Practical Performance of the 2015 American Thyroid Association Guidelines for Predicting Tumor Recurrence in Patients with Papillary Thyroid Cancer in South Korea

Seul Gi Lee,[1] Woo Kyung Lee,[2] Hye Sun Lee,[3] Jieun Moon,[3] Cho Rok Lee,[1] Sang Wook Kang,[1] Jong Ju Jeong,[1] Kee-Hyun Nam,[1] Woong Youn Chung,[1] Young Suk Jo,[2] and Jandee Lee[1]

Background: The 2015 American Thyroid Association (ATA) management guidelines for adult patients with differentiated thyroid cancer propose the use of the modified initial risk stratification and response to therapy re-stratification systems. This study was conducted to validate the practicality of the revised guidelines for predicting tumor recurrence in patients with differentiated thyroid cancer.

Methods: Patients with papillary thyroid cancer ($n = 2425$) who underwent total thyroidectomy with central neck node dissection with or without modified radical neck node dissection at a single institution between October 1985 and July 2009 were retrospectively enrolled. The accuracy of three different risk-stratification strategies for predicting disease-free survival, set out in the 2009 and 2015 ATA management guidelines, was assessed: the initial risk stratification (ATA 2009-RSS), the modified initial risk stratification (ATA 2015-RSS), and the response to therapy re-stratification (ATA 2015-RTR).

Results: After applying the ATA 2015-RSS, 258/1913 (13.5%) of patients originally designated as intermediate risk by the ATA 2009-RSS were designated as low risk. This was mainly due to the small number of metastatic lymph nodes. Recurrence was detected in 136 (5.6%) patients during follow-up. Of the 2425 cases examined, 375 were designated as low risk by the ATA 2009-RSS, with a recurrence rate of 1.1%. However, the ATA 2015-RSS designated 633 (26.1%) cases as low risk, with a recurrence rate of 0.9%. Implementing the ATA 2015-RTR predicted an excellent response in 1597 (65.9%) cases, with a recurrence rate of 1.1%. According to the proportion of variance explained (PVE), the Akaike information criterion, Harrell's c index, and integrated area under the curve, comparing the predictive accuracy of the ATA 2009-RSS, ATA 2015-RSS, and ATA 2015-RTR revealed that the ATA 2015-RTR was a superior predictor of recurrence.

Conclusions: A proportion of patients designated as intermediate risk by the ATA 2009-RSS were designated as low risk by the ATA 2015-RSS. The ATA 2015-RTR may be the most accurate predictor of risk of recurrence.

Keywords: risk stratification, recurrence, practice guidelines, papillary thyroid cancer, predictive accuracy

Introduction

DIFFERENTIATED THYROID CANCER is the most common malignancy in endocrine organs, and its incidence is increasing worldwide (1). The robust increase of thyroid cancer has invited debate about the appropriate management strategy for such patients, particularly those with papillary thyroid carcinoma (PTC). However, the lack of accurate prognostic indicators or markers for predicting tumor progression and recurrence means that appropriate management of PTC is unclear.

Molecular markers and clinical risk assessment with respect to PTC have also been rigorously investigated (2–7). As a result, the clinical implications of pathological parameters such as multifocality, minimally or extensive extrathyroidal extension (ETE), and micro- and macroscopic lymph node (LN) metastases can be used to estimate a continuum of risk, with difference ranges for selected clinicopathologic features. In addition, the *BRAF*^{V600E} and *TERT* promoter mutations have been identified as potential markers of a poor prognosis for PTC, although they are not routinely recommended for initial risk stratification (2,3). Based on these

scientific advances, the newly proposed 2015 American Thyroid Association (ATA) management guidelines indicate the risk of structural disease recurrence associated with selected clinicopathologic features for the patients without structurally identifiable disease after initial therapy (8). Besides initial risk stratification, the new 2015 guidelines suggest that clinicians use the response to therapy re-stratification system attributed to previous studies about the estimated risk of recurrence based on effectiveness of initial therapy (9–11).

Even though the above-mentioned risk-stratification system proposed by the new ATA management guidelines cannot be used to make a definite decision at the preoperative stage, it is postulated that the new guidelines will improve the ability to predict tumor progression and recurrence. To test this, the predictive accuracy of three risk stratification systems was compared: (i) the initial risk stratification set out in the 2009 ATA management guidelines (ATA 2009-RSS), (ii) the modified initial risk stratification set out in the 2015 ATA management guidelines (ATA 2015-RSS), and (iii) the response to therapy re-stratification set out in the 2015 ATA management guidelines (ATA 2015-RTR). To achieve this, data were retrospectively collected from a large number of patients with PTC who underwent thyroidectomy at a single institution, and the predictive accuracy of the three systems was examined using quantitative statistical methods.

