Prognostic factors in differentiated thyroid carcinomas and their implications for current staging classifications

Arja Jukkola, Risto Bloigu1, Tapani Ebeling2, Pasi Salmela2 and Guillermo Blanco

Department of Oncology, Oulu University Hospital, Finland
1Department of Medical Informatics, Oulu University Hospital, Finland
2Department of Internal Medicine, Oulu University Hospital, Finland

(Requests for offprints should be addressed to G Blanco, Department of Clinical Oncology, Oulu University Hospital, PL 22, FIN-90229 Oulu, Finland; Email: guillermo.blanco@ppsph.fi)

Abstract

Differentiated thyroid carcinomas (DTC) (papillary, follicular and follicular type of papillary) have a favourable prognosis, but a proportion of patients develop recurrences and eventually die of the disease. Various prognostic factors have been identified and been used to create the current staging classifications (AGES, AMES, MACIS, EORTC, UICC-TNM). We examined 499 DTC patients retrospectively to validate known prognostic factors that enable them to be recognised as having either a low or a high risk of death related to a recurrence of DTC, by reference to the current staging classifications. Sixty-nine of them (14%) had local or distant recurrences, the mean time to recurrence being 7.7 years. The 10-year disease-free survival rate was 80%, and the ten-year overall survival rate for the entire group was 91%, with a mean survival time of 8.7 years. Male gender, a follicular type of tumour, larger tumour size, extrathyroidal invasion outside the capsule and nodal metastases were all related to a higher incidence of tumour recurrence, and the follicular type of histology, age > 45 years, larger tumour size and local invasion entailed poorer survival. The AMES and to some extent the EORTC classification were not reproducible in this material, mainly because some prognostic variants were no longer encountered or were insufficient in number to allow reliable conclusions to be drawn. The MACIS staging classification leaves the definition of the intermediate and high risk groups too wide and is therefore not very reliable. Pooling of stages I and II improved the relevance of the TNM classification. All the current staging classifications are able to discern a low risk DTC group well. We achieved a highly accurate definition of risk in the present material using only two parameters, age (cut-off value 50 years) and extracapsular invasion of the thyroid gland.

Introduction

The incidence of thyroid carcinoma in Finland has been steadily rising so that, according to the Finnish Cancer Registry (2003), there were 303 new cases in 2001, representing 1.3% of all new cancers recorded in that year and implying a twofold increase for men and an increase of approximately 30% for women during the last decade. Thyroid carcinoma is almost four times as frequent in females as in males in Finland, 6.6/100 000 compared with only 2.1/100 000 in men (Finnish Cancer Registry 2003).

Papillary and follicular thyroid carcinoma, referred to as differentiated thyroid carcinoma (DTC) cover the majority of thyroid carcinoma cases. The prognosis for DTC is usually excellent, but even so a proportion of the patients develop recurrences and eventually die of the disease. Age at the time of diagnosis, gender, histological type, tumour size and extrathyroidal invasion have been found previously to be associated with a poor clinical outcome (Cady & Rossi 1988, De Groot et al. 1990, Hay 1990, Schlumberger & Pacini 1999), while the prognostic impact of regional lymph node metastases is still a

The predictability of the outcome is closely related to these risk factors, which have been used to develop various prognostic staging systems to discern patients with either a low or a high risk of recurrence or DTC-related death. Many scoring or staging systems have been developed for this purpose during the last two decades, but up to now there has been no consensus regarding the clinical relevance of any one staging classification rather than the others. The most commonly used staging classifications are AMES (Cady & Rossi 1988), AGES (Hay et al. 1987), MACIS (Hay et al. 1993), EORTC (Byar et al. 1979), and UICC-TNM (Hermanek & Sobin 1992). These use well-defined prognostic factor values that enable the construction of scores, which in turn can be used to give accurate predictions of the long-term outcome for individual patients. The clinical applicability of these staging classifications for patients with DTC has certain limitations, mainly because they do not reflect the actual clinical presentation of patients with DTC, nor do they consider the impact of early diagnosis and therapy on the outcome of DTC.

This study examines retrospectively 499 cases with DTC treated at our institution during the years 1970–1999 to define more closely the prognostic factors that were related to a poor clinical outcome and to look for a staging classification which will not only facilitate treatment decisions but will also provide a reliable method for comparing results between centres.

