looking at all 10 sensitivity and specificity pairs for each of the 10 Thyrogen Tg cut-offs assuming the Withdrawal reference was 100% sensitive and specific. Table 8.2 presents the sensitivity and specificity values from these calculations. From the table, the sponsor then concluded that Thyrogen Tg levels of 1, 2, and 3 ng/mL were "equivalent" to Withdrawal Tg levels of 2, 5, and 10 ng/mL, respectively. These Tg levels cut-offs were used in the *post hoc* diagnostic utility analyses (see Section 8.2.2.2).

Table 8.2
Study TSH95-0101
Receiver Operator Curve Analysis for
Successfully Ablated, Tg Antibody Negative Patients:
Arms I and II Combined (n=141)

Withdrawal Tg Level (ng/mL)	Thyrogen Tg Level (ng/mL)	Sensitivity (%)	Specificity (%)
10	1	87	69
	2	78	84
	3	72	95
	4	69	96
	5 - 10	67 - 53	96 - 98
5	1	87	77
	2	78	92
	3 - 10	71 - 51	100 - 100
2	1	87	92
	2	75	100
	3 - 10	65 - 47	100 - 100

Source: Table 4M, Volume 1.47, page 47.

8.2.2.2 Diagnostic Utility Analyses

Various diagnostic utility analyses were performed for each treatment arm. The analyses presented the prevalence, sensitivity, specificity, positive and negative predictive values, and accuracy of the various testing modalities compared to several reference standards. Table 8.3 presents the *post hoc* diagnostic utility analyses. The sponsor used the same Tg cut-off level for THST as the Thyrogen Tg cut-off level for these analyses.

Table 8.3
Post Hoc Diagnostic Utility Analysis (For the Detection of Remnant, Recurrent Cancer, or Metastatic Cancer)

Reference Standard for Detecting Thyroid Cancer or Remnant	Diagnostic Test for Detecting Thyroid Cancer or Remnant
Withdrawal Tg ≥ 10 ng/mL **OR** a Withdrawal Diagnostic or Post Therapy Scan Class ≥ 1	1) THST Tg level ≥ 3 ng/mL 2) Thyrogen Tg level ≥ 3 ng/mL 3) Thyrogen Diagnostic WBS ≥ 1 4) Thyrogen Diagnostic WBS ≥ 1 or Thyrogen Tg level ≥ 3 ng/mL
Withdrawal Tg ≥ 5 ng/mL **OR** a Withdrawal Diagnostic or Post Therapy Scan Class ≥ 1	1) THST Tg level ≥ 2 ng/mL 2) Thyrogen Tg level ≥ 2 ng/mL 3) Thyrogen Diagnostic WBS ≥ 1 4) Thyrogen Diagnostic WBS ≥ 1 or Thyrogen Tg level ≥ 2 ng/mL
Withdrawal Tg ≥ 2 ng/mL **OR** a Withdrawal Diagnostic or Post Therapy Scan Class ≥ 1	1) THST Tg level ≥ 1 ng/mL 2) Thyrogen Tg level ≥ 1 ng/mL 3) Thyrogen Diagnostic WBS ≥ 1 4) Thyrogen Diagnostic WBS ≥ 1 or Thyrogen Tg level ≥ 1 ng/mL

Source: February 13, 1998 submission, #009, Volume 1 of 2, Tab 3; and modifications by Statistical Reviewer.

All diagnostic utility analyses pertaining to Tg testing were limited to patients who were successfully ablated and Tg antibody negative. The Thyrogen stimulated Tg levels drawn on the day with the highest proportion of patients were above the cut-off for detecting thyroid cancer, i.e. ≥10 ng/mL, was assessed. All diagnostic utility analyses were conducted to include both high and low risk patients, as determined by the TNM (Tumor-Node-Metastasis) classification system, and to study these groups separately.

