REVIEW

Prognostic Scores for Predicting Recurrence in Patients with Differentiated Thyroid Cancer

Charoonsak Somboonporn

Abstract

Background: Differentiated thyroid cancer (DTC) is a cancer group that shares molecular and cellular origin but shows different clinical courses and prognoses. Several prognostic factors have been reported for predicting recurrence for individual patients. This literature review aimed to evaluate prognostic scores for predicting recurrence of DTC. Materials and Methods: A search of the MEDLINE database for articles published until December 2015 was carried out using the terms "thyroid neoplasms AND (recurrent OR persistent) AND (score OR model OR nomogram)". Studies were eligible for review if they indicated the development of prognostic scoring models, derived from a group of independent prognostic factors, in predicting disease recurrence in DTC patients. Results: Of the 308 articles obtained, five were eligible for evaluation. Two scoring models were developed for DTC including both papillary and follicular carcinoma, one for papillary carcinoma, and the other two for papillary microcarcinoma. The number of patients included in the score development cohort ranged from 59 to 1,669. The number of evaluated potential prognostic factors ranged from 4 to 25. Tumor-related factors were the most common factors included in the final scores, with cervical lymph node metastases being the most common. Only two studies showed internal validation of the derived score. Conclusions: There is a paucity of prognostic scores for predicting disease recurrence in patients with DTC, in particular for follicular thyroid carcinoma. Several limitations of the created scores were found. Performance of the scores has not been adequately studied. Comprehensive validation in multiple cohorts is recommended before widespread use.

Keywords: Thyroid neoplasms - recurrence - prognosis - prognostic scores

Asian Pac J Cancer Prev, **17 (5)**, 2369-2374

Introduction

Thyroid cancer comprises a group of tumors with different characteristics. Differentiated thyroid cancer (DTC) comprises papillary carcinoma and follicular carcinoma, including their variants. Mostly, DTC has an indolent clinical course with low mortality. The 10-year survival rates for papillary and follicular thyroid cancer have been reported to be 93-94% and 84-85%, respectively (Hundahl et al., 1998; Mazzaferri and Kloos, 2001). However, a more aggressive disease with low survival and high recurrence occurs in some patients. Although disease-specific mortality is of great concern, recurrence is another important outcome as a result of a long-term clinical course of the disease. Identifying patients of potentially having high risk of recurrence helps individual management plan.

A wide variety of individual prognostic factors for DTC recurrence have been reported in the literature (Haugen et al., 2016); however, using all of them is not practical. Hence, prognostic scoring models have been developed to ease risk stratification in routine clinical practice. The aim of this review was to identify and evaluate available prognostic scores in predicting disease recurrence in patients with DTC.

Materials and Methods

Eligibility criteria for selecting studies

A review of previously published articles exploring prognostic scores for DTC recurrence was carried out. A search of the MEDLINE database until December 2015 was done using the terms "thyroid neoplasms AND (recurren* OR persisten*) AND (scor* OR model OR nomogram*)" with language limitation to articles published in English. After screening the titles and abstracts, as well as full text articles when needed, the eligible articles were identified for data extraction. The eligibility criteria included studies identifying independent prognostic factors for DTC recurrence using multivariate analysis and creating scoring model from these significant prognostic factors.

Data extraction and presentation

*Department of Radiology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand *For correspondence: charoonsak.som@gmail.com*

These data were extracted and presented for each study: first author's name, publication year, country of origin, characters of study patients, recurrence rate, follow-up period, the number of patients in the score development and validation cohort (if presented), the number of factors included in the scoring model, explored potential prognostic factors, independent prognostic factors obtained from multivariate analysis, and performance of the score (if presented). Individual prognostic factors were categorized according to the

