CENTER FOR DRUG EVALUATION AND RESEARCH

Approval Package for:

Application Number: 20420

Trade Name: GenESA

Generic Name: Arbutamine

Sponsor: Gensia Automedics

Approval Date: September 12, 1997

CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION: 20420

CONTENTS

	Included	Pending Completion	Not Prepared	Not Required
Approval Letter	X			
Tentative Approval Letter				X
Approvable Letter	X			
Final Printed Labeling	X			
Medical Review(s)	X			
Chemistry Review(s)	X			
EA/FONSI	X			
Pharmacology Review(s)	X			
Statistical Review(s)	X			
Microbiology Review(s)	X			
Clinical Pharmacology Biopharmaceutics Review(s)	X			
Bioequivalence Review(s)				X
Administrative Document(s)	X			
Correspondence				

CENTER FOR DRUG EVALUATION AND RESEARCH

Application Number: 20420

APPROVAL LETTER

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

Food and Drug Administration
Rockville MD 20857

NDA 20-420

SEP 12 1997

Gensia Automedics, Inc.
Attention: Ms. Cynthia Luchetti
9360 Towne Centre Drive
San Diego, CA 92121

Dear Ms. Luchetti:

Please refer to your December 20, 1993 new drug application (NDA) submitted under section 505(b) of the Federal Food, Drug, and Cosmetic Act for GenESA (arbutamine hydrochloride injection) System.

We acknowledge receipt of your correspondence and amendments dated May 16, June 2, 19, 24, 25, 27 and 30, July 16, August 6 and September 4 (two), 1997.

This new drug application provides for the use of the GenESA System as an aid in diagnosing the presence or absence of coronary artery disease in patients who cannot exercise adequately when used in conjunction with radionuclide myocardial perfusion imaging or echocardiography.

We have completed the review of this application and have concluded that adequate information has been presented to demonstrate that the drug is safe and effective for use as recommended in the final printed labeling included with your July 16, 1997 submission. Accordingly, the application is approved effective on the date of this letter.

Please make the following changes to your labels and labeling at your next printing:

1. The product name should be expressed as follows on all labeling and labels:

   GenESA (arbutamine hydrochloride injection)
   0.05 mg/mL
   for Intravenous Infusion with the GenESA Device

2. The following sentence should be added at the end of the **DOSAGE AND ADMINISTRATION** section of the labeling:

   > Parenteral drug products should be inspected visually for particulate matter and discoloration prior to administration, whenever the solution and container permit.

   Please delete the incomplete statement under **Syringe and Plunger Rod Assembly** (second bullet).

Please submit one market package of the drug when it is available.

We remind you that you must comply with the requirements for an approved NDA set forth under 21 CFR 314.80 and 314.81.

If you have any questions, please contact:

Mr. Gary Buehler
Regulatory Health Project Manager
(301) 594-5332

Sincerely yours,

Robert Temple, M.D.
Director
Office of Drug Evaluation I
Center for Drug Evaluation and Research

cc:
Original NDA
HF-2/MedWatch (with draft/final labeling)
HFD-2/MLumpkin
HFD-92 (with draft/final labeling)
HFD-101 (with draft/final labeling)
HFD-110
HFD-40 (with draft/final labeling)
HFD-613 (with draft/final labeling)
HFD-735 (with draft/final labeling)
HFD-21 (with draft/final labeling - for drugs discussed at advisory committee meeting)
DISTRICT OFFICE
HFD-810/New Drug Chemistry Division Director
HFD-110/GBuehler/7/17/97;7/18/97
sb/7/17/97;7/21/97
R/D: JShort/7/18/97
RWolters/7/18/97
NOza/7/18/97
ADeFelice/7/18/97
PMarroum/7/18/97
KMhjoob/7/18/97
RFenichel/7/18/97
NMorgenstern/7/18/97

APPROVAL

CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER: 20420

APPROVABLE LETTER

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

Food and Drug Administration
Rockville MD 20857

NDA 20-420
PMA P940001

Gensia, Inc.
Attention: Ms. Cynthia Luchetti
9360 Towne Centre Drive
San Diego, CA 92121

MAY 12 1997

Dear Ms. Luchetti:

Please refer to your December 20, 1993 new drug application (NDA) and premarket approval application (PMA) submitted under section 505(b) and 515(c), respectively, of the Federal Food, Drug, and Cosmetic Act for GenESA (arbutamine) System.

We acknowledge receipt of your correspondence and amendments dated May 30, June 10, 20 and 21, August 13, September 16, November 8 and 15, December 2 (two), 10 and 13, 1996 and January 28, March 10, 11, 18, 20 and 24 and April 21, 1997.

We have completed the review of these applications as submitted with draft labeling and they are approvable. Before the applications may be approved, however, it will be necessary for you to submit the following information:

1. In your November 8, 1996 response you indicated that the

2. On page 85 of volume 7 is a table summarizing the data collected for the GenSEA 1 ml/hr flow rate test report. Please explain how the rate error is calculated and provide the unit (i.e. percentage, ml/hr, etc.) associated with the error.

3. The directions for operating the device do seem to be adequate. We recommend, however, that the descriptive section on arbutamine (including the contraindications, warning and dosage recommendations) be removed and replaced with references to the drug labeling instead. By making references to the drug labeling, future revisions to the drug labeling would not require the device instruction manual to be updated.

4. You have provided a summary of the changes made to the system hardware, system software, closed loop algorithm software and the IV pump. According to the description provided on page 17, 21 and 22 of volume 7, the closed loop algorithm software was at revision 5 in the original NDA submission (December 20, 1993) and no further modification was made since that time. This description appears to be inconsistent with your November 8, 1996, response to question 6c which discusses the upgrade from revision 4 to revision 5 based on more recent study results (volume 7, page 291). Please clarify.

In addition, it will be necessary for you to submit final printed labeling (FPL) for the system. The labeling should be identical in content to the enclosed marked-up draft. If additional information relating to the safety or effectiveness of this system becomes available, revision of the FPL may be required. You will note that we have not, at this time, agreed with your proposal to label the GenESA System for use with echocardiography. This does not reflect a view that echocardiography is not used or perceived as useful with exercise or other stress testing. The only data we have with well-defined and blindly (with respect to angiography) read echocardiograms, however, shows very low (and poorly defined) specificity (study 123) or sensitivity (study 141). We continue to believe that blinded readings with well-defined endpoints are critical to an interpretation of the value of the GenESA-echo combination. We understand that you may want to consider this issue further.

Please submit sixteen copies of the printed labels and other labeling, ten of which are individually mounted on heavy weight paper or similar material.
In addition, please submit three copies of the introductory promotional material that you propose to use for this system. All proposed materials should be submitted in draft or mock-up form, not final print. Please submit one copy to the Division of Cardio-Renal Drug Products and two copies of both the promotional material and the package insert directly to:

Food and Drug Administration
Division of Drug Marketing, Advertising and Communications, HFD-40
5600 Fishers Lane
Rockville, Maryland 20857

Within 10 days after the date of this letter, you are required to amend these applications, notify us of your intent to file an amendment, or follow one of your other options under 21 CFR 314.110. In the absence of such action FDA may take action to withdraw these applications.

The system may not be legally marketed until you have been notified in writing that the applications are approved.

Should you have any questions, please contact:

Mr. Gary Buehler
Regulatory Health Project Manager
Telephone: (301) 594-5332

Sincerely yours,

Robert Temple, M.D.
Director
Office of Drug Evaluation I
Center for Drug Evaluation and Research

Enclosure

CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER: 20420

FINAL PRINTED LABELING

Y36-105-G02

GenESA®

arbutamine HCl 0.05 mg/mL

Sterile Solution for Intravenous Infusion
with the GenESA® Device

DESCRIPTION

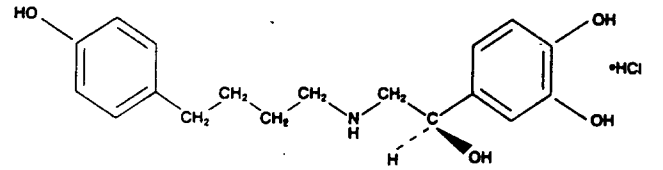
GenESA® (arbutamine hydrochloride, sterile solution for intravenous infusion, 0.05 mg/mL) is a synthetic catecholamine with chronotropic and inotropic properties. Chemically, arbutamine hydrochloride is (R)-4-[1-hydroxy-2-[[4-(4-hydroxyphenyl)-butyl]amino]ethyl]-1,2-benzenediol hydrochloride.

Arbutamine hydrochloride is an off-white amorphous solid, which is freely soluble in water and ethanol, but is practically insoluble in diethyl ether and hexane. It has the following structural formula:

CHEMICAL STRUCTURE

Molecular formula:
$C_{18}H_{23}NO_4 \cdot HCl$

Molecular weight:
365.85



GenESA (arbutamine hydrochloride, sterile solution for intravenous infusion, 0.05 mg/mL) is formulated in an isotonic, buffered vehicle (pH 3.8) in a 20 mL prefilled syringe.

Active Ingredient: Arbutamine Hydrochloride 0.05 mg/mL.

Inactive Ingredients: Sodium Metabisulfite, NF 0.10 mg/mL, Citric Acid Monohydrate, USP 1.40 mg/mL, Trisodium Citrate Dihydrate, USP 0.88 mg/mL, Sodium Chloride, USP 8.50 mg/mL, Disodium Edetate, USP 0.10 mg/mL, and Water for Injection, USP q.s. ad 1.0 mL. The air in the prefilled syringe has been replaced by Nitrogen, NF.

The GenESA® System comprises the cardiac stress agent GenESA (arbutamine hydrochloride) and a drug delivery device, the GenESA® Device. GenESA is intended for direct intravenous infusion ONLY with the GenESA Device. The "GenESA System Directions for Use", a detailed instruction manual provided with each GenESA Device, provides an overview of the GenESA System, full details on how to conduct a pharmacological stress test using the GenESA System, and detailed information on the operation and function of the GenESA Device (see also DOSAGE AND ADMINISTRATION).

CLINICAL PHARMACOLOGY

Mechanism of Action

By increasing cardiac work through its positive chronotropic and inotropic actions, arbutamine acts as a cardiac stress agent to mimic exercise and provoke myocardial ischemia in patients with compromised coronary arteries. It also probably limits regional subendocardial perfusion, and hence tissue oxygenation, by its increase in heart rate (HR). The delivery system adjusts the rate of arbutamine delivery to achieve a selected increase in heart rate.

Arbutamine is a sympathomimetic that exhibits increased selectivity for β-adrenoceptors over α-adrenoceptors in functional assays. The β-agonist activity of arbutamine provides cardiac stress by increasing HR, cardiac contractility and systolic blood pressure.

Some α-receptor activity is retained in vivo, such that the degree of hypotension observed for given chronotropic activity is less with arbutamine than with isoproterenol, a specific β-agonist.

Effects on HR and Blood Pressure

In clinical studies of patients (N=494) with known or suspected coronary artery disease (CAD), GenESA was delivered using the GenESA Device to achieve an identified maximum HR unless an endpoint (pain, ECG changes) was attained sooner. Under these conditions the mean maximum increase in HR was 52 bpm and the mean maximum increase in systolic blood pressure was 36 mmHg.

After termination of arbutamine infusion, HR decreased, with 50% of the HR increase gone by 16 minutes (N=315).

The GenESA System can be programmed to give different rates of HR increase (HR Slope). The maximum increases in HR and systolic blood pressure were independent of the selected rate of HR increase. In patients who underwent a formal comparison of GenESA and exercise testing, effects on HR and BP were similar.

The effects of arbutamine on HR and systolic blood pressure are attenuated by concurrent administration of selective and non-selective β-blockers. Depending on the β-blocker, evidence of this attenuation is still present 23 hours after the last dose. In patients in whom β-blockers had been withdrawn for a minimum of 48 hours prior to receiving arbutamine, the HR and systolic blood pressure responses to arbutamine were similar to those seen in patients who had not received β-blockers for at least 2 weeks prior to arbutamine.

In 19 patients (12 with, and 7 without, a stenosis ≥ 50% in 1 or more major coronary arteries) the cardiac hemodynamic effects of arbutamine were assessed using invasive techniques. Cardiac contractility (left ventricular dP/dt), cardiac output (measured by thermodilution) and total systemic vascular resistance (SVR) were determined at low stress (a mean HR increase from baseline of 25 bpm) and at peak stress (a mean HR increase from baseline of 40 bpm). At low stress, cardiac contractility and output had increased by 66% and 45%, respectively, from baseline, and they rose to 80% and 63% above baseline at peak stress. Total SVR decreased from baseline by 41% at low stress and by 48% at peak stress.

By 15 minutes after the end of the infusion, cardiac contractility and SVR were within approximately 15% of baseline, and cardiac output was 29% above baseline.

Left ventricular ejection fraction (LVEF), assessed by echocardiography, was evaluated in a study of 156 patients with known or suspected CAD tested using arbutamine infusion and exercise. With an increase in HR from baseline of approximately 20 bpm, the relative increase in LVEF was 22%. At the end of the arbutamine infusion, when the mean increase in HR from baseline was 59 bpm, the relative increase in LVEF was 23%. By comparison, the relative increase in LVEF at the end of exercise was 14%.

Pharmacokinetics and Metabolism

Because of sensitivity limitations of the arbutamine assay, the pharmacokinetics of arbutamine in humans have been characterized only for the first 20-30 minutes after the termination of intravenous infusions up to 0.3 μg/kg/min. In a study in 12 healthy men, half-life was about 8 minutes, plasma clearance about 4 L/hr/kg, and volume of distribution 0.74 L/kg. Plasma protein binding was approximately 58%. Arbutamine is mainly (>75%) eliminated by metabolism, principally to methoxyarbutamine, which is found in free or conjugated form in urine. Ketoarbutamine has been tentatively identified as another metabolite. Following intravenous infusion of 14C-arbutamine to healthy men, 84% of the total radioactivity was excreted in the urine within 48 hours, with 9% excreted in feces. Total radioactivity in plasma declined with a half-life of 1.8 hours, probably due to metabolites with longer half-lives than arbutamine. The rapid onset of effect on heart rate after start of arbutamine infusion (approximately 1 minute) and the rapid decline in heart rate following termination of infusion of arbutamine (time for a 50% decrease is 13-16 minutes) suggest that the pharmacological activity resides primarily with the parent compound and not with any of its more long-lived metabolites. Pharmacokinetics have not been characterized in women, the elderly, or different racial groups.

Clinical Trials

The usefulness of diagnostic tests can be defined in various ways. Measures of sensitivity (ability of test to identify diseased patients, in this case the rate of positive stress tests in patients with positive angiograms, or true positives divided by true positives plus false negatives) and specificity (ability of test to identify people without disease, in this case the rate of negative stress tests in patients with negative angiograms, or the true negatives divided by true negatives plus false positives) are frequently used. The problem is that the usefulness of a test can depend not only on sensitivity and specificity but on prevalence of the disease. Thus, for example, even a very sensitive test will be of minimal use in a population where almost all patients have the disease; for example, if 100% of patients have the disease, even 90% sensitivity will mean an "error rate" (declaring no disease when disease was present) in 10% of patients. In addition to sensitivity and specificity, therefore, tests are often described in terms of positive predictive fraction (the rate of correctness of a positive test) and negative predictive fraction (the rate of correctness of a negative test).

In clinical studies, patients underwent coronary angiography and GenESA System testing with radionuclide perfusion imaging (using thallium-201 or technetium-99m sestamibi) or with echocardiography. For purposes of these studies, an angiogram was considered positive if it demonstrated at least one ≥ 50% diameter stenosis of a major coronary artery; the GenESA System test was considered positive if perfusion defects were seen by radionuclide imaging or if wall motion abnormalities were noted on echocardiography, at baseline or during stress.

First for perfusion imaging and second for echocardiography, the following discussion gives both (1) a sensitivity/specificity/positive and negative predictive value analysis for two studies (one in patients with a high risk for CAD, one with a lower risk) and (2) an overall analysis that relates the information provided by the test relative to a prior estimate (based on a standard algorithm) of the likelihood of CAD being present.

1. Perfusion Imaging: Sensitivity/Specificity

The ability of radionuclide tests to predict the results of coronary angiography was assessed in 234 patients enrolled in two studies. In the high-risk study, patients were selected based on coronary angiographic evidence of CAD obtained within 12 weeks prior to the GenESA System test with thallium imaging. Patients were also included if coronary angiography was scheduled within 4 weeks following the GenESA System test. All studies were read without knowledge of other results. In the lower-risk study, patients were selected if coronary angiography had been performed within 12 weeks before or after the GenESA System test with thallium or sestamibi imaging, and results were re-read blindly after the study to give a similar assessment of the test. The results are shown in Table 1.

Table 1
Sensitivity, Specificity and Predictive Fractions for Radionuclide Imaging with the GenESA System

Study	Sensitivity	Specificity	Positive Predictive Fraction	Negative Predictive Fraction
High Risk	97/112 (87%)	2/8 (25%)	97/103 (94%)	2/17 (12%)
Lower Risk	51/81 (63%)	21/33 (64%)	51/63 (81%)	21/51 (41%)
Thallium	10/16 (63%)	7/12 (58%)	10/15 (67%)	7/13 (54%)
Sestamibi	41/65 (63%)	14/21 (67%)	41/48 (85%)	14/38 (37%)

Note that although sensitivity and specificity are in general independent of prevalence, it is possible that in this case prevalence (or, more likely, the presence of various factors related to CAD) does influence the test results, e.g., by giving more false positives in the high risk group (and thus lower specificity).

Note also that in a very high-risk group, use of the test may give less than satisfactory overall advice although (see next section) the ability to predict the results of angiography in any given patient may be improved. Thus, even with a high sensitivity of 87%, 15/112 patients with CAD were identified as not having it, and of 8 patients without CAD, 6 were identified as having it. The test is most helpful where the likelihood that the patient has arterial disease is neither very high nor very low.

2. Perfusion Imaging: Predictive Value of the GenESA System Test

Another approach to considering results of GenESA System testing is to describe the impact of the test result on the estimated likelihood of CAD based on the patients' defined risk, utilizing all available data about the patient. Using an algorithm developed by Pryor, DB, et al. (Am J Med 1983; 75:771-80), the 233 patients with demographic data available who underwent coronary angiography and GenESA System testing assessed with perfusion imaging were categorized as having a low (<20%), intermediate (20-80%) or high (>80%) likelihood of CAD. The characteristics of the three groups are summarized in Table 2 below.

Table 2
Characteristics of 233 Patients Who Each Underwent Coronary Angiography and had a GenESA System Test Assessed with Radionuclide Imaging

Pretest CAD likelihood Group called % (N) of Patients	<20% "Low" 4% (9)	20-80% "Intermediate" 21% (50)	>80% "High" 75% (174)
Age 65 years	11% (1)	18% (9)	36% (63)
Male	22% (2)	64% (32)	86% (150)
Typical Angina	0% (0)	14% (7)	74% (128)
Atypical Angina	11% (1)	25% (50)	20% (34)
Hyperlipidemia	33% (3)	54% (27)	59% (102)
Diabetes	33% (3)	16% (8)	29% (51)
Smoking	33% (3)	32% (16)	45% (78)
Prior MI	0% (0)	18% (9)	51% (88)
MI on ECG	0% (0)	0% (0)	20% (35)
ST-T Abnormality	22% (2)	16% (8)	30% (52)
# Patients with:			
1 Risk Factor	33% (3)	6% (3)	0% (0)
2 Risk Factors	67% (6)	28% (14)	6% (10)
3 Risk Factors	0% (0)	44% (22)	18% (32)
4 Risk Factors	0% (0)	22% (11)	29%(51)
≥5 Risk Factors	0% (0)	0% (0)	47% (81)
Mean (±SD) # Risk Factors/Patient	1.7±0.5	2.8±0.8	4.5±1.4
Angiography positive	33% (3)	56% (28)	78% (135)

As summarized in Table 3 below (and as would be seen with any other less-than-perfect test), the performance of the GenESA System varied from one subgroup to another; it was most uniformly accurate in patients with an intermediate pre-test likelihood of disease.

