Original Article

Balkan Med J

Integrating the Interaction Between Extrathyroidal Extension and Tumor Size to Optimize T Stage in Differentiated Thyroid Cancer

Hui Ouyang[1,2], Xiaolin Dou[1,2], Botao Sun[1,2], Jinlong Chen[1,2], Hongyan Zai[1,2], Xinying Li[1,2], Fada Xia[1,2]

[1]Department of General Surgery, Xiangya Hospital, Central South University, Changsha, China
[2]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

Background: The American Joint Committee on Cancer (AJCC) T stage for differentiated thyroid cancer (DTC) has limited prognostic stratification capability, and the interaction between extrathyroidal extension (ETE) and tumor size remains inadequately characterized.

Aims: To investigate the interaction between ETE and tumor size and to develop a modified T stage with improved prognostic stratification for DTC.

Study Design: Retrospective, population-based cohort study.

Methods: We analyzed 163,616 patients with DTC from the Surveillance, Epidemiology, and End Results database, divided into development (n = 106,516; 2004-2015) and validation (n = 57,100; 2016-2021) cohorts. Disease-specific survival was the primary outcome. We evaluated the interaction between ETE and tumor size on additive and multiplicative scales and developed a modified T stage using recursive partitioning analysis. Performance assessment included discrimination (C-index), calibration (Brier score), and five standard cancer staging criteria.

Results: Significant positive additive interactions were observed between ETE and tumor size, with relative excess risk due to interaction (RERI) increasing from minimal ETE (MinETE) [RERI = 1.90; 95% confidence interval (CI): 1.02-2.78] to Gross4b ETE (RERI = 6.55; 95% CI: 2.74-10.35). In tumors > 2 cm, MinETE was associated with a 4.58-fold increased risk (95% CI: 3.77-5.55), compared with a 2.34-fold increase (95% CI: 2.05-2.66) in smaller tumors, indicating size-dependent prognostic effects. The modified T stage incorporating these interactions demonstrated improved discrimination (C-index: 0.775; 95% CI: 0.770-0.779) compared with the AJCC 8th edition T stage (C-index: 0.766; 95% CI: 0.761-0.770). This represents a statistically significant but numerically modest improvement, with consistent enhancement across all performance metrics. The modified system also improved prediction of lymph node and distant metastases and provided clearer tumor-node-metastasis stage separation.

Conclusion: The modified T stage incorporating size-dependent effects of ETE was associated with improved prognostic stratification compared with the AJCC 8th edition. These findings support consideration of tumor size-ETE interaction in future staging refinements to enable more precise risk stratification and personalized treatment strategies for patients with DTC. Prospective, multicenter validation is warranted to confirm these findings and evaluate their impact on clinical decision-making.

INTRODUCTION

Differentiated thyroid cancer (DTC) generally has an excellent prognosis, with a disease-specific survival (DSS) rate exceeding 90%. However, survival varies significantly across stages, declining from 99.6% in stage I to 48.6% in stage IV,[1] highlighting the need for precise staging to guide treatment and surveillance. In the current American Joint Committee on Cancer (AJCC) 8th edition tumor-node-metastasis (TNM) staging system,[2] the T stage serves as a cornerstone of tumor classification from stages I to IVA, incorporating both tumor size (T1-T3a) and extrathyroidal extension (ETE) extent (T3b-T4b). Notable revisions from the 7th edition[3] include the elimination of minimal ETE (MinETE) from T classification and subdivision of T3 into T3a (> 4 cm, intrathyroidal) and T3b (gross ETE to strap muscles).

Corresponding author: Xinying Li, Department of General Surgery, Xiangya Hospital, Central South University; National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

e-mail: lixinyingcn@csu.edu.cn

Corresponding author: Fada Xia, Department of General Surgery, Xiangya Hospital, Central South University; National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

e-mail: xiafada@xiangya.com.cn

Received: January 27, 2026 **Accepted:** April 16, 2026 **Available Online Date:** • **DOI:** 10.4274/balkanmedj.galenos.2026.2026-1-269

Available at www.balkanmedicaljournal.org

ORCID iDs of the authors: H.O. 0000-0001-7687-1185; X.D. 0009-0003-3972-4049; B.S. 0000-0001-9068-1687; J.C. 0000-0003-0799-045X; H.Z. 0009-0003-8964-1591; X.L. 0000-0001-9835-2191; F.X. 0000-0002-8904-897X.

Cite this article as: Ouyang H, Dou X, Sun B, et al. Integrating the Interaction Between Extrathyroidal Extension and Tumor Size to Optimize T Stage in Differentiated Thyroid Cancer. Balkan Med J. [Epub Ahead of Print]

However, these changes remain controversial due to limited clinical evidence.[4] While some studies suggest that MinETE lacks prognostic significance,[5-10] emerging data indicate an independent association with adverse outcomes,[4,11-18] warranting reconsideration in future staging systems.[11,12,16] These conflicting findings may be partly explained by the interaction between MinETE and tumor size, as recent research suggests that the prognostic implications of MinETE vary according to tumor size.[11,13,19] Notably, the 2015 American Thyroid Association (ATA) guidelines[20] classify patients with MinETE as low-to-intermediate risk regardless of tumor size and generally recommend adjuvant radioactive iodine therapy. This discrepancy in the prognostic interpretation of MinETE between the ATA and AJCC guidelines further underscores the need for a comprehensive reevaluation of the T stage.

Furthermore, the empirical introduction of T3b has sparked considerable discussion.[21-25] A retrospective analysis[25] of 2,579 patients with DTC found that T3b tumors had comparable survival to T2 tumors and showed a trend toward better DSS than T3a tumors, suggesting the need for potential revision of the T3b category. These findings were further corroborated by Song et al.[23] in a cohort study of 3,104 patients with DTC, which demonstrated no significant DSS differences between T3b tumors (≤ 4 cm) and T2 tumors. More recently, Park et al.[21] reported in 6,282 patients with DTC that small T3b tumors (≤ 2 cm) exhibited survival outcomes similar to those of T1 tumors, further challenging the current T stage classification. Collectively, these studies demonstrate the variable prognostic significance of the T3b classification across different tumor sizes, emphasizing the need for a refined T stage that accounts for tumor size-ETE interactions.

This study hypothesized that the current AJCC T stage for DTC could be optimized by considering the interaction between ETE and tumor size. Using a large national cohort, we aimed to (1) assess the interaction between ETE and tumor size on DSS; (2) develop a modified T stage reclassifying MinETE and T3b-T4b based on tumor size; (3) compare the modified T stage with existing T stage systems; and (4) evaluate the clinical utility of the modified T stage by examining its ability to predict lymph node and distant metastases and its role in improving TNM stage stratification.