Table 1. Prognostic Scores for Differentiated Thyroid Cancer Recurrence

First author		Welsch M	Onitilo AA	Buffet C	Niemeier LA	Kim KM
Publication year		2007	2009	2012	2012	2013
Country		Germany	USA	France	USA	Korea
Study patients		DTC	DTC	PMCA	PMCA	PTC
Recurrence rate (%)		18.7	9.6	4.1	10.2	5.7
Follow-up period, year (Mean + SD or mean)		9 ± 5	20	6.5 ± 6.7	5.3* and 4.8**	8.3
Number of patients in the score development cohort		171	614	1,669	59	593
Number of patients in the score validation cohort		NA	595	NA	40	NA
Number of prognostic factors included in the scoring model		25	4	3	4	4
Tumor factors	Tumor histology	o■	o■		o	
	Tumor histopathologic grading	o■				
	Tumor size	●■	●■	o		
	Tumor fibrosis				●■	
	Psammoma bodies				o	
	Anatomic location of primary tumor				o	
	Superficial location of primary tumor				●■	
	Multifocality			●■	●■	
	Distance to resected margin of primary tumor				o	
	Histologic lymph node metastases	●■	●■	●■		
	Preoperative imaging lymph node metastases					●■
	Vascular invasion			o		
	Infiltration of surrounding tissue or ETE	o■			o	●■
	Tumor capsule involvement	o■				
	Thyroid capsule invasion	●■		o		
	Infiltrative border				o	
	Positive resection margin				o	
	Lymph node status morphology	●■				
	Lymph node in PET scan	●■				
	Lymph node in RAI scan	●■				
	Pulmonary metastases morphology	●■				
	Pulmonary metastases in PET scan	●■				
	Pulmonary metastases in RAI scan	o■				
	Bone scan	o■				
	Bone metastases morphology	●■				
	Bone metastases in PET scan	●■				
	Bone metastases in RAI scan	●■				
	Other organ metastases morphology	●■				
	Other organ metastases in PET scan	●■				
	Other organ metastases in RAI scan	●■				
	Distant metastases		●			
Patient factors	Age	o■	●■	o		●■
	Gender	o■	o	●■		●■
	Serum Tg level before RAI treatment	●■				
	Serum Tg level 6 weeks after RAI treatment	●■				
	Current serum Tg level (at any time)	●■				
	Hashimoto's thyroiditis				o	
	BRAF V600E mutation				●■	
Treatment factors	Type of thyroid surgery		o	o		
	Post-thyroidectomy RAI treatment		o	o		
	Initial cervical lymph node surgery			o		
	Initial diagnosis year			o		

DTC, differentiated thyroid cancer; PTC, papillary thyroid carcinoma; PMCA, papillary microcarcinoma; PET, positron emission tomography; RAI, radioactive iodine; Tg, thyroglobulin ; ETE, extrathyroid extension; NA, not applicable due to score validation not performed; *in aggressive group, **in non-aggressive group; o indicates prognostic factors that did not show statistically significant association with recurrence; ● indicates prognostic factors that showed statistically significant association with recurrence; ■ indicates prognostic factors selected into the prognostic score

Table 2. Prognostic Factors and Corresponding Score Points in the Multiparameter Scoring System (Welsch et al., 2007)

