CENTER FOR DRUG EVALUATION AND RESEARCH

APPROVAL PACKAGE FOR:

APPLICATION NUMBER

21-256

Statistical Review(s)

STATISTICAL NDA REVIEW AND EVALUATION
(CLINICAL STUDIES)

NDA: 21-256 **Date: 11/19/2001.**

APPLICANT: ChiRhoClin, Inc.

NAME OF DRUG: Synthetic Human Secretin.

INDICATION: 1) Diagnosis of pancreatic exocrine), 3) Facilitation of retrograde cholengio pancreatography (ERCP). 2) Diagnosis of gastrinoma during endoscopic

USER FEE DUE DATE: 12/14/2001

DRUG CLASSIFICATION: 1P

DOCUMENT REVIEWED: NDA Volumes 1-3, 28 – 30, and 39, Dated June 14, 2001.

MEDICAL REVIEWER: Marcel Barreiro, MD., HFD 180

STATISTICAL REVIEWER: Wen-Jen Chen, Ph.D., HFD 715

STATISTICAL ISSUES: Statistical equivalence claimed based on test of significance.

KEY WORDS/PHRASES: Clinical studies; NDA review; Diagnostic accuracy; Active control/Clinical equivalence; Carry-over effect; Sample size.

1.0 . INTRODUCTION

In the Volume 3 of this NDA submission, the sponsor made the following observations with regard to Secretin:

> Secretin is a gastrointestinal peptide hormone that was first extracted from porcine duodenum by Jorpes & Mutt. The peptide was subsequent]y sequenced and synthesized by Mutt, Bodansky, and their co-workers at the Karolinska Institute. Secretin is a highly purified naturally occurring hormone.
>
> [ ]

In the submitted document, the sponsor proposed the following three indications for the use of synthetic human secretin (sHS): 1. Diagnosis of pancreatic exocrine — 2. Diagnosis of gastrinoma , and 3. Facilitation of during ERCP. Six clinical studies, CRC98-1, CRC98-2, CRC77-2, CRC99-8, CRC99-9, and CRC98-4, were submitted by the sponsor to support the above three proposed indications. Of the six studies, the three studies, CRC98-1, CRC98-2, and CRC99-9 were to support the first indication, two studies, CRC97-2 and CRC99-8, were to support the second indication, and the open label study CRC98-4 was to support the third indication.

2.0 Studies CRC98-1, CRC98-2, and CRC99-9 used for diagnosis of pancreatic exocrine

One notes that study CRC98-1 (NDA 21-136) which provided no information on the tested agent sHS was already submitted by the sponsor on December 1999 to support synthetic porcine secretin (sPS) on the use of diagnosis of pancreatic exocrine — and its statistical review and evaluation report was issued on March 7, 2000. The decision made was that agent sPS was approvable (not yet approved as of this review) for the use of diagnosis of pancreatic exocrine Therefore, study CRC 98-1 is not considered to support the efficacy of sHS used for the diagnosis of pancreatic exocrine and no further review on this study is necessary.

As the reasons stated in the refusing-to-file letter for sHS on the use of diagnosis of pancreatic exocrine — dated on May 11, 2000, Study CRC98-2 is not considered to support the efficacy of sHS used for the diagnosis of pancreatic exocrine — Therefore, for the diagnosis of pancreatic exocrine , — only Study CRC99-9 is considered.

2.1 Background Information for Study CRC99-9

Objectives: The objective of this Phase III study was to obtain comparative pharmacological, diagnostic efficacy, and safety data for synthetic porcine secretin, synthetic human secretin, and biological derived porcine secretin as diagnostic agents in patients with a diagnosis of chronic pancreatitis.

Study Design: This study is a randomized, crossover design comparing a single dose of synthetic porcine secretin (sPS) 0.2 µg/kg, synthetic human secretin (sHS) 0.2 µg/kg, and biologically derived porcine secretin (bPS) 1 CU/kg administered intravenously in 6 volunteer patients with a documented diagnosis chronic pancreatitis. [No information on blinding was provided in the protocol]. Each patient was to undergo 3 secretin stimulation tests separated by at least 24 hours. Pancreatic secretion volume and bicarbonate concentrations are the primary pharmacological end points.

Study Population: Male and female patients, with a diagnosis of chronic pancreatitis documented by a prior secretin stimulation test with bPS [copied from the protocol subsection of 5.1 in page

1811 of Volume 23 submitted by the sponsor] and by clinical and laboratory findings consistent with this diagnosis. Females were required to be of non-childbearing potential or using a medically approved method of contraception. Patients were required to provide written informed consent.

Efficacy Variables: 1. Diagnostic concordance with bPS for chronic pancreatitis. 2. Pharmacological variables: pancreatic juice volume and pancreatic juice bicarbonate concentration.

Pharmacological Evaluation: The GLM procedure from Statistical Analysis System (SAS Institute, Cary, NC) was used to test for treatment and carryover effect. Multiple comparisons and regression procedures were utilized to compare treatments bPS and sHS. Summary and descriptive statistics were used for treatment effects.

Number of Subjects: Six (6) patients were planed and enrolled, completed all 3 tests (sHS, sPS, and bPS) and were analyzed.

2.2.1 Sponsor's Statistical Analysis and Results for Study CRC99-9

Demographics and Baseline Characteristics

Except for the baseline pancreatic secretion volume and pancreatic secretion bicarbonate concentration, the sponsor did not present the results of the statistical analyses on demographic variables and baseline characteristics. The analysis results for the baseline pancreatic secretion volume and pancreatic secretion bicarbonate concentration were demonstrated in the next sub-section for "Results for pharmacological effect analysis".

Analysis of Diagnostic Efficacy Results

The sponsor reported that the six patients tested positive for chronic pancreatitis (HCO_3 concentration < 80 mEQ/L in each aliquot) after stimulation with sHS, sPS, and bPS. There was 100% agreement in the diagnostic results for the diagnosis of chronic pancreatitis among the three drugs sHS, sPS, and bPS.

Results for Pharmacological Effect Analysis

The results for the pharmacological effect analyses in the comparisons among sHS 0.2 μg/kg, sPS 0.2 μg/kg, and bPS 1CU/kg on mean values for pancreatic juice volume and bicarbonate concentration at each fifteen minute interval and for the entire 60 minutes were presented by Table 2.2.1.1 and Table 2.2.1.2, respectively.

Table 2.2.1.1 (Sponsor's) Pancreatic Stimulation Results for 15 Minute Intervals

TREAT		V_B	BC_B	V_15	BC_15	V_30	BC_30	V_45	BC_45	V_60	BC_60
SHS	Mean	27.2	10.0	69.7	43.3	46.2	63.8	45.7	61.8	45.2	59.3
SHS	STD	27.0	13.2	34.4	8.9	35.5	12.9	32.7	14.3	35.1	11.3
SHS	%CV	99.5	131.6	49.4	20.2	75.9	20.2	71.6	23.1	77.6	19.0
sPS	Mean	26.8	4.3	50.7	32.8	75.5	50.3	46.2	53.7	45.2	44.2
sPS	STD	31.2	6.0	29.2	14.9	50.9	15.2	36.4	15.5	23.3	12.9
sPS	%CV	116.4	138.2	57.5	45.5	67.4	30.2	78.9	28.8	51.5	29.3
bPS	Mean	32.7	4.3	55.0	31.2	52.2	44.8	51.0	55.3	42.8	48.3
bPS	STD	25.6	8.0	38.2	9.3	44.3	16.2	30.8	13.2	22.6	20.8
bPS	%CV	78.5	185.6	69.5	30.0	84.9	36.1	60.3	23.8	52.7	43.0
	Prob	0.8687	0.5924	0.5740	0.6248	0.1420	0.3922	0.9981	0.9506	0.8376	0.5908

Source: Sponsor's table in page 553, Volume 29; BC=bicarbonate concentration (mEq/L); V=volume (mL).

Table 2.2.1.2 (Sponsor's) Pancreatic Stimulation Results for 60 Minute Sample

TREAT		V_1_60	BC_1_60	B_TBC	TBC
sHS	Mean	206.7	54.9	12.2	12.2
sHS	STD	119.3	10.2	9.1	9.0
SHS	%CV	57.7	18.7	74.4	73.7
sPS	Mean	217.5	44.4	10.2	10.3
sPS	STD	110.0	12.8	7.6	7.6
sPS	%CV	50.6	28.8	74.8	73.5
bPS	Mean	201.0	44.5	9.6	9.6
bPS	STD	118.6	12.0	6.8	6.6
bPS	%CV	59.0	27.0	70.4	68.5
	Prob	0.7504	0.7265	0.2236	0.2385

Source: Sponsor's table in page 553, Volume 29; BC=bicarbonate concentration (mEq/L); V=volume (mL); TBC=total bicarbonate (mEq).

