Reassessing the NTCTCS Staging Systems for Differentiated Thyroid Cancer, Including Age at Diagnosis

Donald S.A. McLeod,[1,2,3] Jacqueline Jonklaas,[4] James D. Brierley,[5] Kenneth B. Ain,[6] David S. Cooper,[7]
Henry G. Fein,[8] Bryan R. Haugen,[9] Paul W. Ladenson,[7] James Magner,[10] Douglas S. Ross,[11]
Monica C. Skarulis,[12] David L. Steward,[13] Mingzhao Xing,[7] Danielle R. Litofsky,[14]
Harry R. Maxon,[15] and Steven I. Sherman[14]

Background: Thyroid cancer is unique for having age as a staging variable. Recently, the commonly used age cut-point of 45 years has been questioned.

Objective: This study assessed alternate staging systems on the outcome of overall survival, and compared these with current National Thyroid Cancer Treatment Cooperative Study (NTCTCS) staging systems for papillary and follicular thyroid cancer.

Methods: A total of 4721 patients with differentiated thyroid cancer were assessed. Five potential alternate staging systems were generated at age cut-points in five-year increments from 35 to 70 years, and tested for model discrimination (Harrell's C-statistic) and calibration (R^2). The best five models for papillary and follicular cancer were further tested with bootstrap resampling and significance testing for discrimination.

Results: The best five alternate papillary cancer systems had age cut-points of 45–50 years, with the highest scoring model using 50 years. No significant difference in C-statistic was found between the best alternate and current NTCTCS systems ($p = 0.200$). The best five alternate follicular cancer systems had age cut-points of 50–55 years, with the highest scoring model using 50 years. All five best alternate staging systems performed better compared with the current system ($p = 0.003$–0.035). There was no significant difference in discrimination between the best alternate system (cut-point age 50 years) and the best system of cut-point age 45 years ($p = 0.197$).

Conclusions: No alternate papillary cancer systems assessed were significantly better than the current system. New alternate staging systems for follicular cancer appear to be better than the current NTCTCS system, although they require external validation.

Introduction

PAPILLARY AND FOLLICULAR THYROID CANCERS (termed differentiated thyroid cancer) comprise most cases of thyroid malignancy. The majority of differentiated thyroid cancers have an excellent prognosis, although a subset of patients manifest aggressive disease. Optimal treatment requires accurate risk stratification both at diagnosis and throughout the disease course. Thyroid cancer staging at the time of initial treatment is therefore an important prerequisite to informed patient care.

Thyroid cancer is unique among malignancies for including age as a staging variable. Early prognostic studies recognized the association of younger age with excellent

[1]Department of Endocrinology and Diabetes, Royal Brisbane and Women's Hospital, Herston, Australia.
[2]Population Health Department, QIMR Berghofer Medical Research Institute, Herston, Australia.
[3]School of Medicine, University of Queensland, Herston, Australia.
[4]Division of Endocrinology, Department of Medicine, Georgetown University Medical Center, Washington, District of Columbia.
[5]Department of Radiation Oncology, Princess Margaret Hospital, Toronto, Canada.
[6]Department of Internal Medicine, Veterans Affairs Medical Center and University of Kentucky, Lexington, Kentucky.
[7]Division of Endocrinology, Diabetes, and Metabolism, The Johns Hopkins University School of Medicine, Baltimore, Maryland.
[8]Division of Endocrinology and Metabolism, Sinai Hospital, Baltimore, Maryland.
[9]Division of Endocrinology, Metabolism, and Diabetes, University of Colorado School of Medicine, Aurora, Colorado.
[10]Genzyme, a Sanofi Company, Cambridge, Massachusetts.
[11]Thyroid Unit, Massachusetts General Hospital, Boston, Massachusetts.
[12]Diabetes, Endocrinology, Obesity Branch, National Institutes of Health, Bethesda, Maryland.
Departments of [13]Head and Neck Surgery and [15]Nuclear Medicine, University of Cincinnati Medical Center, Cincinnati, Ohio.
[14]Department of Endocrine Neoplasia and Hormonal Disorders, The University of Texas MD Anderson Cancer Center, Houston, Texas.

survival, with some studies suggesting a worse prognosis in patients older than 40 years of age (1–4), or older than 50 years of age (5). In 1983, the second edition of the American Joint Committee on Cancer Manual for Staging Cancer first settled on a dichotomous cut-point of 45 years (6). This found support in a landmark study by Tubiana *et al*. (7), and the authors' group, the National Thyroid Cancer Treatment Cooperative Study (NTCTCS), adopted the cut-point of age 45 years for their unique staging system prior to registry inception in 1986 (Table 1) (8).

Several recent studies have suggested that this dogma of an altered prognosis after the age of 45 years should be reevaluated. In a previous analysis, the authors' group noted that most deaths in thyroid cancer patients appear to occur in those diagnosed after the age of 55 years (9). Other studies have variously suggested under-staging of certain patients younger than 45 years of age (10), or that mortality does not rise before an age at diagnosis of 50 years (11) or 55 years (12). In analysis of data from EUROCARE, the relative risk of death increased in the group of women older than 55 years of age (13).

This controversy is a good opportunity to reevaluate the NTCTCS staging system. In addition, if a different age cut-point could be documented in a large registry such as the NTCTCS, then the information might be useful to the American Joint Committee on Cancer (AJCC)/Union for International Cancer Control (UICC) groups working on the eighth editions of their staging systems.