Prognostic factors	Characteristics	Score points
Gender	Male	1
	Female	-1
Age	For every year < 45	-0.1
	For every year > 45	0.1
Histology	Papillary	-1
	Follicular	1
Primary histopathologic grading	G1	-2
	G2	1
	G3	2
	G4	5
Primary lymph node status (according to histologic report)	No metastases	-1
	Ipsilateral metastases	2
	Contralateral metastases	3
Tumor diameter	For every mm < 20	-0.2
	For every mm > 20	0.2
Infiltration of surrounding tissue	No	0
	Yes	5
Tumor capsule involvement	Yes	-2
	No	1
Thyroid capsule	No contact to capsule	-2
	Contact to capsule	2
	Extension beyond capsule	4
Lymph node status morphology	No metastases	-1
	Ipsilateral metastases	2
	Contralateral metastases	4
Lymph node in PET scan	No FDG uptake	-2
	Positive FDG uptake	2
Lymph node in RAI scan	Positive RAI uptake	-1
	No RAI uptake	2
Pulmonary metastases morphology	No metastases	-1
	One metastasis	3
	≥2 metastases	5
Pulmonary metastases in PET scan	No FDG uptake	-2
	Positive FDG uptake	4
Pulmonary metastases in RAI scan	Positive RAI uptake	-1
	No RAI uptake	4
Bone scan	Negative	-1
	Positive	3
Bone metastases morphology	No metastases	0
	One metastasis	5
	≥2 metastases	7
Bone metastases in PET scan	No FDG uptake	-2
	Positive FDG uptake	4
Bone metastases in RAI scan	Positive RAI uptake	-1
	No RAI uptake	4
Other organ metastases morphology	No metastases	-1
	One metastasis	2
	≥2 metastases	3
Other organ metastases in PET scan	No FDG uptake	-1
	Positive FDG uptake	1
Other organ metastases in RAI scan	Positive RAI uptake	-1
	No RAI uptake	2
Serum Tg level before RAI treatment (ng/mL)	<1	-5
	1-10	0
	10-100	2
	>100	4
Serum Tg level 6 weeks after RAI treatment (ng/mL)	<1	-5
	1-2	0
	2-5	1
	5-20	2
	20-100	4
	>100	6
Serum Tg level at any time of presentation (ng/mL)	<1	-5
	1-2	0
	2-5	2
	5-20	3
	20-100	5
	>100	7

PET, positron emission tomography; RAI, radioactive iodine; Tg, thyroglobulin

concept of "subject-based classification" as tumor-related factors, patient-related factors, and treatment-related factors (Gospodarowicz et al., 2006).

Results

Of the 308 articles screened, five studies were identified and included for review (Welsch et al., 2007; Onitilo et al., 2009; Buffet et al., 2012; Niemeier et al., 2012; Kim et al., 2013). All are retrospective cohort study. The studies were published during 2007 to 2013. Two studies were from the United States (Onitilo et al., 2009; Niemeier et al., 2012), two from Europe (Welsch et al., 2007; Buffet et al., 2012), and one from Asia (Kim et al., 2013). The study size of the score development cohort ranged from 59 to 1,669 patients, and in total consisted of 3,106 participants. The study patients were DTC (including both papillary and follicular carcinoma) in two studies (Welsch et al., 2007; Onitilo et al., 2009), while the remaining three studies included exclusively patients with papillary carcinoma (Buffet et al., 2012; Niemeier et al., 2012; Kim et al., 2013); two of them (Buffet et al., 2012; Niemeier et al., 2012) focused only patients with primary tumor size of ≤1 cm, the so-called papillary microcarcinoma (PMCA).

Table 1 represents a summary of the studies included in this review. Prognostic factors that were found without and with statistically significant association with disease recurrence were indicated as 'open circle' and 'closed circle', respectively. In addition, factors that were selected into the prognostic scoring models were indicated as 'square'. Finally, the number of prognostic factors included in the scoring models ranged from 4 to 25. Most of the independent prognostic factors were tumor pathologic factors, such as cervical lymph node metastases, tumor size, multifocality, and extrathyroid extension (ETE). Age and gender were patient factors that were found significantly associated with recurrence in two of the four studies examining these factors. In one study, BRAF mutation was found to be a prognostic factor for recurrence in PMCA patients. No treatment factors were found to be significantly associated with recurrence outcome.

In three studies, prognostic scores were developed from evaluated prognostic factors that showed statistically significant association with recurrence outcome (Buffet et al., 2012; Niemeier et al., 2012; Kim et al., 2013). However, the other two studies had variation in selecting factors into the final scoring scheme. The proposed score by Welsch et al. in 2007 adopted all potential prognostic factors into the scoring scheme, even though some of them were not statistically significant. In the study by Onitilo et al. in 2009, although showing no significant association with recurrence, tumor histology was forced into the model, while distant metastasis, an independent factor, was excluded from the model.

Application of these five prognostic scoring systems was described as follows

1. Multiparameter scoring system (Welsch et al., 2007): This score was calculated by addition or subtraction of the score points corresponding to each prognostic factor (Table 2). Factors that were not available were scored with the value zero. Five categories of probability of disease recurrence were estimated according to the corresponding cut-off points: 3.9% for scores ≤ 10; 14.1% for scores -10 to 0; 46.2% for scores 0 to 10; 66.7% for scores 10 to 20; and 100% for scores >20.

