CENTER FOR DRUG EVALUATION AND RESEARCH

APPROVAL PACKAGE FOR:

APPLICATION NUMBER

20-372/S-013

Statistical Review(s)

DEPARTMENT OF HEALTH AND HUMAN SERVICES
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH
OFFICE OF BIOSTATISTICS

Statistical Review and Evaluation

CLINICAL STUDIES

NDA:	20-372 SEI-013 Efficacy Supplement
Name of drug:	Myoview(Technetium 99m Tc-Tetrofosmin)
Applicant:	Amersham Health
Indication:	GSPECT Imaging of the Myocardium
Documents reviewed	Hardcopy submission dated April 29 2002
	Volumes 1 & Volumes 36 through 68 & Volume 72
Project manager:	Patricia Stewart (HFD-160).
Clinical reviewer:	Sally Loewke MD (HFD-160)
Dates:	Received 4/29/02; user fee (10 months) 3/01/02
Statistical reviewer:	Mucci, A. G. Ph.D. (HFD-715)
Statistics team leader:	Michael Welch, Ph.D. (HFD-715).
Biometrics division director:	S. Edward Nevius, Ph.D. (HFD-715)
Keywords:	NDA review, clinical study, diagnostic accuracy, non-inferiority

1. Executive summary

1.1 Overview

Myoview scintigraphic imaging of the myocardium is currently indicated for the detection of perfusion changes induced by rest/stress testing in patients with known or suspected coronary artery disease. The efficacy supplement under review is intended as evidence for an extension of this indication to include the assessment of ventricular function (Ejection Fraction; Wall Motion) in similar patients. Pursuant to this objective, the submitted studies (MYO301/MYO303) focus on the efficacy of Core-Center blinded categorical diagnoses (Normal/Abnormal) for these functional endpoints for Myoview GSPECT (ECG gated myocardial perfusion single photon emission computerized tomography). The primary statistical endpoints are Sensitivity, Specificity and Accuracy, with Core-Center blinded MUGA (Multiple gated acquisition) reads providing the Reference Standard. Presumably, some reasonably high Sensitivities and Specificities of Core-Center blinded read Myoview GSPECT with respect to Core-Center blinded read MUGA would qualify Myoview GSPECT as a *substitute* for MUGA for both LVEF and WM diagnoses. However, since the Sponsor merely reported Sensitivities and Specificities without providing criteria for success for Myoview GSPECT when compared to MUGA, and since MUGA is itself subject to error in LVEF and WM disease classifications (a circumstance ignored by the Sponsor), the submitted results reduce to a straightforward report of statistical efficacy endpoints whose values (about 80% for Sensitivity and Specificity; see Appendix) do not provide unequivocal support for the proposed extended indication. The statistical reviewer, in an attempt at addressing these liabilities, took the following approach to this review:

(1): Core-Center MUGA diagnoses were reinterpreted as Comparator diagnoses for evaluation of Core-Center Myoview GSPECT diagnoses, with On-Site MUGA diagnoses substituting for Core-Center MUGA diagnoses as the Reference Standard. This approach allows MUGA vs MUGA error differences in diagnoses to be addressed.

(2): Myoview GSPECT was considered as comparable in performance to MUGA if the 95% two-sided confidence interval for differences, Core-Center Myoview GSPECT minus Core-Center MUGA, for both Sensitivity and Specificity of diagnoses, and with On-Site MUGA as the reference standard, fell above -.10. However, it is to be stressed that this approach constituted an exploratory *choice* for statistical analyses by the statistical reviewer for the evaluation of Myoview GSPECT vs MUGA; the Sponsor provided no comparable and relevant analyses.

1.2 Principal Findings

(a): The non-inferiority criteria above, for both Sensitivity and Specificity for Myoview GSPECT vs MUGA, were met in both studies for both LVEF and WM. This equivalence is taken here as suggestive evidence that Myoview GSPECT is substitutable for MUGA.

1.3 Conclusions and Recommendations

When Core-Center Myoview GSPECT LVEF and WM Sensitivities and Specificities are compared to Core-Center MUGA Sensitivities and Specificities, with On-Site MUGA as the Standard, Myoview GSPECT achieves levels similar to MUGA.