Methods

Registry protocol and data collection

The data collection and analytical methods of the NTCTCS have been described elsewhere (8,9,14–20). Briefly, 11 North American centers contributed patient data, with registration beginning in January 1987 (this data analysis captures patients registered through to 2011). New patients were registered within three months of their initial surgery. Institutional Review Boards (IRB) of contributing centers approved the study, and ongoing oversight of the project occurs through the University of Texas MD Anderson Cancer Center IRB, where the central database is currently maintained.

Management of patients was nonrandomized and was solely at the discretion of their treating physicians on the basis of perceived best practice and clinical need at that period of time at their institution, independent of registry participation. Pre-specified baseline demographic, clinical, histologic, and radiologic data were entered into a PC-based clinical data management system locally (Medlog v2000-2, Incline Village, NV) and transmitted to the central registry database. Clinical status, investigations, and treatments were updated on a yearly basis. Where possible, the causes of death were reviewed and mortality data confirmed through the Social Security Death Index for U.S. patients and the Ontario Registrar General for Canadian patients.

Principles of restaging

The general principles of the NTCTCS staging system reevaluation were:

- Any alteration to the staging system must have been data-driven and internally valid. Any new staging systems devised were to be tested against the current staging system, specifically using statistical methods allowing for significance testing.
- Multiple age cut-points in addition to 45 years were assessed.
- Overall survival was the outcome of interest for primary analysis, to be consistent with the AJCC/UICC systems.
- Any new staging system devised should be easily transferable to Tumor Node Metastasis (TNM) format, for ease of external validation and potential adaption to other data sets.
- Papillary and follicular cancer were assessed separately, consistent with the current system.
- Current NTCTCS staging variables were assessed. The current NTCTCS staging system includes all anatomic/

TABLE 1. CURRENT NATIONAL THYROID CANCER TREATMENT COOPERATIVE STUDY STAGING CLASSIFICATION

	Papillary cancer		Follicular cancer	
	Age <45	*Age ≥45*	*Age <45*	*Age ≥45*
Primary tumor size				
<1 cm	I	I	I	II
1–4 cm	I	II	I	III
>4 cm	II	III	II	III
Primary tumor description				
Microscopic multifocal	I	II	I	III
Macroscopic multifocal or macroscopic tumor capsule invasion	I	II	II	III
Microscopic extraglandular invasion	I	II	I	III
Macroscopic extraglandular invasion	II	III	II	III
Poor differentiation	n/a	n/a	III	III
Metastases				
Cervical lymph node metastases	I	III	I	III
Extracervical metastases	III	IV	III	IV

The final stage is the highest of the individual primary tumor size, primary tumor description, and metastases scores. Age is in years.

pathologic variables collected by the registry (e.g., there is no distinction between central and lateral cervical lymph node status in the data entry; thus any future NTCTCS staging system can only specify general cervical lymph node status, and maximal tumor diameter is recorded as <1 cm, 1–4 cm, or >4 cm).

- No other nonanatomic variables, such as sex, were considered (as per current AJCC/UICC practice), with the exception of an exploratory analysis incorporating the tall-cell variant of papillary thyroid cancer.

Using these principles, alternate potential staging systems were created for testing against the current system.

Specific approaches to creating alternate staging systems

First the data set was separated into patients with papillary thyroid cancer and those with follicular thyroid cancer. Cutpoints were then created for age from 35 to 70 years in five-year increments. Using Kaplan–Meier estimators, univariate overall 5- and 10-year survivals were generated in papillary cancer patients for each potential prognostic variable currently in the NCTCS, at all the age cut-points defined above. Independent significance of each prognostic variable was also assessed at all the age cut-points through Cox models. Based on these assessments, five potential alternate staging systems were created at each age cut-point (rather than one per cut-point, because the estimates may be subject to interaction, confounding, and colinearity). The rules for this were: (a) for each potential alternate staging system, there must be four stages (I–IV); and (b) plausibility of stage progression was ensured (e.g., while microscopic extraglandular extension and macroscopic extraglandular extension could have the same stage designation or macroscopic extension could be staged higher, microscopic extraglandular extension could not be staged above macroscopic).

Method of testing the potential alternate staging systems

To test the potential alternate staging systems for papillary cancer, Kaplan–Meier plots were generated and Cox proportional hazards models were performed on each potential alternate staging system, with the outcome of overall survival. From the potential alternate systems across all age cut-points, the best five were chosen for detailed analysis based on model discrimination to the overall survival data, specifically through the highest Harrell's C-statistic values (analogous to area under the curve in receiver operator characteristic curves) (21). For these highest-scoring staging systems and the current NTCTCS staging system, confidence intervals for the Harrell's C-statistic were generated using bootstrap resampling (22) with 1000 replicates, and hypothesis testing was performed comparing the alternate systems to the current NTCTCS staging system to identify whether any differences in the Harrell's C-statistic were statistically significant. Model calibration was assessed by calculating $R^2 = 1 - \exp([(L_0 - L_p) \times 2/n])$ (23), where L_0 is the log partial likelihood of the null model (no covariates), L_p is the log partial likelihood of the fitted model, and n is the number of subjects. The same assessment was performed for follicular cancer.

Results

There were 4721 patients with differentiated thyroid cancer and complete staging information registered through to 2011, representing a total of 31,356 patient-years of follow-up. Of these, 4159 had a diagnosis of papillary thyroid cancer and 562 had follicular thyroid cancer (including Hürthle cell cancers and poorly differentiated cancers without areas of papillary architecture). Table 2 summarizes demographic and clinical tumor characteristics of included patients. There were 277 deaths in the papillary thyroid cancer patients, and 104 deaths in patients with follicular cancer.

