Estimating Risk of Recurrence in Differentiated Thyroid Cancer After Total Thyroidectomy and Radioactive Iodine Remnant Ablation: Using Response to Therapy Variables to Modify the Initial Risk Estimates Predicted by the New American Thyroid Association Staging System

R. Michael Tuttle,[1] Hernan Tala,[1] Jatin Shah,[2] Rebecca Leboeuf,[1] Ronald Ghossein,[3]
Mithat Gonen,[4] Matvey Brokhin,[1] Gal Omry,[1] James A. Fagin,[1] and Ashok Shaha[2]

Background: A risk-adapted approach to management of thyroid cancer requires risk estimates that change over time based on response to therapy and the course of the disease. The objective of this study was to validate the American Thyroid Association (ATA) risk of recurrence staging system and determine if an assessment of response to therapy during the first 2 years of follow-up can modify these initial risk estimates.
Methods: This retrospective review identified 588 adult follicular cell-derived thyroid cancer patients followed for a median of 7 years (range 1–15 years) after total thyroidectomy and radioactive iodine remnant ablation. Patients were stratified according to ATA risk categories (low, intermediate, or high) as part of initial staging. Clinical data obtained during the first 2 years of follow-up (suppressed thyroglobulin [Tg], stimulated Tg, and imaging studies) were used to re-stage each patient based on response to initial therapy (excellent, acceptable, or incomplete). Clinical outcomes predicted by initial ATA risk categories were compared with revised risk estimates obtained after response to therapy variables were used to modify the initial ATA risk estimates.
Results: Persistent structural disease or recurrence was identified in 3% of the low-risk, 21% of the intermediate-risk, and 68% of the high-risk patients ($p < 0.001$). Re-stratification during the first 2 years of follow-up reduced the likelihood of finding persistent structural disease or recurrence to 2% in low-risk, 2% in intermediate-risk, and 14% in high-risk patients, demonstrating an excellent response to therapy (stimulated $Tg < 1\,ng/mL$ without structural evidence of disease). Conversely, an incomplete response to initial therapy (suppressed $Tg > 1\,ng/mL$, stimulated $Tg > 10\,ng/mL$, rising Tg values, or structural disease identification within the first 2 years of follow-up) increased the likelihood of persistent structural disease or recurrence to 13% in low-risk, 41% in intermediate-risk, and 79% in high-risk patients.
Conclusions: Our data confirm that the newly proposed ATA recurrence staging system effectively predicts the risk of recurrence and persistent disease. Further, these initial ATA risk estimates can be significantly refined based on the assessment of response to initial therapy, thereby providing a dynamic risk assessment that can be used to more effectively tailor ongoing follow-up recommendations.

Introduction

THE LAST SEVERAL YEARS have seen a renewed interest in thyroid cancer management paradigms that base treatment and follow-up recommendations on individualized risk assessments (1–4). This individualized management approach is embraced by the thyroid cancer management guidelines published by the American Thyroid Association (ATA) in which postoperative staging is recommended not only for assessing risk for recurrence and mortality, but also for tailoring decisions regarding both the need for postoperative adjuvant therapy (including need for radioactive iodine [RAI] ablation and degree of thyrotropin [TSH] suppression) as well as the frequency and modality of follow-up studies (5).

Because of its utility in predicting disease-specific mortality, and its widespread use in tumor registries, the ATA

[1]Endocrinology Service, Department of Medicine, [2]Head and Neck Surgery, [3]Pathology, and [4]Epidemiology-Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York.

Table 1. Initial American Thyroid Association Risk of Recurrence Classification

Low risk	Intermediate risk	High risk
All the following are present 　No local or distant metastases 　All macroscopic tumor has been resected 　No invasion of locoregional tissues 　Tumor does not have aggressive histology (e.g., tall cell, insular, columnar cell carcinoma, Hurthle cell carcinoma, follicular thyroid cancer). 　No vascular invasion 　No ^{131}I uptake outside the thyroid bed on the post-treatment scan, if done	Any of the following is present 　Microscopic invasion into the perithyroidal soft tissues 　Cervical lymph node metastases or ^{131}I uptake outside the thyroid bed on the post-treatment scan done after thyroid remnant ablation 　Tumor with aggressive histology or vascular invasion (e.g., tall cell, insular, columnar cell carcinoma, Hurthle cell carcinoma, follicular thyroid cancer)	Any of the following is present 　Macroscopic tumor invasion 　Incomplete tumor resection with gross residual disease 　Distant metastases

guidelines recommend the American Joint Cancer Committee/Union Internationale Contre le Cancer (AJCC/UICC) staging system for use in all patients with differentiated thyroid cancer (5). As a complement to the AJCC/UICC staging system, the ATA guidelines also endorsed the use of other clinicopathologic staging systems to improve prognostication and better tailor follow-up studies for individual patients. Recognizing that the commonly used clinicopathologic staging systems were designed to predict death and not recurrence, the updated ATA guidelines also proposed a novel, but unproven, staging system to predict risk of recurrent/persistent disease in differentiated thyroid cancer (Table 1).

