Modified dynamic risk stratification for predicting recurrence using the response to initial therapy in patients with differentiated thyroid carcinoma

Min Ji Jeon1, Won Gu Kim1, Woo Ri Park1, Ji Min Han1, Tae Yong Kim1, Dong Eun Song3, Ki-Wook Chung2, Jin-Sook Ryu4, Suck Joon Hong2, Young Kee Shong1 and Won Bae Kim1

Departments of 1Internal Medicine, 2Surgery, 3Pathology and 4Nuclear Medicine, Asan Medical Center, University of Ulsan College of Medicine, 88, Olympic-ro 43-gil, Songpa-gu, Seoul 138-736, Korea

Abstract

Objective: A new risk stratification system was proposed to estimate the risk of recurrence in patients with differentiated thyroid carcinoma (DTC) using the response to initial therapy. Here, we describe the modified dynamic risk stratification system, which takes into consideration the status of serum anti-Tg antibody (TgAb), and validate this system for assessing the risk of recurrence in patients with DTC.

Patients and methods: Patients who underwent total thyroidectomy with radioiodine remnant ablation due to DTC between 2000 and 2005 were included. We classified patients into four groups based on the response to the initial therapy (‘excellent’, ‘acceptable’, ‘biochemical incomplete’, and ‘structural incomplete’ response).

Results: The median follow-up period of 715 patients with DTC was 8 years. The response to initial therapy was an important risk predictor for recurrent/persistent DTC. The relative risks (95% CI) of recurrence were 16.5 (6.3–43.0) in the ‘acceptable response’ group, 41.3 (15.4–110.8) in the ‘biochemical incomplete response’ group, and 281.2 (112.9–700.5) in the ‘structural incomplete response’ group compared with the ‘excellent response’ group (P < 0.001, P < 0.001, and P < 0.001 respectively). The disease-free survival rate of the ‘excellent response’ group to initial therapy was 98.3% whereas that of the ‘structural incomplete response’ group was only 6.8%.

Conclusions: Our study validates the usefulness of the modified dynamic risk stratification system including the status of serum TgAb for predicting recurrent/persistent disease in patients with DTC. Personalized risk assessment using the response to initial therapy could be useful for the follow-up and management of patients with DTC.

Introduction

The practical and effective risk stratification system is essential for the individualized assessment and management of patients with differentiated thyroid carcinoma (DTC). Over the last decades, a number of studies have identified various clinicopathological risk factors and suggested staging systems to distinguish low-risk patients from high-risk patients (1, 2, 3, 4, 5, 6). The survival rate for patients with DTC is usually excellent and only a small percentage of patients die due to DTC (7, 8). A recent study estimated the 10-year probability of death resulting from thyroid cancer to be 3% (9). Therefore, the risk stratification system predicting the recurrence of DTC could be more practical for clinicians to identify high-risk patients and to provide more personalized management of the disease.
Metastasis (TNM) staging system is commonly used in clinical practice and to predict the risk of disease-related death (6). However, the TNM staging system is not appropriate for assessing the risk of DTC recurrence (1, 10). To improve prognostication and follow-up management for patients with DTC, the American Thyroid Association (ATA) recommended the use of a new postoperative clinicopathological staging system (1). The ATA risk stratification system is a comprehensive approach that includes postoperative pathology analysis, radioiodine therapy, and serum thyroglobulinemia. However, the ATA risk stratification system neither defines the specific thyroglobulin (Tg) level nor describes serial ultrasonography (US) results for assessment of the risk of recurrence. Moreover, the risk of recurrence and death may change over time depending on the response to therapy and the clinical course of the disease.

Recently, the dynamic risk stratification system, which reassesses the risk of recurrence using the level of serum Tg and results of follow-up diagnostic images as markers of the response to initial surgery and radioactive iodine (RAI) remnant ablation, was described (11, 12). This system classified patients into four categories by disease status during the first 6–24 months after initial treatment. This practical approach effectively estimates the risk of recurrent disease in DTC patients. This system has been validated in three large cohort studies in the United States and Brazil (11, 12, 13). However, this dynamic risk assessment did not take into consideration the status of serum anti-Tg antibody (TgAb) even though 10–25% of patients with DTC have a positive serum TgAb level (14). Delayed risk stratification, proposed by Castagna et al. (15), also did not include patients with a positive serum TgAb level.

