Comment on tables 1-6:

Although there was not a statistically significant difference in the discordance rate between the Thyrogen and WD diagnostic scans, the numerical trend consistently favored the WD-scan. Likewise, although there was not a statistically significant difference between the two dosing regimens in the ability to detect remnants/cancer by diagnostic imaging, there appeared to be a numerical trend favoring arm II(3 dose regimen).

PRIMARY EFFICACY:

Hypothyroid signs and symptoms by Billewicz scale:

As expected, hypothyroid signs and symptoms occurred with significantly greater frequency during the WD phase than during the Thyrogen phase ($p < 0.05$).

SECONDARY EFFICACY:

As expected, patients' quality of life as measured by the SF-36, was maintained during the Thyrogen phase but was significantly reduced ($p < 0.05$) following withdrawal.

Secondary Efficacy:

^{131}I Uptake:

48h thyroid bed uptake was measured by thyroid probe or region of interest (ROI) at the time of Thyrogen and WD diagnostic scanning. In patients with positive scans in arm I, there was no difference in the 48h thyroidal uptake of the Thyrogen and WD scans when measured by either technique (thyroid probe: 27 patients in arm I and ROI: 24 patients in arm I). However, in arm II scan positive patients, then 48 h WD scan thyroidal uptake as measured by the thyroid probe, was significantly greater ($p < 0.05$) than the Thyrogen scan; there was no statistically significant difference in uptake when measured by the ROI technique (arm II: ROI: 27 patients). However, interpretation of these data is limited because of the high proportion of patients who had very minimal thyroid bed uptake.

Thyroid uptake kinetics following Thyrogen and WD scanning was measured in 11 patients (6 in arm I and 5 in arm II) at the NIH. This evaluation is limited in that only 5/11 patients had uptake in the thyroid bed on either scan. Serial ROI measurements of thyroid bed uptake were acquired at 2, 48 and 72 hr after ^{131}I administration. The results were:

Mean/median % ^{131}I thyroid bed uptake was significantly greater ($p = 0.01$) during the WD phase (since mean and median values were similar, only the median values will be specified: Thyrogen phase: median: 0.027% [range: 0.012-0.048%] vs. WD phase: 0.060% [range: 0.028-0.114%])

Mean/median thyroid uptake half-life ($T_{1/2}$) was significantly

longer (p= 0.01) during the WD phase (median- Thyrogen phase: 12.1 hrs. [range: 7.8-15.6 hrs.] vs. WD: 15.2 hrs.[range: 12.0-28.2 hrs.].

Mean/median cumulated activity was significantly greater (p=-0.05) during the WD phase (median- Thyrogen phase: 6.23 [range: 1.25-17.3] vs. WD 9.91 [range: 3.42-19.7].

Whole body retention of ^{131}I:

Was measured at the time of the 48h thyroid bed uptake measurement during both the Thyrogen phase and the WD phase at 3 sites in the TSH95-0101 study (16 patients in arm I and 14 patients in arm II). The median 48 hr % ^{131}I whole body retention was significantly greater (p= 0.01) during the WD phase compared to the Thyrogen phase (median 48h % ^{131}I whole body retention was ~2x greater during the WD phase in arm I compared to the Thyrogen phase and ~2.5x greater during the WD phase in arm II).

Kinetics of ^{131}I whole body retention were measured at the NIH, in 11 and 12 patients, respectively, combining both dosing regimens. Serial measurements of whole body retention were acquired at 2, 48 and 72 hr. after ^{131}I administration following Thyrogen and withdrawal. Results were:

48h whole body retention was ~2 fold significantly greater (p= 0.01) during the WD phase compared to the Thyrogen phase,

whole body retention half-life ($T_{1/2}$) of ^{131}I was significantly longer (p= 0.01) during the WD phase by a factor of 0.38 and

whole body cumulative activity (AUC) was higher for ^{131}I during WD by a factor of 0.41.