Table 3
Predictive Value of GenESA System Testing when used with Radionuclide Imaging

Pretest probability of positive angiogram	N	Positive GenESA Test	Positive GenESA test was correct (angiogram positive)	Negative GenESA Test	Negative GenESA test was correct (angiogram negative)
Low	9	3 (33%)	0 (0%)	6 (67%)	6 (100%)
Intermediate	50	28 (56%)	22 (79%)	22 (44%)	10 (46%)
High	174	135 (78%)	126 (93%)	39 (22%)	7 (18%)

It is difficult for any diagnostic test to contribute information when the pretest probability of disease is extremely low or extremely high. As the pretest likelihood gets higher and higher, a positive test result provides a smaller and smaller increment of information, while a negative test result is more and more likely to be a false negative. Conversely, as the pretest likelihood of disease approaches zero, positive test results are more and more likely to be false positives. These considerations are of course applicable to all diagnostic tests, not just to the GenESA System.

To interpret the data another way, one can estimate the post-test likelihood of CAD, given the pre-test likelihood and the result of a GenESA System test (Diamond GA, et al., NEJM 1979: 300;1350-58). These results are shown in Table 4 for perfusion imaging and confirm the general discussion of the previous paragraph.

Table 4
Post-Test Likelihood of Coronary Artery Disease Given the Pre-Test Likelihood and the Result of GenESA System Testing Assessed with Radionuclide Imaging

Pre-Test Likelihood	Post-Test Likelihood	
	with positive GenESA test	with negative GenESA test
10%	16	4
20%	30	9
30%	43	15
40%	54	22
50%	64	29
60%	72	38
70%	80	49
80%	87	62
90%	94	79

3. Echocardiography: Sensitivity/Specificity

The ability of GenESA System echocardiography tests to predict the results of coronary angiography was assessed in two studies, involving a total of 389 patients. Patients were selected as in the radionuclide studies. The blinded re-reading of the results from the lower-risk study was technically inadequate, and the results shown in Table 5 for that study are based on non-blinded readings.

Table 5
Sensitivity, Specificity and Predictive Fractions for Echocardiography with the GenESA System

Study	Sensitivity	Specificity	Positive Predictive Fraction	Negative Predictive Fraction
High Risk	110/131 (84%)	4/16 (25%)	110/122 (90%)	4/25 (16%)
Lower Risk	137/194 (71%)	32/48 (67%)	137/153 (90%)	32/89 (36%)

4. Echocardiography: Predictive Value of the GenESA System Test

The 381 patients with demographic data available who underwent coronary angiography and GenESA System testing assessed with echocardiography were categorized as was done for perfusion imaging and the characteristics of the three groups are summarized in Table 6 below.

Table 6
Characteristics of 381 Patients Who Each Underwent Coronary Angiography and had a GenESA System Test Assessed with Echocardiography

Pretest CAD likelihood Group called % (N) of Patients	<20% "Low" 3% (13)	20-80% "Intermediate" 57% (103)	>80% "High" 70% (265)
Age 65 years	8% (1)	23% (24)	38% (101)
Male	15% (2)	53% (56)	89% (235)
Typical Angina	0% (0)	18% (19)	79% (210)
Atypical Angina	62% (8)	51% (53)	15% (40)
Hyperlipidemia	46% (6)	57% (60)	62% (165)
Diabetes	0% (0)	16% (17)	20% (53)
Smoking	23% (3)	31% (33)	38% (101)
Prior MI	0% (0)	15% (16)	53% (140)
MI on ECG	0% (0)	0% (0)	28% (73)
ST-T Abnormality	23% (3)	21% (22)	37% (99)
# Pabents with:			
1 Risk Factor	31% (4)	5% (5)	0% (0)
2 Risk Factors	62% (8)	31% (32)	2% (5)
3 Risk Factors	8% (1)	40% (41)	22% (58)
4 Risk Factors	0% (0)	22% (23)	26% (71)
≥5 Risk Factors	0% (0)	2% (2)	49% (131)
Mean (±SD) # Risk Factors/Patient	1.8±0.6	2.9±0.9	4.6±1.3
Angiography positive	39% (5)	55% (57)	78% (207)

As with perfusion imaging and summarized in Table 7 below, the performance of the GenESA System varied from one subgroup to another; it was most uniformly accurate in patients with an intermediate pre-test likelihood of disease.

Table 7
Predictive Value of GenESA System Testing when used with Echocardiography

Pretest probability of positive angiogram	N	Positive GenESA Test	Positive GenESA test was correct (angiogram positive)	Negative GenESA Test	Negative GenESA test was correct (angiogram negative)
Low	13	5 (38%)	2 (40%)	8 (62%)	7 (88%)
Intermediate	103	57 (55%)	47 (83%)	46 (45%)	21 (46%)
High	265	207 (78%)	193 (93%)	58 (22%)	7 (12%)

It is difficult for any diagnostic test to contribute information when the pretest probability of disease is extremely low or extremely high (see earlier discussion of perfusion imaging results). As for perfusion imaging, the post-test likelihood of CAD, given the pre-test likelihood and the result of a GenESA System test, was estimated. These results are shown in Table 8 and confirm the general discussion of the previous paragraph.

Table 8
Post-Test Likelihood of Coronary Artery Disease Given the Pre-Test Likelihood and the Result of GenESA System Testing Assessed with Echocardiography

Pre-Test Likelihood	Post-Test Likelihood	
	with positive GenESA test	with negative GenESA test
10%	16	5
20%	30	10
30%	43	15
40%	54	22
50%	63	30
60%	72	39
70%	80	50
80%	87	63
90%	94	79

INDICATIONS AND USAGE

The GenESA System delivers arbutamine, a catecholamine, through a closed-loop, computer-controlled drug-delivery system to elicit acute cardiovascular responses similar to those produced by exercise. In patients with suspected coronary artery disease (CAD) who cannot exercise adequately, stress induction with the GenESA System is indicated as an aid in diagnosing the presence or absence of CAD.

The effectiveness of the GenESA System has been demonstrated in clinical studies using radionuclide myocardial perfusion imaging to predict the results of coronary angiography. These studies were in patients with high and lower risks of CAD and utilized blinded, central reading of images. Estimates of sensitivity, specificity and predictive values are presented in the "Clinical Trials" section.

Although the effectiveness of the GenESA System was also assessed in similar clinical studies utilizing echocardiography to predict the results of coronary angiography, the blinded, central reading of the images from the lower-risk echocardiography study was technically inadequate. Estimates of sensitivity, specificity and predictive values, based on the non-blinded readings of echocardiograms at the local study sites, are presented for the lower-risk patients (see Clinical Trials). For the study of high-risk patients, the estimates are based on valid, blinded, central reading of images.

Like exercise testing, cardiac stress testing with the GenESA System must always be performed under the direct supervision of a physician, and cardiac emergency equipment and supplies (defibrillator, intravenous β-blocker, etc.) must always be available. Arbutamine must not be administered without use of the GenESA Device.

CONTRAINDICATIONS

Arbutamine is contraindicated in patients with idiopathic hypertrophic subaortic stenosis, in patients with a history of recurrent sustained ventricular tachycardia, in patients with congestive heart failure (NYHA Class III or IV), and in patients who have shown previous manifestations of hypersensitivity to arbutamine. The GenESA System must not be used in the presence of an implanted cardiac pacemaker or automated cardioverter/defibrillator.

WARNINGS

During clinical trials that included 2082 patients with known or suspected coronary artery disease, arbutamine administration was associated with 10 serious adverse events, including 3 episodes of ventricular fibrillation, 1 episode of sustained ventricular tachycardia, 3 episodes of atrial fibrillation (see Table 10 for a summary of all arrhythmias reported as adverse events), 1 myocardial infarction and 2 cases of severe angina. Two of the three cases of ventricular fibrillation occurred after the GenESA Device indicated a plateau in HR response and had terminated arbutamine infusion, but the physician restarted the infusion. There were no deaths.

The incidence of serious adverse events is thus low, less than 0.5%. Nevertheless, the potential information to be gained through the use of arbutamine, delivered using the GenESA Device (see INDICATIONS AND USAGE), must be weighed against the potential risks to each patient.

Arbutamine may precipitate or exacerbate supraventricular and ventricular arrhythmias and its administration is not recommended in patients with a history of sustained arrhythmias of this nature. Given the proarrhythmic effects of certain antiarrhythmic drugs, particularly Class 1 agents such as quinidine, lidocaine and flecainide, arbutamine should not be administered to patients receiving such therapy.

Supraventricular or ventricular arrhythmias can occur during the administration of arbutamine (see ADVERSE REACTIONS) with isolated premature ventricular and atrial contractions being the most frequent arrhythmias. Most arrhythmias were self-limiting and all resolved without sequelae. If any arrhythmias are of clinical concern, drug infusion should be discontinued immediately and appropriate therapy (e.g. intravenous β-blockers - see OVERDOSAGE) administered, if necessary. The GenESA Device is not designed to detect arrhythmias. Appropriate monitoring equipment, such as a diagnostic quality ECG machine, must therefore be used during a GenESA System test. The GenESA Device administers arbutamine based upon HR response and it is possible that, in the presence of an arrhythmia, the GenESA Device may register an inaccurate HR. The ECG should be monitored carefully and appropriate action, including, if necessary, discontinuation of drug infusion, taken in the event of inaccurate HR detection.

Arbutamine may cause rapid increases or paradoxical decreases in HR and systolic blood pressure. Discontinuation of arbutamine infusion results in reversal of these effects. The infusion may be restarted, if considered clinically appropriate (see DOSAGE AND ADMINISTRATION).

The safety of arbutamine administration in patients with recent (within 30 days) myocardial infarction has not been formally evaluated. The administration of arbutamine is not recommended in patients with unstable angina, mechanical left ventricular outflow obstruction such as severe valvular aortic stenosis, uncontrolled systemic hypertension, a cardiac transplant, a history of cerebrovascular accident or peripheral vascular disorder resulting in cerebral or aortic aneurysm. In addition, arbutamine is not recommended in patients with narrow-angle glaucoma or uncontrolled hyperthyroidism.

Arbutamine should not be administered to patients receiving digoxin, atropine (or other anticholinergic drugs) or tricyclic antidepressants. As the dosing of arbutamine is based on the HR response of the patient, the use of atropine to enhance the chronotropic response to arbutamine is not recommended.

Reactions suggestive of hypersensitivity have been reported occasionally with the administration of other catecholamines (such as Dobutrex® (dobutamine)). Like other parenterally administered catecholamines, GenESA contains sodium metabisulfite, a sulfite that may cause allergic-type reactions, including anaphylactic symptoms and life-threatening or less severe asthmatic episodes in certain susceptible individuals. The overall prevalence of sulfite sensitivity in the general population is unknown and probably low. Sulfite sensitivity is seen more frequently in asthmatic than nonasthmatic individuals. No such reactions have been reported with arbutamine.

PRECAUTIONS-(See WARNINGS)

During the administration of arbutamine, as with any parenteral catecholamine, ECG and blood pressure should be continuously monitored. The GenESA Device provides such monitoring capabilities but a diagnostic-quality ECG machine must also be used to monitor the ECG.

Like other catecholamines, arbutamine can produce a transient reduction in serum potassium concentration, rarely to hypokalemic levels. In one study, the transient decrease in serum potassium after arbutamine was greater in patients with arrhythmias (N=169) than in those without arrhythmias (N=72).

Changes in serum potassium in patients with clinically significant arrhythmias were not clearly different from those seen in other patients.

As with other catecholamines, GenESA infusion is associated with a transient increase in corrected QT interval, as measured from the surface ECG. This effect did not appear to be associated with an increased incidence of arrhythmias.

The acute use of the GenESA System for diagnostic testing makes it unlikely that alterations in renal and /or hepatic function will influence the safety and diagnostic efficacy of a GenESA System test.

Carcinogenesis, Mutagenesis, Impairment of Fertility

Arbutamine is intended for single-dose use only and therefore animal carcinogenicity or long-term toxicity studies have not been performed.

Arbutamine was shown to be non-genotoxic in the Ames bacterial reverse mutation assay, with and without S9 mix, and in the mouse micronucleus test. Arbutamine was positive in the human lymphocyte chromosomal aberration assay (>66 μg/mL) and in the mouse lymphoma cell assay (>39 μg/mL).

Studies to determine the effect of arbutamine on the impairment of fertility have not been performed.

Pregnancy: Teratogenic Effects

Pregnancy Category B

Reproduction studies performed in rats and in rabbits at doses up to 0.9 and 0.36 mg/kg/day I.V., respectively (4 and 12 times the maximum recommended human dose on a mg/m² basis), revealed no evidence of harm to the fetus. There are, however, no adequate and well-controlled studies in pregnant women. Because animal reproduction studies are not always predictive of human response, arbutamine should be used during pregnancy only if clearly needed.

Drug Interactions

Beta-adrenergic antagonists may attenuate the response to arbutamine and should be withdrawn, as recommended in the relevant product labeling, at least 48 hours before conducting a GenESA System test. There was no evidence of drug-drug interactions in clinical studies in which arbutamine was administered concurrently with other drugs, including platelet aggregation inhibitors, nitrates, and calcium channel blockers.

ADVERSE REACTIONS

Adverse events were recorded during controlled clinical trials in 2082 patients with known or suspected coronary artery disease. Serious adverse events (ventricular and atrial fibrillation, and severe cardiac ischemia) are described above (see WARNINGS).

The most frequently reported adverse events in the 2082 patients were: tremor (15%), angina pectoris (12%), non-serious cardiac arrhythmias (12%), headache (9%), and hypotension (6%). Adverse events occurring in ≥ 1% of the 2082 patients are shown in Table 9:

Table 9
Incidence of Most Frequent (≥ 1%) Adverse Events with Arbutamine

	Incidence (%) of Adverse Events
Tremor	15
Angina pectoris	12
Cardiac arrhythmias	12
Ventricular	6
Supraventricular	4
Headache	9
Hypotension	6
Chest pain	4
Dizziness	4
Dyspnea	4
Palpitation	4
Flushing	3
Hot flushes	3
Nausea	3
Paresthesia	2
Anxiety	1.9
Pain (non-specific)	1.8
Increased sweating	1.5
Fatigue	1.3
Taste perversion	1.3
Dry mouth	1.1
Hypoesthesia	1.0
Vasodilation	1.0

Other adverse events, considered at least possibly related to arbutamine administration and occurring in <1% of the 2082 patients, and seen at least twice, are listed by body system.

- Cardiovascular: myocardial ischemia (0.1%) - see WARNINGS, ST segment depression (0.6%), hypertension (0.4%).
- Body as a Whole: asthenia (0.4%), malaise (0.2%), rigors (0.2%), back pain (0.1%).
- Central and Peripheral Nervous System Disorders: twitching (0.3%).
- Gastrointestinal System: abdominal pain (0.1%).
- Psychiatric Disorders: nervousness (0.7%), agitation (0.2%).
- Respiratory System Disorders: coughing (0.2%), bronchospasm (0.1%).
- Other: rash (0.2%), abnormal lacrimation (0.1%), application site reaction (0.1%).

Cardiac arrhythmias were reported as adverse events, if symptomatic or considered clinically significant, by the physician supervising the stress test. Overall cardiac arrhythmias, as identified by the investigator as adverse events, are shown in Table 10:

Table 10
Incidence of Arrhythmias Reported as Adverse Events

	Incidence of Arrhythmias Reported as Adverse Events (N=2082)
Total number of patients	251 (12%)
Ventricular	130 (6.2%)
Ventricular fibrillation	3 (0.1%)
Ventricular tachycardia	37 (1.8%)
Other ventricular*	106 (5.1%)
Supraventricular	79 (3.8%)
Supraventricular tachycardia	39 (1.9%)
Atrial fibrillation	20 (1.0%)
Other supraventricular**	24 (1.2%)
Junctional	16 (0.8%)
Bradycardia	23 (1.1%)
Sinus tachycardia	18 (0.9%)
Heart Block†	3 (0.1%)
Sinus arrhythmia	1 (0.05%)

* Includes premature ventricular contractions (PVCs), couplets, triplets (rate ≤ 100 bpm), multifocal PVCs, ventricular bigeminy/trigeminy and idioventricular rhythm

** Includes premature atrial contractions and atrial arrhythmia (coronary sinus rhythm)

† Includes sinoatrial block and right bundle branch block

NOTE: Patients may have experienced more than one arrhythmia

OVERDOSAGE - (See WARNINGS)

Because arbutamine delivery is controlled by the GenESA Device to give a defined increase in heart rate, overdosage is unlikely to occur. The maximum total dose permitted by the GenESA Device is 10 μg/kg. If overdosage occurs it should be short-lived, as arbutamine is metabolized rapidly, and most effects would be extensions of arbutamine's pharmacologic effects.

The symptoms of toxicity are those of catecholamine excess: tremor, headache, flushing, hypotension, dizziness, paresthesia, nausea, hot flushes, angina, increased sweating and anxiety. The positive chronotropic and inotropic effects of arbutamine on the myocardium may cause tachyarrhythmias, hypertension, myocardial infarction and ventricular fibrillation. If arbutamine is ingested, unpredictable absorption may occur from the mouth and gastrointestinal tract.

Treatment - initial actions include discontinuing administration, establishing an airway and ensuring adequate oxygenation and ventilation. Severe signs or symptoms (angina, tachyarrhythmias, ST segment abnormalities, hypotension) may be successfully treated with an intravenous β-blocker, such as metoprolol, esmolol or propranolol, at I.V. doses of 7.5-50 mg, 10-80 mg and 0.5-2 mg, respectively. Other treatment, such as sublingual nitrates, should be used if considered clinically appropriate. Given the rapid elimination of arbutamine, forced diuresis, peritoneal dialysis, hemodialysis, or charcoal hemoperfusion are unlikely to be required for arbutamine overdosage.

DOSAGE AND ADMINISTRATION

Before using the GenESA System, it is essential to read and understand The GenESA System Directions For Use in addition to this section of labeling. The GenESA System Directions for Use describe the complete operating instructions for the GenESA Device and the delivery of arbutamine.

GenESA (arbutamine hydrochloride, sterile solution for infusion, 0.05 mg/mL) must be administered from the prefilled syringe and must not be diluted or transferred to another syringe. GenESA is intended for direct intravenous infusion ONLY with the GenESA Device.

The GenESA Device comprises a single channel ECG (R wave) detector, a non-invasive blood pressure monitor, computer software (closed-loop algorithm) which controls drug delivery, an intravenous syringe pump, display functions and an operator key pad. The GenESA Device individualizes the dosing regimen of arbutamine according to the HR response of the patient using the closed-loop algorithm. The physician selects the desired rate of HR rise (HR SLOPE: either LOW, 4 bpm/min; MEDIUM, 8 bpm/min; or HIGH, 12 bpm/min); alternatively any value from 4-12 bpm/min may be selected and the maximum HR to be achieved (HR TARGET - estimated by the device as (220-age) 85%, or adjusted manually by the operator) for each patient test. The choice of HR Slope should be based upon the desired duration of the test and the rate of HR rise, judged by the physician, to be most appropriate. Clinical data, obtained using a HR Slope of 8 bpm/min, support the use of this MEDIUM slope in a majority of patients.

