Staging Systems for Papillary Thyroid Carcinoma
A Study of 2 Tertiary Referral Centers

Brian Hung-Hin Lang, MS, FRACS,* Sin-Ming Chow, MBBS, FRCR,†
Chung-Yau Lo, MS, FRCS (Edin), FACS,* Stephen C. K. Law, MBBS, FRCR,†
and King-Yin Lam, MD, FRCPA‡

Objective: To find out the most applicable and consistent staging system for papillary thyroid carcinoma (PTC) available in the literature.

Background: The commonly used staging systems for PTC have predicted cancer-specific survival (CSS) well. However, their applicability and generalizability have not yet been evaluated in different clinical settings.

Methods: A MEDLINE search from 1965 to 2005 was carried out to identify different staging systems available in the literature and 9 systems were applicable to 1634 PTC patients within 2 tertiary-referral centers. The CSS of each staging system within individual centers were calculated using Kaplan-Meier method and the CSS of each tumor stage in one individual center was compared with that of the other by log-rank test. In addition, within each center, the predictability of each staging system relative to the others was ranked based on the proportion of variation explained (PVE) value.

Results: Clinicopathologic features, treatment received, and tumor stages were significantly different between the 2 centers. There were also significant differences in CSS within at least one tumor stage between the 2 centers in 8 of the 9 staging systems. The TNM was a highly predictive and consistent staging system within the 2 centers. Although the absolute PVE values differed between the 2 centers, the relative ranking of the 9 staging systems within each center correlated significantly to each other ($P < 0.05$).

Conclusions: Despite referral, treatment, and data collection biases inherent within each center, the TNM system remained to be the most applicable and consistent staging system for PTC in 2 centers managing the same population group.

(*Ann Surg* 2007;246: 114–121)

From the *Division of Endocrine Surgery, Department of Surgery, University of Hong Kong Medical Centre, Queen Mary Hospital, Hong Kong SAR, China; †Department of Clinical Oncology, Queen Elizabeth Hospital, Hong Kong SAR, China; and ‡Discipline of Pathology, School of Medicine, Griffith University, Gold Coast, Australia.
Reprints: Chung-Yau Lo, FRCS, FACS, Division of Endocrine Surgery, Department of Surgery, Queen Mary Hospital, 102 Pokfulam Road, Hong Kong SAR, China. E-mail: cylo@hkucc.hku.hk.

Papillary thyroid carcinoma (PTC) is the commonest type of differentiated thyroid carcinoma (DTC) and accounts for at least 70% of all follicular-cell derived thyroid malignancies.[1] In our locality, its incidence has increased by 2.5 times in the last 20 years,[2] and this trend has been similarly observed in other parts of the world.[3–6] Although the overall prognosis of PTC is good, up to 10% of patients would eventually die of this disease and an even greater proportion would face the morbidity associated with recurrent disease.[1,7,8] A number of studies attempted to identify various reliable clinicopathologic factors and to devise risk-group stratification or staging systems in predicting outcome and facilitating management.[9–13] Stage-specific treatment aims at tailoring to the needs of individuals. Strategy treatment planning selects those at high risk of cancer death for more aggressive surgical and adjuvant treatment while avoiding unnecessary treatment in good-risk patients.[14]

Given the large number of staging systems available in the literature, several studies had been carried out comparing the predictability of these systems in patients with DTC for cancer-specific survival (CSS) across different population groups.[12,15–17] As there is evidence that PTC and follicular thyroid carcinoma differ significantly in clinicopathologic characteristics and outcome, specific attention to these 2 histologic groups individually is warranted.[18,19] To our knowledge, this is the first study attempting to investigate the applicability, generalizability, and predictability of staging systems for PTC in different clinical settings for the same population group.

The present study evaluates different staging systems applicable to PTC based on a comprehensive literature MEDLINE search. Patient characteristics, histologic features of PTC, and follow-up data were provided by the 2 different tertiary-referral centers specialized in managing thyroid carcinoma.

PATIENTS AND METHODS

Two Centers

The 2 participating centers were Queen Mary Hospital (Center A) and Queen Elizabeth Hospital (Center B) in Hong Kong. Center A is a University-based teaching hospital situated in the western part of Hong Kong Island and Center B is a major government-owned hospital located in the central part of the Kowloon peninsula. Both centers have been

tertiary-referral centers for managing thyroid carcinoma at their respective area. However, due to their differences in geographic location, Center A covers mostly referrals from the Hong Kong Island area whereas Center B covers mostly the Kowloon and, to some extent, the New Territories area. Based on the information provided by the Hong Kong Cancer Registry, the caseload in the past 10 years for Center A equated to approximately one sixth of all new thyroid carcinoma cases in the region whereas for Center B the caseload equated to approximately one fourth.[2]

In terms of data collection, both centers began to prospectively establish their thyroid cancer database in the 1990s.[18,20] Both databases collected data of patients managed before the 1990s retrospectively from medical records. All relevant data, including the clinicopathologic parameters, treatment modalities received, postoperative outcome, and follow-up data, were entered. The thyroid cancer database in Center A was maintained and updated by the Department of Surgery, whereas in Center B it was by the Department of Clinical Oncology. For the purpose of the present study, Center B had provided a number of important patient parameters, including clinicopathologic characteristics, treatment modalities received, and follow-up status. These data were used for calculation of various cancer staging, survival analysis, and comparison between centers.

