Clinical Oncology 31 (2019) e11–e20

Contents lists available at ScienceDirect

Clinical Oncology

journal homepage: www.clinicaloncologyonline.net

Original Article

The Use of Post-ablation Stimulated Thyroglobulin in Predicting Clinical Outcomes in Differentiated Thyroid Carcinoma – What Cut-off Values Should We Use?

K.C.W. Wong *, T.Y. Ng *, K.S. Yu *, J.S.S. Kwok †, K.C.A. Chan †, J.J.S. Suen *, S.F. Leung *, A.T.C. Chan *

* *Department of Clinical Oncology, Prince of Wales Hospital, Shatin, Hong Kong, China*
† *Department of Chemical Pathology, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, Hong Kong, China*

Received 2 January 2018; received in revised form 26 September 2018; accepted 1 October 2018

Abstract

Aims: Recently published international guidelines recommended using the stimulated thyroglobulin (sTg) post-radioactive iodine (RAI) ablation, in conjunction with tumour stage, as a risk stratification factor. The choice of cut-off values for sTg, namely 1 and 10 ng/ml, was, however, largely based on the functional sensitivities of the assays used, with relatively few published data addressing the prognostic impact of alternative cut-off values. Our study aims to provide data on the prognostic value of sTg at different levels of sensitivities and specificities.

Materials and methods: We conducted a retrospective review of all adult cases of differentiated thyroid carcinoma receiving RAI ablation at our centre from 2008 to 2010. All patients had sTg measured at around 6 months post-ablation. The functional sensitivity of our assay was 0.5 ng/ml. The outcome was adverse clinical event, defined as cancer-related death, persistent macroscopic disease demonstrable on imaging (including radioisotope scan) and/or receiving further treatment for persistent or recurrent disease. A receiver operating characteristic (ROC) analysis was carried out.

Results: We identified 140 patients treated in the review period, with 106 of them suitable for further analysis. The reasons for exclusion included the presence of anti-thyroglobulin antibodies and medullary or anaplastic histological subtypes. Most (54.7%) had intermediate-risk disease as per the American Thyroid Association classification (2009). The median follow-up duration was 6.4 years; the minimum, excluding deaths, was 5.0 years. ROC analysis showed that the optimal cut-off value of sTg for predicting adverse clinical events was >1.0 ng/ml, associated with a sensitivity of 90.9%, a specificity of 81.0%, a positive predictive value of 55.6% and a negative predictive value of 97.1%.

Conclusion: Based on ROC analysis of sensitivities and specificities, our data showed that a post-ablation sTg value of 1 ng/ml is the optimal cut-off in prognostication of adverse clinical events.

Keywords: Iodine radioisotopes; receiver operating characteristic; thyroglobulin; thyroid carcinoma, non-medullary

Introduction

Differentiated thyroid carcinoma (DTC) can range from a remarkably indolent disease to a very aggressive malignancy with high mortality. Disease-specific mortality as reflected by the American Joint Committee on Cancer (AJCC) tumour, node and metastasis (TNM) staging may not fully correlate with the generally much higher risk of disease recurrence [1]. The American Thyroid Association (ATA) risk stratification system incorporates additional clinical pathological prognostic factors [2]. In the last few years, accumulating evidence suggests that the response to initial treatment, as indicated by thyroglobulin (Tg) and neck ultrasonography in a dynamic risk stratification model, correlates much better with the subsequent clinical outcome [3].

The stimulated thyroglobulin (sTg) level measured at 9–12 months post-radioactive iodine (RAI) ablation is one of the key parameters in dynamic risk stratification [4]. In the studies investigating dynamic risk stratification, the

Author for correspondence: K.C.W. Wong, Department of Clinical Oncology, Prince of Wales Hospital, 30–32 Ngan Shing Street, Shatin, Hong Kong, China.
E-mail address: kenneth.wong@cantab.net (K.C.W. Wong).

https://doi.org/10.1016/j.clon.2018.10.009

cut-off values of sTg were 1 and 10 ng/ml, with the former and the latter values indicating excellent and incomplete responses, respectively [3,5]. However, the choice of the lower cut-off value of 1 ng/ml was largely based on the functional sensitivities of the assays used, and the higher cut-off value of 10 ng/ml was chosen because it coincided with the level at which an empirical dose of RAI would be considered by most physicians [6–10].