8.2.2.3 Additional Diagnostic Utility Analysis

In the March 16, 1998 submission, the sponsor changed the indication and labeling, and withdrew the proposed label claim that Thyrogen Tg levels of 1, 2, and 3 ng/mL corresponded to Withdrawal Tg levels of 2, 5, and 10 ng/mL, respectively.

The current labeling proposed the use of the combination of a Thyrogen Tg test and diagnostic whole body scan, where a Thyrogen Tg level ≥2 ng/mL or a Thyrogen diagnostic whole body scan of class ≥ 1 was indicative of thyroid remnant or cancer in follow-up patients. The reference standard of a Withdrawal Tg level of ≥2 ng/mL or a Withdrawal diagnostic or post-therapy scan of class ≥ 1 was defined as being indicative of thyroid remnant or cancer in follow-up patients. The only justification given for using a Tg level of ≥2 ng/mL (Withdrawal or Thyrogen Tg level) as indicative of thyroid remnant or cancer was that it " was above the variability of the assay and strongly suggests the presence of either thyroid remnant or cancer" (Section 3, page 6, March 16, 1998 submission).

9.0 OTHER STUDY INFORMATION

9.1 Principal Investigators and Blinded Readers

Table 9.1 presents the principal investigators that participated in Studies TSH92-0601 and TSH95-0101. There were 11 principal investigators in Study TSH92-0601 and 14 principal investigators in Study TSH95-0101. These were *a priori* defined centers for each study.

Table 9.1
Principal Investigators for Studies TSH92-0601 and TSH95-0101

Study	Investigator (M.D.)	Site	Location
TSH92-0601	Lewis E. Braverman	Univ. of Massachusetts Medical Center	Worcester, MA
	David S. Cooper	Sinai Hospital	Baltimore, MD
	Gilbert H. Daniels	Massacusetts General Hospital	Boston, MA
	Terry F. Davies	Mt. Sinai Medical Center	New York, NY
	Leslie J. DeGroot	Univ. of Chicago Medical Center	Chicago, IL
	Jeffrey R. Garber	Beth Israel Hospital	Boston, MA
	Ian D. Hay	Mayo Clinic	Rochester, MN
	Paul W. Ladenson	Johns Hopkins Univ. School of Medicine	Baltimore, MD
	Silvina Levis	Univ. of Miami School of Medicine	Miami, FL
	Ernest Mazzaferri	Ohio State University	Columbus, OH
	Bruce D. Weintraub	National Institutes of Health	Bethesda, MD
TSH95-0101	Lewis E. Braverman	Univ. of Massacusetts Medical Center	Worcester, MA
	David S. Cooper	Sinai Hospital	Baltimore, MD
	Gilbert H. Daniels	Massacusetts General Hospital	Boston, MA
	Terry F. Davies	Mt. Sinai Medical Center	New York, NY
	Leslie J. DeGroot	Univ. of Chicago Medical Center	Chicago, IL
	Paul W. Ladenson	Johns Hopkins Univ. School of Medicine	Baltimore, MD
	Ernest Mazzaferri	Ohio State University	Columbus, OH
	Furio Pacini	University of Pisa	Pisa, Italy
	Christoph Reiners		Würzburg, Germany
	Chester Ridgway	Univ. Of Colorado Health Sciences Center	Denver, CO
	Mary Samuels	Oregon Health Sciences University	Portland, OR
	Martin Schlumberger	Institut Gustave Roissy	Villejuif, France
	Steven Sherman	MD Anderson Cancer Center	Houston, TX
	Monica Skarulis	National Institutes of Health	Bethesda, MD

Table 9.2 lists the blinded readers for both studies. The same three blinded readers evaluated the images for both studies.

Table 9.2
Blinded Readers for Studies TSH92-0601 and TSH95-0101

Blinded Readers	Ralph R. Cavalieri, M.D.	V.A. Medical Center	San Francisco, CA
	David V. Becker, M.D.	New York Hospital-Cornell Medical Center	New York, NY
	Harry R. Maxon III, M.D.	Univ. Of Cincinnati Medical Center	Cincinnati, OH

9.2 Patient Enrollment, Accountability, and Evaluability

Tables 9.3 and 9.4 display the disposition of the enrolled patients and the number of patients evaluable for the efficacy and safety analyses in Studies TSH92-0601 and TSH95-0101.