Materials and Methods

Patients

The study enrolled 2425 patients with PTC who underwent total thyroidectomy between October 1985 and July 2009. Of the 2425 patients, 1423 (58.7%) underwent total thyroidectomy with prophylactic or therapeutic ipsilateral central compartment neck node dissection (CCND), and 1002 (41.3%) underwent total thyroidectomy with prophylactic or therapeutic ipsilateral CCND and therapeutic modified radical neck node dissection due to clinically suspicious or pathologically confirmed N1b nodes. The mean follow-up period was 88.0 months (range 60–265 months). The study was approved by the Institutional Review Board, which waived the requirement for informed consent due to the retrospective design of the study.

Postoperative management and follow-up

After excluding patients with pathologic T1N0M0 PTC, 2314 (95.4%) patients received radioactive iodine (RAI) for remnant ablation at four to eight weeks following surgery. During RAI treatment, serum thyroglobulin (Tg) and antithyroglobulin antibody (TgAb) concentrations were measured after stimulation by thyrotropin (TSH), and post-ablation whole-body scans (WBS) were performed.

To assess disease status, serum Tg and TgAb levels were measured at three and six months after initial treatment and then annually thereafter. In order to evaluate the thyroid bed and central and lateral nodal compartments, neck ultrasound was routinely performed at 6–12 months and then periodically thereafter. Follow-up diagnostic RAI WBS was performed at 6–12 months after adjuvant RAI therapy in patients with a high or an intermediate risk of persistent disease. Chest X-ray, computed tomography (CT), or 18-fludeoxyglucose F-positron emission computed tomography (^{18}FDG-PET/CT) was considered when assessing high-risk PTC patients with an elevated serum Tg level (generally serum Tg >10 ng/mL after TSH stimulation or Tg >1 ng/mL after TSH suppression) or increasing TgAb levels with or without negative RAI WBS imaging.

Among the study patients, 387 (15.6%) presented with a biochemical and structural incomplete response with increased serum Tg levels at one year after surgery. Further imaging studies, including neck ultrasound (100%), RAI-WBS (95.4%), chest X-ray (33.6%), chest CT (15.0%), neck CT (14.2%), and PET/CT (22.0%), were performed in accordance with the follow-up protocol. Recurrence was confirmed when disease sites were found to have local LN metastasis, defined as a positive cytology in a fine-needle aspiration biopsy, or when distant metastasis was confirmed by a biopsy or imaging.

Risk stratification

Patients were stratified according to the ATA 2009-RSS and ATA 2015-RSS. The initial risk stratifications were based on postoperative histopathological, laboratory, and imaging findings. However, data regarding $BRAF^{V600E}$ and $TERT$ promoter mutation status were not available because molecular tests for these mutations were not routinely recommended. ATA 2015-RTR was assessed according to the results of imaging and Tg/TgAb levels measured one year after initial therapy.

Statistical analysis

Continuous, quantitative data are reported as the mean ± standard deviation (SD), and categorical, qualitative data as numbers and percentages. Univariate Cox regression analysis was performed to identify predictors of disease-free survival (DFS), and variables with a p-value <0.10 were included in a multivariate Cox regression analysis to identify independent predictors of DFS. DFS was compared between different risk groups (designated using the same risk-stratification system) using Kaplan–Meier analysis with log-rank test.

The performance of each risk-stratification system was compared using four statistical methods: the proportion of variance explained (PVE) (12), the Akaike information criterion (AIC) (13), Harrell's c index (14), and the time-dependent receiver operating characteristic (ROC) curve method (15,16). While PVE statistics can be used to evaluate the impact of the addition of predictors to a hierarchical model on the amount of variation in the outcome that is explained by a model, use of this statistic is not always straightforward. To improve the predictive ability, a modified PVE is used to assess the dimensionality of a reduced correlation matrix. The modified PVE was calculated using the Cox regression model, where G2 is maximum likelihood ratio from the Cox regression model, and n is the total number of patients. The AIC was calculated for each risk stratification system using the following equation:

$$AIC = 2k - 2\ln(L)$$

where k is the number of estimated parameters included in the model, and L is the maximum value of the likelihood

function. Harrell's c index and integrated area under the curve (iAUC) calculated based on the time-dependent ROC curve for the three risk stratification systems were assessed to compare the abilities of the three risk-stratification systems to predict recurrence accurately over the entire follow-up period. The iAUC is a weighted average of the AUC across a particular follow-up period, and measures the predictive accuracy of a model during follow-up. A higher iAUC indicates better predictive accuracy. Differences in iAUC between the three risk-stratification systems were calculated using a bootstrap method, with 1000 re-samplings (17). Two-sided p-values of <0.05 and excluding 0 in the confidence interval (CI) were considered to indicate statistical significance. Statistical analyses were performed using R package v3.1.3.