While these initial staging systems provide an important starting point for risk assessment, they are static representations of the patient at the time of initial therapy and are not designed to be modified over time based on the clinical course of the disease. Further, none of the commonly used staging systems include adequate variables to address the impact of treatment on subsequent outcomes (6,7). Since initial surgery and RAI remnant ablation (RRA) are likely to have a major impact on risk of recurrence and risk of death in thyroid cancer patients, the failure to incorporate variables that assess response to these initial therapies into our staging systems will result in overly pessimistic predictions in high-risk patients having an excellent response to therapy and overly

optimistic estimates in low-risk patients that fail initial therapy. Therefore, it is not surprising that the risk estimates provided by any of the commonly used clinicopathologic staging systems account for only a small proportion of the observed variance in disease-specific survival (8–10).

In an effort to integrate the effect of therapy into our risk assessments, we have built on previously published studies from our center (3,11,12) to develop a risk stratification scheme in which clinical data obtained during the first 2 years of follow-up are used to categorize response to therapy as either excellent, acceptable, or incomplete (4) (Table 2). This is based on previous studies demonstrating that risk of persistent disease or recurrence is altered by the completeness of initial surgery, changes in serum thyroglobulin (Tg) or anti-Tg antibody levels (TgAb) over time (13–18), the degree of RAI avidity of pulmonary metastases (19), the presence of an undetectable stimulated Tg especially when coupled with a negative neck ultrasonography (US) in low-risk patients (20), and 2-deoxy-2-(^{18}F)fluoro-D-glucose positron emission tomography (18-FDG-PET) scanning in high-risk patients (21). Further, many recommendations in the ATA guidelines are defined in terms of initial risk stratification as well as the results from follow-up testing (e.g., recommendation # 45b advises follow-up primarily with a suppressed Tg and physical examination in low-risk patients with undetectable

Table 2. Response to Initial Therapy Definitions (6–24 Months After Radioactive Iodine Ablation)

Excellent response	Acceptable response	Incomplete response
All the following 　Suppressed and stimulated Tg < 1 ng/mL 　Neck US without evidence of disease 　Cross-sectional and/or nuclear medicine imaging negative (if performed)	Any of the following 　Suppressed Tg < 1 ng/mL and stimulated Tg ≥ 1 and <10 ng/mL 　Neck US with nonspecific changes or stable subcentimeter lymph nodes 　Cross-sectional and/or nuclear medicine imaging with nonspecific changes, although not completely normal	Any of the following 　Suppressed Tg ≥ 1 ng/mL or stimulated Tg ≥ 10 ng/mL 　Rising Tg values 　Persistent or newly identified disease on cross-sectional and/or nuclear medicine imaging

Tg, thyroglobulin; US, ultrasonography.

stimulated Tg and negative neck US after total thyroidectomy and RRA) (5).

In this article, we validate the newly proposed ATA risk stratification system for prediction of early recurrence of disease and proceed to demonstrate how these initial risk estimates can be refined by incorporating response to therapy variables in a retrospective review of 588 consecutive thyroid cancer patients followed for a median of 7 years at a single, tertiary-care referral center.

Materials and Methods

Subjects

After obtaining institutional review board (IRB) approval, we retrospectively reviewed the electronic medical records of 710 consecutive patients with differentiated thyroid cancer evaluated at Memorial Sloan Kettering Cancer Center (MSKCC) who had undergone total thyroidectomy and RRA either at MSKCC or elsewhere between January 1994 and December 2004. Of the 710 potentially eligible patients, 588 patients had adequate clinicopathological information to allow for accurate initial risk stratification and determination of clinical status throughout their follow-up. Of the 710 potential patients, 121 were excluded from the study for the following reasons: inadequate follow-up information ($n = 64$), interfering anti-Tg antibodies ($n = 25$), inadequate information for initial staging ($n = 17$), age <18 years at diagnosis ($n = 13$), or anaplastic thyroid cancer histology ($n = 3$). A minimum of 3 years of follow-up was required for entry into the study unless one of the clinical endpoints (recurrence or death) was reached before that time point. Patients <18 years old at diagnosis, or with a histological diagnosis of medullary thyroid cancer or anaplastic thyroid cancer were excluded from the study. Low- to intermediate-risk patients not requiring total thyroidectomy or RRA were not included in this study.

All patients included in the study were receiving TSH suppressive therapy and had at least one neck US performed at our center after initial therapy during the first 2 years of follow-up (most had two or more serial US evaluations) and two or more serum Tg and TgAb determinations obtained on levothyroxine suppression during the first 2 years of follow-up. All patients had a Tg and TgAb levels while on TSH suppressive therapy at the time of final follow-up. Patients with interfering TgAb were excluded (Dynotest-TgS immunoradiometric assay; Brahms, Inc., Berlin, Germany). Complete surgical resection was defined as an extracapsular total thyroidectomy with therapeutic, compartment-oriented neck dissection for suspicious or biopsy-proven metastatic cervical lymphadenopathy. A stimulated Tg value within the first 2 years of follow-up was not a requirement for inclusion in the study but was available in 80% of the patients (see Results section).