Study TSH92-0601 enrolled a total of 152 patients at 11 centers in the United States and lasted from September 9, 1992 to February 2, 1994. Of the 152 patients enrolled, 144 patients completed the

study.

Table 9.3
Study TSH92-0601:
Patient Accountability and Evaluable for Efficacy

Disposition	No. of Patients
Number Enrolled	152
Number Treated with Thyrogen and Evaluable for Safety	152
Number Completed Study	144
Number Who Did Not Complete Study	8
Reason for Discontinuation	
Withdrew from study	1
Thyrogen Study Period Adverse Event	4
Other	3
Number in Intent-to-Treat Population for Efficacy	138

Source: Table 3, Volume 1.52, page 50 and Table 1C, Volume 1.47, page 22.

Study TSH95-0101 enrolled a total of 254 patients at 14 centers, 11 in the United States and 3 in Europe, and lasted from November 8, 1995 to April 4, 1997. There were 123 patients randomized to receive 2 injections of Thyrogen (Arm I) and 131 patients randomized to receive 3 injections of Thyrogen (Arm II).

Table 9.4
Study TSH95-0101:
Patient Accountability and Evaluable for Efficacy

Disposition	No. of Patients Arm I	No. of Patients Arm II
Number Enrolled	123	131
Number Treated with Thyrogen and Evaluable for Safety	117	112
Number Not Treated with Thyrogen	6	19
Reason for Discontinuation		
Withdrew from study	4	11
Non-compliant with study	1	6
Did not meet inclusion criteria	1	2
Number Completed Study	116	110
Number Who Did Not Complete Study	1	2
Reason for Discontinuation		
Withdrew from study	0	1
Non-compliant with study	0	1
Disease progression required other therapy	1	0
Number of Successfully Ablated and Tg Antibody Negative Patients: Evaluable for Efficacy	82	88

Source: Tables 5A, 5B and 5C, Volume 1.56, pages 70-71.

10.0 SPONSOR'S EFFICACY RESULTS

The primary and secondary efficacy results for Studies TSH92-0601 and TSH95-0101 are presented in this section.

10.1 Comparison of Scan Cancer Classification Within Treatment Arm

Table 10.1 presents the results for the primary outcome of the within patient scan evaluation by the blinded readers for both studies.

Table 10.1
Studies TSH92-0601 and TSH95-0101 Primary Efficacy Results of Within Patient Equivalence of Scan Classification: ITT Population Evaluated by Blinded Reader Majority Rule

Scan Classification	TSH92-0601	TSH95-0101 Arm I	TSH95-0101 Arm II
Thyrogen Scan ≥ Withdrawal Scan [95% Confidence Interval]	119/138 (86.2%) [79.1 - 91.3%]	104/113 (92.0%) [85 - 96%]	99/107 (92.5%) [85.4 - 96.5%]
Withdrawal Scan ≥ Thyrogen Scan [95% Confidence Interval]	135/138 (97.8%) [93.3 - 99.4%]	110/113 (97.3%) [91.9 - 99.3%]	102/107 (95.3%) [88.9 - 98.3%]
Concordance	116 (84.1%)	101 (89.4%)	94 (87.9%)
Discordance	22 (15.9%)	12 (10.7%)	13 (12.2%)
Favoring Thyrogen	3 (2.2%)	3 (2.7%)	5 (4.7 %)
Favoring Withdrawal	19 (13.8%)	9 (8.0%)	8 (7.5%)
p-value*	<0.001	0.146	0.581

Source: Table 4A, Volume 1.47, page 36.
* Based on two-tailed sign test.

