Dynamic risk stratification in the follow-up of thyroid cancer: what is still to be discovered in 2017?

Jolanta Krajewska, Ewa Chmielik and Barbara Jarząb

Abstract

The adequate risk stratification in thyroid carcinoma is crucial to avoid on one hand the overtreatment of low-risk and on the other hand the undertreatment of high-risk patients. The question how to properly assess the risk of relapse has been discussed during recent years and resulted in a substantial change in our approach to risk stratification in differentiated thyroid cancer, proposed by the newest ATA guidelines. First initial risk stratification, based on histopathological data is carried out just after primary surgery. It should be emphasized, that a high quality of histopathological report is crucial for proper risk stratification. Next, during the follow-up, patients are restratified considering their response to treatment applied and classified to one of the following categories: excellent response, biochemical incomplete response, structural incomplete or indeterminate response. This new approach is called dynamic risk stratification as, in contrary to the previous rigid evaluation performed at diagnosis, reflects a real-time prognosis and thereby substantially influences and personalizes disease management. In this review, we raise some unresolved questions, among them the lack of prospective studies, fulfilling evidence-based criteria, necessary to validate this model of risk stratification. We also provided some data concerning the use of dynamic risk stratification in medullary thyroid cancer, not yet reflected in ATA guidelines. In conclusion, dynamic risk stratification allows for better prediction of the risk of recurrence in thyroid carcinoma, what has been demonstrated in numerous retrospective analyses. However, the validation of this approach in prospective studies seems to be our task for near future.

Introduction

The adequate risk stratification in a malignant neoplastic disease is crucial to avoid on one hand the overtreatment of low-risk patients and on the other hand the undertreatment of high-risk subjects. This issue is particularly important in thyroid carcinoma, generally characterized by good outcomes with 10-year overall survival (OS) rates for patients with papillary (PTC), follicular (FTC) and medullary thyroid cancer (MTC) of 93, 85 and 75%, respectively (Hundahl et al. 1998).

The question how to properly assess the risk of relapse in differentiated thyroid carcinoma (DTC) has...
been discussed during recent years and finally resulted in a substantial change in our approach to DTC risk stratification, fulfilling evidence-based medicine (EBM) criteria, proposed by the newest ATA guidelines (2015 ATA GLs) (Haugen et al. 2016). A static, single risk evaluation, done until recently only at DTC diagnosis, has been replaced by so-called dynamic risk stratification. First initial risk evaluation, also updated in the current ATA GLs (Table 1), is carried out postoperatively. On the basis of distinct histopathological features and the presence of nodal or distant metastases patients are classified as low, intermediate or high risk. Next, during the further follow-up, they are restratified considering response to treatment applied to one of the following category: excellent, incomplete biochemical, incomplete structural or indeterminate response (Table 2). Such restratification, which included the results of imaging studies (among them neck ultrasound and RAI scintigraphy), serum thyroglobulin (Tg) level (stimulated or suppressed) and anti-Tg antibody level (Haugen et al. 2016) allows for a better prediction of the risk of recurrent disease and to individualize patient management and follow-up. A group of patients, who achieved an excellent response, is characterized by a very low risk of DTC relapse (1–4%) and cancer-specific death (<1%). Among at least 30% of patients among those demonstrating an incomplete biochemical response do not present persistent disease at the end of follow-up without any further treatment except for thyroxine suppressive therapy, 20% show no evidence of disease (NED) after additional therapy, whereas 20% develop structural disease. However, the risk of cancer-related death is still substantially low in this group (<1%). Considering the patients with an incomplete structural response, 50–85% of them have persistent DTC despite treatment. DTC-specific death rates in patients with locoregional disease and distant metastases are 11% and 50%, respectively. Patients, in whom indeterminate response is obtained, are also characterized by a low risk of cancer-related death (<1%). Most of them remain stable or nonspecific changes resolve during further follow-up. However, 15–20% of them will have structural disease (Haugen et al. 2016).

For the first time, this new approach was proposed in Tuttle et al. (2010) (Table 3). The authors retrospectively analyzed medical records of 588 DTC patients with a minimum of 3 years of follow-up. First, these patients were stratified according to the 7th edition of the AJCC/ UICC staging system and the ATA 2009 risk groups (as low, intermediate or high risk) (Cooper et al. 2009). Next, 471 subjects from the study group were restratified on the basis of clinical data obtained during the first 2 years of follow-up (suppressed or stimulated Tg level, imaging studies) and the response to primary treatment with total thyroidectomy and radioiodine (RAI) ablation (excellent, complete biochemical response, complete structural response, or indeterminate response). A dynamic risk stratification (first follow-up, follow-up 3–5 years, follow-up 6–10 years) has been performed postoperatively. The authors have concluded that the ATA 2009 criteria are not sufficiently predictive of the risk of recurrence after initial treatment. More detailed information on the risk of DTC recurrence is obtained using a dynamic follow-up approach. The patients who achieved a complete biochemical response were excluded from this analysis.

Table 1 The comparison of criteria used for ATA 2009 and 2015 initial risk classification.

<table>
<thead>
<tr>
<th>Category</th>
<th>ATA 2009 Cooper et al. (2009)</th>
<th>ATA 2015 Haugen et al. (2016)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATA low risk</td>
<td>1. No local and distant metastases</td>
<td>PTC fulfilling all ATA 2009 criteria and additionally Clinical N0 or ≤5 pathologic N1 micrometastases (&lt;0.2 cm in the largest dimension) Intrathyroidal, encapsulated follicular variant of PTC (NIFTP)*</td>
</tr>
<tr>
<td></td>
<td>2. All macroscopic tumor resected</td>
<td>Intrathyroidal, well differentiated FTC with capsular invasion and/or minimal (&lt;4 foci) vascular invasion Intrathyroidal papillary microcarcinoma, unifocal or multifocal, including BRAF V600E mutated (if known)</td>
</tr>
<tr>
<td></td>
<td>3. No tumor invasion of locoregional tissues or structures</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. No aggressive histology or vascular invasion</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. If RAI given no RAI uptake outside thyroid bed on the first posttreatment WBS</td>
<td></td>
</tr>
<tr>
<td>ATA intermediate risk</td>
<td>1. Microscopic invasion of the tumor into the perithyroidal soft tissues at initial surgery</td>
<td>ATA 2009 criteria 1 and 3 and additionally: hobnail PTC variant PTC with vascular invasion Multifocal papillary microcarcinoma with ETE and BRAF V600E mutated (if known)</td>
</tr>
<tr>
<td></td>
<td>2. Cervical lymph node metastases or RAI uptake outside thyroid bed on WBS done after remnant ablation</td>
<td>Clinical N1 or &gt;5 pathologic N1 with all involved lymph nodes &lt;3 cm, in largest dimension</td>
</tr>
<tr>
<td></td>
<td>3. Tumor with aggressive histology (tall cell, insular, columnar cell variant)</td>
<td></td>
</tr>
<tr>
<td>ATA high risk</td>
<td>1. Macroscopic tumor invasion into the perithyroidal soft tissues (gross ETE)</td>
<td>ATA 2009 criteria 1–3 and additionally Postoperative serum Tg suggestive of distant metastases Pathologic N1 with any metastatic lymph node &gt;3 cm in largest dimension FTC with extensive vascular invasion (&gt;4 foci of vascular invasion)</td>
</tr>
<tr>
<td></td>
<td>2. Incomplete tumor resection</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Distant metastases</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Tg out of proportion to what is seen on posttreatment WBS</td>
<td></td>
</tr>
</tbody>
</table>