MATERIALS AND METHODS

Data source and patient cohort

Data were obtained from the Surveillance, Epidemiology, and End Results (SEER) database. We included adult patients (≥ 18 years) diagnosed with DTC between 2004 and 2021. Cases were identified using ICD-O-3 codes for papillary thyroid carcinoma (PTC; 8050, 8260, 8340-8344, 8350, 8450-8460), follicular thyroid carcinoma (FTC; 8330-8332, 8335, 8339), and oncocytic carcinoma (OCA; 8290). Exclusion criteria were unknown clinicopathological or follow-up data, follow-up < 1 month, or presence of distant metastases (M1) at diagnosis. Of 190,383 initially eligible patients with DTC, 21,499 (11.3%) were excluded due to missing clinicopathological data, primarily unknown ETE status (n = 10,967; 5.8%) and unknown tumor size (n = 6,233; 3.3%), followed by unknown nodal status (n = 3,777; 2.0%). An additional 5,268 patients were excluded due to distant metastases at diagnosis (n = 3,047) or follow-up < 1 month (n = 2,221). The final cohort comprised 163,616 patients, divided into a development cohort (2004-2015; n = 106,516) and a temporal validation cohort (2016-2021; n = 57,100) (Supplementary Figure 1). No a priori sample size calculation was performed; all eligible patients meeting the predefined inclusion criteria were included to maximize statistical power. As the SEER database is publicly available and de-identified, this study was exempt from Institutional Review Board approval.

Variables and outcomes

Collected variables included demographic characteristics (age, sex, race/ethnicity, income, county type, and marital status), clinicopathologic features (histology, tumor size, ETE, and nodal status), treatment variables (surgery and radiotherapy), and follow-up data (survival time and cause of death). The primary exposures of interest were ETE and tumor size. ETE was classified into five categories: no ETE (No), MinETE, gross invasion of pericapsular soft tissue/connective tissue or strap muscles (Gross3b), gross invasion of subcutaneous soft tissues, larynx, trachea, esophagus, or recurrent laryngeal nerve (Gross4a), and gross invasion of the prevertebral fascia or encasement of the carotid artery or mediastinal vessels (Gross4b). The primary outcome was DSS, defined as the time from diagnosis to thyroid cancer-related death or last follow-up.

Interaction analysis between ETE and tumor size

We visualized the relationship between tumor size stratified by ETE and DSS using dependence and partial dependence plots derived from the Cox proportional hazards model. Dependence plots illustrate unadjusted survival trends, whereas partial dependence plots adjust for covariates, reflecting the adjusted effect of the variable.[26,27] We further evaluated the interaction between ETE and tumor size on both multiplicative and additive scales.[28,29] Detailed methods are provided in the Supplementary Methods.

Recursive partitioning analyses for modified T stage development

Recursive partitioning analysis (RPA)[30-32] was used to develop a modified T stage system based on tumor size (≤ 2 cm, 2–4 cm, or > 4 cm) and ETE categories in the development cohort. After RPA, 10-year DSS rates and hazard ratios (HRs) were calculated for each terminal node group. The final modified T stage was constructed by combining terminal groups with homogeneous survival patterns, considering both statistical significance and clinical relevance. Detailed methods are provided in the Supplementary Methods.

Performance evaluation of the modified T stage

We compared the modified T stage with the AJCC 7th and 8th T stages as well as other proposed staging systems (namely, the 2024 Cancers T stage[21] and the 2019 Thyroid T stage[23]), using: (1) survival curves, HRs, and 10-year DSS differences to assess survival discrimination; (2) discrimination (Harrell's C-index) and calibration (Brier score),[26] with 1,000 bootstrap replicates for internal validation; and (3) five widely established staging criteria,[33-35] including hazard consistency,

discrimination, explained variation, likelihood difference, and sample size balance. Detailed methods are provided in the Supplementary Methods.

Generalizability of the modified T stage

First, we compared the performance of the modified T stage with other T stage schemas using the temporal validation cohort (2016-2021). Additionally, we used overall survival (OS) as an alternative endpoint to evaluate the prognostic performance of the modified T classification compared with other proposed T schemas. Furthermore, we conducted subgroup analyses to assess the consistency of the modified T stage across age groups (< 55 and ≥ 55 years) and histological subtypes (PTC, FTC, and OCA). Notably, OS-based analyses and subgroup evaluations were performed in the entire cohort with internal validation using 1,000 bootstrap replicates.

Clinical significance of the modified T stage

To assess the clinical significance of the modified T stage, we examined its association with lymph node and distant metastases as well as its impact on the TNM staging system. Detailed methods are provided in the Supplementary Methods.

Statistical analysis

Categorical variables were summarized as frequencies (%) and compared using the chi-square (χ^2) test. Continuous variables were reported as medians with interquartile ranges or means with standard deviations, as appropriate, and analyzed using the Mann-Whitney U test. Multivariable Cox proportional hazards regression was applied throughout the study, adjusting for age, sex, race/ethnicity, marital status, household income, county type, histologic type, N stage, surgery, and radiotherapy. The proportional hazards assumption was assessed for the interaction covariates (ETE and tumor size) and the modified T stage in the multivariable Cox model using Schoenfeld residual tests. Both individual and global tests were non-significant (all $p > 0.05$), indicating no violation of the assumption. The overall analytical workflow, including interaction analysis, recursive partitioning, performance evaluation, and generalizability assessment, is detailed in the Supplementary Methods. All statistical analyses were performed using R version 4.3.1 (Supplementary Table 1), with statistical significance set at $p < 0.05$.

RESULTS

Patient characteristics

A total of 163,616 patients with DTC were included (Supplementary Figure 1). As shown in Table 1, the development cohort comprised 106,516 patients, with 1,805 (1.7%) disease-specific deaths and a median follow-up of 121 months. The validation cohort comprised 57,100 patients, with 258 (0.5%) deaths and a median follow-up of 32 months. Despite statistically significant differences, the demographic and clinical characteristics were generally comparable between the two cohorts. However, the validation cohort showed notable differences in treatment patterns, with higher rates of lobectomy (21.6% vs. 13.7%) and lower use of radioactive iodine therapy (30.0% vs. 45.9%).

Interaction analysis between ETE and tumor size

Tumor size distribution across ETE categories (Supplementary Figure 2) showed that as ETE severity increased from no ETE to Gross4b, the proportion of small tumors (≤ 2 cm) decreased from 71.6% to 25.3%, whereas the proportion of larger tumors (> 4 cm) increased from 6.7% to 35.1%. Notably, even among Gross4a/4b cases, a substantial proportion of tumors remained small (≤ 2 cm: > 25%; ≤ 4 cm: > 60%).

The dependence plot (Figure 1a) showed that unadjusted 10-year DSS rates declined with increasing ETE severity across all tumor sizes, indicating a size-dependent deterioration in survival. The partial dependence plot (Figure 1b) confirmed these trends after adjustment for covariates. Interaction analysis (Supplementary Table 2) demonstrated significant positive additive interactions, with an increasing relative excess risk due to interaction (RERI) as ETE severity increased (MinETE: 1.90 to Gross4b: 6.55). A sensitivity analysis using restricted cubic splines in a Poisson regression model confirmed that the interaction between ETE and tumor size persisted when tumor size was modeled as a continuous variable (likelihood ratio test for interaction: all $p < 0.001$ across ETE categories; Supplementary Figure 3).