Patients

In Center A, there were a total of 589 patients with a histologic diagnosis of PTC being treated from 1961 to 2001, while in Center B there were 1045 patients from 1966 to 2001. All histologic variants of PTC were included in both centers. To ensure consistency and accuracy, each histologic diagnosis was made according to the standardized criteria approved by World Health Organization.[21] The ethnicity, gender, and age did not differ. The majority of patients in Centers A and B were female (79.6% vs. 82.2%, $P = 0.175$) and ethnic Chinese (93.7% vs. 98.2%, $P = 0.647$). The median age of the cohort in Center A was 43.0 years (range, 10.0–89.0 years) or (mean $\pm$ SD, 45.4 $\pm$ 17.4 years), and in Center B it was 43.9 years (range, 7.7–91.6 years) or (mean $\pm$ SD, 45.2 $\pm$ 15.7 years) ($P = 0.770$).

Management Strategy

Details of the management protocol for patients with PTC in the 2 respective medical centers had been described elsewhere.[18,22,23] Both centers appeared to share a similar surgical strategy. For patients with a preoperative diagnosis of DTC, a total or near-total thyroidectomy was the preferred procedure. For those diagnosed with DTC after a lobectomy, the decision to perform a completion total thyroidectomy and/or to administer radioiodine (RAI) ablation was determined by risk factors such as the patient's age, tumor characteristics, as well as patient's preference. If patients underwent completion total thyroidectomy within 6 months of their initial thyroid surgery, it was considered as part of the initial surgery. Prophylactic central compartment (level VI) lymph node dissection was not practiced routinely in either center. A selective neck dissection was carried out in cases of presence of clinical suspicious lymph nodes or cytologically proven

lateral cervical nodal metastases. At Center A, a complete resection meant a macroscopic tumor clearance of primary tumor based on the decision made by the operating surgeon (ie, a R0/R1 resection); whereas at Center B, it meant both a macroscopic and microscopic tumor clearance of primary tumor based on the description of the histologic report (ie, a R0 resection).

Both centers shared a similar adjuvant postoperative treatment protocol. Patients with at least one or more of the following risk factors would be considered for RAI ablation 4 to 6 weeks after surgical treatment by thyroxine withdrawal: tumor size >1-cm, lymph node metastasis, age older than 40 years, presence of extrathyroidal extension, macroscopic postoperative residual disease in the neck, and/or distant metastasis. The routine practice of performing a neck scan (tracer dose, 10 MBq) to detect any residual thyroid tissue in the neck before RAI ablation was discontinued in the late 1990s. A diagnostic whole-body I^{131} scans (80 MBq) were performed at 3 to 6 months after RAI ablation; 3 GBq (80mCi) I^{131} would be administered as a standard ablative dose while subsequent I^{131} therapy would be performed with 5.5 GBq. An additional 5.5 GBq I^{131} therapy would be administered periodically at 4- to 6-monthly intervals until uptake was no longer visible or disease progressed despite treatment. External radiotherapy to the thyroid bed and its associated lymphatic drainage was given to patients with gross locoregional disease, extensive extrathyroidal tumor extension, incomplete resection, and/or extracapsular lymph node metastasis. Although the above protocol was as strictly followed as possible in both centers, variations in indications of treatment and RAI dosage occurred infrequently because of individual physician and/or patient preference.

Follow-up and Surveillance of Patients

Complete follow-up data of patients were available for the 2 centers. In Center A, the median follow-up period was 83.2 months (range, 0.2–497.3 months; mean $\pm$ SD, 115.6 $\pm$ 97.6 months); and in Center B, it was 109.7 months (range, 0.4–313.7 months; mean $\pm$ SD, 116.1 $\pm$ 71.9 months). In both centers, combined clinic where clinical oncologists and general/endocrine surgeons were established or held regularly to decide on specific patient management. Follow-up visit with routine clinical examinations was conducted at 3-monthly interval in the first 2 years, 6 monthly for the subsequent 3 years and annually thereafter. Thyroglobulin monitoring during follow-up visits had been practiced in both centers since 1990. In addition, chest x-ray and ultrasonography of neck were included during follow-up visits in Center A. Human recombinant TSH was not available during the study period at both centers. Radioactive scans were done in the presence of an elevated thyroglobulin level without thyrotrophin stimulation (>10 μg/L), documented nodal recurrence, or radiologic evidence of recurrence or metastases. The diagnosis of distant metastases on presentation was based on histologic, radiologic, or scintigraphic evidence. Locoregional recurrences were diagnosed by ultrasound, CT, or MRI imaging and frequently confirmed by fine needle aspiration cytology. For both centers, survival data including the cause of death were retrieved from the Hong Kong Hospital Authority ter-

ritory-wide computerized medical system record, death certificates and postmortem examination reports if available.

Application of Staging Systems

A comprehensive MEDLINE search was performed to identify relevant articles from 1965 to 2005 indexed under the key words thyroid carcinoma/cancer, staging, risk stratification, multivariate analysis, or risk factors. The abstracts of all available articles were read and those describing staging systems or risk group stratifications were reviewed in detail. In addition, the bibliographies of relevant articles were thoroughly searched to identify potentially relevant articles not captured in the original MEDLINE search.