The results from Table 2.2.1.1 and Table 2.2.1.2 indicated that the mean values for pancreatic juice volume and bicarbonate concentration at each fifteen-minute interval and for the entire 60 minute sampling period stimulated by the three drugs, sPS, sHS, and bPS, showed no statistically significant differences.

Based on the above non-significance results, in the conclusion for the pharmacological effects, the sponsor purported that sHS, sPS, and bPS have statistically equivalent pharmacological effects.

Adverse Events

There were 2 adverse events (AEs) in 2 patients. Both AEs occurred after administration of bPS. Patients one had mild flushing lasting one minute, which resolved spontaneous. Patients two had mild nausea lasting 3 minutes, which resolved spontaneously. There were no other safety problems observed.

2.3 Reviewer's Analyses and Comments

As indicated in the sponsor's submission, one objective of this study was to demonstrate the comparative pharmacological effects for sHS 0.2 μg/kg, sPS 0.2 μg/kg and bPS 1CU/kg. However, instead of using the preferred confidence interval approach, the sponsor used the non-significant results to claim that the pharmacological effects for sPS, sHS, and bPS were statistically equivalent.

In addition, the contents of the pharmacological variables are the outcomes stimulated by the three diagnostic agents (sHS, sPS, and bPS) and not the blood concentrations of these three drugs. Therefore, these pharmacological variables should be treated as clinical endpoints and the two-sided 95% confidence intervals should have been applied by the sponsor to assess the clinical equivalence with regard to the three agents on the pharmacological effects.

In order to appraise the sponsor's efficacy claim, this reviewer performs the following three analyses on each of the three pharmacological variables, peak bicarbonate concentration (PEAKBC), total bicarbonate (TBC), and total sixty (60) minute pancreatic juice volume (V_1_60): i.) Confidence interval analysis, ii) Graphic display, and iii.) Probability analysis. Variable PEAKBC is defined as the maximum of 0 to 15, 15 to 30, 30 to 45, and 45 to 60 minute bicarbonate concentrations. Due to the small numbers of patients and lack of pre-specification, these three analyses must be viewed as descriptive in nature. Data used in the statistical analyses were copied from the sponsor's submitted document of Volume 29. In addition, due to small sample size (only 6 patients), no sub-group analysis is performed.

i.) Confidence interval analysis

This reviewer calculated the 95% confidence intervals on the differences of the following two treatment effects separately for variables PEACKBC, TBC, and V_1_60: sPS versus bPS (sPS – bPS) and sHS versus bPS (sHS – bPS).

Since the structure of the sponsor's crossover design setting induced non-estimable treatment effects when using the model with parameters of treatment and period effects, the parameters in the model used to calculate the two-sided 95% confidence intervals are treatment effects only.

Table 2.3.1 displays the 95% confidence intervals on the differences of two treatment effects, bPS sample means, and the upper and lower bounds divided by bPS sample means with regard to the three variables, PEACKBC, TBC, and V_1_60, separately for sHS versus bPS (sHS-bPS) and sPS

versus bPS (sPS-bPS).

Table 2.3.1 The 95% confidence intervals of the differences on the two treatment effects
sHS-bPS

VARIABLES	95% CONF. INT.# LWR. BND.[1]	UPR. BND.[2]	BPS SAMPLE MEAN	PERCENTAGE LB/BPS M.[3]	UB/BPS M.[4]
PEAKBC	1.38	12.30	58.8	2.3%	21.0%
TBC	-0.58	5.97	9.51	-6.1%	63.0%
V_1_60	-26.30	37.60	201.00	-13.0%	19.0%

sPS-bPS

VARIABLES	95% CONF. INT.# LWR. BND.[1]	UPR. BND.[2]	BPS SAMPLE MEAN	PERCENTAGE LB/BPS M.[3]	UB/BPS M.[4]
PEAKBC	-7.29	3.62	58.8	-12.4%	6.2%
TBC	-2.46	4.10	9.51	-26.0%	43.0%
V_1_60	-15.47	48.47	201.00	-8.0%	24.0%

#: Confidence Interval; [1]: Lower Bound; [2]: Upper Bound; [3]: Lower Bound/bPS Sample Mean; [4]: Upper Bound/bPS Sample Mean.

For the differences of sHS versus bPS (sHS-bPS), Table 2.3.1 contains the following results:

- The lower and upper bounds for the mean difference (sHS – bPS) of peak bicarbonate concentration are estimated to be 2.3% and 21.0% of the bPS mean, respectively.
- The lower and upper bounds for the mean difference (sHS – bPS) of total bicarbonate are estimated to be –6.1% and 63.0% of the bPS mean, respectively.
- The lower and upper bounds for the mean difference (sHS – bPS) of total volume are estimated to be –13.0% and 19.0% of the bPS mean, respectively.

In response to this reviewer's information request for the previous study CRC98-1 with regard to the delta margin used for the equivalence analysis, the sponsor indicated that 20% of the bPS mean was selected as the delta margin. To be consistent with the previous study, the 20% of bPS mean is again used as the delta margin for this study. One notes that the upper bounds of the 95% two-sided confidence intervals on the differences of sHS versus bPS for peak bicarbonate concentration and total bicarbonate are both greater than the delta margin. Therefore, the two drugs sHS and bPS are not statistically equivalent when assessed by peak bicarbonate concentration and total bicarbonate, especially, for peak bicarbonate, the mean of sHS is actually significantly greater than that of bPS.

For the differences of sPS versus bPS (sPS – bPS), Table 2.3.1 indicates:

- The lower and upper bounds for the mean difference (sPS – bPS) of peak bicarbonate concentration are estimated to be –12.4.0% and 6.2% of the bPS mean, respectively.
- The lower and upper bounds for the mean difference (sPS – bPS) of total bicarbonate are estimated to be –26.0% and 43.0% of the bPS mean, respectively.
- The lower and upper bounds for the mean difference (sPS – bPS) of total volume are estimated to be –8.0% and 24.0% of the bPS mean, respectively;

Similarly, the upper bounds of the 95% two-sided confidence intervals on the differences of sPS versus bPS on total bicarbonate and total volume are greater than the delta margin claimed by the sponsor for the statistical equivalence. Therefore, the two agents sPS and bPS are not statistically equivalent when assessed by peak bicarbonate and total volume.

ii) Graphic Display

The following graphs demonstrate each patient's data on sPS, sHS, and bPS separately for variables PEACKBC, TBC, and V_1_60, using 6 patients from Study CRC99-9. The diagrams for the three variables, PEACKBC, TBC, and V_1_60, are given in Fig. 2.3.1, Fig. 2.3.2, and Fig. 2.3.3, respectively. In each plot, symbols 'S', 'B', and 'H' denote data from sPS, bPS, and sHS, respectively.

Figure 2.3.1

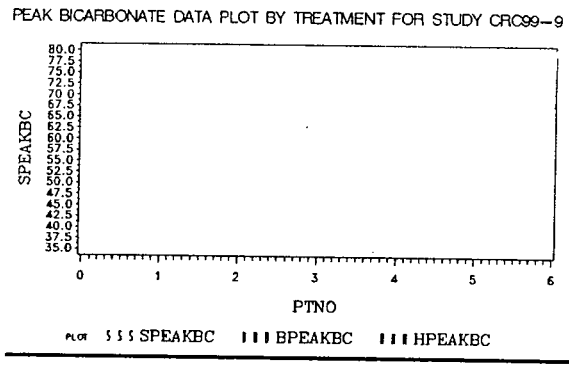

Figure 2.3.2

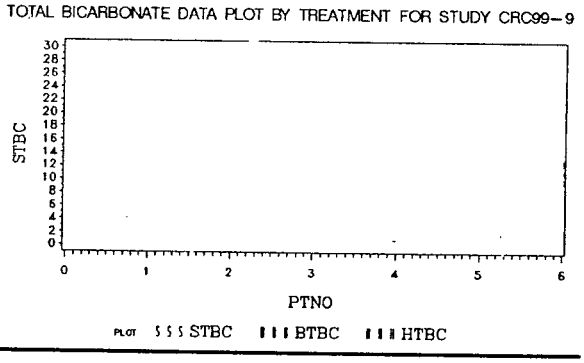

Figure 2.3.3

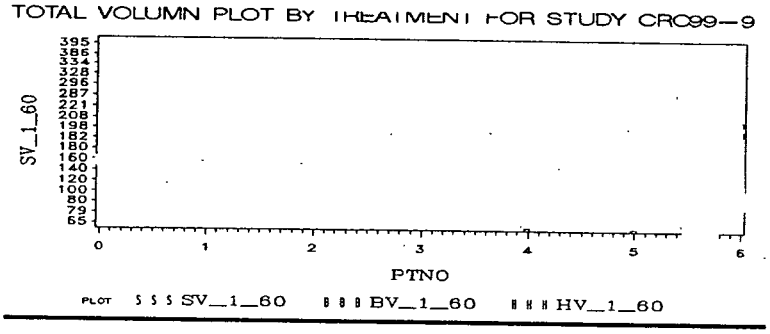

Figure 2.3.1, 2.3.2, and 2.3.3 indicate the following phenomena:

- For each of the three pharmacological variables, the within patient variations of sPS, sHS, and

bPS are in general large, especially for patients 1, 3, and 4 for peak bicarbonate concentration, patients 2 and 3 for total bicarbonate and patients 1, 2, and 4 for total volume.