Results

Demographic and clinicopathologic characteristics of the study patients

The mean age of the 2425 patients was 45.0 ± 12.3 years, and the majority (86.5%) were female. Overall, 1346 (55.5%) patients had stage I disease, and 648 (26.7%) patients had stage III disease. ETE was apparent in 1639 (67.6%) patients, and 2314 (95.4%) received postoperative RAI ablation. All other characteristics are listed in Supplementary Table S1 (Supplementary Data are available online at www.liebertpub.com.thy).

Risk migration after applying ATA 2015-RSS and ATA 2015-RTR

Classification according to the ATA 2009-RSS led to 375 (15.5%), 1913 (78.9%), and 137 (5.6%) patients being assigned to the low-, intermediate-, and high-risk groups, respectively (Supplementary Table S2). However, after application of ATA 2015-RSS, 258 patients in the intermediate-risk group migrated to the low-risk group. This was due to the small volume of metastatic LNs (five or fewer LN micrometastases <0.2 cm). Five patients initially placed in the intermediate-risk group migrated to the high-risk group due to the size of the metastatic LNs (≥3 cm). Finally, the proportion of patients in the low-, intermediate-, and high-risk

groups according to ATA 2015-RSS were 633 (26.1%), 1650 (68.0%), and 142 (5.9%), respectively (Supplementary Table S2).

After implementing ATA 2015-RTR to ATA 2009-RSS, an excellent response was observed in 301/375 (80.3%) patients in the low-risk group, 1251/1913 (65.4%) in the intermediate-risk group, and 45/137 (32.8%) in the high-risk group (Table 1 and Fig. 1). Implementing ATA 2015-RTR to ATA 2015-RSS led to an excellent response in 498/633 (78.7%) patients in the low-risk group, 1053/1650 (63.8%) in the intermediate-risk group, and 46/142 (32.4%) in the high-risk group. Thus, a total of excellent responses was observed in 1597/2425 (65.9%) patients (Table 1 and Fig. 1).

Risk-stratification systems as the predictors of recurrence

Overall, recurrence was detected in 136 (5.6%) patients during follow-up. To verify the accuracy of the three different systems for predicting tumor recurrence, univariate analyses were performed based on clinicopathologic parameters. As expected, sex, TNM stage, tumor size, multiplicity, bilaterality, ETE, high-dose RAI ablation (≥200 mCi), and all three different risk-stratification systems were significant predictors of recurrence (Table 2). Next, multivariate analyses were performed to analyze the relationships between clinicopathologic parameters and the three risk-stratification systems. Interestingly, all three risk-stratification systems predicted tumor recurrence, with strong statistical significance after adjusting for covariates, including age, sex, multiplicity, bilaterality, and adjuvant RAI ablation (Table 2). However, because several factors such as tumor size, ETE, and TNM stage are already included in the ATA 2009-RSS and ATA 2015-RSS, these clinicopathologic parameters were excluded to avoid collinearity.

### Comparison of the ability of the risk-stratification systems to predict recurrence

Kaplan–Meier survival analysis revealed that all three risk-stratification systems had significant differences across all stages ($p < 0.001$; Fig. 2A–C). To examine the correlation between each risk-stratification system and recurrence

TABLE 1. RE-STRATIFICATION OF PATIENTS AT ONE YEAR AFTER SURGERY BASED ON THE ATA 2015-RTR

	ATA 2015-RTR				
	Excellent	Indeterminate	Biochemical incomplete response	Structural incomplete response	Total
ATA 2009-RSS					
Low	301 (80.3%)	52 (13.9%)	20 (5.3%)	2 (0.5%)	375
Intermediate	1251 (65.4%)	357 (18.7%)	272 (14.2%)	33 (1.7%)	1913
High	45 (32.8%)	32 (23.4%)	49 (35.8%)	11 (8.0%)	137
ATA 2015-RSS					
Low	498 (78.7%)	94 (14.8%)	39 (6.2%)	2 (0.3%)	633
Intermediate	1053 (63.8%)	314 (19.0%)	250 (15.2%)	33 (2.0%)	1650
High	46 (32.4%)	33 (23.2%)	52 (36.6%)	11 (7.7%)	142
Total	1597 (65.9%)	441 (18.2%)	341 (14.1%)	46 (1.9%)	2425

