Reviewer's Comments on the ITT Efficacy Results:

Table 3 shows a great variability among the blinded reader' assessments of the percentage of cases in which post-dose images provided additional information over pre-dose, within and between studies. This percentage ranges from 20% to 99% in Study 42,440-3A and ranges from 43% to 72% in Study 42,440-3B. The observed variability in the response rates among the readers (investigators and blinded readers) was reflected on the limits of the confidence intervals as shown in Table 3. The lower limits of the confidence interval which presumably, under repeated sampling, contains the true response rate, range from about 11% to about 97% in Study 42,440-3A and from 32% to 62% in Study 42,440-3B.

Comparing the readers assessments across the two studies the results of Table 3 show that for the investigators as well as the blinded readers, except for Blinded Reader # 2, the percentages of additional information in Study 42,440-3B were higher than those in Study 42,440-3A. The differences range from 17 percentage points for Blinded Reader # 3, to 41 percentage points for Blinded Reader # 4. Blinded Readers # 2 assessments across the two studies, however, were in the opposite directions compared to those of other readers. The difference between the assessments in the two studies (43%-99%) was minus 56 percentage points.

Since Blinded Readers 1 and 2 viewed only static images, and the results of Blinded Reader 2 in Study 42,440-3A were higher than those of other readers but the opposite case occurred in Study 42,440-3B, one might consider the assessments of Blinded Readers 3 and 4 only. Even though this reduces the variability among the readers' assessments, it is difficult to interpret the difference in the response rates for these two readers in 42,440-3A, which were 55% and 20%, respectively (see Table 3).

The large differences in Blinded Readers 3 and 4 assessments might be interpreted to be a result of different readers assessing different aspects of the images. This is addition to assessment being based on personal judgement on the reader's side. However, for blinded readers, these factors may be expected to affect both the pre-dose and post-dose image assessment. Consequently, the blinded readers' assessments are, in this reviewer's opinion, useful measures in looking for additional information.

For efficacy evaluation, the hypotheses testing should be made against the anticipated response rate used in the design of the pivotal studies (i.e., 75%). Similarly, the lower limits of the 95% confidence intervals around the observed response rate should be judged against the claimed 75%. Post-hoc testings against 1% or 50%, when the observed response rates below 75%, does not provide evidence of efficacy. In Section IV.B.I we consider an alternative analysis for the primary endpoint.

The sponsor's efficacy results for the primary endpoint in the per-protocol population are similar to those of the ITT population shown in Table 3. Consequently the above comments also apply to the results of Table 4.

In addition to the per patient analyses of Tables 3 and 4, the sponsor summarized the per protocol efficacy results regarding additional information by pathology type (i.e., intra-luminal and extra-luminal), by the nature of the additional information, and by whether or not this information could have changed the patient's management/therapy. The sponsor's summaries are given in Attachment I (see Tables A1.1, A1.2, and A1.3, respectively). The sponsor's conclusion, as well as that of this reviewer, was that there was no trend in the response to the pathology type.

III.D.I.B. Evaluation of Reader's Agreement:

For assessing reader agreement, the sponsor calculated the Kappa statistic for the blinded readers' responses to the primary efficacy question. Patients with non-missing results, common to every pair of readers were analyzed. The sponsor's estimates of the Kappa statistics are presented in Table 5.

Table 5/ Sponsor's Results

Kappa Statistic Results of Blinded Readers Agreement for the Primary Efficacy Parameter

(Pivotal Studies, SonoRx Patients, Per-protocol Results)

Blinded Reader Pairs	Study 42,440-3A	Study 42.440-3B
Blinded Reader 1 vs Blinded Reader 2	0.017	0.323
Blinded Reader 1 vs Blinded Reader 3	0.059	0.253
Blinded Reader 1 vs Blinded Reader 4	0.013	0.220
Blinded Reader 2 vs Blinded Reader 3	NA [a]	0.130
Blinded Reader 2 vs Blinded Reader 4	NA [a]	0.141
Blinded Reader 3 vs Blinded Reader 4	-0.045	-0.009

Source sponsor Table AB p.84 vol 1.45 study A and Table AB p.81 vol 1.50 for Study B

[a] The Kappa correlation was not applicable as Blinded Reader 2 indicated yes for all patients analyzed.

The Kappa statistics, which measures the blinded readers' agreement after correcting for chance agreement, takes value 0 when the observed agreement is equal to the expected chance agreement, takes value 1 for complete agreement and takes negative values for disagreement. Based on this, the above estimates are not far from what is expected under chance agreement alone.

III.D.I.C. SonoRx Secondary Efficacy Results (Per-Protocol)

The sponsor presented results for per-protocol analyses for the following secondary efficacy parameters: Delineation of anatomy, Effect on investigator's confidence in diagnosis/ pathology: post-dose versus pre-dose; Readers's confidence: no other diagnosis /pathologic process, Post-dose images that helped rule out diagnoses /pathology identified pre-dose; Post-dose images that helped identify diagnoses/ pathology not identified pre-dose; and Impact of gas shadowing artifact. Finally, the sponsor estimated the sensitivity and specificity rates for the pathology diagnosis for the unenhanced ultrasound (pre-dose) and SonoRx (post-dose) in comparison to the comparative procedure conducted prior to the start of the study.

Among the various secondary endpoints the sponsor analyzed, two are closely related to the objectives of the study, the delineation of abdominal anatomy and the detection or exclusion of pathology (see Section II.A). Based on this and the fact that other secondary endpoints are subjective (such as the one measuring the degree of confidence) or that it did not show efficacy results, our discussion primarily will address the sponsor's results for the delineation of abdominal anatomy and the sensitivity and specificity estimates of the pathology diagnosis.

The sponsor's efficacy results shows highly significant p-values, almost across all readers and for the two pivotal studies, for the delineation of most of the following anatomical areas: stomach, gastric wall, pylorus, duodenum, pancreatic head, pancreatic body, pancreatic tail, and, to some extent for the pancreatic duct. The sponsor's efficacy results are presented Table A2.1 (Attachment II). The sponsor reported p-values were based on comparing each readers' pre-dose and post-dose image scores (excellent, good, fair, poor, none) for the delineation of specific abdominal anatomy by the Wilcoxon signed rank test.

Since pre-dose and post-dose image scores are available for every patient, a more appropriate analysis is to view the data as a matched pairs and test for improvement in the post-dose score over the pre-dose score. This analysis is more appropriate in this reviewer opinion than the sign rank test, which is expected to result in many ties for this type of data. We will pursue this alternative analyses in Section IV.B.II.

Concerning the sensitivity and specificity estimates for the pre-dose and post-dose images, the utility of these measure for testing efficacy depends on the availability of reasonably accurate 'gold standard' . For the present NDA. the sponsor used the results of different comparators to form a judgement about

pathology diagnosis. Since different modalities were used for different patients, it is difficult to judge the accuracy of the sensitivity and specificity estimates. In addition, when different modalities, with different expected accuracies, are used for one patient it is not clear how the sponsor decided about the pathology diagnosis. We will re-visit this issue in Section IV.A.IV, but for the present we will report the sponsor's results and comment on the them.

Table 6 presents the sponsor's sensitivity and specificity for pre-dose and post-dose images for pathology diagnosis in comparison to all clinical information available (excluding the ultrasound images).

Table 6/ sponsors Results

Sensitivity and Specificity Estimates of SonoRx patients in the Pivotal Controlled Studies

(Phase III : 42,440-3A and --3B) Per Protocol Population

Measure/ Reader	Study 42.440-3A		Study 42.440-3B	
	Pre-SonoRx	Post-SonoRx	Pre-SonoRx	Post-SonoRx
i) Sensitivity:				
Investigators	46.4% (32/ 69)	56.5% (39/ 69)	36.0% (27/ 75)	52.0% (39/ 75)
Blinded Reader 1	51.6% (33/64)	51.6% (33/64)	34.3% (24/70)	47.1% (33/70)
Blinded Reader 2	37.5% (24/64)	28.1% (18/64)	33.8% (23/68)	42.6% (29/68)
Blinded Reader 3	54.5% (30/55)	45.5% (25/55)	38.0% (27/71)	42.3% (30/71)
Blinded Reader 4	34.5% (19/55)	36.4% (20/55)	29.6% (21/71)	42.3% (30/71)
ii) Specificity:				
Investigators	70.0% (7/ 10)	60.0% (6/ 10)	69.2% (9/ 13)	69.2% (9/ 13)
Blinded Reader 1	33.3% (3/9)	22.2% (2/9)	70.0% (7/10)	60.0% (6/10)
Blinded Reader 2	55.6% (5/9)	44.4% (4/9)	70.0% (7/10)	70.0% (7/10)
Blinded Reader 3	55.6% (5/9)	100% (9/9)	50.0% (5/10)	50.0% (5/10)
Blinded Reader 4	55.6% (5/9)	55.6% (5/9)	60.0% (6/10)	60.0% (6/10)

Source: For Study 42.440-3A Sponsor 's Tables 20.1 , 20.2 A.8.1, A.8.2, B.8.1, B.8.2, C.8.1, C.8.2 Vol 1.45- 1.46 and for Study 42,440-3B Sponsor 's Tables 20.1, 20.2 A.8.1, A.8.2, B.8.1, B.8.2, C.8.1, C.8.2 Vol 1.51-1.52.