- Similarly, for each of the three pharmacological variables, the between patient variations for each of the three agents, sPS, bPS, and sHS, are large as well.
- For peak bicarbonate concentration, of the six sHS data, five are greater than those of bPS; in the same way, for total bicarbonate, four out of six sHS data are greater than those of bPS, indicating sHS has the tendency to generate larger pharmacological values than those of bPS.

Based on the above observations, the following comments are made:

- Due to small sample size (6 patients) and large variation (within and between patients) on the pharmacological data collected by the sponsor for sPS, sHS, and bPS, the sponsor's claim that sHS, sPS, and bPS have statistically equivalent pharmacological effects based on the results of non-significance is not reliable, highly likely due to lake of power.
- In addition, the peak bicarbonate concentration and total bicarbonate data have carried certain evidence that patients stimulated by sHS may generate larger pharmacological values than those of bPS.

iii) Probability analysis

In order to assess the robustness of the 100% agreement reported by the applicant for the diagnosis of chronic pancreatitis by the two diagnostic agents, sHS and bPS, this reviewer calculated the probability of 100% agreement between sHS and bPS in the diagnosis of pancreatitis using six patients, under the assumption that there exits a certain disagreement probability between these two agents.

For example, if 10% of subjects are expected to differ on the tests, the probability of no disagreement in 6 subjects is 53%. For another example, if the disagreement probability for the two diagnostic agents is 30%, the probability of the 100% agreement in the diagnostic results tested by these two agents on the six patients is 12%. These examples suggest that due to small sample size (6 patients), the observed lack of difference can not rule out a relatively large disagreement rate.

2.4 Comments/Conclusions on treatment effects

- The upper bounds of the 95% confidence intervals on the two mean differences (sHS - bPS) for the three variables, peak bicarbonate concentration, total bicarbonate, and total volume, calculated by the sick patients in pivotal study CRC99-9, are estimated to be 21%, 63%, and 19% of the bPS means, respectively. These results indicate that the sHS means for peak bicarbonate concentration, total bicarbonate, and total volume can be greater than those of bPS by up to 21%, 63%, and 19%, respectively.
- In addition, if 20% of the bPS mean was selected as the delta margin by the sponsor's response to the information request letter for the previous Study CRC98-1, the two drugs sHS and bPS are not statistically equivalent when assessed by peak bicarbonate concentration and total

bicarbonate.

- Graphic displays suggest:
  - i.) Due to small sample size (6 patients) and large variation (within and between patients) on the pharmacological data collected by the sponsor for sPS, sHS, and bPS, the sponsor's claim that sHS, sPS, and bPS have statistically equivalent pharmacological effects based on the results of non-significance is not reliable, highly likely due to lack of testing power.
  - ii.) In addition, the peak bicarbonate concentration and total bicarbonate data have carried certain evidence that patients stimulated by sHS may generate larger pharmacological values than those of bPS.
- Under the assumption of 30% disagreement probability for the two diagnostic agents, synthetic human secretin, and biologically derived porcine secretin, the probability of 100% agreement in the diagnostic results tested by these two agents on the six sick patients is 12%. Thus, the observed lack of difference can not rule out a relatively large disagreement rate between these two agents.
- In conclusion, the results of efficacy analyses from the single pivotal study are not statistically persuasive to support the use of synthetic human secretin for the diagnosis of pancreatic exocrine.

3.1 Background information for Studies CRC99-8 & CRC97-2 used for the diagnosis of gastrinoma

Objectives: The objective of these two Phase III studies was to evaluate the diagnostic efficacy and safety of sPS and sHS in comparison to bPS (approved by FDA in 1981) for patients with documented gastrinoma alone or as a component of MEN-I (Type-I Multiple endocrine neoplasia).

Study Design: The two studies were randomized, single blind, active controlled, and single center studies. Study CRC97-2 was a 2-way crossover setting with three patients enrolled while Study CRC99-8 was a 3-way crossover setting with six patients enrolled. The doses administered for bPS, sPS, and sHS, were 2CU/kg, 0.4μg/kg, and 0.4μg/kg, respectively. In the crossover setting, single doses of sPS, sHS, and bPS administered at least 2.5 hours apart.

Study Population: Consisted of male and female of non-childbearing potential with a documented diagnosis of gastrinoma and willing to sign written, informed consent.

Diagnostic Efficacy: Increase in serum gastrin concentrations from baseline was assessed. Diagnostic paradigm of greater than 110 pg/mL increase is the diagnosis of gastrinoma. Gastrin determinations were at 0, 1, 2, 5, 10, 15, 20, and 30 minutes post secretin injection. [However, the sponsor did not specify which time point would be used for the assessment of 110 pg/mL increase.]

Statistical Analysis Methods: The comparison of synthetic porcine secretin (sPS), biologically derived porcine secretin (bPS), and synthetic human secretin (sHS), with respect to gastrin results, was performed by the analysis of variance on data from Study CRC99-8 only. The comparison at baseline used a model that included the factors of product, period, and patient. The sponsor claimed that for the post-baseline time points, the change from baseline was analyzed using a repeated measure analysis of variance. The model included the factors of product, patient, time point, and the product by time point interaction. The significance of the within product change was based on a paired t-test. Due to some patients with very high gastrin results, all p-values reported are based on the use of a $\log_{10}$ data transformation.

Number of Subjects: Six patients have been enrolled for StudyCRC99-8 and completed all three treatments. One of these patients had been previously enrolled but had poor venous access preventing proper blood samples from being obtained, and was withdrawn. Subsequently, the patient was re-enrolled and completed all three tests. In addition, 3 patients from the 2-way crossover amendment to CRC97-2 at — in which sPS and bPS were compared are included in the diagnostic efficacy analyses.

3.2 Sponsor's Statistical Analysis and Results

Demographics and Baseline Characteristics

Except for the baseline gastrin concentration, the sponsor did not present the results of the statistical analyses on demographic variables and baseline characteristics. The analysis results for the baseline gastrin concentration was summarized in the next sub-section for "Results for Analysis on serum gastrin concentrations".

Analysis of Diagnostic Efficacy Results

Six patients in the 3-way crossover study (CRC99-8) with tissue confirmed gastrinoma, had positive diagnostic results after sPS, sHS, and bPS. For the three patients in the 2-way crossover study (Study CRC97-2), two patients, with known gastrinoma, had positive tests after sPS and bPS. A third patient who had previously tested positive with sPS and had subsequently undergone curative resection of a solitary gastrinoma had negative tests after sPS and bPS. There was diagnostic agreement for sPS, sHS and bPS in 6 of 6 patients (3-way crossover) and between sPS and bPS in 9 of 9 patients overall and in 8 of 8 patients remaining positive for gastrinoma.

Finally, the sponsor concluded that there is 100% diagnostic agreement among sPS, sHS, and bPS in terms of producing test results positive for gastrinoma in patients with a tissue diagnosis of gastrinoma.