Data are expressed as number of patients (%).
ATA, American Thyroid Association; ATA 2015-RTR, the ATA response to therapy re-stratification; ATA 2009-RSS, the initial ATA risk-stratification system; ATA 2015-RSS, the modified ATA initial risk-stratification system.
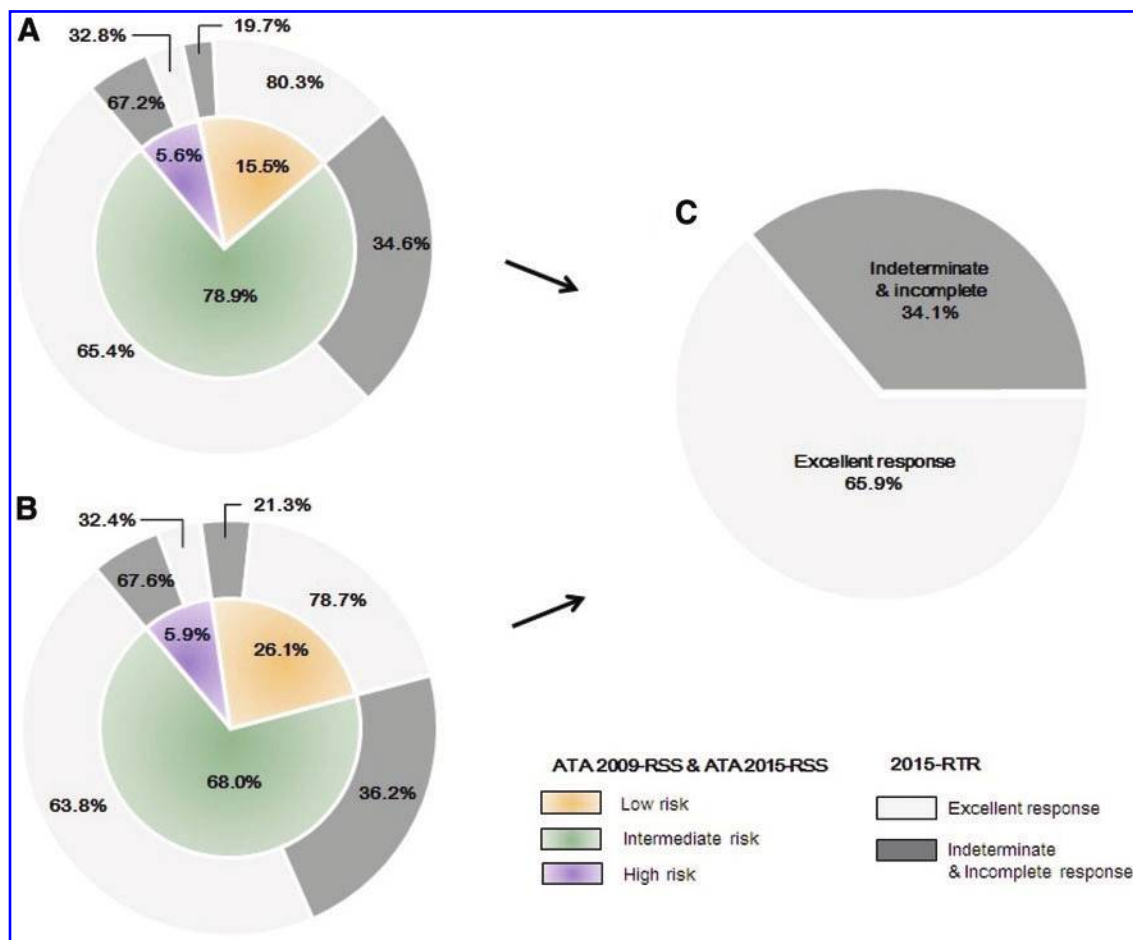
FIG. 1. Distribution of patients, according to (**A**) the initial American Thyroid Association (ATA) risk-stratification system (ATA 2009-RSS), (**B**) the modified ATA initial risk-stratification system (ATA 2015-RSS), and (**C**) the ATA response to therapy re-stratification (ATA 2015-RTR). Color images available online at www.liebertpub.com/thy

prediction, the PVE was calculated for the ATA 2009-RSS, ATA 2015-RSS, and ATA 2015-RTR. The results were 75.13%, 89.35%, and 100.00%, respectively (Table 3). The goodness of fit for predicting recurrence was evaluated by the AIC. The ATA 2009-RSS, ATA 2015-RSS, and ATA 2015-RTR scored 2028.35, 2013.25, and 1829.51, respectively (Table 3). Harrell's c index and iAUC, which evaluated the discriminatory ability of the three risk-stratification systems, were 0.64, 0.67, and 0.83 and 0.63, 0.66, and 0.83 for the ATA 2009-RSS, ATA 2015-RSS, and ATA 2015-RTR, respectively (Table 3). The difference in iAUC was 0.03 [CI 0.014–0.044] between ATA 2009-RSS and ATA 2015-RSS, and 0.17 [CI 0.123–0.217] between ATA 2015-RSS and ATA 2015-RTR. A marked difference in the time-dependent AUC for each risk-stratification system was observed (Fig. 2D). This remained steady throughout the whole follow-up period. Taken together, the results show that although all three risk-stratification systems have the statistical power to predict tumor recurrence, the ATA 2015-RTR is the most accurate.