*The 2015 ATA GL did not discuss noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP). ETE, extrathyroidal extension; FTC, follicular thyroid cancer; PTC, papillary thyroid cancer; RAI, radioiodine; WBS, whole body scan.
DTC risk stratification

Table 3 guidelines. Next, 8–12 months after Negative imaging has been confirmed by a few retrospective Castagna Endocrine-Related Cancer Indeterminate response Structural incomplete biochemical incomplete response Excellent response

Dynamic risk stratification in DTC

Dynamic risk stratification in DTC patients treated with total thyroidectomy and RAI ablation

The idea of dynamic risk stratification, proposed in Tuttle et al. (2010) has been confirmed by a few retrospective analyses, published recently (Table 3). The first data come from an Italian study (Castagna et al. 2011). In fact the authors called this new approach ‘delayed risk stratification’ (DRS) but it unequivocally reflected Tuttle’s idea. Castagna et al. (2011) retrospectively reviewed a group of 512 DTC patients, who were stratified soon after initial therapy (total thyroidectomy and RAI ablation) on the basis of ATA 2009 (Cooper et al. 2009) and ETA 2006 (Pacini et al. 2006) guidelines. Next, 8–12 months after primary treatment, they were subjected to DRS involving their clinical status. Initially, at the time of treatment, 45.1% of patients were classified as ETA low-risk and 54.9% of remaining subjects as ETA high-risk. Simultaneously, 54.9% of patients were classified as ETA low-risk and 45.1% of remaining subjects as ETA high-risk. The authors concluded that a risk-adapted approach to follow-up cannot be based solely on static, initial stratification of the risk that remains unchanged over the life as a newly proposed recurrence staging system effectively predicted the risk of recurrence and persistent DTC (Tuttle et al. 2010).

This review summarizes 7-year experience with new principles of risk stratification. We mainly focused on DTC; however, we decided also to present some preliminary data concerning the use of dynamic risk stratification in MTC.

<table>
<thead>
<tr>
<th>Excellent response</th>
<th>Biochemical incomplete response</th>
<th>Structural incomplete response</th>
<th>Indeterminate response</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total thyroidectomy and RAI ablation Haugen et al. (2016)</strong></td>
<td><strong>Total thyroidectomy alone Momesso et al. (2016)</strong></td>
<td><strong>Lobectomy alone Momesso et al. (2016)</strong></td>
<td><strong>Lobectomy alone Momesso et al. (2016)</strong></td>
</tr>
<tr>
<td>Negative imaging</td>
<td>Nonstimulated Tg &lt; 0.2 ng/mL or Suppressed Tg &lt; 0.2 ng/mL or TSH-stimulated Tg &lt; 1 ng/mL, Undetectable TgAb</td>
<td>Negative imaging</td>
<td>Nonstimulated Tg &lt; 0.2 ng/mL or Suppressed Tg &lt; 0.2 ng/mL or TSH-stimulated Tg &lt; 1 ng/mL, Undetectable TgAb</td>
</tr>
<tr>
<td>Suppressed Tg &lt; 1 ng/mL or Stimulated Tg &lt; 10 ng/mL or Rising TgAb levels</td>
<td>Nonstimulated Tg &lt; 2 ng/mL</td>
<td>Stable nonstimulated Tg &lt; 30 ng/mL</td>
<td>Stable nonstimulated Tg &lt; 30 ng/mL</td>
</tr>
<tr>
<td>Undetectable TgAb</td>
<td>Nonstimulated Tg &gt; 5 ng/mL or Stimulated Tg &gt; 10 ng/mL or Increasing Tg values over the time with similar TSH levels or Rising TgAb levels</td>
<td>Nonstimulated Tg &gt; 10 ng/mL or Increasing Tg values over the time with similar TSH levels or Rising TgAb levels</td>
<td>Nonstimulated Tg &gt; 10 ng/mL or Increasing Tg values over the time with similar TSH levels or Rising TgAb levels</td>
</tr>
<tr>
<td>Nonspecific findings on imaging studies</td>
<td>Structural or functional evidence of disease With any Tg level</td>
<td>Structural or functional evidence of disease regardless of Tg or TgAb</td>
<td>Structural or functional evidence of disease regardless of Tg or TgAb</td>
</tr>
<tr>
<td>Faint uptake in thyroid bed on RAI scanning</td>
<td>With or without TgAb</td>
<td>Nonstimulated Tg &lt; 0.2 ng/mL or Nonstimulated Tg &lt; 0.2 ng/mL or Nonstimulated Tg &lt; 0.2 ng/mL</td>
<td>Nonstimulated Tg &lt; 0.2 ng/mL or Nonstimulated Tg &lt; 0.2 ng/mL or Nonstimulated Tg &lt; 0.2 ng/mL</td>
</tr>
<tr>
<td>Nonstimulated Tg detectable, but &lt;1 ng/mL</td>
<td>Nonstimulated Tg &lt; 1 ng/mL</td>
<td>Nonstimulated Tg &lt; 30 ng/mL</td>
<td>Nonstimulated Tg &lt; 30 ng/mL</td>
</tr>
<tr>
<td>Stimulated Tg detectable, but &lt;10 ng/mL</td>
<td>Nonstimulated Tg &lt; 2 ng/mL</td>
<td>Nonstimulated Tg &gt; 10 ng/mL</td>
<td>Nonstimulated Tg &gt; 10 ng/mL</td>
</tr>
<tr>
<td>or TgAb stable or declining in the absence of structural or functional disease</td>
<td>Nonstimulated Tg 2–10 ng/mL</td>
<td>Nonstimulated Tg &lt; 2 ng/mL</td>
<td>Nonstimulated Tg 2–10 ng/mL</td>
</tr>
<tr>
<td>Nonspecific findings on imaging studies</td>
<td>Nonstimulated Tg 0.2–5 ng/mL</td>
<td>Nonstimulated Tg &lt; 0.2 ng/mL</td>
<td>Nonstimulated Tg 0.2–5 ng/mL</td>
</tr>
<tr>
<td>or RAI scanning</td>
<td>Stimulated Tg &gt; 10 ng/mL or Increasing Tg values over the time with similar TSH levels or Rising TgAb levels</td>
<td>or</td>
<td>or</td>
</tr>
<tr>
<td>Structural or functional disease</td>
<td>Stable nonstimulated Tg &lt; 30 ng/mL</td>
<td>Structural or functional disease</td>
<td>Structural or functional disease</td>
</tr>
<tr>
<td>of disease regardless of Tg or TgAb</td>
<td></td>
<td>regardless of Tg or TgAb</td>
<td>regardless of Tg or TgAb</td>
</tr>
<tr>
<td>Non-specific findings on imaging studies</td>
<td>with similar TSH levels</td>
<td>Non-specific findings on imaging studies</td>
<td>Non-specific findings on imaging studies</td>
</tr>
<tr>
<td>or RAI scanning</td>
<td>or Rising TgAb levels</td>
<td>or TgAb levels stable or declining in the absence of structural or functional disease</td>
<td>or TgAb levels stable or declining in the absence of structural or functional disease</td>
</tr>
<tr>
<td>Faint uptake in thyroid bed on RAI scanning</td>
<td>or Rising TgAb levels</td>
<td>or TgAb levels stable or declining in the absence of structural or functional disease</td>
<td>or TgAb levels stable or declining in the absence of structural or functional disease</td>
</tr>
<tr>
<td>Nonstimulated Tg detectable, but</td>
<td>Structural or functional disease</td>
<td>Nonstimulated Tg 0.2–5 ng/mL</td>
<td>Structural or functional disease</td>
</tr>
<tr>
<td>&lt;1 ng/mL</td>
<td>regardless of Tg or TgAb</td>
<td>or</td>
<td>regardless of Tg or TgAb</td>
</tr>
<tr>
<td>or TgAb stable or declining in the absence of structural or functional disease</td>
<td>Nonstimulated Tg &lt; 0.2 ng/mL</td>
<td>Nonstimulated Tg 2–10 ng/mL</td>
<td>Nonstimulated Tg 2–10 ng/mL</td>
</tr>
<tr>
<td>RAI scanning</td>
<td>or Rising TgAb levels</td>
<td>or TgAb levels stable or declining in the absence of structural or functional disease</td>
<td>or TgAb levels stable or declining in the absence of structural or functional disease</td>
</tr>
</tbody>
</table>