Results for Analysis on serum gastrin concentrations

The results on the comparisons of gastrin concentration among sPS, bPS, and sHS, in Study CRC99-8, using patients who were diagnosed with gastrinoma and evaluated using all three products were concluded below by the sponsor:

1) For baseline data analysis, no statistically significant differences related to the product ($p=0.90$) and period effect ($p=0.72$) were found.
2) For the mean change from the baseline, the three products performed in a similar manner and the overall comparisons on the mean changes from baseline among the three products were not significantly different ($p=0.44$).
3) In the pair-wise treatment efficacy comparisons for sHS, sPS, and bPS at each assessed time point, only the comparison of sHS versus sPS evaluated mean change from baseline change at 15 minutes of injection was significant ($p=0.0274$).
4) For all three products, the median changes from baseline within products were statistically significant at 2 minutes post-baseline ($p=0.015$ for sPS, $p=0.005$ for bPS, and $p=0.035$ for sHS): the median increase for sHS was 371.00 compared to 839.00 for bPS and 463.50 for sPS.
5) For the median change from baseline to 1 minute post-baseline, bPS was statistically significant (179.00).
6) For the median change from baseline to 5 minutes post-baseline, both sPS and bPS were statistically significant (381.00 for sPS and 608.00 for bPS).
7) Finally, for the assessed time points after 5 minutes post-baseline, all of the median changes

from baseline were not statistically significant.

Based on the above results, the sponsor concluded that sPS, sHS and bPS produced equivalent pharmacological responses in terms of serum gastrin in gastrinoma patients. [However, the results for the equivalent phannacological responses in terms of serum gastrin were based on non-significant results of tests.]

Adverse Events

Three patients had a total of 13 adverse events. Subject #1 reported mild tingling in both legs for 2 minutes after administration of sHS, which spontaneously resolved. After administration of sPS, subject #1 complained of moderate sweating of the hands and feet for 6 minutes and 2 brief episodes of burning in the stomach. These were listed as four AEs and resolved spontaneously.

In addition, subject #3 complained of an upset stomach and burning in the stomach (2 AEs) for 5 minutes after sHS, burning in the stomach for 3 minutes after sPS, and burning in the abdomen for 5 minutes after bPS. Finally, subject #5 had a sensation of warmth in the face and abdomen for one minute following bPS (2 AEs) and for 2 minutes following sHS (2 AEs).

3.3 Reviewer's Analysis and Comments

As noted by this reviewer, the sponsor used the non-significant results from the superiority analyses to claim that the three agents (sPS, sHS, and bPS) produced equivalent serum gastrin concentrations for patient with gastrinoma. Since a small number of patients (6) from Study CRC99-8 were used in the superiority analyses, the non-significant results may be due to lack of power to detect the differences among the three agents.

In addition, it is noted that the contents of the serum gastrin concentrations are the outcomes stimulated by the three diagnostic agents (sHS, sPS and bPS) and not the blood concentrations of these three drugs. Therefore, the gastrin concentration should be treated as a clinical endpoint and a two-sided 95% confidence interval approach should have been applied by the sponsor to assess the clinical equivalence with regard to the three agents on serum gastrin concentration.

In order to appraise the sponsor's efficacy claim, this reviewer performs the following three analyses for both studies (CRC99-8 and CRC97-2) on the serum gastrin concentrations i.) confidence interval analysis on the difference of two treatment effects, ii) Graphic display, and iii.) Probability analysis. Due to the small numbers of patients and lack of pre-specification, these three analyses must be viewed as descriptive in nature. Data used in the statistical analyses were copied from the sponsor's submitted document of Volume 29.

i.) Confidence interval analysis on the difference of two treatment effects

It is noted that data for the two studies CRC99-8 and CRC 97-2 were collected from different crossover settings: three treatments and periods for Study CRC99-8 and two treatments and periods for CRC97-2. This reviewer therefore, only uses data from Study CRC99-8 to calculate the two-sided 95% confidence intervals on the differences of mean changes of serum gastrin concentration from baseline to each of 5, 10 and 20 minutes post-baselines for sPS versus bPS (sPS – bPS) and sHS versus bPS (sHS – bPS).

Due to small sample sizes (6 patients), it may lack of power to detect the carry over effects of the three drugs (sPs, sHS, and bPS) on the serum gastrin concentration. In addition, data for the three treatments, sPS, sHS, and bPS, from Study CRC99-8 were not collected from a statistical crossover design. This reviewer therefore, does not perform the carry over effect test and assume no carry over effects for Study CRC99-8. However, since in the sponsor's crossover setting, single doses of sPS, sHS, and bPH administered only 2.5 hours apart, the carryover effects of drugs are dubious. The parameters in the model used to calculate the two-sided 95% confidence intervals consist of treatment effect, period effect, and random subject effect.

Originally, the analyses on the differences of mean changes from baseline to each of 5, 10, and 15 minutes were recommended by the medical reviewer, Dr. Barreiro, due to more than 90% of gastrinoma patients exhibiting a rise in serum gastrin within 15 minutes of secretin administration. However, since at 15 minute post-baseline, data of the three drugs, sPS, bPS, and sHS, for one patient and data of sPS for another patient were missing, the parameters of the differences on the pair-wise treatment effects become non-estimable based on the model proposed above. Thus, the confidence interval analysis excludes data from 15 minute post-baseline. Instead of using transformed data, the model is applied to the raw data.

Table 3.3.1.1 displays the two-sided 95% confidence intervals on the differences of mean changes of gastrin concentration from baseline to each of 5, 10 and 20 minutes post-baselines for sHS versus bPS (sHS – bPS) and sPS versus bPS (sPS – bPS).

Table 3.3.1.1 (Reviewer's) 95% confidence intervals on the differences of mean changes from baseline
sHS versus bPS (sHS – bPS)

P. BASELINE[#]	95% CONF. INT.[#] LWR. BND.[1]	UPR. BND.[2]	BPS LSQ.[&] MEAN	PERCENTAGE LB/BPS M.[3]	UB/BPS M.[4]
5-minute	-1331.0	277.1	1280.0	-104.0%	22.0%
10-minute	-5283.0	2378.0	3255.0	-162.0%	73.1%
20-minute	-3995.0	1349.1	3601.0	-111.0%	37.5%

Table 3.3.1.1 (Reviewer's) 95% confidence intervals on the differences of mean changes from baseline (Continued)

sPS versus bPS (sPS – bPS)

P. BASELINE[#]	95% CONF. INT.[#] LWR. BND.[1]	UPR. BND.[2]	BPS LSQ[&] MEAN	PERCENTAGE LB/BPS M.[3]	UB/BPS M.[4]
5-minute	6030.0	8187.2	1280.0	471.0%	640.0%
10-minute	-2364.0	7915.0	3255.0	-72.6%	243.0%
20-minute	-3812.0	3358.0	3601.0	-106.0%	93.3%

[&]:Least square; [#]: Post-Baseline; [1]: Lower Bound; [2]: Upper Bound; [3]: Lower Bound/bPS Least Square Mean; [4]: Upper Bound/bPS Least Square Mean.

For sHS versus bPS, Table 3.3.1.1 indicates the following results:

- The lower and upper bounds for the difference in the mean change of sHS versus bPS (sHS – bPS) on gastrin concentration from baseline to 5-minute post baseline are estimated to be –104.0% and 22.0% of the bPS mean, respectively.
- The lower and upper bounds for the difference in the mean change of sHS versus bPS (sHS – bPS) on gastrin concentration from baseline to 10-minute post baseline are estimated to be –162.0% and 73.1% of the bPS mean, respectively.
- The lower and upper bounds for the difference in the mean change of sHS versus bPS (sHS – bPS) on gastrin concentration from baseline to 20-minute post baseline are estimated to be –111.0% and 37.5% of the bPS mean, respectively.

For sPS versus bPS, Table 3.3.1.1 indicates the following results:

- The lower and upper bounds for the difference in the mean change of sPS versus bPS (sPS – bPS) on gastrin concentration from baseline to 5-minute post baseline are estimated to be 471.0% and 640.0% of the bPS mean, respectively.
- The lower and upper bounds for the difference in the mean change of sPS versus bPS (sPS – bPS) on gastrin concentration from baseline to 10-minute post baseline are estimated to be –72.6% and 243.0% of the bPS mean, respectively;
- The lower and upper bounds for the difference in the mean change of sPS versus bPS (sPS – bPS) on gastrin concentration from baseline to 20-minute post baseline are estimated to be -106.0% and 93.3% of the bPS mean, respectively;

ii) Graphic display

The following graphs demonstrate the individual sick patient's changes of gastrin concentrations from baseline to each of 5, 10, 15, and 20 minutes post baseline (post-baseline – baseline) for sHS, sPS, and bPS. Data of six sick patients from Study CRC99-8 are used for sHS while data of six sick patients from Study CRC99-8 and two sick patients from Study CRC97-2 are used for sPS and bPS. One patient in Study CRC97-2, who was already recovered from gastrinoma, is not included in the graphic display.