Discussion

This study shows that all three risk-stratification systems that were examined are useful for predicting the risk of re-currence after initial treatment, although the ATA 2015-RTR was the most accurate.

After implementing the ATA 2015-RSS to the ATA 2009-RSS, some patients migrated from the intermediate- to the low-risk group. This is because the revised definition of low risk included a small volume of LN metastases. However, the proportion of such patients identified herein (13.5%) was smaller than that in a previous report showing 30% migration (18). The relatively small number of migrated patients might be explained by the peculiar characteristics of PTC in Korea. The frequencies of ETE and lateral LN metastases are higher in Korea than they are in Western countries (9–11,18–20). Indeed, aggressive features such as multifocality, bilaterality, and central LN metastases are also more common (9–11,18,21). Alternatively, aggressive PTC might have been higher because the study patients were mostly managed at a tertiary referral hospital. This is particularly true for rare or advanced disease seen at referral hospitals. In addition, $BRAF^{V600E}$ mutation status, a causal genetic event in almost 80% of PTC cases in South Korea, was not included in the analysis. If $BRAF^{V600E}$ mutation status had been included, it is quite possible that migration from the intermediate- to the low-risk group might have been even smaller.

It is interesting that the ATA 2015-RTR predicted tumor recurrence more accurately than the ATA 2009-RSS or ATA

Table 2. Univariate and Multivariate Analyses of Baseline Variables for Disease-Free Survival

|	Univariate analysis		Multivariate analysis	
	HR [CI]	p-Value	HR [CI]	p-Value
Age	0.994 [0.980–1.007]	0.3750	0.995 [0.982–1.008]	0.4167
Sex (female)	0.394 [0.271–0.572]	<0.0001	0.464 [0.318–0.677]	<0.0001
TNM stage[a]				
Stage 1	1	—	—	—
Stage 2	18.728 [8.536–41.089]	<0.0001	—	—
Stage 3	0.569 [0.327–0.988]	0.0452	—	—
Stage 4a	2.492 [1.686–3.683]	<0.0001	—	—
Stage 4b	8.326 [2.206–34.214]	0.0033	—	—
Stage 4c	123.811 [60.342–254.037]	<0.0001	—	—
Tumor size[a]	1.517 [1.357–1.695]	<0.0001	—	—
Multiplicity[a]	1.483 [1.061–2.073]	0.0211	—	—
Bilaterality[a]	1.675 [1.198–2.344]	0.0026	—	—
Extrathyroidal extension[a]	3.689 [2.191–6.213]	<0.0001	—	—
Adjuvant RAI ablation[a]				
No.	1	—	—	—
30 mCi	0.933 [0.377–2.307]	0.8801	—	—
150 mCi	1.788 [0.708–4.516]	0.2191	—	—
200 mCi	2.995 [1.055–8.504]	0.0394	—	—
>200 mCi	9.478 [1.837–48.910]	0.0072	—	—
ATA 2009-RSS				
Low risk	1	—	1	—
Intermediate risk	5.048 [1.859–13.703]	0.0015	4.629 [1.702–12.587]	0.0027
High risk	21.240 [7.460–60.470]	<0.0001	18.607 [6.514–53.154]	<0.0001
ATA 2015-RSS				
Low risk	1	—	1	—
Intermediate risk	6.672 [2.719–16.372]	<0.0001	6.337 [2.582–15.557]	<0.0001
High risk	25.655 [9.949–66.157]	<0.0001	23.244 [8.991–60.089]	<0.0001
ATA 2015-RTR				
Excellent response	1	—	1	—
Indeterminate response	4.209 [2.255–7.854]	<0.0001	4.181 [2.240–7.803]	<0.0001
Biochemical incomplete response	13.792 [8.092–23.507]	<0.0001	13.491 [7.912–23.002]	<0.0001
Structural incomplete response	68.030 [38.844–119.145]	<0.0001	63.870 [36.384–112.119]	<0.0001

[a]These variables were not evaluated in the multivariate analysis because they were included as criteria in the risk stratifications.

HR, hazard ratio; CI, confidence interval; TNM, tumor node metastasis; RAI, radioactive iodine.

