

## Long-Term Outcomes Following Therapy in Differentiated Thyroid Carcinoma: NTCTCS Registry Analysis 1987–2012

Aubrey A. Carhill, Danielle R. Litofsky, Douglas S. Ross, Jacqueline Jonklaas, David S. Cooper, James D. Brierley, Paul W. Ladenson, Kenneth B. Ain, Henry G. Fein, Bryan R. Haugen, James Magner, Monica C. Skarulis, David L. Steward, Mingxhao Xing, Harry R. Maxon, and Steven I. Sherman\*

**Context:** Initial treatments for patients with differentiated thyroid cancer are supported primarily by single-institution, retrospective studies, with limited follow-up and low event rates. We report updated analyses of long-term outcomes after treatment in patients with differentiated thyroid cancer.

**Objective:** The objective was to examine effects of initial therapies on outcomes.

**Design/Setting:** This was a prospective multi-institutional registry.

**Patients:** A total of 4941 patients, median follow-up, 6 years, participated.

**Intervention:** Interventions included total/near-total thyroidectomy (T/NTT), postoperative radioactive iodine (RAI), and thyroid hormone suppression therapy (THST).

**Main Outcome Measure:** Main outcome measures were overall survival (OS) and disease-free survival using product limit and proportional hazards analyses.

**Results:** Improved OS was noted in NTCTCS stage III patients who received RAI (risk ratio [RR], 0.66;  $P = .04$ ) and stage IV patients who received both T/NTT and RAI (RR, 0.66 and 0.70; combined  $P = .049$ ). In all stages, moderate THST (TSH maintained subnormal-normal) was associated with significantly improved OS (RR stages I–IV: 0.13, 0.09, 0.13, 0.33) and disease-free survival (RR stages I–III: 0.52, 0.40, 0.18); no additional survival benefit was achieved with more aggressive THST (TSH maintained undetectable-subnormal). This remained true, even when distant metastatic disease was diagnosed during follow-up. Lower initial stage and moderate THST were independent predictors of improved OS during follow-up years 1–3.

**Conclusions:** We confirm previous findings that T/NTT followed by RAI is associated with benefit in high-risk patients, but not in low-risk patients. In contrast with earlier reports, moderate THST is associated with better outcomes across all stages, and aggressive THST may not be warranted even in patients diagnosed with distant metastatic disease during follow-up. Moderate THST continued at least 3 years after diagnosis may be indicated in high-risk patients. (*J Clin Endocrinol Metab* 100: 3270–3279, 2015)

Differentiated thyroid carcinoma (DTC) arising from thyroid follicular cells includes both papillary and follicular histological types, which account for more than 90% of all thyroid cancers. Although most patients diagnosed with DTC have excellent long-term survival, a sig-

nificant proportion may have persistent/recurrent disease, and some eventually die from their thyroid malignancy. Accurate prognostication is necessary to identify which patients may benefit from more or less aggressive therapy, but current staging systems still fail to account for much of

ISSN Print 0021-972X ISSN Online 1945-7197

Printed in USA

Copyright © 2015 by the Endocrine Society

Received February 5, 2015. Accepted July 7, 2015.

First Published Online July 14, 2015

\* Author Affiliations are shown at the bottom of the next page.

Abbreviations: CI, confidence interval; DFS, disease-free survival; DTC, differentiated thyroid cancer; FTC, follicular thyroid cancer; HCC, Hürthle cell variant; OS, overall survival; PTC, papillary thyroid cancer; RAI, radioactive iodine; RR, risk ratio; THST, thyroid hormone suppression therapy; T/NTT, total or near-total thyroidectomy.

the variance in disease outcomes (1–4). As death related to thyroid cancer is uncommon and often occurs years after diagnosis, randomized trials of primary treatment for DTC have been considered impractical. Nonetheless, incidence rates are escalating, and the need for accurate risk stratification and long-term outcome data to support treatment decisions is increasingly relevant both for those with early stage, very low risk tumors, and for those with advanced and metastatic disease.

Standard of care treatments include surgery, radioactive iodine-131 (RAI), and thyroid hormone suppression therapy (THST). In the absence of prospective trials, considerable debate remains as to the appropriate extent of surgery, benefit and dosing of postoperative RAI, and optimal level and duration of THST. Consequently, there has been extensive dependence upon retrospective studies with low event rates and/or expert opinion (5). For example, balancing the potential benefits and risk of more aggressive total or near-total thyroidectomy (T/NTT) remains challenging, especially for patients at very low risk for cancer-specific mortality (6–11). Despite guidelines from professional societies that describe indications for the use of postoperative RAI (5, 12), application in general practice differs widely (13), underscoring the need for stronger evidence. Although the use of THST has been reported to decrease recurrence rates and cancer-related mortality (14, 15), the optimal degree of suppression required to achieve these goals is not clear. The reduction of TSH levels to  $\leq 0.1$  mU/L has been associated with better clinical outcomes in high-risk thyroid cancer patients (16), but the possibility that milder reductions in TSH might offer the same benefits has not been convincingly defined. Earlier registry data from the National Thyroid Cancer Treatment Cooperative Study Group (NTCTCS) (7, 17) suggested that reduced disease-specific mortality rates are seen in high-risk patients whose TSH is suppressed to very low/undetectable concentrations, but low-risk patients had equivalent outcomes whether the serum TSH is completely suppressed or maintained in the low/normal range.

The NTCTCS, formed in 1987, maintains a multicenter registry, currently contributed to by 11 North American institutions with expertise in the treatment of patients with thyroid carcinoma. The registry follows a large cohort of patients with DTC with the primary endpoint of assessing the effects of initial treatment strategies and management on long-term outcomes. The most recent overall analysis

from this registry reported on outcomes through 2001 from nearly 3000 patients with a median follow-up of only 3 years (6). With more than a decade of further observations and nearly 5000 patients with DTC enrolled, we present an updated analysis of outcomes after primary therapy and long-term THST.

## Patients and Methods

### Patients and data collection

The data collection, data management, and analysis methods of the registry have been described in previous publications (3, 7, 17–21). Institutional review boards of participating centers approved the study, and ongoing oversight of the registry occurs through The University of Texas MD Anderson Cancer Center, where the central database is currently managed and maintained.