The diagrams for the four post baselines, 5, 10, 15, and 20 minutes, are given in Fig. 3.3.2.1,

Fig. 3.3.2.2, Fig. 3.3.2.3, and Fig. 3.3.2.4, respectively. In each plot, symbols 'H', 'S', and 'B' denote gastrin changes from baseline to post-baseline for treatments, sHS, sPS, and bPS , respectively. Moreover, patients 7 and 8 in each plot are from Study CRC97-2.

Fig. 3.3.2.1.

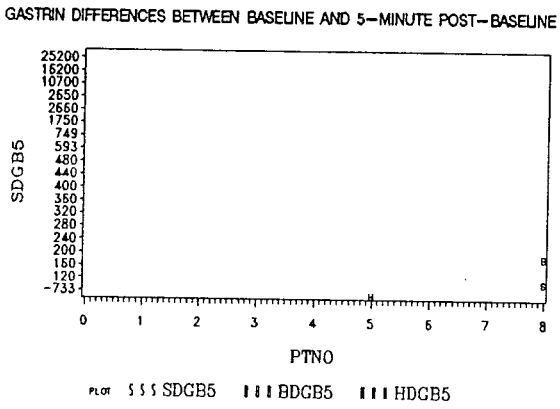

Fig. 3.3.2.2.

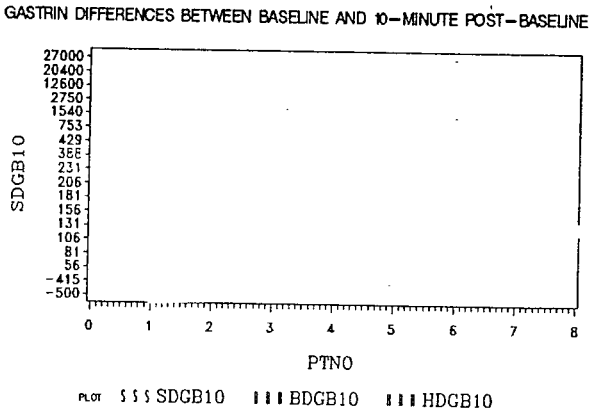

Fig. 3.3.2.3

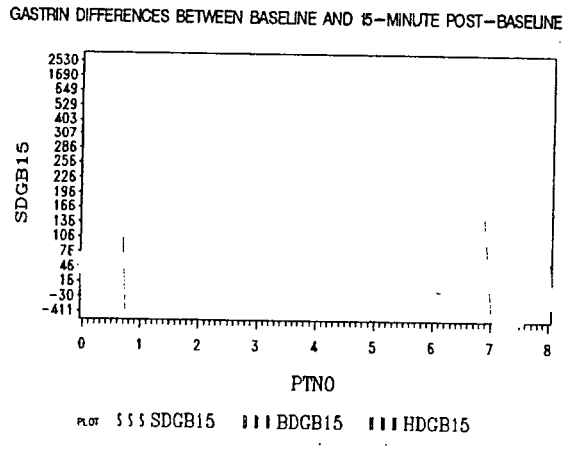

Fig. 3.3.2.4

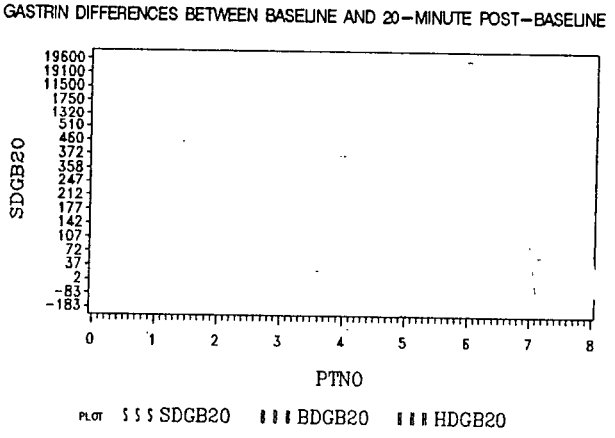

Figure 3.3.2.1, 3.3.2.2, 3.3.2.3, and 3.3.2.4 indicate the following phenomena:

- For each of 5, 10, 15, and 20 minutes post baseline, the within patient variations on the three gastrin changes (sPS, sHs, and bPS) from the baseline are in general large, especially for patients 1, 2, 5, and 6. For instance, for patient 2, the three gastrin changes from baseline to 10 minute post-baseline are –500, 1540, and 2750 pg/mL for bPS, sPS, and sHS, respectively.
- For each of 5, 10, 15, and 20 minute post-baselines, the between patient variations on each of the three gastrin changes from the baseline for sPS, sHS, and bPS, are large as well, especially for patient 6, the gastrin changes from baseline were much higher than those of other patients. For instance, the three gastrin changes from baseline to 10 minute post-baseline for patient 6 are 12600, 20400, and 27000 for sHS, bPS, and sPS, respectively, which are in general at least 50 times greater than those of other patients.

Following the graphic phenomena, it is noted that the variations of the serum gastrin concentration changes from baseline to the four post baselines (5, 10, 15, and 20 minutes) are large for both within and between patients. In addition, as indicated by the sponsor, increase in serum gastrin concentrations from baseline greater than 110 pg/mL is the diagnosis of gastrinoma and gastrin determinations were at 1, 2, 5, 10, 15, 20, and 30 minutes post secretin injection. Based on the above diagnostic criteria, the sponsor claimed that there was 100% diagnostic agreement among sPS, sHS and bPS in 6 out of 6 patients, from Study CRC99-8, in terms of producing test results positive for gatrinoma in patients with a tissue diagnosis of gastrinoma. However, the four figures show that the gastrin changes, for the three drugs (sPS, sHS, and bPS) of each patient, from baseline

to 5, 10, 15, and 20 minute post-baselines for the six patients were not all greater than 110 pg/mL. For instance, for patient 5, the gastrin changes of sHS from baseline to 5, 10, 15, and 20 minute post- baselines were –733, -415, -411, -183, respectively, which were much smaller than 110 and perhaps should not be identified as positive for the diagnosis of gastrinoma. Therefore, the diagnostic efficacy result claimed by the sponsor that 100% of diagnostic agreement among sPS, sHS, and bPS in terms of producing test results positive for gastrinoma in patients with a tissue diagnosis of gastrinoma is not reliable.

Based on the above observations, the following comments are made:

- Due to larg variations for both within and between patients on the serum gastrin concentrations collected for sPS, sHS, and bPS and small sample size, the sponsor's claim that sPS, sHS and bPS produced equivalent pharmacological responses in terms of serum gastrin in gastrinoma patients is not supported. As noted by this reviewer, the statistical equivalence on the serum gastrin concentrations for sHS, sPS, and bPS, purported by the sponsor, was based upon the non-superiority results and not by the clinical equivalence analysis. Therefore, following the above arguments (huge variations and small sample size), the non-significant test on the equality of two agents is very likely due to lake of power.

- Due to not all three gastrin changes (sPS, sHS, and bPS) of each patient from baseline to 5, 10, 15, and 20 minutes post baselines greater than 110 pg/mL, the diagnostic efficacy result claimed by the sponsor that 100% of diagnostic agreement among sPS, sHS, and bPS in terms of producing test results positive for gastrinoma in patients with a tissue diagnosis of gastrinoma is not supported.

iii) Probability analysis

In order to assess the robustness of the 100% agreements reported by the applicant for the diagnosis of gastrinoma for sPS versus bPS and sHS versus bPS, this reviewer performs the following analyses under the assumption that there exits a certain disagreement probability between the two agents:

- For sHS versus bPS, calculates the probability of 100% agreement between sHS and bPS in the diagnosis of gastrinoma, using 6 patients for Study CRC99-8.
- For sPS versus bPS, calculates the probability of 100% agreement between sPS and bPS in the diagnosis of gastrinoma, using 9 patients (6 for Study CRC99-8 and 3 for Study CRC97-2).

sHS versus bPS

Table 3.3.3.1 presents probabilities of 100% agreement between sHS and bPS in the diagnosis of gastrinoma for the six patients, under the assumption that there exists a certain diagnostic disagreement probability between these two agents.

Table 3.3.3.1 (Reviewer's) Probability analysis on the 100% diagnostic agreement

DISAGREEMNENT PROBABILITY FOR SHS VERSUS BPS	PROBABILITY OF 100% AGREEMENT FOR 6 PATIENTS
10%	53%
15%	38%
20%	26%
25%	18%

Table 3.3.3.1 indicates that if 10% of subjects are expected to differ on the two tests (sHS and bPS), the probability of no disagreement in the 6 subjects is 53%. For another example, if the disagreement probability for the two diagnostic agents is 25%, the probability of the 100% agreement in the diagnostic results tested by these two agents on the 6 sick patients is 18%. Table 3.3.3.1 suggests that due to small number of patients tested by the two agents, sHS and bPS, the observed lack of difference can not rule out a relatively large disagreement rate.

sPS versus bPS

Table 3.3.3.2 presents probabilities of 100% agreement between sPS and bPS in the diagnosis of gastrinoma for the nine patients, under the assumption that there exists a certain diagnostic disagreement probability between these two agents.