Reviewer 's Comments on the Sensitivity and Specificity Estimates:

Whereas the investigator's assessments showed increase in the sensitivity estimates in post-dose versus pre-dose for both studies (increase from 46% to 57% in Study 42,440-3A and from 36% to 52% in Study 42,440-3B), the sensitivity estimates based on the blinded readers' assessments were not

consistent across the two studies. Table 6 shows that whereas the post-dose sensitivity estimates were higher than their pre-dose analogues in Study 42,440-3B they were in the opposite direction in Study 42,440-3A. Specificity estimates being based on a small numbers of patients are unreliable for making judgment. In addition, these estimates are not consistent across the two studies.

Overall, having most of the sensitivity estimates below 50%, whether one considers the pre-dose or post-dose images, indicates non-compatibility of the diagnosis based on the ultrasound images with that of the other comparative diagnostic modalities. Sensitivity estimates of 50% indicates that the agreement of the test under consideration with the 'gold standard' in the diagnosis of patients who have the disease occurs at random and is not an indication of the efficacy of the test. However, since the sensitivity estimates of the pre-dose and post-dose images were based on a not well-defined 'gold standard', the utility of using these estimates in making a judgement about efficacy is dubious in this reviewer's opinion.

III.D.I.D. Control Agent Primary Efficacy Results: Additional Post-dose Information Provided Over Pre-dose

The patients' disposition data for the control agent in the two pivotal studies (42,440-3A and 42,440-3B) were given in Table 2. Since, as in SonoRx patients, only images of patients who met the per-protocol analysis criteria were sent to Blinded Readers 3 and 4, the sponsor considered analyses for this population only. Table 7 presents the sponsor's per-protocol efficacy results on whether the post-dose images provided additional information over the pre-dose images, for investigators and the four blinded readers.

Reviewer 's Comments on the Efficacy Results of the Control Agent:

Comparison of the efficacy results of the control agent in Table 7 with their analogues for SonoRx patients in Table 4, shows that only for Blinded Reader 3's assessment in Study 42,440-3B SonoRx provided more additional information than the control agent. The difference in the response rates for this reader (75%-57%) was 18 percentage points in favor of the SonoRx. But all remaining comparisons were in favor of the control agent with difference ranging from 1 percentage point (100% - 99% for Blinded Reader 2, Study 42,440-3A) to 22 percentage points (68% - 44% for Blinded Reader 1 in the same study). Statistical testing for comparison of the SonoRx and control agent response rates is given in Section IV.B.III.

Table 7/ Sponsor's Results

Additional Information Provided Post-dose Over Pre-dose

(Pivotal Studies, Control Agent, Per-protocol Population)

Reader	Study 42.440-3A		Study 42.440-3B	
	n/N[a] (%)	(95% CI)[b]	n/N (%)	(95% CI)
Investigators	14 / 21 (67)	(46.5, 86.8)	14 / 25 (56)	(36.5. 75.5)
Blinded Reader 1	13 / 19 (68)	(47.5, 89.3)	19 / 22 (86)	(72.0. 100)
Blinded Reader 2	19 / 19 (100)	(100, 100)	11 / 22 (50)	(29.1. 70.9)
Blinded Reader 3	12 / 17 (71)	(48.9, 92.2)	12 / 21 (57)	(36.0. 78.3)
Blinded Reader 4	6/ 17 (35)	(12.6, 58.0)	17 / 21 (81)	(64.2. 97.7)

Source: Sponsor Table AI , p. 168, Vol. 1.37

[a] n = number of patients with additional information provided post-dose over pre-dose, and N=number of per- protocol patients;

[b] 95% confidence interval assuming the normal approximation

III.D.II. Supportive Study (Phase III: 42,440-7)

III.D.II.A. Primary Efficacy Results:

The primary efficacy question, as discussed in Section II.A., was whether post-SonoRx images provided more information than the post-water images. The sponsor stated that no evaluation was made for pre-dose images which were collected for the purpose of patient's management.

As in the pivotal studies, the blinded readers who had no knowledge of the study agent administered or patients' clinical information, , according to the sponsor, completed the Case Report Forms, which included an Individual Image Evaluation section (SonoRx and Water images viewed separately) and a Comparison Image Evaluation section (SonoRx and water images viewed side-by-side). Similarly Blinded Readers 1 and 2 evaluated only acceptable static images and Blinded Readers 3 and 4 reviewed both static and video images for the per-protocol population.

The sponsor noted that SonoRx and/or water images from 5 patients (3 ingested <350 mL, 2 did not receive both agents) were inadvertently not provided to Blinded Readers 3 and 4. Consequently, as in the pivotal studies, the sponsor performed an analysis for the primary endpoint, for the ITT population, by assuming water provided more information than SonoRx for the 3 patients, but no analysis was done, for the ITT population, on the secondary efficacy data collected from Blinded Readers 3 and 4. Only analyses for the per-protocol population were carried out for the secondary efficacy parameters.

The primary efficacy parameter analysed in this study was, overall, which images provided more diagnostic information (SonōRx, water, or both are 'equal'). The sponsor's analyses for this primary endpoint were: (i) using the binomial test to test the hypothesis the proportion of patients whose SonoRx images provided more diagnostic information over water against the null hypothesis (equal chance [50%]), (ii) carrying out the test in (i) after splitting the equal responses, in an even fashion, between the SonoRx and water responses. The sponsor called this post-hoc test " equal split test" and (iii) comparing, in a post-hoc manner, the SonoRx and the water responses using the sign test, with responses of equal were considered ties (neither in favour of SonoRx or water) and omitted from the analysis.

The sponsor's primary efficacy results for the intent-to-treat and for the per-protocol populations, are given in Table 8, and 9, respectively. Since the results of the two analyses are similar, this reviewer's comments will address the two tables jointly.

Table 8/ Sponsor's Results

Images Providing more Diagnostic Information: SonoRx, Water, or Equal

(Supportive Study # 42,440-7; ITT population)

	Investigators n / N[a] (%)	Blinded Reader 1 n / N[a] (%)	Blinded Reader 2 n / N[a] (%)	Blinded Reader 3 n / N[a] (%)	Blinded Reader 4 n / N[a] (%)
SonoRx	33/ 53 (62)	25 / 50 (50)	19 / 50 (38)	24 / 47 (51)	18 / 47 (38)
Water	12 / 53 (23)	17/ 50 (34)	19 / 50 (38)	10 / 47 (21)	9 / 47 (19)
Equal	6 / 53 (11)	8/ 50 (16)	12 / 50 (24)	13 / 47 (28)	20 / 47 (43)
Not Done[b]	2/53 (4)				
SonoRx vs 50% test[c]	p=0.098	p=1.000	p=0.119 *	p=1.000	p=0.144 *
Equal split test[d]	p=0.013	p=0.322	p=1.000	p=0.079	p=0.243
Sign test	p=0.008	p=0.280	p=1.000	p=0.024	p=0.122

Source: Sponsor's Table R p. 66, and Table AB, p.83, Vol . 1.55, with some modifications

[a] n = Number of patients with more diagnostic information for given response (SonoRx or water) or number of patients with same amount of information for response of equal, and N = number of of per-protocol patients

b 2 patients who did not receive SonoRx and whose water images were inadvertently not sent to the readers were categorized as 'not done' for the investigator's readings, these were classified as water providing more information in the analysis

c Binomial test for the SonoRx proportion of success (providing more information).

d The equal category was split between SonoRx and water as follows: Investigators: 3 SonoRx and 3 water; Blinded Reader 1: 3 SonoRx and 4 water; Blinded Reader 2: 6 SonoRx and 6 water; Blinded Reader 3: 6 SonoRx and 7 water; Blinded Reader 4: 10 SonoRx and 10 water.

* Wrong direction. the p-value is for testing wether the observed rate is significantly less than 50%.

Table 9/ Sponsor's Results

Images Providing more Diagnostic Information: SonoRx, Water, or Equal

(Supportive Study # 42,440-7; Per-protocol Population)

	Investigators n / Na (%)	Blinded Reader 1 n / Na (%)	Blinded Reader 2 n / Na (%)	Blinded Reader 3 n / Na (%)	Blinded Reader 4 n / Na (%)
SonoRx	31/ 48 (65)	23 / 43 (53)	18 / 47 (38)	24 / 44 (55)	18 / 44 (41)
Water	12 / 48 (25)	13 / 43 (30)	17 / 47 (36)	7 / 44 (16)	6 / 44 (14)
Equal	5 / 48 (10)	7 / 43 ((16)	12 / 47(26)	13 / 44 (30)	20 / 44 (46)
SonoRx [a] vs 50% test [b]	p=0.059	p=0.761	p=0.144 *	p=0.652	p=0.291 *
Equal split test [c]	p=0.013	p=0.222	p=1.000	p=0.023	p=0.096
Sign test	p=0.005	p=0.133	p=1.000	p=0.003	p=0.023

Source: Sponsor's Table Q, p. 65 and Table Z. p.81 Vol. 1.55.