The results for the primary efficacy outcome of Thyrogen scans versus Withdrawal scans classification of cancer are as follow. Of the discordant scan pairs in Study TSH92-0601, the Withdrawal scan was favored over the Thyrogen scan for classifying a higher category of cancer ($p<0.001$, based on a two-tailed sign test). In both arms of Study TSH95-0101, there was no difference in the number of discordant scan pairs between Thyrogen and Withdrawal. That is, when discordant, the number of Thyrogen and Withdrawal scans that showed a higher classification cancer category were not different (Arm I: $p=0.146$, Arm II: $p=0.581$).

10.2 Comparison of Scan Cancer Classification Between Treatment Arms

Table 10.2 presents the results for the primary outcome of the between arm comparison using the blinded reader data for Study TSH95-0101.

Table 10.2
Study TSH95-0101
Between Arm Comparison of Discordant Scan Pairs

Scan Classification	Arm I	Arm II	p-value**
Concordance	101 (89.4%)*	94 (87.9%)	
Favoring Thyrogen	3 (2.7%)	5 (4.7 %)	0.758
Favoring Withdrawal	9 (8.0%)	8 (7.5%)	

Source: Table 5Y, Volume 1.56, page 91.
* Percent of the total number in each arm.
** Based on two-tailed Fisher's Exact test to assess the difference in percent favoring Thyrogen between Arm I and Arm II.

There was no significant difference between the study arms in their ability to detect thyroid cancer or remnants by diagnostic radioiodine imaging with either Thyrogen or Withdrawal scans (p=0.758).

10.3 Diagnostic Utility Analyses

Table 10.3 presents sensitivity and specificity values for the *post hoc* diagnostic utility analyses of three tests compared to three Withdrawal reference standards by study arm. This was a secondary efficacy analysis for Study TSH95-0101 (see Table 8.3 for more detail).

Table 10.3
Study TSH95-0101
Sensitivities and Specificities for Three Tests using Three Reference Standards
Reference Standard: a Withdrawal Tg Level ≥ 10, 5, or 2 ng/mL OR a Withdrawal Diagnostic or Post-Therapy Scan ≥ 1

Test Tg Level and Reference Standard Tg Level Test ng/mL / Reference ng/mL	THST Tg Sensitivity, Specificity (%)		Thyrogen 72 hour Tg Sensitivity, Specificity (%)		Thyrogen 72 hour Tg and 48 hour Thyrogen Scan ≥ 1 Sensitivity, Specificity (%)	
	Arm I	Arm II	Arm I	Arm II	Arm I	Arm II
3 ng/mL / 10 ng/mL	38, 100	46, 92	74, 100	72, 89	94, 93	97, 81
2 ng/mL / 5 ng/mL	40, 96	50, 95	77, 93	79, 91	94, 89	95, 81
1 ng/mL / 2 ng/mL	57, 76	67, 94	79, 91	93, 94	91, 86	99, 83

Source: Tables 4O, 4P, and 4Q, Volume 1.47, pages 51-53.

In general, the sensitivities for Arm II were better than Arm I within each test. Within an arm, the sensitivity for the Thyrogen 72 hour Tg test was better than that for the THST Tg test, and the sensitivity for the combination of the Thyrogen 72 hour Tg and 48 hour scan test was better than that for the Thyrogen 72 hour Tg test.

Finding comparable Thyrogen and Withdrawal Tg levels (see Section 8.2.2.1) was the sponsor analysis presented in support of "equivalence" in Tg levels for the two methods. The sponsor did not further discuss the results for the per protocol utility analyses (see Section 8.2.1.3) to show equivalence of the two methods. The only time the per protocol utility analyses were mentioned was to defend why the diagnostic utility analyses were modified. The sponsor believed that this analysis was "a more accurate and useful evaluation of THST Tg testing, Thyrogen Tg testing, and Thyrogen WBS and Tg testing as compared to a Withdrawal WBS and Tg test" (FAX dated 2-10-98, page 6).