Generation and testing of potential alternate staging systems

The analyses of prognostic variable associations with survival (univariate and multivariate) at each age cut-point are detailed in the accompanying Supplementary Appendices (Supplementary Appendix S1 for papillary cancer and Supplementary Appendix S2 for follicular cancer; Supplementary Data are available online at www.liebertpub.com/thy). Accompanying these are the potential alternate staging systems that were constructed. Harrell's C-statistic and R^2 for each potential alternate staging system are listed in the

TABLE 2. DEMOGRAPHIC AND TUMOR CHARACTERISTICS OF NTCTCS COHORT

Parameter	Papillary cancer	Follicular cancer
Number of patients	4159	562
Years follow-up time (median [IQR])	5.9 [3.0–9.8]	5.6 [3.0–9.4]
Percent female (%)	73.4	67.7
Age (mean ± SD)	44.3 ± 15.3	50.6 ± 16.8
Prior radiation exposure (%)	4.4	4.5
Tumor size (%)		
<1 cm	19.5	2.4
1–4 cm	68.4	64.8
>4 cm	12.1	32.8
Multifocality and capsular invasion (%)		
None	58.4	65.8
Microscopic multifocal	22.7	15.0
Macroscopic multifocal or macroscopic tumor capsule invasion	18.9	19.2
Extraglandular extension (%)		
None	73.7	79.2
Microscopic	12.0	9.1
Gross	14.3	11.7
Poor differentiation (%)	1.0	14.1
Cervical lymph node metastases (%)	35.8	7.8
Distant metastases (%)	5.6	14.4

Papillary cancer includes 3205 conventional papillary cancers, 710 mixed follicular–papillary histology, 108 occult sclerosing, 132 tall-cell cancers, and 4 columnar cell cancers. Follicular cancer includes 320 well-differentiated follicular cell cancers, 192 Hürthle cell cancers, and 50 poorly differentiated cancers.

TABLE 3. BEST ALTERNATE THYROID CANCER STAGING SYSTEMS MODEL ASSESSMENT

Staging system	Harrell's C-statistic [CI]	Difference in C-statistic to current NTCTCS system	p-Value for C-statistic difference	R^2
Papillary cancer				
Current NTCTCS	0.820 [0.792–0.848]	—	—	0.108
Age 45, model A	0.829 [0.800–0.857]	0.009	0.20	0.118
Age 45, model B	0.830 [0.801–0.859]	0.010	0.15	0.118
Age 50, model A	0.831 [0.802–0.859]	0.011	0.22	0.122
Age 50, model B	0.832 [0.803–0.860]	0.012	0.20	0.122
Age 50, model C	0.827 [0.799–0.855]	0.007	0.39	0.117
Follicular cancer				
Current NTCTCS	0.800 [0.764–0.836]	—	—	0.252
Age 50, model C	0.828 [0.795–0.865]	0.028	0.014	0.275
Age 50, model D	0.831 [0.797–0.859]	0.031	0.003	0.271
Age 50, model E	0.831 [0.803–0.865]	0.031	0.004	0.272
Age 55, model A	0.829 [0.792–0.866]	0.028	0.025	0.271
Age 55, model C	0.827 [0.790–0.864]	0.027	0.035	0.268
Papillary cancer (exploratory analysis including tall-cell variant)				
Current NTCTCS	0.820 (0.792–0.848)	—	—	0.108
Age 45, model A	0.830 (0.801–0.858)	0.010	0.15	0.118
Age 45, model B	0.831 (0.802–0.859)	0.011	0.10	0.118
Age 50, model A	0.832 (0.803–0.860)	0.012	0.19	0.123
Age 50, model B	0.833 (0.804–0.862)	0.013	0.17	0.123
Age 50, model C	0.827 (0.799–0.855)	0.007	0.40	0.117

Model details are provided in Supplementary Appendices S1 and S2.

appendixes. Of the five best alternate models for papillary cancer, three models made use of an age cut-point of 50 years, while two models had a cut-point of 45 years. For follicular thyroid cancer, three models had an age cut-point of 50 years, and two models had a cut-point of 55 years (the sixth-best follicular thyroid cancer model had an age cut-point of 45 years). The models with highest Harrell's C-statistic also were the models with the highest R^2 for both papillary and follicular cancer (best alternate models are highlighted in bold font in the Supplementary Appendices).

Papillary cancer comparison of potential alternate systems against the current NTCTCS system

The Harrell's C-statistic and bootstrapped confidence intervals, and R^2 values for the current NTCTCS staging system and the best five alternate papillary thyroid cancer staging systems are detailed in Table 3. Table 3 also gives the p-values for the differences in C-statistic between the current NTCTCS system and each alternate system.

The best alternate staging system was model B, incorporating a cut-point of 50 years (Table 4). The best alternate staging system downstaged a substantial proportion of patients (2803 stage I patients, 866 stage II, 286 stage III, and 126 stage IV compared with 1911 stage I, 1219 stage II, 806 stage III, and 145 stage IV patients under the current staging system). Although the Harrell's C-statistic was numerically higher than the current NTCTCS system and visually the stages are better separated on Kaplan–Meier plots (Fig. 1A current system; Fig. 1B best alternate system), the difference was not statistically significant. The best alternate systems with cut-points age 50 and 45 years were also not significantly different for Harrell's C-statistic ($p = 0.320$).