RAI, radioiodine; Tg, thyroglobulin; TgAb, anti-Tg antibodies.

acceptable or incomplete). At the initial assessment, 23, 50 and 27% of patients from the whole group were classified as low, intermediate and high risk with the estimated risk of persistent structural or recurrent disease of 3, 18 and 66%, respectively. After restratification, in patients showing an excellent response, the likelihood of persistent structural disease or recurrent DTC was reduced from 3 to 2% in a low-risk group, from 18% to 2% in an intermediate-risk group and, what was particularly important, from 66% to 14% in a high-risk group. While in a case of an incomplete response, the likelihood of persistent structural or recurrent DTC increased up to 13, 41 and 79% in low-, intermediate- and high-risk groups, respectively. The authors concluded that a risk-adapted approach to follow-up cannot be based solely on static, initial stratification of the risk that remains unchanged over the life as a newly proposed recurrence staging system effectively predicted the risk of recurrence and persistent DTC (Tuttle et al. 2010).

This review summarizes 7-year experience with new principles of risk stratification. We mainly focused on DTC; however, we decided also to present some preliminary data concerning the use of dynamic risk stratification in MTC.

http://erc.endocrinology-journals.org © 2017 Society for Endocrinology
Printed in Great Britain
Published by Bioscientifica Ltd.
DOI: 10.1530/ERC-17-0270
Printed in Great Britain
### Table 3  
The summary of studies assessing delayed/dynamic risk stratification in differentiated thyroid cancer.

<table>
<thead>
<tr>
<th>Study reference</th>
<th>Study location</th>
<th>Study design</th>
<th>Study sample size</th>
<th>Median length of follow-up</th>
<th>Time of DRS after initial therapy</th>
<th>Proportion of LR/IR/HR patients at diagnosis according to ATA 2009 criteria</th>
<th>Proportion of LR/IR/HR patients after DRS</th>
<th>Treatment response</th>
<th>Outcome</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients treated with total thyroidectomy and RAI ablation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuttle et al. (2010)</td>
<td>USA</td>
<td>Retrospective analysis</td>
<td>588 DTC patients</td>
<td>7 years (1–15)</td>
<td></td>
<td></td>
<td>ATA LR 135 (23%)</td>
<td>DRS ERS 159 (34%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ATA IR 294 (50%)</td>
<td>DRS ARS 95 (20%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ATA HR 159 (27%)</td>
<td>DRS IRS 217 (46%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Castagna et al. (2011)</td>
<td>Italy</td>
<td>Retrospective analysis</td>
<td>512 DTC patients</td>
<td>5.6 years (1.08–52)</td>
<td>8–12 months</td>
<td></td>
<td>ATA LR 244 (47.6%)</td>
<td>DRS LR 353 (68.9%)</td>
<td>ERS: 86% ATA LR</td>
<td>1% ATA LR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ATA IR/HR 268 (52.2%)</td>
<td>DRS HR 159 (31.1%)</td>
<td>12% ATA IR</td>
<td>2% ATA IR</td>
</tr>
<tr>
<td>Hong et al. (2014)</td>
<td>Korea</td>
<td>Retrospective analysis</td>
<td>398 DTC patients</td>
<td>10.7 years (0.7–13.6)</td>
<td>8–15 months</td>
<td></td>
<td>ATA LR 81 (20.4%)</td>
<td>DRS ERS 229 (57.5%)</td>
<td>ERS: 75.3% ATA LR</td>
<td>1% ATA LR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ATA IR 259 (65.1%)</td>
<td>DRS ARS 78 (19.6%)</td>
<td>60.2% ATA LR</td>
<td>2% ATA IR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ATA HR 58 (14.6%)</td>
<td>DRS ARS 62 (15.6%)</td>
<td>20.7% ATA HR</td>
<td>4% DRS ERS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20.1% ATA IR</td>
<td>4% DRS ARS</td>
<td>0% DRS ARS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22.4% ATA HR</td>
<td>0% DRS ARS</td>
<td></td>
</tr>
</tbody>
</table>

**Proportion of LR/IR/HR patients at diagnosis according to ATA 2009 criteria**

- **ERS:** 86% ATA LR
- **DRS ARS:** 14% ATA HR
- **DRS IRS:** 87% DRS ARS

**Proportion of LR/IR/HR patients after DRS**

- **ERS:** 11% ATA LR
- **DRS ARS:** 95 (20%)
- **DRS IRS:** 277 (46%)
<table>
<thead>
<tr>
<th>Study</th>
<th>Patients</th>
<th>Follow-up</th>
<th>Time after therapy</th>
<th>ATA LR (%)</th>
<th>ATA IR (%)</th>
<th>ATA HR (%)</th>
<th>DRS LR (%)</th>
<th>ERS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jeon et al. (2014)</td>
<td>715 DTC</td>
<td>8 years</td>
<td>6–24 months</td>
<td>ATA LR 88</td>
<td>ATA IR 578</td>
<td>ATA HR 49</td>
<td>DRS LR 435</td>
<td>ERS 98%</td>
</tr>
<tr>
<td></td>
<td>patients</td>
<td></td>
<td></td>
<td>(12%)</td>
<td>(81%)</td>
<td>(7%)</td>
<td>(61%)</td>
<td>NED 98%</td>
</tr>
<tr>
<td>Kowalska et al. (2016)</td>
<td>916 DTC</td>
<td>7 years</td>
<td>9–12 months</td>
<td>ATA LR 573</td>
<td>ATA IR 369</td>
<td>ATA HR 185</td>
<td>DRS LR 731</td>
<td>ERS 95.4%</td>
</tr>
<tr>
<td></td>
<td>patients</td>
<td>(1–13)</td>
<td>after RAI treatment</td>
<td>(79.8%)</td>
<td>(40.3%)</td>
<td>(20.2%)</td>
<td>(79.8%)</td>
<td>LR 75.6%</td>
</tr>
<tr>
<td>Morosán et al. (2016)</td>
<td>90 DTC</td>
<td>4 years</td>
<td>After 2 years</td>
<td>ATA LR 42</td>
<td>ATA IR 39</td>
<td>ATA HR 9</td>
<td>DRS LR 535</td>
<td>ERS 74%</td>
</tr>
<tr>
<td></td>
<td>patients</td>
<td>(2–13)</td>
<td>of follow-up</td>
<td>(47%)</td>
<td>(43%)</td>
<td>(10%)</td>
<td>(61%)</td>
<td>LR 36%</td>
</tr>
</tbody>
</table>

