The measure M used to generate the two tables above was developed by the reviewer through examination of the Sponsor's new imputation measures. Table(3.4.5) below compares the statistics derived from one of these imputation schemes to the reviewer's statistics as derived from the measure M. The purpose of this comparison is to provide exploratory evidence that these measures have similar consequences.

**Sponsor's Imputation #4:**

If a Reader scores a vessel Uninterpretable, and the other two Readers score it Interpretable, then the Diagnostic score for the vessel for the Uninterpretable Read is the average of the Diagnostic scores for the Interpretables

If two Readers score a vessel Uninterpretable, then all three readers are given the Diagnostic score for the Interpretable

If all three readers score the vessel Uninterpretable, all three Readers are given the Incorrect (Worst Outcome) Diagnostic score.

**Table(3.4.5)  
Comparison of Sponsor's Imputation #4 and Reviewer's Majority (M) Rule  
(Vessel Weighted Statistics)**

		N	SPONSOR IMPUTATION #4		REVIEWER RULE M	
			Pre	Post	Pre	Post
STUDY#12	Sens	138	.62	.72	.62	.73
	Spec	248	.85	.91	.87	.94
	Acc	249	.82	.89	.83	.92
STUDY#13	Sens	85	.57	.80	.57	.82
	Spec	171	.80	.84	.81	.87
	Acc	172	.78	.85	.78	.86

b(4)

N = Number of Patients

Sponsor's statistics are averages over the three readers

#### 4.0 CONCLUSIONS AND RECOMMENDATIONS

The evidence that the original NDA21711 presented for Superiority of Vasovist Enhanced Image Diagnostics over Baseline Image Diagnostics was partially driven by the Imputation of Incorrect Diagnosis to Uninterpretable Images. Since Baseline Images in these studies presented relatively high percentages of Uninterpretables, the Incorrect Diagnosis default could have introduced a bias against Baseline if some number of Baseline Uninterpretables were contingently rather than intrinsically so classified; for instance, through insufficient reader training in Baseline Image reading or through an underspecified uniform calibration of Baseline imaging. These concerns drove, in part, the Approvable Letter in which new studies and new reads were requested. The Sponsor provided no new studies or reads, but, rather, only exploratory new analyses of the original reads of the original data. These Exploratory analyses suggest that more expert readings could provide contributive evidence for Vasovist Diagnostic Superiority over Baseline, but are insufficient in themselves for Approval. The Sponsor should, at an absolute minimum, provide re-reads of images by new readers so as to minimize the asymmetry in Uninterpretables between Baseline and Enhanced read classifications. The statistical analyses could then be more representative of Vasovist performance.

Appears This Way  
On Original

**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  
11/9/2005 04:37:13 PM  
BIOMETRICS

Mike Welch  
11/10/2005 11:45:08 AM  
BIOMETRICS  
Concur with review.



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

## STATISTICAL REVIEW AND EVALUATION CLINICAL STUDIES

**NDA\Serial Nr:** NDA 21711  
**Drug Name:** Vasovist Injection (gadofosveset trisodium)  
**Indication(s):** MRA Imaging of Blood Vessels  
**Applicant:** EPIX Pharmaceuticals  
**Date(s):** Letter Date: Dec 12, 2003 ; Ext. PDUFA Date Jan 15, 2005  
**Review Priority:** Standard  
**Biometrics Division:** Division of Biometrics II  
**Statistical Reviewer:** A G Mucci, Ph.D. (HFD-715)  
**Concurring Reviewer:** Mike Welch, Ph.D. (HFD-715)  
**Clinical Team:** Tong Li, M.D. (Medical Reviewer) HFD-160  
Zili Li, M.D. (Team Leader) HFD-160  
**Project Manager:** James Moore, CAPT., PHS

**Keywords:** Clinical Studies, NDA Review, Sensitivity, Specificity

<b>1. EXECUTIVE SUMMARY .....</b>	<b>3</b>
<b>1.1 Conclusions and Recommendations .....</b>	<b>3</b>
<b>1.2 Brief Overview of Clinical Studies .....</b>	<b>4</b>
<b>1.3 Statistical Issues and Findings .....</b>	<b>6</b>
<b>2. INTRODUCTION.....</b>	<b>12</b>
<b>2.1 Overview.....</b>	<b>12</b>
<b>2.2 Data Sources.....</b>	<b>18</b>
<b>3. STATISTICAL EVALUATION.....</b>	<b>19</b>
<b>3.1 Evaluation of Efficacy.....</b>	<b>19</b>
<b>3.1.1 Overview of Design and Evaluation of Efficacy for Study #12.....</b>	<b>20</b>
<b>3.1.2 Overview of Design and Evaluation of Efficacy for Study #13.....</b>	<b>26</b>
<b>3.1.3 Overview of Design and Evaluation of Efficacy for Study #14.....</b>	<b>31</b>
<b>3.1.4 Overview of Design and Evaluation of Efficacy for Study #15.....</b>	<b>37</b>
<b>3.1.5 Remarks on Uninterpretable Images Stratified by Center.....</b>	<b>42</b>
<b>3.2 Evaluation of Safety.....</b>	<b>44</b>
<b>4. FINDINGS IN SPECIAL POPULATIONS .....</b>	<b>44</b>
<b>5. SUMMARY AND CONCLUSIONS .....</b>	<b>44</b>
<b>APPENDIX .....</b>	<b>46</b>

## 1. EXECUTIVE SUMMARY

### 1.1 Conclusions and Recommendations

Principal Design: The Sponsor provided four, primary Phase III Diagnostic Imaging studies of Vasovist Enhanced Magnetic Resonance Angiography (MRA) for the evaluation of patients with known or suspected peripheral vascular disease (two studies), renal artery disease (one study), or pedal artery disease (one study). The primary efficacy objective in all four studies was the determination of the presence or absence of stenosis in the vessels under examination. Each study used Baseline MRA (unenhanced) as Comparator, and X-Ray Angiography (XRA) as the Standard of Reference, and each study provided three blinded readers. The Sponsor's goal was to demonstrate improved Sensitivity and Specificity for Enhanced MRA over Baseline MRA. The Sponsor's criteria for improvement in diagnostic performance required that the lower bound of the two-sided 95% Confidence Interval for the difference in performance - Enhanced MRA minus Baseline MRA - exceed zero. The Sponsor did not state that these criteria be met for all readers. In the absence of a clear, protocol statement of minimal criteria for improvement, two sets of criteria were investigated by the Reviewer:

Requirement #1: At least two of the three readers must demonstrate superiority on both statistics (Sensitivity and Specificity).

Requirement #2: At least two of the three readers must demonstrate superiority on one statistic, while the performance of the other statistic not be significantly inferior. Furthermore, if two studies investigate the same vessel group, then the statistic with diagnostic superiority should be the same in both these studies.

Principal Results: The Sponsor's results in three of the four studies (two peripheral, — — achieved Requirement #1, but only under conditions which the Agency found arguable. The problem was that all four studies presented large percentages of uninterpretable vessels for Baseline image reads (ranging from 10% to 40% across studies), as contrasted with less than 2% for Enhanced image reads. Such levels could be consistent with, for instance, (a) inherent limitations in Baseline MRA diagnostics, or (b) an underspecified Baseline imaging protocol. The Sponsor chose to impute incorrect diagnoses to uninterpretable images, a procedure that, if (a) obtains, equates an inherent "non diagnosis" to a wrong diagnosis. It is this imputation scheme, designated as Worst Outcome, which, given the high rate of uninterpretables, ensures successful performance for Enhanced MRA diagnoses over Baseline MRA diagnoses. Since none of the protocols includes rigorous specifications for Baseline imaging procedures, the assumption that Worst Outcome is the appropriate imputation for uninterpretable reads is questionable. The agency review team has therefore provided supplementary analyses and statistics under two alternative schemes: the Interpretables scheme, which confines the analyses to the subset of interpretable reads, and the Pre=Post scheme, which imputes the Enhanced read diagnosis to the Baseline uninterpretable read. The results of the analyses under these alternative imputation schemes were as follows.

b(4)

For each of these Imputation schemes, Interpretables or Pre=Post, none of the four studies met the conditions of Requirement #1, namely, statistical superiority for both statistics for two of the three readers. ▽

b(4)

Further, although Study #12 did meet the first condition for Requirement #2, namely superiority in one statistic, and not inferior for the other, and Study #13 also met this first condition, superiority in Study #12 was provided for Specificity, whereas superiority in Study #13 was provided for Sensitivity, and therefore the second condition under Requirement #2, namely, superiority in the same statistic for studies in the same vessel group, was not met.

Consequently, under each of the alternative imputation schemes ( Interpretables, or Pre=Post), and under either set of success criteria ( Requirement #1 or Requirement #2), the statistics do not support the claim that Vasovist MRA outperforms Baseline MRA.

The statistics which support these conclusions are shown in the table below for the interpretable image results.