2. Alternative quantitative to the TNM scoring system (Onitilo et al., 2009): Prognostic score = sum of the followings: Histopathology = 1 if not papillary thyroid carcinoma, otherwise 0; Age = 4 if ≥ 45 years, otherwise 0; Regional lymph node metastases = 4, otherwise 0; Primary tumor size = 6 if > 4 cm limited to thyroid gland or any tumor with ETE, otherwise 0. Three categories of probability of 10-year disease-free survival (DFS) were estimated according to the corresponding cut-off points: 95% for scores 0 to 5; 70% for scores 6 to 10; and 45% for scores 11 to 15.

3. Scoring system for PMCA (Buffet et al., 2009): Prognostic score = sum of the followings: Gender: female = 0, male = 1; Multifocality: absent = 0, present = 1; Lymph node metastases: absent = 0, present = 3. Three categories of 10-year recurrence probability (with 95% CI) were estimated according to the corresponding cut-off points: 2.7% (1.3 to 4.1%) for scores 0 to 2; 24% (13.7 to 34.3%) for scores 3 to 4; and 42.6% (12.7 to 72.5%) for score 5.

4. Combined molecular-pathological score for PMCA (Niemeier et al., 2012): Prognostic score = 1 if tumor in superficial location of thyroid gland + 1 if BRAF positive + 1 if multifocality present + 1 if extensive fibrosis in primary tumor present. Three categories of recurrence risk were estimated according to the corresponding cutoff points: 0% for scores 0 to 2; 20% for score 3; and 60% for score 4.

5. Clinical prognostic index of University of Yonsei (Kim et al., 2013): Prognostic score = (0.03 x age in years) + 0.8 if male gender + 0.5 if ETE present + 0.7 if preoperative lymph node metastases present. Four categories of 10-year DFS were estimated according to the corresponding cut-off points: 95.6% for scores <1.50; 94.5% for scores 1.50 to 2.29; 85.0% for scores 2.30 to 3.29; and 27.3% for scores ≥3.30.

Discussion

Since combined multiple prognostic factors perform better in predicting the outcome than individual variable does, all potential factors can be evaluated and the factors that have the most degree of association with the outcome are identified. Weight of association of these factors can be used to create a summed prognostic score for predicting the outcome. This review identified five studies attempting to construct the risk prediction model for disease recurrence in patients with DTC.

The review found diversity among patient characteristics, variability in definition of outcome and event rate, and variability in the evaluated prognostic factors and their definitions in each study. Of all five prognostic scores, two were developed in DTC patients, but three scores were studied in purely papillary thyroid

carcinoma; two of these focused only on PMCA patients. None was carried out in patients with purely follicular thyroid carcinoma. In fact, the total number of patients diagnosed with follicular carcinoma included in the two studies (Welsch et al., 2007; Onitilo et al., 2009) was only 225. Caution is needed in applying these scores for predicting outcome in affected patients.

Definition of recurrence outcome varied across the studies. The prognostic score of Welsch et al. reported outcome as recurrence rate, while the score of Kim et al. reported it as disease-free survival. However, the definition of recurrence in both studies was not clearly defined. Onitilo et al. reported outcome as disease-free survival defined as the time to first recurrence; however, the definition of recurrence was not clearly defined. Buffet et al. reported outcome as time to recurrence and clearly defined the definition of recurrence, based on cytological or histological analysis, ^{131}I uptake in the lesion, and predetermined serum thyroglobulin (Tg) level. Niemeier et al. (2012) reported outcome as recurrence rate and defined recurrence based on serum Tg measurements, neck ultrasonography, and clinical examination.