Modified DRS including an additional variable of serum TgAb level effectively predicts recurrent/persistent DTC in patients after TT and RAI ablation.
### Patients treated with lobectomy or total thyroidectomy without RAI ablation

<table>
<thead>
<tr>
<th>Study reference</th>
<th>Study location</th>
<th>Study design</th>
<th>Study sample size</th>
<th>Median length of follow-up</th>
<th>Time of DRS</th>
<th>Proportion of LR/IR/HR patients at diagnosis according to ATA 2009 criteria</th>
<th>Treatment response</th>
<th>Outcome</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tavarelli et al. (2017)</td>
<td>France</td>
<td>Retrospective analysis</td>
<td>560 pT3 PTC patients Group II (328) T &gt; 10 mm + ETI</td>
<td>6.6 years (1.9–92)</td>
<td>6–8 months after initial treatment</td>
<td>Group I (160) T ≤ 10 mm + ETI 10-year DFS 89% Group II (328) T &gt; 10 mm + ETI 10-year DFS 67% Group III (78) T &gt; 40 mm no ETI 10-year DFS 67%</td>
<td>DRS NED 389 (78.4%) DRS persistent disease 107 (21.6%)</td>
<td>DRS NED 389 (78.4%) DRS persistent disease 107 (21.6%)</td>
<td>DFS in patients with NED at DRS Group 1 (160) T ≤ 10 mm + ETI 10-year DFS 89% Group II (328) T &gt; 10 mm + ETI 10-year DFS 67% Group III (78) T &gt; 40 mm no ETI 10-year DFS 67%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study reference</th>
<th>Study location</th>
<th>Study design</th>
<th>Study sample size</th>
<th>Median length of follow-up</th>
<th>Time of DRS</th>
<th>Proportion of LR/IR/HR patients at diagnosis according to ATA 2009 criteria</th>
<th>Treatment response</th>
<th>Outcome</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Momesso et al. (2016)</td>
<td>USA &amp; Brazil</td>
<td>Retrospective analysis</td>
<td>507 DTC patients 320 patients total thyroidectomy</td>
<td>100.5 months (24–510)</td>
<td>During 2 first years of follow-up</td>
<td>ATA(^a) LR 91.4% ATA IR 8.6%</td>
<td>After 2 years</td>
<td>ATA(^a) LR 91.4% ATA IR 8.6%</td>
<td>ATA(^a) LR 91.4% ATA IR 8.6%</td>
</tr>
<tr>
<td>Park et al. (2017)</td>
<td>Korea</td>
<td>Retrospective analysis</td>
<td>357 DTC patients 187 patients total thyroidectomy</td>
<td>8.6 years (6.4–9.5)</td>
<td>During 2 first years of follow-up</td>
<td>ATA(^a) LR 81.9% ATA IR 18.1%</td>
<td>After 2 years</td>
<td>ATA(^a) LR 81.9% ATA IR 18.1%</td>
<td>ATA(^a) LR 81.9% ATA IR 18.1%</td>
</tr>
</tbody>
</table>

\(^a\)Number of patients restratified 471.
ATA high-risk patients, respectively. At this time point, patients showing complete remission were restratified as DRS low-risk category, whereas subjects who had a persistent DTC were involved into DRS high-risk category. Thereby, nearly 50% of patients at first involved to ATA and ETA high-risk groups moved to a low-risk category and nearly 12% of patients from ATA and ETA low-risk groups were reclassified as a high-risk category. In the next step of the analysis, to demonstrate the validity of such re-classification, the authors attempted to correlate ATA, ETA and DRS with final outcomes. Complete remission was achieved in comparable rates of patients from ATA and ETA low-risk groups: 91.4% and 90.8%, respectively, whereas this value was significantly higher (96.6%) in DRS low-risk category. Considering ATA intermediate/high-risk and ETA high-risk groups, complete remission was noticed in 60.8% and 61.6% of patients, respectively. The percentage of patients with complete remission in DRS high-risk category was only in 27.1%. Conversely, 33.9% of patients from ATA intermediate/high-risk group and 32.8% from ETA high-risk group demonstrated persistent DTC. These rates were significantly lower than this one, noticed in DRS high-risk category (66.6%). Interestingly, no significant differences in recurrence rate after a period of complete remission between low- and high-risk groups were observed, regardless of the stratification system used. The authors believed that patients in clinical remission had a good prognosis irrespective of initial risk stage. Finally, to assess the ability to predict the outcomes (remission/persistent DTC), the authors calculated a positive predictive value (PPV), negative predictive value (NPV) and proportion of variance (PVE) for each particular stratification system. The values of all above-mentioned factors achieved for DRS (PPV 72.8%, NPV 96.3% and PVE 62.1%) were significantly better than those related to ATA and ETA initial risk classification (Castagna et al. 2011). Thus, the authors unequivocally concluded about the superiority of DRS over other stratification systems performed at DTC diagnosis.

The superiority of DRS (also these authors used the term ‘delayed’ proposed by the Italian group) was also demonstrated by other analysis of 398 DTC patients, who underwent total thyroidectomy and RAI ablation (Hong et al. 2014). This group had an initial risk evaluation, performed after RAI treatment, on the basis of TNM and ATA 2009 (Cooper et al. 2009). After 8–15 months, it was subjected to DRS and classified as having an excellent, acceptable, biochemical incomplete or structural incomplete response. Using ATA criteria, 20.4% of patients were evaluated as low risk, 65.1% as intermediate risk and 14.6% as high risk. At the time of DRS, 57.5% of patients demonstrated an excellent response, 19.6% acceptable response, 15.6% incomplete biochemical and 7.3% incomplete structural response. Considering ATA low-risk group, an excellent response was stated in 75.3%, acceptable in 16%, incomplete biochemical in 7.4%, and incomplete structural in 1.2% of patients. In ATA intermediate-risk group, 60.2% of patients showed an excellent response, 20.1% acceptable response, 16.2% biochemical incomplete and 22.4% structural incomplete response, whereas in ATA high-risk class, these rates were respectively 20.7, 22.4, 24.1 and 24.1%. At the end of this analysis, the authors compared the association of ATA 2009 risk system and DRS with progression-free survival (PFS). A multivariate analysis, which included DRS and the factors used for ATA risk stratification (extrathyroidal extension, regional lymph node metastases, histologic type) demonstrated a statistically significant association only between DRS and PFS. In another multivariate analysis of ATA risk stratification and the factors involved in DRS (stimulated serum thyroglobulin (Tg) level $\geq$ 1 ng/mL and TSH suppressed Tg level $\geq$ 1 ng/mL), ATA risk stratification did not show any significant association with PFS either but both stimulated and suppressed Tg levels $\geq$ 1 ng/mL did (Hong et al. 2014). In conclusion, according to the data given by ATA, Tg evaluation is crucial for risk assessment.