In summary:

Both the ^{131}I thyroid bed uptake and whole body retention studies demonstrate that, compared to the Thyrogen phase, there is, in the withdrawal phase, significantly greater:
^{131}I thyroid bed uptake, half-life and cumulative activity;
^{131}I whole body retention, half-life and cumulative activity

Additional Analysis:
Comparison of the Duration of Elevated TSH levels between Thyrogen and Withdrawal:

14 days after Thyrogen administration, mean and median endogenous TSH levels did return to suppressed levels. In contrast, 2-4 weeks after withdrawal of THST, TSH levels were documented to be on the rise at the time of ^{131}I administration for the WD scan.

My comments regarding the clinical significance of the differences between withdrawal and Thyrogen pertaining to ^{131}I uptake and whole body retention kinetic studies and the duration of time TSH is elevated as they relate to the scan images:

When the Thyrogen and WD scans were discordant, the proportion of scans that were a higher classification consistently favored WD over Thyrogen. This is probably due to 2 factors:

1. the duration of time TSH is elevated
2. differences in ^{131}I kinetics between the euthyroid (Thyrogen phase) and the hypothyroid (WD phase) state.

Schlumberger demonstrated (JCEM 57(1):148-151, 1983) that radioiodine uptake is related to the rise in TSH and the length of time TSH is elevated. The longer the duration of time TSH is elevated, the higher the ^{131}I uptake. Therefore, scanning performed 7-19 days after T3 withdrawal, was more sensitive in detecting thyroid cancer metastases than scanning after 3 days of bTSH administration. He states: "In conclusion, T3 withdrawal gives higher uptake than bTSH alone and should be preferred for scanning and treating metastases in patients with thyroid carcinoma."

The higher ^{131}I uptake during withdrawal compared to Thyrogen is related to the prolonged period of TSH stimulation and to the slower metabolism and decreased renal clearance of ^{131}I, intrinsic to the hypothyroid state. Consequently, at a given diagnostic activity of ^{131}I, the withdrawal scan is more sensitive than the Thyrogen scan in the ability to detect thyroid remnants and/or cancer.

SECONDARY EFFICACY:

THE KINETIC PROFILE OF THE Tg RESPONSE TO THYROGEN:

This analysis was done to determine the optimal time to measure Tg after Thyrogen had been administered. 3 cohorts of patients were examined: recently thyroidectomized but pre-^{131}I ablation and follow-up patients (surgically and ^{131}I ablated): baseline Tg <10 ng/ml (i.e. Tg on THST <10 ng/ml) and >10 ng/ml. These data demonstrated:

a. Mean Thyrogen Tg values peaked at 72h post the final injection in arm I and at 48-72h in arm II

b. **Mean Tg levels obtained 2-6 weeks after THST withdrawal were consistently higher than after 5 to 9 days of TSH stimulation provided by Thyrogen.**

The sponsor made the comment that compared to arm I, in arm II, there was less variation in the individual patient Thyrogen Tg response at 1, 2 & 3 days after the final injection.

The following tables depict the mean/median TSH and Tg levels on Thyrogen vs. WD in successfully ablated, Tg antibody negative, recently thyroidectomized vs. follow/up patients with baseline Tg <10 ng/ml vs. follow-up patients with baseline Tg <

10 ng/ml:

Arm I: Mean (SD) and Median (Range) TSH (mU/L) and Tg Response (ng/ml):

Patients	Peak Thyrogen TSH mean (SD) median (range)	Peak WD TSH mean (SD) median (range)	48h Thyrogen Tg mean (SD) median (range)	72h Thyrogen Tg mean (SD) median (range)	WD Tg mean (SD) median (range)
Recent n= 10	132.6(46.8) 134.7 (75-207)	59.8 (18.4) 62.5 (36-90)	12.9 (35.2) 1.0 (0.5-113)	12.9 (32.7) 1.1 (0.5-100)	20.8 (56.4) 1.4 (0.8-171)
F/U THST Tg <10 (n=56)	118.3(61.8) 104.4 (53-400)	74.9 (47.3) 64 (20-294)	9.2 (24.1) 1.3 (0.5-166)	8.9 (19.2) 1.2 (0.5-119)	21.7 (37.4) 3.2 (0.5-179)
F/U THST Tg >10 (n= 13)	129.9(95.8) 80.0 (75-369)	56.0 (26.6) 55.5 (25-106)	1095 (1888) 219 (23-5513)	1238 (2053) 258 (21-5993)	2823 (6173) 431 (24-21955)