10.4 Additional Diagnostic Utility Analyses

The sponsor presented sensitivities for detecting thyroid remnant or cancer using a Thyrogen Tg level ≥ 2 ng/mL or a Thyrogen diagnostic scan of class ≥ 1 for each arm of the study. The reference standard was a Withdrawal Tg level ≥ 2 ng/mL OR a Withdrawal diagnostic or post-therapy scan of class ≥ 1. The sensitivities were 88% (50/57) for Arm I and 93% (63/68) for Arm II (see March 16, 1998 submission, pages 13, 14, and 59).

10.5 Hypothyroid Symptoms

In both studies there was a significant difference in symptoms of hypothyroidism, favoring Thyrogen. Patients who remained euthyroid after Thyrogen experienced significantly less

hypothyroid symptoms than during Withdrawal ($p<0.05$ with the Billewicz scale in both studies). This was a primary outcome in Study TSH95-0101 and a secondary outcome in Study TSH92-0601.

10.6 Quality of Life

In both studies there was a significant difference in the quality of life, favoring Thyrogen. Patients who remained euthroid after Thyrogen experienced a significantly better quality of life than during Withdrawal ($p<0.05$ with the SF-36 scale in Study TSH95-0101 and $p<0.0001$ with the short form POMS scale in Study TSH92-0601 only). This was a secondary outcome in both studies.

10.7 Safety Results

There was one serious and severe adverse event noted in this submission per the Medical Reviewer. There was one death in Study TSH92-0601; a patient died from pulmonary embolism and recurrent cancer that infiltrated the trachea. This death was not related to drug. There was one adverse event of concern in Study TSH95-0101; one patient had TSH stimulation induced swelling of metastases in the sternum. For further safety information, refer to the Medical Reviewer's safety review.

10.8 Sponsor's Overall Conclusions

The sponsor claimed that these studies have demonstrated the safety and efficacy of Thyrogen for the proposed indication and that "equivalence" of the clinical utility of Thyrogen and Withdrawal for the diagnosis of thyroid remnants and thyroid cancer was shown in Study TSH95-0101. Their justification was based on the results for the concordance between the Thyrogen and Withdrawal scans and the ROC curve analysis for Tg levels. This "equivalence" was further "verified" by the diagnostic utility analyses which used a test of results the combination of a Thyrogen Tg test and scan compared to the combination of a Withdrawal Tg test and scan. In addition, patients who remained euthyroid after Thyrogen experienced significantly less hypothyroid symptoms and a significantly better quality of life than during Withdrawal.

11.0 REVIEWER'S COMMENTS ON STUDY DESIGN AND ANALYSES

1. The Withdrawal and Thyrogen scans were presented to the blinded reader in a paired fashion, that is, both scans were individually evaluated, one after the other and the readers knew that the scan pair belonged to the same patient. This way of evaluating scans leads to recall or carry-over bias.

For Study TSH95-0101, evidence of bias in favor of concordant scans can be noted from the following four sources.

i. The first was found within the data set of the evaluation of the scan pairs, Volume 1.62, Patient Data Listing 11.1.0, pages 1-78. The blinded readers used comments that indicated that they either remembered what was present on the first scan evaluated or had the first scan available for comparison. These comments included the following: Patient 201, Reader MXN, "Exactly as above plus ?uptake in a 3rd lesion right neck - could be saliva in esophagus"; and Patient 516, Reader MXN, "Same as above".

ii. The second source was from the result of the scan pair analysis for Study TSH92-0601. This analysis resulted in Withdrawal scans being favored over Thyrogen scans for classifying a higher category of cancer.

iii. The third source comes from using the same blinded readers in both studies. Since the same three blinded readers were used, they knew that one intent of the study was to show the "equivalence" of the Thyrogen and Withdrawal scans. In Study TSH92-0601, a CRF question asked to rate the two scans in a side-by-side comparison as being "equivalent" or not. This may have resulted in the large number of scan pairs being rated equivalent that would not have been as rated in a completely randomized read.

iv. The fourth source of bias came from the post-ablation or post-therapy scans and the most recent pre-study scan that were provided to the readers, when available, for their evaluation of the study scans. Since it was not known where these scans were when the study scans were being evaluated, there is no guarantee that the readers did not look at them when evaluating the study scans.