[a] n = Number of patients with more diagnostic information for given response (SonoRx or water) or number of patients with same amount of information for response of equal, and N = number of of per-protocol patients.

[b] Binomial test for the SonoRx proportion of success (providing more information).

[c] The equal category was split between SonoRx and water as follows: Investigators: 2 SonoRx and 3 water; Blinded Reader 1: 3 SonoRx and 4 water; Blinded Reader 2: 6 SonoRx and 6 water: Blinded Reader 3: 6 SonoRx and 7 water; Blinded Reader 4: 10 SonoRx and 10 water.

* Wrong direction, the p-value is for testing wether the observed rate is significantly less than 50%.

Reviewer's Comments on the Efficacy results of the Supportive Study:

There are several points one needs to consider in interpreting the efficacy results in Tables 8 and 9. Among these are the following:

(i) Unlike the response classification in the pivotal studies which was in two categories (provided additional information or not) the response in the supportive study 42,440-7 was classified into three categories (SonoRx image provided more, equal, or less diagnostic information than water). In addition, the comparison in the pivotal studies was post-dose versus pre-dose image, but in the supportive study was post-dose SonoRx versus post-dose water. Furthermore, having a cross-over design for the supportive study suggests the need to examine the results by period for consistency, that is, to look for response-time interaction. This was not the case in the pivotal studies which had a baseline control design.

(ii) Since the 'equal response' category accounts for 30% and 46% of Blinded Readers 3 and 4 responses, respectively, the comparison's results are very sensitive to the way of handling this category. By splitting the response in this category between the other two categories, one in effect is increasing the percentage of success, and consequently, reducing the variance of its estimates and thus leading to more significant results. With sponsor's claim of superiority of SonoRx enhanced images over water image, splitting the 'equal' category is unjustified.

(iii) Use of the sign test in not appropriate for data which have many counts in the 'equal' category as in present sponsor's data. By omitting the response in the 'equal' category when conducting the sign test, one in effect is exaggerating the difference between the two treatments. This might explain the sponsor's reported smaller p-values for the sign test in comparison to those of the binomial tests.

(iv) Concerning interpretation of the results, the reported p-values, in Tables 8 and 9, for the binomial test against 50% are not significant, even though the results show a trend in favor of SonoRx. The difference in the proportion of cases where SonoRx enhanced images provided more information than water image ranged from 0, for Blinded Reader 2, to 39 percentage points, for the investigators in the analysis of the ITT population. Analysis of the per-protocol population showed similar results. Since the observed response rates are lower than the anticipated response rate used in the sample size calculation (70%), the sponsor tested against 50% instead of against 70%. Aside from inappropriate -ness of the sign test as discussed in (iii), the results of this post-hoc test show highly significant p-values for the investigators (p=0.008 and 0.005 for analysis of the ITT and the per-protocol population, respectively). The other p-values for blinded readers, adjusting for multiplicity, would not reach significance, except for Blinded Readers # 3 for the analysis of the per-protocol population.

III.D.II.B. Evaluation of Reader's Agreement:

The sponsor's estimates of the Kappa statistic, for evaluating the agreement of readers' assessments were as follows: for Blinded Reader 1 vs Blinded Readers 2, 3, and 4 were : 0.066, 0.190 and 0.051, respectively; for Blinded Reader 2 vs Blinded Readers 3 and 4 were : 0.25 and 0.029, respectively; and for Blinded Reader 3 vs Blinded Readers 4 was 0.048. These results show some agreement between Blinded Reader 3 assessment and those of Blinded Readers 1 and 2.

III.D.II.C. Secondary Efficacy Results

The sponsor presented results of several secondary endpoints, as those in the pivotal studies. The results for comparing the investigators and blinded readers scores for the SonoRx and water images for delineation of specific abdominal anatomy, by using the Wilcoxon signed rank test, were not consistent across the readers nor across the anatomical areas considered. For the majority of the comparisons the results were not significant. Similarly, for the assessment of the impact of gas shadowing artifact on specified abdominal anatomy for the SonoRx and water images. In addition, in judging for efficacy one might consider adjustments for multiplicity. However, since these are secondary endpoint we will not pursue the multiplicity issue here.

Since pathology diagnosis is one of the objectives of the study we present, as in the pivotal studies, in Table 10 the sponsor's sensitivity and specificity estimates for the SonoRx and water enhanced images relative to the diagnosis of other modalities (see Section II.A for partial listing).

Reviewer's Table 10/ Sponsor's results

Sensitivity and Specificity Estimates of SonoRx and Water patients, Study (42,440-7) , Per- protocol Analysis

Reader/ Measure	Sensitivity		Specificity	
	SonoRx	Water	SonoRx	Water
Investigators	72.7% (32/44)	100% (32/44)	100% (4/4)	100% (4/4)
Blinded Reader 1	56.4% (22/39)	43.5% (17/39)	100% (4/4)	75% (3/4)
Blinded Reader 2	37.2% (16/43)	30.2% (13/43)	100% (4/4)	100% (4/4)
Blinded Reader 3	65% (26/40)	67.5% (27/40)	75% (3/4)	50% (2/4)
Blinded Reader 4	27.5% (11/40)	32.5% (13/40)	100% (4/4)	100% (4/4)

Source: Sponsor's Tables 20.1. 20.2. A.9.1, A.9.2, B.9.1, B.9.2. C.9.1, C.9.2. D.9.1. D.9.2, Vol. 1.55

The results of Table 10 shows that the sensitivity of the SonoRx water images are comparable. In fact the sensitivity of the water images for the investigators readings is higher than that of SonoRx enhanced images. The specificity estimates are unreliable being based on a small number of patients.

IV. Reviewer's Overall Comments and Analyses:

Section IV.A. summarizes this reviewer's comments concerning: the number of patients analyzed, post-hoc analyses versus planned analyses, blinded reader' assessments (1 and 2 vs 3 and 4), the sponsor's comparative procedures used for pathology diagnosis, as well as the reviewer's other concerns. Section VI.B. presents results of analyses which address some of the concerns expressed in Section IV.A.

IV.A. Reviewer's Overall Comments:

IV.A.I. Number of Patients Analyzed:

As previously discussed (Section III.B) the number of patients analyzed shows great variability, in particular for the first pivotal study (# 42,440-3A). The ITT population in this study ranges from 93 patients for the investigators to 76 patients for Blinded Readers 3 and 4. Similarly, the per-protocol

population in this study ranges from 79 patients for the investigators to 64 patients for Blinded Readers 3 and 4. This variability can be attributed to the interpretation of the statement 'images of acceptable technical quality' in the definition of the ITT population (see Section III.B).

Since the purpose of the analysis of the ITT population is, in general, to include all patients who are 'randomized' to the trial, blinded readers should have been given all images even those of unacceptable quality by the investigators. The sponsor's interpretation requires not only the image be of acceptable quality to the investigators, but also to be of acceptable quality to the blinded reader, even if the quality of the images was deemed acceptable by the investigators as well as a technical reviewer before it was presented to the blinded reader.

IV.A.II. Planned vs. Post-hoc Efficacy Analyses:

The sponsor determined the sample size in the pivotal studies, as discussed in Section III.A, by assuming that the SonoRx post-dose images should provide additional information over the pre-dose images in at least 75% of the cases. However, having the observed rate based on the readers' assessments fall below 75%, the sponsor in a post-hoc analysis tested the observed rates against 1% (see page 15 for related discussion). Similarly, calculation of the sample size in the supportive study (# 42,440-7), was based on the assumption that SonoRx images would provide additional information over water images in at least 70% of the cases. But in testing for efficacy the sponsor used 50%. Furthermore, as the results of testing against the 50% were not significant, the sponsor used unplanned analyses, such as sign test or splitting the cases in which the SonoRx and water images gave similar information (see Tables 8 and 9)

Due to the post-hoc nature of the analysis, should one view the reported significant p-values for testing against 1%, or in some cases against 50%, as evidence for efficacy ? For efficacy one needs to show that post-dose images outperform (provides additional information) the pre-dose images for at least certain percentage of the patients. The magnitude of this percentage depends on clinical practice (availability of other agents , etc.). Based on this, the statistical analysis should consist of:

(i) testing whether post-dose image outperform (provide more information) the pre-dose image, and (ii) if the results were in favor of the post-dose image, estimating the magnitude of gain of the provided by post-dose (SonoRx enhanced) over that of the pre-dose, preferably with confidence intervals around the point estimates.

Since there is a possibility that the post-dose image provides similar information to that of the pre-dose image, the sponsor's analysis does not address the efficacy testing as outlined above. In fact, for an observed response rate below 50%, a significant p-value does not imply efficacy for the contrast agent, but might implies the opposite. To explain this, let us assume that the observed response rate for the contrast agent was 45% with highly significant p-values for testing, say, against 40%. Now whether the significant p-value implies efficacy of the contrast agent (i.e., provided additional information over the pre-dose) depends on how the remaining 55% was distributed between the two categories: 'equal response' or 'pre-dose provides something which the post-dose does not'. Clearly, if the equal category accounts for only 1% and pre-dose image provided additional information over the post-dose image in 54% of the cases, then the contrast agent failed to show efficacy.