In this study, we describe the modified dynamic risk stratification system, which includes the level of serum TgAb, and analyze this system for assessing the risk of recurrence/persistence in patients with DTC at a single tertiary referral hospital in Korea. We also evaluated the relative importance of this system for predicting prognosis compared with TNM staging and ATA risk classification.

**Subjects and methods**

**Patients**

From 2000 to 2005, 1296 patients with DTC who underwent first surgery at the Asan Medical Center (Seoul, Korea) were retrieved from a historical cohort. Patients were treated by total or near total thyroidectomy with subsequent RAI ablation therapy according to a protocol established by the Endocrinology Division of the Asan Medical Center (Seoul, Korea) (16). Patients with anaplastic, poorly differentiated, or medullary thyroid carcinoma were excluded. Patients with a tumor size of <1 cm ($n = 366$) or without adequate follow-up data for determining dynamic risk stratification ($n = 215$) were also excluded. A total of 715 patients were eligible for this study. This study was approved by the Institutional Review Board of the Asan Medical Center.

**Follow-up protocol to detect recurrence**

After remnant ablation, all patients were prescribed levothyroxine for TSH suppression and were regularly followed-up with physical examination. The level of serum Tg and serum TgAb was measured every 6–12 months. The level of serum-stimulated Tg after thyroid hormone withdrawal was measured within 2 years after remnant ablation. A TgAb value exceeding 60 IU/ml was considered to be positive for interfering with the Tg measurement. Neck US and diagnostic radioiodine whole-body scan (WBS) were done during the first 6–24 months after initial therapy and repeated at 12- to 24-month intervals. Additional diagnostic imaging studies, such as computed tomography, magnetic resonance scan, or whole-body fluorodeoxyglucose (FDG)–positron emission tomography (PET) scan, were also performed in some patients as needed.

**Definitions of clinical outcomes**

Patients with an undetectable level of serum-stimulated Tg (<1 ng/ml), negative TgAb, and no suspicious structural disease at the end of the follow-up period were considered to have no clinical evidence of disease (NED). Patients were considered to have persistent disease if the suppressed or stimulated Tg level was ≥1 ng/ml or if suspicious structural disease was seen on images. The recurrence was defined as a newly developed biochemical (suppressed or stimulated Tg level ≥1 ng/ml) or structural disease on images after the NED period. Patients with persistent disease were classified as having either biochemical persistent disease or structural persistent disease. Biochemical persistent disease was defined as a persistently elevated suppressed or stimulated Tg level ≥1 ng/ml without structural evidence of disease. Structural persistent disease was defined as pathologically or cytologically proven malignant tissue and/or highly suspicious metastatic lesions by imaging studies such as neck US, WBS, or FDG–PET scan.
Risk stratification by each classification

We categorized eligible patients according to the sixth edition of the AJCC TNM staging system (stages I, II, III, and IV) and the ATA risk classification (low, intermediate, and high risk of recurrence) (1). We also classified patients by the response to initial therapy using the levels of serum-stimulated Tg and TgAb and the results of neck US and diagnostic WBS 6–24 months after initial therapy.