Upon starting the test, the GenESA Device administers a small dose of arbutamine (0.1 μg/kg/min for 1 minute) and measures the patient's HR response. The device then calculates the difference between the desired and actual HR response, and maintains or modifies, as necessary, the infusion rate. The maximum infusion rate delivered by the GenESA Device is 0.8 μg/kg/min and the maximum total dose is 10 μg/kg. The GenESA Device includes a "HOLD HR" feature that, when activated, allows HR to be maintained at approximately that level for up to 5 minutes.

Monitoring features of the GenESA Device include continuous display of ECG, HR, blood pressure and dosing information. In addition, the device has a series of "alerts" that warn of conditions that may require attention and "alarms" that stop drug delivery due to a potential safety hazard. Each alert or alarm provides a visual message and an audible tone to warn the operator. The physician conducting the GenESA System test may manually interrupt the delivery of arbutamine at any time, if clinically appropriate. The infusion of arbutamine may be restarted by the operator if the condition resulting in the interruption of infusion has been corrected, a diagnostic endpoint has not been reached, and it is considered safe and appropriate to do so.

"Heart rate saturation" (a flattening or plateau of the HR response to increasing dose of arbutamine) describes the maximal HR response to arbutamine and is an endpoint of the GenESA System test. If such a flattening or plateau of the HR response is detected when the HR is ≤ 40 bpm above the baseline level, restart of the arbutamine infusion is allowed. If the HR is >40 bpm above baseline and a HR saturation alarm occurs, restart of the arbutamine infusion is prevented by the GenESA Device (since it is unlikely that any further clinically significant increase in HR will occur following restart and there is a potential risk of serious cardiac arrhythmias (see **WARNINGS**)).

The infusion of arbutamine should be terminated when a diagnostic endpoint (e.g. ST segment deviation on ECG) has been reached, if clinically significant symptoms or arrhythmias occur, or if clinically appropriate for any other reason. The GenESA Device will stop drug delivery when the maximum HR limit (HR TARGET) has been reached or after a total of 10 μg/kg arbutamine has been delivered. Following completion of the infusion, the patient should be monitored (using the GenESA Device or other means) until HR and blood pressure have returned to acceptable levels.

For Basic Operating Instructions and other essential information on the use of the GenESA System, see the Quick Reference pull-out cards attached to the GenESA Device.

SYRINGE AND PLUNGER ROD ASSEMBLY:

- Remove the prefilled glass syringe and plunger rod from the package.
- Inspect the grey tip cap and stopper for proper engagement. Evidence of leakage or a loose tip cap may indicate a violation of sterility.

  Note: Parenteral drug products should be inspected visually for particulate matter and discoloration prior to administration.
- Thread the plunger rod into the stopper clockwise until fully seated (approximately one full revolution).
- Hold the syringe vertically (tip UP), remove the tip cap, and manually express any air from the syringe.
- Attach the IV administration set to the syringe luer lock. Use a push and twist (clockwise) action until fully engaged.
- Load the assembled syringe/IV set into the GenESA Device as instructed in The GenESA System Directions For Use.

HOW SUPPLIED

GenESA (arbutamine hydrochloride, sterile solution for infusion, 0.05 mg/mL), is available as a 20 mL prefilled syringe. Each syringe contains 1 mg of arbutamine hydrochloride. Store at 2-8°C. Protect from light. (NDC 0703-1105-01).

CAUTION: Federal (USA) law prohibits dispensing without a prescription.

Gensia Automedics, Inc.
San Diego, California, 92121
Issued: July 1997

GENSIA AUTOMEDICS™

CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER : 20420

MEDICAL REVIEW(S)

MAR 26 1997

Steven M. Rodin, M.D.
Medical Officer

Food and Drug Administration

Division of Cardio-Renal Drugs
Tel 301-594-5377; FAX- 5495

Medical Review of NDA Safety Update

1 Application Identifiers

NDA #:	**20-420**
Drug:	**i.v. arbutamine**
Sponsor:	Gensia Inc.
Proposed indication:	diagnostic adjunct
Pharmacologic type:	nonselective β- and α-adrenergic agonist
Date of submission:	28 January 1997
Review last revised:	26 March 1997

Background:

The submitted safety update has a data cutoff of 11/30/96, and presents cumulative safety data from the 697 patients in Phase 3 studies (studies 122, 123, 127, 128 and 129) and the 1385 patients in Phase 3B studies (studies 132, 135, 136, 137, 138, 139, and 141).

Deaths:

There were 6 deaths in Phase 3B, each being investigator-attributed to causes unrelated to drug:

- An 81 year old female (0141-36A-0710) with severe coronary artery disease (CAD) and unstable angina enrolled in violation of the protocol. The patient developed chest pain and ischemic ST segment depression during the arbutamine infusion. She was treated with metoprolol and nitroglycerin, with resolution several hours after the discontinuation of arbutamine. The patient subsequently died two weeks after the stress test.

- A 62 year old male (0141-05A-1251) with a history of severe 3-vessel CAD had an uneventful (albeit positive) arbutamine test, and 12 days later died at home.

- Patient 0141-67E-5653 was a 68 year old male with history of myocardial infarction (MI). Nine days after an uneventful arbutamine test he underwent coronary bypass, and died on post-operative day 2 due to severe mitral and coronary insufficiency.

- A 67 year old male (0141-68E-5626) had a positive arbutamine test, and 7 days later underwent aortic valve replacement surgery. The prosthetic valve became occluded, the patient experienced an MI and subsequently died.

- A 54 year old female patient (0141-50A-0831) died 4 days after peripheral vascular surgery, 9 days after a negative arbutamine test.
An autopsy revealed the probable cause of death to be adult respiratory distress syndrome or early pneumonia.

- Patient 0141-54A-1031, a 70 year old male with history of MI, died of an acute pulmonary embolus on the second day following surgical fixation of a fractured hip.

Arrhythmias:

In the post Phase 3 studies there was a change in reporting methods such that all arrhythmias were reported, not just those considered clinically significant. In addition, a different (revision 5) infusion algorithm was used in the later studies. Although it is plausible that one or both of these variables could result in higher rates of reported arrhythmias in the later studies, there was another variable which could plausibly reduce these rates, i.e. the restart of arbutamine infusion after a heartrate (HR) "saturation alarm" was prohibited in the more recent studies.

In post-Phase 3 studies, ventricular fibrillation (VF) was reported in 0.1% of patients (vs 0.3% in Phase 3), ventricular tachycardia (VT) was reported in 1.8% of patients (vs 1.3% in Phase 3), supraventricular tachycardia (SVT) was reported in 1.9% of patients (vs 1.0% in Phase 3), and atrial fibrillation was reported in 1.0% of patients (vs 0.6% in Phase 3).

No episodes of VF or sustained VT occurred in the 1236 patients receiving arbutamine in studies which both used infusion algorithm revision 5 and prohibited restarting of arbutamine infusion after a HR "saturation alarm".

Adverse Events resulting in Withdrawal from studies:

Two patients in the Phase 3B studies were withdrawn due to adverse events:

- Patient 0141-54A-1031 (whose narrative is described above under "deaths").

- Patient 0132-01-0013 was not exposed to arbutamine.

Most Commonly reported Adverse Events:

In the cumulative (n=2082) database of Phase 3 and 3B patients the most frequent adverse events (i.e. having reported rates >3%) were tremor (15%), arrhythmias (12%), angina (12%), headache (8%), and hypotension (6%).

COMMENT:

With the exception of the differences described above, in general the adverse events reported in the cumulative database of 2082 patients in Phase 3 and 3B were comparable in rate and type to those reported in the 697 patients in the Phase 3 database.

Steven Rodin, MD 3/26/97

Steven M. Rodin, MD
Medical Officer

Date

cc: HFD-110/ division file, GBuehler, RFenichel; ***no copy to SRodin**

Steven M. Rodin, M.D.
Medical Officer

Food and Drug Administration

Division of Cardio-Renal Drugs
Tel 301-443-0320; FAX-9283

Addendum to Medical Review of NDA:

1 General information

NDA #:	**20-420**
Drug:	**arbutamine**
Sponsor:	Gensia Inc.
Proposed indication:	diagnostic adjunct
Date of main review herein addended:	19 March 1997
Most recent submission:	20 March 1997
Addendum last revised:	27 March 1997

2 The higher prevalence of females in study 141 does not explain the differences in arbutamine-echocardiography test sensitivity or specificity, relative to study 123:

I conveyed in my main review (dated 19 March 1997) that a deeper search of the stress-test literature could be undertaken to query whether, in contrast to my sense of the literature, the "gender" covariate adequately explains (relative to study 123) the 46 percentage point lower mean echo sensitivity and 49 percentage point greater mean echo specificity obtained in study 141.

In their latest submission the sponsor essentially agrees[1] that neither of the conditions were met for reasonably concluding that this covariate reconciles the observed differences in the two studies (these conditions I describe correctly in sections d.i and d.ii of the ~~above~~ below errata). That is, not only does the "gender" covariate not plausibly explain the lower sensitivity, but moreover, it plausibly predicts a specificity difference which would be *directionally opposed* to that observed.

2 Distribution of Pre-treatment covariates in Echo study 141:

My main review noted that the pre-treatment covariates underlying study 141's echo data could potentially be more optimally assessed. An analysis of only patients evaluable via the "Echo-Ischemoid" test criterion is now available. This shows comparable covariate distributions to those depicted on Table 3 of my main review. The new analysis of study 141 again showed (relative to study 123) a somewhat lower prevalence of ventricular wall motion abnormalities (33 vs 39%, respectively), and a higher prevalence of females (30 vs 13%, respectively).

---

[1]They reference one report of "gender"-associated increase in sensitivity and another report of the opposite "gender" effect in studies of various pharmacologic stress imaging modalities, and allude to the known lower specificity of CAD diagnoses in stress-tested women.

DUPLICATE

MAR 20 1997

Steven M. Rodin, M.D.
Medical Officer

Food and Drug Administration

Division of Cardio-Renal Drugs
Tel 301-594-5377; FAX- 5495

Medical Review of NDA Efficacy Data

1 Application Identifiers

NDA #:	**20-420**
Drug:	**i.v. arbutamine**
Sponsor:	Gensia Inc.
Proposed indication:	diagnostic adjunct
Pharmacologic type:	nonselective β- and α-adrenergic agonist
Date of submission:	8 November 1996
Review last revised:	19 March 1997

2. Background:

2.1 Chronology of this NDA's history:

Arbutamine is a catecholamine proposed for use as an adjunct to various cardiac diagnostic techniques for the diagnosis of coronary artery disease (CAD). An NDA was submitted on 23 December 1993 with the proposal for the drug to be marketed as part of an infusion device system called the GenESA® System.

My initial review dated 26 October 1994 evaluated the original NDA submission (there I also proposed an analytic framework for the approval and labelling of diagnostic adjunct drugs).

Dr. Robert Temple issued a not approvable letter on 4/6/95. The first major amendment was subsequently submitted on 25 October 1995, and reviewed by me in a report dated 11 March 1996. The sponsor submitted a retort on 11 April 1996 which makes frequent reference to what they call the "Medical Review"; it it noted that this is a misnomer since their criticisms by and large are not

relevant to the analyses produced by me, the Medical Reviewer of this drug[1].

Dr. Temple sought a blinded, central image interpretation-based re-analysis in an attempt to rectify some of the noted deficiencies of the sponsor's original analysis of study 141. The current submission (a second major amendment dated 8 November 1996) contains the requested re-analysis.

In order to provide the "big picture" context needed for decision-making, I have also included in this present report pertinent new as well as recapitulated analyses of data from previous arbutamine studies.

2.2 **Operational definitions**:

Dysynergy: any of the following ventricular wall motion abnormalities: hypokinesis, akinesis, or dyskinesis (paradoxical systolic motion away from the chamber's center).

Sensitivity: Sensitivity describes the test's capacity for making a correct diagnosis in confirmed cases of *disease*[2]. A high sensitivity test frequently makes the correct (True Positive (TP)) diagnosis among patients with disease, and infrequently makes the incorrect (False Negative (FN)) diagnosis among patients with disease. Sensitivity is calculated as a fraction whose numerator is the number of TP test results, and whose denominator is (assuming exclusion of nondiagnostic test results) the number of patients with confirmed disease. The number of patients with disease is equal to the sum of the number of TP and FN test results, and sensitivity is calculable as TP/(TP + FN). The complement of sensitivity, False Negative Fraction, is the fraction of diagnoses which are *incorrect* among patients with disease, and is calculated as FN/(TP + FN). Sensitivity is equal to 1 minus the False Negative Fraction.

Specificity: Specificity describes the test's capacity for making a correct diagnosis in confirmed cases of *no disease* (rather, none of the disease of interest). A 100% specific test generates positive test results only for the specific reason of disease being truly absent. A high specificity test frequently makes the correct (True Negative (TN)) diagnosis among patient with no disease, and infrequently makes the error of detecting things that are not real disease (such detections are false positive (FP) diagnoses among undiseased patients). Specificity is calculated as a fraction whose numerator is the number of true negative (TN) test results, and whose denominator is the number of patients confirmed to not have the disease of interest. The number of patients with no disease equals the sum of TN and FP test results (assuming exclusion of nondiagnostic test results), and specificity is calculated as TN/(TN + FP). The complement of specificity, False Positive Fraction, is the fraction of diagnoses which are *incorrect* among patients without disease, and is calculated as FP/(TN + FP). Specificity is equal to 1 minus the False Positive Fraction.

---

[1]Gensia has confirmed that their references to "the Medical Review" were broadly alluding to the set of all DCRDP reviews produced by medical people.

[2]as opposed to Specificity, which is the test's capacity for making a correct diagnosis in confirmed cases of *no disease*.

Tracer "redistribution": in this report I use this term simply to describe the empirical phenomenon whereby the size of a region of myocardial hypoperfusion (this region being known as the perfusion "defect") is apparently less during delayed post-stress imaging than it is when observed during peak stress. This term is not, in this report, intended to imply anything about the pathophysiologic basis for the phenomena.

3 Methods of Re-analysis of Study 141: [1]

3.1 Summmary of study design:

The main goal of the re-analysis was to replace the criteria-indeterminate, nonblinded, noncentrally-adjudicated analyses with criteria-specified, blinded, centrally-adjudicated ones.

The sponsor's re-analysis of this study was based on 330 evaluable subjects (about 31% of 1070 enrolled patients with suspected or known CAD qualified on the basis of having an evaluable angiogram within 12 event-free weeks of an arbutamine test). The test had been performed using an iv arbutamine infusion at an intended dose rate ≤0.8 μg/kg/min up to 10 μg/kg/d[3] using "closed-loop" device infusion algorithm revision #5.[4] The retrospective objective of the requested re-analysis was to assess the sensitivity and specificity (for the diagnosis of the presence CAD) of arbutamine tests conducted using Tc^{99m}-sestamibi or $thallium^{201}$ scintigraphy, or 2D-echocardiographic methods.

See my previous review of 11 March 1996 for a detailed discussion of the design of this study.

3.2 Eligibility criteria.

Patients eligible for the retrospective re-analysis were those with angiograms within 12 weeks prior to or following the arbutamine test and without any intervening cardiac events between the arbutamine test and angiography. When available, the angiogram and respective echo and/or scintigraphic (radioisotope-based) data were to be submitted to the central laboratories, in a blinded manner, for evaluation.

[3] as in the previous pivotal studies 122, 123, and 127.

[4] relative to the infusion algorithm revision #4 used in previous studies, with this method infusions were continued to higher levels of achieved heartrate.

3.3 Endpoints.

3.3.1 Descriptions

Test classifications as positive or negative were made on the basis of various test outcome classification criteria (as described below); the concordance of test classifications with the standard classifications of disease generated by coronary angiography was then computed and expressed in conventional terms as sensitivity and specificity measures.

Coronary Angiographic standard of disease:

The presence of significant CAD was defined as coronary angiographic evidence of ≥50% reduction in cross-sectional diameter of any "major" epicardial coronary artery ("major" refers to the left main, left anterior descending, left circumflex, or right coronary arteries or any of their major branches).

Perfusion Scintigraphy test classification schemes:

The principle perfusion scintigraphy endpoints were test sensitivity and specificity for the diagnosis of the presence of epicardial CAD (not the severity of disease or multiplicity of vessels involved, but merely the presence of a stenosis which reduced the cross-sectional diameter of any major epicardial coronary artery by 50% or more).

Regional ventricular hypoperfusion (i.e. a "perfusion defect") was considered to be detected when abnormally low regional ventricular tracer uptake was observed (by at least 2 of 3 blinded observers) on the immediate post-stress perfusion image. For the blinded, centralized re-analysis the threshold for declaring a finding of regionally low tracer uptake was reached when a moderate deficit in observed tracer scintillations (relative to normally perfused regions) was noted in at least 2 of 20 defined ventricular regions, or when a severe deficit was noted in 1 ventricular region.

The criteria used by the sponsor for defining a positive stress-scintigraphy test were:

- what I will call the "Scinti-Ischemoid" criterion[5]:
  with this criterion a test is classified as Positive when a threshold perfusion defect is observed at peak stress, AND a smaller defect is observed in the absence of drug-induced hemodynamic stress [as evidenced by any degree of tracer "redistribution" upon delayed imaging, or (if pre-stressor (rest) or re-injection imaging were to be performed) by a finding of a smaller image defect during rest or re-injection imaging, relative to the size seen at peak stress].

[5]for these and the echo criterion discussed below I adapted the sponsor's term "ischemia" to avoid implying that classifications generated according to this criterion are necessarily valid representations of the presence or absence of ischemic pathophysiology.

- what I call criterion "Scinti-CAD ambiguous"[6] [otherwise termed the "CAD" criterion by the sponsor]:

> with criterion "Scinti-CAD ambiguous" a test is classified as Positive when a threshold perfusion defect is observed at peak stress, irrespective of the observability of that defect in the absence of drug-induced hemodynamic stress (e.g. irrespective of the findings on "redistribution", rest or re-injection images).

Nondiagnostic results were excluded when scintigraphic test classifications were assessed for concordance with the angiographic standard (there was 1 such result for both thallium and sestamibi imaging).

2D-Echocardiography test classification schemes:

The principal 2-dimensional (2D) echocardiographic endpoints were the arbutamine test's sensitivity and specificity for the diagnosis of the presence of epicardial CAD (again, not the severity of disease or multiplicity of vessels involved, but merely the presence of a stenosis which reduced the cross-sectional diameter of any major epicardial coronary artery by 50% or more).

Ventricular dysynergy was considered to be detected when a wall motion abnormality was observed in one or more segments of the anterior wall, or two or more segments of the posterior/inferior wall. Nondiagnostic results were excluded from the echo analyses (e.g. there were 20 such nondiagnostic results excluded during application of test outcome classification criterion "Echo-Ischemoid").

APPEARS THIS WAY
ON ORIGINAL

[6]I have chosen this term to underscore the potential problems whereby classifications generated via this criterion can be biased toward overestimation of the role of the drug stressor per se in eliciting the diagnostic phenomenon (i.e. when such phenomenon are to a total or partial extent observable in the absence of the drug's effect).