Demographic, clinical, histological, and radiological data were collected and entered into a PC-based clinical data management system locally (Medlog, version 2012–5; Incline Village, NV) and transmitted to the managing site. Patient care was determined by the individual physician independent of registry participation. All treatments administered within 6 months of the first surgery were considered “initial therapies,” and follow-up data were reported annually. Disease stage was assigned according to the previously described registry staging system (3) (see [Supplemental Appendix 1](#)); central pathology review was not performed. Clinical status at entry was assigned by the treating physician (at completion of initial therapies) as either disease-free or not disease-free, with remaining tumor extent and/or metastases detailed. Disease monitoring and identification of structural disease recurrence were determined by the treating physicians per local standard of care. Where possible, mortality events were confirmed through the Social Security Death Index for non-Canadian sites and the Office of the Registrar General of Ontario through Cancer Care Ontario for Canadian patients.

### Cohort definitions

The overall cohort contains 4941 patients diagnosed with DTC and registered between January 1987 and November 2012. Within the overall cohort, the THST cohort comprises 3268 patients and only includes those with TSH values recorded at least 50% or more of their follow-up time (7, 17). TSH scores generated for the THST cohort as previously described represent an assessment of therapy throughout the follow-up period (7, 17). Second- or third-generation TSH assays were in use at each institution's clinical laboratory, with functional sensitivities of at least 0.1 mU/L. An undetectable serum TSH was defined by the clinical laboratory at each participating institution. At each follow-up visit, serum TSH levels were categorized into one of four groups and assigned a score: TSH undetectable (TSH score = 1), TSH subnormal but detectable (TSH score = 2), TSH normal

Department of Endocrine Neoplasia and Hormonal Disorders (A.A.C., D.R.L., S.I.S.), The University of Texas MD Anderson Cancer Center, Houston, Texas 77030; Thyroid Unit (D.S.R.), Massachusetts General Hospital, Boston, Massachusetts 02114; Division of Endocrinology (J.J.), Department of Medicine, Georgetown University Medical Center, Washington, DC 20057; Division of Endocrinology and Metabolism (D.S.C., P.W.L., M.X.), The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205; Department of Radiation Oncology (J.D.B.), Princess Margaret Hospital, Toronto, ON M5G 2M9, Canada; Department of Internal Medicine (K.B.A.), Veterans Affairs Medical Center and University of Kentucky, Lexington, Kentucky 40502; Division of Endocrinology and Metabolism (H.G.F.), Sinai Hospital, Baltimore, Maryland 21215; Division of Endocrinology, Metabolism, and Diabetes (B.R.H.), University of Colorado School of Medicine, Aurora, Colorado 80045; Genzyme (J.M.), a Sanofi Company, Cambridge, Massachusetts 02142; Diabetes, Endocrinology, Obesity Branch (M.C.S.), National Institutes of Health, Bethesda, Maryland 20892; Departments of Head and Neck Surgery (D.L.S.) and Nuclear Medicine (H.R.M.), University of Cincinnati Medical Center, Cincinnati, Ohio 45219

(TSH score = 3), and TSH elevated (TSH score = 4). To account for the variations in serum TSH levels over time, a mean TSH score was then calculated from all recorded TSH scores throughout the course of treatment and follow-up. Patients were grouped for analysis according to their mean TSH score and categorized as follows: patients with a mean TSH score of 1.0–1.99, which corresponded to aggressive THST (undetectable to subnormal TSH levels); patients with a mean TSH score of 2.0–2.99, which corresponded to moderate THST (subnormal to normal TSH levels); and patients with a mean TSH score of 3.0–4, which corresponded to nonsuppressed THST (normal to elevated TSH levels) throughout the course of follow-up. To examine the effect of THST on outcomes after diagnosis of distant metastatic disease, whether at initial staging or during subsequent follow-up, mean TSH scores were also calculated, starting from the date of metastatic disease detection through the remaining course of their follow-up.

### Statistical analysis

In descriptive analyses, DTC histologies included papillary thyroid cancer (PTC), follicular thyroid cancer (FTC), or Hürthle cell variant (HCC). All DTC patients were aggregated for outcomes analysis. Endpoints examined were overall survival (OS) and disease-free survival (DFS). In our analysis of DFS, recurrent disease and death from any cause were considered events. Recurrent disease was defined as structural evidence of disease determined either radiographically or by pathology. Recurrence was analyzed only among patients who were considered disease-free at entry based on local clinical practice patterns. Of note, in no analysis of outcome after primary treatment was a significant benefit seen in disease-specific survival that was not also seen in OS. Therefore, only OS analyses are presented. Therapies examined as potential predictors of outcomes were a lesser surgical resection vs T/NTT, no RAI vs RAI within 6 months postoperatively, and increasing degrees of THST during follow-up as reflected in TSH scores. All patients with follow-up were included in OS analysis, whereas only those patients reported as having no residual disease after initial therapy were included in DFS analysis. Nominal data were examined using  $\chi^2$  analyses. Univariate predictors of OS and DFS were determined using product-limit survival analysis and the log-rank statistic. The relative contribution of each univariate predictor, represented with a relative risk ratio (RR) and corresponding 95% confidence interval (CI), was determined using proportional hazards multivariate regression modeling. Throughout this report, a  $RR < 1$  indicates an improved outcome associated with the more extensive initial therapy (ie, greater surgical extent, RAI administration) and/or higher degree of THST.

Potential selection bias in the use of therapies was examined using propensity analysis (22, 23). Covariates applied to the regression model as potential contributors to discrepant therapy application were sex, age, histology, tumor size, extraglandular invasion, neck metastases, surgical extent, reporting institution, and year of diagnosis. *P* values of .05 or less were considered statistically significant. All analyses were performed using the SAS JMP 10 statistical software package (version 10.0.0).