**Table (1.1.1)**  
**Statistics on Interpretables by Reader**

	READER A				READER B				READER C			
	PRE	POST	DIFF	95% CI	PRE	POST	DIFF	95% CI	PRE	POST	DIFF	95% CI
<b>Study #12</b>												
Sens.	.69	.82	.12	(.06, .3)	.68	.74	.05	(-.01, .11)	.51	.61	.10	(-.05, .15)
Spec.	.88	.85	-.04	(-.07, -.01)	.88	.94	.05	(.03, .07)	.93	.96	.03	(.01, .05)
<b>Study #13</b>												
Sens.	.61	.85	.24	(.17, .32)	.66	.84	.18	(.10, .26)	.56	.71	.15	(.06, .24)
Spec.	.89	.81	-.08	(-.11, -.05)	.85	.84	-.02	(-.05, .01)	.94	.93	-.01	(-.03, .01)

b(4)

Key:

	Superior
	Not Inferior
	Inferior

## 1.2 Overview of Clinical Studies

The EPIX NDA#21711 contains four Phase III studies dedicated to analyses of the clinical efficacy of Vasovist (MS-325) enhanced MRA for diagnoses of disease in several vessel groups. These studies are:

**Study MS-325-12:** A study of Vasovist enhanced MRA for the evaluation of patients with known or suspected Peripheral Vascular Disease.

**Study MS-325-13:** A second study of Vasovist enhanced MRA for the evaluation of patients with known or suspected Peripheral Vascular Disease.

**Study MS-325-14:** A study of Vasovist enhanced MRA for the evaluation of patients with known or suspected Renal Artery Disease.

**Study MS-325-15:** A study of Vasovist enhanced MRA for the evaluation of patients with known or suspected Pedal Artery Disease.

The design and the results for these four trials are described and examined individually directly below. The material directly below provides an overview of the common design elements for these trials and the principal results concerning Efficacy. In all that follows the four Phase III trials will be denoted as Study12 , Study#13 , Study#14 and Study#15. The four phase III trials have the following common design:

**Common Study Objective:** To evaluate the performance of a 0.03 mmol/kg dose of MS-325-Enhanced MRA when compared to pre-contrast (baseline) MRA, using X-ray Angiography (XRA) as the Standard of Reference (SOR), for the diagnosis of vessel disease in patients with known or suspected disease in a designated vessel bed ( typically consisting of four to eight vessels.)

**Common Imaging and Image Read Design:** These are Cross-Over Designs in which patients underwent Baseline MRA, MS-325 (Vasovist) Enhanced MRA, and Standard of Reference XRA. The Baseline and Vasovist MRA were performed open label and in sequence during one imaging session; the XRA imaging was performed within 30 days, but no sooner than 3 days, of the MRA imaging. The imaged vessels were evaluated for levels of stenosis. The stenosis level for any given vessel was the largest stenosis value found in the vessel. The reads relevant to the Primary Efficacy Analysis protocol were performed as follows: The MRA images were read independently by three blinded readers. The individual images examined by each blinded reader were images randomized with respect to patient, side (left or right), and sequence (baseline or enhanced). The XRA images were read independently by two blinded readers ( different from the MRA readers), with a third independent and blinded Adjudicator brought in whenever the diagnoses from the original two readers were contradictory with respect to binary decisions based on the primary endpoints. The Primary Efficacy Endpoint was vessel level Stenosis, which was defined to be the presence of a stenosis level of at least

50% in the vessel. For each read ( Baseline MRA, Vasovist Enhanced MRA, XRA) each vessel was assigned one of three values: Stenosis, No Stenosis, Uninterpretable.

**Common Primary Efficacy Endpoints:** The Primary Efficacy Endpoints were “vessel weighted” Sensitivity and Specificity. Vessel Level Sensitivity is the proportion of vessels (across all patients) identified as stenosed by XRA which were correctly identified as stenosed by MRA. Vessel Level Specificity is the proportion of vessels (across all patients) identified as non-stenosed by XRA which were correctly identified as non-stenosed by MRA. Uninterpretable vessels were assigned “Worst Outcome” values by the Sponsor; that is, the binary assignment was opposite the XRA assignment.

**Common Primary Efficacy Objectives:** The Primary Efficacy Objectives were:

(A):The Rejection of the Null Hypothesis of Equality of Baseline MRA Sensitivity to Vasovist Enhanced MRA Sensitivity.

(B):The Rejection of the Null Hypothesis of Equality of Baseline MRA Specificity to Vasovist Enhanced MRA Specificity.

These hypotheses were evaluated using a McNemar statistic which corrected for within-patient dependencies among vessel diagnostics. Significance was set at the .05 level. Sample sizes were set to achieve at least 80% power for 10% to 20% increases in Vasovist Enhanced Sensitivity and Specificity over Baseline Sensitivity and Specificity, under study-specific assumptions on percentages of vessels that would be inaccessible for XRA, and study-specific assumptions on vessel level disease prevalence. ( Details for the individual studies are in Section(1) below .) No assumptions were made concerning the expected percentages of MRA uninterpretable vessels.

**Note:** No hypotheses were provided regarding minimal performance levels for Vasovist MRA performance, for instance, that Sensitivity and Specificity achieve, say, 90% lower confidence values in excess of .80. The explicit Efficacy Objectives captured by rejection of the Null Hypotheses, along with the particulars on sample size determinations, can be conservatively interpreted as indications that the Sponsor expected Vasovist Enhanced MRA to increase both Sensitivity and Specificity by at least 10% over Baseline, but no specific hypotheses were provided for testing such improvements. In effect, the Sponsor’s criteria for a “Win” are explicitly reducible to rejection of the Null Hypotheses of Equality of Baseline MRA performance with Vasovist Enhanced MRA performance. It is to be understood that the differences, Enhanced over Baseline statistics, necessary for rejection of Equality must be *positive* differences, so that a “Win” requires that the lower limit of the two-sided 95% CI for differences for Enhanced over Baseline performance for both Sensitivity and Specificity exceed zero.

### 1.3 Statistical Issues and Findings

#### Criteria for Efficacy Evaluations in this Statistical Review

As described immediately above, the Sponsor's Efficacy Objectives translate into the following general criterion for a "Win" for the Sponsor:

The two-sided 95% Confidence Intervals for both Vasovist Sensitivity minus Baseline Sensitivity and Vasovist Specificity minus Baseline Specificity must have lower limits greater than zero. That is, (A) and (B) below must both obtain:

(A): Lower Limit of the 95% CI for Enhanced Sensitivity - Baseline Sensitivity  $> 0$   
and

(B): Lower Limit of the 95% CI for Enhanced Specificity - Baseline Specificity  $> 0$

These criteria, in turn, require strengthening as follows:

**Win Criterion:** At an absolute minimum, (A) and (B) must jointly hold for at least two of the three readers.

**Caveat:** All four trials present a significant problem which must be dealt with before the statistics can be evaluated with respect to the Win Criterion. The problem is the large percentages of Baseline uninterpretable images. The average numbers per study range from 10% to 40%, as compared with less than 2% for Enhanced uninterpretable images. There are two immediate ways to interpret this asymmetry:

*The occurrence of high percentages of uninterpretable images at Baseline is an intrinsic feature of baseline imaging.*

Or:

*The occurrence of high percentages of uninterpretable images at Baseline is a consequence of an underspecified Baseline imaging protocol which allowed for suboptimal Baseline imaging.*

If the latter possibility held, the Sponsor's Worst Outcome Imputation for Uninterpretables ( the outcome opposite to the XRA truth) would introduce a bias in favor of Vasovist. Since the second possibility cannot be discounted, a statistical analysis restricted to the Worst Outcome Imputation would not constitute an exhaustive and objective examination of the submitted results. For this reason this review will expand its analyses to include examination of the Sensitivity and Specificity statistics for three distinct Imputation schemes:

**(A): The Sponsor's chosen Worst Outcome Scenario**, in which all uninterpretables – baseline or enhanced image – were classified as incorrectly diagnosed. The rationale for this imputation rests on the assumption that large levels of uninterpretable imagings are evidence of intrinsic limitations in the Baseline diagnostic imaging technique.

(B): **The Interpretables Scenario**, in which the statistics were calculated only for vessels which were interpretable both at baseline and post-contrast. This scenario avoids the entire problem of imputation, but ignores the information in the relatively large subset of interpretable Vasovist images whose corresponding baselines were uninterpretable images. This imputation scheme corrects for the possibility that the Baseline imaging potential was underutilized.

(C): **Pre=Post Scenario**, in which the uninterpretable pre-images are assigned the same diagnoses as their post-injection interpretable counterparts. This scenario is consistent with the Null Hypothesis of equality of pre- and post-injection diagnostic statistics. In situations where the baselines have very high percentages of uninterpretables, this imputation scheme makes a good deal of sense. However, in cases where the percentages are significant, but not overwhelming, this scheme ignores the accumulated evidence gathered from the analyses of the joint pre- and post-injection performance on interpretables.

**Statistical Review Protocol:** This statistical review will consist principally of several tables which present the Sensitivity and Specificity statistics for Baseline and Enhanced diagnostics for the above described three imputation schemes. As would be expected, the advantages of Vasovist diagnostics over Baseline diagnostics diminish as one moves from Imputation Scheme(A) to Imputation Scheme(C). In particular, the Win Criteria, which are achieved in all four trials for the Sponsor's Imputation Scheme(A), are not achieved for either of the other two schemes in any of the four studies.

Table (1.3.1) below presents the Reader by Reader performances for Sensitivity and Specificity in the four Phase III Trials for each of the three Imputation schemes for Uninterpretables. These performances are evaluated here strictly with respect to the Sponsor's proposed criteria for a "Win", namely the rejection of the stipulated Null Hypotheses of Equality of performance, understood here as the determination of a lower limit for the two-sided 95% CI for Enhanced minus Baseline performance that exceeds zero.

A "Win" for a Reader occurs if both Hypotheses are rejected ( at 2-sided .05 Level)

Y means a Win obtains for the Reader; N means a Win doesn't obtain for the Reader

A Win for an Imputation Scheme requires, at a minimum, that at least two of the three Readers provide Wins simultaneously for both Sensitivity and Specificity.