2. Statistical Review and Evaluation of Evidence

2.1 Introduction and Background

Myoview enhanced GSPECT Myocardial Imaging is currently approved with the following indication:

Myoview is indicated for scintigraphic imaging of the myocardium following separate administration under exercise and/or resting conditions. It is useful in the delineation of regions of reversible myocardial ischemia in the presence or absence of infarcted myocardium. Myoview is also indicated for scintigraphic imaging of the myocardium to identify changes in perfusion induced by pharmacologic stress in patients with known or suspected coronary artery disease.

The Sponsor proposes the following addition to this indication:

Myoview is also indicated for the assessment of ventricular function in patients being evaluated for heart disease and/or ventricular function.

This additional indication is functional, and the supporting evidence for this addition provided in the two trials (MYO301/MYO303) herein examined consists of evaluations of the functional endpoints of *Left Ventricular Ejection Fraction* (LVEF) and *Wall Motion* (WM).

The Phase III studies MYO301 and MYO303 had identical designs and objectives. These common designs and objectives are outlined directly below.

Title: An Open-Label, Multicentre, Phase III Trial Evaluating Ventricular Function as Assessed by LVEF and Wall Motion Using Technetium-99m Tetrofosmin (Myoview) Gated SPECT Imaging.

Objective: To evaluate the diagnostic utility of Tc-99m Tetrofosmin (Myoview)gated SPECT (GSPECT) Imaging in assessing LVEF and WM.

Patient Population: Stable subjects who are being evaluated for known or suspected cardiac disease and/or ventricular function and who have been referred for exercise stress myocardial perfusion imaging.

Sample Size: MYO301 enrolled 145 subjects of whom 127 were evaluable by core center blinded reads for LVEF and 124 were evaluable for core center blinded reads for WM. MYO303 enrolled 191 subjects of whom 170 were evaluable for core center blinded reads for LVEF and 171 were evaluable for core center blinded reads for WM.

Imaging Modalities: The Test diagnostic was Core-Center independent blinded read evaluation of Myoview GSPECT Imagings; the Reference Diagnostic was Core-Center consensus read evaluation of MUGA Imagings.

Imaging Sequence: Subjects underwent, typically on the same day, low dose rest Myoview GSPECT Imaging (9-12mCi), followed by high dose exercise stress Myoview GSPECT Imaging (15-24mCi). The Myoview GSPECT Images constitute the Test Images for LVEF and WM evaluations. One to five days subsequent to Myoview GSPECT Imaging the subjects underwent rest MUGA Imaging, utilizing [redacted] RBC labelled red blood cells (15-20mCi). The MUGA Images constitute the Reference Images for LVEF and WM evaluations.

Efficacy Endpoints: The primary diagnostic efficacy endpoints are LVEF and subject level WM. The efficacy of these endpoints for Myoview GSECT Imaging was assessed by the Sponsor through comparisons to LVEF and subject level WM as determined by consensus Core-center MUGA reads. For primary analyses, both LVEF and WM were collapsed into the categories of Normal/Abnormal, and the principal statistics for Myoview GSPECT assessment were Accuracy, Sensitivity, and Specificity for these binary classifications with respect to the reference standard of Core-Center MUGA reads. The Myoview GSPECT reads were provided by three independent blinded readers; the MUGA reads were three reader consensus reads. The LVEF reads for both GSPECT and MUGA were *evaluations* of LVEF values provided by a Core-Center imaging specialist. This evaluation resulted in acceptance of the imaging specialist's LVEF calculation in virtually all cases in both studies; thus, the three independent blinded GSPECT LVEF reads reduced essentially to one read.

Study Hypotheses: The Sponsor did not construct the trials with Efficacy hypotheses for Myoview GSPECT Accuracy, Sensitivity, Specificity as their focus. Thus, there were no pre-chosen values for Accuracy, Sensitivity, and Specificity with respect to which the trial results were to be tested so as to determine study success or failure. The Sponsor did remark that the sample sizes were adequate to distinguish between a Null Hypothesis that LVEF Sensitivity (or Specificity) = .60 and an Alternative Hypothesis that LVEF Sensitivity (or Specificity) = .75 with significance = .05 and power = .72.
This level of efficacy required at least 51 normal subjects and 51 abnormal subjects. Similarly, 102 subjects provide that the Null Hypothesis that WM Accuracy = .60 is distinguishable from the Alternative Hypothesis that WM Accuracy = .75 with significance = .05 and power = .94. However, the Sponsor stated that the sample sizes were not determined with respect to significance and power for hypotheses, but rather, were determined principally so as to ensure that the estimates of these primary statistical measures were within .10 of the true values, given that the true values were in the neighborhood of .75 to .85. Thus, the studies were not geared towards examination of

any particular efficacy hypotheses, but, rather, were designed merely to ensure reasonable accuracy for the calculated values of the primary statistics.
Principal Statistical Reviewer Concerns: The Sponsor's conclusion regarding the trial results is: *The present studies demonstrate that Myoview MPI (Myocardial Perfusion Imaging) GSPECT provides very reliable assessment of LV function (LVEF and WM) for use in clinical management of patients with known or suspected heart disease.*