Table 3.3.3.2 (Reviewer's) Probability analysis on the 100% diagnostic agreement

DISAGREEMNENT PROBABILITY FOR SPS VERSUS BPS	PROBABILITY OF 100% AGREEMENT FOR 9 PATIENTS
10%	39%
15%	23%
20%	13%
25%	8%

Table 3.3.3.2 indicates that if 10% of subjects are expected to differ on the tests, the probability of no disagreement in 9 subjects is 39%. For another example, if the disagreement probability for the two diagnostic agents is 25%, the probability of the 100% agreement in the diagnostic results tested by these two agents on the nine patients is 8%. Table 3.3.3.2 suggest that due to small number of patients tested by the two agents, sPS and bPS, the observed lack of difference can not rule out a relatively large disagreement rate.

3.4 Comments/Conclusions on treatment effects

- The lower bounds of the 95% confidence intervals for the difference in the mean change of sHS versus bPS (sHS – bPS) on gastrin concentration from baseline to 5, 10, and 20 minutes post baselines were –104.0%, -162.0%, and –110.0%, of the bPS means, respectively. These results show that the sHS mean changes from baseline to 5, 10, and 20 minutes are less than those of bPS by up to 104%, 162%, and 110%, respectively, indicating that agent sHS intends to generate lower mean changes of serum gastrin concentration from baseline than those of bPS. It follows that the efficacy result that two agents, sHS and bPS, produced equivalent pharmacological responses in terms of serum gastrin concentration in gastrinoma patients claimed by the sponsor based on the non-superiority results is not reliable, highly likely due to lack of power to detect the difference.
- The upper bounds of the 95% confidence intervals for the difference in the mean change of sPS versus bPS (sPS – bPS) on serum gastrin concentration from baseline to 5, and 10 minutes post baselines were 640.0% and 243.0% of the bPS means, respectively. These results show that the sPS mean changes from baseline to 5 and 10 minutes are greater than those of bPS by up to 640.0% and 243.0% respectively, indicating that agent sPS intends to generate higher mean changes of serum gastrin concentration from baseline than those of bPS. The efficacy result that two agents, sPS and bPS, produced equivalent pharmacological responses in terms of serum gastrin in gastrinoma patients claimed by the sponsor based on the non-superiority results is not reliable, highly likely due to lack of power to detect the difference.
- Graphic displays suggest:

i.) As noted by this reviewer, the statistical equivalence on the serum gastrin concentrations for sHS, sPS, and bPS, purported by the sponsor, was based upon the non-superiority results and not by the clinical equivalence analysis. Due to large variations for both within and between patients on the serum gastrin concentrations collected for sPS, sHS, and bPS and small sample size, the sponsor's claim that sPS, sHS and bPS produced equivalent pharmacological responses in terms of serum gastrin in gastrinoma patients is not supported. Following the above arguments (huge variations and small sample size), the non-significant test on the equality of two agents is highly likely due to lake of power.

i.) Due to not all three gastrin changes (sPS, sHS, and bPS) of each patient from baseline to 5, 10, 15, and 20 minutes post baselines greater than 110 pg/mL, the diagnostic efficacy result claimed by the sponsor that 100% of diagnostic agreement among sPS, sHS, and bPS in terms of producing test results positive for gastrinoma in patients with a tissue diagnosis of gastrinoma is not reliable.

- Under the assumption of 25% disagreement probabilities for sPS versus bPS and sHs versus bPS, the probability of 100% agreement in the diagnostic results were 8% for sPS versus bPs tested by nine patients and 18% for sHS versus bPS tested by six patients. Therefore, due to small number of patients tested by the two agents, sHs versus bPS or sPS and bPS, the observed lack of difference can not rule out a relatively large disagreement rate.
- In conclusion, the results of efficacy analyses from studies Study CRC99-8 and Study CRC97-2 are not statistically persuasive to support the use of sHS and sPS on the Diagnosis of gastrinoma.

4.1 Background information for Study CRC98-4 (Amendment) used for facilitation of pancreatic and bile duct cannulation during ERCP

Objectives: The objective of this Phase III study was to obtain efficacy and safety data for sHS to facilitate cannulation of the minor duct in patients with pancreas divisum during ERCP procedure.

Study Design: This is a randomized, double-blind, crossover multi-center study for the use of sHS (0.2 μg/kg) to facilitate minor pancreatic duct cannulation in patients with pancreas divisum during ERCP procedure.

Twenty-seven patients were enrolled into this study and evaluated with an untreated control. Of theses 27 patients, 3 were successfully cannulated and were not studied further. Of the 24 subjects for whom cannulation was not successful, the orifice was visualized without treatment for 8 patients. The orifice was not visualized for the remaining 16 patients. Each of these 24 patients was then randomized to receive either placebo or secretin. If the cannulation was not successful on the first treatment, the opposite treatment was then used.

Study Population: 1. Patients suspected diagnosis of pancreas divisum during ERCP. 2. Males and females of non-childbearing potential without active acute pancreatitis or known sensitivity to secretin and not pregnant or nursing.

Diagnostic Efficacy and Statistical Methods: Comparative efficacy for minor pancreatic duct localization and successful cannulation during ERCP procedure. McNemar's test was to be applied to analyze the cannulation outcomes among three treatment groups: sHS, untreated control, and placebo.

Number of Subjects: Twenty patients (20) were planned based on agreements with FDA. Twenty-seven (27) patients (all patient population) were enrolled and 24 patients were randomized (randomized patient population) to receive either sHS or placebo as first treatment.

4.2 Sponsor's Statistical Analysis and Results

Demographics and Baseline Characteristics

The sponsor did not present data or results of the statistical analyses on demographic variables and baseline characteristics.

Analysis of Diagnostic Efficacy Results

Table 4.2.1 presents the results of the comparisons on the successful rates for sHS versus untreated control and sHS versus placebo using all patient population (ALLP), randomized population (RANP), and all available data population (AADP). The sponsor indicated that all available data

population consisted of data observed for each patient. However, for the other two populations, the unobserved (missing) data were replaced by "unsuccessful" outcome.

Table 4.2.1 (Sponsor's) Summary of facilitated cannulation success rates of outcomes

	UNT.CONTRL[1] (U)	PLACEBO (P)	SECRETIN (S)	P- VALUE U VS. S	P VS. S
AADP	11.1% (3/27)	13.3% (2/15)	66.7% (16/24)	< 0.0001*	0.18
ALLP	11.1% (3/27)	7.4% (2/27)	59.3% (16/27)	0.004*	0.001*
RANP	0.0% (0/24)	8.3% (2/24)	66.7% (16/24)	< 0.0001*	0.001*

[1]: Untreated Control Group. *: Significance at 0.05 significance level by McNemar's test.
Note: Missing Data in populations ITTP and RANP were replaced with unsuccessful outcome.

Table 4.2.1 showed that except for the placebo versus sHS using AADP population, the cannulation successful rate of sHS was significantly greater than those of untreated control and placebo groups.

Based on the results of Table 4.2.1, the sponsor concluded that synthetic human secretin (sHS) is an effective agent ı

Adverse Events

The sponsor reported that no adverse events were observed in this study.