An ‘excellent response’ was defined as a negative serum-stimulated Tg level (<1 ng/ml) with the absence of serum TgAb (<60 IU/ml) and no suspicious metastatic lesions by imaging. An ‘acceptable response’ was defined as negative findings by imaging with a low serum-stimulated Tg level (1–10 ng/ml) or the presence of serum TgAb (≥60 IU/ml) because patients with positive serum TgAb showed a similar disease-free survival (DFS) rate as patients with a low serum-stimulated Tg level (1–10 ng/ml) (Fig. 1). A ‘biochemical incomplete response’ was defined as a high serum-stimulated Tg level (≥10 ng/ml) without any evidence of structural metastatic lesions by imaging. A ‘structural incomplete response’ was defined as the presence of a known metastatic lesion or a newly developed suspicious metastatic lesion by imaging. We adopted the definition of the dynamic risk stratification system by Tuttle et al. (11, 12). However, we also took into consideration the levels of serum TgAb during our analysis. We did not include the level of suppressed Tg or changes in the Tg level as a risk classifier. This modified definition of dynamic risk stratification system is shown in Table 1.

Statistical analysis

We used R version 2.13 and R libraries survival, car and Cairo to analyze data and draw graphs (R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org). Continuous variables were presented as medians with interquartile range (IQR) and categorical variables were presented as numbers with percentages.

The primary endpoint of this study was the structural recurrence or persistence after initial therapy, pathologically proven malignant tissue and/or metastatic lesions in other distant organs by imaging studies.

A Cox proportional hazard model was used to evaluate the risk of structural recurrence/persistence of disease. DFS curves were constructed using the Kaplan–Meier method, and the log rank test was used to evaluate differences in DFS between groups. The relative importance of each risk stratification was determined by calculating the proportion of variation in survival time explained (PVE) by

Table 1 Definition of the modified dynamic risk stratification by the response to the initial therapy.

<table>
<thead>
<tr>
<th>Excellent response</th>
<th>Acceptable response</th>
<th>Biochemical incomplete response</th>
<th>Structural incomplete response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stimulated Tg &lt;1 ng/ml</td>
<td>Stimulated Tg 1–10 ng/ml or positive TgAb</td>
<td>Stimulated Tg ≥10 ng/ml</td>
<td>Suspicious recurrent or persistent disease on images</td>
</tr>
<tr>
<td>Negative TgAb</td>
<td>No evidence of recurrent or persistent disease on images</td>
<td>No evidence of recurrent or persistent disease on images</td>
<td></td>
</tr>
</tbody>
</table>

*The cutoff value of TgAb was 60 IU/ml. A TgAb value ≥60 IU/ml was considered to be positive.

*Images: diagnostic whole-body scan or neck ultrasonography or any cross-sectional images during first 2 years after remnant ablation.
the maximum likelihood ratio ($G^2$) from the Cox model ($PVE = 1 - \exp (-G^2/2n)$) (4, 17). The PVE (%) ranges from 0 to 100 and the higher PVE value suggests the more accurate predictability of risk of recurrence. DFS curves were also constructed using the Kaplan–Meier method. All $P$ values were two sided with $P<0.05$ considered statistically significant.

### Results

The median follow-up period of 715 patients was 8 years (IQR 6.3–9.3). The median age of the patients was 46.2 years (IQR 37.1–54.7) and 622 patients (87%) were female. The median dose of administered radioiodine was 5.6 GBq. The majority of patients had papillary thyroid carcinoma (PTC, 95%) including variant types of papillary carcinoma, while a small percentage of patients had follicular thyroid carcinoma (2%), Hürthle cell carcinoma (1%), or a mixed type of thyroid carcinoma (2%). At the time of last follow-up, 552 patients (88%) were NED, 156 patients (22%) had persistent disease, and seven patients (1%) showed recurrent disease (Table 2). The primary endpoint, structural recurrent/persistent disease, was found in 132 patients (18%) during the follow-up period. Most of the structural recurrent/persistent disease was confined to the cervical lymph nodes in these patients ($n = 100; 76\%$). Seven patients (1%) with structural persistent disease died due to the progression of DTC. The median period from initial therapy to the detection of structural recurrent/persistent disease was 2.8 years (IQR 1.4–4.8).