The sponsor's analysis, unfortunately, does not address efficacy in the way summarized above. Furthermore, it is not clear how one interprets a negative response to the question which its response was used for efficacy analysis (question 13 of the Case Report Form). Specifically, if the post-dose image did not provide additional information it is not clear whether this implies that the two images provided similar information or pre-dose image provides more information the post-dose image. In a response to this reviewer's inquiry about this point, the sponsor's reply of January 17, 1997, suggests to take the response to Question 13 in conjunction with that for question 14, which asks whether pre-dose image provide information that the post-dose image did not. In Section V.B we consider the responses to both question in carrying the proposed statistical analysis.

IV.A.III. Blinded Readers Assessments (Blinded Readers 1 and 2 vs Blinded Readers 3 and 4):

The sponsor provided Blinded Readers 1 and 2 with static images only even though the plan was to provide both static and video images. After finding the results of these readers' evaluations were inconsistent with those of Phase II studies (used in planning the pivotal studies) the sponsor decided to have two additional blinded readers (Blinded Readers 3 and 4) for each study and to provide them with both static and video images.

Aside from adherence to the study protocol, the inclusion of the additional blinded readers raises issue concerning the validity of the statistical inference. In theory, inference based on these readers assessments should be viewed as a conditional inference since these assessments occurred after other assessment were made. This reviewer, however, is unaware of any statistical methodology which could adjust the p-value for penalty. Aside from the statistical concerns about the validity of the inference, the inclusion of two additional blinded readers suggests, in light of the wide variability in the four

readers's assessments, to focus the analysis on the assessments of a subset of the four readers. If this seems reasonable, should one consider Blinded readers 1 and 2 or Blinded Readers 3 and 4 ? Any of these two choices has its pros and cons. Blinded Readers 1 and 2 were to assess the images but they were not given all available data, on the other hand, Blinded readers 3 and 4 underwent training and given both types of images. Following the discussion of Section II.D.I.A this reviewer, regardless of the statistical issue raised, is inclined to give more weight to the assessments of Blinded readers 3 and 4 even though this reviewer's analysis of the primary endpoint in Section VI.B will be given for all blinded readers and investigators.

IV.A.IV. Sponsor's Comparative Procedures Used for Pathology Diagnosis:

The sponsor's estimates of the sensitivity and specificity, for the blinded readers, for the unenhanced (pre-dose) and enhanced (SonoRx) were based on the sponsor's judgement concerning the pathology diagnosis based on the results of the comparative procedures conducted prior to the start of the study. There are several issues raised by the Medical Officer, R. Yeas, M.D., in a meeting on 5/20/97, concerning the way in which these procedures were used to form a 'standard' for pathology diagnosis (see Section III.D.II.B for related discussion).

IV.A.V. Reviewer's Other Concerns:

In addition to the issues raised in the previous sections, there are other points which should be addressed. Among these are:

i) Consistency of the efficacy results across the study centers
ii) Robustness of the efficacy results with respect to the change in the number of patients whose images were not readable or missing
iii) As the supportive study (Study # 42,440-7) should play a role in determining efficacy, one needs to address the independence of this study from the pivotal studies.

Concerning consistency of the efficacy results across each study' centers, in light of the large variations among the readers' assessments for the study as whole, it is expected that the variability among the study centers to be even greater. The sponsor's results for testing homogeneity by using the Chi-square test were significant (the p-value for Study 42,440-3A was 0.003). However, one can not put much weight on the significance of the p-values since many centers have small numbers of patients. In Study 42,440-3A, for example, the number of patients in these centers ranges from 2 patients (Center # 9414) to 16 patients (Center # 9411). Clearly, one might combine centers with small number of

patients provided that these centers have results in the same direction. However, the comparison is complicated due to having 5 different assessments (readers) for each patient's image with wide variability in their assessments. Based on these points, we will not pursue this issue further, instead one might focus on the variability among the reader's assessment for each study as whole.

Concerning robustness of the efficacy results to number of patients analyzed, the sponsor excluded 9 patients from Study 42,440-3A whose images were not readable or missing (as indicated, the sponsor did not specify the number of patients in each category). Clearly, the way in which the missing values are treated will have impact on the efficacy results of this study.

Regarding independence of the supportive study from the pivotal studies, Table 11 presents the list of participants involved in more than one study.

Table 11 / Reviewer's Table

List of participants in more than one study, classified by their role in each study

	Study 42.440-3A	Study 42.440-3B	Study 42.440-7
Blinded readers	L. Tannenbaum. M.D.(2)*	R. Barr, M.D. (1)*	D. Rubin. M.D. (1)* L. Tannenbaum, M.D. (2)*
Investigators	E. Bluth. M.D. (2)* A. Lev-Toaff. M.D. (8)*	D. Rubin. M.D.(2)* P. Lund. M.D. (4)*	R. Barr, M.D. (1)* E. Bluth. M.D. (2)* A. Lev-Toaff. M.D. (3)*
Technical Reviewer	B. Goldberg, M.D. (1)*	P. Lund. M.D.(1)* B. Goldberg. M.D.(2)*	B. Goldberg, M.D. (1)*

* Numbers in parentheses indicate the investigator's or Blinded Reader's order in the study

Source of data: Sponsor's listing: pp.1-2, vol.1.45, pp.1-2, vol.1.50. pp.1-2, vol.1.55.

It can be seen from Table 11 that Blinded Readers 1 & 2 in the Supportive Study 42,440-7 were, respectively, an investigator and a blinded reader in the two pivotal studies. Also, 3 of the investigators in the Supportive study were investigators in the pivotal studies. Also the three studies have the same technical reviewer.

IV.B. Reviewer's Analyses:

Subsections IV.A.I and IV.A.II presents results of re-analyses for the primary and secondary endpoints in the pivotal studies and Subsection IV.A.III compares the SonoRx primary endpoint with that of the control agent in the pivotal studies as well as in the supportive study.

IV.B.I. Primary Endpoint: Percentage of Patients with Additional Information:

Following the discussion in Section IV.A.I, our statistical analysis consists of the following two steps:

(i) Testing whether the percentage for which SonoRx enhanced images provides additional information over the pre-dose image is significantly different from the percentage for which pre-dose image provides more information. Specifically, let Ps be the percentage of patients for whom post-dose image provided additional information over the pre-dose image, and let Pr be the percentage of patients for whom pre-dose image provided additional information over the post-dose image, then we consider testing the hypothesis:

$$H_0 : Ps = Pr \quad \text{vs.} \quad H_1 : Ps \neq Pr$$

and (ii) if the null hypothesis is rejected we provide point estimate and 95% confidence intervals for the difference Ps - Pr. The lower limit of the 95% confidence intervals gives the minimum gain, in terms of additional information, one expects to have as a result of using the contrast agent over that of the of the pre-dose image.

We carry out the above analysis for the ITT as well as the per-protocol populations. Since the McNemar's test we are using to test the above hypothesis is not sensitive to changes in the number of observations on the diagonal of the 2x2 table, i.e., for which the pre-dose image provides 'equal' information to post-dose image, we will consider for the ITT analysis the same number of patients used by the sponsor. For patients with missing data (mainly because the ingested less than 350 mL of SonoRx, 12 in Study 42,440-3A and 4 in Study 42,440-3B) we assign the worst scores. That is, we assume pre-dose image provided more information than the post dose-image. Since for this analysis there are cells with small frequencies we report p-values calculated by exact method, using the statistical software StatXact.

Table 12 presents the results of the ITT analysis for primary endpoint in the pivotal studies and Table 13 presents the corresponding results for per-protocol analysis.

Table 12/ Reviewer's Table
Comparison of Additional Information Provided For Post-dose Over Pre-dose and Pre-dose Over Post-dose (Pivotal Studies, Control Agent, ITT Population *)

	Study 42.440-3A				Study 42.440-3B			
Reader	Ns/N (Ps) [1]	Nr/N (Pr) [2]	p-value Ps = Pr [3]	d = Ps - Pr (95% C.I.)	Ns/N (Ps)	Nr/N (Pr)	p-value Ps = Pr [3]	d = Ps - Pr (95% C.I.)
Investigators	54/ 93 (58)	6/ 93 (6)	< 0.0001	52 (41. 62)	67/ 94 (71)	2/ 94 (2)	< 0.0001	69 (59. 78)
Blinded Reader 1	30/ 85 (35)	5/ 85 (6)	< 0.0001	29 (20. 40)	59/ 84 (70)	0/ 84 (0)	< 0.0001	70 (59. 80)
Blinded Reader 2	72/ 85 (85)	0/ 85 (0)	< 0.0001	85 (75. 92)	36/ 84 (43)	8/ 84 (10)	< 0.0001	33 (23, 44)
Blinded Reader 3	41/ 76 (54)	15/ 76 (20)	0.0007	34 (24. 46)	61/ 85 (72)	6/ 85 (7)	< 0.0001	65 (54, 75)
Blinded Reader 4	15/ 76 (20)	12/ 76 (16)	0.7011	4 (01. 11)	52/ 85 (61)	4/ 85 (5)	< 0.0001	56 (45. 67)

* patients with missing data (16 patients) were assigned the worst score, i.e. pre-dose provide 'more information' than post -dose.