The primary analytical framework for interaction analysis was the cause-specific Cox proportional hazards model, from which HR-based RERI estimates (Supplementary Table 2) and RPA were derived. Complementary sensitivity analyses included risk-based RERI estimation using cumulative incidence functions (CIFs) (Supplementary Table 3) and restricted cubic spline Poisson regression with tumor size treated as a continuous variable (Supplementary Figure 3), to evaluate robustness under competing-risk and continuous-variable frameworks, respectively. The RERI strata included sufficient event counts for stable estimation. The reference strata comprised 433 disease-specific deaths among 65,295 patients (no ETE, ≤ 2 cm) and 441 deaths among 25,014 patients (no ETE, > 2 cm). The ETE-exposed strata ranged from 23 deaths among 236 patients (Gross4b, ≤ 2 cm) to 244 deaths among 1,297 patients (Gross4a, > 2 cm), with detailed distributions reported in Supplementary Table 3. Furthermore, a cumulative incidence-based analysis (Supplementary Table 3) was performed using two complementary approaches: the Aalen-Johansen CIF, which accounts for non-cancer death as a competing event and served as the primary estimate, and the Kaplan-Meier method, retained as a secondary reference given its known tendency to overestimate cumulative incidence when competing risks are treated as censored. Both approaches yielded concordant positive additive interactions on the absolute risk scale [all 95% bootstrap confidence interval (CI) excluding zero], confirming robustness of the interaction under a competing-risk framework.

TABLE 1. Patient Characteristics in the Development and Temporal Validation Cohorts.

Characteristics	Overall n = 163,616[1]	Development cohort (diagnosis years 2004-2015) n = 106,516[1]	Validation cohort (diagnosis years 2016-2021) n = 57,100[1]	*p* value[2]
Age (years)	50.00 (39.00, 61.00)	50.00 (39.00, 60.00)	51.00 (39.00, 62.00)	< 0.001
Age category (years)				< 0.001
< 55	99.581 (60.9%)	66.264 (62.2%)	33.317 (58.3%)	
≥ 55	64.035 (39.1%)	40.252 (37.8%)	23.783 (41.7%)	
Sex				< 0.001
Female	125.558 (76.7%)	82.370 (77.3%)	43.188 (75.6%)	
Male	38.058 (23.3%)	24.146 (22.7%)	13.912 (24.4%)	
Race/ethnicity				< 0.001
Non-Hispanic White	105.969 (64.8%)	71.627 (67.2%)	34.342 (60.1%)	
Non-Hispanic Black	10.135 (6.2%)	6.489 (6.1%)	3.646 (6.4%)	
Non-Hispanic Asian or Pacific Islander	17.770 (10.9%)	10.914 (10.2%)	6.856 (12.0%)	
Non-Hispanic other/unknown	2.102 (1.3%)	1.175 (1.1%)	927 (1.6%)	
Hispanic	27.640 (16.9%)	16.311 (15.3%)	11.329 (19.8%)	
Marital status				< 0.001
Married	101.219 (61.9%)	67.296 (63.2%)	33.923 (59.4%)	
Single	33.825 (20.7%)	20.636 (19.4%)	13.189 (23.1%)	
Divorced/widowed/separated	20.702 (12.7%)	13.703 (12.9%)	6.999 (12.3%)	
Unknown	7.870 (4.8%)	4.881 (4.6%)	2.989 (5.2%)	
Household income				< 0.001
Low	45.602 (27.9%)	32.629 (30.6%)	12.973 (22.7%)	
Middle	65.806 (40.2%)	43.114 (40.5%)	22.692 (39.7%)	
High	52.208 (31.9%)	30.773 (28.9%)	21.435 (37.5%)	
County type				0.852
Urban	147.863 (90.4%)	96.250 (90.4%)	51.613 (90.4%)	
Rural	15.753 (9.6%)	10.266 (9.6%)	5.487 (9.6%)	
Histologic types				< 0.001
PTC	151.896 (92.8%)	98.641 (92.6%)	53.255 (93.3%)	
FTC	8.228 (5.0%)	5.481 (5.1%)	2.747 (4.8%)	
OCA	3.492 (2.1%)	2.394 (2.2%)	1.098 (1.9%)	
Tumor size (mm)	14.00 (7.00, 25.00)	14.00 (7.00, 25.00)	14.00 (8.00, 25.00)	< 0.001
Tumor size (cm)				< 0.001
≤ 2 cm	112.122 (68.5%)	73.567 (69.1%)	38.555 (67.5%)	
2-4 cm	38.171 (23.3%)	24.175 (22.7%)	13.996 (24.5%)	
> 4 cm	13.323 (8.1%)	8.774 (8.2%)	4.549 (8.0%)	
Extrathyroidal extension				< 0.001
No	142.464 (87.1%)	90.309 (84.8%)	52.155 (91.3%)	
MinETE	9.020 (5.5%)	7.100 (6.7%)	1.920 (3.4%)	
Gross3b	7.794 (4.8%)	6.196 (5.8%)	1.598 (2.8%)	
Gross4a	3.251 (2.0%)	2.066 (1.9%)	1.185 (2.1%)	
Gross4b	1.087 (0.7%)	845 (0.8%)	242 (0.4%)	

TABLE 1. Continued.

Characteristics	Overall n = 163,616[1]	Development cohort (diagnosis years 2004-2015) n = 106,516[1]	Validation cohort (diagnosis years 2016-2021) n = 57,100[1]	*p* value[2]
N stage				< 0.001
N0	50.578 (30.9%)	32.518 (30.5%)	18.060 (31.6%)	
Nx (without node examined)	79.866 (48.8%)	53.817 (50.5%)	26.049 (45.6%)	
N1a	21.908 (13.4%)	13.668 (12.8%)	8.240 (14.4%)	
N1b	11.264 (6.9%)	6.513 (6.1%)	4.751 (8.3%)	
Surgery				< 0.001
Lobectomy	26.899 (16.4%)	14.573 (13.7%)	12.326 (21.6%)	
Total thyroidectomy	132.674 (81.1%)	89.726 (84.2%)	42.948 (75.2%)	
Less than lobectomy	4.043 (2.5%)	2.217 (2.1%)	1.826 (3.2%)	
Radiotherapy				< 0.001
Isotopes	66.077 (40.4%)	48.939 (45.9%)	17.138 (30.0%)	
Beam/implants	2.669 (1.6%)	2.049 (1.9%)	620 (1.1%)	
None/unknown	94.870 (58.0%)	55.528 (52.1%)	39.342 (68.9%)	
Follow-up (months)	88.00 (40.00, 137.00)	121.00 (90.00, 159.00)	32.00 (15.00, 49.00)	< 0.001
Death				< 0.001
Alive	150.183 (91.8%)	94.559 (88.8%)	55.624 (97.4%)	
Death of cancer	2.063 (1.3%)	1.805 (1.7%)	258 (0.5%)	
Death of other causes	11.370 (6.9%)	10.152 (9.5%)	1.218 (2.1%)	

[1]Median (Q1, Q3); n (%).
[2]Wilcoxon rank sum test; Pearson's chi-squared test. PTC, papillary thyroid carcinoma; FTC, follicular thyroid carcinoma; MinETE, minimal extrathyroidal extension.