Risk stratification

Each patient ($n = 588$) was risk-stratified using the 7th edition of the AJCC/UICC staging system (stage I, II, III, or IV) and the newly proposed ATA risk of recurrence stratification system (low-, intermediate-, or high-risk of recurrence; Table 1) (5).

Only those patients who had a stimulated Tg determination during the first 2 years of follow-up ($n = 471$) could be used in the analysis of response to therapy (Table 2). All clinical data obtained during the first 2 years of follow-up were used to assess the response to initial therapy (total thyroidectomy and RAI ablation) as either excellent, acceptable, or incomplete using the re-stratification scheme proposed by our group (4).

Laboratory studies

Between 1994 and 1997, a variety of Tg assays were used with functional sensitivities of approximately 1 ng/mL (used as part of the 2 year response to therapy assessment in 128 patients, 21 of whom had a stimulated Tg value <1 ng/mL). Starting in 1998, all Tg values were measured using the Dynotest-TgS immunoradiometric assay (Brahms, Inc.; functional sensitivity 0.6 ng/mL normalized to CRM 457).

Follow-up

Patients were usually followed every 6 months during the first year and at 6–12 month intervals thereafter at the discretion of the attending physician based on the risk of the individual patient and the clinical course of the disease.

Clinical endpoints

Patients were considered to have no clinical evidence of disease (NED) at final follow-up if they had a suppressed serum Tg < 1 ng/mL, no detectable TgAb, and no structural evidence of disease. Patients with suppressed Tg values ≥1 ng/mL, stimulated Tg values ≥2 ng/mL, or any evidence of disease on cross-sectional imaging (US, computed tomography scan, or magnetic resonance imaging), functional imaging (RAI scan or 18-FDG-PET scan), or biopsy-proven disease (cytology or histology) were considered as having persistent disease. A recurrence was defined as new biochemical (suppressed Tg ≥1 ng/mL, and/or stimulated Tg ≥2 ng/mL), structural, or functional evidence of disease that was detected following any period of NED. Disease-specific mortality was also an endpoint (patients dying of unrelated conditions had the final endpoint determined based on data available before their demise).

Patients with persistent disease were further classified as having either biochemical evidence of disease (elevated basal or stimulated Tg values alone without structural correlate) or structural evidence of disease on imaging. Patients were considered to have structural evidence of disease if any of the following conditions were met: (i) positive cytology/histology, or (ii) highly suspicious lymph nodes or thyroid bed nodules on the neck US (hypervascularity, cystic areas, heterogeneous content, rounded shape, or enlargement on the follow-up), or (iii) findings on RAI scans, 18-FDG-PET scans, or other cross-sectional imaging highly suspicious for metastatic disease.

Statistical methods

Continuous data are presented as means and standard deviations with median values. Categoric comparisons were performed with the Fischer's exact test. Analysis was performed using SPSS software (version 16.0.1: SPSS, Inc., Chicago, IL). The proportion of variance explained (PVE) is

estimated using Nagelkerke's method and the RSQUARE option in PROC LOGISTIC of SAS/SAT (22,23).

Results

The demographics, clinical features, risk stratification, and final outcomes for each of the 588 patients included in this study are presented in Table 3. As expected, the majority of patients had papillary thyroid cancer (85%), were female (68%), and were classified as either having a low (23%) or intermediate (50%) risk for recurrence based on the new ATA

TABLE 3. DESCRIPTION OF THE COHORT

		n
Age (years)		
Mean ± SD	46 ± 15	588
Median	46	
Range	18–83	
Gender		
Female	68%	400
Histology		
Papillary	85%	500
Poorly differentiated	6%	35
Follicular	4%	24
Hurthle cell	5%	29
131I activity for ablation (mCi)		
Mean ± SD	134 ± 72	588
Median	114	
AJCC stage		
I	48%	281
II	12%	71
III	15%	89
IVa	12%	71
IVb	1%	5
IVc	12%	71
ATA initial risk classification		
Low	23%	135
Intermediate	50%	294
High	27%	159
Response to therapy classification		
Excellent	34%	159
Acceptable	20%	95
Incomplete	46%	217
Stimulated Tg done (first 2 years)		
Yes	80%	470
No	20%	118
Follow-up Duration (years)		
Mean ± SD	8 ± 3	588
Median	7	
Range	1–15	
Clinical status after initial therapy		
No evidence of disease	52%	305
Biochemical evidence of persistent disease	19%	108
Structural evidence of persistent disease	28%	167
Recurrent disease	1%	8
Status at final follow-up		
No evidence of disease	67%	394
Persistent/recurrent disease	28%	165
Death of disease	5%	29

$n = 588$.
ATA, American Thyroid Association; SD, standard deviation; AJCC, American Joint Cancer Committee.

risk stratification system. The relatively high proportion of patients classified as having high-risk of recurrence (27%) probably reflects the selection bias of patients referred to our cancer center. During follow-up, only 1% of the patients developed recurrent disease after a period of NED, whereas 28% of the patients had structural evidence of persistent disease and 19% had biochemical evidence of persistent disease (without a structural correlate) as the best response to initial therapy (total thyroidectomy and RAI ablation). However, at the time of last follow-up (a median of 7 years after initial therapy, and after additional therapy was given to some patients), 67% of the patients were NED, 28% had persistent biochemical, structural, or recurrent disease, and 5% had died of thyroid cancer.