The introduction of bias through paired readings of the Withdrawal and Thyrogen scans compromises the results of any analysis based on image outcomes.

2. The results of the primary endpoint for scan concordance/discordance are not valid from a clinical perspective. Thyrogen is indicated as an alternative to Withdrawal for radio nuclide imaging and/or thyroglobulin testing for the diagnosis of thyroid remnants and metastatic cancer. Since there was no independent evaluation of the Thyrogen scans, there is no information about the diagnostic merits of the Thyrogen scans alone.

3. Studies TSH92-0601 and TSH95-0101 are not independent. The same three blinded readers and 7 of the same Principal Investigators were used in both studies (see Tables 9.1 and 9.2).

4. The primary analysis for both Studies TSH92-0601 and TSH95-0101 should have been based on a primary endpoint of the number of discordant scan pairs for Thyrogen and Withdrawal. Per the Medical Reviewer, the clinically relevant scan pair results are the discordant scan pairs, not the number of scan pairs where the Thyrogen scan was rated greater than or equal to the Withdrawal scan. Using the sponsor's proposed primary endpoint, bias may also have been introduced in Study TSH95-0101 through the opportunity of having less scan pairs being evaluated as having Withdrawal being superior to Thyrogen.

5. The sample size for Study TSH95-0101 should have been based on the number of discordant scan pairs. A McNemar test based sample size estimate should have been used to adequately power the study for demonstrating a statistical difference in the number of discordant scan pairs between Thyrogen and Withdrawal. The resulting small number of discordant pairs may have been a reason why the sign test for the discordant scan pairs between Thyrogen and Withdrawal was not statistically significant in Study TSH95-0101.

6. The use of additional information for the panel mediation resulted in conclusions that were not based solely on the information used in the initial blinded evaluation of the scans. In these cases, image outcomes were confounded with information from other modalities.

7. The scans for the blinded reads were pre-selected as being adequate for evaluation by the study monitor without any standardized criteria specified. This way of selecting images for blinded reads may introduce bias when the criteria are not available for evaluation.

8. The per protocol indication for Thyrogen was for the diagnosis of metastatic cancer in follow-up thyroid cancer patients. After performing the per protocol diagnostic utility analyses for Study TSH95-0101, the sponsor discovered that there was "an unacceptable number of false positive diagnoses of metastatic cancer when comparing Withdrawal scans and Tg levels to the reference standard of a post-therapy scan" (FAX from sponsor dated 2-10-98). The sponsor then decided to change the indication to using Thyrogen for the diagnosis of thyroid remnants and well-differentiated thyroid cancer (local and metastatic).

9. The first change in the indication for Thyrogen led to a number of "best fit" *post hoc* diagnostic utility analyses to justify the equivalence of the clinical utility of Thyrogen to Withdrawal. Initially, the reference standard for the diagnostic utility analyses was a Withdrawal Tg cut-off of ≥ 10 ng/mL or a Withdrawal diagnostic or therapeutic scan ≥2 to diagnose the presence of metastatic cancer. Per the Medical Reviewer, a Withdrawal Tg ≥ 10 ng/mL is the demarcation line for metastatic disease. The indication was later changed to the diagnosis of thyroid remnants and thyroid cancer but there was no clear Tg cut-off value that could be used.

Since the sponsor intended for Thyrogen treatment to replace Withdrawal treatment, the cut-offs for the Tg level were prospectively planned to be the same for both treatments. That is, a one-on-one comparison of the 2, 5, and 10 ng/mL prospectively defined cut-offs for both treatments was to have been evaluated and reported. Instead, a "best fit" diagnostic utility analysis was performed to identify true positives by lowering the threshold for the Thyrogen Tg cut-off levels. Thus it was reported that Thyrogen Tg levels of 1, 2, and 3 ng/mL were comparable to Withdrawal Tg levels of 2, 5, and 10 ng/mL, respectively. Per the Medical Reviewer, Thyrogen was less sensitive than Withdrawal in stimulating Tg production and this was the intrinsic problem with the product; any cut-off used is immaterial to the clinical utility of Thyrogen.