N: (S) means No Win for Sensitivity; N: (Sp) means No win for Specificity

N: (S, Sp) means No Win for both Sensitivity and Specificity

**Basic Result for NDA#21711: None of the Studies provides a Win for the Interpretables Scheme or the Pre=Post Scheme.**

**Table (1.3.1)  
Win Profile by Study and Reader**

	RDR(A)	RDR(B)	RDR(C)
STUDY#12			
Worst Outcome	Y	Y	Y
Interpretables	N: (Sp)	N: (S)	Y
Pre=Post	N: (Sp)	N: (S)	Y
STUDY#13			
Worst Outcome	Y	Y	Y
Interpretables	N: (Sp)	N: (Sp)	N: (Sp)
Pre=Post	N: (Sp)	N: (Sp)	N: (Sp)

b(4)

Table 1.3.1 should be understood as the critical table for determination of success or failure of these studies. If the Worst Outcome Imputation Scheme is legitimate, then the sponsor wins on Study#12, Study#13, and Study#14. If the Worst Outcome Imputation Scheme is not legitimate, and study success criteria default to the Interpretables Scheme or the Pre=Post Scheme, then none of the studies achieves a "Win".

The next table presents a "smoothed" overview of the Confidence Interval Statistics that determined the entries in Table (1.3.1). These CI's represent averages over readers, and it is to be understood that this table, which is not reader specific, does not provide the individual confidence intervals which determined the classifications in the table above, although these reader smoothed confidence intervals, being sufficiently similar to the individual reader confidence intervals, should strongly suggest the contents in Table (1.3.1). The reader-specific confidence intervals are found in Section 3.1 below.

**Appears This Way  
On Original**

**Table (1.3.2)**  
**Averaged Sensitivities and Specificities and their Confidence Intervals**  
**(N = Number of Vessels)**

STUDY#12	N	Sensitivity			CI		Specificity				
		Pre	Post	Diff			N	Pre	Post	Diff	CI
Sponsor	237	.57	.72	.15	(.10, .20)		1409	.78	.91	.13	(.09, .16)
Interpretables	(210)	.63	.72	.09			(1223)	.90	.92	.02	
Pre=Post	237	.64	.72	.08	(.04, .12)		1409	.90	.91	.01	(-.01, .03)*
STUDY#13											
Sponsor	146	.53	.79	.26	(.21, .31)		1018	.74	.84	.10	(.05, .15)
Interpretables	(125)	.61	.80	.19			(835)	.89	.86	-.03*	
Pre=Post	146	.64	.79	.15	(.09, .22)		1018	.88	.84	-.04	(-.07, -.01)*

b(4)

Remarks: Reader-averaged Confidence Intervals are not provided for the Interpretables Scheme since the subsets of Interpretables changed from reader to reader.

The average numbers and percentages of Baseline Interpretable vessels are displayed in the Table (1.3.3), below, stratified by disease status, Gender, and Age Group. ( Enhanced Interpretables consistently exceeded 98% and are not displayed.) It should be noticed that there are two columns for Study#14. Column(A) displays a "derived" percentage for Interpretables; column (B) displays the real percentage for Interpretables. These distinctions will be described in the *Overview of the Design for Study#14* in Section 3.1 below.

**Appears This Way  
On Original**

**Table (1.3.3)**  
**Average levels of Interpretable Vessels by Study and Category**

Category	Study#12	Study#13
Overall	89%	84%
Stenosis	90%	87%
No Stenosis	87%	84%
Male	90%	84%
Female	88%	85%
Age<65	93%	83%
Age>=65	87%	85%

b(4)

The percentages of uninterpretables vary considerably from Center to Center as well as among readers. One possible explanation for these levels of variation by Center would be differences in imaging procedures from center to center. This issue remains to be investigated. (See table in Section 3.1.5.)

**Statistical Background for the Vessel Weighted Primary Endpoints**

Let T represent the "Reference" binary (XRA) vessel specific diagnosis that contributes to the statistic of some category of interest. Thus, if the particular category of interest is Sensitivity, then T=1 if significant stenosis is present in the vessel, while T =0 if significant stenosis is absent in the vessel; if the particular category of interest is Specificity, then T=1 if significant stenosis is absent in the vessel, while T =0 if significant stenosis is present in the vessel. ( There could, of course, be more than two categories. For instance, Stenosis could be partitioned into four categories: No Stenosis, Low Stenosis, Moderate Stenosis, High Stenosis. In such cases, T =1 whenever the XRA vessel diagnosis falls into the chosen category, T =0 whenever the XRA diagnosis falls into any of the other three categories.)

Let X represent the corresponding binary variable for Pre-Contrast MRA diagnoses, and let Y represent the corresponding binary variable for Post-Contrast MRA diagnoses, and let  $Z = Y - X$ . Now assume there are N patients with J vessels each. Set:

$T(k,j)$  = Binary Reference Diagnosis for vessel j for patient k

(a):  $X(k,j)$  = Corresponding Pre-Contrast Binary MRA diagnosis for vessel j for patient k

(b):  $Y(k,j)$  = Corresponding Post-Contrast Binary MRA diagnosis for vessel j for patient k

(c):  $Z(k,j) = Y(k,j) - X(k,j)$

In the statistics below,  $W(k,j) = (a) \text{ or } (b) \text{ or } (c)$  ( Fix one of these.)

Next:

$$V(k) = \sum_1^J T(k, j) = \# \text{ of vessels with diagnosis } T=1 \text{ by XRA for patient } k$$

$$U(k) = \sum_1^J W(k,j)T(k,j) \# \text{ of vessels where } W(k,j) = T(k,j)=1$$

Note that, when the variable T signifies the presence of Stenosis , then:

If  $W = X$ , then U measures the number of vessels among those vessels classified with Stenosis by the Reference for which the Pre-Contrast MRA diagnosis is also Stenosis.

If  $W = Y$ , then U measures the number of vessels among those vessels classified with Stenosis by the Reference for which the Post-Contrast MRA diagnosis is also Stenosis.

If  $W = Z$ , then U measures the difference,  $Y-X$  , over the vessels classified as having Stenosis by the Reference.

When the variable T signifies the absence of Stenosis, then:

If  $W = X$ , then U measures the number of vessels among those vessels classified with No Stenosis by the Reference for which the Pre-Contrast MRA diagnosis is also No Stenosis.

If  $W = Y$ , then U measures the number of vessels among those vessels classified with No Stenosis by the Reference for which the Post-Contrast MRA diagnosis is also No Stenosis.

If  $W = Z$ , then U measures the difference,  $Y-X$  , over the vessels classified as having No Stenosis by the Reference.

The Sponsor's Vessel Weighted Statistic is ( taken over N patients):

$$Q = \sum_1^N U(k) / \sum_1^N V(k)$$

When  $T = 1$  means Stenosis, then Q represents:

Vessel Weighted Pre-Contrast Sensitivity when  $W = X$

Vessel Weighted Post-Contrast Sensitivity when  $W = Y$

Vessel Weighted Post-Contrast Sensitivity minus Pre-Contrast Sensitivity when  $W = Z$

When  $T = 1$  means No Stenosis, then Q represents:

Vessel Weighted Pre-Contrast Specificity when  $W = X$   
Vessel Weighted Post-Contrast Specificity when  $W = Y$   
Vessel Weighted Post-Contrast Specificity minus Pre-Contrast Specificity when  $W = Z$

The Sponsor's Primary Efficacy evaluations were based on the function  $Q$ . The Sponsor rightly emphasized that the numerator of  $Q$  does not consist of independent ( vessel by vessel ) observations, and neither does the denominator, and moreover, the numerator is not independent of the denominator. If there were true independence at all levels, then the Sponsor's Null Hypotheses could be evaluated with the standard McNemar statistic. In order to accommodate this complication, the Sponsor introduced "Clustering" assumptions, involving various dependence parameters among a patient's true vessel diagnoses and also among the patient's MRA vessel diagnoses. These assumptions lead to a modified McNemar statistic, the length of whose confidence intervals reflect the levels of dependence. ( Intuitively, positive correlations between MRA vessel diagnoses for a particular patient, and also between XRA vessel diagnoses for that patient, would likely increase the length of the confidence intervals.) The Reviewer could not determine the reliability of these assumptions, and therefore sought out a different approach with fewer assumptions. This approach is detailed in the Appendix. It depends only on the Delta Method. *It will be noted here that this approach led to exactly the same confidence intervals as did the Sponsor's approach.*

## **2. Introduction**

### **2.1 Overview**

The EPIX NDA#21711 contains four Phase III studies dedicated to analyses of the clinical efficacy of Vasovist (MS-325) enhanced MRA for diagnoses of disease in several vessel groups. These studies are:

**Study MS-325-12:** A study of Vasovist enhanced MRA for the evaluation of patients with known or suspected Peripheral Vascular Disease.

**Study MS-325-13:** A second study of Vasovist enhanced MRA for the evaluation of patients with known or suspected Peripheral Vascular Disease.

**Study MS-325-14:** A study of Vasovist enhanced MRA for the evaluation of patients with known or suspected Renal Artery Disease.

**Study MS-325-15:** A study of Vasovist enhanced MRA for the evaluation of patients with known or suspected Pedal Artery Disease.