A total of 17 different staging systems were identified. In the order of their year of publication, they were the European Organization for Research and Treatment of Cancer classification (EORTC),[9] the Age, Grade, Extent and Size classification (AGES),[10] the Age, Metastases, Extent and Size system (AMES),[11] the Clinical Class system (Clinical Class),[24] the DNA, Age, Metastases, Extent and Size system (DAMES),[25] the Metastases, Age, Completeness of Surgery, Invasion and Size system (MACIS),[26] the Sex, Age and Grade system (SAG),[27] the Ohio State University system (OSU),[28] the Noguchi classification (Noguchi),[29] the Grade, Age, Metastases, Extent and Size classification (GAMES),[30] the University of Münster system (Münster),[31] the National Thyroid Cancer Treatment Cooperative Study classification (NTCTCS),[12] the University of Alabama and M. D. Anderson system (UA&MDA),[32] the University of Murcia system (Murcia),[33] the AJCC/UICC, 6th edition, Tumor, Node and Metastasis system (TNM),[13] the Cancer Institute Hospital system (CIH),[34] and the Ankara Oncology Training and Research Hospital system (Ankara).[17]

Patients from each center were staged according to the method described in the original publication. Overall, only 9 of the 17 staging systems were applicable to both centers because information for staging calculation was not available for some systems. Specific reasons for nonapplicability of these systems were the lack of data on tumor grading (AGES), DNA ploidy (DAMES), nuclear atypia (SAG), gross lymph node involvement (Noguchi), microscopic and macroscopic tumor multifocality (NCTTCS), histologic PTC variants (Murcia), size of lymph node metastases (CIH), and tumor angioinvasion (Ankara).

Statistical Analysis

The end-point for the present analysis was cause-specific survival (CSS). χ^2 tests or Fisher exact test tests were used for comparison of dichotomous variables, while Mann-Whitney U test was used for continuous variables. For each staging system, CSS was calculated by the Kaplan-Meier method and compared by the log-rank test. Using Cox proportional hazards analysis, the relative importance of each staging system was determined by calculating the proportion of variation in survival time explained (PVE). PVE ranges from 0% to 100% with increasing values indicating better predictability. The staging system with the largest PVE value would imply the best predictive system on CSS and statistical comparison of PVEs for different staging systems using the

computationally intensive technique was not performed.[15] Each staging system was ranked relative to each other based on the PVE value. To determine PVE, a mathematical formula was used: $PVE = 1 - \exp(-G^2/n)$, where G^2 is the maximum likelihood ratio that is determined by analysis of χ^2 associated with the null hypothesis (ie, that all predictor variables have coefficients of 0) and n is the total number of valid cases in the study.[35] Correlation in the ranking of staging systems between the 2 centers was performed by Spearman rank test. $P < 0.05$ was considered to indicate statistical significance. Statistical analyses were performed using the SPSS for Windows 11.0 computer software (SPSS Inc., Chicago, IL)

RESULTS

Table 1 shows a comparison of patient clinicopathologic features and treatment received between Centers A and B. The mean tumor size in Center A was greater than that in Center B (2.57 ± 1.75 cm vs. 2.30 ± 1.72 cm, $P < 0.001$) and the proportion of tumors measuring less than 2.0 cm was greater in Center B than that of Center A (44.2% vs. 40.6%, $P = 0.038$). Although the proportion of cervical lymph node metastases was greater in Center A (41.9% vs. 31.8%, $P < 0.001$), there were a greater proportion of patients with distant metastases (3.8% vs. 1.7%, $P = 0.007$), extrathyroidal tumor involvement (44.2% vs. 38.0%, $P = 0.006$), and incomplete surgical resection (23.2% vs. 7.8%, $P < 0.001$) in Center B. The proportion of patients treated by unilateral thyroid resection or lobectomy was more frequent at Center A (9.8% vs. 4.1%, $P < 0.001$), whereas RAI ablation was more readily administered at Center B (81.3% vs. 60.8%, $P < 0.001$). When only those who underwent bilateral thyroid resections were considered, RAI ablation was still more readily administered at Center B (84.3% vs. 67.1%, $P < 0.001$). For those patients with incomplete resection, the frequency of postoperative external-beam irradiation was similar between the 2 centers (Center B vs. Center A; 50.0% vs. 60.7%, $P = 0.177$).

Survival Analysis and Staging

The CSS rates in Center A at 5, 10, and 15 years were 95.1%, 91.3%, and 87.7%; whereas in Center B, they were 96.5%, 94.3%, and 92.7%, respectively. The difference in CSS between the 2 centers was statistically significant ($P = 0.005$).

Table 2 shows a comparison of the distribution of patients according to the tumor stages or risk group stratification using the 9 different staging systems. Only the AMES and GAMES staging systems demonstrated no significant difference in tumor stratification between the 2 centers, whereas the other 7 staging systems demonstrated significant differences. A comparison of CSS within each tumor stage or risk group between the 2 centers demonstrated that all staging systems except TNM demonstrated that there were significant differences in CSS between 2 centers in at least one of the tumor stages or risk subgroups. Figures 1 and 2 show the CSS curves by TNM for Centers A and B, respectively.