Patients and methods

Patients

The patient characteristics are described in Table 1. Four hundred and ninety-nine cases of well differentiated thyroid carcinoma (DTC), 366 (73%) of the papillary type, 47 (10%) of the follicular type and 86 (17%) of the follicular papillary type (Table 1), treated in Northern Finland during the years 1970–1999, were analysed in this retrospective study. The average population of the Oulu and Lapland regions together is about 500,000. The majority of the patients were women (82%) (Table 1), and the mean age was 47 years (range 14–86). The mean age of the patients with papillary carcinoma was 46 years (range 14–86), that of the follicular carcinoma group was 50 years (range 19–83), while that of the patients with the follicular variant of papillary carcinoma was the lowest, 45 years (range 15–75). The highest incidence of DTC was at 41–50 years in the females and at 51–60 years in the males. The mean follow-up time was 8.7 years (range 0.16–32).

Hashimoto’s thyroiditis was found in 18% of the patients and struma nodosa in 33%. Achieved Basedow disease occurred only in 0.3%. Other cancers developed in 3.2% of the cases, but were restricted entirely to those patients with the papillary type. In 4% of the cases with papillary carcinoma, DTC was also found among close relatives.

An overwhelming majority of the tumours were classified as T1 (<1 cm) or T2 (1–4 cm) (95%) (Table 1), representing 96% of the cases in the papillary group, 81%
in the follicular group and 99% in the follicular variant of papillary thyroid carcinoma. The proportion of tumours less than 1 cm was uneven, being 28% in the papillary group, but only 3% in the follicular group, while the proportion in the group with the follicular variant of papillary carcinoma was quite similar to that in the papillary group, 25%. Attachment to the thyroid capsule was detected in a quarter of the patients, and infiltration through the capsule occurred in 24%. Invasion of the extracellular tissue was found in the lymph vessels (13%), muscle (9%) or blood vessels (9%). The main metastatic site was the regional lymph nodes (11%) (Table 1).

Incidental tumours were recognised in approximately 8% of the cases undergoing thyroidectomy for reasons other than carcinoma.

Total thyroidectomy and iodine-131 ablation treatment to destroy the thyroid remnant was the main therapy used in this population. Patients with nodal metastases in the neck region, as determined by ultrasound and needle biopsy or during operation, underwent neck surgery. Serum thyroglobulin and thyrotrophin (TSH) concentrations (intended suppressed values TSH $\leq 0.05$ mU/l) were measured during the follow-up. Total thyroidectomy was performed on 426 of the surgical patients (87%), subtotal thyroidectomy in 43 cases (9%) and lobectomy in 17 cases (4%) respectively. Only 4 patients (1%) were considered inoperable. No data were available on 9 patients. Only 62 patients (12%) underwent neck surgery. Ninety-two percent of the patients showed thyroid tissue remnants after scanning with 5 mCi iodine-131 after surgery, but only in 1% was there iodine positivity outside the neck region. The patients exhibiting thyroid tissue remnants were treated with iodine-131 ablation of 100 mCi if they had negative lymph nodes, or of 150 mCi if the lymph nodes were positive for malignancy. In almost half of the cases, two ablative procedures were required to destroy the remaining thyroid tissue. In 4% of the patients, external radiotherapy to the neck region was used because of inoperable local disease or because of iodine-131 negativity in the thyroid remnant or in lymph nodes in the neck region that were positive for malignancy.

Mild transient hypocalcaemia was found in 134 patients after surgery, and 13% of them needed calcium substitution. Unilateral recurrent laryngeal nerve palsy was detected in 3.5% of the cases and bilateral recurrent laryngeal nerve palsy in five cases.

The study proposal was approved by the Ethical Committee of the Faculty of Medicine, University of Oulu and Oulu University Hospital, and the Finnish Ministry of Health gave permission for the further use of patients’ tissue samples.

---

**Statistical analyses**

Standard tests of statistical significance were used: Student’s $t$-test, Mann–Whitney, Pearson and Chi-Square. Survival curves (Kaplan–Meier) were compared using the log rank, Breslow or Taritone-Ware test. $P < 0.05$ was considered statistically significant. Cox regression analysis was used to find significant predictors of survival in the different staging classifications. The proportion of deaths due to thyroid carcinoma was too low to create a reliable multivariate analysis.