The design and the results for these four trials are described and examined individually directly below. The material directly below provides an overview of the common design

elements for these trials and the principal results concerning Efficacy. In all that follows the four Phase III trials will be denoted as Study12 , Study#13 , Study#14 and Study#15. The four phase III trials have the following common design:

**Common Study Objective:** To evaluate the performance of a 0.03 mmol/kg dose of MS-325-Enhanced MRA when compared to pre-contrast (baseline) MRA, using X-ray Angiography (XRA) as the Standard of Reference (SOR), for the diagnosis of vessel disease in patients with known or suspected disease in a designated vessel bed ( typically consisting of four to eight vessels.)

**Common Imaging and Image Read Design:** These are Cross-Over Designs in which patients underwent Baseline MRA, MS-325 (Vasovist) Enhanced MRA, and Standard of Reference XRA. The Baseline and Vasovist MRA were performed open label and in sequence during one imaging session; the XRA imaging was performed within 30 days, but no sooner than 3 days, of the MRA imaging. The imaged vessels were evaluated for levels of stenosis. The stenosis level for any given vessel was the largest stenosis value found in the vessel. The reads relevant to the Primary Efficacy Analysis protocol were performed as follows: The MRA images were read independently by three blinded readers. The individual images examined by each blinded reader were images randomized with respect to patient, side (left or right), and sequence (baseline or enhanced). The XRA images were read independently by two blinded readers ( different from the MRA readers), with a third independent and blinded Adjudicator brought in whenever the diagnoses from the original two readers were contradictory with respect to binary decisions based on the primary endpoints. The Primary Efficacy Endpoint was vessel level Stenosis, which was defined to be the presence of a stenosis level of at least 50% in the vessel. For each read ( Baseline MRA, Vasovist Enhanced MRA, XRA) each vessel was assigned one of three values: Stenosis, No Stenosis, Uninterpretable.

**Common Primary Efficacy Endpoints:** The Primary Efficacy Endpoints were “vessel weighted” Sensitivity and Specificity. Vessel Level Sensitivity is the proportion of vessels (across all patients) identified as stenosed by XRA which were correctly identified as stenosed by MRA. Vessel Level Specificity is the proportion of vessels (across all patients) identified as non-stenosed by XRA which were correctly identified as non-stenosed by MRA. Uninterpretable vessels were assigned “Worst Outcome” values by the Sponsor; that is, the binary assignment was opposite the XRA assignment.

**Common Primary Efficacy Objectives:** The Primary Efficacy Objectives were:

(A):The Rejection of the Null Hypothesis of Equality of Baseline MRA Sensitivity to Vasovist Enhanced MRA Sensitivity.

(B):The Rejection of the Null Hypothesis of Equality of Baseline MRA Specificity to Vasovist Enhanced MRA Specificity.

These hypotheses were evaluated using a McNemar statistic which corrected for within-patient dependencies among vessel diagnostics. Significance was set at the .05 level.

Sample sizes were set to achieve at least 80% power for 10% to 20% increases in Vasovist Enhanced Sensitivity and Specificity over Baseline Sensitivity and Specificity, under study-specific assumptions on percentages of vessels that would be inaccessible for XRA, and study-specific assumptions on vessel level disease prevalence. ( Details for the individual studies are in Section(1) below .) No assumptions were made concerning the expected percentages of MRA uninterpretable vessels.

**Note:** No hypotheses were provided regarding minimal performance levels for Vasovist MRA performance, for instance, that Sensitivity and Specificity achieve, say, 90% lower confidence values in excess of .80. The explicit Efficacy Objectives captured by rejection of the Null Hypotheses, along with the particulars on sample size determinations, can be conservatively interpreted as indications that the Sponsor expected Vasovist Enhanced MRA to increase both Sensitivity and Specificity by at least 10% over Baseline, but no specific hypotheses were provided for testing such improvements. In effect, the Sponsor's criteria for a "Win" are explicitly reducible to rejection of the Null Hypotheses of Equality of Baseline MRA performance with Vasovist Enhanced MRA performance. It is to be understood that the differences, Enhanced over Baseline statistics, necessary for rejection of Equality must be *positive* differences, so that a "Win" requires that the lower limit of the two-sided 95% CI for differences for Enhanced over Baseline performance for both Sensitivity and Specificity exceed zero.

#### **Criteria for Efficacy Evaluations in this Statistical Review**

As described immediately above, the Sponsor's Efficacy Objectives translate into the following general criterion for a "Win" for the Sponsor:

The two-sided 95% Confidence Intervals for both Vasovist Sensitivity minus Baseline Sensitivity and Vasovist Specificity minus Baseline Specificity must have lower limits greater than zero. That is, (A) and (B) below must both obtain:

(A): Lower Limit of the 95% CI for Enhanced Sensitivity - Baseline Sensitivity > 0  
and

(B): Lower Limit of the 95% CI for Enhanced Specificity - Baseline Specificity > 0

These criteria, in turn, require strengthening as follows:

**Win Criterion:** At an absolute minimum, (A) and (B) must jointly hold for at least two of the three readers.

**Caveat:** All four trials present a significant problem which must be dealt with before the statistics can be evaluated with respect to the Win Criterion. The problem is the large percentages of Baseline uninterpretable images. The average numbers per study range from 10% to 40%, as compared with less than 2% for Enhanced uninterpretable images. There are two immediate ways to interpret this asymmetry:

*The occurrence of high percentages of uninterpretable images at Baseline is an intrinsic feature of baseline imaging.*

Or:

*The occurrence of high percentages of uninterpretable images at Baseline is a consequence of an underspecified Baseline imaging protocol which allowed for suboptimal Baseline imaging.*

If the latter possibility held, the Sponsor's Worst Outcome Imputation for Uninterpretables ( the outcome opposite to the XRA truth) would introduce a bias in favor of Vasovist. Since the second possibility cannot be discounted, a statistical analysis restricted to the Worst Outcome Imputation would not constitute an exhaustive and objective examination of the submitted results. For this reason this review will expand its analyses to include examination of the Sensitivity and Specificity statistics for three distinct Imputation schemes:

**(A): The Sponsor's chosen Worst Outcome Scenario**, in which all uninterpretables – baseline or enhanced image – were classified as incorrectly diagnosed. The rationale for this imputation rests on the assumption that large levels of uninterpretable images are evidence of intrinsic limitations in the Baseline diagnostic imaging technique.

**(B): The Interpretables Scenario**, in which the statistics were calculated only for vessels which were interpretable both at baseline and post-contrast. This scenario avoids the entire problem of imputation, but ignores the information in the relatively large subset of interpretable Vasovist images whose corresponding baselines were uninterpretable images. This imputation scheme corrects for the possibility that the Baseline imaging potential was underutilized.

**(C): Pre=Post Scenario**, in which the uninterpretable pre-images are assigned the same diagnoses as their post-injection interpretable counterparts. This scenario is consistent with the Null Hypothesis of equality of pre- and post-injection diagnostic statistics. In situations where the baselines have very high percentages of uninterpretables, this imputation scheme makes a good deal of sense. However, in cases where the percentages are significant, but not overwhelming, this scheme ignores the accumulated evidence gathered from the analyses of the joint pre- and post-injection performance on interpretables.

**Statistical Review Protocol:** This statistical review will consist principally of several tables which present the Sensitivity and Specificity statistics for Baseline and Enhanced diagnostics for the above described three imputation schemes. As would be expected, the advantages of Vasovist diagnostics over Baseline diagnostics diminish as one moves from Imputation Scheme(A) to Imputation Scheme(C). In particular, the Win Criteria, which are achieved in all four trials for the Sponsor's Imputation Scheme(A), are not achieved for either of the other two schemes in any of the four studies.

Table (2.1.1) below presents the Reader by Reader performances for Sensitivity and Specificity in the four Phase III Trials for each of the three Imputation schemes for

Uninterpretables. These performances are evaluated here strictly with respect to the Sponsor's proposed criteria for a "Win", namely the rejection of the stipulated Null Hypotheses of Equality of performance, understood here as the determination of a lower limit for the two-sided 95% CI for Enhanced minus Baseline performance that exceeds zero.

A "Win" for a Reader occurs if both Hypotheses are rejected ( at 2-sided .05 Level)

Y means a Win obtains for the Reader; N means a Win doesn't obtain for the Reader

A Win for an Imputation Scheme requires, at a minimum, that at least two of the three Readers provide Wins simultaneously for both Sensitivity and Specificity.

N: (S) means No Win for Sensitivity; N: (Sp) means No win for Specificity

N: (S, Sp) means No Win for both Sensitivity and Specificity

***Basic Result for NDA#21711:None of the Studies provides a Win for the Interpretables Scheme or the Pre=Post Scheme.***

**Table (2.1.1)  
Win Profile by Study and Reader**

	RDR(A)	RDR(B)	RDR(C)
<b>STUDY#12</b>			
Worst Outcome	Y	Y	Y
Interpretables	N: (Sp)	N: (S)	Y
Pre=Post	N: (Sp)	N: (S)	Y
<b>STUDY#13</b>			
Worst Outcome	Y	Y	Y
Interpretables	N: (Sp)	N: (Sp)	N: (Sp)
Pre=Post	N: (Sp)	N: (Sp)	N: (Sp)

b(4)

Table 2.1.1 should be understood as the critical table for determination of success or failure of these studies. If the Worst Outcome Imputation Scheme is legitimate, then the

sponsor wins on Study#12, Study#13, and ——— If the Worst Outcome Imputation Scheme is not legitimate, and study success criteria default to the Interpretables Scheme or the Pre=Post Scheme, then none of the studies achieves a “Win”.

b(4)

The next table presents a “smoothed” overview of the Confidence Interval Statistics that determined the entries in Table (2.1.1). These CI’s represent averages over readers, and it is to be understood that this table, which is not reader specific, does not provide the individual confidence intervals which determined the classifications in the table above, although these reader smoothed confidence intervals, being sufficiently similar to the individual reader confidence intervals, should strongly suggest the contents in Table (2.1.1). The reader-specific confidence intervals which determined Table (2.1.1) are found in Section 3.1.