Arm II: Mean (SD) and Median (Range) Tg Response (ng/ml):

Patients	Peak Thyrogen TSH mean (SD) median (range)	Peak WD TSH mean (SD) median (range)	48h Thyrogen Tg mean (SD) median (range)	72h Thyrogen Tg mean (SD) median (range)	WD Tg mean (SD) median (range)
Recent n= 12	110.0(43.2) 99.8 (64-218)	63.7 (34.8) 57.0 (22-125)	91.2 (260) 3.4 (0.5-870)	75.2 (221) 3.6 (0.5-773)	74.5 (226) 7.8 (0.5-790)
F/U THST Tg <10 (n= 49)	108.5(47.4) 86.0 (50-221)	65.2 (32.6) 60.2 (21-193)	9.2 (26.8) 1.2 (0.5-171)	9.1 (26.9) 1.2 (0.5-177)	22.6 (51.3) 1.8 (0.5-249)
F/U THST Tg >10 (n= 21)	80.3 (29.7) 75 (26-184)	65.0 (34.5) 64 (11-164)	20,444 (33731) 2011 (21-146758)	28,300 (48,470) 2225 (20-169107)	29,454 (46,861) 6692 (54-174701)

Comment on the above tables:

Although the mean and median TSH levels were higher on Thyrogen than on WD, the mean and median Tg levels were higher on WD in both treatment arms. (Note, the medians are a more accurate measure than the means here, due to the wide variation in the TSH and Tg levels).

At the end of this review are enclosed the sponsor's figures of the mean serum TSH levels vs. days after Thyrogen administration for arms I and II. These figures indicate that the duration of time the mean serum TSH level remains ≥20 mU/L is for ~ only 5 days after Thyrogen administration in arm I and for ~ only 3 days in arm II.

It has been established that the elevation of Tg is a function of the rise in TSH as well as the duration of time TSH is elevated. The mean/median Tg levels on WD are higher than those on Thyrogen due to the prolonged period of TSH stimulation following THST WD compared to only an acute rise in TSH on Thyrogen.

THE DIAGNOSTIC UTILITY OF THYROGEN Tg ALONE AND COMBINED WITH THE THYROGEN SCAN TO DETECT METASTATIC DISEASE USING AS THE REFERENCE STANDARD A POST-THERAPY SCAC CLASS > 2 OR A LYMPH NODE BIOPSY POSITIVE FOR CANCER POST STUDY:

Prospectively, the purpose of the diagnostic utility analysis was to compare Tg on Thyrogen to WD to Tg on THST **to detect metastatic cancer** at Tg cut-offs, 2, 5 and 10 ng/ml, using a post-therapy scan class ≥ 2 as the reference standard.

Using this reference standard is problematic for 2 reasons:

1. In the majority of the patients, post-therapy scans were not done. Specifically, in arm I, post-therapy scans were performed in only 30/78 successfully ablated patients who received Thyrogen; and only 9 of these were class ≥ 2 (9/78= 12%). In arm II, post-rx. scans were performed in only 41/86 successfully ablated patients who received Thyrogen; 23 of these were class ≥ 2 (23/86= 27%).

2. In arm I, 56% (19/34) successfully ablated, Tg antibody negative patients had WD Tg ≥ 10 ng/ml but were either not treated (and, therefore, there was no post-rx. scan) or the post-rx. scan was negative. The corresponding # (%) in arm II was 7/40 (18%). By the reference standard, all these patients would be classified as false positives, even though they had a clinically significant Tg elevation.