Follicular cancer comparison of potential alternate systems against the current NTCTCS system

The Harrell's C-statistic and bootstrapped confidence intervals, and R^2 values for the current NTCTCS staging system and the best five alternate follicular thyroid cancer staging systems are detailed in Table 3. Table 3 also gives the p-values for the differences in C-statistic between the current NTCTCS system and each alternate system.

All five best alternate follicular thyroid cancer staging systems had statistically significantly better discrimination than the current NTCTCS system. The numerically best alternate staging system had a cut-point of 50 years (model E; Table 5). The best alternate staging system again downstaged

TABLE 4. BEST ALTERNATE STAGING SYSTEM FOR PAPILLARY THYROID CANCER (AGE 50, MODEL B)

	Age <50	Age ≥50
Primary tumor size		
<1 cm	I	I
1–4 cm	I	II
>4 cm	I	III
Primary tumor description		
Macroscopic multifocal or macroscopic tumor capsule invasion	I	II
Microscopic extraglandular invasion	I	II
Macroscopic extraglandular invasion	I	III
Metastases		
Cervical lymph node metastases	I	II
Extracervical metastases	II	IV

The final stage is the highest of the individual primary tumor size, primary tumor description, and metastases scores. Age is in years.
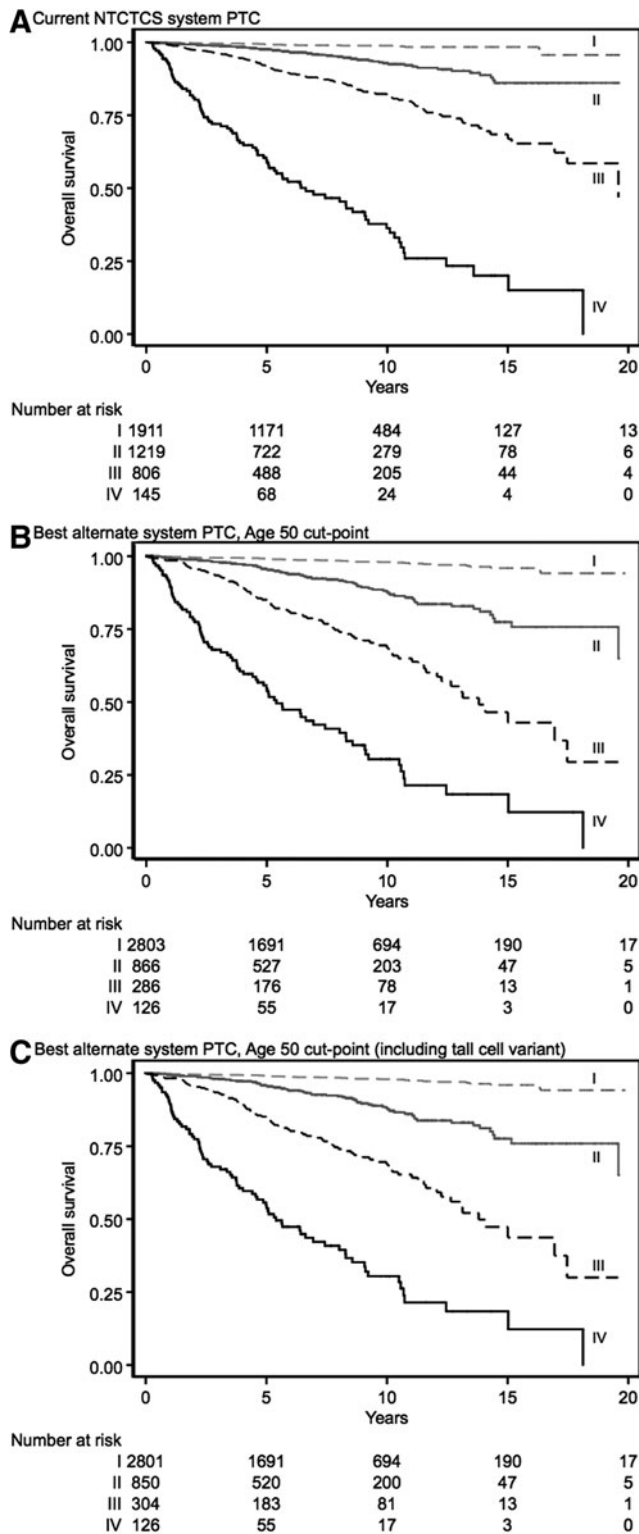
FIG. 1. Kaplan–Meier plots of overall survival for (**A**) current papillary thyroid cancer National Thyroid Cancer Treatment Cooperative Study (NTCTCS) staging system, (**B**) best alternate staging systems (with cut-point at age 50 years), and (**C**) exploratory analysis best alternate system.

TABLE 5. BEST ALTERNATE STAGING SYSTEM FOR FOLLICULAR THYROID CANCER (AGE 50, MODEL E)

	Age <50	Age ≥50
Primary tumor size		
<1 cm	I	I
1–4 cm	I	II
>4 cm	I	III
Primary tumor description		
Microscopic multifocal	I	II
Macroscopic multifocal or macroscopic tumor capsule invasion	I	II
Microscopic extraglandular invasion	I	II
Macroscopic extraglandular invasion	I	II
Poor differentiation	III	III
Metastases		
Cervical lymph node metastases	II	III
Extracervical metastases	III	IV

The final stage is the highest of the individual primary tumor size, primary tumor description, and metastases scores. Age is in years.

a substantial proportion of patients (233 stage I patients, 124 stage II, 129 stage III, and 70 stage IV compared with 144 stage I, 67 stage II, 270 stage III, and 75 stage IV patients under the current staging system). The Kaplan–Meier plots for the current follicular thyroid cancer and best alternate staging systems are shown in Figure 2A and B, respectively. When testing the best alternate system (cut-point age 50 years) with the best system using a cut-point of age 45 years (an identical staging system except the age cut-point), there was no significant difference (difference in Harrell's C-statistic = 0.006; $p = 0.197$). Testing the best alternate system (cut-point age 50 years) with the current papillary thyroid cancer NTCTCS system in follicular thyroid cancer patients (to assess the need for separate ongoing staging systems) suggested that the best alternate system (cut-point age 50 years) was significantly better (difference in Harrell's C-statistic = 0.02; $p = 0.048$).