TABLE 1. A Comparison of Patient Characteristics of Papillary Thyroid Carcinoma in the 2 Centers

Clinicopathologic Variable	Center A (n = 589) [No. (%) patients]	Center B (n = 1045) [No. (%) patients]	P
Age at diagnosis			0.343
<45 yr	209 (35.5)	546 (52.2)	
≥45 yr	380 (64.5)	499 (47.8)	
Gender			0.175
M	120 (20.4)	184 (17.6)	
F	469 (79.6)	861 (82.4)	
Size			0.038
<2.0 cm	239 (40.6)	462 (44.2)	
2.0–4.0 cm	268 (45.5)	475 (45.5)	
>4.0 cm	82 (13.9)	108 (10.3)	
Cervical lymph node metastases			<0.001
No	342 (58.1)	713 (68.2)	
Yes	247 (41.9)	332 (31.8)	
Distant metastases			0.007
No	579 (98.3)	1005 (96.2)	
Yes	10 (1.7)	40 (3.8)	
Extrathyroidal involvement			0.006
No	365 (62.0)	583 (55.8)	
Yes	224 (38.0)	462 (44.2)	
Multifocality			0.644
No	397 (67.4)	693 (66.3)	
Yes	192 (32.6)	352 (33.7)	
Completeness of resection*			<0.001
No	46 (7.8)	242 (23.2)	
Yes	543 (92.2)	803 (76.8)	
Extent of surgery			<0.001
Total thyroidectomy	510 (86.6)	916 (87.7)	
Subtotal thyroidectomy	21 (3.6)	86 (8.2)	
Lobectomy	58 (9.8)	43 (4.1)	
Radioiodine ablation			<0.001
No	231 (39.2)	195 (18.7)	
Yes	358 (60.8)	850 (81.3)	
External-beam irradiation to neck			0.360
No	514 (87.3)	897 (85.8)	
Yes	75 (12.7)	148 (14.2)	

*At Center A, complete resection was defined by macroscopic tumor clearance of primary tumor from thyroid bed; whereas at Center B, it was defined by both macroscopic and microscopic clearance of primary tumor.

Predictability of Staging Systems in the Two Centers

Table 3 shows a comparison of the PVE value and relative ranking of the 9 staging systems within each of the centers. There were variations in PVE values between the 2 centers. In terms of ranking of systems based on the highest PVE value, the top 3 ranked systems were similar in the 2 centers, namely, in the order of MACIS, TNM, and EORTC. The rankings of the 9 staging systems between the 2 centers correlated with each other significantly (R = 0.929, $P < 0.001$).

DISCUSSION

A well-accepted staging system not only provides clinicians and patients with useful prognostic information but also facilitates management and standardizes cancer information exchange between different medical centers.[36,37] However, to fulfill these functions, a staging system should not only be evaluated for its ability in predicting CSS (ie, predictability) but also for its applicability and generalizability in different clinical settings.[38,39] The TNM has been widely accepted for PTC staging because it is simple and has also been extensively used in other disease sites. However, its staging accuracy has been shown to vary with different disease sites as well as with the experience of the stagers.[38]

To evaluate the applicability and generalizability of these different staging systems for PTC, the present study purposefully chose to compare 2 different tertiary-referral centers specialized in managing thyroid carcinoma within the same population group. There was no attempt at standardizing or adjusting the cancer coding of the 2 thyroid cancer databases, and these databases had been maintained and updated independently by 2 different centers. As a proof of the lack of standardization, the patient characteristics and treatment received in the present study were comparable to those previously published at their respective center.[18,22,23]

From the results, there appeared to be a number of significant differences in patient characteristics between the 2 centers. These differences included the tumor size as well as the proportions of cervical lymph node and distant metastases, extrathyroidal involvement, and completeness of resection. Treatment received including the extent of surgery and the administration of RAI ablation was also found to be significantly different between the 2 centers. Hypothetically, all these may have been attributed to referral biases inherent within each center.[40] As Center A was essentially an endocrine surgical unit with its main expertise in surgical treatment and Center B was a clinical oncology unit with its main expertise in radiation oncology, there could have been significant differences in referral pattern leading to a different case mix. The significant differences in the extent of surgery and RAI ablation between the 2 centers supported this hypothesis. In Center B, the proportion of patients who received RAI ablation was well over 80% compared with that of only 60% to 70% in Center A ($P < 0.001$). Similarly, the proportion of patients who underwent either a total or subtotal thyroidectomy in Center B was close to 95%; whereas in Center A, the proportion was around 85%. This could be attributed to the intent of referring patients to Center B for radioiodine therapy only after bilateral thyroidectomies. After all, unlike other centers,[41] neither Center A nor B advocated large thyroid remnant ablation. An alternative reason might have been due to the subtle differences in indications for completion total thyroidectomy after a lobectomy (a treatment bias). However, the proportion of external-beam radio-

TABLE 2. A Comparison of Stage Stratification and Cause-Specific Survivals of Papillary Thyroid Carcinoma Between the 2 Centers