Despite a negative post-therapy scan, metastatic disease was confirmed by lymph node biopsy post study (performed due to WD Tg > 30 ng/ml) in 3 patients (2 in arm I and 1 in arm II) and by a positive chest CT post study (2 patients in arm II). Therefore, in total, 11 patients in arm I and 26 in arm II had confirmed metastatic disease (by post-rx. scan or post-study positive lymph node bx. or chest CT scan).

Using Tg cut-offs of 1, 2, 3, 5 and 10 ng/ml, the number of patients in whom metastatic disease was missed (false negatives) by a Tg on THST vs. 72 h Thyrogen Tg vs. WD Tg was as follows by treatment arm: (note: the n in each column represents the total number of successfully ablated, Tg antibody negative patients):

False Negative Thyroglobulin Levels: Tg on THST vs. 72h Thyrogen vs. WD in successfully ablated, Tg antibody negative patients with metastatic disease confirmed by post-rx. scan or + lymph node biopsy or + chest CT scan:

	ARM I (n=11 metastatic disease)			ARM II (n=26 metastatic disease)		
Tg cut-offs ng/ml	Tg on THST	72 hr. Thyrogen Tg*	WD Tg	Tg on THST	72 hr. Thyrogen Tg	WD Tg
≥ 1	3	0	0	2	0	0
≥ 2	4	0	0	4	0	0
≥ 3	5	0	0	5	1	0
≥ 5	5	0	0	6	2	0
≥ 10	6	4	1	8	4	0

* 72 hr. value used because that is when Thyrogen Tg level peaked

Comments on the above table:

1. At any given cut-off, in either treatment arm, the false negatives (FNS) were highest on Tg on THST and lowest on WD.

2. Combining the false negatives on WD Tg (since the WD phase was identical in both treatment arms), there was only 1 FN patient at the 10 ng/ml Tg cut-off. (This patient's WD Tg was 9.0 ng/ml).

3. Thyrogen Tg failed to detect confirmed metastatic disease in 4/11 patients (36%) in arm I at the 10 ng/ml cut-off and in 7/26 (27%) patients in arm II at Tg cut-offs as low as 3 ng/ml.

Thyrogen Tg levels vs. WD Tg levels in the 8 patients who were FNS on Thyrogen:

ARM I			ARM II		
Patient ID #	72 h Thyrogen Tg (ng/ml)	Wd Tg (ng/ml)	Patient ID #	72 h Thyrogen Tg (ng/ml)	WD Tg (ng/ml)
# 202			# 311		
# 310			# 1706		
# 1426			# 1228		
# 210			# 1713		

APPEARS THIS WAY
ON ORIGINAL

False Negative Thyrogen Tg plus scan vs. WD Tg plus scan in successfully ablated, Tg antibody negative patients with metastatic disease confirmed by post-rx. scan or + lymph node biopsy or + chest CT scan:

	ARM I(n=11 metastatic disease		ARM II(n=26metastatic disease	
Tg cut-offs ng/ml	72 hr Thyrogen Tg + Thyrogen Scan	WD Tg + WD Scan	72 hr Thyrogen Tg + Thyrogen Scan	WD Tg + WD Scan
≥ 1	0	0	0	0
≥ 2	0	0	0	0
≥ 3	0	0	1	0
≥ 5	0	0	2	0
≥ 10	3	0	4	0

Comments on the above table:

1. The WD scan was positive for metastatic disease in the 1 patient whose WD Tg was 9.0 ng/ml. Therefore, there were no FNS on WD when the WD Tg and scan were combined.

2. Combining both treatment arms, the addition of the Thyrogen scan reduced the # of FNS by only 1 patient. Hence, the Thyrogen scan failed to detect metastatic disease in 7 of these 8 patients. (Note: although, the WD scan was positive for metastatic disease in only 1 of these 7, the WD Tg was > 10 ng/ml in all of these patients).