We first evaluated the ability of the AJCC staging system to predict the clinically relevant endpoints of NED, persistent disease, and recurrent disease in the entire group of patients ($n = 588$, Table 4). While AJCC stage IV patients were less likely to be NED (20%) and more likely to have persistent structural disease (68%) than stage I–III patients ($p < 0.01$), the likelihood of developing recurrent disease detected after a period of NED was similar in all AJCC stages (1%–2%). Further, the prevalence of persistent structural disease was higher in stage II patients (34%) than stage III patients (14%). Evaluation of the stage II patients with persistent disease reveals that 92% (34/37) of them were <45 years old at diagnosis with metastatic disease at presentation (M1). Additionally, the other three stage II patients with persistent disease had relatively high-risk tumors (55 year old with 2 cm papillary thyroid cancer, 50 year old with 3 cm Hürthle cell carcinoma, and 60 year old with 4 cm poorly differentiated thyroid cancer). Finally, 26/28 patients who died of thyroid cancer were AJCC stage IV (1 stage II, 1 stage III).

When classified based on the ATA risk stratification system ($n = 588$ patients), 3% of the low-risk patients, 21% of the intermediate-risk patients, and 68% of the high-risk patients ($p < 0.001$) had persistent structural or recurrent disease detected during follow-up (Table 5). In addition, 11% of low-risk, 22% of intermediate-risk, and 18% of high-risk patients had biochemical evidence of persistence without structurally identifiable disease. In total, 14% of the low-risk, 43% of the intermediate-risk, and 86% of the high-risk patients ($p < 0.001$) had biochemical or structural evidence of persistent or recurrent disease following total thyroidectomy and RRA. All patients dying of thyroid cancer had been classified as having a high risk of recurrence.

Re-stratification based on response to initial therapy (total thyroidectomy and RRA) in the 471 patients with stimulated Tg values available at the 2-year follow-up time point demonstrated that only 4% (6/159) of the excellent responders had recurrent disease that was identified 4–11 years after initial therapy (5 with Tg elevations without structurally identifiable disease and 1 with loco-regional lymph node recurrence detected by neck US with suppressed Tg < 0.6 ng/mL and no detectable TgAb) (Table 6). None of the patients with an acceptable response to therapy demonstrated either persistent structural disease or recurrence during follow-up. However, 13% of the acceptable responder patients had biochemical evidence of persistent disease during follow-up without evidence of structurally identifiable disease. Finally, only 4% (8/217) of the incomplete responders were rendered NED without additional therapy. In each of these eight patients, a

TABLE 4. CLINICAL OUTCOMES FOLLOWING INITIAL THERAPY FOR EACH AMERICAN JOINT CANCER COMMITTEE STAGE

	n = 588			
Clinical outcome following initial therapy	AJCC I (n = 281)	AJCC II (n = 71)	AJCC III (n = 89)	AJCC IV (n = 147)
No evidence of disease (n = 305)	66% (186)	48% (34)	63% (56)	20% (29)
Persistent disease, biochemical evidence only (n = 108)	23% (64)	18% (12)	21% (18)	10% (14)
Persistent disease, structurally identifiable (n = 167)	10% (28)	34% (25)	14% (13)	68% (101)
Recurrent disease (n = 8)	1% (3)	0% (0)	2% (2)	2% (3)

gradual decline in serum Tg values over many years without additional therapy rendered them NED at final follow-up. Persistent structural disease and persistent biochemical disease was present significantly more often in the incomplete responders than the excellent or acceptable responders ($p < 0.001$).

The impact of the response to initial therapy re-stratification is most apparent in patients initially classified as intermediate or high risk of recurrence (Table 7). In these patients, an excellent response to therapy results in a significant decrease in the likelihood of having persistent structural disease or recurrent thyroid cancer (from 18% to 2% in intermediate-risk and 66% to 14% in high-risk patients). Likewise, an incomplete response to therapy is associated with an increased likelihood of having persistent structural disease or recurrence in each of the initial risk categories (3% to 13% in low-risk, 18% to 41% in intermediate-risk, and 66% to 79% in high-risk patients).