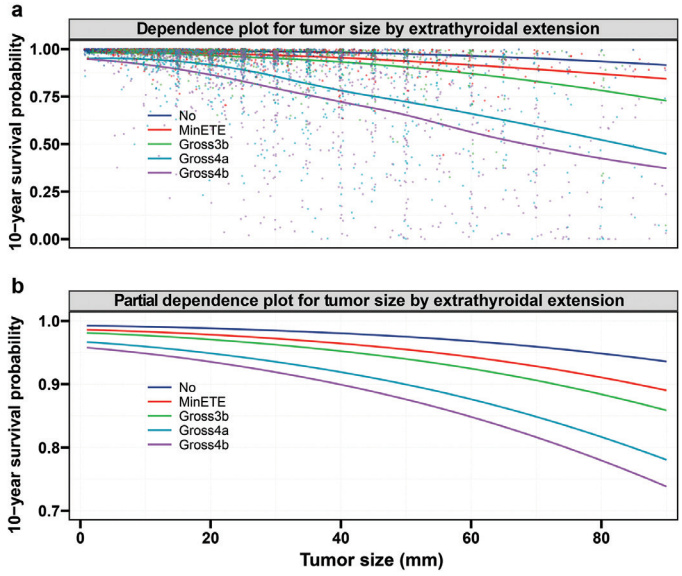

FIG. 1. Dependency plots and partial dependency plots for extrathyroidal extension across tumor size predicting 10-year survival probability. Dependence plot (a) showing unadjusted survival trends across tumor sizes stratified by ETE categories. Each point represents an individual patient, with locally estimated scatterplot smoothing curves fitted to visualize survival trends across five ETE categories. Partial dependence plot (b) demonstrating adjusted survival probabilities after controlling for age, sex, race/ethnicity, marital status, household income, county type, histologic type, N stage, surgery, and radiotherapy. ETE, extrathyroidal extension; MinETE, minimal extrathyroidal extension.

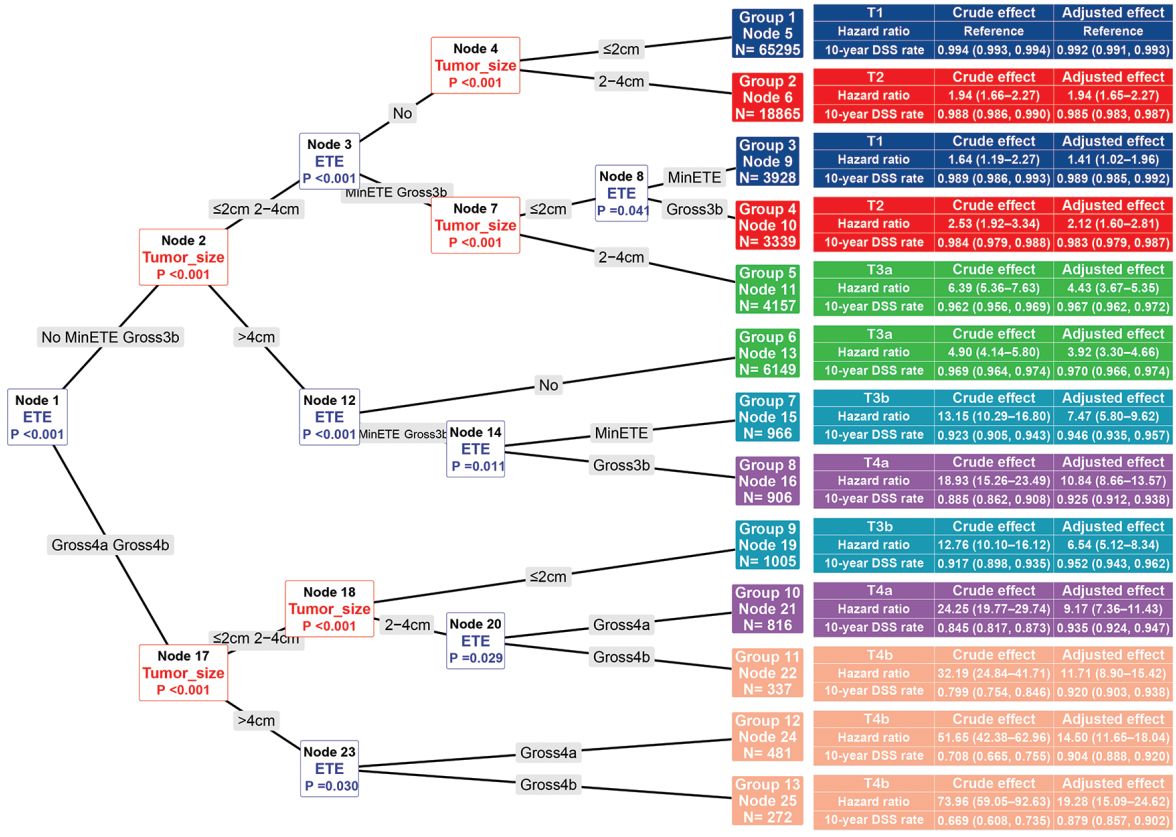

FIG. 2. Development of modified T stage through recursive partitioning analysis in the development cohort. This classification tree illustrates the hierarchical stratification of patients based on extrathyroidal extension and tumor size categories using recursive partitioning analysis. The algorithm recursively divided the cohort to maximize survival differences between subgroups, with splitting criteria determined by log-rank tests with Bonferroni adjustment. Internal nodes display the splitting variable (ETE or tumor size) and the corresponding Bonferroni-adjusted *p* value, while terminal nodes show the patient number in each final group. The 13 terminal node groups identified by the analysis were subsequently combined into six modified T stage categories (T1 through T4b) based on similar survival patterns and clinical relevance, as indicated by color coding: T1 (light blue), T2 (blue), T3a (green), T3b (yellow), T4a (orange), and T4b (red). ETE, extrathyroidal extension; N, number of patients; DSS, disease-specific survival; HR, hazard ratio; CI, confidence interval; inETE, minimal extrathyroidal extension.

RPA for modified T-stage development

RPA (Figure 2) stratified patients into 13 terminal node groups based on ETE and tumor size. The event counts (disease-specific deaths/total) across terminal nodes ranged from 40/3,928 (Group 3) to 433/65,295 (Group 1), with all nodes containing ≥ 40 events. Firth penalized Cox regression applied to the four nodes with fewer than 80 events produced HR estimates consistent with standard Cox estimates, confirming the stability of effect estimates (Supplementary Table 4). After combining groups with similar survival patterns, we developed a modified T stage comprising six distinct groups (Supplementary Figures 4-6 and Figure 2). The modified T stage and its changes relative to the AJCC 8th edition are summarized in Supplementary Table 5: T1 included tumor ≤ 2 cm with no ETE or MinETE (unchanged); T2 included tumor 2-4 cm with no ETE, or tumor ≤ 2 cm with Gross3b, incorporating tumors ≤ 2 cm with Gross3b and excluding 2-4 cm tumors with MinETE; T3a included tumor > 4 cm with no ETE, or tumor 2-4 cm with MinETE or Gross3b, incorporating tumors 2-4 cm with MinETE or Gross3b and excluding tumor > 4 cm with MinETE; T3b included tumor > 4 cm with MinETE, or tumor ≤ 2 cm with Gross4a or Gross4b, incorporating size-specific ETE patterns; T4a included tumor > 4 cm with Gross3b, or tumor 2-4 cm with Gross4a, incorporating tumor > 4 cm with Gross3b and excluding ≤ 2 cm or > 4 cm tumors with Gross4a; and T4b included tumor 2-4 cm with Gross4b, or tumor > 4 cm with Gross4a or Gross4b, incorporating Gross4a > 4 cm and excluding tumor ≤ 2 cm with Gross4b. The Sankey diagram (Supplementary Figure 7) showed that patients initially classified as T1, T2, and T3a without ETE in the AJCC 8th edition generally retained their classification in the modified T stage. However, substantial reclassification was observed among patients with ETE, with smaller tumors more often downstaged and larger tumors more often upstaged.