[1] Ns is the number of patients for whom post-dose image provide additional information over pre-dose image and N is the total number of patients analyzed. Ps is the corresponding percentage.

[2] Nr is the number of patients for whom pre-dose image provide additional information over post-dose image and N is the total number of patients analyzed, Pr is the corresponding percentage

[3] Exact p-value (McNemar test) calculated by using StatXact.

Table 13/ Reviewer's Analysis
Comparison of Additional Information Provided For Post-dose Over Pre-dose and Pre-dose Over Post-dose (Pivotal Studies, Control Agent, Per-Protocol Population)

	Study 42,440-3A				Study 42,440-3B			
Reader	Ns/N (Ps) [1]	Nr/N (Pr) [2]	p-value Ps = Pr [3]	d = Ps - Pr (95% C.I.)	Ns/N (Ps)	Nr/N (Pr)	p-value Ps = Pr [3]	d = Ps - Pr (95% C.I.)
Investigators	46/ 79 (58)	5/ 79 (6)	< 0.0001	52 (40, 63)	62/ 88 (70)	1/ 88 (1)	< 0.0001	69 (59, 79)
Blinded Reader 1	29/ 73 (40)	5/ 73 (7)	< 0.0001	33 (22, 45)	55/ 80 (69)	0/ 88 (0)	< 0.0001	65 (54, 75)
Blinded Reader 2	63/ 73 (86)	0/ 73 (0)	< 0.0001	86 (76, 93)	32/ 78 (41)	8/ 78 (10)	0.0002	31 (21, 42)
Blinded Reader 3	41/ 63 (65)	2/ 63 (3)	<.0007	62 (49, 74)	61/ 81 (75)	2/ 81 (2)	< 0.0001	73 (62. 82)
Blinded Reader 4	15/ 64 (23)	0/ 64 (0)	0.0001	23 (14. 36)	52/ 81 (64)	0/ 81 (0)	< 0.0001	64 (53. 75)

[1] Ns is the number of patients for whom post-dose image provide additional information over pre-dose image and N is the total number of patients analyzed. Ps is the corresponding percentage.

[2] Nr is the number of patients for whom pre-dose image provide additional information over post-dose image and N is the total number of patients analyzed. Pr is the corresponding percentage.

[3] Exact p-value (McNemar test) calculated by using StatXact.

It can be seen from the results of Tables 12 and 13 that, whether one considers the ITT or the per-protocol population, the post dose image provides, significantly additional information over the pre-dose, for almost all readers (except Blinded Reader 4 in the 42,440-3A). For the ITT analysis, in which worst case considered for the missing data, the magnitude of gain as result of using SonoRx ranges from 4 percentage points for Blinded Reader 4 in Study 42,440-3A to 65 percentage points for Blinded Readers 3 in Study 42,440-3B. For the per-protocol analysis the corresponding gain ranges from 23 percentage points to 73 percentage points for the same two readers in the two studies.

IV.B.II. Secondary Endpoint: Delineation of Abdominal Anatomy

The sponsor's analyses for the secondary endpoints show highly significant results for the delineation of most the 8 anatomical areas shown in Table A2.1(Attachment II). But the Case Report Form shows that the evaluation was planned for 17 anatomical areas. The Medical Officers, Yaes, R., M.D. and Jones, E., M.D., recommended that one might consider, in addition to those areas analyzed by the sponsor, the superior mesenteric artery and para-aortic lymph nodes. The justification for this was that since the time between taking the contrast agent and imaging is about 10 minutes, it is unlikely that the agent has reached to other areas. Based on this, we will re-analyze the sponsor's data on delineation for the following 10 anatomical areas: stomach, gastric wall, pylorus, duodenum, pancreatic head, pancreatic body, pancreatic tail, pancreatic duct, superior mesenteric artery and para-aortic lymph nodes.

Following this reviewer's comments on the sponsor's results for this endpoint (see Section III.D.II.C) we consider, as in the analysis of the primary endpoint, analysis of changes in the post-dose score from that of the pre-dose score. Our analysis addresses the question of whether there is an improvement in the categorical score (none, poor, fair, good, excellent) as results of taking the contrast agent. The approach we consider can be viewed as a generalization of the McNemar test for matched pairs. The same test can be used to test for marginally homogeneity (see: User Manual, StatXact -Turbo, 1992, pp. 118-119) . To describe our test, let us define P_{ij} to be the probability that a patient have pre-dose score (i) and post-dose score (j) for i, j=1,2,...5. Then our analysis consists of the following two steps:
(i) testing the hypothesis:

$$H_0 : P_{ij} = P_{ji} \quad \text{vs.} \quad H_1 : P_{ij} \geq P_{ji} \quad \text{for} \quad i < j$$
$$\text{or } H_1' : P_{ij} \leq P_{ji} \quad \text{for} \quad i < j$$

with at least strict inequality for at least one (i,j) and,
(ii) we desire, when the null hypothesis is rejected, to have point estimate and 95% confidence intervals for the magnitude of improvements in the score (post-dose over pre-dose) . However, since the

magnitude of improvement from pre-dose score (i) to post-dose score (j) could takes values in the ranges (-4, 4), which is not easy to interpret for this subjective measure, we consider the following simplified score for improvement:

$$\text{Improvement Score} = \begin{cases} 1 \text{ if post-dose score is greater than pre-dose score,} \\ 0 \text{ if post-dose score = pre-dose} \\ -1 \text{ if post dose score is less than pre-dose score} \end{cases}$$

Then the magnitude of gain can be evaluated from the non-zero scores. As in the analysis of the primary endpoint, we provide a point estimate for the magnitude of gain in score as well as 95% confidence intervals for each of the 10 anatomical areas considered in the analysis. However, since the number of these comparisons is large, we present results for the ITT analyses for Blinded Readers 3 and 4 only. Selecting of these readers assessment is based on the discussion in Section IV.A. III. Here for patients with missing data we assumed the neutral case (i.e., the pre-dose score is similar to that of the post-dose). Since some cells have small frequencies we consider, as for the primary endpoint, exact p-values, calculated by StatXact. Tables 14 and 15 present , respectively, the results of the analyses for Study 42,440-3A and Study 42,440-3B.

It can be seen from Tables 14 and 15 that, for both pivotal studies and for both blinded readers, the post-dose image score is significantly higher than their corresponding pre-dose image score for the following anatomical areas: stomach, gastric wall, pylorus and duodenum. These p-values remain significant even if one adjusts for multiplicity. The magnitude of improvements for these areas range from 26 percentage points to 61 percentage points in Study 42,440-3A and from 50 percentage points to 75 percentage points in Study 42,440-3B . For the pancreatic areas (head, body tail and duct) the results are not consistent across the two studies and across the two readers. Study 42,440-3B shows significant results for the pancreatic areas but only to a less extent in Study 42,440-3A. For the superior mesenteric artery and para-aortic lymph nodes only the evaluation of Blinded Readers #3 in Study 42,440-3B shows significant results.

Table 14/ Reviewer's Analysis

Analysis of Pre-dose and Post-dose image Scores for Certain Anatomical Areas

(Secondary Endpoint, Pivotal Study 42,440-3A, ITT Population *)

	Blinded Reader # 3				Blinded Reader # 4			
Anatomical Area	Ns/N (Ps) [1]	Nr/N (Pr) [2]	p-value Ps = Pr [3]	d = Ps - Pr [4] (95% C.I.)	Ns/N (Ps)[1]	Nr/N (Pr) [2]	p-value Ps = Pr [3]	d = Ps - Pr [4] (95% C.I.)
Stomach	47/ 76	1/ 76	<0.0001	0.61 (0.49. 0.72)	41 / 76	6/ 76	<0.0001	0.46 (0.35. 0.58)
Gastric Wall	46/ 76	1 / 76	<0.0001	0.59 (0.57, 0.70)	37/ 76	6/ 76	<0.0101	0.41 (0.30. 0.53)
Pylorus	44/ 76	1 / 76	<0.0001	0.57 (0.45, 0.68)	30/ 76	10/ 76	<0.0001	0.26 (0.17, 0.38)
Duodenum	43/ 76	0 / 76	<0.0001	0.57 (0.45. 0.68)	28/ 76	8/ 76	0.0003	0.26 (0.17. 0.38)
Pancreatic Head	11/ 76	2 / 76	0.0066	0.12 (0.06, 0.21)	25/ 76	10/ 76	0.0106	0.20 (0.11, 0.30)
Pancreatic Body	8/ 76	1/ 76	0.0273	0.09 (0.04. 0.18)	16/ 76	9/ 76	0.0883	0.09 (0.04. 0.18)
Pancreatic Tail	28/ 76	2/ 76	<0.0001	0.34 (0.24, 0.46	19/ 76	18/ 76	0.7016	0.01 (.00. 0.07)
Pancreatic Duct	3/ 76	0/ 76	0.2500	0.04 (0.01, 0.11)	21/ 76	13/ 76	0.9128	0.11 (0.05. 0.20)
S. Mes. Artery	3/ 76	1/ 76	0.5000	0.03 (0.00, 0.09)	18/ 76	15/ 76	0.9153	0.04 (0.01, 0.11)
Para- Lymph Nodes	2/ 76	0/ 76	0.5000	0.03 (0.00, 0.09)	1/ 76	2/ 76	1.000	-0.01 (0.00, -0.07).