2015-RSS. Moreover, the ATA 2015-RTR categorized 65% of all patients into the ''excellent response'' group, with a low rate of recurrence (1.1%), similar to the 1.1% and 0.9% rates predicted by the ATA 2009-RSS and ATA 2015-RSS, respectively. However, the ATA 2009-RSS classified 15.5% of patients as low risk, and the ATA 2015-RSS classified 26.1% as low risk. Therefore, the ATA 2015-RTR might be a better and more cost-effective predictor of tumor recurrence. In fact, 65.4% of patients in the intermediate-risk group and 32.8% in the high-risk group (according to the ATA 2009-RSS) were placed in the ''excellent response'' group by the ATA 2015-RTR. Similarly, 63.8% of patients in the intermediate-risk group and 32.4% in the high-risk group according to the ATA 2015-RSS were reclassified as excellent responders. Taken together, these data suggest that the effectiveness of therapy is a crucial factor affecting PTC recurrence, although the clinicopathologic behavior of PTC might also be an important factor. Moreover, it was noteworthy that compared with initial stratification systems, the ATA 2015-RTR had the advantage of one year response to therapy.

Considering that the current ATA guidelines do not currently mandate prophylactic CCND, particularly for T1–2

PTC, an attempt was made to reanalyze the accuracy of risk stratification assuming that prophylactic CCND had not been performed on these patients. Data were filtered to exclude the possibility of upstaging PTC due to routine prophylactic CCND, and a new presumption model of risk stratification was designed. In the new model, all those with prophylactic CCND in T1–2 PTC were reclassified as N0. After applying the new presumption model, Kaplan–Meier survival analysis showed there were significant differences between ATA 2009-RSS and ATA 2015-RSS across all stages ($p < 0.001$; Supplementary Fig. S1A and B). To examine the prediction accuracy of each risk-stratification system, PVE, AIC, iAUC, and Harrell's c index were also examined for the ATA 2009-RSS and ATA 2015-RSS (Supplementary Fig. S1C and Supplementary Table S3). According to the results of the new presumption model, if the possibility of PTC upstaging due to prophylactic CCND was excluded, the ATA 2015-RTR was significantly more accurate than the ATA 2009-RTR.

Four different types of statistical analyses were used to compare the predictive accuracy of the three risk-stratification systems. Because the PVE identifies differences between predicted risk and actual risk, it is normally expressed as a
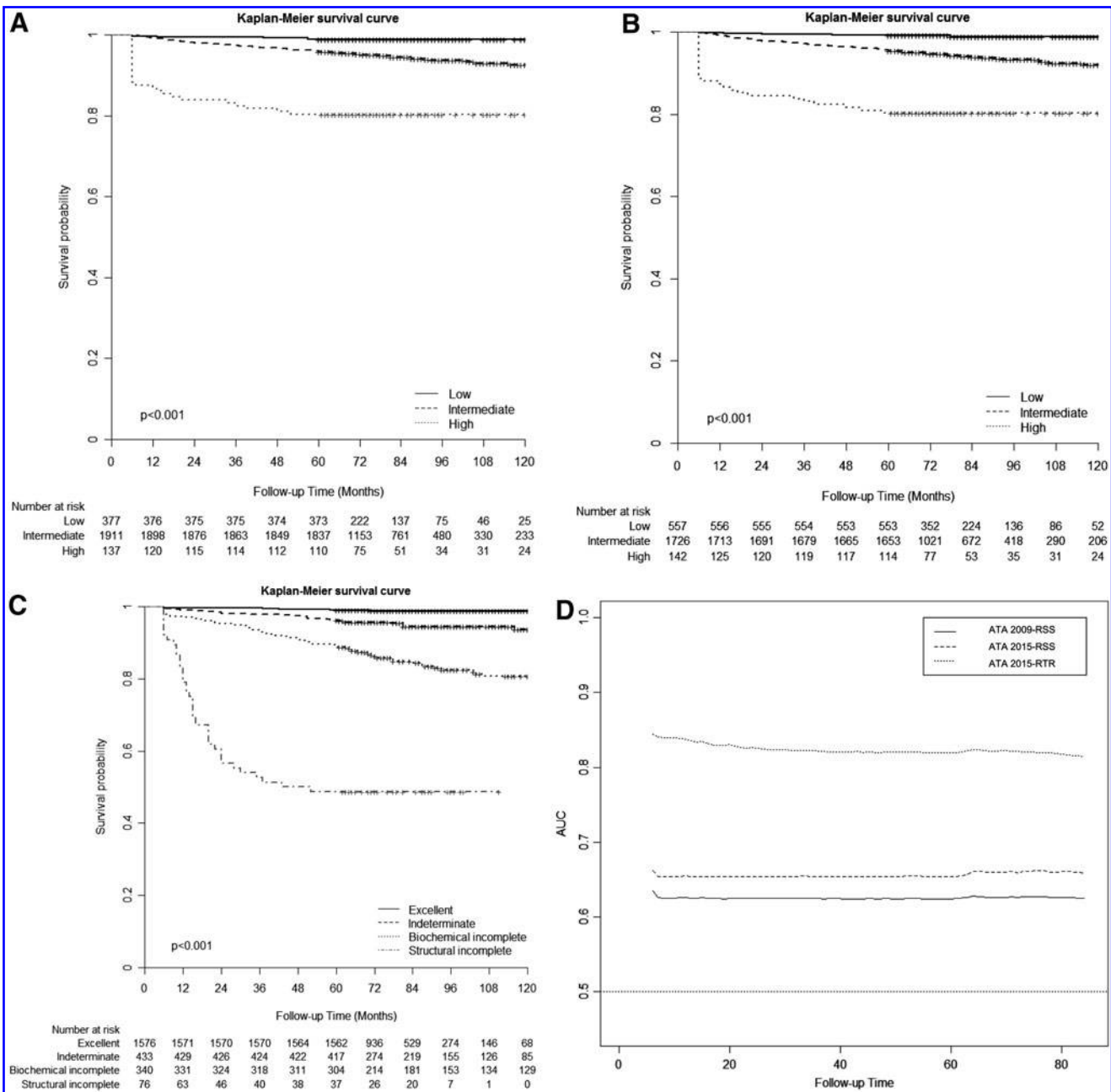
FIG. 2. Probability of disease-free survival according to (**A**) the ATA 2009-RSS, (**B**) the ATA 2015-RSS, and (**C**) the ATA 2015-RTR. The three systems showed statistically significant differences across all groups of patients. (**D**) Predictive accuracy for disease-free survival: time-dependent area under the curve according to follow-up time.