In our response to therapy assessments, both the suppressed and stimulated Tg values play a central role in risk re-stratification. To determine if the additional clinical data (negative cross sectional and/or functional imaging) used in the excellent response to therapy definition was improving the ability to identify patients at lowest risk of recurrence, the likelihood of being NED at the final follow-up was calculated

for each of the following possible definitions of response to therapy assessed at the 2 years of follow-up point: suppressed Tg < 1 ng/mL alone, stimulated Tg < 1 ng/mL alone, and the composite excellent response to therapy endpoint combining clinical data (cross sectional and/or functional imaging) with either a suppressed or stimulated Tg < 1 ng/mL (Table 8). In each of the ATA risk categories, the composite excellent response to therapy endpoint using as part of its definition a stimulated Tg < 1 ng/mL provided the highest likelihood of being NED at the final follow-up. However, using a suppressed Tg < 1 ng/mL (with no stimulated Tg assessment) as part of the definition of excellent response at the 2-year follow-up time point also provided clinically meaningful information about the likelihood of being NED at the last follow-up, in both the ATA low-risk (94%) and intermediate-risk (90%) patients.

The ability of the ATA staging system and the response to therapy re-stratification system to predict recurrent/ persistent disease was assessed by determining the PVE for each system. The ATA risk stratification system was able to account for 34% (PVE = 0.34) of the observed variance in predicting recurrent/persistent disease. However, the response to therapy re-stratification system was able to account for 84% (PVE = 0.84) of the variance observed.

Discussion

Our data demonstrate for the first time that the newly proposed ATA risk of recurrence stratification system effectively defines short-term risk of having recurrent or persistent structural disease in differentiated thyroid cancer patients treated with total thyroidectomy and RRA over a 7-year follow-up period. These initial risk estimates can be used to guide management recommendations during the first 1–2 years of follow-up. However, the intensity and modality of long-term follow-up studies should be tailored to revised risk estimates that are based on response to therapy assessments, rather than the initial static risk estimates that do not account for the effects of treatment on clinically relevant outcomes.

Consistent with the previous recent publications (24,25), our data also demonstrate that the AJCC staging is not the best tool for predicting the risk of recurrent/persistent disease. This is most readily apparent in the high rate of persistent/ recurrent disease seen in stage II patients. This is almost

TABLE 5. CLINICAL OUTCOMES FOLLOWING INITIAL THERAPY FOR AMERICAN THYROID ASSOCIATION RISK CATEGORIES

	n = 588		
Clinical outcome following initial therapy	Low (n = 136)	Intermediate (n = 291)	High (n = 161)
No evidence of disease (n = 305)	86% (117)	57% (166)	14% (22)
Persistent disease, biochemical evidence (n = 108)	11% (15)	22% (64)	18% (29)
Persistent disease, structurally identifiable (n = 167)	2% (3)	19% (56)	67% (108)
Recurrent disease (n = 8)	1% (1)	2% (5)	1% (2)

Table 6. Re-Stratification Based on Response to Initial Therapy

	n = 471		
Clinical outcome after re-staging	Excellent response (n = 159)	Acceptable response (n = 95)	Incomplete response (n = 217)
No evidence of disease (n = 245)	96% (153)	87% (84)	4% (8)
Persistent disease, biochemical evidence (n = 96)	0% (0)	13% (11)	39% (85)
Persistent disease, structurally identifiable (n = 124)	0% (0)	0% (0)	57% (124)
Recurrent disease (n = 6)	4% (6)	0% (0)	0% (0)

Table 7. Impact of Response to Initial Therapy Assessment on Initial Estimates of Risk

	ATA initial risk of recurrence classification (n = 471)		
	Low	Intermediate	High
Initial estimate of risk of persistent structural or recurrent disease	3% (3/104)	18% (43/241)	66% (83/126)
Modified estimate of risk of persistent structural or recurrent disease based on response to initial therapy			
Excellent response (n = 159)	2% (1/59)	2% (2/86)	14% (2/14)
Acceptable response (n = 95)	0% (0/30)	0% (0/56)	0% (0/9)
Incomplete response (n = 217)	13% (2/15)	41% (41/99)	79% (81/103)

Table 8. Using Thyroglobulin and Response to Initial Therapy Assessments at the 2-Year Follow-Up Time Point to Predict Likelihood of Being No Clinical Evidence of Disease at Final Follow-Up

Initial risk stratification	Response to therapy variables during first 2 years of follow-up	NED at final follow-up
Low risk (n = 104)	Suppressed Tg < 1 ng/mL alone	84%
	Stimulated Tg < 1 ng/mL alone	89%
	Excellent response (imaging negative[a] and suppressed Tg < 1 ng/mL)	94%
	Excellent response (imaging negative[a] and stimulated Tg < 1 ng/mL)	97%
Intermediate risk (n = 241)	Suppressed Tg < 1 ng/mL alone	74%
	Stimulated Tg < 1 ng/mL alone	80%
	Excellent response (imaging negative[a] and suppressed Tg < 1 ng/mL)	90%
	Excellent response (imaging negative[a] and stimulated Tg < 1 ng/mL)	94%
High risk (n = 126)	Suppressed Tg < 1 ng/mL alone	39%
	Stimulated Tg < 1 ng/mL alone	55%
	Excellent response (imaging negative[a] and suppressed Tg < 1 ng/mL)	80%
	Excellent response (imaging negative[a] and stimulated Tg < 1 ng/mL)	82%

n = 471 with both suppressed and stimulated Tg values available for analysis.
[a]Negative imaging: normal neck US in all patients. In addition, any other functional or cross-sectional imaging obtained at the discretion of the treating physician was interpreted as having no evidence of persistent/recurrent thyroid cancer.
NED, no clinical evidence of disease.