Exploratory analysis—tall-cell variant of papillary cancer

As an exploratory analysis, the inclusion of the tall-cell variant of papillary thyroid cancer in potential staging systems was also examined (analogous to the way in which the staging system for follicular thyroid cancer includes the poorly differentiated subtype). This analysis was performed after finding that patients recorded as having tall-cell variant had an independently worse prognosis, particularly when diagnosed after 50 years (see Supplementary Appendix S3). The best amended staging system with tall-cell variant incorporated is similar to the staging system shown in Table 4, but tall-cell variant is added under primary tumor description with patients <50 years being stage I and patients ≥50 years being accorded stage III disease. The best alternate staging system in this case was model B, using the cut-point of age 50 years. The associated Kaplan–Meier plot is shown in Figure 1C. As can be seen from this figure, more patients are classified as stage III compared with the best alternate staging system not using tall-cell variant (Fig. 1B). This approach
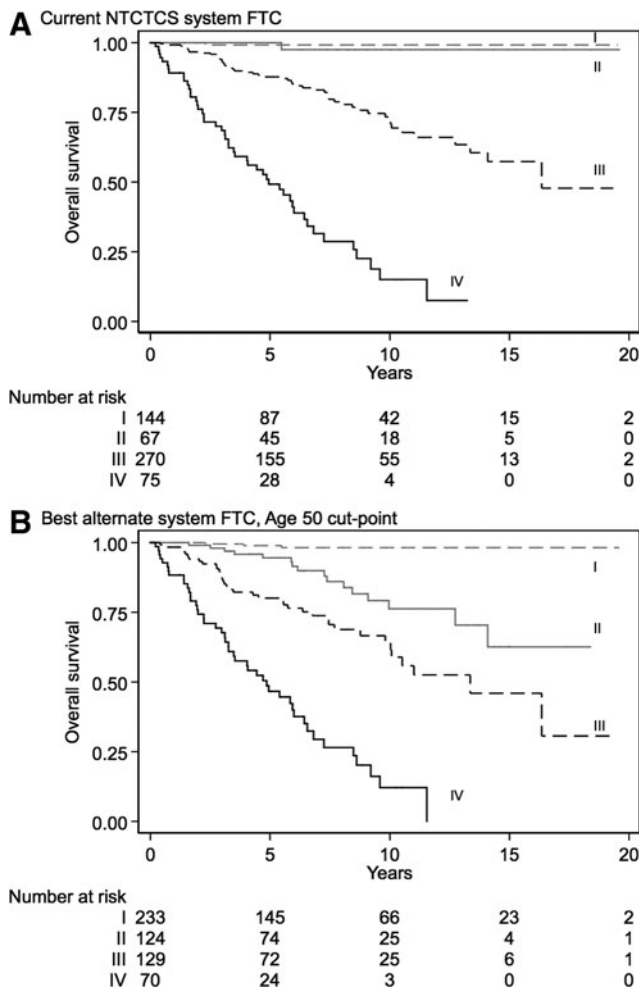
FIG. 2. Kaplan–Meier plots of overall survival for (**A**) current follicular thyroid cancer NTCTCS staging system versus (**B**) the best alternate staging system (with cut-point aged 50 years).

resulted in a marginally improved Harrell's C-statistic over the best alternate papillary thyroid cancer staging systems not incorporating the tall-cell variant (see Table 3), although it was not significantly better than the current NTCTCS papillary cancer staging system.

Discussion

Potential revision of staging system

Papillary thyroid cancer. For papillary thyroid cancer, alternate staging systems were constructed that predicted overall survival well on the NTCTCS data set. The best alternate staging systems had marginally higher numerical scores than the current staging system for discrimination and calibration, although none appeared significantly better. Given the lack of a statistically significant difference at this time, there is currently no plan to alter the NTCTCS papillary thyroid cancer staging system that was first devised in 1986, thereby allowing ease of backward comparability cross-analyses and publications.

Further use of subtypes of papillary cancer was not pursued in the main analysis because of potential lack of standardi-

zation of the histologic classification of papillary thyroid cancer subtypes across centers and eras, and the inability to review retrospectively pathology centrally across the NTCTCS cohort. This uncertainty about classification of the tall-cell variant is common to many analyses that are nevertheless reported and show a worse prognosis associated with tall-cell variant. An example is an analysis where use of the Surveillance, Epidemiology, and End Results (SEER) database precludes any central review of pathology (24). This decision not to incorporate tall-cell variant does not greatly alter the summary measures of model performance in the present results because subtypes such as the tall-cell variant are rare. In addition, adding a second nonanatomical factor in addition to age could further complicate stage group definitions. However, visually, the Kaplan–Meier plots show clearer separation of stages II and III (Fig. 1C), and this use of tall-cell variant could potentially improve prognostication in larger databases. Moreover, for an individual patient, including important histologic subtypes may improve prediction of a patient's prognosis, and the merits of this approach should be examined in future analyses of stage and proposals for thyroid cancer–specific prognostic factor tools.