**Results**

**Validation of predictors of DTC recurrence**

Sixty-nine patients out of the 499 had local or distant recurrences (14%), the average time to recurrence being 7.7 years. The 10-year disease-free survival rate in the whole group was 80%. Gender predicted the recurrence of DTC ($P = 0.003$), males having a poorer prognosis than females (Table 2), but age at onset did not predict recurrence ($P = 0.6$). The follicular type of tumour entailed a higher risk of recurrence ($P = 0.01$). The majority of the patients (70%) had tumours <1–4 cm (T2) in diameter, and larger tumours carried a higher risk of recurrence ($P = 0.0001$). Local invasion, in the form of capsule invasion ($P = 0.0001$), lymph vessel invasion ($P = 0.0001$), blood vessel invasion ($P = 0.03$) and muscle invasion ($P = 0.0001$), was also found to be associated with a higher incidence of tumour recurrence, but this was not the case with attachment of the tumour to the thyroid capsule alone. The 14 patients with nodal metastases had a significantly shorter recurrence-free survival than those without nodal disease ($P = 0.0001$) (Table 2).

Four hundred and twenty-six patients underwent total thyroidectomy and the rest underwent either hemithyroidectomy or subtotal resection, with the exception of four patients who had an inoperable disease. The patients treated with total thyroidectomy had a significantly better recurrence-free survival rate than the other surgical cases ($P = 0.03$), but the patients on whom neck surgery was performed for primary nodal metastases ($P = 0.0001$) and those who received radiotherapy to the neck region had poorer recurrence-free survival rates ($P = 0.007$) (Table 2).

**Validation of factors for overall survival**

The 5-year overall survival rate for the entire DTC group was 95%, and the 10-year survival rate was 91%. The mean survival time was 8.7 years. In the univariate analysis the follicular type of histology ($P = 0.01$), age $>45$ years ($P = 0.0001$), larger tumour size ($P = 0.0001$) and local invasion, i.e. invasion beyond the capsule ($P = 0.0001$), lymph vessel invasion ($P = 0.008$), blood
vessel ($P = 0.0007$) and muscle ($P = 0.0001$) infiltration entailed a shorter survival. Patients treated with total thyroidectomy had a better prognosis, as did those who did not receive radiotherapy. The patients with primary distant metastases had a poor survival outcome. Gender, presence of nodal metastases or neck surgery did not affect survival (Table 2).

The AMES classification

The first report from the Lahey Clinic Foundation (Cady et al. 1979) used three prognostic factors for the analysis of risk in cases of DTC, namely age (cut-off 40 years for men and 50 years for women), gender and histological type: follicular vs papillary. Applying the same prognostic factors as stated in the first publication of the AMES
staging classification (Cady et al. 1979) to this present population, it was possible to discern a low-risk group with a survival rate of 100% at 20 years, an intermediate-risk group with a survival rate of 84% and a high-risk group with a survival rate of 54% (Fig. 1A). The second report from the same institution (Cady & Rossi 1988) added two further parameters that were important for the prognosis: the size of the tumour and the local and distant extents of the disease, thus achieving a more accurate prognosis and distinguishing a low-risk group and a high-risk group with survival rates of 84% and 54%, respectively.
risk group. Unfortunately, the addition of a tumour size over 5 cm and the presence of distant metastases, requirements for the elaboration of the expanded risk definition as stated in the second AMES staging classification (Cady & Rossi 1988), were not applicable to our population, so that it was not possible to reproduce this classification, which had to be disregarded.

The MACIS scoring classification
This classification used standard survival analysis to estimate the percentage of patients surviving DTC at specific times after the initial treatment, together with a comparison of survival curves and a multivariate analysis to define the variables examined as possible predictors of DTC survival (Hay et al. 1993). After validation of many of the prognostic variables in a set of patients, five variables emerged as useful predictors of survival: age (cut-off 40 years), size of the tumour, incomplete operation, extrathyroidal invasion and distant metastases. These variables were used to construct a system for assigning performance scores and defining four prognostic groups (0–6.9, 7.0–7.9, 8.0–8.9, 9.0+), with an accurate DTC mortality rate for each at 20 years of follow-up. When our population was assessed according to these variables and the scoring system used to create the MACIS staging classification, four risk groups emerged: a low-risk group (score 0–6.9) with a survival rate of 94% at 20 years, a high-risk group (9+) with a survival rate of 58%, and two other risk groups (7-7.9 and 8-8.9) that lay in between but had less clear, superimposed survival curves, both with a survival rate of 83% (Fig. 1B).