entirely due to young patients with distant metastases at diagnosis (M1) who are very unlikely to die of thyroid cancer, but quite likely to have persistent disease during the first several years of follow-up (26). Therefore, while the AJCC staging system provides valuable information regarding overall survival, it does not provide adequate information with regard to risk of recurrent/persistent disease.

The ATA risk stratification system is a static representation of risk based on clinicopathologic features available at the time of initial evaluation. By overlaying a response to therapy assessment on the initial risk stratification estimates, a more accurate and individualized risk assessment can be formulated for each patient during long-term follow-up. From a clinical management perspective, the most useful, dynamic risk estimates are obtained when response to therapy assessments are used to refine initial risk estimates.

The change in risk estimates as a function of response to therapy is most notable in the ATA intermediate-risk group where an estimated risk of recurrent or persistent structural disease of 18% decreases to as little as 2% in patients having an excellent response to therapy. Similarly, the few high-risk patients with an excellent response to therapy see the risk of recurrent or persistent structural disease decrease from 66% to 14%. On the contrary, the small number ATA low-risk patients who have an incomplete response to therapy see the risk of having persistent structural disease at the time of final follow-up rise from the initial 3% estimate to 13% based on the lack of response to total thyroidectomy and adjuvant RAI therapy. The 1%–2% risk of recurrence seen in the ATA low- and intermediate-risk patients with an excellent response to therapy is very consistent with the findings of Pacini *et al.* in low-risk patients with undetectable stimulated Tg values and negative neck US obtained during the first year of follow-up (20).

Since we defined recurrence as newly detected biochemical or structural disease following any period of no evidence of disease, only patients achieving an excellent response to therapy (suppressed serum Tg < 1 ng/mL, a stimulated Tg < 2 ng/mL (if done), and no structural evidence of disease) could be subsequently classified as having disease recurrence. This definition differs from the language commonly used in clinical practice where newly identified structural disease (even in patients with persistent biochemical evidence of disease) is often referred to as a recurrence. However, we feel that structural disease identified in the follow-up of patients not previously rendered free of disease is better described as persistent rather than recurrent disease.

Our data show that a composite clinical endpoint that includes a stimulated Tg value <1 ng/mL in addition to negative imaging is associated with the highest likelihood of being NED at final follow-up. However, our data also demonstrate that a rather high likelihood of being NED can also be achieved in low-risk (94%) and intermediate-risk (90%) patients based on negative neck US and a suppressed Tg < 1 ng/mL at the 2-year follow-up point. Therefore, the decision regarding whether a stimulated Tg value is required in an individual patient must be based the perceived need for achieving the highest possible likelihood of being NED, the functional sensitivity and reliability of the available Tg assay, the quality of the neck US, the expense of the stimulated Tg testing, and the inconvenience of the additional office visits required for the stimulated testing.

As the PVE analysis demonstrates, the ATA risk system only accounts for 34% of the variance in predicting recurrent/persistent disease. This value is consistent with the PVE values reported for predicting death in many of the commonly used staging systems [e.g., American Joint Cancer Committee (AJCC), metastasis, age, completeness of resection, invasion, size (MACIS), Memorial Sloan Kettering Cancer Center (MSKCC), and National Thyroid Cancer Treatment Cooperative study], which typically range from 5% to 30% depending on the cohort analyzed and the method used for calculation (8–10). It is not surprising that the addition of response to therapy assessment markedly improves the predictive ability of a staging system since much of the variability in outcome can be appreciated in the first 2 years of follow-up. In our data, re-stratification based on response to therapy assessment increased the PVE to 84%, allowing for much more confident prediction of outcomes for individual patients.

Potentially important variables that could account for some of the variation not explained by current staging systems could include (i) the impact of additional therapies given after the 2 year re-stratification, (ii) complications of therapy that alter either overall survival or our ability to detect recurrent disease, and (iii) the influence of initial genotype of tumor aggressiveness, response to therapy, or ability to detect recurrent disease (e.g., downregulation of sodium iodine symporter and Tg production seen with tumors harboring the BRAF mutation) (27).