There are good reasons for taking a conservative approach to revaluating cancer staging systems, and only adopting a new alternate system based on rigorous evidence. In the current case, while the new potential alternate papillary thyroid cancer staging systems scored numerically higher on the Harrell's C-statistic and R^2, none was demonstrably better than the old system, and there appear to be minimal differences in overall model fit for staging systems using either a 45 or 50 years cut-point for papillary thyroid cancer. In addition, the current NTCTCS and best alternatives had excellent discrimination, as evidenced by their Harrell's C-statistic being >0.80 (25). Retaining the current well-performing staging system, especially when faced with a lack of statistically superior alternative, allows continuity of analysis with previous studies, and easier benchmarking and comparison.

Recent studies have questioned the legitimacy of the age 45 years cut-point for thyroid cancer staging systems (11,12). The best alternate papillary cancer staging systems had dichotomous age cut-points of either 45 or 50 years, with very little objective difference in model performance. This indicates that current thyroid cancer staging systems likely have their age cut-point close to optimized. It is not surprising that the overall model fit is similar at these age cut-points because the cut-points are close together and mortality between ages 45 and 50 years is still low. Thus, most patients would have been staged the same under both systems. In the absence of a definitively better overall new staging system, there might still be an argument for altering future AJCC/IUCC staging systems. The present data show that a staging system using older cut-points of 50 or even 55 years can also generate excellent discrimination. Under current staging systems, patients aged just above 45 years may be "up-staged" because of tumor features, when in fact their mortality risk remains low. These patients therefore face potential harms from being "up/over-staged" if clinicians use the staging results to prescribe more aggressive adjuvant therapy (e.g., more aggressive primary surgery or radioiodine ablation therapy decisions based on cancer stage). A change in staging system may therefore be defensible if new staging systems using different age cut-points are at least as good as old

systems, yet expose fewer patients to unnecessary treatment. The authors are not aware of any decision analysis studies that have investigated this important question.

Follicular thyroid cancer. In contrast to the papillary thyroid cancer analysis, the new alternate follicular cancer staging systems performed significantly better than the current NTCTCS system. Prior to adopting a revised staging system for follicular thyroid cancer (e.g., model E, age 50; Table 5), the authors hope to perform external validation to confirm these results in an independent data set. Practically speaking, overall model characteristics were similar for cut-points between 45 and 55 years. As discussed above, considerations other than overall model fit may be appropriate in deciding age cut-points for future thyroid cancer staging systems. The current analysis also shows that the best alternate follicular cancer system (model E, age 50) had significantly higher discrimination than the current papillary thyroid cancer NTCTCS system when applied to follicular thyroid cancer patients. Separate staging systems for patients with papillary and follicular thyroid cancers will therefore apply.

Strengths and weaknesses

The robust and data-driven assessment of the staging system strengthened the analysis. A quantitative assessment of the models (Harrell's C-statistic) was also used, which enabled significance testing between the staging systems prognostic discrimination. The NTCTCS data set is large and mature, providing opportunity to assess staging systems in a cohort of more than 31,000 patient-years.

Quantitative comparison of survival models is a young and evolving field. Future statistical progress may yield new and improved methods in comparing discrimination and calibration for staging systems. At present, the authors are not aware of any available methods to compare significance of differences in model calibration generally, and few methods are available to test calibration in prognostic models with only four risk levels, as in the present case. The available calibration method, R^2, has drawbacks, including lack of comparability of values between different study populations because of differential censoring (23). As an illustration, the R^2 values that were calculated were much higher in the follicular cancer models because mortality was substantially higher in the follicular cancer cohort, leading to less-frequent censoring of patients.

The decision was made to reanalyze the staging system using a traditional TNM-like system with a dichotomous age cut-point, as opposed to a prognostic scoring system, such as the Metastasis, Age, Completeness of resection, Invasiveness, and Size (MACIS) score (26), or a more complicated TNM-like system with multiple age cut-points. The categorical data format of the registry does not easily lend itself to production of a continuous scoring system such as the MACIS score. It was also reasoned that reassessment of age cut-points in TNM-like differentiated thyroid cancer staging could inform future AJCC/UICC systems. While it is unlikely that sharp demarcation in thyroid cancer biology and thus prognosis exists by age, a dichotomous age cut-point for TNM-like staging systems is favored over a more complicated system with multiple age cut-points for simplicity and usability. However, it is possible that staging systems using multiple cut-points could better predict individual patient prognoses; this was not assessed this in the present analysis.

Staging thyroid cancer at disease onset also fails to take into account the initial response to treatment, which has been shown to be important in determining risk of persistent disease and future recurrence (27). However, all anatomic staging is performed at diagnosis, and this characteristic of staging systems is unlikely to change in the near future. Furthermore, anatomic staging and response to treatment criteria offer complimentary disease information, and both approaches should continue to be assessed and optimized.