## Results

### Description of cohorts

Among the 4941 DTC patients in the overall cohort, 88% had PTC, 8% had FTC, and 4% had HCC. Median

follow-up duration was 6 years (range, 0–25 y) with a total of 34 631 person-years of documented follow-up. Only 94 patients, 1.9% of the cohort, lacked any follow-up information. The overall cohort characteristics are summarized in Table 1. The median follow-up time in years (range) by stage was as follows: stage I, 6.6 (0–25.0); stage II, 6.0 (0–24.3); stage III, 6.5 (0–23.0); and stage IV, 4.2 (0.1–24.0). Among the 3649 patients considered disease-free after initial therapy, 933 (26%) were diagnosed with structural recurrent disease occurring a median of 1.2 years (range, 0.2–21) after diagnosis, either detected by imaging or confirmed by pathology report. Local or regional recurrence accounted for 74% of recurrences (60% regional, 14% local), and distant metastases accounted for 11% of recurrences. In 16% of patients reported to have recurrence, the site was not specified. Five-year OS after recurrence was 91% for either local or regional recurrent disease, as compared with 81% after recurrence as distant metastases.

Characteristics of the THST cohort are summarized in Table 1. Among the 3238 patients in this cohort, the mean number of TSH values per patient was 6.2 (range, 1–24; median, 6.0). The clinical characteristics of this THST cohort were compared with a remaining cohort in which sufficient TSH data were not reported (Table 1). Among those patients examined in OS analysis, the “remaining cohort” had more FTC/HCC patients than the THST cohort and more stage IV patients, was less likely to have received RAI, and had more deaths. Among those examined in DFS analysis, the “remaining cohort” had fewer patients over 45 years of age, had more FTC/HCC patients, was less likely to have received RAI, and had fewer deaths than the THST cohort. Of the remaining cohort, 40% had no TSH values recorded.

Initial disease stage was a significant predictor of OS as well as DFS (Figure 1, A and B). No significant difference was observed in analysis of treatment outcomes among the histological subtypes. Therefore, all three histological subtypes were combined for subsequent analyses, as in previous NTCTCS reports.

### Effect of surgical extent and RAI in the overall cohort by stage

In univariate analyses in stage I patients (Supplemental Appendices 2 and 3), T/NTT and RAI were each associated with decreased DFS but without change in OS. Propensity analysis was subsequently performed to understand the impact of variations of clinicopathological presentation within the stage I group. No difference was demonstrated in DFS among the strata according to surgical extent (Supplemental Appendix 4); a similar lack of difference was seen in propensity analysis for RAI treat-

**Table 1.** Comparison of Clinical Characteristics of Analyzed Cohorts

Parameter	OS				DFS			
	Overall Cohort	THST Cohort	Remaining Cohort	P Value	Overall Cohort	THST Cohort	Remaining Cohort	P Value
n	4941	3238	1703		3649	2472	1177	
Sex								
Male	27	27	29		25	75	74	
Female	73	73	71	.10	75	25	26	.66
Age at diagnosis								
<45 y	52	52	53		53	52	56	
≥45 y	48	48	47	.38	47	48	44	.05
Histology								
PTC	88	89	86		89	89	87	
FTC	8	8	9		7	7	8	
HCC	4	4	5	.007	4	4	5	.024
NTCTCS stage								
I	43	44	43		49	48	51	
II	27	27	26		30	31	29	
III	24	25	24	.004	20	21	20	.38
IV	5	4	7		0	0	0	
Surgical treatment								
NTT	86	86	85		86	86	86	
Other	14	14	15	.17	14	14	14	.75
RAI administered activity, mCi								
Yes	74	75	71		72	73	69	
≤30	8	8			9	9	8	
31–75	7	8			9	10	4	.0055
>75	58	58			54	53	56	
No	26	25	29	.0017	28	27	31	
TSH score category								
1.0–1.9		32				30		
2.0–2.9		59				61		
3.0–3.9		9				9		
Deaths, n	426	210	216	<.0001	214	115	99	<.0001
Recurrences, n					933	668	265	.003
Follow-up time, y median (range)	6.2 (0–25.0)	6.9 (0–25.0)	4.6 (0–24.6)		5.5 (0–25.0)	6.1 (0.2–25.0)	4.2 (0–24.6)	

Data are expressed as percentage of patients unless specified otherwise. All parameter comparisons used Pearson's  $\chi^2$  test.

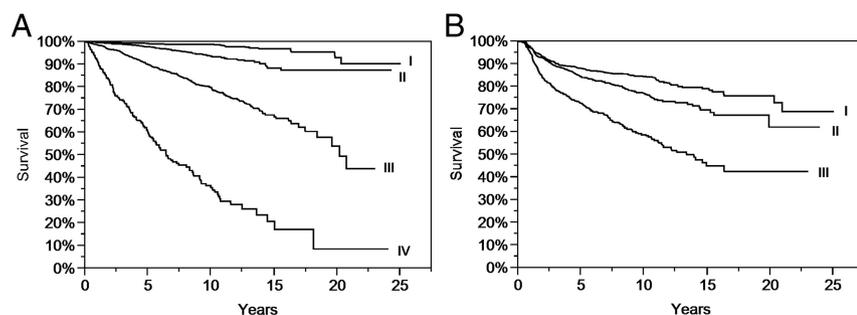
ment application (Supplemental Appendix 5). In the logistic regression models that were the bases for the propensity analyses, significant covariates associated with T/NTT were histology, tumor size, neck metastases, reporting institution, and diagnosis year. Significant covariates associated with administration of RAI were age,

tumor size, extraglandular invasion, neck metastases, surgical extent, reporting institution, and diagnosis year.

By univariate analysis, T/NTT was significantly associated with improved OS in stage III patients, with similar RRs in stages II and IV that did not meet significance (Supplemental Appendix 2). RAI was also associated with

improved OS in stage III patients, with similar but nonsignificant RRs in stages II and IV (Supplemental Appendix 3). Additionally, RAI was associated with improved OS in stage III patients, with similar but nonsignificant RRs in stages II and IV (Supplemental Appendix 3). Additionally, RAI was associated with improved DFS in stage II and with similar RR, although not significantly different, in stage III.

Table 2 summarizes subsequent multivariate analysis of T/NTT and RAI in the overall cohort. Although T/NTT no longer remained independently predictive of improved OS among stage III patients, RAI re-



**Figure 1.** A, Product limit estimates (with log-rank statistic) of OS after diagnosis of DTC by registry at entry ( $P < .0001$ ). B, Product limit estimates (with log-rank statistic) of DFS after diagnosis of DTC by registry at entry ( $P < .0001$ ).