4.3 Reviewer's Analyses and Comments

In order to authenticate the efficacy claim made by the sponsor, this reviewer comments on i.) issue of the study design and ii.) issue of the statistical analysis.

i.) Issue of the study design

As indicated by the study design, all enrolled 27 patients were first evaluated with an untreated control. Then, those 24 patients (8 patients with visualized orifices) with cannulation failed using untreated control were randomized to either placebo or sHS. However, from the sponsor's table on "listing of individual data" in page 789 of Volume 29, one notes that of the 8 patients with visualized orifices, only two patients (25%) were assigned to placebo. This randomization process results in 50% (6 out of 12 patients) of patients in sHS group with visualized orifices versus only 17% (2 out of 12 patients) of those patients in placebo group. It follows that the patients may not be properly randomized to one of the two treatment (placebo or sHS) groups.
In addition, this reviewer also notes that of the 27 enrolled patients, 11 patients (41%) had visualized orifices when assessed without treatment. However, the twelve (12) patients randomized to receive placebo as first treatment had no visualized orifices, including those two patients with visualized orifices when evaluated by untreated control. Consequently, all twelve patients were assessed failure in cannulation procedure. Apparently, even though patients were randomized, rather than a double blind design, it was, in reality, an open label study since the investigator could

identify the treatment (placebo or sHS) for each patient treated by the volume of pancreatic juice produced. Therefore, the assessment on the success of cannulation may be biased toward in favor of the tested drug sHS. For the impact of the biased assessment on this small sample size study, refer to the sensitivity analysis presented in the next subsection ii.).

ii.) Issue of the statistical analysis

Firstly, as noted from the sponsor's table on "listing of individual data" in page 789 of Volume 29, nine (9) out of twelve (12) patients randomized to receive treatment sHS first were successful in cannulation and were not continued to perform ERCP procedure using placebo, based on the assessment rule set up by the study design. Accordingly, data for the 9 patients assessed by placebo as the second treatment were missed because of ERCP procedure not being performed with placebo treatment. In order to avoid the biases induced by missing data replaced with "unsuccessful" outcome, the result for the comparison on the success rates for sHS versus placebo using McNemar's test (proposed by the sponsor in the protocol) should be based on AADP population. From the sponsor's analysis results, the success rate of treatment sHS is not significantly different from that of placebo (p=0.18).

Secondarily, in order to assess the impact of the biased assessment toward in favor of sHS commented in subsection i.), this reviewer performs the sensitivity analysis by Fisher exact test using data from the first period of the cross-over design only. Following the sponsor's table on "listing of individual data" in page 789 of Volume 29, one notes that of the twelve patients randomized to receive sHS as first treatment, 9 patients succeeded in the cannulation procedure while for placebo, non of the 12 patients succeeded. Table 4.3.2.1 presents the results of the sensitivity analysis to evaluate the impact of increasing the number of patient success for the cannulation procedure in the placebo group.

Table 4.3.2.1 (Reviewer's) Assessment on the impact of patient success in placebo group

SHS		PLACEBO		FISHER EXACT TEST
FAILUR	SUCCESS	FAILURE	SUCCESS	P- VALUE
3	9	12	0	0.0003
3	9	11	1	0.0028
3	9	10	2	0.012
3	9	9	3	0.04
3	9	8	4	0.10

Table 4.3.2.1 shows that due to small sample size, Fisher exact p-value of one success (p=0.0028) in the placebo group is 10 times to that of none success (p=0.0003). Since only one study was conducted to assess the efficacy of sHS in the use of facilitation of cannulation, instead of using 0.05 significance level, a much smaller significance level (for example 0.005) is recommended to compare the treatment effects. The paper entitled "A comment on replication, P-value and evidence" written by Steven N. Goodman, published by Statistical in Medicine, Vol. 11, 875-879 (1992), indicates that if the result from one study trial shows significant treatment effect under much

smaller significance level, it provides stronger evidence (higher power) that the second study, if conducted, will be also shown significant treatment effect under α-significance level of 0.05. Thus, for this one pivotal study, if 0.005 significance level is used for the comparisons between sHS and placebo, then only two patients in the placebo group misidentified as failure will result in a non-significance difference for the efficacy comparison between sHS and placebo in cannulation procedure. As a result, due to one study with small sample size, the impact of the biased assessment toward in favor of the tested drug sHS can not be ignored.

Finally, by the sponsor's study design, patients assigned to treatment sHS were patients after being assessed failure by the untreated control. It follows that the p-value ($p<0.0001$) for McNemar's test in the comparison of sHS versus untreated control was calculated only using patients assessed failure by the untreated control. Therefore, the study design planed by the sponsor is not appropriate to compare the efficacy on the facilitation of cannulation for sHS versus untreated control.

4.4 Overall Conclusions for Study CRC98-4

- Due to 50% (6 out of 12 patients) of patients in sHS group with visualized orifices identified during the untreated control phase versus only 17% (2 out of 12 patients) of those patients in placebo group, patients for this small study may not have been properly randomized to treatment groups (placebo or sHS).
- Rather than a double blind design, this was an open label study since the investigator could identify the treatment (placebo or sHS) for each patient received by the volume of pancreatic juice produced. Therefore, the assessment on the success of cannulation may be biased toward in favor of the tested drug sHS.
- In order to avoid the biases induced by missing data replaced with unsuccessful outcomes and by the statistical methods proposed by the sponsor in page 746 of Volume 29, the result for the comparison on the success rates of cannulation for sHS versus placebo using McNemar's test should be based on the AADP population. From the sponsor's McNemar's test result, the success rate of treatment sHS is not significantly different from that of placebo ($p=0.18$).
- In addition, given only the single study with small sample size, the impact of the possibly biased assessment in favor of the tested drug sHS can not be ignored.
- It is noted that p-value ($p<0.0001$) for McNemar's test in the comparison of sHS versus untreated control was calculated only using patients assessed failure by the untreated control. Thus, the study design planed by the sponsor is not appropriate to compare the efficacy on the facilitation of cannulation for sHS versus untreated control.

In conclusion: Due to issues with respect of the study design and statistical analysis, the use of synthetic human secretin on facilitation [redacted] during ERCP is not supported by the data based on the one study CRC98-4 submitted by the sponsor.

5.0 Overall Conclusions of treatment effects

- The results of efficacy analyses from the single pivotal study StudyCRC99-9 fail to provide adequate statistical evidence to support the use of synthetic human secretin on the diagnosis of pancreatic exocrine —
- The results of efficacy analyses from studies Study CRC99-8 and Study CRC97-2 fail to provide adequate statistical evidence to support the use of synthetic human secretin and synthetic porcine secretin on the diagnosis of gastrinoma.
- Due to issues with respective to the study design and statistical analysis, the use of synthetic human secretin on facilitation [redacted] during ERCP is not supported by the data based on the single study CRC98-4.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

Wen-Jen Chen
11/19/01 07:44:17 PM
BIOMETRICS

Mike Welch
11/20/01 11:28:33 AM
BIOMETRICS
Concur with review. Also see secondary review.

Secondary Statistical Review

NDA 21-256

Drug: Synthetic Human Secretin
Sponsor: ChiRhoClin
Indications: (1) Diagnosis of Pancreatic Exocrine
(2) Diagnosis of Gastrinoma
(3) Facilitation
Reference: (1) Primary statistical review by Wen-Jen Chen, PhD (HFD-715)
(2) Primary medical review (draft) by M.A.Barreiro, MD (HFD-180)
Reviewer: Mike Welch, PhD (HFD-715)

Introduction

The primary statistical review concludes that the sponsor's studies fail to provide adequate statistical evidence in support of the intended indications. I basically concur with that result. However, it can be argued that the diagnostic trials were fundamentally flawed from a clinical design perspective and thus unable to achieve their objectives, irrespective of small study size. This review provides some comments in this regard. Some general principles governing diagnostic clinical trials are given in the appendix.

Pancreatic Exocrine Disease Studies

Study CRC 99-9

For this single study, the intended indication for secretin stimulation testing (SST) using Synthetic Human Secretin (sHS) is diagnosis of pancreatic exocrine Six patients with documented diagnosis of Chronic Pancreatitis (CP) were randomized in a crossover study to received three stimulation tests using the new agent, sHS; an unapproved comparator agent, Synthetic Porcine Secretin (sPS); and an approved comparator agent, Biologic Porcine Secretin (bPS). The principal outcomes, based on SST, are pancreatic secretion volume and biocarbonate concentrations at specific timepoints; the latter is intended to classify the disease as present if its level is below the threshold value of 80 mEq/L. All three tests gave positive outcomes in all six subjects. The implication is that such a result is sufficient to support the diagnostic claim.

Study CRC 98-2

In this study, twelve patients, disease confirmed by previous SST with bPS, were randomized to either sPS or sHS in a crossover fashion. Outcome measures are the same as described above,and both tests gave positive results for all patients. However, the historic use of bPS is not considered a valid comparator by the Division, and since sPS is not approved, the trial is not considered supportive.