As shown in Table 3, the TNM staging system classified patients into stage I ($n = 348; 49\%$), stage II ($n = 32; 4\%$), stage III ($n = 271; 38\%$), and stage IV ($n = 64; 9\%$). DFS rates were 83.0% in stage I, 58.8% in stage II, 85.6% in stage III, and 22.3% in stage IV at the end of the follow-up period (Fig. 2A, $\chi^2 = 10.06, P = 0.002$). Structural recurrent/persistent disease was more prevalent in stage II patients compared with stage III patients because 17 patients (53%) in stage II had distant metastasis. The PVE value of the TNM staging system was 8.68%, and it was the lowest value among the three risk stratification systems (Table 3).

When we classified patients according to the ATA risk classification, 88 patients (12%) belonged to the low-risk group, 578 patients (81%) belonged to the intermediate-risk group, and 49 patients (7%) belonged to the high-risk group. The risk of recurrence/persistence was significantly higher in intermediate-risk (HR = 15.34, $P = 0.007$) and high-risk groups (HR = 106.16, $P < 0.001$) than in the low-risk group. The ATA risk classification clearly predicted the risk of DTC recurrence/persistence in our study subjects, especially in the low-risk group. DFS rates were 98.7% in stage I ($n = 348; 49\%$), 78.7% in the intermediate-risk group, and 49 patients (7%) belonged to the high-risk group. The risk of recurrence/persistence was significantly higher in intermediate-risk (HR = 15.34, $P = 0.007$) and high-risk groups (HR = 106.16, $P < 0.001$) than in the low-risk group. The ATA risk classification clearly predicted the risk of DTC recurrence/persistence in our study subjects, especially in the low-risk group. DFS rates were 98.7% in stage I ($n = 348; 49\%$), 78.7% in the intermediate-risk group, and 33.8% in the high-risk group at the end of the follow-up period according to the modified dynamic risk stratification. The number of patients in the four groups was as follows: an excellent response was obtained in 143 patients (20%), a biochemical incomplete response was obtained in 49 patients (7%), and a structural incomplete response was obtained in 88 patients (12%). DFS rates were 98.3% in patients with an excellent response, 71.6% in patients with an acceptable response, and 54.2% in patients with a biochemical incomplete response, which were also constructed using the Kaplan–Meier method. All $P$ values were two sided with $P<0.05$ considered statistically significant.

### Table 2 Description of study cohort.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Median (IQR)</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>46.2 (37.1–54.7)</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>622 (87)</td>
<td>93 (13)</td>
</tr>
<tr>
<td>Male</td>
<td>5.6 (3.7–7.4)</td>
<td></td>
</tr>
<tr>
<td>Follow-up duration (years)</td>
<td>8 (6.3–9.3)</td>
<td></td>
</tr>
<tr>
<td>Histology</td>
<td>643 (90)</td>
<td></td>
</tr>
<tr>
<td>Papillary</td>
<td>38 (5)</td>
<td></td>
</tr>
<tr>
<td>Follicular</td>
<td>18 (2)</td>
<td></td>
</tr>
<tr>
<td>Hürthle cell</td>
<td>4 (1)</td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>12 (2)</td>
<td></td>
</tr>
<tr>
<td>TgAb status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive TgAb at the time of ablation</td>
<td>110 (15)</td>
<td></td>
</tr>
<tr>
<td>Positive TgAb at 6–24 months after initial treatment</td>
<td>35 (5)</td>
<td></td>
</tr>
<tr>
<td>Extent of surgery</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total thyroidectomy alone</td>
<td>99 (14)</td>
<td></td>
</tr>
<tr>
<td>Total thyroidectomy with central neck dissection</td>
<td>499 (70)</td>
<td></td>
</tr>
<tr>
<td>Total thyroidectomy with lateral neck dissection</td>
<td>117 (16)</td>
<td></td>
</tr>
<tr>
<td>Clinical status at the end of period</td>
<td>552 (88)</td>
<td></td>
</tr>
<tr>
<td>No evidence of disease</td>
<td>29 (4)</td>
<td></td>
</tr>
<tr>
<td>Biochemical persistent disease</td>
<td>127 (18)</td>
<td></td>
</tr>
<tr>
<td>Structural persistent disease</td>
<td>7 (1)</td>
<td></td>
</tr>
<tr>
<td>Recurrence</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