10. The chosen reference standards were not adequate to address the sponsor's analyses of finding comparable Tg levels between Thyrogen and Withdrawal. Depending on which reference standard was used, more weight was given to the classification of a patient having the disease from either the result of a scan (diagnostic or therapeutic) or to the result of a specific Tg level cut-off. For example, per protocol reference standard 2 was defined as:

> A Withdrawal post therapy scan class ≥ 2 **or** a negative post therapy scan with Withdrawal Tg ≥ 10 ng/mL **or** if no post therapy scan, a Withdrawal Tg ≥ 10 ng/mL and a decision by the physician to treat the patient (e.g. ^{131}I therapy, surgical dissection, external radiation, etc.).

Using this to define Thyrogen Tg levels comparable to Withdrawal Tg levels was not adequate because those patients who had a positive Withdrawal post therapy scan did not contribute information about their Withdrawal Tg level to the analysis and thus biased the results. This attempt to find the "best fit" of the data to the Withdrawal cut-off for Tg resulted in using no information from the Withdrawal Tg value by using the scan result. Subsequent *post hoc* analyses used similar reference standards and are also biased. For example, *post hoc* reference standard 1 was defined as:

> A Withdrawal Tg ≥ 10 ng/mL or a Withdrawal Diagnostic or Post Therapy Scan Class ≥ 1.

This *post hoc* reference standard, in particular, has another limitation that further biases the results. Even if a patient had a Withdrawal Tg ≥10 ng/mL, there was no information provided that the Tg

level was the basis used to classify the patient if their scan was rated as greater than or equal to 1.

11. The sponsor assumed that the Withdrawal reference standard had a sensitivity, specificity, and accuracy of 100%. The sponsor stated in Section 3 on page 5 of the March 16, 1998 submission that "the withdrawal diagnostic ^{131}I scan is not a perfect gold standard, as a false negative rate of up to 35% has been reported." Such a false negative rate in the reference standard can induce large bias in test result accuracy.

12. The reference standard used the Withdrawal scan classification assigned by the blinded readers. Since the blinded reader scan results are biased, the reference standard is also biased. Data for the reference standard should have been evaluated by an independent source that did not participate in the conduct of the study or evaluate any study images for efficacy results.

13. The sponsor combined the data from both arms to perform the ROC curve analyses to find the "optimal" cut-offs that were then used to calculate diagnostic utility analyses for each arm. Their rationale was, "to have the largest sample size" (Volume 1.56, Section 5.4.4, page 96). This rationale is not appropriate when each arm was to be analyzed separately to determine which dosing regimen was best to use.

14. Due to the many sources of uncorrectable bias this reviewer feels that no reanalysis of the study outcomes is warranted.

12.0 REVIEWER'S CONCLUSIONS

The sponsor claimed that Studies TSH92-0601 and TSH95-0101 have demonstrated the safety and efficacy of Thyrogen. While there were no serious safety problems, efficacy has not been shown in these studies. Bias in the scan evaluation and the many *post hoc* analyses did not constitute adequate analyses to demonstrate that Thyrogen therapy was equivalent to Withdrawal therapy to warrant Thyrogen acting as an alternative therapy to Withdrawal.

13.0 RECOMMENDATION

From a statistical standpoint, the sponsor has not provided adequate and well controlled studies that show substantial evidence for efficacy in support of their intended claims.