Korean authors evaluated dynamic risk stratification in a group of 715 DTC patients after total thyroidectomy and RAI ablation (Jeon et al. 2014). The authors called their approach a ‘modified dynamic risk stratification’ because they focused also on the status of anti-Tg antibodies in their risk assessment. Like in other papers, at DTC diagnosis, the patients were categorized according to ATA 2009 criteria (Cooper et al. 2009) as low risk (12%), intermediate risk (81%) and high risk (7%). Next, 6–24 months after an initial therapy, they were reclassified to one of the following category: excellent response, acceptable response, biochemical incomplete response and structural incomplete response: 61, 20, 7 and 12% of patients, with disease-free survival (DFS) rates of 98.3, 71.6, 54.2 and 6.8%, respectively. PVE value of this modified dynamic risk stratification was significantly higher than ATA and TNM risk stratification systems: 44.59% vs 12.14% and 8.68% respectively (Jeon et al. 2014).

Another paper, confirming the relevance of DRS, involved into a retrospective analysis a group of 916 DTC patients treated with total thyroidectomy and RAI ablation in a single institution between the years 2000 and 2013.
(Kowalska et al. 2016). Likewise, the patients underwent an initial staging according to ATA 2009 (Cooper et al. 2009) and ETA 2006 (Pacini et al. 2006) guidelines and were restaged 9–12 months after primary treatment (DRS) on the basis of response criteria proposed in Momesso & Tuttle (2014) as having an excellent response, biochemical incomplete response, structural incomplete response or indeterminate response. Patients, in whom an excellent response was demonstrated, were restaged by DRS as low risk, whereas remaining subjects as high risk. The majority of patients (62%) presented stage I at DTC diagnosis. Using ATA 2009 criteria 59.6, 34.6 and 5.8% of the patients were initially stratified as low, intermediate and high risk, respectively. While according to ETA classification, 20.9% were considered as very low risk, 37.3% as low risk and 41.8% as high risk. Nine to twelve months later, 79.8% of patients achieved an excellent treatment response (DRS low risk), whereas remaining 20.2% of patients were restaged at as DRS high-risk (among them 3.3% with biochemical incomplete response, 6.4% with incomplete structural response and 10.5% with indeterminate response). At the final follow-up, 83.6% of patients were free of disease, 12% had persistent DTC and 4.4% died, among them 1.9% due to DTC. However, there is no information to which risk-group the patients with cancer-related death were classified. Similar to Castagna study, Kowalska et al. (2016) calculated PPV, NPV and PVE. Their results were concordant with the data reported in Castagna et al. (2011), where DRS showed the highest ability in a prediction of DTC outcomes, with PPV, NPV and PVE 56.8, 98.5 and 56.7%, respectively (Kowalska et al. 2016).

Recently, a dynamic prediction of the risk of recurrence has been applied in a group of older DTC patients above the age of 60 years (Morosán et al. 2016). Ninety patients at median age of 65 years (range 62–82), treated with total thyroidectomy and RAI ablation between 2000 and 2013, and with follow-up period longer than 2 years, were included into a retrospective analysis. These patients were initially stratified using ATA 2009 criteria (Cooper et al. 2009), whereas their response to treatment was evaluated 2 years later. According to ATA 2009 risk classification 47% of patients were initially allocated to low-risk, 43% to an intermediate-risk and 10% to high-risk groups. After 2 years, an excellent response was stated in 74, 36 and 0%, an indeterminate response in 24, 51 and 100% and an incomplete response in 2, 13 and 0% of patients from ATA low-, intermediate- and high-risk groups, respectively. At the end of follow-up, 95% of patients from ATA low-risk group demonstrated NED, whereas in the ATA high-risk group, 100% of subjects showed persistent DTC. Considering ATA intermediate-risk group 61% of patients were free of disease and 39% had persistent disease. When this group was restaged 2 years later using response criteria, 100% of patients with an excellent response and only 61% of subjects with indeterminate/incomplete responses were disease free at final follow-up (Morosán et al. 2016).

The most recent study analyzed the impact of DRS on prognosis in intermediate-risk PTC (Tavarelli et al. 2017). This retrospective analysis included 560 pT3 PTC patients treated with total or near-total thyroidectomy and RAI ablation in a single institution. The study population was split into three groups according to tumor size and the presence of extrathyroidal invasion (ETI): group 1 (n = 160) pT3 ≤ 10 mm with ETI; group 2 (n = 328) pT3 >10 mm with ETI and group 3 (n = 78) pT3 due to tumor size >4 cm without ETI. At initial presentation, group 2 demonstrated more often lymph node metastases and extranodal extension. Lymph node metastases and multifocality were more frequent in group 1. The differences between the groups were significant. Moreover, patients from group 3 received a significantly higher RAI activity. Ten-year DFS was 89, 67, and 82% in groups 1, 2 and 3, respectively. Lymph node involvement, especially in the lateral compartment (N1b), male sex and group 2 had a negative significant impact on DFS. The patients were reevaluated 6–8 months after an initial treatment. Clinical remission was obtained in 78.4% of patients, whereas 21.6% had persistent disease. In patients with NED at the six- to eight-month check-up, 10-year DFS significantly improved in all groups: from 89% to 98% in group 1, from 67% to 96% in group 2 and from 82% to 90.5% in group 3. Considering patients with persistent PTC at check-up none of the patients from group 1 were disease free. DFS decreased to 19% in group 2 and to 57% in group 3 (Tavarelli et al. 2017).

**Dynamic risk stratification in DTC patients treated with total thyroidectomy or lobectomy, without RAI ablation**

To date, the results of two studies, evaluating the utility of dynamic risk stratification in patients treated only with surgery (total thyroidectomy or lobectomy) without RAI ablation, have been published. Criteria on how to classify treatment responses in these patients, proposed in Momesso et al. (2016), are given in Table 2.

The first one included 507 DTC patients, among whom, 187 underwent lobectomy (L) and 320 underwent total thyroidectomy (TT) (Momesso et al. 2016). First, after the operation, patients were categorized according to ATA
2015 modified criteria (Table 1) as low, intermediate and high risk and next, during the first 2 years of follow-up, they were reclassified according to treatment response (excellent, incomplete biochemical, incomplete structural or indeterminate). Considering L and TT patients alone, there were 81.9% and 91.4% of ATA low-risk patients and 18.1% and 8.6% of ATA intermediate-risk patients, respectively. There were no ATA high-risk patients in both groups, what is in accordance to indications for lobectomy. An excellent response was obtained in 64.2%, incomplete biochemical response in 9.1%, incomplete structural response in 2.1% and indeterminate response in 24.6% of L patients. These values were 16.5, 2.2, 1.9 and 79.4% for TT patients, respectively. Such high percentage of patients demonstrating an indeterminate response is surprising. The authors explained it by the presence of nonspecific US findings, nonstimulated Tg value between 0.2 and 5.0 ng/mL, and positive stable or declining anti-Tg antibodies. As it could be expected the risk of DTC relapse was low in patients with an excellent and indeterminate response but very high in patients demonstrating an incomplete structural response: 100% in both groups. Eight TT patients and 21 L patients required additional treatment. All but one showed NED at final follow-up (Momesso et al. 2016).