As with all retrospective studies, our data have several important limitations. While a median follow-up period of 7 years is a reasonable starting point for assessment of clinical outcomes (especially the risk of persistent disease), longer studies will be required to assess the risk of late recurrence. It is likely that a small number of patients classified as NED at 7 years may develop clinically significant recurrent disease in the years to come. Indeed, the few patients who recurred after showing an excellent response to initial therapy did not manifest clinically evident disease until between 4 and 11 years after initial therapy. Since follow-up recommendations were based on individual physician preference and not on a prospective protocol, the intensity and frequency of follow-up studies varied from patient to patient with more intensive testing being done in the highest risk patients. This would likely lead to an increased sensitivity for detection of persistent/recurrent disease in high-risk patients than the less rigorous testing paradigm often used in low-risk patients. Further, the exclusion of pediatric patients and patients with interfering antibodies are also a limitation of this study.

Even though all the patients in this study were followed here at MSKCC, many of them had their initial surgeries and RRA done before being referred to our center. Therefore, many important aspects of their initial therapy cannot be assessed in detail. These include many critical factors that have a major impact on the effectiveness of the initial therapy, including completeness of initial surgery, intraoperative decision making with regard to extent of resection with respect to both extrathyroidal disease and lymph node metastases, adherence to a low-iodine diet before ablation, adequacy of TSH stimulation for ablation, administered activity of RAI, interpretation of diagnostic and follow-up whole-body RAI scans, and the need for external beam irradiation therapy. Thus, while this study does address the impact of initial therapy

(surgery and RRA) in a large cohort of patients followed at a single institution, additional studies will be required to determine if meticulous surgical resection, with or without RRA, can achieve similar (or better) clinical outcomes than those reported in this series of patients who received initial therapy in centers with a varying degree of expertise in thyroid cancer management.

A final limitation of this study is that the definitions of the response to therapy (excellent, acceptable, and incomplete) used in this article can only be applied to patients rendered athyreotic with total thyroidectomy and RRA. This limitation is not meant to imply that low- or intermediate-risk patients always require remnant ablation. Quite to the contrary, we routinely follow low- to intermediate-risk patients treated with either lobectomy or total thyroidectomy without RRA using physical examination findings, follow-up neck US, and serial Tg values to assess response to therapy. Further studies are warranted to address the usefulness of the response to therapy assessment in patients not treated with RAI.

In summary, our data confirm the ATA risk classification as a good initial predictor of recurrent/persistent disease in differentiated thyroid carcinoma. However, our data also indicate that a risk-adapted approach to follow-up cannot be based solely on static, initial estimates of risk that remain unchanged over the life of the patient. It is necessary to incorporate response to therapy variables into our risk stratification systems to provide accurate, dynamic risk estimates for individual patients. These changing risk estimates can then be used to intelligently modify follow-up management plans over time. This management approach will allow us to better tailor the risk/benefits of treatment, testing, and follow-up to the risks of recurrent/persistent disease and death for each individual patient.

Acknowledgment

Disclosure Statement

H.T., J.S., R.G., M.G., J.F., G.O., and A.S. have nothing to declare. R.L. receives research support from Genzyme Corporation. R.M.T. is a consultant to and has received honoraria from the Genzyme Corporation.