| Staging System | Center A (n = 589) | | | Center B (n = 1045) | | | Difference in Stage Stratification Between the 2 Centers (*P*) | Difference in Cancer-Specific Survival in Different Stages Between the 2 Centers (*P*) |
| | No. (%) Patients | Cancer-Specific Survival | | No. (%) Patients | Cancer-Specific Survival | | | |
		10-yr (%)	15-yr (%)		10-yr (%)	15-yr (%)		
EORTC							0.005*	
I	314 (53.3)	99.0	98.1	511 (48.9)	100.0	100.0		0.021
II	120 (20.4)	97.8	91.2	273 (26.1)	96.4	94.2		0.304
III	107 (18.2)	70.0	59.6	201 (19.2)	86.9	76.5		0.008*
IV	48 (8.1)	50.3	33.5	57 (5.5)	56.7	56.7		0.753
V	0 (0.0)	0	0	3 (0.3)	33.3	33.3		NA
AMES							0.847	
Low risk	446 (75.7)	97.3	95.0	789 (75.5)	99.4	98.5		0.026*
High risk	143 (24.3)	69.5	61.7	256 (24.5)	78.7	71.7		0.113
Clinical Class							<0.001*	
I	218 (37.0)	98.0	97.0	443 (42.4)	98.9	98.1		0.271
II	132 (22.4)	97.0	92.9	128 (12.2)	97.9	95.8		0.221
III	229 (38.9)	82.7	75.3	434 (41.5)	92.0	89.0		0.004*
IV	10 (1.7)	45.0	45.0	40 (3.8)	60.5	60.5		0.372
MACIS							0.011*	
I	437 (74.2)	98.9	97.7	732 (70.0)	99.5	98.9		0.054
II	50 (8.5)	86.8	68.3	139 (13.3)	100.0	97.0		<0.001*
III	54 (9.2)	64.2	55.1	80 (7.7)	75.1	75.1		0.117
IV	48 (8.1)	42.1	33.7	94 (9.0)	58.6	44.6		0.371
OSU							<0.001*	
I	52 (8.8)	100.0	100.0	127 (12.5)	100.0	97.1		0.408
II	279 (47.4)	96.6	95.0	406 (38.9)	98.9	98.0		0.049*
III	248 (42.1)	84.8	78.3	471 (45.1)	91.8	89.4		0.001*
IV	10 (1.7)	45.0	45.0	41 (3.9)	60.5	60.5		0.682
GAMES							0.126	
Low risk	289 (49.1)	98.8	98.8	497 (47.6)	100.0	99.2		0.088
Intermediate risk	242 (41.1)	88.1	81.9	467 (44.7)	93.7	92.8		0.001*
High risk	58 (9.8)	64.8	48.6	81 (7.8)	62.5	62.5		0.592
Münster							0.001*	
Low risk	360 (61.1)	96.5	94.2	573 (54.8)	98.6	97.4		0.020*
High risk	229 (38.9)	80.1	72.8	472 (45.2)	89.4	86.6		0.005*
UA & MDA							0.021*	
Low risk	374 (63.5)	99.1	98.4	635 (60.8)	99.6	99.0		0.235
Intermediate risk	205 (34.8)	77.4	67.8	370 (35.4)	89.0	84.5		<0.001*
High risk	10 (1.7)	45.0	45.0	40 (3.8)	60.5	60.5		0.682
TNM							0.004*	
I	364 (61.8)	98.6	98.6	665 (63.6)	99.7	98.6		0.295
II	81 (13.8)	94.8	94.8	45 (4.3)	95.0	95.0		0.144
III	35 (5.9)	90.1	83.6	178 (17.0)	98.2	95.5		0.150
IV[†]	109 (18.5)	61.7	40.6	157 (15.0)	63.4	56.4		0.596

*Statistically significant.
[†]Comprised Stages IVA, IVB, and IVC.
EORTC indicates European Organization for Research and Treatment of Cancer; AMES, Age, Metastases, Extent and Size; MACIS, Metastases, Age, Completeness of surgery, Invasion and Size; OSU, Ohio State University; GAMES, Grade, Age, Metastases, Extent and Size; Münster, University of Munster; UA & MDA, University of Alabama and M.D. Anderson; TNM, Tumor, Node and Metastasis; intermed-risk, intermediate risk group; NA, not applicable.

therapy was similar between the 2 centers and this remained so even when those with incomplete resection were considered only, indicating that a similar policy was probably used for this adjuvant treatment modality. The extraordinarily high percentage of incomplete resection in Center B compared with Center A (23.2% vs. 7.8%, *P* < 0.001) might have been attributed to the less stringent definition of incomplete resection in Center B.

Interestingly, though, the tumor size was larger and the incidence of lymph node metastases was higher in Center A.
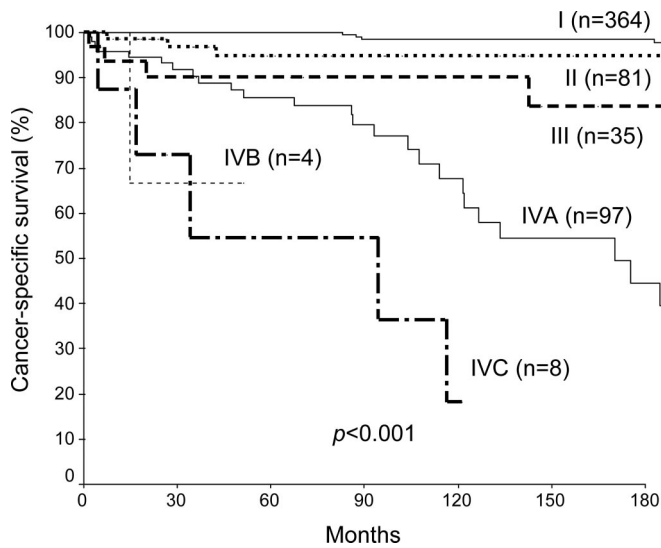
FIGURE 1. Actuarial 15-year cancer specific survival of 589 papillary thyroid carcinoma in Center A by the AJCC/UICC, 6th edition, TNM staging system.
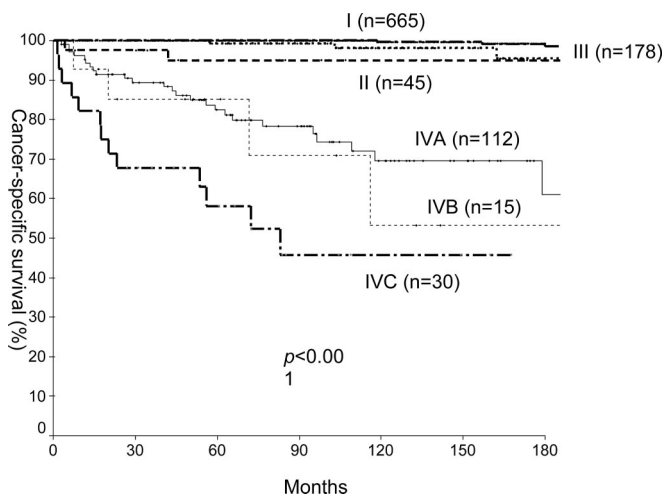
TABLE 3. The Value of Proportion of Variance Explained and Ranking of the 9 Applicable Staging Systems in Papillary Thyroid Carcinoma for the 2 Centers