IQR, interquartile range.
response, and 6.8% in patients with a structural incomplete response (Fig. 2C, \( \chi^2 = 609.9, P < 0.001 \)). The risk of structural recurrent/persistent disease was significantly higher in acceptable responders (HR = 16.51, \( P < 0.001 \)), biochemical incomplete responders (HR = 41.33, \( P < 0.001 \)), and structural incomplete responders (HR = 281.18, \( P < 0.001 \)) than in excellent responders. This risk stratification system had the highest PVE value (44.59%) among the three risk stratification systems (Table 3).

Furthermore, this PVE value was higher than the PVE value of dynamic risk stratification for patients that were negative for TgAb (Table 3).

We also applied this modified dynamic risk stratification system in each subgroup of patients according to the ATA risk classification. This dynamic risk stratification predicted recurrence/persistence in patients with intermediate-risk (\( P < 0.001 \), Fig. 2E) and high-risk groups (\( P < 0.001 \), Fig. 2F). As shown in Table 3, patients belonging to the intermediate-risk group by the ATA classification were reclassified as an excellent response (\( n = 352, 61\% \)), an acceptable response (\( n = 127, 22\% \)), a biochemical incomplete response (\( n = 45, 8\% \)), and a structural incomplete response (\( n = 54, 9\% \)). The DFS rate was 97.9% in excellent responders, 71.9% in acceptable responders, 54.9% in biochemical incomplete responders, and 1.9% in structural incomplete responders (Fig. 2E).

### Discussion

This study demonstrated that the modified dynamic risk stratification system, which takes into consideration the level of serum TgAb, effectively estimated the risk of recurrence/persistence in patients with DTC. Most excellent responders did not experience structural recurrent/persistent disease, whereas it was confirmed in 93.2% of patients with structural incomplete response at the end of the follow-up period. The ATA risk classification also effectively predicted the low risk of recurrent/persistent disease. There was only one recurrence in 88 patients belonging to the low-risk classification. However, our modified dynamic risk stratification reclassified the risk of recurrent/persistent disease in intermediate- and high-risk groups according to the ATA risk classification. The TNM staging system was not appropriate for estimating the risk of recurrence/persistence in our study and this finding was consistent with previous reports (1, 18). We also calculated PVE values, and the modified dynamic risk stratification system had the highest PVE value (44.59%) among the three risk stratification systems. Additionally, this modified dynamic risk stratification system was applied in each subgroup of patients according to the ATA risk classification, and it predicted recurrence/persistence in patients with intermediate-risk and high-risk groups (\( P < 0.001 \)).
European Journal of Endocrinology

We also suggested that the level of serum results owing to TgAb interference with immunometric limited in the presence of TgAb because of false-negative the prognostic value of serum Tg measurement was be a prognostic indicator during the early postoperative periodic. Therefore, we included the level of serum TgAb in

The level of serum Tg, and the results of cross-sectional images system included the clinical status of the disease, the TgAb and a change in the serum TgAb concentration could correspond to the time of initial treatment. (A) AJCC/TNM staging (P=0.015). (B) ATA risk classification system (P<0.001). (C) Modified dynamic risk stratification (P<0.001).

Our modified dynamic risk stratification and validated this system in our cohort (22).

The prevalence of positive serum TgAb in patients with DTC is 10–25% and is about twofold higher than in the general population (14, 21). The timing of TgAb measurement is likely to affect the prevalence of positive serum TgAb level because TgAb decreases and eventually disappears as it removes the antigenic stimulus from Tg after initial therapy (22). A previous study reported that TgAb continuously decreased in most patients after surgery and the prevalence of positive serum TgAb decreased to <10% after 3 years (23). In this study, the prevalence of positive serum TgAb was 15% at the time of the remnant ablation and decreased to 5% at the time of restratification, 6–24 months after initial treatment. Thus, the rate of TgAb positivity in this study may be underestimated because of a delay in TgAb measurement.