**Table (2.1.2)**  
**Averaged Sensitivities and Specificities and their Confidence Intervals**  
**(N = Number of Vessels)**

STUDY#12	N	Sensitivity				Specificity				
		Pre	Post	Diff	CI	N	Pre	Post	Diff	CI
Sponsor	237	.57	.72	.15	(.10, .20)	1409	.78	.91	.13	(.09, .16)
Interpretables	(210)	.63	.72	.09		(1223)	.90	.92	.02	
Pre=Post	237	.64	.72	.08	(.04, .12)	1409	.90	.91	.01	(-.01, .03)*
STUDY#13										
Sponsor	146	.53	.79	.26	(.21, .31)	1018	.74	.84	.10	(.05, .15)
Interpretables	(125)	.61	.80	.19		(835)	.89	.86	-.03*	
Pre=Post	146	.64	.79	.15	(.09, .22)	1018	.88	.84	-.04	(-.07, -.01)*

b(4)

Remarks: Reader-averaged Confidence Intervals are not provided for the Interpretables Scheme since the subsets of Interpretables changed from reader to reader.

The average numbers and percentages of Baseline Interpretable vessels are displayed in Table (2.1.3), below, stratified by disease status, Gender, and Age Group. ( Enhanced Interpretables consistently exceeded 98% and are not displayed.)

b(4)

distinctions will be described in section 3.1.3.

These

b(4)

**Table (2.1.3)**  
**Average levels of Interpretable Vessels by Study and Category**

Category	Study#12	Study#13
Overall	89%	84%
Stenosis	90%	87%
No Stenosis	87%	84%
Male	90%	84%
Female	88%	85%
Age<65	93%	83%
Age>=65	87%	85%

b(4)

## 2.2 Data Sources

The data sets on which the statistical analyses were based are:

MS325\_Supp\_Data.xpt (Arrived in EDR on 12-12-2003)

Upon FDA request, the Sponsor later provided an amplified version of this data set which incorporated variables required for analyses involving uninterpretable reads. The later data is

MS325\_Supp\_Data\_v2.xpt (Arrived in EDR on 9-02-2004)

The history behind the request and delivery of the expanded data set is as follows:

The FDA Medical Officer made a formal request for an expanded data set for the four primary studies on July 29 2004. This new data set was to include variables which directly indicated if the MRA blinded read for any vessel was interpretable or uninterpretable. The original data set provided such information only indirectly, and ambiguously; the only indicator of uninterpretability was a default stenosis level: if a vessel was uninterpretable, and if the XRA registered Stenosis, then the MRA read for the stenosis level was 0% ; if the XRA registered No Stenosis, then the MRA read for the stenosis level was 100%. The difficulty here was that it was impossible to determine the uninterpretability from this information since the values 0% or 100% for the MRA reads could have been real values rather than default values. The original data set did not include the uninterpretability flag because the Sponsor's statistics were limited to the Worst Outcome scheme, and the above described default stenoses levels, chosen as antithetical to the XRA, were adequate for evaluation of these statistics. The statistical reviewer was not aware that the appropriate uninterpretability flag was not included in the

original data until late July 2004 when the decision was made to conduct analyses for alternative imputation schemes. Since the Agency considered it essential that alternative imputation schemes be investigated, the uninterpretable had to be identified.

EPIX sent the requested data set on August 30, 2004, and its reception in EDR was reported to the statistical reviewer on September 2, 2004. The Sponsor reported that an earlier submission (mid August 2004) was rejected by EDR because of a formatting problem, consequently the data had to be sent again. This two week delay was critical, given the original PDUFA date for an October 15, 2004 closure. The extended PDUFA date reflects the Agency's judgment that more time was necessary for an adequate statistical analysis of the requested data set.

### **3. STATISTICAL EVALUATION**

#### **3.1 Evaluation of Individual Studies**

##### **3.1.1 Overview of Design and Evaluation of Efficacy for Study#12**

**Study Title:** A Multicenter, Comparative, Phase III Study to Determine the Safety and Efficacy of MS-325-Enhanced MRA for Evaluation of Aortoiliac Occlusive Disease in Patients with Known or Suspected Peripheral Vascular disease.

**Study Objective:** To evaluate the diagnostic performance of a 0.03 mmol/kg dose of MS-325-Enhanced MRA versus pre-contrast MRA with X-ray Angiography (XRA) as the Standard of Reference (SOR) in the evaluation of aortoiliac occlusive disease in patients with peripheral vascular disease.

**Imaging and Image Read Design:** This was a Cross-Over Design in which patients underwent Baseline MRA, MS-325 (Vasovist) Enhanced MRA, and XRA. The Baseline and Vasovist MRA were performed open label and in sequence during one imaging session; the XRA imaging was performed within 30 days but no sooner than 3 days of the MRA imaging. Up to seven peripheral vessels were imaged ( three on the left, four on the right). The reads relevant to the Primary Efficacy analysis protocol were performed as follows: The MRA images were read independently by three blinded readers. The individual images examined by a blinded reader were images randomized with respect to patient, side (left or right) and sequence (baseline or enhanced). The XRA images were read independently by two blinded readers ( different from the MRA readers.) Since the XRA reads constituted the Standard of Reference, divergences in XRA diagnoses were resolved by an independent third, adjudicating , XRA blinded reader. ( This adjudication was reserved for binary decisions so that a majority rule could be employed; details are presented below under Primary Efficacy Variables.)

**Primary Efficacy Variable:** The Primary Efficacy variable was vessel level Stenosis. A vessel was diagnosed as stenosed if the most severe diameter of stenosis in the vessel was at least 50%. If, for any particular vessel, a reader judged the image quality

inadequate for the primary efficacy variable diagnosis, the vessel was recorded as "uninterpretable". Thus, ( for non-missing data) each vessel was assigned a value Stenosis, Non-Stenosis, or Uninterpretable for the primary variable.

**Primary Efficacy Endpoints:** The Primary Efficacy Endpoints were "vessel weighted" Sensitivity and Specificity. Vessel Level Sensitivity is the proportion of vessels (across all patients) identified as stenosed by XRA which were also correctly identified as stenosed by MRA. Vessel Level Specificity is the proportion of vessels (across all patients) identified as non-stenosed by XRA which were also correctly identified as non-stenosed by MRA. The Protocol specified that, in all MRA diagnoses, uninterpretable vessels were assigned "worst outcome" values for the blinded readers; that is, the assignment was opposite the XRA assignment.

**Primary Efficacy Objective:** The Primary Efficacy Objective was rejection of the Null Hypothesis of Equality of Baseline MRA Sensitivity to Vasovist Enhanced MRA Sensitivity and rejection of the Null Hypothesis of Equality of Baseline MRA Specificity to Vasovist Enhanced MRA Specificity. These hypotheses were evaluated using a modified McNemar statistic which corrected for within-patient dependencies among diagnostics. The modified McNemar, like the standard McNemar, is a chi-squared statistic with one degree of freedom. Significance was set at the .05 level. Sample sizes of 288 and 263 patients were set to achieve 90% power for Sensitivity and Specificity, respectively, under the assumption that baseline Sensitivity and Specificity were .70, and Enhanced Sensitivity and Specificity were .80, and that about 80% of the vessels would not have clinically significant stenosis. These sample sizes also reflected several assumptions on diagnostic dependencies proper to the modified McNemar statistic, and an expected rate of about 15% for XRA non-assessable vessels. No assumptions regarding MRA uninterpretable image levels were provided.

### ***Principal Efficacy Results for Study#12***

**Investigational Sites:** Patients were enrolled into 28 sites, in the US or in Canada, between June 21 1999 and September 20 2001.

**Patient Disposition for Efficacy:** A total of 315 patients were enrolled into the study. There were 41 early discharges, largely for non-compliance or withdrawal of consent. Of the remaining 274 patients, 251 had evaluable XRA results and were the patients included in the primary efficacy analysis.

**Primary Efficacy Endpoint Results:** The principal problem encountered in Study#12 was the imbalance between baseline and enhanced image diagnostics in percentages of vessels classified as uninterpretable. Baseline reads averaged 13% for uninterpretable vessels, compared to an average of about 1% for Enhanced reads. The fact that one in seven vessels was uninterpretable for baseline reads posed a major efficacy analysis problem. One can either restrict analyses to the subset of interpretable images, or extend the analysis to the entire data set through imputation of diagnoses for the uninterpretable images. The statistics for three imputation methods are provided in this review:

(A): The Sponsor's chosen Worst Outcome scenario, in which all uninterpretables – baseline or enhanced image – were classified as incorrectly diagnosed. The rationale for this imputation rests with the assumption that large levels of uninterpretable imagings are evidence of intrinsic limitations in the diagnostic technique.

(B): The Interpretables scenario, in which the statistics were calculated only for vessels which were interpretable both at baseline and post-contrast. This scenario avoids the problem of imputation, but ignores the information in the relatively large subset of interpretable Vasovist images whose baselines were uninterpretable images.