The TNM classification
The TNM classification is a good attempt at conciliating and unifying thyroid carcinoma with the International Classification Guide for Tumours (Hermanek & Sobin 1992). This classification integrates all histological variants of thyroid carcinoma, but a separate stage group for papillary and follicular carcinoma together is recommended. It takes into account age (cut-off 45 years), size of the tumour, nodal involvement and distant metastases. The staging classification defines the low-risk group well in our population, with a survival rate of 100% at 20 years (stage I), and also distinguishes the highest risk group (stage IV), as expected, as it has no survivors after the same length of time, but again the intermediate-risk groups are less well defined (stage II and stage III), with survival curves that are superimposed and represent survival rates of 74% and 85% respectively at 20 years (Fig. 1C). No deaths were detected in patients <45 years, which is why these patients are not included in Fig. 1C.

A proposal for risk definition - age and extracapsular involvement
By taking into account two relevant factors defined by univariate analysis in the present material, age (<50 years and ≥50 years) and extracapsular involvement, as the sole prognostic variables for discerning low-risk and high-risk categories it was indeed possible in a very simple way to recognise a low-risk category with a survival rate of 89% at 20 years and a high-risk category with a survival rate of 60% (Fig. 1D). Comparisons of the factors used in these staging classifications and the Cox regression analysis for the risk of death in relation to the classifications are presented in Tables 3 and 4 respectively.

Discussion
In general, DTC has a favourable prognosis, but it is of great importance to identify at the time of diagnosis those patients who have a high risk of progressive disease and DTC-related death. The importance of recognising prognostic variables is also relevant for the optimal management of DTC, e.g. the extent of thyroid surgery and the indications for postoperative radioiodine therapy. A recognised prognostic classification is also critical for the comparison of treatment results. Gender proved here to be of prognostic value for disease-free survival, which was shorter for males than for females. Our patients older than 45 years also had a poorer disease outcome, but advanced age did not predict shorter recurrence-free survival (Table 2). Gender has been found to be an independent prognostic factor (Cunningham et al. 1990), but controversial results have also been published (Gilliland et al. 1997). Earlier studies have shown that the prognosis is poor in older patients with DTC (Rigway 1992, Schlumberger & Pacini 1999). Lymph node metastases were a risk factor for recurrence, and they were most numerous in the papillary type of carcinoma (Table 2), as described earlier (Mazzaferri & Jhiang 1994, Schlumberger 1998). In some, but not all, studies (Hay 1990, Mazzaferri & Jhiang 1994), local lymph node involvement has been associated with an increased risk of tumour recurrence and also with DTC-related mortality (Mazzaferri & Jhiang 1981, Mizukami et al. 1992, Schlumberger & Pacini 1999). The prognostic impact of regional lymph node metastases still remains a controversial issue. The presence of primary distant metastases beyond the neck entails a poorer prognosis, and patients treated with radiotherapy to the neck region at the time of diagnosis also have a poor prognosis (Table 2).

The accuracy of the preoperative diagnosis of thyroid malignancy has substantially improved over the years, thanks to the combined use of ultrasound and fine needle biopsy, and this has markedly influenced the recognition
of malignancy at earlier stages (Mäkäräinen 1986). The change towards earlier diagnosis may have implications for the use of the current staging classifications. Here, we set out to explore the predictability of the outcome by using defined valid prognostic variables applied to the current staging classifications, which in general serve well to define prognostic groups, although the applicability of these classifications to the present DTC material is less reliable, especially with regard to the definition groups with a high and an intermediate risk of DTC-related death. It has been suggested that one of the criteria for a good staging system is that the survival curves for the stage groups are well separated (Hannequin et al. 1986, Brierley et al. 1997).

When the prognostic factors in this population are assessed in terms of the first AMES classification (Cady et al. 1979), it is possible to discern three risk groups - low, intermediate and high (Fig. 1A, Table 3). The extended AMES staging classification which takes into account tumour size and the extent of the disease (Cady & Rossi 1988) was not reproducible in this material, mainly because some prognostic variants were no longer encountered or the numbers were insufficient to allow reliable conclusions to be drawn.