TABLE 3. COMPARISON OF THE PERFORMANCE AND PREDICTIVE ACCURACY
OF THE ATA 2009-RSS, ATA 2015-RSS, AND ATA 2015-RTR

	PVE(%)	AIC	Harrell's c index (SE)	iAUC [CI]
ATA 2009-RSS	75.13	2028.85	0.64 [0.599–0.672]	0.63 [0.595–0.667]
ATA 2015-RSS	89.35	2013.25	0.67 [0.630–0.701]	0.66 [0.624–0.694]
ATA 2015-RTR	100.00	1829.51	0.83 [0.792–0.868]	0.83 [0.794–0.865]

PVE, proportion of variance explained; AIC, Akaike information criterion; SE, standard error; iAUC, integrated area under the curve.

range (from 0 to 100) (12). The 100% PVE for the ATA 2015-RTR indicated that this system exactly matches the actual risk and explains all of the variance. Because AIC measures the relative quality of statistical models and provides a means for model selection, it was used to compare the different models; a smaller AIC indicates a preferred model. Here, it was found that the ATA 2015-RTR had the lowest AIC value. Harrell's c index and iAUC were evaluated for discrimination and predictive ability of the survival model. The bootstrapping method compares predictive accuracy over the entire follow-up period. Again, the ATA 2015-RTR showed the highest Harrell's c index and the highest iAUC. Thus, the ATA 2015-RTR is the most accurate risk-stratification system, although it cannot be used at the preoperative stage.

Conclusion

To the best of the authors' knowledge, this is the first large cohort study to compare the predictive accuracy of the ATA 2009-RSS, ATA 2015-RSS, and ATA 2015-RTR. It was found that 13.5% of patients assessed as intermediate risk by the ATA 2009-RSS were assessed as low risk by the ATA 2015-RSS. In addition, the ATA 2015-RTR is likely to be the most accurate predictor of recurrence, although the data did not include $BRAF^{V600E}$ and $TERT$ promoter mutation status, and the mean follow-up period was relatively short (88.0 months). Further studies should examine the clinical utility of the ATA 2015-RTR.

Acknowledgments

Author Disclosure Statement

No competing financial interests exist.