Performance evaluation of the modified T stage

The modified T stage demonstrated superior prognostic stratification compared with existing systems, with clear separation in DSS between stages (Supplementary Figure 8). HRs and 10-year survival differences across T categories further supported this finding (Figure

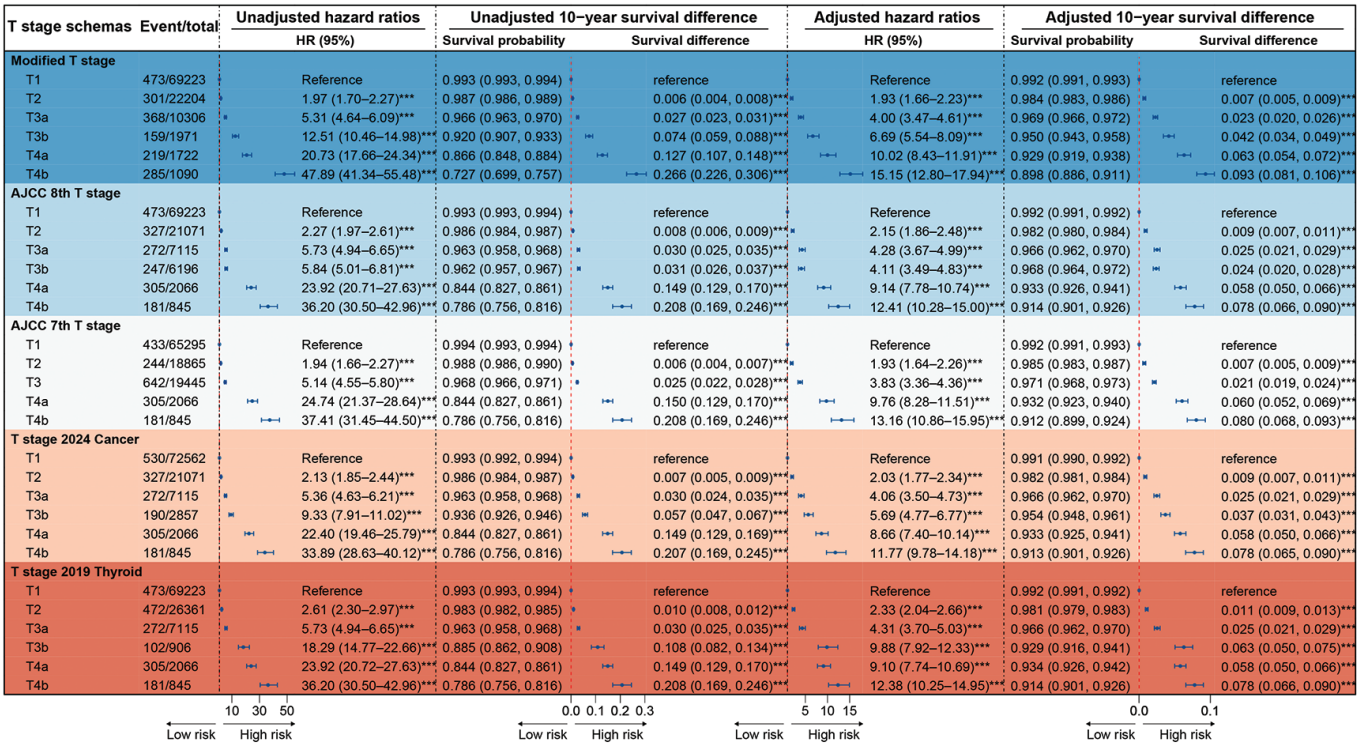

FIG. 3. Hazard ratios and 10-year survival difference of disease-specific survival for different T stages in the development cohort. The adjusted analyses were derived from multivariate Cox models adjusting for age, sex, race/ethnicity, marital status, household income, county type, histologic type, N stage, surgery, and radiotherapy. AJCC, American Joint Committee on Cancer; HR, hazard ratio; ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

3). Notably, the AJCC 8th edition T stage showed poor discrimination between T3a and T3b (HR: 5.73 vs. 5.84). In contrast, both the 2024 cancer T stage (which recategorized Gross3b tumors ≤ 2 cm as T1) and the 2019 thyroid T stage (which recategorized Gross3b tumors ≤ 4 cm as T2) demonstrated improved separation between T3a and T3b by downstaging smaller tumors with Gross3b classification.

Additionally, the modified T stage exhibited superior discriminative and predictive performance, with the highest concordance index (C-index: 0.775, 95% CI: 0.770-0.779) and the lowest Brier score (0.0115, 95% CI: 0.0108-0.0122) in the development cohort. This performance was maintained in both internal bootstrap validation (C-index: 0.764; Brier score: 0.0138) and temporal validation (C-index: 0.821; Brier score: 0.0038) (Table 2). Although these improvements were statistically significant, they were numerically modest compared with those of the AJCC 8th edition. Moreover, the modified T stage consistently outperformed existing systems across five staging performance criteria (Supplementary Table 6).

Generalizability of the modified T stage

In the temporal validation cohort, the modified T system demonstrated superior discriminative and predictive performance as well as overall performance across five cancer staging evaluation criteria (Table 2 and Supplementary Table 6). Although higher absolute C-index values were observed in this cohort (0.821 for the modified T stage), this likely reflects shorter follow-up duration and fewer events rather than inherently better model performance. Importantly, the modified T stage maintained its relative advantage over the AJCC 8th edition (C-index: 0.821 vs. 0.815) in a different treatment era characterized by higher lobectomy rates and reduced use of radioactive iodine. When OS was used as an alternative outcome, the modified T classification continued to show consistent stepwise risk stratification and superior performance (Supplementary Figure 9; Supplementary Tables 7 and 8). Furthermore, subgroup analyses by age and histological subtype confirmed its superiority over the AJCC 8th edition T stage (Supplementary Figures 10 and 11; Supplementary Table 9).

Clinical significance of the modified T stage

The modified T stage demonstrated improved discrimination for lymph node and distant metastases, avoiding the reversal patterns observed in the AJCC 8th edition (Supplementary Figures 12-15). In addition, a modified TNM staging system was constructed (Supplementary Figure 16), which provided clearer survival stratification (Figure 4 and Supplementary Figure 17) and outperformed the AJCC 8th TNM system in both the overall cohort and internal bootstrap validation (Supplementary Table 10). The Sankey diagram indicated that the modified TNM system maintained similar proportions in stage I (74.58% vs. 73.67%) while achieving more refined stratification in advanced stages (Supplementary Figure 18).