The criteria used for classifying stress-echocardiography tests as positive were:

- criterion "Echo-Ischemoid"[7]:

  With this criterion a test is classified as positive when new OR worsened (each relative to pre-stress) dysynergy is observed during arbutamine exposure.

- criterion "Echo-CAD ambigous"[8]:

  With this criterion a test is classified as positive when dysynergy is observed EITHER prior to arbutamine exposure, OR during arbutamine exposure (both new and worsened abnormalities, relative to pre-drug).

3.3.2 **Endpoint measurement methods**

Coronary Angiographic methods:

The methods of angiographic data aquisition and processing were not prespecified or standardized across centers, and the details of site-specific methodology were not captured.

For the blinded, centralized re-analysis angiograms were evaluated using qualitative methods and, when possible, quantitative methods. Coronary artery stenoses were evaluated using quantitative coronary angiographic methods whenever possible, but angiograms submitted on videotape were evaluated by only qualitative analysis. The centrally-interpreted qualitative results represented the average of the percent stenosis measurements of two readers blinded to knowledge of the test results or any other clinical data.

Perfusion Scintigraphy methods:

The methods of tracer administration, scintigraphic image aquisition, and data processing were not prespecified or standardized across centers; and the specific details of site-specific methodology were not captured. Two nuclear cardiologists independently interpreted scintigraphic data without knowledge of patient history or the results of other diagnostic procedures. If consensus was not reached, a third nuclear cardiologist provided an interpretation and the consensus score was documented. The re-analysis plan specified that scans would be deemed nondiagnostic if unacceptable on the basis of image clarity or background uniformity. See below for scintigraphic endpoint definitions.

---

[7]called "echo ischemia" by the sponsor.

[8]called "echo CAD" by the sponsor.

Echocardiographic methods:

The methods of echo image aquisition, and data processing were not prespecified or standardized across centers, and the specific details of site-specific methodology were not captured. It is known that the rest-peak stress echocardiographic method was used in which ventricular wall motion data were first aquired under resting conditions, and again under conditions of peak arbutamine-induced hemodynamic stress.

For the blinded, centralized re-analysis digitized loops of baseline (pre-arbutamine) echocardiograms were to be submitted whenever possible. Videotape was to be submitted if echo data were either not digitized or digitized using equipment or media not compatible with the Central Laboratory or if data were only collected on videotape. Echocardiograms were analyzed centrally by two independent observers who were unaware of the results of other procedures. A third observer was used to break ties whenever there was discrepancy between the first two observers. For each set of images the left ventricle was divided into 16 segments and each interpretable segment was assigned a score according to the follow categories: 1 = normal or hyperkinetic, 2= hypokinetic, 3= akinetic, 4= dyskinetic. The ventricular wall was divided into a total of 16 segments. Echo tests were considered nondiagnostic if less than 13 of 16 segments were adequately visualized (and no positive findings were made). See below for echocardiographic endpoint definitions.

4 Integrated Results (current and prior data submissions):

In this re-analysis of study 141 the sponsor retrieved data from over 80% of the qualified patients were retrieved, thus capturing 330 patients with evaluable angiograms within 12 event-free weeks of the arbutamine test. This evaluable, centrally-analyzed, blinded dataset (cutoff date of 11/1/95) forms the basis for the analyses evaluated in this report[9].

The dataset from the previous studies (123 and 127) which is most comparable to the evaluable dataset reported in study 141 is the intent-to-treat/completers dataset (this included a small number of protocol violators whose violations do not have plausible impact on these estimates).

4.1 Distribution of Covariates:

4.1.1 Pre-treatment covariates:

Shown below is the distribution of pre-treatment covariates in the dataset upon which the optimal thallium analyses were based in studies 141 and 127. No imbalances are noted which would plausibly have substantive impact on the prevalence-independent estimates of test concordance with angiography.

[9]small numbers of patients had imaging studies using different modalities than the majority, and these small sample experiences are not discussed in this report.

Table: 1

Pre-treatment characteristics of subjects
in THALLIUM studies 141 vs 127

(dataset: patients evaluable vis a vis the "CAD" criterion):

	THALLIUM samples under comparison	
	study 141 THALL "CAD" criterion	study 127 THALL "CAD" criterion
Male (%)	75	83
Female (%)	25	17
Age mean (yr)	63	60
Caucasian (%)	86	98
Black (%)	11	2
Other racial group (%)	4	1
rate of angiographic CAD (one or more ≥50% diameter-narrowing stenosis) (%)	57	93
weight mean (lb)	180	194

[source: table 3, fax submission dated 3/17/97]

Shown immediately below is the distribution of pre-treatment covariates in the sestamibi portion of study 141.

Table: 2

Pre-treatment characteristics of subjects
in SESTAMIBI study 141

(dataset: evaluable vis a vis either "Scint-Ischemoid or -CAD" criteria):

	study 141 Sestamibi
Male (%)	74
Female (%)	26
Age mean (yr)	57
Caucasian (%)	85
Black (%)	11
Other racial group (%)	5
rate of angiographic CAD (one or more ≥50% diameter-narrowing stenosis) (%)	76
weight mean (lb)	180

[source: table 2, fax submission 3/16/97]

Shown below is the distribution of pre-treatment covariates in echo study 123 vs the echocardiographic portion of study 141. This is presumably a slightly "rough take" on a more relevant dataset description which is still pending (that being the set upon which the optimal echo analyses were based, i.e. patients evaluable vis a vis the "Echo-Ischemoid" criterion).

Based on this sort-of-rough take on the data, the echo dataset in study 141 (relative to that in study 123[10]) showed a somewhat lower prevalence of ventricular wall motion abnormalities (35 vs 40%, respectively), a somewhat higher prevalence of females (30 vs 17%, respectively), a somewhat lower prevalence of angiographically proven disease (77 vs 90%, respectively). Note that pre-treatment covariates underlying the echo data are potentially more optimally assessed on the basis of considering only patients evaluable via the "Echo-Ischemoid" test criterion, because it is only this criterion which is free of what I term the "drug overcrediting" bias (see more on this point below).

Table: 3

"Sort-of-rough" take on pre-treatment characteristics of subjects in studies 141 vs 123 vs 127:

(dataset: evaluable vis a vis either "Echo-Ischemoid or -" criteria):

	ECHO datasets	
	study 141 ECHO	study 123 ECHO
Male (%)	70	83
Female (%)	30	17
Age mean (yr)	59	60
Caucasian (%)	83	89
Black (%)	13	6
Other racial group (%)	4	5
rate of angiographic CAD (one or more ≥50% diameter-narrowing stenosis) (%)	77%	90
rate of baseline (pre-drug) wall motion abnormality (%)	35	40
weight mean (lb)	not reported	179

[sources- table 2 of submission dated 3/16/97, and my previous report]

Datasets used for this table were the evaluable/blinded/centralized dataset from study 141 vs the blinded/centralized/intent-to-treat datasets from study 127 vs study 123. Note that the descriptions generated did not restrict consideration to those patients evaluable vis a vis the "Ischemoid"-based echo positivity criterion.

[10]I focus on this comparison because of disparities in results of these particular datasets.

4.1.2 Treatment covariate:

Study 141 employed an arbutamine infusion algorithm (revision #5) which differed from the algorithm revision #4 used in previous studies 123 and 127. With the revision #5 method, infusions were continued to higher levels of achieved heartrate. As expected, this more frequently resulted in patients attaining the conventional target heartrate (HR)[11] in study 141 than, for example, in study 123 (the respective rates were here 20 vs 8%). Assuming (not unreasonably) that drug-elicited tachycardia is of some value as index of the intensity of diagnosis-eliciting effect for this pharmacologic stress test, the imbalance in the rate of attainment of target HR has potential significance for derived estimates of diagnostic performance[12].

4.2 Concordance of test classifications with the Angiographic standard

Note that the most comparable of the available datasets for comparing study 141 results to those of previous pivotal studies are the evaluable completers dataset in study 141 vs the intent-to-treat completers[13] datasets in studies 123 or 127. I present the results of central/blinded analyses which were based on specified criteria for test outcome classifications (called the "central" analyses by the sponsor), rather than the nonblinded/noncentrally adjudicated analyses which were based on criteria-indeterminate test classifications ("local" analyses). The latter analyses were seriously flawed as critiqued below and in my previous review[14].

Test concordances with the angiographic standard are also summarized below for the previous pivotal studies 123 and 127, in order to provide the necessary perspective for integrative conclusions.

---

[11]defined as 85% of the age-predicted maximal HR.

[12]as further discussed in the comments section below.

[13]although "evaluable" datasets were constructed by the sponsor in studies 123 and 127, the intent-to-treat datasets are optimal there since they maximize sample sizes and do not plausibly introduce significant impairment of signal fidelity given that the "unevaluable" completers of arbutamine tests in those studies were subjects with relatively minor protocol violations.

[14]review of 11 March 1996.

4.2.1 **Pooled Thallium results**:

I pooled the thallium data in the table below (see Appendix 1 for unpooled Thallium results as well).

Table: 4

Arbutamine-THALLIUM tests' Sensitivity and Specificity, vis a vis the diagnosis of the angiographic presence of a 50% diameter-narrowing epicardial coronary stenosis (pooled results of studies 141 and 127)

test classification criterion: "Scinti-CAD"
datasets: blinded/centrally-interpreted (both studies);
evaluable (study 141), intent-to-treat/arbut test completers (127)

Metric	*pooled Arbut-THALLIUM* (criterion "Scinti-CAD")
SENSITIVITY mean & 95% CI	84% (76-90%) [n= 128]
SPECIFICITY mean & 95% CI	45% (23-69) [n= 20]

[source= table 1 of sponsor's fax submitted 3/17/97]

4.2.2 **Sestamibi results:**

Table: 5

Arbutamine-SESTAMIBI tests' Sensitivity and Specificity, vis a vis the diagnosis of the angiographic presence of a 50% diameter-narrowing epicardial coronary stenosis (study 141)

test classification criterion: "Scinti-CAD"
datasets: blinded/centrally-interpreted/evaluable

Metric	*Arbut-SESTAMIBI* study 141 crit. "Scinti-CAD"
SENSITIVITY mean & 95% CI	63% (50-75%) [n= 65]
SPECIFICITY mean & 95% CI	67% (43-85) [n= 21]

[source= table 6 of fax submitted 3/17/97]

4.2.3 **Unpooled Echocardiography results:**

With regard to the echo analyses, the assumption of study homogeneity was not well met insofar as there was imbalance in a pre-treatment covariate known to impact echo-based specificity estimates (according to the "sort-of-rough" take on covariate distribution this was the case with respect to the pre-test rate of ventricular wall motion abnormality)[15]. I elected not to pool the arbutamine-echo data across studies 141 and 123, given this, and the fact that separate analysis of the individual echo studies led to noteable evidence of disparate results, evidence that could only be obscured by any efforts to pool the data.

Table: 6

Arbutamine-ECHO tests' Sensitivity and Specificity, vis a vis the diagnosis of the angiographic presence of a 50% diameter-narrowing epicardial coronary stenosis (unpooled results of studies 141 and 123)

test classification criteria: "Echo-Ischemoid"
datasets: blinded/centrally-interpreted (both studies);
evaluable (study 141), intent-to-treat/arbut test completers (study 123)

Metric	*Arbut-ECHO*	
	study 141 crit.:"Echo-Ischemoid"	study 123 crit.: "Echo-Ischemoid"
SENSITIVITY mean & 95% CI	30% (23-39%) [n= 142]	76% (67-83%) [n= 127]
SPECIFICITY mean & 95% CI	80% (64-91%) [n= 39]	31% (11-59%) [n= 16]

[source= tables 3 & 5 of fax submitted 3/10/97]

[15]note that this covariate is not established as capable of biasing thallium-based estimates.

4.3 Reviewer's Conditional Probability Analyses

4.3.1 Bayesian method used:

Using the conditional probability model published by Diamond & Forrester[16], (equations are shown in the table below), I generated model-based Bayesian estimates of the incremental contribution of Arbutamine tests' to the clinical characterization of patients' probability of angiographic CAD. I input mean sensitivities and specificities into the model, thus assuming that these point estimates are adequate representations of the tests' ability to provide concordance with the angiographic standard of disease. More conservative assessments of incremental utilities of the test can be obtained by instead assuming that the lower 95% confidence bounds provide the best estimates of sensitivity and specificity.

Table: 7

Equations used in the Bayesian model of Diamond & Forrester

given a POSITIVE test result, the Post-Test likelihood of Disease= $\frac{\text{(pretest likelihood)(Sensitivity)}}{\text{((pretest likelihood)(Sensitivity)) + ((1- pretest likelihood)(1- Specificity))}}$
given a NEGATIVE test result, the Post-Test likelihood of Disease = $\frac{\text{(pretest likelihood)(1- Sensitivity)}}{\text{((pretest likelihood)(1- Sensitivity)) + ((1- pretest likelihood)(Specificity))}}$

[16]N Eng J Med 300:1350-1358,1979,

4.3.2 **Results of Conditional Probability analyses:**

Table: 8

Model-based estimates of Arbutamine tests' incremental contribution to the clinical characterization of a patient's likelihood of CAD

scintigraphic test classification criterion: "Scinti-CAD"
echo test classification criterion: "Echo-Ischemoid"

Imaging mode and data source	PRE-test probability of CAD: Low denotes: 20% Intermed " : 50% High denotes: 80%	Change from pre-test probability of CAD (expressed as # of percentage points), given a test result which is:	
		POSITIVE	NEGATIVE
Arbut-**SESTA** study 141	LOW	+12	-8
	INTERMED	+16	-14
	HIGH	+8	-11
Arbut-**THALL** (pools studies 141 & 127)	LOW	+8	-12
	INTERMED	+10	-24
	HIGH	+6	-21
Arbut-**ECHO** i) study 141	LOW	+7	-2
	INTERMED	+10	-3
	HIGH	+6	-2
ii) study 123	LOW	+2	-4
	INTERMED	+2	-6
	HIGH	+2	-4

These model-based estimates of changes from prior probability utilized the above-described equations of Diamond & Forrester (quattro files mean/arb_echo.wq1 & mean/arbscint.wq1). The input sensitivities & specificities were means obtained from blinded, "central" test classifications, using classification criteria "Scinti-CAD" or "Echo-Ischemoid" (as defined earlier in this report), & either evaluable datasets (study 141), or (in studies 123 & 127) intent-to-treat datasets of arbutamine test completers (irrespective of whether completed exercise tests). SESTA= Tc^{99m}-sestamibi; THALL= $thallium^{201}$; ECHO= 2D echocardiography.

For the arbutamine-echo studies these Bayesian analyses are consistent with what is suggested by simple contemplation of the disparate, study-dependent distribution of concordance errors. Study 141-based echo point estimates show low sensitivity/moderately high specificity, while the point estimates from study 123 show a diametrically opposed error distribution (i.e. moderately high sensitivity/low specificity). In order to more closely evaluate the implications of this from a conditional probability perspective, a somewhat "higher resolution" analysis is provided below. Here the categorical pre-test disease likelihoods were divided into 5 rather than 3 levels. The analysis was otherwise conducted in the same manner as described for the previous table.

Table: 9

Model-based estimates of Arbutamine-ECHO tests' incremental contribution to the clinical characterization of a patient's likelihood of CAD

echo test classification criterion: "Echo-Ischemoid"

Arbut-**ECHO** study	PRE-test probability of CAD: Low denotes: 20% Low-Medium denotes: 35% Intermed denotes: 50% High-Medium denotes: 65% High denotes: 80%	Change from pre-test probability of CAD (expressed as # of percentage points), given a test result which is:	
		POSITIVE	NEGATIVE
study 141	LOW	+7	-2
	LOW-MEDIUM	+10	-3
	INTERMED	+10	-3
	HIGH-MEDIUM	+9	-3
	HIGH	+6	-2
study 123	LOW	+2	-4
	LOW-MEDIUM	+2	-6
	INTERMED	+2	-6
	HIGH-MEDIUM	+2	-6
	HIGH	+2	-4

5 Comments and Conclusions

5.1 Incorporating pretest probabilities into a rational evaluation of arbutamine

The sponsor proposes to label the sensitivity and specificity according to pre-test probability, with the latter estimated using a conventional algorithm. This approach is, at best, not as illuminating as the Bayesian perspective I brought to bear in the above conditional probability analyses.

The sponsor implies[17] that the metrics, sensitivity and specificity, are dependent upon pre-test likelihood, and in so doing suggests something beyond what evidence has shown. Ironically, I feel partly (yet unwittingly) a stimulus to their presenting those metrics as if they were a function of pre-test disease likelihood insofar as I once pointed to the compelling evidence that the specificity of ventricular wall motion-based methods of diagnosing CAD is dependent on the presence of pre-test wall motion abnormality. It should be noted that in the current NDA only the echocardiographic data fall into this category so any extrapolations to scintigraphy data need be recognized to be presumptive. At most it may be reasonable to assume that arbutamine-echo test specificity has some degree of relationship to the pre-test likelihood of disease. Yet there are multiple reasons for baseline wall motion abnormality (the covariate of known importance), some (such as concomitant stenoses) plausibly having straightfoward correlation with pre-test likelihood, other reasons (e.g., remote infarct with currently normal artery lumen after spontaneous or iatrogenic thromolysis, angioplasty or surgical bypass) less plausibly are high-fidelity markers of current risk of signficant stenosis.

Moreover there have been no data or argument offered in support of a postulate that arbutamine-echo sensitivity is dependent on any identified baseline covariated, or any combination of covariates which correlate with the pre-test likelihood of disease. Similarly, no support has been given to the postulate that thallium or sestamibi-based sensitivity or specificity is dependent on any identified baseline covariate, or any combination of covariates which are themselves correlated with the pre-test likelihood of disease.

My approach to integrating the pre-test probability data attempted to bring as much decision analysis rigor to the pre-test probability data as is possible, by inputting these (along with the other requisite inputs) into the Bayesian model published by Diamond & Forrester. I offer that analysis as a means of exploration of the data's implications; I am not advocating for giving the model's output more than semi-quantitative consideration (I have yet to thoroughly ascertain what is known of the model's validity). To employ this approach is not actually something entirely new to my arbutamine reviews, but rather represents a quantitative framing of previously offered qualitative perspectives on tests which are useful for purposes of diagnostic inclusion vs diagnostic exclusion.

My approach is noteably different than the sponsor's one of describing, with questionable validity and vague purpose, the angiographic concordances as a presumed function of pre-test covariates. I input into the Bayesian equation the experimentally-derived estimates of sensitivity and specificity (based on the whole evaluable sample, not on a subgroup), and nomogram-derived estimates of a given subject's

[17] see their labelling proposal.

pre-test likelihood of disease. What is generated is an estimate of the post-test likelihood of disease given the conditions of either a positive or negative test outcome classification. Knowledge of this is necessary for one to estimate what is without question the clinically most critical parameter for a diagnostic test. Whatever one wants to call this parameter (e.g., "incremental probability change"), it is that estimable entity which alone describes whether the test-facilitated knowledge of the post-test likelihood of disease is very much different then the knowledge of disease likelihood that was available to the diagnostician *a priori* .