The proportion of patients in low- and high-risk groups according to the ATA risk classification was...
Table 4: The clinical outcome according to the modified dynamic risk stratification.

<table>
<thead>
<tr>
<th></th>
<th>Excellent response</th>
<th>Acceptable response</th>
<th>Biochemical incomplete response</th>
<th>Structural incomplete response</th>
</tr>
</thead>
<tbody>
<tr>
<td>NED</td>
<td>428 (98)</td>
<td>106 (74)</td>
<td>13 (27)</td>
<td>5 (6)</td>
</tr>
<tr>
<td>Biochemical, persistent disease</td>
<td>0</td>
<td>11 (8)</td>
<td>17 (35)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Structural, persistent disease</td>
<td>0</td>
<td>26 (18)</td>
<td>19 (39)</td>
<td>82 (93)</td>
</tr>
<tr>
<td>Recurrence</td>
<td>7 (2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

NED, no evidence of disease.

12 and 7% respectively in this study. These values are much lower than previous reports using the dynamic risk stratification system (11, 12, 13). We excluded patients with microcarcinomas in this study because of the controversy in using remnant RAI ablation for patients with microcarcinoma (24, 25, 26). Moreover, we only included patients who underwent surgery between 2000 and 2005 because routine neck US was performed during follow-up after 2000. This could be the reason for the low number of patients in low- and high-risk groups. More intensive examination of pathological specimens could also be another possible reason for the low number of patients in the low-risk group. Identification of small lymph node metastases, single microscopic extrathyroidal extensions, or vascular invasions can change the risk of recurrence from low to intermediate unnecessarily (18). In this regard, the dynamic risk stratification system is a simple and practical system that could overcome weak points of initial pathological staging by reflecting clinical disease status with highly sensitive neck US and the level of serum Tg and TgAb. This approach may help to tailor management of the disease.

This study has several limitations. First, we could not validate the dynamic risk stratification system for patients who did not receive RAI ablation therapy or for those who did lobectomy because we only included patients who underwent total or near total thyroidectomy followed by RAI ablation. This personalized risk assessment using the response to initial therapy could be useful for the follow-up and management of patients with DTC.

In conclusion, we propose the modified dynamic risk stratification by including an additional variable of serum TgAb level to evaluate the risk of recurrent/persistent DTC. We have shown that this modified dynamic risk stratification effectively predicts the recurrence/persistence of DTC in patients who received total thyroidectomy followed by RAI ablation. This personalized risk assessment using the response to initial therapy could be useful for the follow-up and management of patients with DTC.

Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding
This study was supported by the National Research Foundation (NRF) of Korea (Research grant number NRF-2012R1A1A2038383).

References
12 Vaisman F, Tala H, Grewal R & Tuttle RM. In differentiated thyroid cancer, an incomplete structural response to therapy is associated with significantly worse clinical outcomes than only an incomplete thyroglobulin response. Thyroid 2011 21 1317–1322. (doi:10.1089/thy.2011.0232)
15 Castagna MG, Maino F, Cipri C, Belardini V, Theodoropoulou A, Cevenini G & Pacini F. Delayed risk stratification, to include the response to initial treatment (surgery and radiiodine ablation), has better outcome predictivity in differentiated thyroid cancer patients. European Journal of Endocrinology 2011 165 441–446. (doi:10.1530/EJE-11-0466)
18 Randolph GW, Duh QY, Heller KS, LiVolsi VA, Mandel SJ, Steward DL, Tufano RP & Tuttle RM. 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 extranodal extension. Thyroid 2012 22 1144–1152. (doi:10.1089/thy.2012.0043)
21 Spencer CA. Challenges of serum thyroglobulin (Tg) measurement in the presence of Tg autoantibodies. Journal of Clinical Endocrinology and Metabolism 2004 89 3702–3704. (doi:10.1210/jc.2004-0986)

Received 26 June 2013
Revised version received 4 September 2013
Accepted 2 October 2013