Prospectively, A SECOND DIAGNOSTIC UTILITY ANALYSIS WAS TO BE DONE COMPARING THYROGEN Tg TO Tg ON THST, TO **DETECT METASTATIC CANCER** USING AS THE REFERENCE STANDARD: POST-THERAPY SCAN CLASS > 2, OR A WD Tg > 10 ng/ml AND A DECISION TO TREAT THE PATIENT.

This reference standard is problematic for the following reasons:

1. Only 12% of patients in arm I and 27% in arm II had post-rx. scans class ≥2

2. A significant number of patients with WD Tg ≥10 ng/ml were not treated, and, therefore, they would be classified as false positives while those who were treated for a WD Tg >10 ng/ml would be classified as true positives even if the post-rx. scan was negative.

Since this analysis had been prospectively defined, the FN results will be presented here at Tg cut-offs of 2, 5 and 10 mg/ml:

False Negative Tg levels: Tg on THST vs. 72h Thyrogen Tg in successfully ablated, Tg antibody negative patients: Reference standard is a post-rx. scan class ≥2 or a WD Tg ≥10 ng/ml and a decision to treat the patient:

	ARM I		ARM II	
Tg cut-offs (ng/ml)	Tg on THST n= 54*	72 h Thyrogen n=54*	Tg on THST n= 55*	72 h Thyrogen n= 57*
≥ 2	3	0	3	0
≥ 5	4	0	4	2
≥ 10	5	3	6	4

*= n based on the prospectively agreed upon definition of a successfully ablated patient: one who had undergone a NT or TT (-/+ subsequent radioiodine ablation) and had < 1% radioiodine thyroidal bed uptake on the most recent scan prior to study entry).

Comment on table: The FNS on Thyrogen are lower than for Tg on THST at all cut-offs.

Combining the Thyrogen scan with the Thyrogen Tg in this analysis, reduced the FNS by only 1 patient.

POST-HOC RATIONALE FOR REDEFINING THE REFERENCE STANDARD AND HOW THE THYROGEN TG -/+ SCAN DIAGNOSTIC UTILITY ANALYSES WERE TO BE CONDUCTED:

The sponsor's position was that the objective of the diagnostic utility analyses of Tg alone and combined with the scan should be the ability to detect thyroid **remnants and cancer.** Therefore, both withdrawal and post-therapy scans class ≥ 1 were included in the reference standard.

The sponsor then made the following additional points:

1. choosing a WD Tg ≥ 10 ng/ml as part of the reference standard, excludes clinically relevant elevations in WD Tg levels < 10 ng/ml as being cancer
2. clinically relevant Tg levels above a reference cut-off but below 10 ng/ml were being considered as false positives

Another problem, was that a WD Tg ≥ 10 ng/ml was being considered a false positive if the patient was not treated.

A fundamental problem, and one not pointed out by the sponsor, was that:

Several patients in this study had Thyrogen Tg levels < 10 ng/ml when the WD Tg was > 10 ng/ml, with scan detectable cancer in some of these. This means that these patients who had localized or metastatic thyroid cancer visualized on scan would fail to be detected by Thyrogen Tg if the same cutoff were used to compare Thyrogen Tg to WD Tg. The sponsor, therefore, made a downward adjustment in the cut-off

used for Thyrogen relative to WD. They used receiver operator curves (ROCs) to determine how much of an adjustment needed to be made to provide the optimal fit (in terms of sensitivity and specificity) of the Thyrogen Tg database in this study to the WD Tg and scan data in this study. The sponsor then used these best fit Thyrogen Tg cut-offs to proceed with the Thyrogen diagnostic utility analyses.

The following data supports my above statements:

In arm I, of 34 successfully ablated, Tg antibody negative patients with WD Tg > 10 ng/ml, 10 (29%) had corresponding Thyrogen Tg levels < 10 ng/ml. 5 (33%) of the 15 patients who had WD Tg > 10 ng/ml and a positive scan (class ≥1), had a corresponding Thyrogen Tg levels < 10 ng/ml.