References

1. Hay ID 2007 Management of patients with low-risk papillary thyroid carcinoma. Endocr Pract 13:521–533.
2. Mazzaferri EL 2007 Management of low-risk differentiated thyroid cancer. Endocr Pract 13:498–512.
3. Shaha AR, Shah JP, Loree TR 1997 Low-risk differentiated thyroid cancer: the need for selective treatment. Ann Surg Oncol 4:328–333.
4. Tuttle RM 2008 Risk-adapted management of thyroid cancer. Endocr Pract 14:764–774.
5. Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel SJ, Mazzaferri EL, McIver B, Pacini F, Schlumberger M, Sherman SI, Steward DL, Tuttle RM 2009 Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. Thyroid 19:1167–1214.
6. Byar DP, Green SB, Dor P, Williams ED, Colon J, van Gilse HA, Mayer M, Sylvester RJ, van Glabbeke M 1979 A prognostic index for thyroid carcinoma. A study of the E.O.R.T.C. Thyroid Cancer Cooperative Group. Eur J Cancer 15:1033–1041.
7. Lang BH, Lo CY, Chan WF, Lam KY, Wan KY 2007 Staging systems for papillary thyroid carcinoma: a review and comparison. Ann Surg 245:366–378.
8. Brierley JD, Panzarella T, Tsang RW, Gospodarowicz MK, O'Sullivan B 1997 A comparison of different staging systems predictability of patient outcome. Thyroid carcinoma as an example. Cancer 79:2414–2423.
9. Sherman SI, Brierley JD, Sperling M, Ain KB, Bigos ST, Cooper DS, Haugen BR, Ho M, Klein I, Ladenson PW, Robbins J, Ross DS, Specker B, Taylor T, Maxon HR, 3rd 1998 Prospective multicenter study of thyroid carcinoma treatment: initial analysis of staging and outcome. National Thyroid Cancer Treatment Cooperative Study Registry Group. Cancer 83:1012–1021.
10. Verburg FA, Mader U, Kruitwagen CL, Luster M, Reiners C 2010 A comparison of prognostic classification systems for differentiated thyroid carcinoma. Clin Endocrinol (Oxf) 72:830–838.
11. Shah JP 2008 Exploiting biology in selecting treatment for differentiated cancer of the thyroid gland. Eur Arch Otorhinolaryngol 265:1155–1160.
12. Shaha AR, Shah JP, Loree TR 1996 Risk group stratification and prognostic factors in papillary carcinoma of thyroid. Ann Surg Oncol 3:534–538.
13. Castagna MG, Brilli L, Pilli T, Montanaro A, Cipri C, Fioravanti C, Sestini F, Capezzone M, Pacini F 2008 Limited value of repeat recombinant human thyrotropin (rhTSH)-stimulated thyroglobulin testing in differentiated thyroid carcinoma patients with previous negative rhTSH-stimulated thyroglobulin and undetectable basal serum thyroglobulin levels. J Clin Endocrinol Metab 93:76–81.
14. Chiovato L, Latrofa F, Braverman LE, Pacini F, Capezzone M, Masserini L, Grasso L, Pinchera A 2003 Disappearance of humoral thyroid autoimmunity after complete removal of thyroid antigens. Ann Intern Med 139:346–351.
15. Kloos RT, Mazzaferri EL 2005 A single recombinant human thyrotropin-stimulated serum thyroglobulin measurement predicts differentiated thyroid carcinoma metastases three to five years later. J Clin Endocrinol Metab 90:5047–5057.
16. Mazzaferri EL, Robbins RJ, Spencer CA, Braverman LE, Pacini F, Wartofsky L, Haugen BR, Sherman SI, Cooper DS, Braunstein GD, Lee S, Davies TF, Arafah BM, Ladenson PW, Pinchera A 2003 A consensus report of the role of serum thyroglobulin as a monitoring method for low-risk patients with papillary thyroid carcinoma. J Clin Endocrinol Metab 88:1433–1441.
17. Spencer CA 2000 Serum thyroglobulin measurements: clinical utility and technical limitations in the management of patients with differentiated thyroid carcinomas. Endocr Pract 6:481–484.
18. Toubeau M, Touzery C, Arveux P, Chaplain G, Vaillant G, Berriolo A, Riedinger JM, Boichot C, Cochet A, Brunotte F 2004 Predictive value for disease progression of serum thyroglobulin levels measured in the postoperative period and after (131)I ablation therapy in patients with differentiated thyroid cancer. J Nucl Med 45:988–994.
19. Durante C, Haddy N, Baudin E, Leboulleux S, Hartl D, Travagli JP, Caillou B, Ricard M, Lumbroso JD, De Vathaire

F, Schlumberger M 2006 Long-term outcome of 444 patients with distant metastases from papillary and follicular thyroid carcinoma: benefits and limits of radioiodine therapy. J Clin Endocrinol Metab **91**:2892–2899.

20. Pacini F, Molinaro E, Castagna MG, Agate L, Elisei R, Ceccarelli C, Lippi F, Taddei D, Grasso L, Pinchera A 2003 Recombinant human thyrotropin-stimulated serum thyroglobulin combined with neck ultrasonography has the highest sensitivity in monitoring differentiated thyroid carcinoma. J Clin Endocrinol Metab **88**:3668–3673.

21. Robbins RJ, Wan Q, Grewal RK, Reibke R, Gonen M, Strauss HW, Tuttle RM, Drucker W, Larson SM 2006 Real-time prognosis for metastatic thyroid carcinoma based on 2-[18F]fluoro-2-deoxy-D-glucose-positron emission tomography scanning. J Clin Endocrinol Metab **91**:498–505.

22. Nagelkerke NJD 1991 A note on a general definition of the coefficient of determination. Biometrika **78**:691–692.

23. SAS Institute I 2002–2004 SAS 9.1.3 Help and Documentation, SAS Institute, Inc., Cary, NC.

24. Baek SK, Jung KY, Kang SM, Kwon SY, Woo JS, Cho SH, Chung EJ 2010 Clinical risk factors associated with cervical lymph node recurrence in papillary thyroid carcinoma. Thyroid **20**:147–152.

25. Orlov S, Orlov D, Shaytzag M, Dowar M, Tabatabaie V, Dwek P, Yip J, Hu C, Freeman JL, Walfish PG 2009 Influence of age and primary tumor size on the risk for residual/recurrent well-differentiated thyroid carcinoma. Head Neck **31**:782–788.

26. Mazzaferri EL, Kloos RT 2001 Clinical review 128: current approaches to primary therapy for papillary and follicular thyroid cancer. J Clin Endocrinol Metab **86**:1447–1463.

27. Knauf JA, Fagin JA 2009 Role of MAPK pathway oncoproteins in thyroid cancer pathogenesis and as drug targets. Curr Opin Cell Biol **21**:296–303.

Address correspondence to:
R. Michael Tuttle, M.D.
Endocrinology Service
Department of Medicine
Memorial Sloan-Kettering Cancer Center
Zuckerman Building, Room 834
1275 York Ave.
New York, NY 10021

E-mail: tuttlem@mskcc.org