The main analysis was limited to variables currently in the NTCTCS staging systems. However, there may be subtypes of differentiated thyroid cancer associated with poor prognosis, such as the tall-cell variant of papillary thyroid cancer (24,28,29). As mentioned, examination of subtypes of papillary cancer beyond a preliminary analysis was not pursued because of potential lack of standardization of the histologic classification, inability to submit pathology to central review, the rarity of tall-cell variant, and the anticipated lack of support for adding a second nonanatomical factor in addition to age. A similar argument could be made for not adding sex to staging systems, despite some evidence this is a relevant variable (13). However, for an individual patient, including important histologic subtypes, and considering sex, may assist clinicians to predict a patient's prognosis, and the merits of this approach should be examined in future analyses of stage and proposals for thyroid cancer specific prognostic factor tools such as MACIS.

This NTCTCS analysis incorporates patients cared for from 1987 through to 2011 in 11 institutions. There have been significant changes in the presentation and management of thyroid cancer over this period (30,31), which may mean that patients diagnosed today could have a different disease trajectory than historic patients captured by the registry. In addition, developing treatments such as kinase inhibitors used to delay disease progression (32,33) or for re-differentiation therapy (34) may substantially alter prognosis for high-risk patients in the future. The large number of centers contributing to this registry is both a potential weakness and strength. While this would inevitably lead to variations in patient assessment, treatment, and follow-up for the included patients, it is also possible that any resultant variation better reflects outcomes in the wider community, as opposed to a single-institution registry with uniform but potentially nongeneralizable follow-up protocols.

Future analyses and further review of NTCTCS staging system

External validation of these results is encouraged. If other sizable data sets find that alternate age 50 (or 55) years cut-points for papillary and follicular cancer staging systems also perform numerically or significantly better than the current NTCTCS systems, strong consideration will be given to adopting the new systems. Such validation may also be considered by the AJCC/UICC in developing the eighth editions of their staging systems.

It is planned to perform analyses using the new alternate staging systems in relation to prognosis after treatment in order to determine whether the new alternate staging systems

better separate prognosis by the type of surgery, presence or absence of radioiodine ablation, or degree of thyrotropin suppression. If this were the case, it would add further value to the new staging systems.

Acknowledgments

Author Disclosure Statement

References

1. Crile G Jr, Hazard JB 1953 Relationship of the age of the patient to the natural history and prognosis of carcinoma of the thyroid. Ann Surg **138:**33–38.

2. Woolner LB, Beahrs OH, Black BM, Mc CW, Keating FR Jr 1961 Classification and prognosis of thyroid carcinoma. A study of 885 cases observed in a thirty year period. Am J Surg **102:**354–387.

3. Halnan KE 1966 Influence of age and sex on incidence and prognosis of thyroid cancer. Three hundred forty-four cases followed for ten years. Cancer **19:**1534–1536.

4. 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.

5. Franssila KO 1975 Prognosis in thyroid carcinoma. Cancer **36:**1138–1146.

6. Beahrs OH, Myers MH 1983 Manual for Staging of Cancer. Second edition. Lippincott, Williams, & Wilkins, Philadelphia, PA.

7. Tubiana M, Schlumberger M, Rougier P, Laplanche A, Benhamou E, Gardet P, Caillou B, Travagli JP, Parmentier C 1985 Long-term results and prognostic factors in patients with differentiated thyroid carcinoma. Cancer **55:**794–804.

8. 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.

9. Jonklaas J, Nogueras-Gonzalez G, Munsell M, Litofsky D, Ain KB, Bigos ST, Brierley JD, Cooper DS, Haugen BR, Ladenson PW, Magner J, Robbins J, Ross DS, Skarulis MC, Steward DL, Maxon HR, Sherman SI 2012 The impact of age and gender on papillary thyroid cancer survival. J Clin Endocrinol Metab **97:**E878–887.

10. Tran Cao HS, Johnston LE, Chang DC, Bouvet M 2012 A critical analysis of the American Joint Committee on Cancer (AJCC) staging system for differentiated thyroid carcinoma in young patients on the basis of the Surveillance, Epidemiology, and End Results (SEER) registry. Surgery **152:**145–151.

11. Bischoff LA, Curry J, Ahmed I, Pribitkin E, Miller JL 2013 Is above age 45 appropriate for upstaging well-differentiated papillary thyroid cancer? Endocr Pract **19:**995–997.

12. Mazurat A, Torroni A, Hendrickson-Rebizant J, Benning H, Nason RW, Pathak KA 2013 The age factor in survival of a population cohort of well differentiated thyroid cancer. Endocr Connect **2:**154–160.

13. Micheli A, Ciampichini R, Oberaigner W, Ciccolallo L, de Vries E, Izarzugaza I, Zambon P, Gatta G, De Angelis R, Group EW 2009 The advantage of women in cancer survival: an analysis of EUROCARE-4 data. Eur J Cancer **45:**1017–1027.

14. Cooper DS, Specker B, Ho M, Sperling M, Ladenson PW, Ross DS, Ain KB, Bigos ST, Brierley JD, Haugen BR, Klein I, Robbins J, Sherman SI, Taylor T, Maxon HR 3rd 1998 Thyrotropin suppression and disease progression in patients with differentiated thyroid cancer: results from the National Thyroid Cancer Treatment Cooperative Registry. Thyroid **8:**737–744.

15. Taylor T, Specker B, Robbins J, Sperling M, Ho M, Ain K, Bigos ST, Brierley J, Cooper D, Haugen B, Hay I, Hertzberg V, Klein I, Klein H, Ladenson P, Nishiyama R, Ross D, Sherman S, Maxon HR 1998 Outcome after treatment of high-risk papillary and non-Hürthle-cell follicular thyroid carcinoma. Ann Intern Med **129:**622–627.