5.2 **Regarding the totality of the Arbutamine-Thallium evidence**:

a. With these newly analyzed data from study 141 the sample size upon which reliable specificity estimation can be based is increased considerably (by a factor of 2.5).

b. The newly optimized analysis of study 141's arbutamine-thallium test outcome classifications (i.e. the criteria-specified/central/blinded analysis) demonstrates angiographic concordances which are consistent with those of previous study 127 (see Appendix 1 for illustration of the overlapping of confidence intervals for both sensitivity and specificity).

b. The specificity estimates obtained from the pooled data show a mean specificity of arbutamine-thallium testing that was higher than that observed in study 127, and a lower 95% confidence bound for specificity that was also higher than that observed in study 127).

c. Overall, the arbutamine-thallium testing method showed reasonably high sensitivity and low specificity. The Bayesian analyses illuminate what is qualitatively predictable for a test with such characteristics[18]. That is, in in patients with intermediate or high pre-test risk, a negative result of such a test is predicted to confer a nontrivial (21-24 percentage point) marginal decrease in disease probability, with patients at low prior risk experiencing less such marginal utility from the test. Moreover, the model predicts (as was also expected) that positive test results would have less utility (that utility which is provided by these is related to their imparting, principally in patients at intermediate prior risk, the knowledge of a modest (10 percentage point) increase in likelihood of disease.

5.3 **Regarding the Arbutamine-Sestamibi findings**:

a. The re-analysis of study 141's arbutamine-sestamibi results convincingly demonstrate concordances with angiography that afford the test a degree of clinical utility.

b. The arbutamine-sestamibi testing method showed moderate mean sensitivity and moderate mean specificity. Although the incremental utility of a test with these characteristics is (relative to the thallium example) more difficult to mentally model, one would at least be able to venture that similarly sized effects on disease probability would be "produced by" positive and negative test outcomes. This was confirmed by formal Bayesian analyses. These predicted that a negative result of

[18]see the related qualitative comments in my initial review of 26 October 1994.

such a test would confer a modest (11-14 percentage point) marginal decrease in disease probability, but that patients with low prior risk would experience less such marginal utility. The model additionally predicted that positive test results would also have marginal utility in imparting (principally in patients at low or intermediate prior risk), the knowledge of a modest (12-16 percentage point) increase in disease likelihood).

5.4 **Regarding the totality of the Arbutamine-echocardiography evidence**:

5.4.1 **Existence of important between-study disparities in echo results**:

By employing the best matched datasets (i.e. the evaluable dataset in study 141 vs the intent-to-treat dataset in study 123), and undertaking the analysis which is least vulnerable to bias and noise[19] (i.e. the criteria-specified/central/blinded analysis), it became evident that the mean sensitivity found in study 141 is much lower (30% vs 76% in study 123, respectively), that the mean specificity in study 141 is much higher (80% vs 31% in study 123, respectively), and that the study's 95% confidence intervals neither overlap for the sensitivity or the specificity estimates.

If the disparity in submitted arbutamine-echo results is not reconcilable by data or reasoned judgement, I would argue that this would pose a significant barrier to the approval and labelling of arbutamine for use with this imaging modality. My argument would essentially be that:

> *unless the totality of the estimates of sensitivity at least "point in the same direction" and the totality of the estimates of specificity at least "point in the same direction", it may not be clear which directionality is characteristic of any vector(s) which describes the marginal utility of arbutamine-echocardiography testing, nor in which patient population (if any) such vector(s) demonstrates a usefully large size.*

Given this concern, I undertook a deep assessment of the reconcilability of these data.

5.4.2 **Critique of the sponsor's attempts at general[20] explanations regarding echo study disparities**:

a. The sponsor contends that the retrospective nature of the re-analysis of study 141 somehow introduces a flaw into the derived estimates. I do not identify anything compelling in this argument. If "local" image interpretors readers had data to submit to the retrospective "central" analysis they were to do so. The retrospective element may have contributed to what was an incomplete data capture, but we have been submitted no evidence that the nature of the data capture was biased.

---

[19]here I refer to the interobserver variance-based noise introduced by failure to use centrally adjudicated image interpretations in the "local" analyses.

[20]general, as opposed to other explanations (critiqued below) which are aimed solely at reconciling just one of the two between-study discrepancies.

b. The sponsor contends that inferior image display tools (e.g. analog vs digital image displays) used in some study 141 sites flawed the derived concordance estimates, relative to those obtained in study 123. Unfortunately the study 141 results are not sufficiently internally consistent to adequately support that this is a wise perspective. For example, on the basis of this contention one would have quite reasonably expected specificity estimates to have suffered by any use of less than optimal image display methodology in study 141 (state-of-the-art, dual, side-by-side digital displays of rest and stress images from multiple imaging planes are arguably important to the avoidance of interpretive artefacts and clasification errors) whereas it was an apparently greatly superior specificity which was observed in this study.

c. The sponsor contends that the discrepancy is less apparent when employing the criteria-indeterminate/unblinded/noncentrally adjudicated ("local") analysis of study 141 data, and essentially proposes that we focus on the "local" analysis[21]. This proposal is lacking in scientifically defensible rationale.

We reasonably should maintain appreciable concern for the potential for even unintentional bias to have influenced the test interpretations at multiple nonblinded levels (i.e. the interpretors of noninvasive tests were not blinded to the results of angiography or other clinical data, angiography interpretors were not blinded to the results of noninvasive tests or other clinical data). There are no quality controls to prevent a reported test interpretation from being finalized only after angiogram results are known.

Moreover, the labelling that could be written on the basis of the "local" analyses criteria-indeterminate classifications would be rendered so low in quality as to be arguably approval-limiting. The concordance that a given positivity criterion will produce, vis a vis the standard for a given disease's diagnosis, is entirely dependent on the basic structure and threshold-levels of these definitional elements, as amply validated in a body of literature using analyses of Receiver Operating Characteristics. An equally well-validated point is that the distribution of error in concordance (i.e. whether a test has high sensitivity and low specificity, or medium sensitivity and medium specificity, or high sensitivity and high specificity, etc. etc.) is fully dependent on such things as the arbitrary choice of criterion cutoff points. The particular distribution of error in concordance, it is equally well established, determines the direction and extent of any marginal utility the test has in making an incremental contribution to the estimation of a patient's probability of disease. If test positivity criteria are indeterminate (as in the case of the "local" analyses of study 141), then although one can generate a seemingly useful table displaying the concordance between the results of the resulting classifications and the standard for disease (and thus generate seemingly useful descriptions of sensitivity and specificity, and seemingly useful descriptions of marginal test utilities) one is left unable to describe what a future user of the test should apply as test classification criteria in order to obtain, on average, the same concordances and test utilities.

---

[21]since the test positivity criteria were there indeterminate one cannot exclude that we are being asked to compare apples with oranges, i.e. a comparison of the concordances generated by one test classification criterion in study 141 vs those generated by a different (albeit determinate) classification criterion in study 123.

In summary, the noted discrepancies are detected after employing the most rigorous of the available analyses, i.e. the criteria-specified/central/blinded analysis is the one least vulnerable to bias, least vulnerable to noise, and most amenable to labelling.

5.4.3 Critique of sponsor's attempts to account for the much lower point estimate of echo specificity in study 123:

a. *Refutation of the argument for a putative bias causing underestimation of specificity in study 123:*

The sponsor maintains that the relatively lower extent of capture of disease-negative patients in study 123 causes a systematic bias towards underestimation of specificity in study 123. For the sake of argument I will accept their premise that a pattern of angiographic under-referral operates in the research community to limit the number of arbutamine test-negative patients who go on to receive angiography.[22] I further accept, again for the sake of argument[23], their implicit premise that the basis for the smaller size of the sample of proven disease-negative patients in study 123 (relative to the size of that sample in study 141) is a more extensive pattern of angiographic under-referral of arbutamine test-negative patients in study 123, relative to study 141. I beg to differ with the sponsor's assertion that this referral pattern is known to produce a bias towards underestimation of test specificity, an assertion rendered with direct reference to arguments originally put forth in an influential publication by the Cedars Sinai group[24].

I have completed an analysis (the likes of which has not, to my knowledge, previously been published) which finds such an assertion to be unsupported by the argument rendered in that paper. In brief, the original publication reported a trend toward an apparent lowering over time of experimentally-determined point estimates of the specificity of a cardiac diagnostic test. This test, like the echo tests used in the arbutamine study, employed ventricular wall motion analyses for classification of test outcomes.

[22]in the case of an investigational test, this referral behavior ironically begs the question of just how nonerroneous the disease-negative test classifications are, relative to the standard for the diagnosis of interest.

[23]these premises are accepted for the purpose of the dialectic, not necessarily because we have knowledge that they are true.

[24]Rozanski et. al. N Engl J Med 1983; 309:518-22.

The authors put forth the view[25] that *as the extent of angiographic under-referral of test-negative subjects increases, the specificity point estimate generated by an investigative study becomes putatively biased towards under-estimation.*
In putative support of that interpretation the author's elaborated u-upon a thought experiment based on a scenario of total nonreferral to angiography of any patients with negative test results. Their portrayal of the outcome of that thought experiment was as follows:

> ... "if only positive test responders were to be referred to coronary angiography, then the test would appear to be 100 per cent sensitive (since all patients with disease would have a positive test) and 0 per cent specific (since all those without disease would also have a positive test)."

I have detected an artefact produced by the authors' handling of missing data in that thought experiment. They twice imputed (in place of a missing datum) a datum with an arbitrary numerical value of zero. This imputation method implicitly contained two logically flawed assumptions: i) that a zero false negative rate has the same epistemological implications as does an indeterminate false negative rate, and; ii) that a zero true negative rate is epistemologically interchangable for an indeterminate true negative rate.

For illustration of the numeric aspects of my refutation, see the 2 X 2 concordance table, and sensitivity/specificity equations presented below. It is evident that when the authors impute a value of zero for the missing description of FN, the expression for sensitivity [i.e., the ratio described by the formula: TP/(TP + FN)] becomes artefactually reduced to TP/TP, resulting in a sensitivity estimate of 100%. When imputing a value of zero for the missing value of TN, upon calculating specificity [i.e., the ratio described by the formula: TN/(TN + FP)] the authors similarly arrived at an artefactual estimate of zero percent.

The alternative data analysis method I am suggesting seeks to provide artefact-free descriptions of test classification outcomes in the author's experiment, given that the requisite data are totally missing for two of the cells in the 2 X concordance 2 table, and that imputing zero values is based on indefensible logic which results in undesirable artefact.

Applying my suggested analysis method one finds that the number of false negative classifications

---

[25]Granted, in addition to this view, another explanation is provided by the authors' for the finding that some specificity estimates were much higher than others. Although I consider it sound, this additional explanation is irrrelevant to comparisons of 123 vs 141 (because each enrolled patients with suspected CAD as opposed to "normal" subjects), and only pertinent to comparisons of the "normalcy" metric from arbutamine study 128 to studies 123 or 141. Employed in my previous arbutamine reviews, this argument supports the conclusion that the "normalcy metric" over-estimates that specificity observed in patients referred with suspicion of CAD, when (as is usually the case) the average prevalence of a particular specificity-degrading covariate (i.e. baseline wall motion abnormality) is lower in the "normal" subjects than in the CAD-suspect patients.

produced by the test, and the estimate of sensitivity which depends on this number, to both be simply indeterminate under the conditions given. Similarly, the number of true negative disease test classifications and the resulting specificity are, under the conditions given, wholly indeterminate.

Considering their thought experiment as given, I find no other interpretation supportable than that the angiographic under-referral pattern, when complete, causes test specificity (and test sensitivity) to be indeterminate.

Table: 10

An attempt to draw concordance between test results and angiography standard for the diagnosis of CAD, under the conditions of the hypothetical experiment of Rozanski et. al. (i.e. *total* nonreferral to angiography of any patient with a negative test result).

		Results of Angiography	
		Negative	**Positive**
results of Hypothetical Test	**Negative**	data with which to describe the # of True Negatives (TN) are made unavailable by the total nonreferral	data with which to describe the # of False Negatives (FN) are made unavailable by the total nonreferral
	Positive	FP= observed # of False Positive tests	TP= observed # of True Positive tests

Table: 11

Standard equations used to calculate test sensitivity and specificity

Sensitivity= TP/(TP + FN)
Specificity= TN/(TN + FP)

Abbreviation key: TP= True Positive; FN = False Negative; TN= True Negative; FP= False Positive.

b. *Critique of other potential accountings for much lower point estimate of specificity in study 123*:

i. It is reasonable for the observed 51 percentage point difference in specificity point estimates to be **only in small part** attributed to the higher prevalence of pre-drug wall motion abnormalities in study 123 (40 vs 35% in study 141, respectively, according to the "sort-of-rough" take on the baseline covariate data[26]), since this difference plausibly resulted in a not large number of excess false positives. Importantly, there is still a largely unaccounted for specificity difference after taking this variable into consideration.

ii. There is no evidence that differences in the intensity of arbutamine's diagnosis-eliciting effect account for the specificity difference. In fact, one may have expected that the lower rate of attainment of target HR in study 123 would have caused specificity to have been relatively higher in this study than in 141 (whereas the opposite finding was made).

5.4.4 **Critique of potential accountings for the much higher point estimate of echo sensitivity in study 123**:

a. There is no evidence that differences in the intensity of arbutamine's diagnosis-eliciting effect account for this difference either. In fact, one may have expected that the higher rate of attainment of target HR in study 141 would have caused sensitivity to have been relatively higher in this study than in study 123 (whereas the opposite observation was made).

b. I requested the sponsor to scrutinize the data for signals of any higher prevalence of endogenously collateral-rich, surgically graft-based, or angioplasty-based revascularized lesions which in theory could have caused more frequent false negatives (and lower sensitivity) in study 141, relative to study 123. They found no evidence to support that this was the case.

c. The extraordinarily narrow 95% confidence bands for study 141's arbutamine-echo sensitivity estimate (23-39%) are very suggestive that the echo portion of this study was well executed.

d. The echo dataset in study 141, relative to that in study 123, showed a somewhat higher prevalence of females (30 vs 17%, respectively, as per the "sort-of-rough" take on the baseline covariate data). A more refined covariate description has been sought of the sponsor, but it is nonetheless clear that consideration of the best description of the distribution of this covariate would not adequately reconcile the two studies unless one were shown data or perhaps compelling argument to support that the degree of difference found in the covariate's distribution was both:

i. plausibly responsible for one or more of the relative estimate differences in the echo results of study 141 (i.e. the observed lower sensitivity and/or the higher estimate of ~~sensitivity~~ specificity), AND

ii. not plausibly predictive of findings which contradict that which was observed (e.g. predicting that study 141's ~~sensitivity~~ specificity would be lower, whereas it was actually higher, relative to study 123).

[26]as presented in recent and past submissions by the sponsor.

My ad hoc recollection of the stress-echo literature does not suggest that the "gender" covariate will be sufficiently explanatory[27]

5.4.5 **Conclusions regarding the reconcilability of disparate echo data**:

The disparity in submitted arbutamine-echo results has yet to be found reconcilable by data or reasoned judgement. As I intimated previously, this is reasonably construed to seriously undermine the basis for approval and labelling of arbutamine for use with this imaging modality.

The totality of the estimates of arbutamine-echo sensitivity do not even "point in the same direction", nor do the totality of the estimates of arbutamine-echo specificity. Consideration of the disparate outputs of the conditional probability model reveals great ambiguity as to the size and directionality characteristic of any vector which describes the marginal utility of arbutamine-echo testing, and ambiguity as to the patient population (if any) in which this vector(s) demonstrates an adequately clinically useful size.

5.5 **The issue of "drug overcrediting" bias**:

The diagnostic rationale for using such "stressor" drugs as arbutamine is to produce sufficient diagnosis-eliciting phenomena[28] to detect those significant coronary stenoses which do not measurably perturb cardiac perfusion or function in a resting subject, and thus are not detectable by imaging performed in the absence of drug effect. As I suggested with somewhat different language in previous reviews, some criteria for classification of test outcomes are potentially confounded in such a way as to render concordance estimates vulnerable to what I have termed the "drug overcrediting" bias. These criteria are potentially biased toward overestimating the role of the drug stressor in eliciting diagnostic imaging findings which are to some extent (total or partial) observable in the absence of the drug's effect on the basis of the imaging method's inherent discriminative utility. Such phenomenology is illustrated by the case of the hugely occluding coronary stenosis which imparts a diagnostic abnormality of ventricular wall motion and/or perfusion despite the employment of a drug adjunct[29]-free imaging method.

To reduce (or hopefully, to eliminate) the potential for bias towards overestimation of the diagnosis-eliciting utility of the adjunct drug per se, I have sought to differentiate test positivity criteria which are not confounded in the way I am referring to here. For the arbutamine-echo data, this task is not difficult. As in previous reviews, I here based my echo analyses on the test outcome classification

---

[27]unfortunately, the review deadline has thusfar not allowed for efforts at objectifying this recollection, but the sponsor is being queried for their feedback.

[28]such as, for example, tachycardia or coronary hyperemia.

[29]by "drug adjunct" I mean the "stressor" agent as opposed to a radiopharmaceutical or nonisotopic echocontrast agent potentially employed during pharmacologic stress testing.

criterion which I more recently re-named "Echo-Ischemoid"[30]. The Echo-Ischemoid criterion is, unlike the Echo-CAD criterion elsewhere used by the sponsor, unconfounded by the attributing of utility to arbutamine for diagnoses obtainable with imaging performed at rest, i.e. prior to arbutamine administration.

Unfortunately, we do not have access to a similarly unconfounded criterion for a positive scintigraphic (thallium and sestamibi) test, being limited by the study designs used in this development program. Yet I do not suggest that the sponsor be viewed as deficient in this regard, given the unlikelihood that any of the parties to this development program previously focused attention on this potential confounder.

Given the scintigraphic study designs we are working with, it remains optimal to base scintigraphic analyses on the "Scinti-CAD" test outcome classification criterion[31] which I focused on in previous reviews, rather than the "Scinti-Ischemoid" criterion. The latter criterion does not (for either thallium or sestamibi images) provide an adequately reliable surrogate[32] for resting (pre-drug) data, while the former has the advantage of providing a potentially larger sample size upon which to base concordance estimates.

Although I did allude to the limitations of the tracer "redistribution" surrogate in my previous review of 11 March 1996, the emphasis I expressed there is one which has shifted for me upon further consideration. I now consider it more defensible to emphasize that tracer "redistribution" unfortunately does not provide a useful handle on the extent of "drug overcrediting bias", and that in the absence of an adequate rest-scintigraphy surrogate there is little basis for relying on redistribution findings for estimation of the extent of "drug overcrediting bias" (if any) which impacted the concordance estimates in arbutamine-thallium studies[33].

---

[30]I purposefully modify the sponsor's term "ischemia" to avoid implying that classifications so generated are necessarily valid representations of the presence or absence of ischemic pathophysiology. All that is truly known at the end of these arbutamine-echo experiments the extent to which the test outcome classifications were concordant with a standard for the diagnosis of the presence of an anatomic lesion; we have not gained knowledge of the mechanistic basis for any concordance, or the degree of concordance that is provided relative to an agreed-upon standard for the diagnosis of the functional state known as ischemia.

[31]this classifies outcomes based on immediated post-stress images, irrespective of what is observed with delayed imaging.

[32]This criterion employs a finding, known as tracer "redistribution, which it is not a sufficiently reliable surrogate for the status of rest images. For technical reasons which will not be detailed here, when using these tracers the finding of apparently partially redistributing or apparently nonredistributing post-stress perfusion abnormalities is not adequately predictive of the presence of an observable perfusion abnormality at rest.

[33]I take this same view to the arbutamine-sestamibi data.

6 RECOMMENDATIONS !

I make the following recommendations contingent upon our finding, in the final analysis, that clinical issues related to safety[34], device-performance, or data integrity do not prove to adversely impact the approvability of this NDA.