The results of the second study have just been published (Park et al. 2017). The authors retrospectively analyzed a group of 357 DTC patients after lobectomy or TT. As in the above-mentioned study, patients were initially evaluated according to ATA 2015 criteria (Table 1) as low or intermediate risk, 52.4% and 47.6%, respectively. DFS and recurrence risk did not differ significantly between ATA low and intermediate-risk groups. DRS, carried out during the first 2 years of follow-up classified 71.7% of patients as having an excellent response, 18.5% with an indeterminate response, 8.4% with a biochemical incomplete response and 1.4% of remaining patients with an incomplete structural response. The differences in DFS regarding treatment responses were statistically significant. The relative risk (RR) of DTC relapse among patients with an indeterminate response was nearly 2 times higher, with a biochemical incomplete response nearly 21 times higher and with a structural incomplete response nearly 243 times higher comparing to patients showing and an excellent response (Park et al. 2017).

**Dynamic risk stratification in MTC**

Following good clinical conclusions with the use of dynamic risk stratification system in DTC, there were also attempts to use it in MTC. Likewise in DTC, TNM classification allows a better prediction of cancer-related death than the risk of MTC relapse. Moreover, multiple studies, recently published demonstrated a role of calcitonin and carcinoembryonic antigen (CEA) doubling time as an important prognostic factor (Gawlik et al. 2010, Wells et al. 2015). Therefore, changes in our approach to risk stratification in MTC seem to be necessary, although they have not been recommended yet by the newest MTC ATA GL published in 2015 (Wells et al. 2015). The results of two retrospective analysis (Table 4), reported during recent 2 years, evaluating dynamic risk stratification are encouraging; however, the data require a validation in prospective studies (Lindsey et al. 2015, Kwon et al. 2016). Both studies based the evaluation of treatment response on the following criteria: serum calcitonin and CEA concentration (undetectable in an excellent response, elevated in a biochemical incomplete response), presence of structural disease (absent in an excellent and incomplete biochemical response; present in an incomplete structural response regardless of calcitonin and CEA concentration).

The first study included 287 patients, both with hereditary and sporadic MTC (Lindsey et al. 2015). First risk stratification was carried out using TNM/AJCC system (7th edition). Twenty-three percent of patients were staged AJCC I, 11% AJCC II, 7% AJCC III and 59% AJCC IV. Next, they were subjected to dynamic risk stratification including the data obtained during the first year of initial therapy. Excellent response was obtained in 27% of patients from the whole group, whereas 35% and 38% of patients achieved biochemical or structural incomplete response, respectively. Patients showing an excellent treatment response were very likely to have NED (78%). Only 15% of patients from this group experienced a biochemical recurrence, 4% structural recurrence and 3% died during the further follow-up. Conversely, only 1% of patients with an incomplete biochemical response achieved NED at final follow-up. The risk of cancer-related death among patients with an incomplete structural response was 56% over a median of five years of follow-up (Lindsey et al. 2015).

The second retrospective analysis with a very similar study pattern to the previous one involved 120 MTC patients (Kwon et al. 2016). At MTC onset, 37% of patients showed AJCC I stage, 16% AJCC stage II, 13% AJCC stage III and 34% AJCC IV. During one-year follow-up, 70% of patients demonstrated an excellent response, 23% biochemical incomplete response, whereas 7% of the remaining patients had structural disease. There were
### Table 4  The summary of studies evaluating dynamic risk stratification in medullary thyroid cancer.

<table>
<thead>
<tr>
<th>Study reference</th>
<th>Study sample size</th>
<th>Median length of follow-up</th>
<th>Time of DRS</th>
<th>Proportion of AJCC I/AJCC II/AJCC III/AJCC IV patients at MTC diagnosis</th>
<th>Proportion of patients with ERS/IBRS/ISRS after DRS</th>
<th>Treatment response</th>
<th>Outcome</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lindsey et al. (2015)</td>
<td>287 MTC patients</td>
<td>5.2 years (0.2–33.2)</td>
<td>Risk reassessment was based on the data obtained during first year of initial therapy</td>
<td>DRS ERS 65 (23%) AJCC I; 31 (11%) AJCC II; 19 (7%) AJCC III; 164 (59%) AJCC IV</td>
<td>DRS IBRS 102 (35%) DRS ISRS 108 (38%)</td>
<td>ERS: AJCC I 80%</td>
<td>NED: AJCC I 71% AJCC II 29% AJCC III 5% AJCC IV 3% Persistent biochemical disease AJCC I 18% AJCC II 58% AJCC III 48% AJCC IV 21% Persistent structural disease AJCC I 12% AJCC II 3% AJCC III 32% AJCC IV 61% ISRS: AJCC I 18% AJCC II 3% AJCC III 26% AJCC IV 34% Recurrent MTC: Biochemical only AJCC I 18% AJCC II 10% AJCC III 0% AJCC IV 2% Structurally identifiable disease AJCC I 13% AJCC II 0% AJCC III 5% AJCC IV 0% MTC-related death AJCC I 2% AJCC II 10% AJCC III 16% AJCC IV 40%</td>
<td>DRS that uses response to therapy variables to adjust risk estimates over time provides more useful clinical prognostic information than static initial anatomic staging in MTC.</td>
</tr>
</tbody>
</table>
Kwon et al. (2016) 120 MTC patients 6.2 years (4.1–9.2)
Korea Retrospective analysis

Risk reassessment was based on the data obtained during the first year of initial therapy

AJCC I 44 (37%) DRS ERS 84 (70%)
AJCC II 20 (16%) DRS IBRS 28 (23%)
AJCC III 15 (13%) DRS ISRS 8 (7%)
AJCC IV 41 (34%)

ERS:
AJCC I 98%
AJCC II 89%
AJCC III 67%
AJCC IV 33%
IBRS:
AJCC I 2%
AJCC II 0%
AJCC III 0%
AJCC IV 17%

DRS based on the best response to the initial therapy can provide useful prognostic information in addition to initial TNM staging for predicting of mortality, as well as the likelihood of NED in MTC.

NED:
AJCC I 93%
AJCC II 95%
AJCC III 60%
AJCC IV 17%

Persistent biochemical disease
AJCC I 0%
AJCC II 0%
AJCC III 0%
AJCC IV 36%
Persistent structural disease disease
AJCC I 0%
AJCC II 0%
AJCC III 0%
AJCC IV 17%

Recurrent MTC:
Biochemical only
AJCC I 5%
AJCC II 0%
AJCC III 7%
AJCC IV 14%
Structurally identifiable disease
AJCC I 0%
AJCC II 0%
AJCC III 0%
AJCC IV 2%
MTC-related death
AJCC I 0%
AJCC II 0%
AJCC III 0%
AJCC IV 14%

Number of patients stratified 279.
AJCC, American Joint Cancer Committee; ERS, excellent response; DRS, dynamic risk stratification; IBRS, incomplete biochemical response; ISRS, incomplete structural response; MTC, medullary thyroid cancer; NED, no evidence of disease.
significant differences in survival both according to AJCC and DRS. However, a higher value of PVE for DRS (49.1%) suggested its better usefulness in predicting the likelihood of NED than TNM system (PVE 28.7%) (Kwon et al. 2016).

**Nomenclature**

The statements ‘ongoing risk stratification’ and ‘dynamic risk stratification’ are used interchangeably in this paper. However, some authors cited in this manuscript also used ‘DRS’, what may be interpreted as a contrast to the ‘initial risk stratification’.