Staging System	Center A (n = 589)		Center B (n = 1 045)	
	PVE (%)	Ranking	PVE (%)	Ranking
EORTC	16.6	3	12.8	3
AMES	10.5	5	10.2	5
Clinical class	9.6	7	6.7	7
MACIS	18.7	1	16.8	1
OSU	7.7	8	6.7	8
GAMES	10.0	6	10.6	4
Münster	6.1	9	3.9	9
UA & MDA	14.0	4	9.5	6
TNM	17.9	2	15.1	2

PVE indicates proportion of variance explained; EORTC, European Organization for Research and Treatment of Cancer; AMES, Age, Metastases, Extent and Size; MACIS, Metastases, Age, Completeness of surgery, Invasion and Size; OSU, Ohio State University; GAMES, Grade, Age, Metastases, Extent and Size; Münster, University of Münster; UA & MDA, University of Alabama and M. D. Anderson; TNM, Tumor, Node and Metastasis.

FIGURE 2. Actuarial 15-year cancer-specific of 1045 papillary thyroid carcinoma in Center B by the AJCC/UICC, 6th edition, TNM staging system.

Apart from the genuine clinicopathologic differences on presentation, the actual measurement of tumor size might have differed because this was not standardized in the 2 centers. The incidence of lymph node metastases also depended on a number of variables, including the quality of the pathology reporting as well as the surgical strategy or sampling technique, which an individual surgeon or a group of surgeons adopted at the time of operation. The incidence of lymph node metastases had been reported as high as 80% for PTC.[42,43] In principle, although the overall surgical strategy might appear similar in the 2 centers, the actual procedure and operative findings might have varied widely between individual surgeons and centers (a treatment bias). These variations in clinical practice could possibly be minimized in the future by establishing a well-accepted thyroid cancer treatment protocol or guideline, which is currently unavailable in our region.

When patient stratification into different risk-groups or stages was compared within each of the 9 staging systems for the 2 centers, significant differences were shown in 7 except the AMES and GAMES systems (Table 2). These significant differences in tumor stages were possibly a reflection of the underlying differences in patients' clinicopathologic characteristics as each staging system used different patient and tumor variables. The 2 staging systems, which failed to show a difference between the 2 centers, might be least accurate in reflecting the differences in tumor and patient risk profiles. Nevertheless, it was unclear which center had a worse tumor risk profile based on the proportion of patients stratified into different tumor stages.

In the present study, we did not purposely compare survival outcome between the 2 centers as there would be a multitude of factors accounting for the difference, including patients' selection, referral pattern, differences in clinicopathologic features, quality of histopathology reporting, cancer coding or data entry, treatment philosophy, monitoring and treatment of recurrence, as well as comprehensiveness of follow-up. However, since the stage-to-stage CSS in TNM was comparable and the distribution of stages was different between the 2 centers (Table 2), patients' selection, referral pattern, and coding differences were potential attributing factors. In addition, since PTC has a long clinical course, management differences could also influence the outcomes significantly. However, more data on the relapse pattern and the subsequent management would be required before such conclusion could be made and these are beyond the scope of the present study. Furthermore, it is also not our intention to compare survival outcome related to treatment alone without considering other important outcome parameters, such as the quality of life issues and the incidence of morbidity from aggressive surgical or adjuvant treatment.

When CSS within the same tumor stage or risk-group was compared between the 2 centers, there were significant differences in 8 of the 9 staging systems in at least one of the tumor stages or risk groups. These significant differences in CSS within at least one stage of the 8 staging systems could be conceptualized as a stage migration problem because the CSS within the same tumor stage or risk-group should have been similar in the 2 centers regardless of the underlying patient characteristics or distribution of tumor stages.[39] Of course, such difference may be attributed to the different treatment received by the patients within the 2 centers because tumor stages were calculated based on the operative and histologic findings and not the treatment received. Another plausible explanation might have been related to the differences in cancer coding as this was not standardized between the 2 centers. A staging system with a less consistently coded prognostic feature tended to generate significantly different CSS in different clinical settings. Some prognostic features were probably more likely to be less consistently coded because they were influenced by subjective interpretation. An example would be incomplete resection (in MACIS) where at Center A, it was based on an intraoperative decision by the surgeon and at Center B, it was essentially based on the description of the histopathology report. Of course, this type of consistency in cancer coding could also arise from a variety of other reasons but practicality, generalizability, and clear staging instructions are important prerequisites.[38,39] The coding instructions in TNM might have been simpler, less ambitious and easier for clinicians in both centers to follow and apply. Even though the PVE values in MACIS were higher than those in TNM for Centers A and B (ie, better predictability for CSS), the TNM was the only staging system, which did not have significant stage-to-stage CSS difference between the 2 centers (Table 2); therefore, we concluded that the TNM was the most reliable, practical, and consistent staging system for PTC.