**Table 2.** Multivariate Analyses of Outcomes After Initial Treatment Therapies, Overall Cohort

	OS				DFS			
	RR	95% CI	P RR <sup>a</sup>	P Model <sup>b</sup>	RR	95% CI	P RR <sup>a</sup>	P Model <sup>b</sup>
Stage I								
Any RAI vs none	0.79	0.35–1.89	.58	.50	1.79	1.28–2.56	.0005	<.0001
T/NTT vs other	2.04	0.65–9.09	.24		1.52	0.96–2.50	.07	
Stage II								
Any RAI vs none	0.67	0.36–1.28	.22	.13	0.70	0.49–1.01	.053	.11
T/NTT vs other	0.64	0.32–1.41	.25		0.92	0.58–1.52	.72	
Stage III								
Any RAI vs none	0.66	0.46–0.98	.04	.01	0.84	0.57–1.28	.40	.36
T/NTT vs other	0.70	0.47–1.08	.10		0.79	0.53–1.25	.31	
Stage IV								
Any RAI vs none	0.70	0.46–1.10	.12	.049				
T/NTT vs other	0.66	0.41–1.11	.11					

<sup>a</sup> P value of the RR related to the parameter tested.

<sup>b</sup> P value for the proportional hazards model that tested RAI vs none and T/NTT vs other.

mained a significant predictor. Among stage IV patients, the model that combined T/NTT and RAI was associated with improvement in OS, although neither parameter was independently significant. Neither treatment demonstrated an improvement in DFS.

### Effect of surgical extent, RAI, and THST in the THST cohort

In the smaller THST cohort (Supplemental Appendices 6 and 7), the association of RAI with reduced DFS did not reach statistical significance among stage II patients. T/NTT and RAI were associated with significantly improved OS among stage III patients, and RAI was associated with improved OS in stage IV. DFS appeared to be worse in stage I patients treated with more extensive surgery and/or RAI.

THST was associated with improved OS and DFS across all stages for mean TSH scores in the moderate range. However, no further improvement in OS or DFS in any stage was demonstrated by TSH levels averaging in the aggressive range (Supplemental Appendix 8 and Figure 2, A–D).

To examine further the predictive value of THST and stage as independent covariates, these parameters were entered into a proportional hazards regression model (Table 3). Each stage had a significantly higher risk of death and recurrent disease than the prior stage, demonstrating the independent as well as combined association of stage and THST with OS and DFS.

To assess the role of longitudinal THST in treating patients with distant metastases, mean TSH scores were separately calculated beginning from the first report of metastatic disease. Mean TSH scores after diagnosis of metastatic disease in the moderate range were associated with a significant survival benefit (Supplemental Appen-

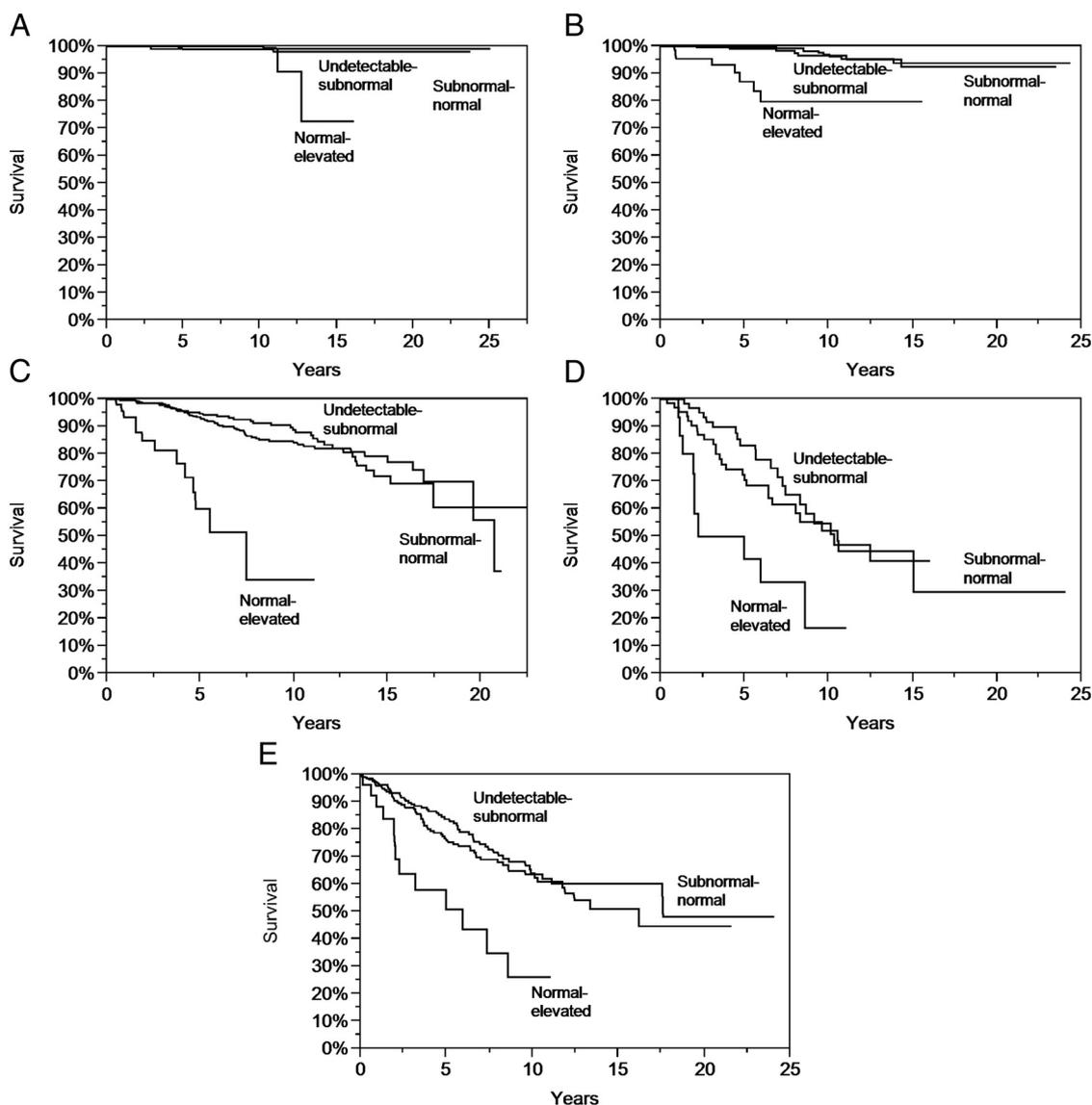
dix 9). No further improvement in OS was observed with TSH levels more aggressively suppressed to the undetectable to subnormal range (Figure 2E).