(C): Pre=Post scenario, in which the uninterpretable pre-images are assigned the same diagnoses as their post-injection interpretable counterparts. This scenario is consistent with the Null Hypothesis of equality of pre- and post-injection diagnostic statistics. In situations where the baselines have very high percentages of uninterpretables, this imputation scheme makes a good deal of sense; when these percentages are large, but not overwhelming, this scheme ignores the accumulated evidence gathered from the analyses of the joint pre- and post-injection performance on interpretables.

**Definitions for Primary Statistical Endpoints:**

The primary statistical endpoints were “Vessel Weighted” Sensitivity and Specificity. These are defined directly below:

Let  $C1 = \# \text{ Vessels diagnosed with Stenosis by XRA}$

Let  $A1 = \# \text{ Vessels correctly diagnosed with Stenosis by Baseline MRA}$

Let  $B1 = \# \text{ Vessels correctly diagnosed with Stenosis by Enhanced MRA}$

Then

Vessel Weighted Baseline MRA Sensitivity =  $A1/C1$

Vessel Weighted Vasovist MRA Sensitivity =  $B1/C1$

Likewise

Let  $C2 = \# \text{ Vessels diagnosed with No Stenosis by XRA}$

Let  $A2 = \# \text{ Vessels correctly diagnosed with No Stenosis by Baseline MRA}$

Let  $B2 = \# \text{ Vessels correctly diagnosed with No Stenosis by Enhanced MRA}$

Then

Vessel Weighted Baseline MRA Specificity =  $A2/C2$

Vessel Weighted Vasovist MRA Specificity =  $B2/C2$

These statistics were calculated for each reader. An overview of these statistics for all three imputation schemes is presented in Table (3.1.1.2) below, where the statistics are averaged over the three blinded readers. It is to be noted here, once more, that the Sponsor's statistics were based on the Worst Outcome scenario. The critical review issues revolve around the validity of this imputation scheme. There are several reasonable objections to this scheme.

First, there is the possibility that the high percentages of uninterpretables among Baseline images are not intrinsic to Baseline imaging, but might instead be the result of imaging without adherence to a fixed imaging protocol; next, there is the absence of documented evidence within this trial that a clinically standard repeat imaging at baseline was performed when the first baseline was uninterpretable; an absence of repeat imagings would allow for a bias in favor of the enhanced images. Such repeat imaging does not appear to be plausible for enhanced images because of the narrow post-injection window of opportunity during which images are to be acquired, and the possible safety concerns attendant upon repeat injections.

**Note:** If, however, high levels of uninterpretability were intrinsic to baseline, it would be reasonable to assume possible correlations between uninterpretability of images for any particular patient and that patient's profile – gender, age, health status (stenosis, no stenosis.) There is, instead, some statistical evidence suggestive of randomness in the occurrence of uninterpretable images:

**Table (3.1.1.1)**  
**Percentages of Uninterpretable Images for Health, Gender, and Age Strata**  
**Study #12**

<b>Stratum</b>	
Overall	89%
Stenosis	90%
No Stenosis	87%
Male	90%
Female	88%
Age < 65	93%
Age ≥ 65	87%

**Preliminary remarks Concerning Table (3.1.1.2)**

Table (3.1.1.2) provides an overview of Vessel Weighted Sensitivity and Specificity for the three imputation schemes. The listed values are averages over the three blinded MRA readers. It should be noted that the 95% confidence intervals for Post vs Pre Injection

differences for the Interpretables scheme is not provided. The computation of the CI for the average for this scheme presents technical difficulties, having to do with the fact that the number of vessels whose truth values enter into the calculation changes from reader to reader.

**Table (3.1.1.2)**  
**Sensitivity and Specificity for Baseline and Enhanced MRA for Study#12**  
**(Averaged over the Three Readers)**

	#Vessels	Percent of Total	Baseline	Enhanced	Difference	95% CI
<b>Sensitivity</b>						
Worst Outcome	237	100%	.57	.72	.15	(.10, .20)
Interpretables	210	89%	.63	.72	.09	
Pre=Post	237	100%	.64	.72	.08	(.04, .12)
<b>Specificity</b>						
Worst Outcome	1409	100%	.78	.91	.13	(.09, .16)
Interpretables	1223	87%	.90	.92	.02	
Pre=Post	1409	100%	.90	.91	.01	(-.01, .03)

**Principal Tables:** The critical tables for evaluation of the Primary Statistical Endpoints are tables (3.1.1.3) and (3.1.1.4) below which list statistics by reader. The “Win” scenario for the Sponsor under any Imputation scheme requires that the lower limit of the two-sided 95% CI for the difference, Enhanced MRA value minus Baseline MRA value, exceed zero for both Sensitivity and Specificity for at least two readers. As the tables demonstrate, this “Win” scenario does not obtain under the Interpretables or the Pre=Post scenario. (The asterisk next to a CI indicates a failure.)

**Table (3.1.1.3)**  
**Sensitivity by Reader for Study #12**

	Reader A			Reader B			Reader C		
	Pre	Post	Diff	Pre	Post	Diff	Pre	Post	Diff
Worst Outcome	.62	.80	.18 (.10, .26)	.67	.73	.06* (0.0, .12)	.42	.61	.19 (.12, .26)
Interpretables	.69	.82	.12 (.06, .18)	.68	.74	.05* (-.01, .11)	.51	.61	.10 (.05, .15)
Pre=Post	.70	.80	.10 (.04, .16)	.69	.73	.05* (-.02, .12)	.53	.61	.08 (.03, .13)

Appears This Way  
 On Original

**Table (3.1.1.4)**  
**Specificity by Reader for Study #12**

	Reader A			Reader B			Reader C		
	Pre	Post	Diff	Pre	Post	Diff	Pre	Post	Diff
Worst Outcome	.75	.84	.09 (.05, .13)	.85	.93	.08 (.05, .11)	.75	.95	.20 (.16, .24)
Interpretables	.88	.85	-.04* (-.07,-0.1)	.88	.94	.05 (.03, .07)	.93	.96	.03 (.01, .05)
Pre=Post	.88	.85	-.04* (-.06,-.02)	.88		.05 (.03, .07)	.93	.95	.02 (.01, .03)

A final table for Study#12 is presented below. This table provides a more detailed look at the distribution of stenosis levels for Truth vs Baseline MRA and for Truth vs Enhanced MRA. The stenosis levels are partitioned into four separate categories: 0% to 9% ; 10% to 49% ; 50% to 90%; 91% to 100%. The vessels which contributed to this table were restricted to interpretables. The percentages are averages over readers.

**Table (3.1.1.5)**  
**XRA vs MRA Stenosis Levels – Study #12**

	Baseline MRA (4328 Reads)				Enhanced MRA (4901 Reads)			
	[0, 9]	[10,49]	[50, 90]	[91, 100]	[0, 9]	[10,49]	[50, 90]	[91, 100]
T = [0, 9]	81%	12%	5%	2%	75%	20%	5%	
T = [10,49]	58%	24%	15%	3%	40%	43%	17%	
T = [50, 90]	27%	20%	44%	9%	12%	24%	59%	5%
T = [91, 100]	7%	3%	10%	80%	4%	3%	17%	76%

Comments:

(1): The percentages are row percentages. Thus:

% of Baselines in [0, 9] when XRA is in [0,9] is 81%

(2): Two measures, one of Concordance, the other of Strong Discordance, can be derived from this table:

**Concordance Measure:**

Percent Agreement of Baseline MRA with XRA = 63%

Percent Agreement of Enhanced MRA with XRA = 66%

**Strong Discordance Measure:**

% of Baselines with less than 10% Stenosis when XRA has at least 50% Stenosis = 21%

% of Enhanced with less than 10% Stenosis when XRA has at least 50% Stenosis = 10%

*Thus, although the Agreement of Baseline MRA with XRA is essentially the same as the Agreement of Enhanced MRA with XRA, the Baselines are twice as likely to register with very low levels of Stenosis when XRA levels are at least 50%.*

### **3.1.2 Overview of Design and Evaluation of Efficacy for Study #13**

**Study Title:** A Multicenter, Phase III Study to Determine the Safety and Efficacy of MS-325-Enhanced MRA in Patients with Suspected Peripheral Vascular disease.

**Study Objective:** To evaluate the diagnostic performance of a 0.03 mmol/kg dose of MS-325-Enhanced MRA versus pre-contrast MRA with X-ray angiography (XRA) as the Standard of Reference (SOR) in the evaluation of peripheral vascular disease.

**Imaging and Read Design:** This was a Cross-Over Design in which patients underwent Baseline MRA, MS-325 (Vasovist) Enhanced MRA, and XRA. The Baseline and Vasovist MRA were performed open label and in sequence during one imaging session; the XRA imaging was performed within 30 days but no sooner than 3 days of the MRA imaging. Up to seven peripheral vessels were imaged ( three on the left, four on the right). The reads relevant to the Primary Efficacy analysis protocol were performed as follows: The MRA images were read independently by three blinded readers. The individual images examined by a blinded reader were images randomized with respect to patient, side (left or right) and sequence (baseline or enhanced). The XRA images were read independently by two blinded readers ( different from the MRA readers.) Since the XRA reads constituted the Standard of Reference, divergences in diagnoses were resolved by an independent third, adjudicating , XRA blinded reader. ( This adjudication was reserved for binary decisions so that a majority rule could be employed; details are presented below under Primary Efficacy Variables.)