The Statistical Reviewer infers from this statement that, with respect to LVEF and WM evaluations, the Sponsor is proposing that the trial results provide sufficient evidence that Myoview GSPECT MPI can serve as a reliable *substitute* for MUGA evaluations of these functional endpoints. However, since Sensitivity/Specificity/Accuracy for Core-Center Myoview GSPECT EF and WM were typically in the .80 range with respect to Core-Center MUGA, it would follow that if MUGA diagnoses constituted a very highly reliable reference standard, then Myoview GSPECT diagnoses could be wrong in one in every five cases. This error rate seems inconsistent with the Sponsor's statement above regarding Myoview GSPECT reliability. In fact, the 80% figures actually achieved could, at best, legitimize Myoview GSPECT as a substitute for MUGA if MUGA itself presented with errors of *replicability of diagnosis* of a similar magnitude. This line of reasoning, (absent from the Sponsor's submission), is elaborated upon below, and underlies the Statistical Reviewer's procedures (as contrasted with the Sponsor's procedures) in analyzing these trials.

2.2 Statistical Evaluation of Evidence on Efficacy

The Statistical Reviewer's procedure for evaluation of the results in these two trials will take the following form:

Sensitivity/Specificity/Accuracy will continue to serve as the primary efficacy endpoints, but instead of a direct comparison of Core-Center Myoview GSPECT to Core-Center MUGA for evaluation of these endpoints, the following approach will be used:

(a): On-Site MUGA data will substitute for Core-Center MUGA data as the Reference Standard.

(b): Core-Center MUGA data will serve as the Comparator.

(c): Myoview GSPECT data will constitute Test data.

Remarks

(1):This procedure is not intended to imply that On-Site MUGA results are more "accurate" than Core-Center MUGA results; the intention, rather, is to gauge, in the only way that appears possible with the existing study data, the extent to which discrepancies between MUGA and Myoview GSPECT are not necessarily indicative of Myoview

GSPECT misdiagnoses, but could just as well represent levels of non-replicability for MUGA diagnoses themselves. Effectively, the 80% agreement between Core-Center MUGA diagnoses and Core-Center GSPECT diagnoses would not register unfavorably for GSPECT if the agreement between On-Site MUGA diagnoses and Core-Center MUGA diagnoses were also in the neighborhood of 80%.

(2): The incorporation of On-Site MUGA data as a Reference demanded that the patient population be reduced to include only those patients for whom both On-Site and Core-Center results were available. Thus:

Study MYO301 was reduced from 127 evaluables to 120 evaluables

Study MYO303 was reduced from 170 evaluables to 153 evaluables

The *primary* analysis of the LVEF and WM data on these patients will be restricted to the following *subject level* comparisons:

On –Site *categorical* MUGA EF to Core-Center *categorical* MUGA EF

On–Site *categorical* MUGA EF to *categorical* Myoview GSPECT EF for Reader#1

On-Site *categorical* MUGA WM to Core-Center *categorical* MUGA WM

On-Site *categorical* MUGA WM to *categorical* Myoview GSPECT majority read WM

The *subject level categorical* classifications are defined directly below:

Reduction of Diagnoses to categorical form:

The reduction of the continuous range LVEF scores to subject level categorical form is accomplished as follows:

Normal if LVEF $\geq$ 50 ; Abnormal if LVEF π 50

The WM classifications and their reduction to subject categorical form are as follows:

(1): Each of five regions (Apical, Inferior, Septal, Anterior, Lateral) is classified into one of four categories: Normal, Hypokinesis, Akinesis, Dyskinesis.

(2): Each such regional classification is then collapsed:

Normal = Normal ; *Abnormal* = Hypokinesis, Akinesis, Dyskinesis.