6.1. Recommendations regarding Approval:

6.1.1 Arbutamine-Thallium testing:

I recommend approval of arbutamine as an adjunct for use with thallium-based perfusion scintigraphy for the purposes of diagnosing epicardial CAD in patients who cannot exercise adequately.

6.1.2 Arbutamine-Sestamibi testing:

I recommend approval of arbutamine as an adjunct for use with sestamibi-based perfusion scintigraphy for the purposes of diagnosing epicardial CAD in patients who cannot exercise adequately.

6.1.3 Arbutamine-2D Echocardiographic testing:

Until and unless re-consideration of my aforementioned conclusions is compelled by data or arguments of the type elicited of the sponsor, I recommend nonapproval of this agent as an adjunct for use with 2 dimensional echocardiography for the purposes of diagnosing epicardial CAD.

6.2 Recommendations regarding Labelling:

6.2.1 Statement of Indication(s)

I recommend the we avoid use of the language, "indicated for the evaluation of CAD" which is found, for example, in the i.v. dipyridamole label. By use of the nonspecific term "evaluation of CAD" we give the appearance of having approved each of the multitude of applications of stress testing that are found in clinical use,[35] many of which are not adequately supported by the submitted arbutamine data. We should specify exactly which diagnostic use has been adequately supported by the submitted data. To not do so will leave us again vulnerable to exerting efforts in the future to address advertising excesses.

---

[34]at their face the data submitted in a new safety update do not give any preliminary suggestion of approval-limiting issues, but my final review of those data is pending.

[35]for example, uses such as for the putative diagnoses of ischemia, myocardial viability, "area-at-risk", and coronary flow reserve deficiency, as well as various prognostic applications in patients with known CAD diagnoses.

6.2.2 **Rational labelling of relationship of test utility & test interpretation to pretest risk**:

I recommend employing simple language (such as that used in my comments about Bayesian implications) which conveys the well-demonstrated importance of users' incorporating pre-test likelihoods into their thinking about the potential utility of and/or interpretation of arbutamine tests.

For the reasons enumerated above, I recommend against acceptance of the sponsor's proposal to write a label which reports sensitivities and specificities according to pre-test likelihood of disease.

6.2.3 **Avoiding misleading references to predictive values**

I recommend against writing a label which makes any reference to positive or negative predictive values, insofar as the prevalence-dependence of these metrics makes them of no generalizable descriptive use. Moreover it is advisable to dissuade any potential advertising pitches aimed at marketing the apparent benefit of the tests high positive predictive value (a characteristic largely engendered by the high disease prevalence in the tested populations).

These indices are population-specific, for example, the reported positive predictive values were biased towards high numbers by the high disease prevalence in the studied populations. The inclusion of positive predictive value invites mistaken expectations on the part of clinicians who apply the test to less highly selected populations, and invites the misleading use of the metric in drug advertising.

6.2.4 **Labelling of Arbutamine test use immediately following exercise**:

Given another recent experience consulting to the marketing division in regard to advertising for a pharmacologic stressor drug, I recommend that the label describe the potential (albeit theoretical) for risks of rapid sequential exposure to exercise stress followed by arbutamine, and suggest the imposition of a delay between exercise and the subsequent administration of arbutamine.

6.2.5 **Future class-labelling recommendation**:

It is recommended that in the future the labelling of this entire class of drugs be simultaneously revised to disclose that the underlying scintigraphy study designs do not allow estimation of the extent to which diagnoses were obtainable in the absence of the drug stressor.
Such information is inherently useful to users, and also serves to encourage future pharmacologic stress-scintigraphy study designers to undertake to eliminate the potential "drug overcrediting" bias.

Steven Rodin, MD 3/19/97

Steven M. Rodin, MD Date
Medical Officer

cc: HFD-110/ division file, GBuehler, RFenichel, RLipicky; ***no copy to SRodin**
cc: HFD-101/ RTemple
cc: HFD-710/ Mahjoob, Chi.

Appendix 1:

A.1 Unpooled Arbutamine-Thallium data

A.1.1 Concordance of test classifications with the Angiographic standard

Table: 12

Arbutamine-THALLIUM tests' Sensitivity and Specificity, vis a vis the diagnosis of the angiographic presence of a 50% diameter-narrowing epicardial coronary stenosis (unpooled results of studies 141, and 127)

test classification criteria: "Scinti-CAD"
datasets: blinded/centrally-interpreted (both studies); evaluable (study 141), intent-to-treat/arbut test completers (127)

Metric	*Arbut-THALLIUM*	
	study 127	study 141
SENSITIVITY mean & 95% CI	87% (79-92%) [n= 112]	63% (35-85%) [n= 16]
SPECIFICITY mean & 95% CI	25% (3-65%) [n= 8]	58% (28-85%) [n= 12]

[source= tables 4 & 5 of submission 3/17/97]

MAR 2 6 1996

Steven M. Rodin, M.D.
Medical Officer

Food and Drug Administration

Division of Cardio-Renal Drugs
Tel 301-443-0320; FAX-9283

Medical Review of NDA Safety Update

1 Application Identifiers

NDA #:	**20-420**
Drug:	**i.v. arbutamine**
Sponsor:	Gensia Inc.
Proposed indication:	diagnostic adjunct
Pharmacologic type:	nonselective β-adrenergic agonist, and low potency α-adrenergic agonist
Date of Safety updates:	25 October 1995, and 18 July 1995
Latest data submission:	14 March 1996
Review last revised:	26 March 1996

2 Findings:

2.1 Background:

The original safety review of Dr. Frank included 804 subjects evaluated up to April 1994. The sponsor has since submitted updated adverse event (AE) information which brings the safety database to 1037 subjects through a data cutoff point of 4/1/95 [submissions dated 7/18/95 and 10/25/95].

Most of the studies not previously reported used device algorithm revision #5[1] which continued infusions to higher levels of achieved heartrate than did the algorithm which was used in pivotal trials (i.e. revision #4).

What are summarized here are the aggregate event rates inclusive of the previous and updated safety data.

2.2 Deaths:

There are still no deaths reasonably attributed to arbutamine.

One newly reported death was secondary to peripheral arterial bypass graft rupture 4 days post-surgery, one was secondary to myocardial infarction (MI) occurring 1 day after aortic valve replacement, and a newly reported fatal cardiac arrest occurred 17 days after exposure to arbutamine.

2.3 Serious Nonfatal AE:

The cases of AE which were investigator-classified as serious and attributable to arbutamine are described here (see also the summary tabulation on page 4). For each of these serious AE the patient was reported to have manifested a recovery from the event.

2.3.1 Ventricular tachyarrythmias:

There were four serious cases of ventricular tachyarrythmia (3 ventricular fibrillation (Vfib) and 1 sustained ventricular tachycardia (VT) case), described as follows:

Patient 122-24-1206 (case 1) was a 50 year old male with history of prior MI and triple vessel coronary artery disease (CAD). The arbutamine infusion was discontinued secondary to chest pain and EKG changes. During recovery the patient developed atrial fibrillation (Afib) which deteriorated to VT and Vfib. Vfib was successfully defibrillated and recovery was reportedly uneventful.

---

[1]studies 132, 138, 139, 140, 141, and 142 used this revision #5.

[cont] **Ventricular tachyarrythmias**:

Patient 122-18-1291 (case 2) was a 58 year old male with CAD. Heartrate (HR) reached a plateau, and a "saturation alarm" sounded. Arbutamine infusion was nonetheless restarted and Vfib subsequently occurred. A first Vfib episode spontaneously reverted to sinus tachycardia, but 3 additional episodes in the subsequent 4 minutes required defibrillation. The patient was hospitalized for 48 hours and was discharged reportedly without further treatment or sequelae.

Patient 136-01-0002 (case 3) was a 69 year old male with history of MI and triple vessel CAD. HR reached a plateau and a "saturation alarm" sounded, yet the arbutamine infusion was restarted with subsequent onset of Vfib. There were 5 episodes of Vfib over a 10 minute period. All were reportedly responsive to defibrillation. The patient received antiarrhythmic pharmacotherapy and reportedly recovered without sequelae.

Patient 139-02-0026 (case 4) was a 63 year old male with history of MI. The arbutamine test was terminated following asymptomatic, sustained (>30 beat duration) VT of two minutes duration. Metoprolol was administered and recovering was reportedly uneventful.

Other Ventricular tachyarrythmias:

The total rate of any reported drug-associated ventricular tachyarrythmia (Vfib or VT) was 1.9%. These included nonsustained, or spontaneously resolving VT with < 30 total ectopic beats.

2.3.2 **Atrial fibrillation**:

There were two reported cases of serious Afib attributable to drug, described as follows:

Patient 122-03-0098 (case 5) was a 58 year old male with CAD. HR reached a plateau with sounding of the "saturation alarm", yet the arbutamine infusion was restarted and asymptomatic Afib subsequently occurred and lasted for 3 hours before uneventful pharmacologic conversion.

Patient 127-59-0269 (case 6) was 67 year old male with a history of MI. After 8 minutes of arbutamine the patient developed sustained AF which spontaneously converted to sinus rhythm. The patient was observed overnight and discharged without further events.

Other atrial arrhythmias:

The reported rate of any Afib was 0.8%. The reported rate of supraventricular tachycardia[2] (SVT) was 1.3%.

[2]it appears that by this the sponsor means SVT other than Afib.

[cont] **Serious Nonfatal AE:**

2.3.3 Myocardial Infarction:

Patient 122-23-1778 (case 7) had an acute MI approximately 20 hours after arbutamine exposure. There was no description of the phenomenology observed during drug exposure.

2.3.4 Unstable angina:

Patient 122-26-1162 (case 8) manifested unstable angina during physical exertion 1 hour after arbutamine exposure.

2.3.5 Summary Tabulation:

The following table summarizes the above described cases of serious AE.

Table: 1

Summary of Serious AE investigator-attributed to Arbutamine

Case number	*Subject identifier*	*Event*	*Comment*
1	122-24-1206	Vfib	--
2	122-18-1291	Vfib	occurred after infusion restart following a HR plateau
3	136-01-0002	Vfib	occurred after infusion restart following a HR plateau
4	139-02-0026	sustained VT	--
5	122-03-0098	Afib	occurred after infusion restart following a HR plateau
6	127-59-0269	Afib	--
7	122-23-1778	MI	occurred 20 hours after end of infusion
8	122-26-1162	Unstable angina	occurred 1 hour after end of infusion

[source: modification of table 14, pg 102, submission dated 10/25/95]

2.4 Common Adverse Events:

The most frequently reported AE are on the next page. Among the hypotension cases 81% required intervention (i.e. discontinuation of infusion and/or administration of medical therapy).

[cont] **Common Adverse Events:**

Table: 2

Most frequent (≥3%) Adverse Events

Event	*Number (%)* *[n= 1,037]*	
Any event	526	(51%)
Tremor	162	(16%)
Headache	104	(10%)
Arrhythmias	99	(10%)
Hypotension	59	(6%)
"Dizziness" (undefined)	55	(5%)
Flushing	53	(5%)
Paresthesia	33	(3%)
Nausea	30	(3%)
Anxiety	27	(3%)

[source: modification of table 13, pg 100 of vol 1, submission dated 10/25/95]

3 **Comments:**

a. The aggregate AE rates in this update are generally comparable to those reported in the previous review. There is now a somewhat clearer sense of the proportion of observed arrhythmias which was associated with restart of infusion following a HR plateau.

b. The sponsor has pursued development with a different infusion algorithm than the one with which the bulk of the safety data were obtained. This new algorithm (revision #5) is intended to achieve a higher HR and thus presumably involves exposure to higher and/or more prolonged levels of arbutamine. It is plausible that revision #5 will prove to have a higher rate of drug-related AE, although it remains to characterize this with certainty. Large safety differences between these algorithms are not apparent in the relatively small sample sizes presently reported.

c. A previous reviewer reported the comparative safety of adenosine stress testing yet, because the necessary information was in a state of obscurity at the time, that description is not correct. I have uncovered evidence of reported cases of adenosine-associated arrhythmic deaths in the Adenoscan® NDA database.

Steven Rodin, MD 3/26/96

Steven M. Rodin, MD Date

cc: HFD-110/ division file, GBuehler, RFenichel ***no copy to Rodin**

MAR 1 1 1996

Steven M. Rodin, M.D.
Medical Officer

Food and Drug Administration

Division of Cardio-Renal Drugs
Tel 301-443-0320; FAX-9283

Medical Review of Major NDA Amendment

1 Application Identifiers

NDA #:	**20-420**
Drug:	**i.v. arbutamine**
Sponsor:	Gensia Inc.
Proposed indication:	diagnostic adjunct
Pharmacologic type:	nonselective β- and α-adrenergic agonist
Date of major Amendment:	25 October 1995
Review initiated:	22 January 1996
Latest data submission:	7 March 1996
Review last revised:	11 March 1996

2 Table of Contents:

3 Background:

3.1 Recap of General background:

Arbutamine is a catecholamine proposed for use as an adjunct to various cardiac diagnostic techniques (electrocardiography (EKG), 2-dimensional (2D) echocardiography, and thallium^{201} perfusion imaging)[1] for the diagnosis of coronary artery disease (CAD). An NDA was submitted on 12/23/93 with the proposal for the drug to be marketed as part of a drug-device system called the GenESA® System. In this system a feedback controlled ("closed-loop") device aims to monitor heart rate (HR) response, and to adjust the rate of drug infusion in order to achieve a pre-specified rate of HR rise and maximum HR.

My review of the NDA's efficacy and clinical pharmacology data was filed on 10/26/94. The pivotal efficacy studies were EKG study 122, echo study 123, and thallium study 127. The safety data were then reviewed by Dr. Karen Frank, and infusion device data were evaluated by the Center for Devices and Radiologic Health. Dr. Temple (ODE I, CDER) issued a not approvable letter on 4/6/95.

3.2 Review Timeline:

The sponsor submitted this major NDA amendment on 10/25/95, and subsequently made numerous additional submissions (many of which I requested) through 3/7/96.

I was fully engaged in a 1P priority NDA review at the time of submission of this NDA amendment. I began this present review on 1/22/96. The evidence and arguments submitted in the sponsor's major amendment (and highly relevant subsequent submissions up to 3/7/96) are here evaluated for their bearing on issues identified in the nonapproval letter of 4/6/95.

APPEARS THIS WAY
ON ORIGINAL

[1] in the pivotal perfusion imaging study (127) the only tracer used was thallium^{201}.

3.3 **Operational definitions**:

Gold standard:

In this NDA the presence of significant CAD was defined as coronary angiographic evidence of ≥50% reduction in cross-sectional diameter of any "major" epicardial coronary artery ("major" refers to the left main, left anterior descending, left circumflex, or right coronary arteries or any of their major branches).

Sensitivity:

Sensitivity describes the test's capacity for making a correct diagnosis in confirmed cases of *disease*[2]. A high sensitivity test frequently makes the correct (True Positive (TP)) diagnosis among patients with disease, and infrequently makes the incorrect (False Negative (FN)) diagnosis among patients with disease. Sensitivity is calculated as a fraction whose numerator is the number of TP test results, and whose denominator is (assuming exclusion of indeterminate test results) the number of patients with confirmed disease. The number of patients with disease is equal to the sum of the number of TP and FN test results, and sensitivity is calculable as TP/(TP + FN). The complement of sensitivity, False Negative Fraction, is the fraction of diagnoses which are *incorrect* among patients with disease, and is calculated as FN/(TP + FN). Sensitivity is equal to 1 minus the False Negative Fraction.

Specificity:

Specificity describes the test's capacity for making a correct diagnosis in confirmed cases of *no disease* (rather, none of the disease of interest). A 100% specific test generates positive test results only for the specific reason of disease being truly absent. A high specificity test frequently makes the correct (True Negative (TN)) diagnosis among patient with no disease, and infrequently makes the error of detecting things that are not real disease (such detections are false positive (FP) diagnoses among undiseased patients). Specificity is calculated as a fraction whose numerator is the number of true negative (TN) test results, and whose denominator is the number of patients confirmed to not have the disease of interest. The number of patients with no disease equals the sum of TN and FP test results (assuming exclusion of indeterminate test results), and specificity is calculated as TN/(TN + FP). The complement of specificity, False Positive Fraction, is the fraction of diagnoses which are *incorrect* among patients without disease, and is calculated as FP/(TN + FP). Specificity is equal to 1 minus the False Positive Fraction.

[2]as opposed to Specificity, which is the test's capacity for making a correct diagnosis in confirmed cases of *no disease*.

4 Sponsor's Argument-based rebuttals:

4.1 Normalcy metric as estimator of Specificity in Patients:

Issue: In my original review[3] I described the methodologic limitations of the normalcy metric obtained in an essentially normal population.[4] I posited that the normalcy metric was plausibly overestimating the specificity of arbutamine testing that was achievable in patients with suspected CAD.

The sponsor retorts that normalcy is a valid surrogate for specificity measures in patients with suspected disease, and that we misunderstood the demographic covariate distribution of the population in study 128.

Normalcy is defined as the ratio of negative test outcomes to all test outcomes in a population presumed (not unreasonably) to be free of the disease of interest. In such a population the negative test outcomes are assumed (not unreasonably) to be true and the positive test outcomes are assumed to be false, so the the normalcy ratio (i.e. negative outcomes to all outcomes) reduces to a specificity measure [TN/(TN + FP)], with the major difference being that the population in which the estimate is obtained is not one of suspected CAD patients.

In the original NDA database, irrespective of the diagnostic imaging modality used, the point estimates of normalcy (from study 128) were not only uniformly higher than the point estimates of specificity obtained in patients (pivotal studies 122, 123, and 127), but there was not even overlap of the 95% confidence intervals for the two metrics. This observation engendered my concern that the subjects in study 128 may have been quite different from the patients in the pivotal trials.

Based on my view that the cardiac comorbidity factors well known to be specificity-degrading are plausibly more prevalent in CAD-suspected patients than they are in essentially normal populations, I proposed that a finding of only few false positive results in essentially normal subjects is no assurance that false positivivity will not be far more frequent in patients for whom the test is intended.[5] I had in mind such comorbidities as resting ventricular wall motion abnormality, and prior myocardial infarct (MI), since these are recognized to reduce the specificity (i.e. increase the false positive rate) for the diagnosis of CAD by echocardiography or thallium imaging[6].

---

[3]see pages 10-14, and 87 of my report.

[4]I use the term "normal" to conveniently refer to persons either confirmed or reasonably presumed to be free of the disease of interest.

[5]here referring to those with a clinical suspicion of active coronary disease.

[6]Some false positive thallium scans occur when necrotic myocardium fails to take up tracer despite the infarct-related artery (IRA) bearing no critical stenosis. Post-MI, in many the IRA is <50% stenosed either because of an intervening angioplasty or because there was only ever a subcritical plaque upon which a thrombus formed transiently, and then was lysed.

Comments (re normalcy metric):

After studying this major efficacy amendment I requested further characterization of the comparative distribution of specificity-degrading (i.e. false positive-eliciting) comorbidity factors in study 128 subjects vs the CAD-suspected patient populations. The new data[7] indicate that, in fact, there was a much lower prevalence of false positive-eliciting comorbidity factors in the population in study 128 than in patient samples in the pivotal studies. In normalcy study 128 only less than 2% of subjects manifested resting echocardiographic wall motion abnormalities, as opposed to 40% of patients in pivotal echocardiographic study 123. Similarly, in normalcy study 128 only less than 5% of subjects manifested baseline EKG-Q waves, as opposed to 28% of patients in pivotal patient studies.