In arm II, of 40 successfully ablated, Tg antibody negative patients with WD Tg > 10 ng/ml, 9 (23%) had corresponding Thyrogen Tg levels < 10 ng/ml. 8 (24%) of the 33 patients who had WD Tg > 10 ng/ml and a positive scan (class ≥1), had corresponding Thyrogen Tg levels < 10 ng/ml.

In summary, using a 10 ng/ml Tg cut-off to compare Thyrogen Tg to WD Tg would have resulted in Thyrogen Tg missing 13/48 (27%) patients with cancer, either localized or metastatic, detectable on scan, in both treatment arms combined in this study.

Specific Tg data were provided for the 4 patients in each treatment arm who had WD Tg > 10 ng/ml and metastatic disease by scan but Thyrogen Tg was < 10 ng/ml (refer to lower table on page 30 of this review). Below is listed the study data for the remaining 5 patients who had a WD Tg > 10 ng/ml and a class 1 WD scan but the Thyrogen Tg was < 10 ng/ml:

Rx.Arm	Patient ID #	WD Tg (ng/ml)	WD scan class	Post-rx. scan class	72h Thyrogen Tg	Thyrogen scan class
I	# 1517		1	1		0
II	# 103		1	1		1
II	# 208		1	Not rx.[a]		1
II	# 720		1	1		1
II	# 1614		1	1		1

a: comment made by investigator that patient would have been treated had he known the Tg measured centrally was > 10 ng/ml (the local Tg, upon which treatment decisions were based, was 9.2 ng/ml).
Note the negative Thyrogen scan in patient # 1517.

HOW THE POST-HOC REFERENCE STANDARD WAS DEFINED AND ANALYZED:

The reference standard was redefined as follows:
Thyroid Remnant or Cancer was Present: if the study WD Tg was ≥ Y

ng/ml **or** the study scan (either WD or post-therapy) was positive (class ≥ 1);
Thyroid Remnant or Cancer was Absent: if the study WD Tg was < Y ng/ml **and** the study scan was negative
Reference Standard Tg values: separate values for Y = 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 ng/ml.

Since the WD Tg and scan study data were included in the reference standard, along with the post-therapy scan data, there are no false positive or false negative WD Tg levels or WD scans. This is because remnants/cancer can only be present or absent by the reference standard, ie. there are only TPs and TNs on WD, no FPS or FNS. Therefore, sensitivity, specificity, positive and negative predictive value and accuracy are all 100% on WD (see tables 1B and 2B below for arms I and II, respectively.

Tables 1A and 2A below, classify the # of successfully ablated, Tg antibody negative patients who had remnants/cancer by the WD/post-rx. scan class and WD Tg levels of 1, 2, 3, 5 and 10 ng/ml.

1.A. ARM I. # succ. abl. pts with neg. or + WD dx or post-therapy scans at various Tg cut-offs (if a post-therapy scan was done, this result was used):

WD dx. or post-rx. scan class	Tg ≤ 1 ng/ml	Tg ≥1 - <2 ng/ml	Tg ≥2 - <3 ng/ml	Tg ≥3 - <5 ng/ml	Tg ≥5- <10 ng/ml	Tg ≥ 10 ng/ml
0	14	8	3	3	2	19[a]
1	6	5	1	1	1	7[b]
2	0	0	0	0	1[c]	3
≥ 3	0	0	0	0	0	5

a= patient 604: WD Tg= 2,037 ng/ml; the dx. scan was negative; post-study neck dissection revealed 6/6 lymph nodes positive for metastatic papillary cancer
b= patient 1314: WD Tg= 388 ng/ml; the post-therapy scan was class 1; post WD surgery revealed suprasternal and mediastinal malignant lymph nodes
c = patient 201: WD Tg= 9.0 and the dx. and post-therapy scans were class 2B.
Therefore, at a Tg cutoff of 10 ng/ml, FN rate= 1/79 pts.= 1.3%.