**Primary Efficacy Variable:** The Primary Efficacy variable was vessel level Stenosis. A vessel was diagnosed as stenosed if the most severe diameter of stenosis in the vessel was at least 50%. If, for any particular vessel, a reader judged the image quality inadequate for the primary efficacy variable diagnosis, the vessel was recorded as “uninterpretable”. Thus, ( for non-missing data) each vessel was assigned a value Stenosis, Non-Stenosis, or Uninterpretable for the primary variable.

**Primary Efficacy Endpoints:** The Primary Efficacy Endpoints were “vessel weighted” Sensitivity and Specificity. Vessel Level Sensitivity is the proportion of vessels (across all patients) identified as stenosed by XRA which were also correctly identified as stenosed by MRA. Vessel Level Specificity is the proportion of vessels (across all patients) identified as non-stenosed by XRA which were also correctly identified as non-stenosed by MRA. In all MRA diagnoses, uninterpretable vessels were assigned “worst outcome” values for the blinded readers; that is, the assignment was opposite the XRA assignment.

**Primary Efficacy Objective:** The Primary Efficacy Objective was rejection of the Null Hypothesis of Equality of Baseline MRA Sensitivity to Vasovist Enhanced MRA Sensitivity and rejection of the Null Hypothesis of Equality of Baseline MRA Specificity to Vasovist Enhanced MRA Specificity. These hypotheses were evaluated using a modified McNemar statistic which corrected for within-patient dependencies among diagnostics. The modified McNemar, like the standard McNemar, is a chi-squared statistic with one degree of freedom. Significance was set at the .05 level. Sample sizes of 121 and 110 patients were set to achieve 90% power for Sensitivity and Specificity, respectively, under the assumption that baseline Sensitivity and Specificity were .70, and Enhanced Sensitivity and Specificity were .85, and that 80% of the vessels would not have clinically significant stenosis. These sample sizes also reflected several assumptions on diagnostic dependencies proper to the modified McNemar statistic; these assumptions were the same as were stated in the Studt#12 Design.

### ***Principal Efficacy Results for Study#13***

**Investigational Sites:** Patients were enrolled into 17 sites, in the US or in Germany, Scotland, Australia and Argentina, between December 10 2001 and October 28 2002.

**Patient Disposition for Efficacy:** A total of 178 patients were enrolled into the study. There were only 3 patients who did not complete the study.

### ***Principal Results for Sensitivity and Specificity.***

The principal problem encountered in Study#13 was the imbalance between baseline and enhanced image diagnostics in percentages of vessels classified as uninterpretable. Baseline reads averaged 16% for uninterpretable vessels, compared to an average of about 2% for Enhanced reads. The fact that about one in six vessels was uninterpretable for baseline reads posed problems for diagnostic imputations. The statistics for three imputation methods are provided in this review:

(A): The Sponsor's chosen Worst Outcome scenario, in which all uninterpretables – baseline or enhanced image – were classified as incorrectly diagnosed. The rationale for this imputation rests on the assumption that large levels of uninterpretable imagings are evidence of intrinsic limitations in the diagnostic technique.

(B): The Interpretables scenario, in which the statistics were calculated only for vessels which were interpretable both at baseline and post-contrast. This scenario avoids the entire problem of imputation, but ignores the information in the relatively large subset of interpretable Vasovist images whose baselines were uninterpretable images.

(C): Pre=Post scenario, in which the uninterpretable pre-images are assigned the same diagnoses as their post-injection interpretable counterparts. This scenario is consistent with the Null Hypothesis of equality of pre- and post-injection diagnostic statistics. In situations where the baselines have very high percentages of uninterpretables, this imputation scheme makes a good deal of sense; when the percentages are large, but not

overwhelming, this scheme ignores the accumulated evidence gathered from the analyses of the joint pre- and post-injection performance on interpretables.

**Definitions for Primary Statistical Endpoints:**

The primary statistical endpoints were “Vessel Weighted” Sensitivity and Specificity. These are defined directly below:

Let  $C1 = \#$  Vessels diagnosed with Stenosis by XRA

Let  $A1 = \#$  Vessels correctly diagnosed with Stenosis by Baseline MRA

Let  $B1 = \#$  Vessels correctly diagnosed with Stenosis by Enhanced MRA

Then

Vessel Weighted Baseline MRA Sensitivity =  $A1/C1$

Vessel Weighted Vasovist MRA Sensitivity =  $B1/C1$

Likewise

Let  $C2 = \#$  Vessels diagnosed with No Stenosis by XRA

Let  $A2 = \#$  Vessels correctly diagnosed with No Stenosis by Baseline MRA

Let  $B2 = \#$  Vessels correctly diagnosed with No Stenosis by Enhanced MRA

Then

Vessel Weighted Baseline MRA Specificity =  $A2/C2$

Vessel Weighted Vasovist MRA Specificity =  $B2/C2$

These statistics were calculated for each reader. An overview of these statistics for all three imputation schemes is presented in Table (3.1.2.2) below, where the statistics are averaged over the three blinded readers. It is to be noted here, once more, that the Sponsor’s statistics were based on the Worst Outcome scenario. The critical review issues revolve around the validity of this imputation scheme. First, there is the possibility that the high percentages of uninterpretables among Baseline images are not intrinsic to Baseline imaging, but might instead be the result of imaging without adherence to a fixed imaging protocol; next, there is the absence of documented evidence within this trial that a clinically standard repeat imaging at baseline was performed when the first baseline was uninterpretable; an absence of repeat imagings would allow for a bias in favor of the enhanced images. Such repeat imaging does not appear to be plausible for enhanced images because of the narrow post-injection window of opportunity during

which images are to be acquired, and the possible safety concerns attendant upon repeat injections.

**Note:** If, however, high levels of uninterpretability were intrinsic to baseline, it would be reasonable to assume possible correlations between uninterpretability of images for any particular patient and that patient's profile – gender, age, health status (stenosis, no stenosis.) There is, instead, some statistical evidence suggestive of randomness in the occurrence of uninterpretable images:

**Table (3.1.2.1)  
Percentages of Uninterpretable Images for Health, Gender, and Age Strata  
Study #13**

<b>Stratum</b>	
Overall	84%
Stenosis	87%
No Stenosis	84%
Male	84%
Female	85%
Age<65	83%
Age≥65	85%

**Preliminary remarks Concerning Table (3.1.2.2)**

(1): Table (3.1.2.2) provides an overview of Vessel Weighted Sensitivity and Specificity for the three imputation schemes. The listed values are averages over the three blinded MRA readers. It should be noted that the 95% confidence intervals for Post vs Pre Injection differences for the Interpretables scheme is not provided. The computation of the CI for the average for this scheme presents technical difficulties, having to do with the fact that the number of vessels whose truth values enter into the calculation changes from reader to reader.

**Table (3.1.2.2)  
Sensitivity and Specificity for Baseline and Enhanced MRA for Study#13  
( Averaged over the Three Readers)**

	#Vessels	Percent of Total	Baseline	Enhanced	Difference	95% CI
<b>Sensitivity</b>						
Worst Outcome	146	100%	.53	.79	.26	(.21, .31)
Interpretables	125	86%	.61	.80	.19	
Pre=Post	146	100%	.64	.79	.15	(.09, .22)
<b>Specificity</b>						
Worst Outcome	1018	100%	.74	.84	.10	(.05, .15)
Interpretables	835	82%	.89	.86	-.03	
Pre=Post	1018	100%	.88	.84	-.04	(-.07, -.01)

**Principal Tables:** The critical tables for evaluation of the Primary Statistical Endpoints are tables (3.1.2.3) and (3.1.2.4) below which list statistics by reader. The “Win” scenario for the Sponsor under any Imputation scheme requires that the lower limit of the two-sided 95% CI for the difference, Enhanced MRA value minus Baseline MRA value, exceed zero for both Sensitivity and Specificity for at least two readers. As the tables demonstrate, this “Win” scenario does not obtain under the Interpretables or the Pre=Post scenario. (The asterisk next to a CI indicates a failure.)

**Table (3.1.2.3)  
Sensitivity By Reader for Study #13**

	Reader A			Reader B			Reader C		
	Pre	Post	Diff	Pre	Post	Diff	Pre	Post	Diff
Worst Outcome	.52	.83	.31 (.21, .41)	.60	.84	.24 (.14, .34)	.49	.70	.22 (.11, .33)
Interpretables	.61	.85	.24 (.16, .32)	.66	.84	.18 (.10, .26)	.56	.71	.15 (.06, .24)
Pre=Post	.63	.83	.20 (.12, .28)	.69	.84	.15 (.08, .22)	.61	.70	.10 (0.0, .20)

**Table (3.1.2.4)  
Specificity By Reader for Study #13**

	Reader A			Reader B			Reader C		
	Pre	Post	Diff	Pre	Post	Diff	Pre	Post	Diff
Worst Outcome	.71	.80	.09 (.03, .15)	.74	.83	.09 (.04, .14)	.78	.90	.12 (.07, .17)
Interpretables	.89	.81	-.08* (-.11, -.05)	.85	.84	-.02* (-.05, .01)	.94	.93	-.01* (-.03, .01)
Pre=Post	.87	.80	-.07* (-.10, -.04)	.85		-.02* (-.05, .01)	.92	.90	-.02* (-.04, 0.0)

A final table for Study#13 is presented below. This table provides a more detailed look at the distribution of stenosis levels for Truth vs Baseline MRA and for Truth vs Enhanced MRA. The stenosis levels are partitioned into four separate categories: 0% to 9% ; 10% to 49% ; 50% to 90%; 91% to 100%.