Diagnostic Issues

A measure of simple concordance or agreement between the two tests (sHS and bPS) is not a valid measure of diagnostic performance, since both tests may agree but both be in error, and if they disagree, the correct result is unknown. However, assuming that bPS results constitute a truth standard may not be appropriate. Although extensive, successful clinical use of a test may warrant its use in a trial as a truth standard, it is not clear if that is the case here. That is, interpreting concordance in these studies as a measure of sHS test sensitivity would need be supportable by documented clinical use. Even conceding that the patients really have the disease in question, one can only conclude that sHS test sensitivity is higher than 60% (lower bound of one-sided, 95% confidence interval for six successes in six trials.)[1]

However, the failure of these studies as diagnostic trials results from the evidently inappropriate patient sample. The small sample size is obviously an issue here as the reliability of any outcome in only six patients is subject to question. However, just as important in test validation, is the issue of whether the patients themselves represent the correct patient population. Since they were previously confirmed by the comparator procedure; they are not representative. Since no disease-negative cases were studied, the sample is further unrepresentative. However, disregarding patients' prior work-up procedures; it is not clear they had the appropriate range of symptomotology for which the test will be used. Thus, even interpreting concordance as sensitivity would be clearly misleading, as the test (or any diagnostic test) would most likely gives falsely optimistic indication of its potential, given advanced disease cases.

Given an appropriate group of patients (e.g., patients with and without clear symptoms who are suspected but not yet confirmed of CP) and given a suitable standard of truth method, a trial can give reliable estimates for both sensitivity and specificity of the new procedure, and these values can form the basis for other measures of clinical utility that are desirable in a product label. However, the precise role of the test, anticipated in clinical practice, needs to be clearly described in the submission and should be part of the planned indication. For example, the use of SST as a diagnostic tool to rule out disease; to rule in disease; to send patients for other testing; to use as a stand-alone procedure; or to use as an adjunctive test – needs to be clear in study planning as the relative importance of the diagnostic endpoints would depend on how the test is to be employed. Moreover, the risk potential for a diagnostic test cannot be known unless the misclassification errors of the test are adequately estimated – this implies estimating both the false positive and false negative rates of the test. In these studies, neither the patient population or the precise role in clinical use has been made clear in order to justify the study design and results.

Granted, the studies attempt to compare two, maybe very similar, SST procedures that may differ mainly in their use of the type of secretin. Thus, one might argue from a chemical or pharmacological point of view that the bicarbonate outcomes will be similar enough to yield the same diagnosis. Yet a correct 'bioequivalent' type approach was not employed in these studies, and such a conclusion cannot be based on the data from these studies. Even at face value, the data show large variability. The primary reviewer's analysis of the SST outcomes indicates, for example, that the mean sHS

[1] If all 18 subjects in both studies were deemed true positives, then the lower bound for sHS test sensitivity would be about 86%

outcome can be 21% greater than that for bPS. This difference is enough to change a positive bPS diagnosis, based on a value of about 66 mEq/L, to a negative diagnosis based on sHS. This also brings into question the role of the current threshold values in the clinical setting and its validity in using a new form of secretin, as well as the question that has been unanswered in these studies: which test is correct?

Gastrinoma Studies

Study CRC 99-8

This single center crossover study randomized six patients to receive sPS, sHS, and bPS. According to the medical officer's review, only five of the six patients were reported to have documented tissue diagnosis for gastrinoma. A positive SST diagnosis of gastrinoma was based on a minimum of 110 pg/ml increase from baseline in serum gastrin concentration at any measured time point after secretin administration. All six subjects demonstrated increases in serum gastrin exceeding this value, although, two patients (excluding the one without a tissue diagnosis) were not tested in accordance with the protocol-specified design and may have had inadequate wash-out duration, according to the medical reviewer's comments. (Another study, CRC 97-2 applied sPS and bPS to three subjects, and consequently provides no information regarding sHS).

Comment

This reviewer's above concerns regarding diagnostic validity of the pancreatic disease studies apply equally well for this indication. The sample size of only six subjects, (three, if one excludes protocol violations) again clearly precludes any reliable judgement of test performance. As with the other studies, sHS test validity has not been adequately shown in these trials to support a diagnostic indication.

Summary

From a clinical and statistical design perspective, the diagnostic efficacy of SST using sHS has not been demonstrated for either diagnostic indication. The trials employ a non-representative patient population of so few subjects that both bias and lack of precision in results are clearly evident.

If only pharmacological studies are needed to support sHS as a replacement for bPS, then the studies need to apply the correct design and statistical methods to show the relevant parameters are 'bioequivalent' according to prospectively defined criteria. However, to validate the use of sHS in the detection of the indicated disease conditions, good clinical trials are necessary.

The design of such trials would depend on the anticipated clinical population and management decisions that could result from possible test outcomes; the latter clinical decisions lead to the necessary test performance endpoints that appropriately reflect diagnostic risk. Such studies would not have to be large, but in addition to the above, sample size would depend on the prevalence of the disease in the studied population.

APPENDIX

Some Principles for Diagnostic Trials

Generally, in a well-controlled clinical study of a new diagnostic agent or test, diagnostic performance is measured by the proportion of test-positive cases who actually have the disease (sensitivity) as well as the proportion of test-negative cases who do not have the disease (specificity). In such a study, the new test procedure is performed concurrently with a comparison test where the latter is often defined by the standard of care. The diagnostic outcomes of both test procedures is based on the presence or absence of disease determined by an acceptable truth standard procedure. The different test methods (new test and comparison test) would be applied in a random fashion and blindly interpreted to avoid bias.

An independent truth panel can be used to provide a suitable final clinical diagnosis or truth standard, in the absence of more definitive results (e.g. biopsy) but the panel should be blinded to outcomes from study modalities; patient follow-up may be required in test-negative cases to determine true course of disease. All subjects, regardless of test outcome should be evaluated by a truth procedure, and principles of intention to treat (diagnose) should be applied. The objectives of the study needs to be formulated as statistical hypotheses with appropriate significance level and power to show superiority (or non-inferiority) of the new diagnostic procedures, and an appropriate statistical analysis plan should be developed.

Study subjects are representative of the population of patients in whom the test procedure will be used; otherwise the trial outcomes and associated measures of clinical utility cannot be extrapolated to the intended clinical setting. A test used in a patient group consisting of normal subjects and subjects with advanced disease will give upwardly biased estimates of test performance (sensitivity and specificity) but may fail to be a useful diagnostic tool in clinical practice. Using Chronic Pancreatitis (CP) as an example, an appropriate patient mix might include certain symptomatic and asymptomatic patients for whom a diagnosis of CP needs to be established.

In the event there is no well-defined standard of care to represent a comparison test procedure, then the study outcomes would consist of only the diagnostic measures of the new test. These could be compared to clinically relevant thresholds pre-specified in the protocol. However, such studies are viewed as historically controlled and may be problematic. The study, as proposed, implies that the diagnostic measures are to be compared to some pre-defined values.

In some clinical trial settings, it may be recognized that the true nature of the disease is not attainable, and only the results from the new test and a comparison procedure can be measured. If the comparison test is 'sufficiently valid' it may assume the role of a truth standard, and measures of sensitivity and specificity of the new procedure can be obtained and interpreted *with respect to the comparator test*. One should recognized, however, that misclassification errors associated with this surrogate truth standard induce biases in the diagnostic measures being estimated in the study.

Another study approach in this regard is to acknowledge that the new test and comparator test are simply being compared, without acknowledgement of the true disease state. Such a study might be called an 'agreement' trial, and the endpoint of interest would be concordance of test results. However, percent agreement or concordance is not a measure of diagnostic validity. For example, two tests may agree but both be wrong; or agreement may be high while either test sensitivity or specificity is unacceptably low. Percent agreement is often referred to as *accuracy;* which is a misnomer.

Clinical utility of a new diagnostic test can often be better measured in terms of the ability of a test to predictive disease status; for example, the probability that a patient has the disease given a positive test is termed positive predictive value and is a function of test sensitivity and specificity as well as prevalence of the disease in the clinically appropriate population. Similarly, negative predictive values are useful for clinical use. Other measures of clinical utility include the likelihood ratios which (for a positive test result) reflects how much one's prior odds of disease is increased after seeing a positive test result. These are useful measures, however, only if the study produces unbiased and precise estimates of test sensitivity and specificity.

An important property of any new diagnostic test is the decision criteria, based on test result, that is used to classify the disease. Scoring a diagnostic test outcome may, for example, involve assigning a rating according to some scale established for *level of disease suspicion.* This procedure is commonly used, for example, in the evaluation of medical images. When the diagnostic test yields, say a particular laboratory value, that is measured on some numerical scale, the study needs to establish an appropriate cut-off value which partitions the set of possible values into disease and not-disease for classification. Whether the test result is categorical or continuous, this can be accomplished by computing pairs of sensitivity and specificity values, one pair for each possible cut-off value. The set of all such pairs form a receiver operating characteristic (ROC) curve, and the modeling and estimation of such curves, though complicated statistically and computationally, are important tools for validating diagnostic tests.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

Mike Welch
11/26/01 03:29:56 PM
BIOMETRICS