However, the TNM staging system is not without its pitfalls. For examples, some authors suggested that an ideal system should assign patients into uniform-size stage groupings.[44] In the present study, over 60% of patients were assigned to the stage I, whereas only 10% to 20% of patients were assigned to the other stages in both centers. An important feature missing in the present TNM system is an objective quantitation of postoperative residual disease as this is a well-known and significant prognostic factor for PTC but has yet to be incorporated into the current TNM system.[19]

One of the limitations with the present study is that there is no universally accepted and objective measurement for predictability. Apart from PVE, a number of other statistical methods had been put forward, but none has been shown to be superior to PVE.[15] To date, the PVE remained the most accepted measurement of predictability and had been used in many recent comparative studies.[17,45–48] It is interesting to note that despite minor variations in PVE values between the 2 centers, the relative ranking of the staging systems remained remarkably similar and were significantly correlated ($P < 0.001$). Indeed, the top-three ranked staging systems (MACIS, TNM, and EORTC) in PTC were the same in the 2

centers. In other words, our results appeared to suggest that, although the actual predictability (as measured by PVE) of a staging system could vary between the 2 centers, its relative performance of these staging systems remained constant. Therefore, despite the significant differences between the 2 centers as aforementioned, its relative performance of different staging systems remained similar. Our findings could provide an explanation for the variability of PVE values in several studies and yet their top ranked staging systems appeared relatively similar.[15,16]

CONCLUSION

Both tertiary-referral centers managing and studying PTC in the same population group demonstrated significant differences in patient characteristics, treatment modalities, and overall survival outcomes. These differences could possibly be explained by differences in referral pattern, patients' selection, cancer coding as well as treatment philosophy between the 2 centers. By comparing the currently available staging and risk-group stratification systems, the present study was able to suggest that the TNM was the most applicable and consistent staging system for PTC within the 2 centers.

REFERENCES

1. Sherman SI. Thyroid carcinoma. *Lancet*. 2003;361:501–511.
2. Cancer incidence and mortality in Hong Kong 1983–2003. *Hong Kong Cancer Registry*. Available at: http://www.ha.org.hk/cancereg/stat.asp. Accessed May 15, 2006.
3. Haselkorn T, Bernstein L, Preston-Martin S, et al. Descriptive epidemiology of thyroid cancer in Los Angeles County, 1972–1995. *Cancer Causes Control*. 2000;11:163–170.
4. Hodgson NC, Button J, Solorzano CC. Thyroid cancer: is the incidence still increasing? *Ann Surg Oncol*. 2004;11:1093–1097.
5. Reynolds RM, Weir J, Stockton DL, et al. Changing trends in incidence and mortality of thyroid cancer in Scotland. *Clin Endocrinol (Oxf)*. 2005;62:156–162.
6. Burgess JR, Tucker P. Incidence trends for papillary thyroid carcinoma and their correlation with thyroid surgery and thyroid fine-needle aspirate cytology. *Thyroid*. 2006;16:47–53.
7. Mazzaferri EL, Kloos RT. Current approaches to primary therapy for papillary and follicular thyroid cancer. *J Clin Endocrinol Metab*. 2001; 86:1447–1462.
8. Eustatia-Rutten CFA, Corssmit EPM, Biermasz NR, et al. Survival and death causes in differentiated thyroid carcinoma. *J Clin Endocrinol Metab*. 2006;91:313–319.
9. Byar DP, Green SB, Dor P, et al. A prognostic index for thyroid carcinoma: a study of the EORTC Thyroid Cancer Cooperative Group. *Eur J Cancer*. 1979;15:1033–1041.
10. Hay ID, Grant CS, Taylor WF, et al. Ipsilateral lobectomy versus bilateral lobar resection in papillary thyroid carcinoma: a retrospective analysis of surgical outcome using a novel prognostic scoring system. *Surgery*. 1987;102:1088–1095.
11. Cady B, Rossi R. An expanded view of risk-group definition in differentiated thyroid carcinoma. *Surgery*. 1988;104:947–953.
12. Sherman SI, Brierley JD, Sperling M, et al. Prospective multicenter study of thyroid carcinoma treatment: initial analysis of staging and outcome. *Cancer*. 1998;83:1012–1021.
13. Greene FL, Page DL, Fleming ID, et al., eds. *AJCC Cancer Staging Handbook*, 6th ed. New York: Springer-Verlag; 2002.
14. Shaha AR. Implications of prognostic factors and risk groups in the management of differentiated thyroid cancer. *Laryngoscope*. 2004;114: 393–402.
15. Brierley JD, Panzarella T, Tsang RW, et al. A comparison of different staging systems. Predictability of patient outcome: thyroid carcinoma as an example. *Cancer*. 1997;79:2414–2423.

16. Passler C, Prager G, Scheuba C, et al. Application of staging systems for differentiated thyroid carcinoma in an endemic goiter region with iodine substitution. *Ann Surg*. 2003;207:227–234.

17. Yildirim E. A model for predicting outcomes in patients with differentiated thyroid cancer and model performance in comparison with other classification systems. *J Am Coll Surg*. 2005;200:378–392.

18. Chow SM, Law SC, Au SK, et al. Differentiated thyroid carcinoma: comparison between papillary and follicular carcinoma in a single institute. *Head Neck*. 2002;24:670–677.

19. Lang BH, Lo CY, Chan WF, et al. Prognostic factors in papillary and follicular thyroid carcinoma: their implications for cancer staging. *Ann Surg Oncol*. 2007;14:730–738.