**Appears This Way  
On Original**

11 Page(s) Withheld

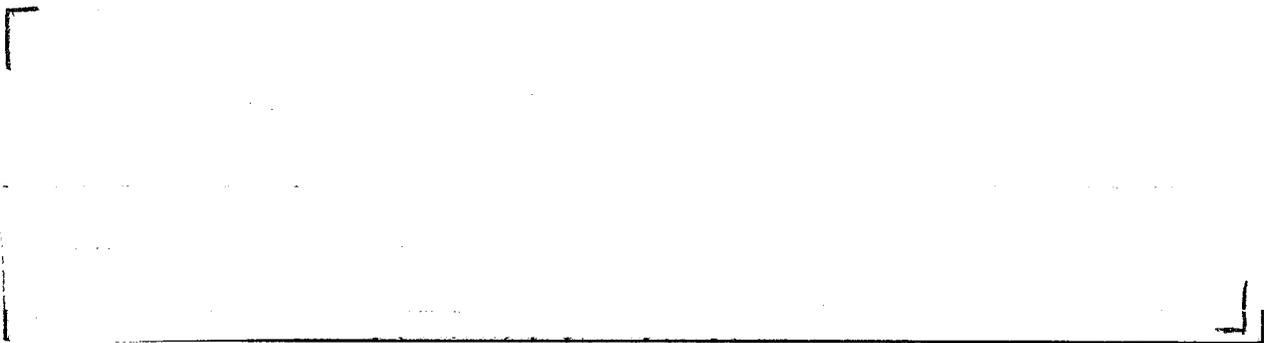
Trade Secret / Confidential (b4)

Draft Labeling (b4)

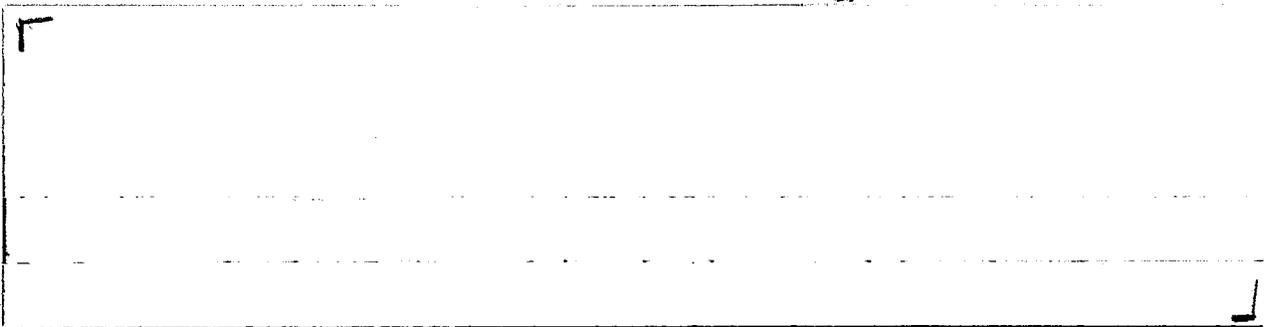
Draft Labeling (b5)

Deliberative Process (b5)

Withheld Track Number: Statistical- 1



b(4)



b(4)

### 3.1.5 Remarks on Uninterpretable Images Stratified by Center

Table (3.1.5.1) below provides some evidence that the percentages of uninterpretables vary considerably from Center to Center. One possible explanation for these levels of variation would be differences in imaging procedures from center to center. This issue remains to be investigated.

Appears This Way  
On Original

**Table (3.1.5.1)**  
**Percentages of Uninterpretable Images by Center**  
 (Percentages refer to % of Vessels)  
 (For Studies 12, 13, 14, centers with 10 or more patients are presented)  
 (For Study 15, centers with 7 or more are presented)

Center	N (Patients)	% Stenosis	Non-interpretable				
			Reader A	Reader B	Reader C	Average	Rank
<b>Study # 12</b>							
38	10	4%	0	0	0	0%	1
13	16	11%	4%	0	4%	3%	2
27	41	12%	1%	0	7%	3%	3
40	33	13%	29%	3%	31%	21%	7
20	40	15%	9%	3%	10%	7%	4
34	13	17%	32%	3%	38%	24%	8
21	13	18%	14%	8%	9%	10%	5
19	36	24%	16%	5%	12%	11%	6
All others	99		13%	5%	20%		
<b>Study #13</b>							
66	18	3%	38%	21%	20%	26%	6
83	30	10%	16%	12%	12%	13%	4
68	30	14%	4%	14%	3%	7%	2
114	16	15%	15%	13%	32%	20%	5
108	13	16%	16%	2%	11%	10%	3
67	29	17%	6%	3%	5%	5%	1
All others	42		35%	24%	29%		

b(4)

### 3.2 Evaluation of Safety

Refer to the Medical Officer's review.

### 4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

There are no special subpopulations requiring analysis.

### 5. SUMMARY AND CONCLUSIONS

The Sponsor provided four primary Phase III Diagnostic Imaging studies of Vasovist Enhanced MRA for the evaluation of patients with known or suspected peripheral vascular disease (two studies), renal artery disease (one study), or pedal artery disease (one study). The primary efficacy objective in all four studies was the determination of the presence or absence of significant stenosis in the vessels under examination. Each study used Baseline (non-contrast) MRA as Comparator, and XRA as the Standard of Reference, and the images for each study were evaluated by three blinded readers. The Sponsor's goal was to demonstrate improved Sensitivity and Specificity for Enhanced MRA over baseline MRA. The Sponsor's criteria for Improvement in diagnostic performance required that the lower value for the two-sided 95% Confidence Interval for the difference, Enhanced MRA measure minus baseline MRA measure, must exceed zero for both Sensitivity and Specificity. The Sponsor did not specify how achievement of these criteria were to be verified with respect to the three readers; for instance, it was not specified that all three readers had to meet the criteria. In the absence in the submission of a clear statement from the Sponsor regarding reader-based criteria for success, the statistical reviewer, after consultation with the Division review team, decided that a reasonable requirement would be that two of the three readers achieved the criteria for improvement.

The Sponsor's results in three of the four studies (two peripheral, — achieved this goal, but only under conditions which the Agency found troublesome. The problem was the following: The most significant feature in all four studies was the percentage of uninterpretable vessels for Baseline image reads – 10% to 40% over all studies, all readers, as contrasted with less than 2% for Enhanced image reads. Such levels could be consistent with: (a) inherent limitations in Baseline MRA diagnostics, or (b) an underspecified Baseline imaging protocol. The Sponsor chose to impute incorrect diagnoses to uninterpretable images, a procedure consistent with (a); it is this imputation scheme, designated as Worst Outcome, which ensures successful performance for Enhanced MRA diagnoses over Baseline MRA diagnoses. The protocols for the various studies do not include rigorous specifications for Baseline imaging; consequently the assumption that Worst Outcome is the appropriate imputation for uninterpretable reads is questionable. The review team has therefore provided supplementary analyses and statistics under two alternative schemes: the Interpretables scheme, which confines the analyses to the subset of interpretable reads, and the Pre=Post scheme, which imputes the

b(4)

Enhanced read diagnosis to the Baseline uninterpretable read. Under each of these schemes, and with respect to the requirement that Superiority be achieved for both Sensitivity and Specificity for at least two of the three readers, the statistics do not support the claim that Vasovist MRA outperforms non-contrast MRA.

Appears This Way  
On Original

Appears This Way  
On Original

## APPENDIX

### General Delta Method Approach

Note that the sequence of vectors  $(U(1),V(1)), (U(2),V(2)), \dots (U(N),V(N))$  is an i.i.d. sequence of bivariate vectors with a common distribution represented by some vector  $(U,V)$  where:

$$(U_0, V_0) = \text{Mean of } (U, V)$$

$$\Gamma = \text{Covariance Matrix of } (U, V) = \begin{bmatrix} \sigma^2(U) & C(U,V) \\ C(U,V) & \sigma^2(V) \end{bmatrix}$$

$$\text{Let } C(U,V) = R \sigma(U) \sigma(V)$$

Then, from the Delta Method:

$$Q = \sum U(I) / \sum V(I) \approx \text{Normal} (M, S^2) \text{ where}$$

$$M = U_0 / V_0 \text{ and}$$

$$S^2 = (U_0 / V_0)^2 \left\{ \sigma^2(U) / (U_0)^2 + \sigma^2(V) / (V_0)^2 - 2R \sigma(U) \sigma(V) / U_0 V_0 \right\} / N$$

In this Review, all the statistics on  $Q$ , in particular the 95% CI's for the several primary statistics – Sensitivities and Specificities for Baseline and Enhanced reads; Enhanced versus Baseline differences in Sensitivities and Specificities - were evaluated using Normal  $(M, S^2)$ , with

$$U_0 \approx \sum U(k) / N ; V_0 \approx \sum V(k) / N$$

$$\sigma^2(U) \approx \sum_1^N (U(k) - U_0)^2 / (N-1)$$

$$\sigma^2(V) \approx \sum_1^N (V(k) - V_0)^2 / (N-1)$$

$$C(U,V) \approx \sum_1^N (U(k) - U_0)(V(k) - V_0) / N$$

The advantage of this approach lies in its circumvention of all assumptions concerning the diagnostic dependencies “within patient”; it simply lets the U and V statistics do the work.

Appears This Way  
On Original

Appears This Way  
On Original

**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  
12/30/04 01:31:49 PM  
BIOMETRICS

Mike Welch  
12/31/04 07:42:26 PM  
BIOMETRICS  
Concur with review.