* patients with missing data (about 4 patients) assumed to have no improvement (i.e, the pre-dose equal the post-dose score).

[1] Ns is the number of patients for whom post-dose image score is higher than pre-dose image score and N is the total number of patients analyzed, Ps is the corresponding perceptage.

[2] Nr is the number of patients for whom pre-dose image is higher than post-dose image score and N is the total number of patients analyzed, Pr is the corresponding percentage

[3] the p-value is for testing the hypothesis in (i) concerning p_{ij}

[4] the improvement score discussed before used for these computations

Table 15/ Reviewer's Analysis

Analysis of Pre-dose and Post-dose image Scores for Certain Anatomical Areas

(Secondary Endpoint, Pivotal Study 42,440-3B, ITT Population *)

	Blinded Reader # 3				Blinded Reader # 4			
Anatomical Area	Ns/N (Ps) [1]	Nr/N (Pr) [2]	p-value Ps = Pr [3]	d = Ps - Pr [4] (95% C.I.)	Ns/N (Ps) [1]	Nr/N (Pr)[2]	p-value Ps = Pr [3]	d = Ps - Pr [4] (95% C.I.)
Stomach	65/ 85	1/ 85	<0.0001	0.75 (0.65. 0.84)	60/ 85	0/ 85	<0.0001	0.71 (0.60. 0.80)
Gastric Wall	62/ 85	1 / 85	<0.0001	0.72 (0.61, 0.81)	65/ 85	1 / 85	<0.0101	0.75 (0.65. 0.84)
Pylorus	51/ 85	2 / 85	<0.0001	0.58 (0.46. 0.68)	54/ 85	0 / 85	<0.0001	0.64 (0.52, 0.74)
Duodenum	45/ 85	1/ 85	<0.0001	0.52 (0.41. 0.63)	42/ 85	0/ 85	<0.0001	0.49 (0.38. 0.60)
Pancreatic Head	44/ 85	1 / 85	<0.0001	0.51 (0.40. 0.62)	16/ 85	0 / 85	<0.0001	0.19 (0.11. 0.29)
Pancreatic Body	37/ 85	1/ 85	<0.0001	0.42 (0.32, 0.54)	11/ 85	0/ 85	0.0010	0.13 (0.07. 0.22)
Pancreatic Tail	43/ 85	0 / 85	<0.0001	0.51 (0.40, 0.62	24/ 85	0 / 85	<0.0001	0.28 (.19. 0.39)
Pancreatic Duct	36/ 85	0/ 85	<0.0001	0.42 (0.32, 0.54)	6/ 85	0/ 85	0.0313	0.07 (0.03. 0.15)
S. Mes. Artery	28/ 85	0 / 85	<0.0001	0.33 (0.23. 0.44)	0/ 85	0 / 85	n/a [3]	0.00 (0.00. 0.04)
Para- Lymph Nodes	28/ 85	1/ 85	<0.0001	0.32 (0.22. 0.43)	1/ 85	0/ 85	1.000	0.01 (0.00. 0.07).

* patients with missing data (about 4 patients) assumed to have no improvement (i.e, the pre-dose equal the post-dose score).

[1] Ns is the number of patients for whom post-dose image score is higher than pre-dose image score and N is the total number of patients analyzed, Ps is the corresponding percentage.

[2] Nr is the number of patients for whom pre-dose image is higher than post-dose image score and N is the total number of patients analyzed, Pr is the corresponding percentage

[3] No discordant to calculate p-values

[3] the p-value is for testing the hypothesis in (i) concerning p_{ij}

[4] the improvement score discussed before used for these computations

IV.B.III. Comparison of the SonoRx Primary Endpoint With That of the Control Agent.

Since, as stated previously, the pivotal studies were not planned for this efficacy evaluation, this analysis can be viewed as a hypothesis generating. In this analysis we compare the SonoRx response rates analyzed by the sponsor with their analogues for the control agent in the same study.

Table 16, compares the per-protocol SonoRx response (percentage of patients for whom SonoRx provided additional information) with their analogue for the control agent in the same study.

Table 16/ Reviewer's Analysis

Comparison of the Percentages for Which Post-dose Image Provided Additional Information Over Pre-dose For SonoRx and Control Agent (Pivotal Studies/ Per-protocol Analysis)

	Study 42.440-3A			Study 42.440-3B		
Reader	SonoRx n/Na (%)	Control Agent n/Na (%)	p-value: χ^2 (Fisher's)	SonoRx n/Na (%)	Control Agent n/Na (%)	p-value χ^2 (Fisher's)
Investigators	46/79 (58)	14/21 (67)	0.495 (0.618)	66/88 (75)	14 25 (56)	0.068 (0.082)
Blinded Reader 1	32/73 (44)	13/19 (68)	0.058 (0.073)	55/80 (69)	19 22 (86)	0.107 (0.115)
Blinded Reader 2	72/73 (99)	19/19 (100)	0.500 (1.00)	33/78 (42)	11 22 (50)	0.507 (0.628)
Blinded Reader 3	42/ 64 (66)	12/17 (71)	0.711 (0.779)	61/81 (75)	12 21 (57)	0.102 (0.111)
Blinded Reader 4	15/64 (23)	6/17 (35)	0.299 (0.358)	52/81 (64)	17 21 (81)	0.153 (0.193)

The results of Table 16 shows that the difference between the SonoRx and control agent response rates in the same study is not significant.

Table 17 presents the results of a similar comparison to that of Table 16 for the supportive study. As this study was intended for comparing SonoRx response with that of the water we consider both the ITT and per-protocol analyses.

Table 17/ Reviewer's Analysis

Comparison of SonoRx Response (% patients with Post-dose Provided Additional Information Over Pre-dose) With Their Analogue For water in the Supportive Study.

Analysis/Reader	Response			p-value: χ 2 (Fisher's)		
Intent-to-Treat	SonoRx n/Na (%)	Water n/Na (%)	Equal Response	SonoRx vs. Water	Split Equal Category	Equal with Water
Investigators	33/53 (62)	12/53 (23)	6/53 (11)	< 0.001 (<.001)	< 0.001 (<.001)	0.004 (0.006)
Blinded Reader 1	25/50 (50)	17/50 (34)	8/50 (16)	0.107 (0.156)	0.111 (0.161)	1.00 0(1.000)
Blinded Reader 2	19/50 (38)	19/50 (38)	12/50 (24)	0.899 (1.00)	0.903 (1.00)	0.017 (0.027)
Blinded Reader 3	24/47 (51)	10/47 (21)	13/47 (28)	0.003 (0.005)	0.008 (0.013)	0.810 (1.000)
Blinded Reader 4	18/47 (38)	9/47 (19)	20/47 (43)	0.041 (0.067)	0.065 (0.098)	0.024 (0.039)
Per- protocol						
Investigators	31/48 (65)	12/48 (25)	5/48 (10)	< 0.001 (<.001)	< 0.001 (<.001)	0.004 (0.008)
Blinded Reader 1	23/43 (53)	13/43 (30)	7/43 (16)	0.030 (0.049)	0.054 (0.084)	0.514(0.667)
Blinded Reader 2	18/47 (38)	17/47 (36)	12/47 (26)	0.804 (1.00)	0.810 (1.00)	0.024 (0.039)
Blinded Reader 3	24/44 (55)	7/44 (16)	13/44 (30)	<0.001 (<.001)	<0.001 (0.001)	0.393 (0.523)
Blinded Reader 4	18/44 (41)	6/44 (14)	20/44 (46)	0.004 (0.008)	0.011 (0.019)	0.090 (0.135)

The analyses in Table 17, shows that the results are sensitive to the way of handling the 'equal response' category. Furthermore, the results vary by reader and population analyzed. Overall, there is a trend in favor of SonoRx, but the difference in the response rates does not reach statistical significance.

V. Safety/ Pediatric Use:

The sponsor compared the adverse events by body system and costar term for the SonoRx and the control agent. The sponsor's analysis shows that the difference in the adverse events between the two treatment groups was not statistically significant. In addition, the sponsor's classification of the adverse events by their severity shows that almost all of the adverse events were not serious.

The contrast agent was not intended for pediatric use. The inclusion criteria specifies that patients should be of age 18 years or older to be eligible for entry in the the study.