(3): A subject is then classified as follows:

Normal if all regions are Normal ; *Abnormal* if at least one region is Abnormal
The reduction of the three reader comparison of MUGA EF to GSPECT EF to the single comparison of MUGA EF to Reader#1 GSPECT EF is justified by the fact that the three Myoview GSPECT EF reads consisted largely in independent reader acceptance of an EF evaluation provided, in each trial, by a single technician. In fact, over both studies, in the 10% of cases in which one or more readers disagreed with the technician and requested a re-evaluation of EF, there wasn't a single case in which the second read issued in a reclassification of EF from normal to abnormal, or vice-versa.

As regards WM analyses, Table(1) below serves as justification for the reduction of the comparisons between the MUGA WM diagnoses and the three independent blinded read GSPECT WM diagnoses to a single comparison between MUGA and majority read Myoview GSPECT.

TABLE(1)
AGREEMENT LEVELS FOR GSPECT WM READS

STUDY	Reader1vs Reader2		Reader1vs Reader3		Reader2 vs Reader3	
	AGREE	KAPPA	AGREE	KAPPA	AGREE	KAPPA
MYO301	.81	.61	.82	.64	.78	.57
MYO303	.89	.78	.90	.80	.90	.80
	Reader1 vs Majority Read		Reader2 vs Majority Read		Reader3 vs Majority Read	
	AGREE	KAPPA	AGREE	KAPPA	AGREE	KAPPA
MYO301	.93	.86	.88	.76	.90	.81
MYO303	.94	.88	.95	.89	.90	.79

2.3 Statistical Reviewer's Findings:

As explained above, the Sponsor provided no criteria for success for the proposed use of Myoview GSPECT as a substitute for MUGA for evaluation of the functional endpoints of LVEF and WM. The statistical reviewer proposes the following criteria:

Myoview GSPECT categorical diagnoses of LVEF and WM can serve as a substitute for MUGA diagnoses of LVEF and WM if the differences in Sensitivities and Specificities between Core-Center Myoview GSPECT and Core-Center MUGA (with On-Site MUGA as reference standard) are "close", in the explicit senses defined below:

(a): The 95% two sided confidence interval for the difference Myoview GSPECT LVEF Sensitivity –MUGA LVEF Sensitivity has lower bound > -.10.

(b): The 95% two sided confidence interval for the difference Myoview GSPECT LVEF Specificity –MUGA LVEF Specificity has lower bound > -.10.

(c): The 95% two sided confidence interval for the difference Myoview GSPECT WM Sensitivity –MUGA WM Sensitivity has lower bound > -.10.

(d): The 95% two sided confidence interval for the difference Myoview GSPECT WM Specificity –MUGA WM Specificity has lower bound > -.10.

The statistical analyses below provide confidence intervals consistent with these criteria. The relevant confidence intervals are presented after Table(4) and Table(5).

TABLE(2)
LVEF Statistics for Individual and Combined Trials

	LVEF SENSITIVITY			LVEF SPECIFICITY		
	N	SPECT	MUGA	N	SPECT	MUGA
STUDY301	71	.77	.86	48	.96	.90
STUDY303	92	.88	.86	58	.91	.90
COMBINED	163	.83	.86	106	.93	.90

TABLE(3)
WM Statistics for Individual and Combined Trials

	LVEF SENSITIVITY			LVEF SPECIFICITY		
	N	SPECT	MUGA	N	SPECT	MUGA
STUDY301	51	.94	.92	64	.84	.84
STUDY303	59	.90	.75	90	.86	.89
COMBINED	110	.92	.83	154	.85	.87

Table(4) below details the direct comparisons of Blinded MUGA EF diagnoses to Blinded Myoview SPECT EF diagnoses for the Combined trials, with On-Site MUGA EF as the Reference Standard.

TABLE(4)
Combined Trials Results for LVEF : Myoview GSPECT vs Core-Center MUGA
On-Site MUGA as Reference Standard
Cell Entries are Numbers of Patients

	On-Site MUGA EF = 0		On-Site MUGA EF = 1	
	SPECT EF= 0	SPECT EF = 1	SPECT EF= 0	SPECT EF = 1
Blinded MUGA EF = 0	91	4	8	15
Blinded MUGA EF = 1	8	3	19	121

Results from Table(4):

Sensitivity of Blinded MUGA EF = .86

Sensitivity of Blinded Myoview GSPECT EF = .83

The 95% confidence interval for the difference is (-.095 , +.045)

Specificity of Blinded MUGA EF = .93

Specificity of Blinded Myoview GSPECT EF = .90

The 95% confidence interval for the difference is (- .026 , + .101)

Table(5) below details the direct comparisons of Blinded MUGA WMEF diagnoses to Blinded Myoview SPECT WM diagnoses, with On-Site MUGA WM as the Reference Standard.