**Dynamic risk stratification in the light of the newest ATA guidelines**

All retrospective studies discussed above clearly demonstrated that the use of dynamic risk assessment allowed for a better prediction of DTC recurrence risk. This idea has been reflected in the newest, updated 2015 ATA GLs (Haugen et al. 2016). The 2015 ATA GLs clearly emphasize that, although this initial staging system allows for a valuable assessment of DTC recurrence risk, it provides static information based only on data available at DTC onset and therefore it is not sufficient for a lifelong follow-up. All retrospective analyses discussed above, as well as our personal clinical experience, prove that substantial proportion of patients initially classified as an intermediate and high risk achieve complete remission (or using current term: an excellent response) and become low risk of DTC recurrence. Thus, there is a necessity to restratify them appropriately taking into consideration their response to treatment applied. It has been proposed to carry out such restratification during the first 2 years of follow-up and at any point later. As it has been confirmed by the papers, already cited by us, the risk of DTC recurrence in a particular patient varies according to the achieved best response. Noteworthy, criteria for treatment response published in the current 2015 ATA GLs (Table 2) are in fact related only to patients who underwent TT and RAI ablation (Haugen et al. 2016). To restratify the subjects treated with TT or lobectomy alone, one should base on the criteria proposed in Momesso et al. (2016) (Table 2). Likewise, this evaluation considers nonstimulated serum Tg level, TgAb level and imaging studies (Momesso et al. 2016). These Tg cut-off values are different than these proposed for patients after TT and RAI ablation and have not been validated by an independent study yet.

There are still some unanswered questions that in our opinion need to be resolved in the nearest future.

**Histopathological controversies regarding 2015 ATA initial risk stratification**

First important information, required for an adequate risk stratification, should be provided by a postoperative histopathological report that has to include basic tumor features necessary for TNM staging and additional data regarding vascular invasion, the number of invaded vessels, the number of examined lymph nodes, the number metastatic lymph nodes and their size of as well as presence/absence of extranodal extension. Histopathologic variants of PTC or FTC should also be identified.

Surprisingly, 2015 ATA GLs did not specify poorly differentiated thyroid cancer (PDTC), although this type of thyroid cancer is obviously characterized by a high risk of relapse and a bad prognosis. Moreover, Hurthle cell carcinoma (HCC, oncocytic variant of follicular carcinoma) was mentioned in the description of recommendation 46 concerning the basic principles of histopathological evaluation, but it was not present in the graph of ‘Risk of Structural Disease Recurrence’. Encapsulated HCC with an extensive vascular invasion (EVI) implies a high risk of recurrence observed in 64% of such tumors (Xu et al. 2015). This paper also reported a high risk of recurrence in patients without distant metastases at the time of diagnosis. HCCs with EVI relapsed even vascular invasion counted 4 or 7 foci and capsular invasion was unifocal (Xu et al. 2015).

AJCC/UICC stratification is strongly recommended for each newly diagnosed DTC patient to predict DTC mortality. Simultaneously with TNM classification, the patients should be subjected to an initial stratification to evaluate the risk of persistent or recurrent DTC. The updated guidelines still recommend the ATA 2009 initial clinic-pathologic risk stratification system (Cooper et al. 2009), currently completed with additional variables such as the extent of lymph node involvement, mutational status and/or vascular invasion and focus on few significant factors such as clinical and pathological metastatic lymph node involvement, extension of vascular invasion, multifocality of PTC, encapsulated follicular variant of papillary thyroid carcinoma and, considering FTC, minimally invasive follicular thyroid carcinoma (Table 1).

The 2015 ATA GL still classify patients as a low, intermediate and high risk of recurrence. Patients having
intrathyroidal tumor, ≤5 lymph node micrometastases, below <0.2 cm in diameter fulfill a low-risk criteria only when they are diagnosed with classical of follicular PTC variant or minimal invasive FTC. The intermediate-risk group involve PTC tumors with aggressive histology, RAI-avid DTC cervical foci outside thyroid bed, microscopic extrathyroidal extension, vascular invasion or >5 metastatic lymph nodes 0.2–3 cm in size. Patients showing aggressive PTC variants also belong to this group. Unfortunately, ATA do not specify the risk for PTC diffuse sclerosing variant, because its prognostic implication is controversial. This variant is characterized by a higher rate of local and distant metastases at presentation and lower DFS than classical PTC. Nevertheless, the overall mortality is rather low, with 10-year disease-specific survival of nearly 93%. However, for patients’ safety, we propose to classify this variant to a high-risk category. Finally, the patients with gross extrathyroidal extension, lymph node metastases >3 cm in diameter, distant metastases (M1) or after incomplete tumor resection constitute a high-risk group. The risk of recurrence varies according to these groups. One could guess, because ATA does not define it, that it is lower than 10% for low-risk subjects, between 10 and 30% for intermediate-risk patients and above 30% for the high-risk group. We believe that the risk has to be unequivocally established by prospective studies. Interestingly, the risk of structural disease recurrence is depicted rather as a continuum than clearly separated 3 groups.

It has to be stressed that both clinical and pathological features listed in 2015 ATA GL risk stratifications systems may be subjective and vary in evaluation from institution to institution. Accumulation of pathological features, which are subjective in interpretation without established standardized criteria, may impact on clinical management.

Minimal extrathyroidal extension (mETE) is an important factor present in the ATA Risk Stratification System and Risk of Structural disease recurrence. It points pT3 tumors to an intermediate-risk group. ETE is characterized by carcinomatous involvement of the sternothyroidal muscle or perithyroidal soft tissue. This factor is evaluated and diagnosed by a microscopic examination. Individual pathologists’ interpretation and disagreement on histological criteria for mETE influences inter-observer variation in identification of mETE. The study performed among 11 thyroid experienced pathologists, based on analysis of 69 scanned slides selected from potentially controversial cases with mETE, presented that the overall strength of agreement in identifying mETE was slight (kappa coefficient factor 0.14). Most pathologist interpreted as mETE perithyroidal fat, skeletal muscle, nerve and thick-walled vessel involvement. Less than half of pathologists also reported using desmoplasia as a criterion for diagnosis of mETE. Inter-observer concordance of the identification of mETE between expert pathologists was moderate to poor (Su et al. 2016). Thus, one has to be aware that mETE is a controversial prognostic factor in PTC and in ATA risk stratifications schemes.

Clinical N0 or N1 evidence is identified by imaging or either palpation preoperatively or intraoperatively. Sonographic examination of lymph nodes among others depends on clinician’s skills, sonographic equipment and patient’s body. These factors are difficult to standardize. Reproducibility of palpation performed by surgeon preoperatively or intraoperatively is related to different factors, which made this examination subjective.

Surgical completeness of lymph node dissection, macroscopic examination and careful microscopic assessment influence the number of positive lymph nodes. On the other hand, micrometastases can be missed by a pathologist. On each stage, both surgical and pathological examinations errors can be made and preclude adequate lymph node assessment (Urken et al. 2016).