VI. Overall Summary/ Conclusion:

The sponsor presented the results of two pivotal studies and one supportive study in support of their claim that SonoRx enhanced images provide additional information over the pre-dose (unenhanced) images. This statistical review raised several issues:

(i) The 'Intent-to-Treat' population was not specifically defined to determine the acceptability of the images; and consequently the sponsor's analyses did not include all eligible images.
(ii) The originally selected two blinded readers were not provided with all information as planned in study protocol; two additional blinded readers employed after results of the first two blinded readers were found unacceptable. This raises statistical issues about the validity of the inference as well as about the choice of reader assessments to be considered for the efficacy evaluation.
(iii) Formulation of the primary efficacy question, Question 13 of the Case Report Form, was not appropriate for addressing efficacy, and consequently, the data analyzed by the sponsor and the statistical test were inadequate for efficacy evaluation. Specifically, if the response to whether the post-dose image provided additional information over the pre-dose image was negative, it is not clear whether this imply that the two images provided similar information or pre-dose image provides more information the post-dose image.
(iv) The sponsor carried out a post-hoc analysis in testing for the response rate against 1% and 50% instead of the anticipated rate in the study plan (75%). Also, for analyses of data from the supportive study, the sponsor used unplanned statistical methods.
(v) Sensitivity and specificity estimates of the pre-dose and post-dose images are possibly subject to large bias due to the way in which different comparative modalities were used for pathology diagnosis.
(vi) Finally, the criteria used for image assessments was, to large extent, subjective and this resulted in wide variability among the blinded readers' assessments, as measured by the Kappa statistics.

With these concerns in mind, the findings of the pivotal studies can be summarized by the following. Considering the primary efficacy endpoint, overall, there is an evidence that the SonoRx enhanced images provided additional information over the pre-dose (unenhanced) images. The magnitude of the gain in using SonoRx over that of the pre-dose image, based on to this reviewer's re-analyses is summarized in the following table:

Percentage of patients for whom SonoRx enhanced images provided additional information over the pre-dose (unenhanced) images

	Study 42.440-3A		Study 42.440-3B	
Reader	Intent-to-Treat * % (95% C. I.)	Per-protocol % (95% C. I.)	Intent-to-Treat * % (95% C. I.)	Per-protocol % (95% C. I.)
Blinded Reader 3	34 (24. 46)	62 (49. 74)	65 (54. 75)	73 (62. 82)
Blinded Reader 4	4 (1. 11)	23 (14. 36)	56 (45. 67)	64 (53. 75)

* We considered the sponsor's interpretation for the ITT population. Patients with missing data were assigned the worst score.

The lower limits of the confidence intervals for the response rates presented in this table are much lower than the anticipated primary endpoint (75%) which were used for determining the sample size of the pivotal studies.

The results of the analysis for the secondary endpoint indicate that SonoRx enhanced images show significantly improved delineation for the following anatomical areas: stomach, gastric wall, pylorus and duodenum. These results are consistent over the two pivotal studies and for Blinded readers 3 and 4. Furthermore, these results remain significant even if one adjusts the p-value for multiplicity. The difference in the percentage of images which recorded a higher post-dose (enhanced) score compared to the pre-dose (unenhanced) score for these areas anatomical areas ranges from 26 to 61 percentage points in Study 42.440-3A and from 50 to 75 percentage points in Study 42,440-3B (see Tables 14 and 15). For the pancreatic areas (head, body, tail and duct) the results show a trend in favor of SonoRx, but they were not consistent across the two studies and for both blinded readers. For the superior mesenteric artery and para-aortic lymph nodes only the assessment of Blinded Readers 3 in Study 42,440-3B shows significant results.

Concerning the utility of SonoRx enhanced image for the detection or exclusion of pathology, comparison of the sponsor's sensitivity and specificity estimates for the pre-dose and post-dose images shows there is no significant differences in both pivotal studies.

For the supportive study, the results of comparing the SonoRx enhanced images and water images, were sensitive to the way of handling the cases in which SonoRx and water images provided similar information, since the 'equal response' category is relatively large in some readers' assessment. Overall, the results though not reaching significance, show a small trend in favor of SonoRx. The response rate in the supportive study is again much below the anticipated rate (70%) which was used in the sample size calculations.

VII. Recommendations:

From statistical perspective this NDA submission provides some, but not substantial, evidence in support of efficacy. The results of the statistical analysis for the second pivotal study (# 42,440-3B) shows significant difference in the percentage of patients for whom SonoRx enhanced images provide additional information over that of the pre-dose (unenhanced) images. These results for the primary endpoint are consistent across Blinded Readers 3 and 4 who assessed both static and video images. The results for the first pivotal study (# 42,440-3A) are weaker and not consistent across the two blinded readers but are significant for Blinded Reader #3. These results remain significant when 'worse' case outcomes are imputed for 12 patients with missing data. However, response rates (SonoRx enhanced images provide additional information over the pre-dose image) for any study and for any blinded reader do not reach 75%, the planned lower limit of the 95% confidence intervals.

For the secondary endpoints, the results of the sponsor's and reviewer's analyses support the sponsor's claim about improvement of delineation of the following anatomical areas: stomach, gastric wall, pylorus and duodenum. Concerning the utility of SonoRx enhanced image for the detection or exclusion of pathology, comparison of the sponsor's sensitivity and specificity estimates for the pre-dose and post-dose images show there is no significant difference in both pivotal studies.

For the supportive study, the results of the analyses of the primary endpoint, though not reaching significance, show a small trend in favor of SonoRx. The response rate in the supportive study is again much below the anticipated rate (70%). Results of the analyses for secondary endpoints are inconclusive.

Although the results of the statistical analysis show trends in favor of SonoRx, the clinical importance of these results needs to be judged by the Medical Division.

/S/

Mohamed Al-Osh, Ph.D 1/18/97
Mathematical Statistician

Concur: /S/ 1/18/97

Michael Welch, Ph.D.
Team Leader

/S/ 1/18/97

Nancy Smith, Ph.D.
Division Director

cc:
Archival
NDA 20,773
HFD-160/Dr. Love/ Dr. Raczkowski/ Dr. Jones/ ~~Dr. Yuen~~/ Ms. Jordan
HFD-160/File Copy
HFD-344/ Dr. Lisook
HFD-720/Dr. Smith /Dr. Welch /Dr. Al-Osh
HFD-720/File Copy
HFD-720 /Chron. Copy

This review contains 48 pages of text and attachments.

Attachment I :

Table A1.1/ Sponsor's Results

Summary of the Classification of Additional Information by Pathology Type

(Intra-luminal and extra-luminal), Pivotal Studies, Per-protocol Analysis

Reader	Intra-luminal			Extra-luminal		
	42,440-3A n/N[a](%)	42,440-3B n/N (%)	Combined n/N (%)	42,440-3A n/N (%)	42,440-3B n/N (%)	Combined n/N (%)
Investigators	15/29 (52)	27/29 (93)	42/58 (72)	29/50 (58)	36/53 (68)	65/103 (63)
Blinded Reader 1	12/26 (46)	24/27 (89)	n/a[b]	17/46 (37)	31/50 (62)	n/a
Blinded Reader 2	25/26 (96)	13/27 (48)	n/a	46/46 (100)	21/48 (44)	n/a
Blinded Reader 3	15/22 (68)	23/28 (82)	n/a	28/41 (68)	40/50 (80)	n/a
Blinded Reader 4	4/22 (18)	16/28 (57)	n/a	9/41 (22)	33/50 (66)	n/a

Above data derived from Summary Tables 2.2.1 and 2.2.2 (Investigator readings only).

[a] n=number of patients with additional information provided post-dose over pre-dose
N=number of per-protocol patients for the respective pathology type. N for each pathology includes the following number of patients with both intra-luminal and extra-luminal pathology:

42,440-3A	Investigators	7 patients
	Blinded Readers 1-4	5 patients
42,440-3B	Investigators	6 patients
	Blinded Readers 1-4	6 patients

The following number of patients could not be categorized into either of these categories (intra- or extra- luminal) and have been excluded from this table:

42,440-3A	Investigators	7 patients
	Blinded Readers 1-4	6 patients
42,440-3B	Investigators	12 patients
	Blinded Readers 1-4	9 patients

[b] n/a: not applicable as different blinded readers were used in the two pivotal studies.

Table A1.2/ Sponsor's Results

Nature of Additional Information Provided Post-dose Over Pre-dose

(Pivotal Studies, Per-protocol Analysis)

Reader	Study 42,440-3A				Study 42,440-3B			
	Improved delineation of abdominal anatomy n/N[a] (%)	Improved confidence in exclusion of pathology n/N (%)	Improved delineation of pathology n/N (%)	Improved evaluation of extent of disease or pathology n/N (%)	Improved delineation of abdominal anatomy n/N[a] (%)	Improved confidence in exclusion of pathology n/N (%)	Improved delineation of pathology n/N (%)	Improved evaluation of extent of disease or pathology n/N (%)
Investigators	32/46 (70)	17/46 (37)	19/46 (41)	5/46 (11)	57/66 (86)	39/66 (59)	32/66 (48)	17/66 (26)
Blinded Reader 1	23/32 (72)	15/32 (47)	9/32 (28)	1/32 (3)	55/55 (100)	37/55 (67)	9/55 (16)	7/55 (13)
Blinded Reader 2	72/72 (100)	72/72 (100)	0/72	0/72	30/33 (91)	10/33 (30)	5/33 (15)	2/33 (6)
Blinded Reader 3	40/42 (95)	15/42 (36)	5/42 (12)	0/42	56/61 (92)	45/61 (74)	17/61 (28)	12/61 (20)
Blinded Reader 4	15/15 (100)	0/15	0/15	0/15	52/52 (100)	24/52 (46)	7/52 (14)	2/52 (4)

Above data derived from 42,440-3A and -3B clinical reports. Table continued on next page.