It is clear that a much lower prevalence of specificity-degrading comorbidity factors existed in study 128 than in pivotal patient studies. These confounding variables are a plausible basis for the normalcy metric having systematically overestimated that specificity which was achievable in patients in the pivotal trials, i.e. patients with with suspected CAD for whom the test would be intended in any marketed use.

These new data only further substantiate my previously posited arguments. The sponsor's contention that the normalcy metric in study 128 was an adequate surrogate for specificity in patients is not compelling.

APPEARS THIS WAY
ON ORIGINAL

APPEARS THIS WAY
ON ORIGINAL

[7]submissions of 1/23/96, and 3/7/96.

4.2 Ethical Considerations in Applying or Not Applying Methods of better estimating Specificity:

Issue: An issue of major concern to this Division and Office is the extremely wide variance of the specificity estimates generated in the arbutamine NDA. The sample sizes of angiogram-negative patients were small, so a potential approach to improving the precision of the estimates would be to further investigate the drug by increasing the sample size of angiogram-negative patients. The sponsor contends in this efficacy amendment that because of the potential risks of the validating angiography procedure it is "ethically inappropriate to perform coronary angiography in the face of a negative non-invasive test".

Comments: I am unable to identify a reasonable basis for perceiving an ethical breach in the performance of a consented-to coronary angiography procedure performed for the purpose of validating a negative outcome of an *investigational* test. In contrast, careful consideration of the matter illuminates the need to further explore the ethical dimension of failing to perform such validation of investigational tests.

Let us reflect on the investigational nature of the test in order to underscore the objective limits of knowledge one has of the utility of the test during the test's development. It appears that one of the conceptual flaws made by opposing ethical formulations of this issue is an inadequate concession to the epistemological fact that an investigational test is not reasonably construed to bear accurate diagnostic information in the absence of validation of that assumption.
The sponsor's proposal that it is unethical to perform a potentially risky validation of a negative [investigational] test result assumes that one can reliably construe a negative result of this investigational test as being uniformly truly negative. It is clear that this construction begs the very question being asked by the investigation, that the history of diagnostic research points to the implausibility of a test generating negative results which are only ever true and never false,[8] and that the data thusfar obtained for arbutamine testing show an appreciable false negative rate at the upper bound of the 95% confidence intervals (this limit ranged from in pivotal studies).

What then is the ethical basis for further investigating arbutamine testing by means of a larger sample of angiogram-negative patients? The first basis is the one which justified the approval of the first IND research project for this product, despite the research having had the potential for introducing risks to subjects. That is, the already available diagnostic methods are suboptimal in their information value, safety, convenience, and/or applicability to patients with suspected CAD who are unable to exercise, and the new research has the potential to adequately characterize a clinically useful method.

As in the case of many investigational therapeutic agents, there are at least two hoped-for benefits, but only one definite benefit associated with the conduct of the research. There exists the hoped-for patient benefit of the test providing relatively unerroneous and unambiguous diagnostic information for application to the individual patient, but the reality of this benefit is poorly known during the

[8]when such results (i.e. 100% sensitivity) have been claimed it was generally because there was acceptance of a rather high rate of falsely *positive* test results.

[cont] **Ethical Considerations**:

validation process. Is it ethical to inform a patient that, independent of the design of the research, a benefit that can potentially accrue to he or she is the benefit of diagnostic information about their suspected disease provided by the as-yet-unvalidated test? The most defensible position is that the potential for such a patient-accrued benefit can only be ethically asserted in a consent form if the design of the research is adequate to at least diagnostically validate[9] the information rendered by the test, and to establish its safety profile.

Societal benefit also weighs heavily in these considerations. In the conduct of diagnostic research there exists the definite societal benefit of obtaining data which convincingly answers whether or not the new method is useful with respect to diagnostic error and safety. There is also the hoped-for social benefit of ultimately discovering that the answer to this question affirms the utility of the as-yet-unvalidated test. Is it ethical to inform a patient that, independent of the design of the research, these benefits can definitely or potentially accrue to society? Again, the most defensible position is that the potential for such social benefits can only be ethically asserted in a consent form if the design of the research is adequate to validate the diagnostic truthfulness of the test (and its safety).

It follows that from the perspective of research *benefit* alone, when a patient consents to enroll on a basis which assumed that the study design was adequate to validate the benefits, it can be reasonably argued that it is not ethically appropriate to forego methodologic steps which are necessary for such validation. Similarly, from the perspective of research *risk* alone, when a patient consents to enroll on a basis which assumed that the research was adequate to validate the risks, it can be reasonably argued that it is not ethically appropriate to forego methodologic steps which are necessary for such validation.

Why would one posit that gold-standard validation is necessary for an adequate validation of benefit? Consider what happens when a development program fails to fully perform such validation (e.g., because of incomplete discouragement of the post-test referral biases which may manifest in the clinical community). There are predictable results of a partial validation failure wherein a fair number, but not all, of the negative investigational test results remain unvalidated: this reduces the sample size of confirmed cases of no disease[10] and thus degrades the resulting estimate of specificity such that its variance becomes so large as to become unreliable. A complete validation failure (wherein all of the negative test results remain unvalidated) entirely undermines the goal of characterizing specificity as well as the goal of characterizing sensitivity. When one cannot distinguish the true negatives from the false negatives neither sensitivity or specificity are calculable. All that can then be characterized is the predictive value of a positive test. The utility of this metric

---

[9]issues of the clinical significance of diagnostically valid information also have substantial importance, but will not be touched upon here. Thus the term "benefit" in this discussion does not imply the clinical benefit of a correct diagnostic classification.

[10]this reasonably assumes that the test has more than zero sensitivity, and thus some of the cases of negative test results actually do not have disease.

[cont] **Ethical Considerations**:

for characterizing the intrinsic discriminative properties of the test is grossly limited by its dependence on disease prevalence. This dependence results in skewing towards high positive predictive value in scenarios like this where the only confirmed cases are cases wherein disease is present, i.e. where the prevalence of confirmed disease is high.

Apart from the undermining of the opportunity to characterize benefit, when a study fails to diagnostically validate the test results this invites the potential for considerable patient risk whereby a physician acts upon the presumed truth of test results which are actually diagnostic errors. For example, if the test falsely declares that disease is absent this would plausibly lead a physician to withold, on the basis of misinformation, a useful intervention. Conversely, if the test falsely declares the disease to be present it is plausible that the physician would, on the basis of this misinformation, undertake diagnostic and/or therapeutic interventions which can provide no benefit, and yet which confer the potential risks of uncomfortable and/or harmful adverse effects for the patient.

This does not deny that there are documented potential risks of diagnostically validating the test results by a gold standard (which in this case is invasive coronary angiography). However, these are unfortunately very difficult to avoid in the conduct of information-reliable diagnostic research. Moreover, these are the risks which, in an adequately informed and consenting subject, are ethically counterbalanced by the definite societal benefit and potential patient benefit which accrue *only by* validation of the truthfulness of the diagnostic information being rendered by the investigational test.

It is recognized that drug developers face a reality wherein investigators at times resist the referring of patients with negative test results to a confirming gold standard procedure (post-test referral bias). In some cases this might be on the basis of a preference for relying on the results of an approved noninvasive test. If that preference is convincingly data-supported then this might require reconsideration of the depth of rationale for undertaking the new research. However, in the case of methodology for the diagnosis of CAD the NDA experience in this division indicates to me that adequate and well-controlled data do not necessarily suppport that conventional methods are so error-free as to be justifiably relied upon in this way. The extent to which these data are known to the physician community, as opposed to numerous optimistic published data of oftentimes lesser quality, appears to be quite limited.

5 Sponsor's Data-based rebuttals:

5.1 Data from adequate and well-controlled trials:

5.1.1 Prevalence of Near-threshold CAD among patients with False Positive Thallium tests:

Issue: The sponsor proposes that false positive test results to some appreciable extent occurred in patients with angiographic disease near the threshold definition of abnormal (i.e. nearly 50% lesions). Their view is that these false positives are attributable to incorrect classifications by the angiographic standard, i.e. that these cases represent clinically important coronary disease-associated cardiac dysfunction which angiography failed to detect. To support this view they submit descriptions of the rate of "near threshold" (40-49%) epicardial stenoses among patients with false positive arbutamine-thallium tests[11].

New analyses: Note that these are intent-to-treat data obtained from an adequate and well-controlled trial (study 127) in which angiograms were to have been evaluated without the knowledge of the imaging test results or any other clinical data.

Six false positive arbutamine-thallium results were obtained among 8 patients with negative angiograms. Reportedly 2 of these 6 cases had blinded angiogram results showing 40-49% reduction in cross-sectional diameter of a major epicardial coronary artery, and are reputed to represent failures of the gold standard to define clinically significant disease.

After studying this portion of the efficacy amendment I recognized that the process of inferring that the 40-49% lesion functionally contributed to the imaging abnormality would be flawed if the image findings were in regions other than the ones subtended by the nearly critically stenosed artery. Therefore I requested additional information about the spatial correlation between image abnormality and near threshold lesion for these cases. These data revealed that in neither of the 2 thallium cases of putative failure of the gold standard did the location of "near threshold" anatomic CAD correlate well with the ventricular region of reduced thallium activity.[12] That is, in both cases the thallium scan defects were observed in what was interpreted at the core laboratory to be the usual supply of the right coronary artery, while the "near threshold" angiographic stenosis was in the left anterior descending artery.

[11]echocardiographic data from study 123 were also analyzed in this way, and are presented in section 7 below.

[12]the source of these data is table 1, pg 6 of sponsor's fax of 3/4/96 at 9:34 am; also page 4 of hardcopy submission dated 3/4/96.

Comments:

Only 2 cases of "near threshold" disease were found among the patients with false positive arbutamine-thallium tests and neither one had a convincing correlation between the location of the thallium scan abnormality and the location of epicardial subthreshold narrowing. While the assessment of function-anatomy correlation is not infallible (insofar as there is inter-patient variability in the region of myocardium supplied by a given epicardial artery), conventional generalizations are recognized[13], and it is these which I applied.

The submitted evidence is sufficiently weak to discount the proposal that false positive thallium tests were largely attributable to the presence of coronary disease which was *actively* functionally significant despite being below the defined angiographic threshold.

The more plausible causes of false positive arbutamine-thallium scans are extracardiac tissue attenuation artefacts, or the presence of inactive disease such as remote, but quiescent MI (as suggested by the prevalence of resting EKG-Q waves in this study). When false positives result from these there is little reason to consider the classification of the test to be an error of the angiographic gold standard since the test did not correctly characterize the current anatomic status of the infarct-related artery.

APPEARS THIS WAY
ON ORIGINAL

APPEARS THIS WAY
ON ORIGINAL

[13]such as those summarized on page 88-89 of Hurst's The Heart. R. Schlant-editor. 8th edition. McGraw Hill, New York.

5.1.2 Impact of more broadly defining a Positive Arbutamine-Thallium test:

Issue: The sponsor proposes that higher arbutamine-thallium test utility can be achieved if the imaging results are integrated with EKG and chest pain symptom findings. They undertook what were essentially analyses of diagnostic sensitivity after expanding the previous definition of a positive noninvasive test from one which produced a positive imaging finding to one which produces either a positive image, a positive EKG, *or* exertional chest pain[14].

After studying the resultant data I needed to request a characterization of the impact of this definition-broadening on specificity, because the submitted sensitivity data were uninterpretable in their absence.

New analyses: These are analyses based on blinded, centrally adjudicated EKG readings, but unblinded assessments of chest pain by site investigators in thallium study 127. Patients with indeterminate (nondiagnostic) imaging results were included[15] (this being inconsistent with the approach used for sensitivity and specificity calculations in the original NDA submission).

After expanding the definition of a positive test the resultant mean sensitivity (relative to that obtained with the un-expanded definition) for arbutamine-thallium increased (not unexpectedly) from 78% to 93%, but at the important cost of a reduction in mean specificity from 25% to 20% (the 95% CI for specificity is 3-56% with the more inclusive definition of a positive test)[16] [source: page 11 of sponsor's fascimile submission of 3/4/95 at 9:34am; reproduced on page 9 of hardcopy submission of 3/4/95].

Comments:

Given that EKG and symptoms findings are known to have low specificity for the diagnosis of CAD, it was predictable that the inclusion of these parameters in the definition of a positive noninvasive test would only attain a higher sensitivity at the expense of a reduction in specificity. Insofar as low mean specificity (and highly variable estimates thereof) has thusfar been a major criticism of arbutamine tests for CAD, these methods which only further reduce specificity make no substantive regulatory contribution.

---

[14]the sponsor renamed this metric "clinical utility" to distinguish it from the prespecified sensitivity analyses.

[15]on this basis 12 additional echocardiogram patients and 8 additional thallium patients (2 of which had negative angiograms) are included in the new analysis.

[16]directionally and quantitatively similar changes were observed with arbutamine-echocardiography. See section 7 below.

5.1.3 **Sensitivity in patients able to exercise only submaximally:**

Issue:

The patients studied in this NDA were able to exercise whereas the proposed indication is for patients unable to exercise. The sponsor provides analyses which aim to support the view that arbutamine test sensitivity is not influenced by a subject's level of ability to exercise.

New analysis:

The sponsor examined test sensitivity among patients with moderate impairment in the ability to exercise, i.e those able to exercise but unable to reach 85% of the maximal age-predicted heartrate during this exertion. The mean sensitivity of arbutamine-thallium testing in such patients was reportedly the same in the 23 patients unable to attain target HR *during exercise*, as it was in the 89 who attained target during exercise[17].

Comments:

Within the limits of the power of the reported analysis, the sensitivity of arbutamine-thallium testing did not appear to be grossly related to the patients' ability to achieve target HR during exercise. Specificity data were not reported.

To comfortably extrapolate test performance in patients able to exercise (albeit submaximally) to those unable to exercise one needs consider the features of the two populations. In my stress-test experience the most common reasons for a patient's nonattainment of target heartrate during exercise have been hemodynamic deconditioning,[18] or the onset of diagnostically nonspecific but conventionally exercise-limiting events such as EKG changes. Patients completely unable to exercise are generally sufferers of peripheral vascular disease, diabetes mellitus-associated lower extremity amputation, or chronic obstructive lung disease. Judgements about the validity of this extrapolation are possible, but their reliability is not clear.

[17]for arbutamine-echocardiography in study 123 the reported sensitivity was 71 vs 78% in those unable (n=42) vs. those able (n=85) to achieve the exercise target HR, respectively.

[18]in the deconditioned patient it is fatigue which frequently results in discontinuation of the test prior to the attainment of target HR, even though the HR manifested at a given exercise duration is higher for the deconditioned subject than for the conditioned subject.

5.2 Perfusion Imaging Data from other trials:

5.2.1 Study 141:

Issue:

The sponsor proposes that concerns about the wide variance of specificity estimates in the original NDA can be mitigated by consideration of newly submitted test results in angiogram-negative patients in study 141.

New Data:

The sponsor has submitted a preliminary and unplanned interim analysis of study 141, inclusive of subjects whose case report forms were received by the sponsor by 11/1/95. This multicenter study was not designed to rigorously support the diagnostic efficacy of arbutamine. This is an ongoing study of patients with suspected CAD which has the prespecified objective of assessing safety and "the overall function and convenience of the GenESA System for conducting a stress test". Investigators can choose the imaging modality at their discretion[19]. While this study of patients employs some of the design features of the pivotal NDA studies (e.g. it administered iv arbutamine via a "closed-loop" infusion at an intended dose rate ≤0.8 μg/kg/min up to 10 μg/kg/d), it differs in important respects. Unlike the previously reported pivotal trials:

a. the arbutamine device infusion algorithm was not the revision #4 which was used in pivotal trials, but instead was algorithm revision #5. This method continued infusions to higher levels of achieved heartrate than did the previous method.

b. there was no formal blinding of noninvasive image interpretors to the angiographic or other clinical data.

c. there was no formal blinding of angiogram interpretors to the noninvasive imaging or other clinical data.

APPEARS THIS WAY
ON ORIGINAL

[19]in addition to thallium and Tc^{99m}-sestamibi data there were interim EKG, and 2D-echocardiography data submitted.

[cont] **Study 141**:

d. there were no data available[20] which allow estimation of the proportion of CAD-diagnostic perfusion scan abnormalities which were detectable under resting conditions in the absence of arbutamine (i.e. the proportion of CAD diagnoses which were obtainable had placebo been used as stressor agent). As a result one cannot exclude that a large proportion of correct diagnoses of the presence of CAD were obtainable in the absence of the stressor.

e. the criteria of noninvasive image and angiographic data interpretation were not prespecified or standardized across centers[21]. At individual sites the image interpretations were not adjudicated by a panel of observers, and no central core-lab adjudication of angiogram results was undertaken.

f. the methods of tracer administration, noninvasive image and angiographic data aquisition, and data processing were not prespecified or standardized across centers, and the details of site-specific methodology were not captured.

g. much of the perfusion imaging data was obtained using only Tc^{99m}-sestamibi as tracer[22]. It is not justified to pool the results obtained with these 2 tracers. As was predicted by consideration of tracer physics and detector instrumentation principles, direct comparator exercise trials in the published literature have suggested that Tc^{99m}-sestamibi has at least somewhat higher mean sensitivity and specificity than has thallium for the diagnosis of a 50% angiographic coronary stenosis[23].

h. it is unclear whether the sponsor's decision to submit this interim analysis was subject to any study selection bias[24].

---

[20]even if one could retrospectively capture the necessary comparisons of stress data with resting (or delayed post-stress) data, it is not clear that they would be interpretable because somewhat ambiguous dual isotope (thallium stress vs Tc^{99m}-sestamibi post-stress) methods were apparently in use.

[21]the sponsor has not even captured the details of the cite-specific diagnostic criteria.

[22]this tracer is FDA approved. One additional patient was studied with yet another tracer,

[23]the initial 3 published trials showed this despite not even optimizing the image aquisition and processing parameters to fully exploit the unique physical and kinetic properties of Tc^{99m}-sestamibi [source: Marcus et. al. Cardiac Imaging. W. B. Saunders Company. 1991, page 1103].

[24]clarification has been sought as to whether there exist other as yet unsubmitted study results.

[cont] **Study 141**:

Thallium tests in study 141:

At their face the arbutamine-thallium tests for the purported diagnosis of CAD showed a mean sensitivity of 46% (95% CI= 19-75%), and a mean specificity of 67% (95% CI= 22-96%).

Table: 1

Reported sensitivity and specificity (for the purported diagnosis of CAD) of Arbutamine-THALLIUM tests

[interim results of study 141]

Metric	*Thallium-based perfusion imaging*
Sensitivity	46% 95% CI= 19-75% [n= 13]
Specificity	67% 95% CI= 22-96% [n= 6]

[source: pg 2 of fax submission dated 3/7/96]

There was reportedly 1 indeterminate result excluded from the above analyses.

[cont] **Study 141**:

Tc^{99m}-Sestamibi results:

At their face the arbutamine-Tc^{99m}Sestamibi tests for the purported diagnosis of CAD showed a mean sensitivity of 82% (95% CI= 71-90%), and a mean specificity of 60% (95% CI= 26-88%).

Table: 2

Reported Sensitivity and specificity (for the purported diagnosis of CAD) of Arbutamine-Tc^{99m}-SESTAMIBI tests

[interim results of study 141]

Metric	*Arbutamine-Sestamibi (Tc^{99m}) perfusion imaging*
Sensitivity	82% 95% CI= 71-90% [n= 67]
Specificity	60% 95% CI= 26-88% [n= 10]

[source: pg 2 of fax submission dated 3/7/96]

There were reportedly less than 2 indeterminate tracer results, and these were excluded from the above analyses.