Another important issue is the lack of pathological criteria for evaluation of extranodal extension (ENE). ENE is a factor, which increases the risk of recurrences, therefore, the standardization of pathological designation of ENE in the future studies is particularly important (Urken et al. 2016). Ito et al. (2009) analyzed ENE in 2 age groups of patients >55 years old and <55 years old. In the first group, only large lymph node metastases in the lateral cervical compartment affected DFS, whereas in the younger one the number of metastatic nodes, ENE and size of metastatic nodes influenced DFS (Ito et al. 2009). According to the College of American Pathologists Protocol for the Examination of Specimens from Patients with carcinomas of the Thyroid Gland from 2017 (Seethala et al. 2017) the presence or absence of ENE should be noted in a pathology report, but the details of pathological criteria required for diagnosis are not explained. The agreement between 11 experts in thyroid pathology in histological diagnosis of ENE was only fair with kappa coefficient factor of 0.35. The study showed that pathologists utilized different histologic criteria for ENE diagnosis. While most pathologists used a criterion ‘tumor beyond nodal capsule’, some of them needed desmoplastic stromal reaction outside node to diagnose ENE. The full agreement was not reached in identification of perinodal skeletal muscle, nerve or thick-walled vessel.
(Du et al. 2016). The lack of concordance in identifying ENE could impact on the disease prognosis, patient care and interpretation of clinical studies that evaluating the influence of ENE on the disease prognosis. Noteworthy, the level of identification of ENE in a group of general pathologists could be lower than one in the cited paper.

Taking into a consideration the recent data regarding a possible impact of BRAF (Xing et al. 2013, 2015) and TERT mutations (Melo et al. 2014, Xing et al. 2014) on DFS and PTC disease-specific mortality, the authors of the guidelines believe that mutational status may improve risk estimation but when interpreted in the context of other clinical and histopathological risk factors. However, this statement requires a prospective validation.

The lack of prospective studies

According to the 2015 ATA GL, an initial recurrence risk estimates should be continually modified during follow-up, because the risk of recurrence and disease-specific mortality can change over the time as a function of the clinical course of the disease and the response to the therapy. This is a strong recommendation. However, the quality of evidence is low because to date there is no prospective study, which validates the data coming from retrospective analyses discussed above. The question arising is whether the use of dynamic risk stratification is safe in patients initially categorized as high risk? The diagnosis of an excellent response in these patients, demonstrating according to TNM classification a higher risk of cancer-related death, may influence their further follow-up to be less strict. Such situation may result in a delay in the diagnosis of DTC recurrence. Therefore, the necessity to carry out prospective studies is our important task for the nearest future.

Dynamic or delayed risk stratification

The idea proposed by Tuttle et al. (2010) and currently recommended by the newest 2015 ATA GL concerned a continuous = dynamic risk stratification to evaluate recurrence risk and adjust patient’s management to it. However, some papers, analyzed in this review, proposed the term ‘DRS’ (Castagna et al. 2011, Hong et al. 2014, Kowalska et al. 2016) probably in opposite to an ‘initial risk stratification’. One should ask the question whether ‘DRS’ means exactly the same as ‘dynamic risk stratification’. DRS in fact reflects a single-point risk assessment, whereas dynamic risk stratification is a real-time evaluation during the whole follow-up that incorporates an individual response to therapy, what is necessary for the personalized approach to ongoing management. Tuttle et al. (2010) proposed to base the assessment of treatment response on clinical data obtained during first 2 years of follow-up but 2015 ATA GL suggest to describe the clinical status at any point during follow-up.

Differences between published studies

There are some issues concerning retrospective studies, described in this review that has to be discussed.

All are single institution studies so a selection bias has to be considered. The time, when the treatment response was evaluated varies across the studies (Table 3). In some papers the authors analyzed all clinical data obtained during first 2 years of follow-up to assess treatment response (Tuttle et al. 2010, Momesso et al. 2016, Park et al. 2017), whereas the others carried out risk restratification in different single time points between 8 and 24 months after initial therapy (Castagna et al. 2011, Hong et al. 2014, Jeon et al. 2014, Kowalska et al. 2016, Morosan et al. 2016). One may ask the question which data are more accurate for risk restratification: these collected during the first 2 years of follow-up or those obtained during single assessment between 8 and 24 months after initial therapy?

Another important issue is related to different follow-up protocols used in distinct centers and outcome variables of reported studies. The number of follow-up visits and diagnostic examinations may influence the diagnosis of cancer recurrence. Considering outcome variables, some of papers analyzed the percentage of patients with NED or persistent disease, recurrence rates, and the number of cancer-related deaths (Tuttle et al. 2010, Castagna et al. 2011, Kowalska et al. 2016, Morosan et al. 2016, Momesso et al. 2016, Park et al. 2017), whereas other studies performed PFS or DFS analyses (Hong et al. 2014, Jeon et al. 2014, Tavarelli et al. 2017) to compare the patients with different stages of initial stratification or treatment response. We have to be aware that important limitation of PFS or DFS is how closely patients are followed. If visits are rare the diagnosis of cancer recurrence may be delayed.

Summary

Dynamic risk stratification substantially changes our approach to DTC patients. It allows for a real-time adequate risk evaluation during the whole follow-up and influences our therapeutic decisions. It should be emphasized here, that a high quality of both histopathological and
biochemical examinations is crucial for a proper risk stratification. An excellent response to treatment justifies less strict DTC monitoring and decreases the degree of TSH suppression. It seems to be particularly important regarding the data published by Dutch authors few years ago. They demonstrated that DTC patients exhibited a higher risk of cardiovascular and all-cause mortality, which both were affected by the level of TSH suppression (Klein Hesselink et al. 2013). Conversely, patients with biochemical and structural incomplete responses require TSH suppression, ongoing observation and imaging studies. Moreover, in a case of structural incomplete response, an additional therapy (RAI or other) may be necessary. Regarding an indeterminate response the guidelines recommend to continue observation with serum Tg repeated assessments and serial imaging of nonspecific lesions (Haugen et al. 2016).

These new ATA GLs have been widely discussed and influenced our management in DTC. Although their adaptation to Polish epidemiological situation was required, ATA 2015 GLs became a basis on which the Polish National Guidelines have been recently updated and published (Jarzab et al. 2016).

Conclusions

Dynamic risk stratification allows for better prediction and individualized the risk of recurrence in DTC, what has been demonstrated in numerous retrospective analyses. However, the validation of this approach in prospective studies seems to be our task for near future.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of this review.

Funding

The paper was supported by the Polish National Center of Research and Development MILESTONE project – Molecular diagnostics and imaging in individualized therapy for breast, thyroid and prostate cancer, grant no. STRATEGMED-2/267398/4/NCBR/2015.

References

Castagna MG, Malino F, Cipri C, Belardini V, Theodoropoulou A, Cevonini 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. European Journal of Endocrinology 165 441–446. (doi:10.1530/EJE-11-0466)


Lindsey SC, Gandy L, Palmer F & Tuttle RM 2015 Response to initial therapy predicts clinical outcomes in medullary thyroid cancer. Thyroid 25 242–249. (doi:10.1089/thy.2014.0277)

promoter mutations are a major indicator of poor outcome in differentiated thyroid carcinomas. *Journal of Clinical Endocrinology and Metabolism* [99](E754-E765). (doi:10.1210/jc.2013-3734)


Urken ML, Haser GC, Likhterov I & Wenig BM 2016 The impact of metastatic lymph nodes on risk stratification in differentiated thyroid cancer: have we reached a higher level of understanding? *Thyroid* [26](481–488). (doi:10.1089/thy.2015.0544)


Received in final form 10 July 2017
Accepted 18 August 2017
Accepted preprint published online 18 August 2017