/S/ 4-20-98

Sonia Castillo, Ph.D.
Mathematical Statistician, HFD-720

Concur:

/S/ 4/20/98

Michael Welch, Ph.D.
Acting Division Director, HFD-720

cc:
Archival NDA 20-898
HFD-510/S. Sobel/D. Orloff/J. Temeck/S. McCort/Div. File
HFD-344/T. Ju
HFD-720/File Copy/M. Welch/S. Castillo
HFD-715/File Copy/Chron./E. Nevius

S. Castillo/x73085/WordPerfect/4/20/98
This review contains 36 pages of text, tables, and figures.

APR 22 1998

ADDENDUM TO STATISTICAL REVIEW

NDA 20-898
Thyrogen® (thyrotropin alfa)
Genzyme Corporation
Document reviewed: Letter submitted by the sponsor dated 4-17-98.
Medical Reviewer: Jean Temeck, M.D., HFD-510
Statistical Reviewer: Sonia Castillo, Ph.D., HFD-720

The sponsor submitted a re-analysis in support of the relationship between Thyrogen and Withdrawal Tg levels in successfully ablated and Tg antibody negative patients. A linear regression analysis of the $\log_{10}$ transformed Tg data was presented for each arm of the study. Table 1 presents the estimated regression line and correlation coefficient for each study arm.

Table 1

Arm	Regression Line	R^2
I	$\log_{10}$(Thyrogen Tg at 72 hrs.)= -0.1612 + 0.9053* $\log_{10}$(Withdrawal Tg)	0.880
II	$\log_{10}$(Thyrogen Tg at 72 hrs.)= -0.1797 + 0.9631* $\log_{10}$(Withdrawal Tg)	0.943

The sponsor concluded that on the log scale, the plot of Withdrawal Tg versus Thyrogen Tg demonstrated a linear relationship. That is, "the Tg value after Thyrogen increased consistently (albeit lower) with the increase in Withdrawal values and was highly correlated." The sponsor also concluded that there was an approximate 1:2 relationship between the Thyrogen Tg level to the Withdrawal Tg level. This was arrived at by using the best fit based on the log scale data, and then transforming the fitted values back to the original scale. The sponsor concluded that "when a patient has a Thyrogen Tg of 5 ng/mL, this would correspond to a withdrawal Tg of approximately 10 ng/mL."

This analysis is both misrepresentative and not consistent with previous sponsor analyses supporting the comparability of Thyrogen and Withdrawal Tg values. The raw Tg level data is highly variable and a log scale transform of the raw data greatly reduces the inherent variability of the data. By minimizing the amount of data variability, the data is better suited to linear regression analysis because it now better conforms to the assumption of constant variance. Concluding that the best fit results of the log scale data is applicable to the raw scale, as the sponsor has concluded, is erroneous. The transformed data may give higher R^2 values but the underlying inherent variability still remains when the untransformed data is reported for clinical use.

The conclusion of this log scale linear regression is not consistent with the previous analyses in support of the comparability of Tg values, that is the ROC curve analyses (See Statistical Review Section 8.2.2.1). The sponsor concluded that a Thyrogen Tg of 3 ng/mL was comparable to a Withdrawal Tg of 10 ng/mL. This is not the same as the log scale analysis above where a Thyrogen Tg of 5 ng/mL was comparable to a Withdrawal Tg of 10 ng/mL. The sponsor has thus shown

inconsistent results in evaluating association between Withdrawal Tg and Thyrogen Tg.

In addition, it makes better clinical sense to predict what the Withdrawal Tg level would have been for a given Thyrogen Tg level. The sponsor analysis presented the predicted Thyrogen Tg level for a given Withdrawal Tg level.

In conclusion, this analysis of the log scale Tg data has not provided evidence in support of the comparability of Withdrawal and Thyrogen Tg values in patients who are successfully ablated and Tg antibody negative.

/S/ 4-22-98

Sonia Castillo, Ph.D.
Mathematical Statistician, HFD-720

Concur:

/S/ 4/22/98

Michael Welch, Ph.D.
Acting Division Director, HFD-720

cc:
Archival NDA 20-898
HFD-510/S. Sobel/D. Orloff/J. Temeck/S. McCort/Div. File
HFD-344/T. Ju
HFD-720/File Copy/M. Welch/S. Castillo
HFD-715/File Copy/Chron./E. Nevius

S. Castillo/x73085/WordPerfect/4/20/98
This review contains 2 pages.