The candidate prognostic factors included for score development varied widely among studies. The majority of them were tumor factors, such as tumor histology, grading, tumor size, ETE, or cervical lymph node metastases. Patient factors were also evaluated, including age, gender, serum Tg level at various time points, and BRAF mutation. Only a few treatment factors were studied, such as type of thyroidectomy and ^{131}I treatment. Among all studies, most of the independent prognostic factors were tumor pathologic factors and they were in accordance with those previously identified in the literature, such as cervical lymph node metastases (Schneider, et al., 2013; Lee et al., 2014), tumor size (Ito et al., 2012), multifocality (Qu et al., 2014), and extrathyroid extension (Kim et al., 2014). This was also the case for patient factors, such as age and gender (Kim et al., 2014), and postoperative serum Tg level (Hasbek et al., 2014). BRAF mutation included in the combined molecular-pathological score for PMCA (Niemeier et al., 2012) has recently been reported as a new and promising prognostic factors related to tumor aggressiveness and high recurrent outcome (Kim et al., 2012; Fernandez et al., 2013; Xing et al., 2015). It should be noted that although none of treatment factors were found to be significantly associated with recurrence, only two studies included factors of this domain in their score development cohorts (Onitilo et al., 2009; Buffet et al., 2012).

Most of the prognostic scores attributed to only 3 or 4 high-impact predictors, which are easy to use, and are therefore practical in routine clinical use (Onitilo et al., 2009; Buffet et al., 2012; Niemeier et al., 2012; Kim et al., 2013). The prognostic score which accounted for several predictors as the multiparameter scoring system seems not easy to use, although the authors claimed that only the available predictors could be used to calculate the summed score for individual patient (Welsch et al., 2007). In addition, the scoring models, in which the total score was calculated from the point scores that were integers, would likely ease the use of the scores in the real-world clinical practice (Onitilo et al., 2009; Buffet et al., 2012; Niemeier et al., 2012).

The sample size and the number of outcome events are crucial issues for prognostic studies because they affect the regression model assumption. Normally, the models require an adequate number of outcome events for each predictor. The rule of thumb dictates that at least 10 outcome events should be obtained for one independent prognostic factor (Guyatt, 2006). All studies in this review consisted of relatively low recurrent rates, ranging from 4.1 - 18.7% (or 6 - 68 events), likely due to short follow-up period, with the mean of only 5.3 - 9 years. This raised the concern as to whether or not these studies had adequate statistical power (Guuatt, 2006), particularly in the study exploring a large number of prognostic factors (Welsch et al., 2007).

Good model performance is important for widespread clinical application of the score. Variety in reporting score performance was observed, such as percent of true or false classification of the outcome (Welsch et al., 2007), % sensitivity and % specificity (Niemeier et al., 2012), or proportion of variant explained (Kim et al., 2013). The remaining two studies reported no information about model performance (Onitilo et al., 2009; Buffet et al., 2012). No studies reported score performance in terms of discrimination ability of the score, as measured by concordance C-index, and the score calibration (Pencina and D'Agostino, 2004).

Internal and external validation of the derived score is another aspect in assessing prognostic score utilization. Two studies reported results of internal validation of their prognostic scores, in 595 patients in the study of Onitilo et al. and 40 patients in the study of Niemeier et al., respectively. None of the evaluated scores performed external validation. Therefore, further studies should be carried out by performing external validation of these available scores in a different patient setting with an adequate number of sample size, such as a least 100 participants (Vergouwe et al., 2005).

There are some limitations of the review. Robust conclusion cannot be made regarding the prognostic score for predicting DTC recurrence due to varying definitions of outcomes and inconsistent use of prognostic factors. Since searching of articles was only done in MEDLINE database, there is a possibility that some eligible studies indexed only in other databases may be missed. Language bias may be present because the search strategy was limited to article published in English.

In conclusion, there is a paucity of data on prognostic scores for predicting disease recurrence in patients with DTC. Several limitations of the proposed scores were found. Performance was not adequately studied. Further studies with comprehensive internal and external validation in multiple cohorts are recommended before widespread use.

Acknowledgements

This review was supported by research fund from the Royal Golden Jubilee Ph.D. scholarship from the Thailand Research Fund and the Faculty of Medicine, Khon Kaen

University, Thailand.

References

Buffet C, Golmard JL, Hoang C, et al (2012). Scoring system for predicting recurrences in patients with papillary thyroid microcarcinoma. *Eur J Endocrinol*, **167**, 267-75.