Treatments significantly associated with outcomes by univariate analyses were examined in multivariate models to identify independent predictors and provide adjusted RRs (Table 4). Across all stages, only moderate THST was an independent predictor of both improved OS and DFS. No additional benefit in OS or DFS was noted from more aggressive THST.

### Outcomes associated with continued THST during follow-up

To evaluate the optimal duration of THST, we examined the effect of continuing THST beyond 1, 3, and 5 years of follow-up by calculating mean TSH scores over the course of each patient's remaining follow-up (Supplemental Appendices 10–12). In multivariate analysis including initial disease stage, continued moderate THST was associated with better OS and DFS for at least 3 years. After 5 years of follow-up, our analysis did not show improved outcomes with continued moderate THST, but considering both the smaller number of patients and fewer events in this group, continued surveillance would be necessary to draw meaningful conclusions regarding duration of THST beyond 5 years.

To identify the contribution of initial disease stage and THST as independent predictors, both parameters were entered into a proportional hazards regression model (Supplemental Appendices 13–15). After 1 year, initial stage and moderate THST were both independent predictors of OS. After 3 years, initial stage and moderate THST remained independent predictors of OS, and after 5 years of follow-up, although initial stage remained a predictor



**Figure 2.** A, Product limit estimates (with log-rank statistic) of OS according to mean TSH category among stage I patients ( $P = .002$ ). B, Product limit estimates (with log-rank statistic) of OS according to mean TSH category among stage II patients ( $P < .0001$ ). C, Product limit estimates (with log-rank statistic) of OS according to mean TSH category among stage III patients ( $P < .0001$ ). D, Product limit estimates (with log-rank statistic) of OS according to mean TSH category among stage IV patients ( $P = .003$ ). E, Product limit estimates (with log-rank statistic) of OS according to mean TSH category after diagnosis of distant metastases ( $P < .0003$ ).

of OS, no benefit was observed with any particular subsequent degree of TSH suppression.

## Discussion

This analysis of a large multicenter registry examines demographics, initial therapy, and outcomes of 4941 patients with DTC. As compared with the previous registry report (7), more than 2000 additional patients have been included, median duration of follow-up has been extended from 3 to 6 years, and total documented follow-up has increased from 10 994 to 34 631 patient-years. The THST cohort now comprises two-thirds of the overall cohort, with twice as many patients and average number of TSH

values per patient. However, the cohort is otherwise comparable with previous reports, with similar characteristics of gender, age of diagnosis, histological subgroups, stage at entry, initial surgical treatment, or RAI-administered activity (Table 1). Of note, fewer patients received aggressive THST and more patients received moderate THST than in the previous reports, suggesting a trend to more conservative therapy. Outcomes analyses are strengthened by recording more than twice the number of deaths and nearly four times the number of recurrences.

Our current analyses support previously reported findings of improved outcomes with T/NTT and RAI in stage III and IV patients. For lower-risk groups, we previously reported a benefit in OS in stage II patients who underwent

**Table 3.** Outcomes Associated With Mean TSH Scores: Multivariate Analysis of Stage and THST

	OS				DFS				
	RR	95% CI	P RR <sup>a</sup>	P Covariate <sup>b</sup>	RR	95% CI	P RR <sup>a</sup>	P Covariate <sup>b</sup>	P Model <sup>c</sup>
Stage				<.0001				<.0001	<.0001
I/II	4.38	2.17–9.57	<.0001		1.50	1.14–1.95	.0034		
III/II	5.30	3.48–8.38	<.0001		2.06	1.58–2.71	<.0001		
IV/III	3.30	2.36–4.56	<.0001						
Mean TSH score				<.0001				<.0001	
2.0–2.9 vs 3.0–4.0	0.17	0.12–0.27	<.0001		0.32	0.24–0.44	<.0001		
1.0–1.9 vs 2.0–2.9	0.95	0.70–1.30	.77		1.25	0.98–1.59	.07		

<sup>a</sup> P value of the RR of the level of the parameter tested.

<sup>b</sup> P value of the parameter within the proportional hazards model.

<sup>c</sup> P value for the proportional hazards model that tested stage and mean TSH score.

T/NTT (RR = 0.57). In our current analysis, the effect size did not appreciably change (RR = 0.56), but the survival advantage did not reach statistical significance. Despite the addition of many more patients and events (now totaling 51 deaths among 1104 patients in this stage), the CI is wider, suggesting that other factors may account for this increased variability. Consistent with our previous report, RAI is associated with OS benefit in stage III patients, with a tendency toward benefit in stage IV patients; this effect persists, independent of the extent of surgery. In the smaller THST cohort, postoperative treatment with RAI also conferred significant OS benefit among stage III and IV patients in univariate analysis. However, in contrast with our previous report, further analysis of the THST cohort demonstrated that RAI is not an independent pre-

dictor of OS in high-risk patients; in fact, when examined in a regression model, only THST remained a significant and independent predictor of OS among patients in stages II, III, and IV. Given that the reporting institution was one of the most significant covariates associated with both RAI and surgery, it is likely that considerable local selection bias contributed to the choice of both surgical extent and RAI use in stage I patients.