16. Jonklaas J, Sarlis NJ, Litofsky D, Ain KB, Bigos ST, Brierley JD, Cooper DS, Haugen BR, Ladenson PW, Magner J, Robbins J, Ross DS, Skarulis M, Maxon HR, Sherman SI 2006 Outcomes of patients with differentiated thyroid carcinoma following initial therapy. Thyroid **16:**1229–1242.

17. Ross DS, Litofsky D, Ain KB, Bigos T, Brierley JD, Cooper DS, Haugen BR, Jonklaas J, Ladenson PW, Magner J, Robbins J, Skarulis MC, Steward DL, Maxon HR, Sherman SI 2009 Recurrence after treatment of micropapillary thyroid cancer. Thyroid **19:**1043–1048.

18. Jonklaas J, Cooper DS, Ain KB, Bigos T, Brierley JD, Haugen BR, Ladenson PW, Magner J, Ross DS, Skarulis MC, Steward DL, Maxon HR, Sherman SI 2010 Radio-iodine therapy in patients with stage I differentiated thyroid cancer. Thyroid **20:**1423–1424.

19. McLeod DS, Cooper DS, Ladenson PW, Ain KB, Brierley JD, Fein HG, Haugen BR, Jonklaas J, Magner J, Ross DS, Skarulis MC, Steward DL, Maxon HR, Sherman SI 2014 Prognosis of differentiated thyroid cancer in relation to serum TSH and thyroglobulin antibody status at time of diagnosis. Thyroid **24:**35–42.

20. Carhill AA, Litofsky DR, Ross DS, Jonklaas J, Cooper DS, Brierley JD, Ladenson PW, Ain KB, Fein HG, Haugen BR, Magner J, Skarulis MC, Steward DL, Xing M, Maxon HR, Sherman SI 2015 Long-term outcomes following therapy in differentiated thyroid carcinoma: NTCTCS Registry Analysis 1987-2012. J Clin Endocrinol Metab [Epub ahead of print]; JC20151346.

21. Harrell FE, Jr., Califf RM, Pryor DB, Lee KL, Rosati RA 1982 Evaluating the yield of medical tests. JAMA **247:** 2543–2546.

22. Efron B 1979 Bootstrap methods: another look at the jackknife. Ann Stat **7:**1–26.

23. Hosmer DW, Lemeshow S, May S 2008 Applied Survival Analysis: Regression Modeling of Time-to-Event Data. Second edition. Wiley-Interscience, Hoboken, NJ.

24. Kazaure HS, Roman SA, Sosa JA 2012 Aggressive variants of papillary thyroid cancer: incidence, characteristics and predictors of survival among 43,738 patients. Ann Surg Oncol **19:**1874–1880.

25. Harrell FE 2001 Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer-Verlag, New York, NY.

26. Hay ID, Bergstralh EJ, Goellner JR, Ebersold JR, Grant CS 1993 Predicting outcome in papillary thyroid carcinoma: development of a reliable prognostic scoring system in a cohort of 1779 patients surgically treated at one institution during 1940 through 1989. Surgery **114:**1050–1057; discussion 1057–1058.

27. Tuttle RM, Tala H, Shah J, Leboeuf R, Ghossein R, Gonen M, Brokhin M, Omry G, Fagin JA, Shaha A 2010 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. Thyroid **20:**1341–1349.

28. Johnson TL, Lloyd RV, Thompson NW, Beierwaltes WH, Sisson JC 1988 Prognostic implications of the tall cell variant of papillary thyroid carcinoma. Am J Surg Pathol **12:**22–27.

29. Morris LG, Shaha AR, Tuttle RM, Sikora AG, Ganly I 2010 Tall-cell variant of papillary thyroid carcinoma: a matched-pair analysis of survival. Thyroid **20:**153–158.

30. Elaraj DM, Clark OH 2007 Changing management in patients with papillary thyroid cancer. Curr Treat Options Oncol **8:**305–313.

31. Nixon IJ, Ganly I, Patel SG, Palmer FL, Whitcher MM, Ghossein R, Michael Tuttle R, Shaha AR, Shah JP 2012 Changing trends in well differentiated thyroid carcinoma over eight decades. Int J Surg **10:**618–623.

32. Brose MS, Nutting CM, Jarzab B, Elisei R, Siena S, Bastholt L, de la Fouchardiere C, Pacini F, Paschke R, Shong YK, Sherman SI, Smit JW, Chung J, Kappeler C, Pena C, Molnar I, Schlumberger MJ, investigators D 2014 Sorafenib in radioactive iodine-refractory, locally advanced or metastatic differentiated thyroid cancer: a randomised, double-blind, phase 3 trial. Lancet **384:**319–328.

33. Schlumberger M, Tahara M, Wirth LJ, Robinson B, Brose MS, Elisei R, Habra MA, Newbold K, Shah MH, Hoff AO, Gianoukakis AG, Kiyota N, Taylor MH, Kim SB, Krzyzanowska MK, Dutcus CE, de las Heras B, Zhu J, Sherman SI 2015 Lenvatinib versus placebo in radioiodine-refractory thyroid cancer. New Engl J Med **372:**621–630.

34. Ho AL, Grewal RK, Leboeuf R, Sherman EJ, Pfister DG, Deandreis D, Pentlow KS, Zanzonico PB, Haque S, Gavane S, Ghossein RA, Ricarte-Filho JC, Dominguez JM, Shen R, Tuttle RM, Larson SM, Fagin JA 2013 Selumetinib-enhanced radioiodine uptake in advanced thyroid cancer. New Engl J Med **368:**623–632.

Address correspondence 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