We also had difficulties in reproducing the results when using the EORTC or AGES classifications. EORTC yields five risk groups based on age, gender, cell type, presence of anaplastic carcinoma, tumour size and

<table>
<thead>
<tr>
<th>Classification</th>
<th>No. of patients</th>
<th>Risk of death odds ratio (95% confidence interval)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TNM</td>
<td>411</td>
<td>1</td>
<td>0.002</td>
</tr>
<tr>
<td>Stage I and II</td>
<td>380</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Stage III</td>
<td>16</td>
<td>14.4 (2.6–7.9)</td>
<td>0.002</td>
</tr>
<tr>
<td>Stage IV</td>
<td>15</td>
<td>83.2 (23–298)</td>
<td>0.0001</td>
</tr>
<tr>
<td>AMES</td>
<td>498</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>280</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td>189</td>
<td>25.7 (3.4–195)</td>
<td>0.002</td>
</tr>
<tr>
<td>High</td>
<td>29</td>
<td>52.2 (6.2–438)</td>
<td>0.0001</td>
</tr>
<tr>
<td>MACIS</td>
<td>498</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0–6.99</td>
<td>419</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>7–7.99</td>
<td>43</td>
<td>3.5 (1.1–11)</td>
<td>0.03</td>
</tr>
<tr>
<td>8–8.99</td>
<td>20</td>
<td>7.2 (1.9–26)</td>
<td>0.003</td>
</tr>
<tr>
<td>9+</td>
<td>16</td>
<td>25 (8.3–78)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Extracapsular invasion and age</td>
<td>498</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>450</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>48</td>
<td>11.6 (4.8–27)</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
number of metastatic sites expressed in terms of individual scores which distinguish survival curves (Byar et al. 1979), but because of the difficulties in matching the variables used in this classification, it was also disregarded.

Four risk groups emerged in our population when we used the MACIS classification (Table 3, Fig. 1B). This staging classification uses prognostic scores to create intermediate risk groups, albeit with statistical significance but too close to each other, making the definition of the intermediate-risk group too broad and not very reliable.

The TNM classification meets with the same difficulties. This defines the low-risk group (stage I) and the high-risk group (stage IV) well in our population, but not the intermediate-risk group (stages II and III) (Table 3, Fig. 1C). The only prognostic variable for patients aged under 45 years is the presence or absence of distant metastases, whereas for those over 45 years of age the classification relies on the size of the tumour, extension of the disease to the lymph nodes and the presence of distant metastases. The risk of recurrence may be underestimated in young patients because they are classified as having stage I disease in the absence of distant metastases even in the presence of locally aggressive tumours. The new 2002 proposal differs substantially from the 1992 classification, the main differences being related to extrathyroidal extension of the disease, nodal involvement and, to some extent, tumour size. The clinical application of the new proposal could only be tested in a prospective study.

The use of only two parameters, age (<50 years and ≥50 years) and extracapsular invasion of the thyroid gland, allowed us to achieve an appropriate and easy definition of risk in this set of patients (Table 3, Fig. 1D). A very close attempt to ours has been made at the Institute Gustave-Roussy (the IGR classification), where a highly significant definition of risk in DTC was obtained using only age and malignancy grade (Schlumberger & Pacini 1999).

Although all the current classifications are able to discern a low-risk group with an excellent prognosis, early diagnosis means that many prognostic variables are no longer found in our population, making the predictability of the traditional staging classifications less accurate today than at the time when they were developed. Brierley et al. (1997) have reported for a population representing the years 1958–1985 that there was no difference in the abilities of the top five classifications (AGES, TNM, EORTC, MACIS and AMES) to predict the prognosis for patients with differentiated thyroid carcinoma, and no statistically significant superiority of any system over the TNM classification was found. A staging classification needs to identify patients for whom there is a high risk of failure. Most differentiated thyroid carcinoma patients nowadays have a stage I–II disease and the proportion of high-risk patients is expected to be small. The differences in patient distribution between the staging classifications reflect problems in applying a classification developed for one population to another (Brierley et al. 1997).

The majority of the patients presented here could be defined as having a low risk of recurrence or of DTC-related death, a finding that raises another important question: the need to avoid overtreating low-risk DTC patients and to consider the use of less radical procedures.

References


Hay ID, Grant CS, Taylor WF & McConahey WM 1987 Ipsilateral lobectomy versus bilateral lobar resection in