In agreement with previous reported registry findings, RAI was not associated with OS benefit in low-risk patients, and as we have previously noted (7), RAI appeared to be associated with worse DFS in stage I patients. Other large studies have also demonstrated that postoperative RAI therapy has no significant effect on long-term outcomes in most low-risk DTC patients (24–27). One pos-

**Table 4.** Multivariate Analyses of Outcomes After Initial Treatment Therapies, THST cohort

	OS				DFS			
	RR	95%CI	P RR <sup>a</sup>	P Model <sup>b</sup>	RR	95%CI	P RR <sup>a</sup>	P Model <sup>b</sup>
Stage I								
RAI vs none	0.53	0.12–2.38	.41	.10	1.79	1.14–2.86	.01	<.0001
T/NTT vs other	1.28	0.29–9.09	.76		2.17	1.16–4.55	.01	
1.0–1.9 vs 2.0–2.9	2.08	0.38–11.1	.37		1.28	0.82–2.00	.27	
2.0–2.9 vs 3.0–4.0	0.10	0.02–0.60	.01		0.35	0.21–0.61	.0004	
Stage II								
RAI vs none	1.92	0.65–6.67	.24	<.0001	0.94	0.58–1.59	.82	.07
T/NTT vs other	1.33	0.40–6.25	.66		1.92	0.88–5.00	.10	
1.0–1.9 vs 2.0–2.9	1.20	0.36–3.57	.75		1.18	0.71–1.85	.52	
2.0–2.9 vs 3.0–4.0	0.05	0.02–0.15	<.0001		0.37	0.19–0.76	.009	
Stage III								
RAI vs none	0.87	0.50–1.61	.65	<.0001	1.02	0.60–1.85	.96	<.0001
T/NTT vs other	0.57	0.34–1.00	.05		0.76	0.44–1.41	.36	
1.0–1.9 vs 2.0–2.9	0.90	0.56–1.41	.66		1.19	0.73–1.89	.48	
2.0–2.9 vs 3.0–4.0	0.15	0.08–0.29	<.0001		0.19	0.11–0.34	<.0001	
Stage IV								
RAI vs none	0.63	0.33–1.30	.20	.032				
T/NTT vs other	1.03	0.48–2.56	.93					
1.0–1.9 vs 2.0–2.9	0.94	0.51–1.72	.84					
2.0–2.9 vs 3.0–4.0	0.30	0.14–0.73	.01					

<sup>a</sup> P value of the RR of the parameter tested.

<sup>b</sup> P value for the proportional hazards model that tested RAI vs none, T/NTT vs other, and mean TSH score category.

sible explanation may be selection bias influenced by a number of factors the treating physician takes into account when making the decision to administer RAI, as suggested in recent survey data (13). Given the apparent worse outcomes in stage I with RAI, we performed propensity analysis attempting to account for potential covariates; we found, with the exception of sex and histology, all covariates examined influenced the likelihood of a patient receiving RAI, including the individual reporting institution. Thus, the worse DFS seen in stage I patients treated with RAI appears to be a function of the underlying criteria used by treating physicians to select patients for RAI. Importantly, however, no stage I subgroup is identified that demonstrates any improved outcome after RAI, providing further support for limited RAI adjuvant therapy in stage I patients, tailoring to individual risk levels (5).

Despite the potential for risk associated with THST (28–37), we report for the first time, in multivariate analysis of primary treatments for DTC across all stages, that only THST was associated with both improved stage-adjusted OS and DFS, and when further examining the degree of THST, aggressive THST conferred no additional survival advantage as compared with moderate THST. Previous analysis noted only OS benefit with moderate THST, whereas current analysis also demonstrates a DFS advantage. In contrast with our previous report, aggressive suppression conferred no additional survival advantage in the high-risk stages as compared with moderate THST. Even limiting the analysis to patients with distant metastatic disease, maintaining mean TSH levels in the moderate range rather than aggressive is still associated with the best outcomes.

These observations greatly strengthen recent reports that suggest a lack of benefit from more aggressive THST. One retrospective analysis of patients with advanced thyroid cancer with serum TSH concentrations that were above 0.1 mIU/L showed decreased disease-specific survival compared with those with a serum TSH concentration less than 0.1 mIU/L (16). However, a recent observational study demonstrated decreased survival of DTC patients when their serum TSH was undetectable as compared with those who had less suppressed TSH levels, noting that the risk of cardiovascular and all-cause mortality was increased independent of age, sex, and cardiovascular risk factors (38). Furthermore, a retrospective analysis of several hundred patients followed for 9 years reported extremely low mortality and recurrence rates if median serum TSH was maintained moderately suppressed with concentrations of less than 2 mIU/L as compared with more aggressive levels of suppression (39). In a small prospective randomized trial of THST in DTC (40), patients were randomized to suppressed TSH or normal reference

range TSH, and after a mean follow-up of nearly 7 years, no differences in DFS were observed between the treatment groups. However, most patients in this study underwent lobectomy without RAI ablation.

There are limitations in analysis of the registry database that have been previously noted. We acknowledge that the nature of a multicenter longitudinal registry has limitations such as incomplete data collection, under- or overrepresentation of certain sociodemographic groups, and limited geographical/population coverage. There exists the potential for institutional bias related to the treating physician's characterization of disease status at entry or selection of treatment, which may have resulted in overcharacterization of disease-free status after initial treatment leading to higher classification of recurrent disease. However, we feel this potential bias has been minimized due to the number of participating institutions. Furthermore, primary conclusions drawn from the registry analysis depend on OS rather than DFS, negating the potential bias in assessing disease-free status. Another possible limitation of our analysis may be that all histological classifications of PTC and FTC have been merged together because we were unable to identify differences in treatment effects between the histological groups. However, given changing patterns of histo- and cytopathological diagnoses over the past several decades, documented inter-institutional variations in distinguishing PTC from FTC, and commonalities in management of both PTC and FTC, combining all DTC tumors was performed for this registry analysis to provide the strongest power to detect treatment effects. Lastly, there may have been selection bias due to the fact that the "remaining cohort" differed from the THST cohort; however, we were limited to the data available for analysis of TSH suppression. Overall, despite the aforementioned limitations, we feel our registry results remain relevant and generalizable to a global population, and unique from other larger cancer registries in the inclusion of detailed long-term data that allow examination of DFS in addition to OS.

This analysis of the larger, more mature registry database extends and refines earlier observations regarding the impact of initial therapies on patient outcomes and further justifies the need for prospective, long-term, controlled studies. Our data confirm prior observations regarding survival benefit in high-risk groups treated with T/NTT and RAI as well as a lack of benefit of postoperative RAI therapy in low-risk patients. Importantly, we demonstrate that aggressive THST confers no additional improvement in OS and DFS compared with moderate THST, in contrast with our earlier support for aggressive THST in higher-risk patients. This observation is extended to support only moderate THST even in patients with distant meta-

static disease, regardless of whether diagnosed initially or subsequently during longitudinal follow-up. Additionally, we report for the first time that continued moderate THST is associated with improved OS and DFS for at least 3 years after diagnosis. Finally, we have demonstrated that the initial NTCTCS stage remains predictive of both OS and DFS throughout at least the first 5 years of follow-up.