TABLE 2. Comparison of C-Index and Brier Score for Disease-Specific Survival Across Different T Stage Schemas.

	Development cohort		Internal validation with 1000 bootstrap replicates		Temporal validation cohort	
T stage schemas	C-index	Brier score	C-index	Brier score	C-index	Brier score
Modified T stage	0.775 (0.770-0.779)	0.0115 (0.0108-0.0122)	0.764 (0.760-0.767)	0.0138 (0.0137-0.0139)	0.821 (0.815-0.826)	0.0038 (0.0032-0.0044)
AJCC 8th T stage	0.766 (0.761-0.770)	0.0118 (0.0106-0.0130)	0.756 (0.754-0.759)	0.0141 (0.0140-0.0142)	0.815 (0.810-0.821)	0.0040 (0.0033-0.0046)
AJCC 7th T stage	0.768 (0.763-0.772)	0.0118 (0.0105-0.0131)	0.756 (0.754-0.758)	0.0142 (0.0140-0.0143)	0.815 (0.810-0.820)	0.0040 (0.0033-0.0046)
T stage 2024 cancers	0.763 (0.759-0.768)	0.0118 (0.0105-0.0131)	0.756 (0.753-0.758)	0.0142 (0.0141-0.0143)	0.813 (0.807-0.819)	0.0040 (0.0033-0.0046)
T stage 2019 thyroid	0.766 (0.761-0.771)	0.0118 (0.0105-0.0130)	0.756 (0.754-0.759)	0.0140 (0.0139-0.0141)	0.815 (0.809-0.820)	0.0040 (0.0033-0.0046)

AJCC, American Joint Committee on Cancer.

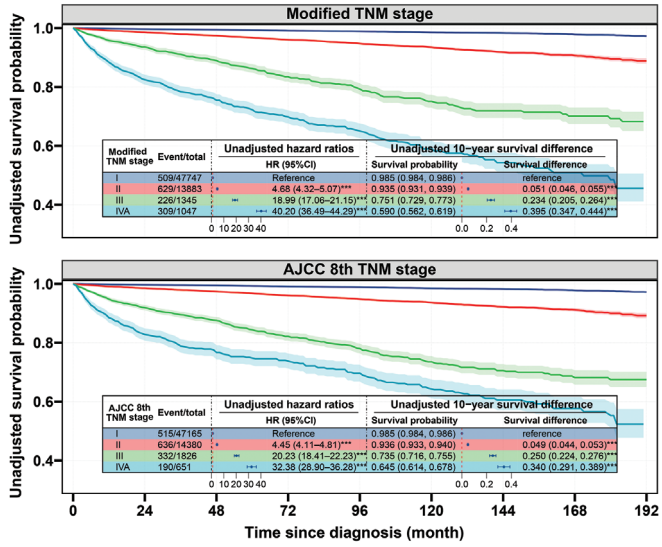

FIG. 4. Unadjusted disease-specific survival curves for modified TNM stage and AJCC 8th TNM stage. Unadjusted Kaplan-Meier survival curves compared the modified TNM staging system (upper panel) with the AJCC 8th edition TNM staging system (lower panel) for disease-specific survival in patients older than 55 years with non-metastatic differentiated thyroid cancer. TNM, tumor-node-metastasis; AJCC, American Joint Committee on Cancer; HR, hazard ratio; CI, confidence interval. ***$p <$ 0.001, **$p < 0.01$, *$p < 0.05$.

DISCUSSION

In this large-scale population-based study, we developed a modified T stage that incorporates significant interactions between ETE and tumor size, demonstrating improved prognostic stratification for DTC. The modified T stage addresses several key limitations of the AJCC 8th edition: (1) it recognizes the prognostic significance of MinETE in a tumor size-dependent manner; (2) it provides more precise prognostic stratification for Gross3b tumors by incorporating tumor size; and (3) it offers the first evidence that the prognostic impact of Gross4a/4b may vary according to tumor size. The clinical utility of the modified T stage is further supported by its superior ability to predict lymph node and distant metastases as well as its improved prognostic stratification when integrated into the TNM framework.

Our findings indicate that the adverse impact of ETE on DSS increases progressively with tumor size, with the RERI rising from 1.90 for MinETE to 6.55 for Gross4b. This size-dependent effect explains why patients with small tumors and ETE (≤ 2 cm and ≤ 4 cm) exhibited relatively favorable outcomes, highlighting that ETE should not be assessed independently of tumor size. These results are consistent with previous studies[11,13,19,21-25] reporting variability in the prognostic effect of ETE across tumor sizes; however, our study provides a more comprehensive quantification of these interactions across all ETE categories in a large population-based cohort.

Using RPA, we developed a modified T classification system that demonstrates a consistent stepwise decline in survival with

increasing T stage. To facilitate clinical interpretation, the modified T categories were aligned with the AJCC 8th edition terminology (T1, T2, T3a, T3b, T4a, and T4b). However, substantial heterogeneity persists within the T4b category, where tumors of different sizes exhibit markedly different survival outcomes, suggesting that further subcategorization may improve prognostic precision.

A major modification in the proposed T stage is the reintroduction of MinETE, which was excluded from the T classification in the AJCC 8th edition, primarily based on single-center studies with limited sample sizes.[5-10] In contrast, large population-based studies[15,16,18] have demonstrated that MinETE is independently associated with worse DSS. Moreover, earlier studies that did not identify a significant association may not have adequately accounted for the interaction between MinETE and tumor size. Recent studies[11,13,19] further support the notion that the prognostic impact of MinETE is dependent on tumor size. This is also supported by the 2015[20] and 2025[36] ATA guidelines, which classify tumors with MinETE as low-to-intermediate risk regardless of size and recommend adjuvant radioactive iodine therapy. In our view, tumors with MinETE represent a transitional phase in disease progression between tumors without ETE and those with gross ETE, warranting consideration for reintegration into the AJCC staging framework.

Another significant modification in the proposed T stage is the reclassification of Gross3b, Gross4a, and Gross4b ETE according to tumor size. Previous studies[21,23] have attempted to optimize the classification of Gross3b tumors. Song et al.[23] downstaged tumors ≤ 4 cm with Gross3b to T2 in a cohort of 3,104 patients with DTC, whereas Park et al.[21] downstaged tumors ≤ 2 cm with Gross3b to T1 in an analysis of 6,282 DTC patients. Although both studies reported improved prognostic precision compared with the AJCC 8th edition, our analysis showed that these proposed T stage modifications did not yield significant performance improvements over the AJCC 8th edition. This may be partly due to the broad grouping of Gross3b tumors without further stratification by tumor size in those studies. In our modified T stage, reclassification of Gross3b tumors ≤ 2 cm as T2 and those measuring 2-4 cm as T3a improved prognostic precision compared with previously proposed staging systems (Table 2). Additionally, to our knowledge, this is the first study to demonstrate that the prognostic impact of Gross4a and Gross4b varies significantly according to tumor size. Small tumors (≤ 2 cm) with Gross4a/4b showed favorable outcomes (10-year DSS > 90%), suggesting that treatment intensity in this subgroup should be further evaluated in prospective studies.