TABLE(5)
Combined Trials Results for WM : Myoview GSPECT vs Core-Center MUGA
On-Site MUGA as Reference Standard
Cell Entries are Numbers of Patients

	On-Site MUGA WM = 0		On-Site MUGA WM = 1	
	SPECT WM=0	SPECT WM =1	SPECT WM= 0	SPECT WM = 1
Blinded MUGA WM = 0	119	15	6	13
Blinded MUGA WM = 1	12	8	3	88

Results from Table(5):

Sensitivity of Blinded MUGA WM = .83

Sensitivity of Blinded Myoview GSPECT WM = .92

The 95% confidence interval for the difference is (+.022 , +.160)

Specificity of Blinded MUGA WM = .87

Specificity of Blinded Myoview GSPECT WM = .85

The 95% confidence interval for the difference is (-.086 , +.047)

Note: The reviewer's analyses required comparisons of Core-Center data with On-Site data. In order that these analyses be carried out, the Sponsor's On-Site data was restricted to patients for whom On-Site data was available. It is important that this slightly reduced data set provide Core-Center statistics sufficiently close to the Core-Center statistics provided by the Sponsor for the larger data set. Tables presenting comparisons for these statistics are provided in the Appendix.

2.4 Conclusions and Recommendations

Core-Center Myoview GSPECT Imaging provides Sensitivities and Specificities for categorical diagnoses of the functional endpoints of LVEF and WM comparable to and not inferior to Core-Center MUGA Sensitivities and Specificities, when both of these modalities are evaluated with respect to the reference standard of on-site MUGA diagnoses. On-Site MUGA, while not an independent standard of truth, nonetheless provides a means by which the variations in MUGA diagnoses themselves can be assessed, in this case through direct comparisons of its diagnoses with Core-Center MUGA diagnoses. The analyses provided in this review can then be interpreted as exploratory evidence that Myoview GSPECT diagnoses for LVEF and WM are consistent with MUGA diagnoses for LVEF and WM to fairly much the same extent that independent MUGA diagnoses are consistent with one another. These results, though arrived at through post-hoc analyses, are suggestive of statistically significant levels of agreement between Myoview GPSECT diagnoses and MUGA diagnoses for the categories Normal/Abnormal for the functional endpoints of LVEF and WM. It is to be stressed, however, that the Sponsor has not provided hypotheses for agreement levels between GSPECT and MUGA, and that, given the fact that MUGA is itself a less than perfect standard for these functional endpoints, appropriate tests for agreement between these modalities, namely tests which are evaluated with respect to a reliable truth standard, remain somewhat elusive.

A G Mucci Ph. D.
Statistical Reviewer

Concur: Dr Welch

cc: HFD-160/S Loewke/P Stewart

Appendix –

The tables below provide comparisons between the Sponsor's and the Reviewer's Core-Center LVEF and WM statistics . (the Core-Center statistics are the statistics which use Core-Center MUGA as the reference standard; On-Site MUGA is not involved.) Since the reviewer's exploratory approach was confined to the subset of patients for whom On-Site data was available, it was important that the relevant Core-Center statistics for these patients be similar. (Note: The Sponsor's statistics, in each study, are the averages over the three blinded readers.)

Table (A1)
Comparisons of Sponsor vs Reviewer LVEF Statistics

	MYO301			MYO303	
	Sponsor (N=127)	Reviewer (N=119)		Sponsor (N=168)	Reviewer (N=150)
Sensitivity	.87	.89		.81	.80
Specificity	.77	.77		.84	.86
Agreement	.82	.82		.83	.83
Prevalence	.45	.45		.40	.43

Table (A2)
Comparisons of Sponsor vs Reviewer WM Statistics

	MYO301			MYO303	
	Sponsor (N=124)	Reviewer (N=115)		Sponsor (N=166)	Reviewer (N=149)
Sensitivity	.81	.83		.85	.89
Specificity	.75*	.84		.80	.81
Agreement	.81	.83		.82	.84
Prevalence	.49	.50		.37	.36

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
2/27/03 12:13:59 PM
BIOMETRICS

S. Edward Nevius
2/27/03 02:40:36 PM
BIOMETRICS
Concur with review.