## Appendices

1. Registry staging classification.
2. Outcomes following T/NTT: univariate analysis, overall cohort.
3. Outcomes following RAI: univariate analysis, overall cohort.
4. Propensity score analysis of surgical extent for stage I patients, overall cohort.
5. Propensity score analysis of (any) RAI therapy for stage I patients, overall cohort.
6. Outcomes following T/NTT: univariate analysis, THST cohort.
7. Outcomes following RAI: univariate analysis, THST cohort.
8. Outcomes associated with mean TSH scores: univariate analysis.
9. Outcomes (in patients with distant metastases) associated with mean TSH scores: univariate analysis.
10. Outcomes associated with mean TSH scores: univariate analysis following 1 year follow-up.
11. Outcomes associated with mean TSH scores: univariate analysis following 3 years follow-up.
12. Outcomes associated with mean TSH scores: univariate analysis following 5 years follow-up.
13. Multivariate analysis of mean TSH score categories and stage following 1-year follow-up.
14. Multivariate analysis of mean TSH score categories and stage following 3-year follow-up.
15. Multivariate analysis of mean TSH score categories and stage following 5-year follow-up.

## Acknowledgments

We, as principal investigators at each NTCTCS institution, thank the physicians and staff members who participated in the management and follow-up of these patients. We acknowledge the substantial contributions of the institutional research staff members who collected and submitted the data. We also appreciate the considerable assistance provided by Jeffrey Cui for the management of the NTCTCS databases. Finally, we acknowledge the efforts of the numerous physicians and scientists whose

contributions were critical to either the creation or the maintenance of the registry effort for many years.

Address all correspondence and requests for reprints to: Steven I. Sherman, MD, Department of Endocrine Neoplasia & Hormonal Disorders, The University of Texas MD Anderson Cancer Center, 1400 Pressler, Unit 1461, Houston, TX 77030. E-mail: [sisherma@mdanderson.org](mailto:sisherma@mdanderson.org).

The NTCTCS has been supported in part by research grants from Genzyme, a Sanofi company, and Pfizer, and by the University of Texas MD Anderson Cancer Center Support Grant (NCI Grant P30 CA016672).

Disclosure Summary: A.A.C., D.S.C., D.R.L., J.J., J.D.B., H.R.M., M.X., H.G.F., and M.C.S. have nothing to declare. K.B.A. has received research grant support from Genzyme, a Sanofi company. B.R.H. has received research funding from Veracyte and Genzyme, a Sanofi company, and a one-time honorarium. J.M. is an employee of Genzyme, a Sanofi company, and a shareholder in Sanofi. D.S.R. has received payments for consulting or honoraria from Genzyme, a Sanofi company, Novo Nordisk, Bayer/Onyx, and Eisai. D.L.S. has received payments for consulting or honoraria from Genzyme, a Sanofi company, Novo Nordisk, Bayer/Onyx, and Eisai. S.I.S. has research support from Genzyme, a Sanofi company, and Pfizer; consulting relationships with Bayer, Eli Lilly, Eisai, Exelixis, Novo Nordisk, and Veracyte; and has received honoraria from Genzyme and Onyx.

## References

1. Brierley JD, Panzarella T, Tsang RW, Gospodarowicz MK, O'Sullivan B. A comparison of different staging systems predictability of patient outcome. Thyroid carcinoma as an example. *Cancer*. 1997;79:2414–2423.
2. Kilfoy BA, Zheng T, Holford TR, et al. International patterns and trends in thyroid cancer incidence, 1973–2002. *Cancer Causes Control*. 2009;20:525–531.
3. Sherman SI, Brierley JD, Sperling M, et al. Prospective multicenter study of thyrocarcinoma treatment: initial analysis of staging and outcome. National Thyroid Cancer Treatment Cooperative Study Registry Group. *Cancer* 1998;83:1012–1021.
4. Elisei R, Molinaro E, Agate L, et al. Are the clinical and pathological features of differentiated thyroid carcinoma really changed over the last 35 years? Study on 4187 patients from a single Italian institution to answer this question. *J Clin Endocrinol Metab*. 2010;95:1516–1527.
5. Cooper DS, Doherty GM, Haugen BR, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2009;19:1167–1214.
6. Hundahl SA, Cady B, Cunningham MP, et al. Initial results from a prospective cohort study of 5583 cases of thyroid carcinoma treated in the United States during 1996. U.S. and German Thyroid Cancer Study Group. An American College of Surgeons Commission on Cancer Patient Care Evaluation study. *Cancer*. 2000;89:202–217.
7. Jonklaas J, Sarlis NJ, Litofsky D, et al. Outcomes of patients with differentiated thyroid carcinoma following initial therapy. *Thyroid*. 2006;16:1229–1242.
8. Grant CS, Stulak JM, Thompson GB, Richards ML, Reading CC, Hay ID. Risks and adequacy of an optimized surgical approach to the primary surgical management of papillary thyroid carcinoma treated during 1999–2006. *World J Surg*. 2010;34:1239–1246.
9. Al-Saif O, Farrar WB, Bloomston M, Porter K, Ringel MD, Kloos