20. Lorentz TG, Lau PW, Lo CY, et al. Multivariate analysis of risk factors influencing survival in Chinese with papillary thyroid cancer. *World J Surg*. 1994;18:547–550.

21. DeLellis RA, Lloyd RV, Heitz PU, et al., eds. *World Health Organization Classification of Tumours: Pathology and Genetics of Tumours of Endocrine Organs*. France: International Agency for Research on Cancer; 2004.

22. Lo CY, Chan WF, Lam KY, et al. Optimizing the treatment of AMES high-risk papillary thyroid carcinoma. *World J Surg*. 2004;28:1103–1109.

23. Lang BH, Lo CY, Chan WF, et al. Classical and follicular variant of papillary thyroid carcinoma: a comparative study on clinicopathological features and long-term outcome. *World J Surg*. 2006;30:752–758.

24. DeGroot LJ, Kaplan EL, McCormick M, et al. Natural history, treatment, and course of papillary thyroid carcinoma. *J Clin Endocrinol Metab*. 1990;71:414–424.

25. Pasieka JL, Zedenius J, Auer G, et al. Addition of nuclear DNA content to the AMES risk-group classification for papillary thyroid carcinoma. *Surgery*. 1992;112:1154–1160.

26. Hay ID, Bergstralh EJ, Goellner JR, et al. 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*. 1993;114:1050–1058.

27. Akslen LA. Prognostic importance of histologic grading in papillary thyroid carcinoma. *Cancer*. 1993;72:2680–2685.

28. Mazzaferri EL, Jhiang SM. Long-term impact of initial surgical and medical therapy on papillary and follicular thyroid cancer. *Am J Med*. 1994;97:418–428.

29. Noguchi S, Murakami N, Kawamoto H. Classification of papillary cancer of the thyroid based on prognosis. *World J Surg*. 1994;18:552–558.

30. Shaha AR, Loree TR, Shah JP. Intermediate-risk group for differentiated carcinoma of the thyroid. *Surgery*. 1994;116:1036–1041.

31. Lerch H, Schober O, Kuwert T, et al. Survival of differentiated thyroid carcinoma studied in 500 patients. *J Clin Oncol*. 1997;15:2067–2075.

32. Beeken S, Roye D, Weiss H, et al. Extent of surgery for intermediate-risk well-differentiated thyroid cancer. *Am J Surg*. 2000;179:51–56.

33. Ortiz Sebastian S, Rodriguez Gonzalez M, Parilla Paricio P, et al. Papillary thyroid carcinoma: prognostic index for survival including the histological variety. *Arch Surg*. 2000;135:272–277.

34. Sugitani I, Kasai N, Fujimoto Y, et al. A novel classification system for patients with PTC: addition of the new variables of large (3 cm or greater) nodal metastases and reclassification during the follow-up period. *Surgery*. 2004;135:139–148.

35. Schemper M, Stare J. Explained variation in survival analysis. *Stat Med*. 1996;15:1999–2012.

36. Greene FL. Cancer staging, prognostic factors, and our surgical challenges. *Am Surg*. 2005;71:615–620.

37. Cooper DS, Doherty GM, Haugen BR, et al. Management guidelines for patients with thyroid nodules and differentiated thyroid carcinoma. *Thyroid*. 2006;16:109–142.

38. Brierley JD, Catton PA, O'Sullivan B. Accuracy of recorded tumor, node, and metastasis stage in a comprehensive cancer center. *J Clin Oncol*. 2002;20:413–419.

39. Patel SG, Shah JP. TNM staging of cancers of the head and neck: striving for uniformity among diversity. *CA Cancer J Clin*. 2005;55:242–258.

40. Sherman SI. Toward a standard clinicopathologic staging approach for differentiated thyroid carcinoma. *Semin Surg Oncol*. 1999;16:12–15.

41. Leblanc G, Tabah R, Liberman M, et al. Large remnant ^{131}I ablation as an alternative to completion/total thyroidectomy in the treatment of well-differentiated thyroid cancer. *Surgery*. 2004;136:1275–1280.

42. Goropoulos A, Karamoshos K, Christodoulou A, et al. Value of the cervical compartments in the surgical treatment of papillary thyroid carcinoma. *World J Surg*. 2004;28:1275–1281.

43. Pereira JA, Jimeno J, Miquel J, et al. Nodal yield, morbidity, and recurrence after central neck dissection for papillary thyroid carcinoma. *Surgery*. 2005;138:1095–1101.

44. Au SK, Law CK, Foo W, et al. In-depth evaluation of the AJCC/UICC 1997 staging system of nasopharyngeal carcinoma: prognostic homogeneity and proposed refinements. *Int J Radiat Oncol Biol Phys*. 2003;56:413–426.

45. D'Avanzo A, Ituarte P, Treseler P, et al. Prognostic scoring systems in patients with follicular thyroid cancer: a comparison of different staging systems in predicting the patient outcome. *Thyroid*. 2004;14:453–458.

46. Lo CY, Chan WF, Lam KY, et al. Follicular thyroid carcinoma: the role of histology and staging systems in predicting survival. *Ann Surg*. 2005;242:708–715.

47. Barbet J, Campion L, Kraeber-Bodere F, et al. Prognostic impact of serum calcitonin and carcinoembryonic antigen doubling-times in patients with medullary thyroid carcinoma. *J Clin Endocrinol Metab*. 2005;90:6077–6084.

48. Boulos DL, Groome PA, Brundage MD, et al. Predictive validity of five comorbidity indices in prostate carcinoma patients with curative intent. *Cancer*. 2006;106:1804–1814.