References

1. Chen AY, Jemal A, Ward EM 2009 Increasing incidence of differentiated thyroid cancer in the United States, 1988–2005. Cancer **115:**3801–3807.
2. Xing MZ, Westra WH, Tufano RP, Cohen Y, Rosenbaum E, Rhoden KJ, Carson KA, Vasko V, Larin A, Tallini G, Tolaney S, Holt EH, Hui P, Umbricht CB, Basaria S, Ewertz M, Tufaro AP, Califano JA, Ringel MD, Zeiger MA, Sidransky D, Ladenson PW 2005 *BRAF* mutation predicts a poorer clinical prognosis for papillary thyroid cancer. J Clin Endocrinol Metab **90:**6373–6379.
3. Xing MZ, Liu RY, Liu XL, Murugan AK, Zhu GW, Zeiger MA, Pai S, Bishop J 2014 *BRAF V600E* and *TERT* promoter mutations cooperatively identify the most aggressive papillary thyroid cancer with highest recurrence. J Clin Oncol **32:**2718–2726.
4. Mazzaferri E 2007 Management of low-risk differentiated thyroid cancer. Endocr Pract **13:**498–512.
5. Roti E, Uberti ECD, Bondanelli M, Braverman LE 2008 Thyroid papillary microcarcinoma: a descriptive and meta-analysis study. Eur J Endocrinol **159:**659–673.
6. Riemann B, Kramer JA, Schmid KW, Dralle H, Dietlein M, Schicha H, Sauerland C, Frankewitsch T, Schober O; MSDS Study Group 2010 Risk stratification of patients with locally aggressive differentiated thyroid cancer. Results of the MSDS trial. Nuklearmedizin **49:** 79–84.
7. Randolph GW, Duh QY, Heller KS, LiVolsi VA, Mandel SJ, Steward DL, Tufano RP, Tuttle RM; American Thyroid Association Surgical Affairs Committee's Taskforce on Thyroid Cancer Nodal Surgery 2012 The prognostic significance of nodal metastases from papillary thyroid carcinoma can be stratified based on the size and number of metastatic lymph nodes, as well as the presence of extra-nodal extension. Thyroid **22:**1144–1152.
8. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, Pacini F, Randolph GW, Sawka AM, Schlumberger M, Schuff KG, Sherman SI, Sosa JA, Steward DL, Tuttle RM, Wartofsky L 2016 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid **26:**1–133.
9. Castagna MG, Maino F, Cipri C, Belardini V, Theodoropoulou A, Cevenini G, Pacini F 2011 Delayed risk stratification, to include the response to initial treatment (surgery and radioiodine ablation), has better outcome predictivity in differentiated thyroid cancer patients. Eur J Endocrinol **165:**441–446.
10. Vaisman F, Momesso D, Bulzico DA, Pessoa CH, Dias F, Corbo R, Vaisman M, Tuttle RM 2012 Spontaneous remission in thyroid cancer patients after biochemical incomplete response to initial therapy. Clin Endocrinol (Oxf) **77:**132–138.
11. 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.
12. Schemper M 1993 The relative importance of prognostic factors in studies of survival. Stat Med **12:**2377–2382.
13. Akaike H 1981 Citation classic—a new look at the statistical-model identification. CC/Eng Tech Appl Sci **51:** 22–22.
14. Harrell FE, Lee KL, Mark DB 1996 Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med **15:**361–387.
15. Heagerty PJ, Lumley T, Pepe MS 2000 Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics **56:**337–344.
16. Heagerty PJ, Zheng Y 2005 Survival model predictive accuracy and ROC curves. Biometrics **61:**92–105.
17. Efron B 1986 How biased is the apparent error rate of a prediction rule? J Am Stat Assoc **81:**461–470.

18. Pitoia F, Jerkovich F, Urciuoli C, Schmidt A, Abelleira E, Bueno F, Cross G, Tuttle RM 2015 Implementing the modified 2009 American Thyroid Association risk stratification system in thyroid cancer patients with low and intermediate risk of recurrence. Thyroid **25:**1235–1242.

19. Cho BY, Choi HS, Park YJ, Lim JA, Ahn HY, Lee EK, Kim KW, Yi KH, Chung JK, Youn YK, Cho NH, Park DJ, Koh CS 2013 Changes in the clinicopathological characteristics and outcomes of thyroid cancer in Korea over the past four decades. Thyroid **23:**797–804.

20. Suh YJ, Kwon H, Kim S-j, Choi JY, Lee KE, Park YJ, Park DJ, Youn Y-K 2015 Factors affecting the locoregional recurrence of conventional papillary thyroid carcinoma after surgery: a retrospective analysis of 3381 patients. Ann Surg Oncol **22:**3543–3549.

21. Pitoia F, Bueno F, Urciuoli C, Abelleira E, Cross G, Tuttle RM 2013 Outcomes of patients with differentiated thyroid cancer risk-stratified according to the American Thyroid Association and Latin American Thyroid Society risk of recurrence classification systems. Thyroid **23:**1401–1407.

Address correspondence to:
Young Suk Jo, MD, PhD
Department of Internal Medicine
Open NBI Convergence Technology Research Laboratory
Severance Hospital, Yonsei Cancer Center
Yonsei University College of Medicine
03772, Seoul
Korea

E-mail: joys@yuhs.ac

or

Jandee Lee, MD, PhD
Department of Surgery
Open NBI Convergence Technology Research Laboratory
Severance Hospital, Yonsei Cancer Center
Yonsei University College of Medicine
03722, Seoul
Korea

E-mail: jandee@yuhs.ac