- RT. Long-term efficacy of lymph node reoperation for persistent papillary thyroid cancer. *J Clin Endocrinol Metab.* 2010;95:2187–2194.
10. Tuttle RM, Ball DW, Byrd D, et al. Thyroid carcinoma. *J Natl Compr Canc Netw.* 2010;8:1228–1274.
  11. Udelsman R, Shaha AR. Is total thyroidectomy the best possible surgical management for well-differentiated thyroid cancer? *Lancet Oncol.* 2005;6:529–531.
  12. Pacini F, Schlumberger M, Dralle H, Elisei R, Smit JW, Wiersinga W. European consensus for the management of patients with differentiated thyroid carcinoma of the follicular epithelium. *Eur J Endocrinol.* 2006;154:787–803.
  13. Haymart MR, Muenz DG, Stewart AK, Griggs JJ, Banerjee M. Disease severity and radioactive iodine use for thyroid cancer. *J Clin Endocrinol Metab.* 2013;98:678–686.
  14. Mazzaferri EL, Kloos RT. Clinical review 128: current approaches to primary therapy for papillary and follicular thyroid cancer. *J Clin Endocrinol Metab.* 2001;86:1447–1463.
  15. McGriff NJ, Csako G, Gourgiotis L, Lori CG, Pucino F, Sarlis NJ. Effects of thyroid hormone suppression therapy on adverse clinical outcomes in thyroid cancer. *Ann Med.* 2002;34:554–564.
  16. Diessl S, Holzberger B, Mäder U, et al. Impact of moderate vs stringent TSH suppression on survival in advanced differentiated thyroid carcinoma. *Clin Endocrinol (Oxf).* 2012;76:586–592.
  17. Cooper DS, Specker B, Ho M, et al. Thyrotropin suppression and disease progression in patients with differentiated thyroid cancer: results from the National Thyroid Cancer Treatment Cooperative Registry. *Thyroid.* 1998;8:737–744.
  18. McLeod DS, Cooper DS, Ladenson PW, et al. Prognosis of differentiated thyroid cancer in relation to serum thyrotropin and thyroglobulin antibody status at time of diagnosis. *Thyroid.* 2014;24:35–42.
  19. Ross DS, Litofsky D, Ain KB, et al. Recurrence after treatment of micropapillary thyroid cancer. *Thyroid.* 2009;19:1043–1048.
  20. Jonklaas J, Noguerras-Gonzalez G, Munsell M, et al. The impact of age and gender on papillary thyroid cancer survival. *J Clin Endocrinol Metab.* 2012;97:E878–E887.
  21. Taylor T, Specker B, Robbins J, et al. Outcome after treatment of high-risk papillary and non-Hürthle-cell follicular thyroid carcinoma. *Ann Intern Med.* 1998;129:622–627.
  22. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med.* 1998;17:2265–2281.
  23. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med.* 1997;127:757–763.
  24. Hundahl SA, Fleming ID, Fremgen AM, Menck HR. A National Cancer Data Base report on 53,856 cases of thyroid carcinoma treated in the U.S., 1985–1995. *Cancer.* 1998;83:2638–2648.
  25. Hay ID, Thompson GB, Grant CS, et al. Papillary thyroid carcinoma managed at the Mayo Clinic during six decades (1940–1999): temporal trends in initial therapy and long-term outcome in 2444 consecutively treated patients. *World J Surg.* 2002;26:879–885.
  26. Durante C, Attard M, Torlontano M, et al. Identification and optimal postsurgical follow-up of patients with very low-risk papillary thyroid microcarcinomas. *J Clin Endocrinol Metab.* 2010;95:4882–4888.
  27. Vaisman F, Shaha A, Fish S, Michael Tuttle R. Initial therapy with either thyroid lobectomy or total thyroidectomy without radioactive iodine remnant ablation is associated with very low rates of structural disease recurrence in properly selected patients with differentiated thyroid cancer. *Clin Endocrinol (Oxf).* 2011;75:112–119.
  28. Mercurio G, Panzuto MG, Bina A, et al. Cardiac function, physical exercise capacity, and quality of life during long-term thyrotropin-suppressive therapy with levothyroxine: effect of individual dose tailoring. *J Clin Endocrinol Metab.* 2000;85:159–164.
  29. Hoftijzer HC, Heemstra KA, Corssmit EP, van der Klaauw AA, Romijn JA, Smit JW. Quality of life in cured patients with differentiated thyroid carcinoma. *J Clin Endocrinol Metab.* 2008;93:200–203.
  30. Biondi B, Fazio S, Cuocolo A, et al. Impaired cardiac reserve and exercise capacity in patients receiving long-term thyrotropin suppressive therapy with levothyroxine. *J Clin Endocrinol Metab.* 1996;81:4224–4228.
  31. Biondi B, Fazio S, Carella C, et al. Cardiac effects of long term thyrotropin-suppressive therapy with levothyroxine. *J Clin Endocrinol Metab.* 1993;77:334–338.
  32. Bauer DC, Rodondi N, Stone KL, Hillier TA. Thyroid hormone use, hyperthyroidism and mortality in older women. *Am J Med.* 2007;120:343–349.
  33. Flynn RW, Bonellie SR, Jung RT, MacDonald TM, Morris AD, Leese GP. Serum thyroid-stimulating hormone concentration and morbidity from cardiovascular disease and fractures in patients on long-term thyroxine therapy. *J Clin Endocrinol Metab.* 2010;95:186–193.
  34. Sugitani I, Fujimoto Y. Effect of postoperative thyrotropin suppressive therapy on bone mineral density in patients with papillary thyroid carcinoma: a prospective controlled study. *Surgery.* 2011;150:1250–1257.
  35. Sawin CT, Geller A, Wolf PA, et al. Low serum thyrotropin concentrations as a risk factor for atrial fibrillation in older persons. *N Engl J Med.* 1994;331:1249–1252.
  36. Bauer DC, Ettinger B, Nevitt MC, Stone KL. Risk for fracture in women with low serum levels of thyroid-stimulating hormone. *Ann Intern Med.* 2001;134:561–568.
  37. Cooper DS, Biondi B. Subclinical thyroid disease. *Lancet.* 2012;379:1142–1154.
  38. Klein Hesselink EN, Klein Hesselink MS, de Bock GH, et al. Long-term cardiovascular mortality in patients with differentiated thyroid carcinoma: an observational study. *J Clin Oncol.* 2013;31:4046–4053.
  39. Hovens GC, Stokkel MP, Kievit J, et al. Associations of serum thyrotropin concentrations with recurrence and death in differentiated thyroid cancer. *J Clin Endocrinol Metab.* 2007;92:2610–2615.
  40. Sugitani I, Fujimoto Y. Does postoperative thyrotropin suppression therapy truly decrease recurrence in papillary thyroid carcinoma? A randomized controlled trial. *J Clin Endocrinol Metab.* 2010;95:4576–4583.