1.B. ARM I: Use of WD Tg + scan (either WD or post-therapy) as the Ref. Std. to Dx Remnants/Cancer @ Tg cut-offs 1-10 ng/ml:
(Note: Disease (remnants and/or cancer) is present= true positive, if the WD Tg is ≥ the given Tg cut-off **or** the WD or post-therapy scan is class ≥ 1. Disease is absent= true negative

if the WD Tg is < the given cut-off **and** there is a negative WD diagnostic **and** post-therapy scan.

pts. who were TP or TN for detection of remnants/cancer at Tg cut-offs between 1-10 ng/ml using as the reference standard: WD Tg ≥ the given cut-off or a positive scan (dx. or post-therapy scan ≥ class 1):

	cut-off 1 ng/ml	cut-off 2 ng/ml	cut-off 3 ng/ml	cut-off 5 ng/ml	cut-off 10 ng/ml
Eligible	81	81	81	81	81
Prevalen.	67	59	56	53	51
FP	0	0	0	0	0
FN	0	0	0	0	0
TP	67	59	56	53	51
TN	14	22	25	28	30
Sensit.	100%	100%	100%	100%	100%
Specific.	100%	100%	100%	100%	100%
+ Pred.va	100%	100%	100%	100%	100%
- Pred.va	100%	100%	100%	100%	100%
Accuracy	100%	100%	100%	100%	100%

2.A. data for following table was derived from ARM II of study TSH 95-0101:

succ. abl. pts with neg. or + WD dx or post-therapy scans at various Tg cut-offs (if a post-therapy scan was done, this result was used):

WD dx. or post-rx. scan class	Tg < 1 ng/ml	Tg ≥ 1- <2 ng/ml	Tg ≥ 2- <3 ng/ml	Tg ≥3- <5 ng/ml	Tg ≥ 5- <10 ng/ml	Tg ≥ 10 ng/ml
0	15	4	3	1	5	6*
1	7	3	1	3	5	10
2	0	0	0	0	0	5
≥ 3	0	0	0	0	0	18

* 3 of these 6 patients had either positive chest CT for lung metastases or cancerous lymph nodes on biopsy post study

2.B. ARM II: Use of WD Tg + Scan (either WD or post-rx.) as the Ref. Std. to Dx. Remnants/Cancer @ Tg cut-offs 1-10 ng/ml:

pts. who were TP or TN for detection of remnants/cancer at

Tg cut-offs between 1-10 ng/ml using as the reference standard: WD Tg ≥ the given cut-off or a positive scan (dx. or post-rx. scan ≥ class 1):

	cut-off 1 ng/ml	cut-off 2 ng/ml	cut-off 3 ng/ml	cut-off 5 ng/ml	cut-off 10 ng/ml
Eligible	87	87	87	87	87
Prevalen.	73	69	66	65	60
FP	0	0	0	0	0
FN	0	0	0	0	0
TP	73	69	66	65	60
TN	14	18	21	22	27
Sensit.	100%	100%	100%	100%	100%
Specific.	100%	100%	100%	100%	100%
+ Pred. va	100%	100%	100%	100%	100%
- Pred. va	100%	100%	100%	100%	100%
Accuracy	100%	100%	100%	100%	100%

2.B. Similar to arm I, 100% operating characteristics (sensitivity, specificity, positive and negative predictive value and accuracy) were assigned to withdrawal because it was used as the reference standard.

RECEIVER OPERATOR CURVES (ROCs)

Having redefined the reference standard as above, the sponsor then proceeded to use receiver operator curves to determine the best post-hoc fit of the Thyrogen Tg database in this study to the study WD Tg and study scan (both WD and post-rx.) database.