Fernandez IJ, Piccin O, Sciascia S, et al (2013). Clinical significance of BRAF mutation in thyroid papillary cancer. *Otolaryngo Head and Neck Surg*, **148**, 919-25.

Gospodarowicz MK, O'Sullivan B, Koh ES (2006). Prognostic factors: principles and applications. In: 'Prognostic factors in cancer. 3rd ed', Eds Gospodarowicz MK, O'Sullivan B, Sobin LH. John Wiley & Sons, New Jersey, 23-38.

Guyatt G (2006). Determining prognosis and creating clinical decision rules. In 'Clinical epidemiology: how to do clinical practice research.3rd ed', Eds Haynes RB, Sackett DL, Guyatt G, et al. Lippincott Williams & Wilkins, Philadelphia, 323-55.

Hasbek Z, Turgut B, Kilicli F, Altuntas EE, Yucel B (2014). Importance of postoperative stimulated Tg level at the time of 131I ablation therapy for differentiated thyroid cancer. *Asian Pac J Cancer Prev*, **15**, 2523-7.

Haugen BR, Alexander EK, Bible KC, et al (2016). 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*, **26**, 1-133.

Hundahl SA, Flemming ID, Fremgen AM, Menck HR (1998). A National Cancer Data Base report on 53,856 cases of thyroid carcinoma treated in the U.S., 1985-1995. *Cancer*, **83**, 2638-48.

Ito Y, Kudo T, Kihara M, et al (2012). Prognosis of low-risk papillary thyroid carcinoma patients: its relationship with the size of primary tumors. *Endocr J*, **59**, 119-25.

Kim SJ, Lee KE, Myong JP, et al (2012). BRAF V600E mutation is associated with tumor aggressiveness in papillary thyroid cancer. *World J Surg*, **36**, 310-7.

Kim KM, Park JB, Bae KS, et al (2013). Clinical prognostic index for recurrence of papillary thyroid carcinoma including intraoperative findings. *Endocr J*, **60**, 291-7.

Kim WY, Kim HY, Son GS, Bae JW, Lee JB (2014). Clinicopathological, immunohistochemical factors and recurrence associated with extrathyroidal extension in papillary thyroid microcarcinoma. *J Cancer Res Ther*, **10**, 50-5.

Lee J, Song Y, Soh EY (2014). Prognostic significance of the number of metastatic lymph nodes to stratify the risk of recurrence. *World J Surg*, **38**, 858-62.

Mazzaferri EL, Kloos RT (2001). Current approaches to primary therapy of papillary and follicular thyroid cancer. *J Clin Endocrinol Metab*, **86**, 1447-63.

Niemeier LA, Kuffner Akatsu H, Song C, et al (2012). A combined molecular-pathologic score improves risk stratification of thyroid papillary microcarcinoma. *Cancer*, **118**, 2068-77.

Onitilo AA, Engel JM, Lundgren CI, et al (2009). Simplifying the TNM system for clinical use in differentiated thyroid cancer. *J Clin Oncol*, **27**, 1872-8.

Pencina MJ, D'Agostino RB (2004). Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med*, **23**, 2109-23.

Qu N, Zhang L, Ji QH, et al (2014). Number of tumor foci predicts prognosis in papillary thyroid cancer. *BMC Cancer*, **14**, 914.

Schneider DF, Mazeh H, Chen H, Sippel RS (2013). Lymph node ratio predicts recurrence in papillary thyroid cancer. *Oncologist*, **18**, 157-62.

Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD (2005). Substantial effective sample sizes were required for external validation studies of predictive logistic regression models, *J Clin Epidemiol*, **58**, 475-83.

Welsch M, Abeln M, Zaplatnikov K, et al (2007). Multiparameter scoring system for the prognosis of differentiated thyroid cancer. *Nuklearmedizin*, **46**, 257-62.

Xing M, AlzahraniAS, Carson KA, et al (2015). Association between BRAF V600E mutation and recurrence of papillary thyroid cancer. *J Clin Oncol*, **33**, `42-50.