[a] n=number of patients with respective nature of additional information

N=Number of per-protocol patients with additional information.

Investigators/blinded readers may have indicated more than one response per patient.

Table A1.3/ Sponsor's Results

Additional Information That Could have Changed The Patients Management/ therapy

(Pivotal Studies, Per-protocol Analysis)

Reader	Study 42,440-3A	Study 42,440-3B	Combined
	n/N[a] (%)	n/N (%)	n/N (%)
Investigators	17/46 (37)	32/66 (48)	49/112 (44)
Blinded Reader 1	21/32 (66)	34/55 (62)	n/a[b]
Blinded Reader 2	0/72	33/33 (100)	n/a
Blinded Reader 3	21/42 (50)	52/61 (85)	n/a
Blinded Reader 4	0/15	5/52(10)	n/a

Above data derived from 42,440-3A and -3B clinical reports.

[a] n=number of patients with additional information that could have changed patient management/therapy.
N=number of per-protocol patients with additional information.

[b] n/a: not applicable as different Blinded Readers were used in the two pivotal studies.

Attachment II :

Table A.2.1/ Sponsor's Results

Summary of the Delineation of Abdominal anatomy: Post-dose compared to Pre-dose

Good and Excellent Ratings Combined for the Stomach, Gastric Wall, Pylorus, Duodenum and Pancreas

SonoRx Patients in (Pivotal Studies; SonoRx Patients, Per-Protocol Analysis)

	Study 42.440-3A			Study 42.440-3B		
Anatomical Area: Stomach	Post-dose n/N (%)	Post-dose n/N (%)	p-value[b]	Post-dose n/N (%)	Post-dose n/N (%)	p-value[b]
Investigators	3/79 (4)	30/79 (38)	0.0001	3/87 (3)	42/87 (48)	0.0001
Blinded Reader 1	3/73 (4)	22/73 (30)	0.0001	11/80 (14)	44/80 (55)	0.0001
Blinded Reader 2	28/73 (38)	64/73 (88)	0.0001	0/76	0/76	0.0001
Blinded Reader 3	7/64 (11)	38/64 (59)	0.0001	15/81 (19)	48/80 (60)	0.0001
Blinded Reader 4	4/64 (6)	20/64 (31)	0.0001	4/81 (5)	28/81 (35)	0.0001
	Study 42.440-3A			Study 42.440-3B		
Anatomical Area: Gastric Wall	Post-dose n/N (%)	Post-dose n/N (%)	p-value[b]	Post-dose n/N (%)	Post-dose n/N (%)	p-value[b]
Investigators	12/79 (15)	42/79 (53)	0.0001	4/87 (5)	30/87 (34)	0.0001
Blinded Reader 1	16/73 (22)	19/73 (26)	NS	11/80 (14)	44/80 (55)	0.0001
Blinded Reader 2	30/73 (41)	63/73 (86)	0.0001	1/75 (1)	1/76 (1)	NS
Blinded Reader 3	8/64 (13)	40/64 (63)	0.0001	14/81 (17)	55/80 (69)	0.0001
Blinded Reader 4	6/64 (9)	20/64 (31)	0.0001	7/81 (9)	38/81 (47)	0.0001

Table A.2.1 Continued

	Study 42.440-3A			Study 42.440-3B		
Anatomical Area: Pylorus	Post-dose n/N (%)	Post-dose n/N (%)	p-value[b]	Post-dose n/N (%)	Post-dose n/N (%)	p-value[b]
Investigators	13/77 (17)	38/77 (49)	0.0001	1/87 (1)	24/87 (28)	0.0001
Blinded Reader 1	18/73 (25)	29/73 (40)	0.0015	3/80 (4)	34/80 (43)	0.0001
Blinded Reader 2	34/73 (47)	66/73 (90)	0.0001	0/75	0/76	0.0001
Blinded Reader 3	5/64 (8)	33/64 (52)	0.0001	20/81 (25)	55/81 (68)	0.0001
Blinded Reader 4	13/64 (20)	21/64 (33)	0.0021	6/81 (7)	31/81 (38)	0.0001
	Study 42.440-3A			Study 42.440-3B		
Anatomical Area: Duodenum	Post-dose n/N (%)	Post-dose n/N (%)	p-value[b]	Post-dose n/N (%)	Post-dose n/N (%)	p-value[b]
Investigators	6/78 (8)	20/78 (26)	0.0001	2/88 (2)	29/88 (33)	0.0001
Blinded Reader 1	2/73 (3)	4/73 (5)	0.0109	2/80 (3)	23/80 (29)	0.0001
Blinded Reader 2	17/73 (23)	44/73 (60)	0.0001	0/75	0/76	0.0001
Blinded Reader 3	1/64 (2)	23/64 (36)	0.0001	11/81 (14)	32/81 (40)	0.0001
Blinded Reader 4	5/64 (8)	10/64 (16)	0.0001	1/81 (1)	15/81 (19)	0.0001

Table A.2.1 Continued

	Study 42.440-3A			Study 42.440-3B		
Anatomical Area: Pancreatic Head	Post-dose n/N (%)	Post-dose n/N (%)	p-value[b]	Post-dose n/N (%)	Post-dose n/N (%)	p-value[b]
Investigators	29/77 (38)	48/77 (62)	0.0001	32/88 (36)	65/88 (74)	0.0001
Blinded Reader 1	38/73 (52)	38/73 (52)	NS	26/80 (33)	48/80 (60)	0.0001
Blinded Reader 2	47/73 (64)	56/73 (77)	0.0001	16/78 (21)	20/77 (26)	0.0001
Blinded Reader 3	41/64 (64)	48/64 (75)	0.0054	34/81 (42)	61/81 (75)	0.0001
Blinded Reader 4	21/64 (33)	27/64 (42)	0.0075	51/81 (63)	62/81 (77)	0.0001
	Study 42.440-3A			Study 42.440-3B		
Anatomical Area: Pancreatic Body	Post-dose n/N (%)	Post-dose n/N (%)	p-value[b]	Post-dose n/N (%)	Post-dose n/N (%)	p-value[b]
Investigators	46/77 (60)	62/77 (81)	0.0001	44/87 (51)	68/87 (78)	0.0001
Blinded Reader 1	48/73 (66)	52/73 (71)	NS	42/80 (53)	67/80 (84)	0.0001
Blinded Reader 2	65/73 (89)	71/73 (97)	0.0002	15/78 (19)	22/77 (29)	0.0001
Blinded Reader 3	50/64 (78)	54/64 (84)	0.0273	47/81 (58)	67/81 (83)	0.0001
Blinded Reader 4	36/64 (56)	39/64 (61)	NS	40/81 (49)	50/81 (62)	0.0010

Table A.2.1 Continued

	Study 42.440-3A			Study 42.440-3B		
Anatomical Area: Pancreatic Tail	Post-dose n/N (%)	Post-dose n/N (%)	p-value[b]	Post-dose n/N (%)	Post-dose n/N (%)	p-value[b]
Investigators	11/77 (14)	30/77 (39)	0.0001	7/87 (8)	34/87 (39)	0.0001
Blinded Reader 1	20/73 (27)	23/73 (32)	0.0216	2/80 (3)	21/80 (26)	0.0001
Blinded Reader 2	37/73 (51)	53/73 (73)	0.0001	3/78 (4)	5/77 (6)	0.0035
Blinded Reader 3	17/64 (27)	31/64 (48)	0.0001	21/81 (26)	38/81 (47)	0.0001
Blinded Reader 4	6/64 (9)	6/64 (9)	NS	14/81 (17)	27/81 (33)	0.0001
	Study 42.440-3A			Study 42.440-3B		
Anatomical Area: Pancreatic Duct	Post-dose n/N (%)	Post-dose n/N (%)	p-value[b]	Post-dose n/N (%)	Post-dose n/N (%)	p-value[b]
Investigators	17/77 (22)	27/77 (35)	0.0001	12/88 (14)	25/88 (28)	0.0001
Blinded Reader 1	8/73 (11)	12/73 (16)	NS	24/80 (30)	44/80 (55)	0.0001
Blinded Reader 2	58/73 (79)	66/73 (90)	0.0001	4/73 (5)	7/71 (10)	NS
Blinded Reader 3	54/64 (84)	55/64 (86)	NS	41/81 (51)	59/81 (73)	0.0001
Blinded Reader 4	8/64 (13)	7/64 (11)	NS	16/81 (20)	19/81 (24)	0.0313

Source: above data derived from Summary Tables 3.1 to 3.5.

[a] n=number of patients with good or excellent rating and N=number of patients with scores for the respective anatomical areas

[b] $p \leq 0.05$ denotes statistically significant better delineation from pre-dose to post-dose based on all rating (excellent, good, fair, poor and none) using the Wilcoxon signed rank test. In all cases in which the excellent and good ratings pre-dose and post-dose were identical, the improvements in other ratings post dose indicated a statistically significant improvement in delineation scores. NS denotes not significant ($p>0.05$).