ROCs are a plot of sensitivity vs. 1 minus the specificity. The 72 h Thyrogen Tg levels were used to generate these curves because this was the time point identified by the Tg kinetic profile as the time when Thyrogen Tg levels peaked. The reference standard used to generate these curves was:
Thyroid Remnant or Cancer was Present: if the study WD Tg was ≥ Y ng/ml **or** the study scan (either WD or post-therapy) was positive (class ≥ 1);
Thyroid Remnant or Cancer was Absent: if the study WD Tg was < Y ng/ml **and** the study scan was negative
Reference Standard Tg values: separate values for Y = 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 ng/ml.
Thyrogen Tg and THST Tg test values: Test Thyrogen and Tg on THST levels of 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 ng/ml.

Using both the study WD Tg levels and the scans, the sponsor generated ROCs for withdrawal at each reference Tg

cut-off of 1-10 ng/ml. Then, combining study Thyrogen Tg levels from both treatment arms, receiver operator curves were generated for Thyrogen at cut-offs of 1-10 ng/ml to compare to each WD reference standard, to determine which Thyrogen Tg cut-off level provided the best fit post-hoc (in terms of sensitivity and specificity) to the respective WD reference standard. This analysis was again repeated for Tg on THST.

RESULTS OF THE RECEIVER OPERATOR ANALYSES:

The WD Tg cut-offs of 2 and 10 mg/ml generated curves that were distinct from the curves generated for reference values of 3-9 ng/ml. The curve generated for a WD Tg level of 5 ng/ml was chosen to represent the medial of the 3-9 ng/ml curves. The 72 h Thyrogen Tg levels in this study that provided the best fit post-hoc (in terms of sensitivity and specificity) to each of the 2, 5 and 10 ng/ml WD reference standards, were 1, 2 and 3 ng/ml Thyrogen Tg cut-off levels, respectively. Please see the table below which indicates the sensitivity and specificity of the full range of Thyrogen Tg cut-off levels of 1-10 ng/ml as they correlated to the 2, 5 and 10 ng/ml WD reference standards:

Results of the Receiver Operator Curves for Thyrogen Tg:

WD Tg (ng/ml)	Thyrogen Tg (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

Assigning 100% sensitivity to the reference standard as the sponsor did, a Thyrogen Tg of 5-10 ng/ml is only 53-67% as sensitive as a withdrawal Tg of 10 ng/ml or a positive WD/post-rx. scan to detect thyroid remnants or cancer. In this study, a

Thyrogen Tg cut-off level of 3 ng/ml gave the **best post hoc fit** to a WD Tg cut-off of 10 ng/ml or a positive WD scan; a Thyrogen Tg of 2 to a WD Tg of 5 ng/ml and a Thyrogen Tg of 1 to a WD Tg of 2 ng/ml. **As noted above, these specific Thyrogen Tg cut-offs of 1, 2 and 3 ng/ml represent only the best post-hoc fit of the Thyrogen Tg database in this study to withdrawal Tg cut-offs of 2, 5 and 10 ng/ml and to a positive WD/post-rx.scan.**

The results of the receiver operator curves for Tg on THST are:

WD Tg (ng/ml)	Tg on THST (ng/ml)	Sensitivity (%)	Specificity (%)
10	1	67	78
	2	48	94
	3	42	96
	4	41	98
	5-10	40-34	98-100
5	1	64	79
	2	46	96
	3-10	40-32	98-100
2	1	62	84
	2	43	97
	3-10	38-30	100-100

Comment on the above table:

Overall, Tg on THST is slightly more specific but less sensitive compared to Thyrogen Tg. A 1 ng/ml cut-off for Tg on THST provides the best fit compared to the WD reference standards of 2, 5 and 10 ng/ml. However, in the diagnostic utility analyses which follow, the sponsor did not use the best fit for Tg on THST, but they did for Thyrogen Tg, thus magnifying the difference between the two, in favor of Thyrogen.

Now that the ROCs had identified the Thyrogen Tg cut-offs that would give the best fit (optimal sensitivity and specificity) to the WD reference standard, they used these best fit Thyrogen Tg cut-offs and the same reference standard to perform the Thyrogen diagnostic utility analyses.