Generalizability—the ability of a classification system to perform consistently across different populations and settings—is a key consideration in oncology guidelines.[31] Our modified T stage demonstrated robust generalizability, with superior performance in internal and temporal validation as well as across different outcomes (DSS and OS) and patient subgroups. Furthermore, the clinical utility of a staging system depends on its ability to inform clinical decision-making.[31] In this regard, the modified T stage showed improved predictive performance for lymph node and distant metastasis compared with the AJCC 8th edition, providing more reliable risk stratification for clinical management. For example, patients with tumors ≤ 2 cm and Gross3b, now classified as T2, may benefit from reconsideration of lobectomy in patients with Gross3b PTC measuring 1-4 cm[24] and raising the hypothesis that management strategies in this subgroup could be further evaluated in prospective trials. Importantly, the modified TNM classification system also demonstrated improved stage separation and prognostic accuracy, which may support more refined risk-adapted trial design and follow-up strategies, pending prospective validation.

Notably, although the improvement in the overall model C-index (0.766-0.775) appears numerically modest, the clinical relevance of the modified T stage is more meaningfully reflected in its improved correlation with lymph node and distant metastasis patterns. Specifically, it eliminates the reversal patterns observed in the AJCC 8th edition and demonstrates a biologically coherent, monotonically increasing association between T category and metastatic risk. Furthermore, when integrated into the TNM framework, the modified TNM stage showed enhanced discrimination and clearer stage separation compared with the AJCC 8th edition TNM system, suggesting that improvements at the T-descriptor level translate into meaningful gains in the overall staging system.

This study has several limitations. The retrospective nature of cancer registry data may introduce coding and reporting biases. Patients with missing clinicopathological data were excluded (11.3% of the initially eligible cohort; Supplementary Figure 1), primarily due to unknown ETE status (5.8%) and unknown tumor size (3.3%). In the SEER registry, missing data largely arise from institutional reporting variability rather than patient-level clinical factors, suggesting a missing-at-random mechanism; however, non-random missingness cannot be entirely excluded. The SEER database lacks recurrence data, limiting assessment of the modified system's performance in predicting disease recurrence, and does not include information on molecular markers (BRAF, RAS mutations), aggressive histopathological variants, vascular invasion, tumor laterality, or detailed treatment variables (RAI dosing, extent of lymph node dissection) that may influence outcomes. Future validation studies using institutional databases that incorporate these variables would allow a more comprehensive prognostic evaluation. Assessment of MinETE is also subject to substantial interobserver variability, and prospective validation using standardized pathological assessment protocols is warranted. A Fine-Gray subdistribution hazard sensitivity analysis confirmed that the relative ordering of all terminal RPA node groups was preserved under a competing-risks framework (Supplementary Figure 6). The cause-specific hazard approach was retained as the primary analysis, as it is recommended for etiological investigations. As an observational study, these findings establish associations rather than causal relationships; therefore, any clinical implications regarding treatment modification should be considered hypothesis-generating and require validation in prospective comparative trials. Heterogeneity within the modified T4b category also warrants further subcategorization to improve prognostic precision. All 13 terminal RPA nodes contained at least 40 disease-specific death events (Supplementary Figures 5 and 6). Firth penalized Cox regression applied to nodes with fewer than 80 events confirmed that the original HR estimates were not materially affected by small-sample bias (Supplementary Table 4). Nevertheless,

wider CIs in smaller subgroups warrant cautious interpretation and external validation. Although tumor size stratification into three categories (≤ 2 cm, 2-4 cm, and > 4 cm) was effective, more granular stratification may be explored in future studies. Of note, the multiple subgroup analyses and performance comparisons in this study were exploratory in nature and should be interpreted with caution, as no adjustment for multiple testing was performed. Despite these limitations, the study has several strengths, including a large population-based cohort with long follow-up, enabling temporal validation and comprehensive subgroup analyses. The detailed ETE data available in SEER allowed the first comprehensive evaluation of interactions between different degrees of ETE and tumor size.

In conclusion, this large-scale population-based study identified a significant positive interaction between ETE and tumor size in DTC. The modified T stage, which incorporates these interactions, demonstrated improved prognostic stratification and enhanced clinical utility compared with existing staging systems. These findings support consideration of incorporating the size-dependent effects of ETE into future revisions of the AJCC staging system. Prospective external validation across diverse populations and clinical settings is warranted to confirm these findings and to evaluate their impact on clinical decision-making.

Ethics Committee Approval: Not applicable.

Informed Consent: Not applicable.

Data Sharing Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authorship Contributions: Concept- H.O.; Design- H.O.; Supervision- X.L., F.X.; Funding- X.L., F.X.; Materials- H.O.; Data Collection or Processing- H.O.; Analysis and/or Interpretation- H.O.; Literature Review- H.O., X.D., B.S., J.C., H.Z.; Writing- H.O., X.L., F.X.; Critical Review- X.D., B.S., J.C., H.Z., X.L., F.X.

Conflict of Interest: The authors declare that they have no conflict of interest.

Funding: This work was supported by the National Natural Science Foundation of China (grant number: 82573294, 82370795 and 82270835).

Supplementary Methods: https://balkanmedicaljournal.org/img/files/Supple%20Materials%20and%20Methods%202026.2025-12-103.pdf

Supplementary Figures: https://balkanmedicaljournal.org/img/files/Supple%20Figure-2026-1-269.pdf

Supplementary Tables: https://balkanmedicaljournal.org/img/files/Supple%20Table-2026-1-269.pdf