Comments (study 141):

1. The employment of a different-than-pivotal-trial infusion algorithm (which reportedly attains higher heartrates) makes it uncertain as to whether the diagnostic efficacy (or safety) results of even a rigorously designed version of study 141 could be validly extrapolated to the pivotal trials.

2. Even at their face the reported arbutamine-thallium results do not mitigate concerns about the low and highly variable specificity estimate obtained for arbutamine-thallium imaging in pivotal trial 127. Even at their face these tests only yielded a higher than previous mean specificity estimate (50% vs 25% in study 127) by sacrificing mean sensitivity (50% vs 87% in study 127), and the lower bound of the specificity confidence interval is still quite low (12%).

3. Since this was not an adequate trial to support an indication for pharmacologic stress testing, it is not actually reasonable to take these results at their face. Although some of the design defiencies would only have plausibly resulted in increased variance, others (such as the absence of blinding) discourage any confidence in excluding a role for bias (even unintentional) in the process of data interpretation. Although in 64% of thallium studies the image in which the image was reportedly *acquired* prior to angiography, there are no quality controls to prevent the reported scan interpretation from being finalized only after angiogram results are known.

4. Another confounding factor in interpretation of the sensitivity and specificity estimates is that it is unknown whether angiographic definitions of CAD were predominantly based on a criterion such as 75% stenosis, rather than the 50% stenosis threshold used in the pivotal studies.

5. Although at their face the reported Tc^{99m}-sestamibi results are perhaps more encouraging for the prospect of improving the methodology of the arbutamine test, there is no convincing basis for applying these results to anything other than hypothesis generation. The same design deficiencies discussed above make this an inadequate trial upon which to base an approval for arbutamine-Tc^{99m}Sestamibi stress testing.

APPEARS THIS WAY
ON ORIGINAL

5.2.2 **Study 130**:

Issue: The sponsor proposes that concerns about the wide variance of specificity estimates in the original NDA can be allayed by integration of newly submitted arbutamine-Tc^{99m}Sestamibi test results in study 130.

New Data:

Now submitted are final results which supplement the interim results of this study, as submitted in the original NDA (see my original review, pg 25-27)[25]. The main objective of this study was to assess the effects of arbutamine on central hemodynamics in patients with suspected CAD. It was not designed to rigorously support the diagnostic efficacy of arbutamine, and those imaging data which were obtained actually differed from pivotal thallium study 127 in important respects:

a. rather than thallium, the perfusion tracer was Tc^{99m}-sestamibi.

b. there was no formal blinding of noninvasive image interpretors to the angiographic or other clinical data.

c. all patients had coronary angiography performed first, and then in the immediate post-angiography setting arbutamine was administered[26].

d. perfusion imaging was performed according to a patient selection process the nature of which is not completely known. One important selection criterion is known: only patients who manifested an increase of ≥30 bpm over pre-drug HR were even eligible for tracer infusion.

e. the criterion for a positive perfusion scintigram for CAD appears to have been retrospectively defined.

f. the sample size was quite small (n=19).

g. Tc^{99m} sestamibi was injected into the right ventricle, as opposed to via a peripheral vein[27].

---

[25]note that in contrast to a finding which I originally gleaned from the original submissions, it is not correct that the definition of a positive sestamibi test for CAD was wholly invalid in this study, although it was retrospective.

[26]it is not clear whether the pivotal trial infusion algorithm (#4) was used.

[27]this may be of little significance given the preclinical evidence of rapid myocardial extraction of this tracer.

[cont] **Study 130**:

Those patients receiving any Tc^{99m} sestamibi tracer were administered 15-25 mCi, followed by single photon emission computed tomography (SPECT) ventricular perfusion imaging 30-60 minutes later, and resting re-injection images at 24-48 hours post-stress. The angiographic definition of CAD was identical to that used in the pivotal studies.

Tc^{99m}Sestamibi results in study 130:

At their face the arbutamine-Tc^{99m}Sestamibi tests for the diagnosis of CAD showed a mean sensitivity of 75%[28] (95% CI= 35-97%) in a sample of 8 patients with confirmed disease, and a mean specificity of 71% (95% CI= 29-96%) in a sample of 7 patients confirmed to have no significant angiographic disease [see the table below].

Table: 3

Sensitivity and specificity (for the diagnosis of CAD) of regional ventricular hypoperfusion observed immediately after peak hemodynamic stress with Arbutamine-Tc^{99m}-SESTAMIBI imaging

[study 130]

Metric	*Arbutamine-Sestamibi-Tc99m perfusion imaging*
Sensitivity	75% 95% CI= 35-97% [n= 8]
Specificity	71% 95% CI= 29-96% [n= 7]

[source: page 3 of fax submission of 3/5/96 at 10:36 Pacific time]

No perfusion images were classified as indeterminate (after application of a retrospective criterion for a nondiagnostic test).

[28]to an appreciable extent the true CAD diagnoses were obtainable by imaging at rest in the absence of Arbutamine. In half of the six patients with true positive perfusion defects observed at peak stress, at least one of the defects was also observed on resting tracer-reinjection images obtained 24-48 hours after arbutamine exposure [source: pg 1 of submission dated 3/5/96].

[cont] **Study 130**:

Comments (study 130):

The reported Tc^{99m}-sestamibi results of study 130 are only adequate for generating the hypothesis that an arbutamine-Tc^{99m}Sestamibi method may have clinical promise. On the basis of principles of radionuclide physics and detector instrumentation, and direct tracer comparator trials[29], these results (even taken on their face) cannot be reasonably extrapolated to support the thallium results of pivotal study 127.

Moreover, it is not justified to take these results on their face because the study design is inadequate to demonstrate diagnostic efficacy. Among the most important limitations are that the findings are potentially subject to interpretation bias, and patient selection bias, and to an appreciable extent the true diagnoses of the presence of CAD were plausibly obtainable by imaging at rest in the absence of arbutamine.

[29]as described in the comments to study 141 above, initial comparator trials suggest that Tc^{99m}-sestamibi has at least somewhat higher mean sensitivity and specificity than has thallium for the diagnosis of CAD.

6 Reviewer's Further Analysis of Sensitivity in Pivotal Thallium Trial 127:

Issue: No estimate has thusfar been generated of the extent to which correct diagnoses of the presence of CAD were obtainable (i.e. the extent to which sensitivity was attainable) by means of thallium imaging performed at rest in the absence of Arbutamine in study 127[30].

New analyses: Further analysis demonstrate that in 59% of the 97 patients with true positive thallium perfusion defects observed at peak stress, at least one of the defects was to some extent observed on delayed "redistribution" images obtained 3-5 hours after arbutamine exposure [source: pg 1 of submission dated 3/5/96].

Comments:

1. The thallium scans obtained 3-5 hours post-arbutamine are generally reasonable surrogates for resting, pre-arbutamine scans. From the vantage point of the pharmacologic adjunct, this is a sufficient delay for drug action to have disappeared (given arbutamine's 8 minute elimination half-life and comparable pharmacodynamic half-life)[31].

From the vantage point of the tracer, in most cases a 3-5 hour delay is sufficient for that thallium which had initially distributed into extracardiac compartments (after injection at peak stress) to recirculate and redistribute into any cardiac regions which are physiologically perfusable at rest. However, the absence of complete "redistribution" has been elsewhere shown to overestimate the proportion of resting perfusion defects which would have been detectable had exogenous thallium been re-administered at rest. The new analyses therefore suggest that to an appreciable extent (although probably modestly less than 59%) the test's sensitivity for the diagnosis of the presence of any CAD arises from the generating of true positive diagnoses which would have plausibly been obtained using even placebo as hemodynamic stressor agent.

2. The presence of so many true positive *resting* perfusion abnormalities is an indicator that the spectrum of patients sampled was over represented by those with high-grade lesions which caused resting perfusion abnormalities. Since the lesser (yet still clinically and angiographically significant, e.g. 51% stenosed) lesions are the only ones which require a hemodynamic stressor in order to detect a perfusion abnormality, a sample which does not adequately contain such cases does not fully challenge the sensitivity of the pharmacologic stress test.

[CONTINUED NEXT PAGE]

[30]an analogous issue was previously discussed in my original review of the arbutamine-echocardiography (study 123) data.

[31]see the discussion of study 108 in my original NDA review.

[cont] **Reviewer's Further Analysis of Sensitivity in Trial 127**:

3. The implication of this new analysis is that it remains unclear whether the dosing regimen of arbutamine used in pivotal study 127 was adequate for producing sufficient diagnosis-eliciting phenomena (such as coronary hyperemia) to detect the angiographically and clinically significant lesions which do not impair perfusion at rest.

4. It can be argued that arbutamine-thallium testing is possibly useful for more than the diagnosis of the presence of any CAD. There are clinicians who, for example, draw inferences about ischemia and myocardial viability from thallium redistribution data. However, such diagnoses were not diagnostically validated by the sponsor. Moreover, where literature reports are available for the validation of such diagnoses with respect some other non-reinjection stress-test methods, there is no absence of clinically important diagnostic error.

7 More Arbutamine-Echocardiography Findings in Pivotal trial 123:

7.1 False Positive Arbutamine-Echocardiography tests: Prevalence of Near-threshold disease:

Issue: The sponsor proposes that false positive arbutamine-echocardiography tests in study 123 were sometimes attributable to there being functionally significant, but anatomically subthreshold CAD.

New analyses: Seven of the 11 false positive cases had blinded angiogram results showing >40% (but less than 50%) reduction in cross-sectional diameter of a major epicardial coronary artery, and in 6 of these 7 (all but patient 34-5)[32] the location of the echocardiographic abnormality (i.e. an apparent stress-induced worsening of ventricular dysynergy) correlated reasonably well with the anatomic location of the lesion [source: table 1, pg 5 of fax dated 3/4/96]. However, I requested a further examination of the data which revealed that in 5 of these 6 cases there were observable pre-stress wall motion abnormalities (and these are recognized to be major cause of false positive stress-echocardiograms).

Comments: It is far more plausible that the observed resting abnormalities resulted from remote and now quiescent MIs, rather than being a real-time consequence of coronary lesions which obstruct the lumen by less than 50% (such subthreshold lesions, if they were ever to be flow-limiting, would be expected to be so limitin during stress, but not under resting conditions). The data do not support the proposal that these 6 false positive test results were attributable to the presence of CAD which was actively functionally significant.

7.2 Impact of more broadly defining a Positive Arbutamine-Echocardiography test:

Issue: The sponsor proposes that higher arbutamine-echocardiography test utility can be achieved if the imaging results are integrated with EKG and chest pain symptom findings.

Analysis: For echocardiography study 123, after expanding the definition of a positive noninvasive test from one which produced a positive imaging finding to one which produces either a positive image, a positive EKG, or exertional chest pain, the resultant mean sensitivity (relative to that obtained with the un-expanded definition) for arbutamine-echocardiography increased (not unexpectedly) from 76 to 93%, but at the cost of a reduction in mean specificity from 31 to 25% (the 95% CI for specificity is then 7-52%). [source: sponsor's fascimile submission of 3/4/95 at 9:34am, page 11; also page 9 of hardcopy submission of 3/4/95].

Comments: These findings are neither unexpected nor contributory to any refutation of the regulatory concerns about the low point estimate and wide confidence interval surrounding the NDA's original estimate of the specificity of arbutamine-echocardiography.

[32]in this case the region perfused by the 41% stenosed right coronary artery lesion would not usually include the location of observed wall motion abnormality, i.e. the anterior left ventricular wall.

8 Arbutamine-echocardiography and Arbutamine-EKG data from Study 141:

8.1 Arbutamine-Echocardiogram tests in study 141:

New analyses: At their face the arbutamine-echocardiogram tests for the putative diagnosis of CAD showed a mean sensitivity of 58% (95% CI= 48-68%), and a mean specificity of 72% (95% CI= 51-88%). See the table below. Although I define a positive echocardiogram for the diagnosis of the presence of CAD by the finding of stress-induced ventricular dysynergy,[33] there is no means for establishing the diagnostic criteria used in this study.

Table: 4

Sensitivity and specificity (for the putative diagnosis of CAD) of 2D echocardiography [interim results of study 141]

Metric	*Arbutamine-stress echocardiography*
Sensitivity	58% 95% CI= 48-68% [n= 131]
Specificity	72% 95% CI= 51-88% [n= 25]

[source: page 13 of fascimile submission dated 3/4/96; and table ?]

There were reportedly 4 indeterminate echocardiograms which were excluded from the above analyses.

[33]defined as cardiac wall motion abnormalities including hypokinesis (reduced wall motion), akinesis (absent wall motion), or dyskinesis (paradoxical systolic motion away from the chamber's center).

[cont] **Arbutamine-Echocardiogram tests in study 141:**

Comment:

a. At their face, the mean specificity is higher and the mean sensitivity is lower than in study 123, although the 95% CI do overlap. However, this was not an adequate trial to support an indication for pharmacologic stress testing, so the results are not reasonablly taken at their face (see the above discussion in section 5.2.1 of issues of blinding and modification of infusion algorithm). Moreover, resting data are not available to sort out whether in this study there was a lower prevalence of baseline wall motion abnormalities which then caused fewer false positives (and thus higher specificity), or whether there was higher prevalence of collaterized or surgically bypassed lesions which caused more frequent false negatives (and lower sensitivity). Neither can one exclude that the apparent differences in study outcomes are attributable to differences in the diagnostic thresholds (test "cutoff points"), because we do not know what diagnostic criteria were used in study 141.

8.2 Arbutamine-EKG tests in Study 141:

New analyses: At their face the arbutamine-EKG tests for the putative diagnosis of CAD showed a mean sensitivity of 37% (95% CI= 28-46%), and a mean specificity of 77% (95% CI= 58-90%). There were reportedly 13 indeterminate EKGs which were excluded from these analyses. See the table below.

Table: 5

Mean sensitivity and specificity (for the diagnosis of CAD) of *stress-induced* EKG ST-segment depression or elevation [interim results of study 141]

metric	***Arbutamine-EKG***
Sensitivity	37% 95% CI= 28-46% [n= 131]
Specificity	77% 95% CI= 58-90% [n= 30]

[source: fax submission dated 3/4/96; and table ?]

Comment: (study 141 EKG data):

Even taken at their face, the data do not support that this diagnostic method has a useful degree of discriminative power.

9 Conclusions::

1. The newly submitted data and arguments fail to convincingly mitigate the concerns which led to nonapproval of this agent.

2. An additional concern is herein raised: it remains unclear whether the dosing regimen of arbutamine used in pivotal study 127 was adequate to sufficiently produce the diagnosis-eliciting phenomena (such as those related to coronary hyperemia) which are required to detect the presence of clinically and angiographically significant lesions which do not impair perfusion at rest.

3. The sponsor's challenges in furthering their development program to the stage of approveability are to some extent dependent on their success in changing the investigator behavior which leads to post-test referral bias. The encouragement of a productive shift in this investigator dynamic is potentially amenable to proactive agency effort at publically presenting and inviting open discussion of its understanding of the implications of such biases, and our experience with the nonutility of the specificity surrogate known as the normalcy metric.

APPEARS THIS WAY
ON ORIGINAL

Steven M. Rodin, MD 3/11/96

Steven M. Rodin, MD Date
Medical Officer

cc: HFD-110/ division file, Buehler, Fenichel, Lipicky; ***no copy to Rodin**
cc: HFD-710/ Mahjoob, Hung, Chi.

APPEARS [illegible]
ON ORIGI[illegible]

tables only: PIVOTAL studies: Results Overview :

For the diagnosis of epicardial CAD, the utility of arbutamine, according to imaging modality, is shown in the table below.

Table A:

Overall results of arbutamine tests in pivotal patient studies
(studies 122, 123, and 127)

study	*sample sizes*	*% achieving "target HR"*	***EKG***		***Thallium-201 scintigraphy***	
			SENS	*SPEC*	*SENS*	*SPEC.*
#127	n= 10-124	3%	**51%**	**80%** (**44**-98%)	**87%**	**25%** (**3**-65%)
#122	n= 19-201	6%	**49%**	**74%** (**49**-91%)	--	--
#123	n= 11-127	8%	**49%**	**82%** (**48**-98%)	--	--

Mean sensitivities (SENS) and specificities (SPEC) for the diagnosis of CAD were determined from intent-to-treat datasets. The 95% confidence intervals are depicted in parentheses. Each study utilized a "closed loop" arbutamine administration.

PIVOTAL studies: Results Overview [continued]:

Echocardiographic ventricular dysynergy (study 123):

Table 28: B ;

Mean sensitivity and specificity, for the diagnosis of CAD,
of **Resting vs Stress-induced ventricular dysynergy** [study 123]

Metric	Resting (pre-stress) echocardiography	Arbutamine-stress echocardiography
Sensitivity	**40%** [n=127]	**76%** [n=127]
Specificity	**75%** (**48**-93%) [n=11]	**31%** (**11**-59%) [n=16]

[source: modification of tables in addenda dated 8/12/94, and 9/7/94]

Shown in parentheses are 95% confidence intervals. The analyses are based on an intent to treat dataset of patients with interpretable echocardiograms.

tables only- PIVOTAL studies: Results: [continued]

Arbutamine vs exercise:

Table 2: C:

Arbutamine vs relatively low-level exercise tests for the diagnosis of CAD in pivotal patient studies

study	*metric*	***EKG***		***Echocardiography***		***Thallium scintigraphy***	
		Arbutamine	*Exercise*	*Arbutamine*	*Exercise*	*Arbutamine*	*Exercise*
#122	Sensitivity	**49%**	**46%**	--	--	--	--
	Specificity	**74%** (**49**-91%)	**84%** (**60**-97%)	--	--	--	--
#123	Sensitivity	**49%**	**47%**	**76%**	**77%**	--	--
	Specificity	**82%** (**48**-98%)	**64%** (**31**-89%)	**31%** (**11**-59%)	**55%** (**23**-83%)	--	--
#127	Sensitivity	**51%**	**43%**	--	--	**87%**	**94%**
	Specificity	**80%** (**44**-98%)	**80%** (**28**-100%)	--	--	**25%** (**3**-65%)	**80%** (**28**-100%)

Shown are the mean sensitivities and specificities for the diagnosis of CAD (intent-to-treat). The 95% confidence intervals are depicted in parentheses.

APR 11 1995

Steven M. Rodin, M.D.
Medical Officer

Food and Drug Administration

Division of Cardio-Renal Drugs
Tel 301-443-0320; FAX-9283

Addendum #1 to Medical Review of NDA:

1 General information

NDA #:	**20-420**
Drug:	**arbutamine**
Sponsor:	Gensia Inc.
Proposed indication:	diagnostic adjunct
Addendum last revised:	11 April 1995

2 Erratum:

In my discussion of the limitations of study 128 (found on page 14 of my original review dated 10/26/94) footnote #6 should be disregarded as it offered an interpretation which is not relevant to thallium imaging. Clinical thallium methods do not have adequate spatial resolution to detect subcritical coronary stenoses via the identification of discrete subendocardial hypoperfusion.

Steven Rodin, MD
Steven M. Rodin, MD
Medical Officer

4/11/95
Date

cc: R.Fenichel/HFD-110; R.Lipicky/HFD-110; R.Temple/HFD-100; G.Buehler /HFD-110; HFD-110 division file (NDA 20-420); *no copy to S.Rodin