REFERENCES

1. Kim M, Kim WG, Oh HS, et al. Comparison of the Seventh and Eighth Editions of the American Joint Committee on Cancer/Union for International Cancer Control Tumor-Node-Metastasis Staging System for Differentiated Thyroid Cancer. *Thyroid*. 2017;27:1149-1155. [CrossRef]
2. Amin MB, Edge SB, Greene FL, et al. AJCC cancer staging manual. 8th ed. New York: *Springer*; 2017. [CrossRef]
3. Perrier ND, Brierley JD, Tuttle RM. Differentiated and anaplastic thyroid carcinoma: Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin*. 2018;68:55-63. [CrossRef]
4. Tran B, Roshan D, Abraham E, et al. An analysis of the american joint committee on cancer 8th edition T staging system for papillary thyroid carcinoma. *J Clin Endocrinol Metab*. 2018;103:2199-2206. [CrossRef]
5. Hay ID, Johnson TR, Thompson GB, Sebo TJ, Reinalda MS. Minimal extrathyroid extension in papillary thyroid carcinoma does not result in increased rates of either cause-specific mortality or postoperative tumor recurrence. *Surgery*. 2016;159:11-19. [CrossRef]
6. Shin JH, Ha TK, Park HK, Ahn MS, Kim KH, Bae KB, Kim TH, Choi CS, Kim TK, Bae SK, Kim SH. Implication of minimal extrathyroidal extension as a prognostic factor in papillary thyroid carcinoma. *Int J Surg*. 2013;11:944-947. [CrossRef]
7. Chéreau N, Buffet C, Trésallet C, et al. Does extracapsular extension impact the prognosis of papillary thyroid microcarcinoma? *Ann Surg Oncol*. 2014;21:1659-1664. [CrossRef]
8. Jin BJ, Kim MK, Ji YB, Song CM, Park JH, Tae K. Characteristics and significance of minimal and maximal extrathyroidal extension in papillary thyroid carcinoma. *Oral Oncol*. 2015;51:759-763. [CrossRef]
9. Woo CG, Sung CO, Choi YM, et al. Clinicopathological significance of minimal extrathyroid extension in solitary papillary thyroid carcinomas. *Ann Surg Oncol*. 2015;22(Suppl 3):728-733. [CrossRef]
10. Arora N, Turbendian HK, Scognamiglio T, et al. Extrathyroidal extension is not all equal: implications of macroscopic versus microscopic extent in papillary thyroid carcinoma. *Surgery*. 2008;144:942-948. [CrossRef]
11. Marongiu A, Nuvoli S, De Vito A, Vargiu S, Spanu A, Madeddu G. Minimal extrathyroid extension (mETE) as the only risk factor in patients with papillary thyroid carcinoma (PC): its clinical impact on recurrence and outcome during long-term follow-up. *Biomedicines*. 2024;12:350. [CrossRef]
12. Seifert R, Schäfers MA, Heitplatz B, Kerschke L, Riemann B, Noto B. Minimal extrathyroid extension in papillary micro carcinoma of the thyroid is an independent risk factor for relapse through lymph node and distant metastases. *J Nucl Med*. 2021;62:1702-1709. [CrossRef]
13. Danilovic DLS, Castroneves LA, Suemoto CK, et al. Is there a difference between minimal and gross extension into the strap muscles for the risk of recurrence in papillary thyroid carcinoma? *Thyroid*. 2020;30:1008-1016. [CrossRef]
14. Xiang J, Wang Z, Sun W, Zhang H. The new T3b category has clinical significance? SEER-based study. *Clin Endocrinol (Oxf)*. 2021;94:449-459. [CrossRef]
15. Bortz MD, Kuchta K, Winchester DJ, Prinz RA, Moo-Young TA. Extrathyroidal extension predicts negative clinical outcomes in papillary thyroid cancer. *Surgery*. 2021;169:2-6. [CrossRef]
16. Liu Z, Huang Y, Chen S, et al. Minimal extrathyroidal extension affects the prognosis of differentiated thyroid cancer: Is there a need for change in the AJCC classification system? *PLoS One*. 2019;14:e0218171. [CrossRef]
17. Bouzehouane N, Roy P, Decaussin-Petrucci M, et al. Prognostic impact of microscopic extra-thyroidal extension (mETE) on disease free survival in patients with papillary thyroid carcinoma (PTC). *Cancers (Basel)*. 2022;14:2591. [CrossRef]
18. Youngwirth LM, Adam MA, Scheri RP, Roman SA, Sosa JA. Extrathyroidal extension is associated with compromised survival in patients with thyroid cancer. *Thyroid*. 2017;27:626-631. [CrossRef]
19. Forleo R, Grani G, Alfò M, et al. Minimal extrathyroidal extension in predicting 1-year outcomes: a longitudinal multicenter study of low-to-intermediate-risk papillary thyroid carcinoma (ITCO#4). *Thyroid*. 2021;31:1814-1821. [CrossRef]
20. Haugen BR, Alexander EK, Bible KC, et al. 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*. 2016;26:1-133. [CrossRef]
21. Park J, An S, Bae JS, Kim K, Kim JS. Impact of tumor size on prognosis in differentiated thyroid cancer with gross extrathyroidal extension to strap muscles: redefining T3b. *Cancers (Basel)*. 2024;16:2577. [CrossRef]
22. Park J, Kang IK, Bae JS, Kim JS, Kim K. Clinical significance of tumor size in gross extrathyroidal extension to strap muscles (T3b) in papillary thyroid carcinoma: comparison with T2. *Cancers (Basel)*. 2022;14:4615. [CrossRef]
23. Song E, Lee YM, Oh HS, et al. A relook at the T stage of differentiated thyroid carcinoma with a focus on gross extrathyroidal extension. *Thyroid*. 2019;29:202-208. [CrossRef]
24. Song E, Kim WW, Jeon MJ, et al. Clinical significance of gross invasion of strap muscles in patients with 1- to 4-cm-sized papillary thyroid carcinoma undergoing lobectomy. *Ann Surg Oncol*. 2019;26:4466-4471. [CrossRef]
25. Tam S, Boonsripitayanon M, Amit M, et al. Survival in differentiated thyroid cancer: comparing the AJCC cancer staging seventh and eighth editions. *Thyroid*. 2018;28:1301-1310. [CrossRef]

26. Ning C, Ouyang H, Shen D, et al. Prediction of survival in patients with infected pancreatic necrosis: a prospective cohort study. *Int J Surg*. 2024;110:777-787. [CrossRef]
27. Kabelitz MA, Hartl L, Schaub G, et al. Identification of optimal portal pressure decrease to control ascites while minimizing HE after TIPS: A multicenter study. *Hepatology*. 2025;82(5):1172-1186. [CrossRef]
28. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. *Int J Epidemiol*. 2012;41(2):514-520. [CrossRef]
29. Liu Y, Chen Z, Sun B, Ouyang H. Age-dependent effects of surgical approach in T3b differentiated thyroid carcinoma: a population-based analysis using machine learning. *Endocr Relat Cancer*. 2025;32:e250417. [CrossRef]
30. Huang YQ, Huang ZN, Hong QQ, et al. Development of a staging system for hepatoid adenocarcinoma of the stomach based on multicenter data: a retrospective cohort study. *Int J Surg Lond Engl*. 2025;111:718-727. [CrossRef]
31. Rami-Porta R, Nishimura KK, Giroux DJ, et al. The international association for the study of lung cancer lung cancer staging project: proposals for revision of the TNM stage groups in the forthcoming (ninth) edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2024;19:1007-1027. [CrossRef]
32. Plichta JK, Thomas SM, Hayes DF, et al. Novel prognostic staging system for patients with de novo metastatic breast cancer. *J Clin Oncol*. 2023;41:2546-2560. [CrossRef]
33. Li WZ, Wu HJ, Lv SH, et al. Assessment of survival model performance following inclusion of Epstein-Barr virus DNA status in conventional TNM staging groups in Epstein-Barr virus-related nasopharyngeal carcinoma. *JAMA Netw Open*. 2021;4:e2124721. [CrossRef]
34. O'Sullivan B, Huang SH, Su J, et al. Development and validation of a staging system for HPV-related oropharyngeal cancer by the International Collaboration on Oropharyngeal cancer Network for Staging (ICON-S): a multicentre cohort study. *Lancet Oncol*. 2016;17:440-451. [CrossRef]
35. Xu W, Shen XW, Su J, O'Sullivan B, Huang SH. Refining evaluation methodology on TNM stage system: assessment on HPV-related oropharyngeal cancer. *Austin Biom Biostat*. 2015;2:1014. [CrossRef]
36. Ringel MD, Sosa JA, Baloch Z, et al. 2025 American Thyroid Association management guidelines for adult patients with differentiated thyroid cancer. *Thyroid*. 2025;35:841-985. [CrossRef]