There are relatively few published data addressing specifically the prognostic impact of alternative cut-off values of sTg. A spectrum of cut-off levels ranging from 0.3 to 10 ng/ml had been suggested by this limited number of prognostication studies. In the present study, we carried out a receiver operating characteristic (ROC) analysis to establish the sTg level with the optimal balance between sensitivity and specificity.

Patients and Methods

Eligibility

We conducted a retrospective review of all cases of DTC who had received post-surgical RAI ablation (i.e. first RAI treatment post-surgery) at our centre during a 3-year period from January 2008 to December 2010. All patients had prior total thyroidectomy, with or without central compartment or lateral neck dissection. Excluded cases were those with age at diagnosis < 18 years or medullary/anaplastic histology. Thirty-four patients were not included in the main ROC analysis because of one or more of the following reasons:

- presence of anti-Tg antibodies, as they could interfere with the measurement of Tg ($n = 18$);
- sTg was not measured within the time frame of 5–12 months post-ablation and/or inadequate stimulation of Tg as reflected by a thyroid-stimulating hormone (TSH) level of <30 mIU/l ($n = 20$) [4];
- recombinant TSH (rhTSH) was used in place of thyroxine withdrawal prior to sTg measurement ($n = 1$);
- presence of poorly differentiated histological elements ($n = 3$).

Treatment and Follow-up Protocol

RAI ablation was given under either endogenous TSH (TSH $\geq$ 30 mIU/l) or rhTSH stimulation. The activity given was 3100 MBq (about 80 mCi, in 91% of patients); those patients with large thyroid remnants were, at the discretion of the attending specialist, given a lower activity of RAI to minimise RAI-induced thyroiditis. A whole body RAI scan was carried out at 5–9 days post-ablation. Thyroxine was prescribed at a dose level to suppress the TSH, usually to $<$0.1 mIU/l initially.

All patients had sTg measured at 6 months post-ablation, under endogenous TSH stimulation (TSH $\geq$ 30 mIU/l); anti-Tg antibody levels were evaluated at the same time (please refer to the section on eligibility). Some patients' sTg was already undetectable at the time of ablation and they only had unstimulated Tg (uTg) checked at 6 months post-ablation; the post-ablation sTg was assumed to be undetectable in these patients ($n = 3$) [11,12].

Measurement of Outcomes

The main outcome measure was adverse clinical event (ACE), defined as the presence of one or more of the following:

- disease-specific mortality;
- persistent macroscopic local-regional disease or distant metastases, which were apparent on clinical examination and/or imaging (including radioisotope scan), with or without histological or cytological proof;
- receiving further RAI treatment with uptake outside the thyroid bed on the post-treatment scan, i.e. those patients with uptake confined to the thyroid bed were regarded as having incomplete ablation and not considered as having an ACE;
- receiving subsequent external radiotherapy for macroscopic recurrent disease local-regionally or at distant sites;
- undergoing further surgery;
- receiving palliative chemotherapy or targeted treatments.

Thyroglobulin and Anti-thyroglobulin Assays

Both the serum Tg and serum anti-Tg autoantibody were assayed using the Immulite 1000 analyser (Siemens AG, Munich, Germany; before August 2011, with functional sensitivity of 0.5 μg/l) or Immulite 2000 XPi analyser (Siemens; after August 2011, with functional sensitivity of 0.9 μg/l). Duplicate measurements were carried out for all serum Tg analyses. The serum Tg assay was a solid-phase, chemiluminescent immunometric assay standardised against certified reference material for human Tg (CRM457). The serum anti-Tg autoantibody assay was a solid-phase, enzyme-labelled, chemiluminescent sequential immunometric assay calibrated to World Health Organization 1st IRP 65/93. Serum samples were frozen at −20°C before sample analysis. An sTg level below the detection limit of the assay was operationally defined as zero.

Statistical Analysis

An ROC analysis was carried out on the 106 eligible-for-analysis patients, to determine the optimal sTg level that would predict subsequent ACEs with the highest sensitivity and specificity (the optimal cut-off). In other words, the optimal cut-off was determined by identifying the point on the ROC curve that was furthest away from chance. This would translate into identifying the point furthest away

from the diagonal line on the ROC curve. This analysis was repeated for each ATA risk group. Positive predictive (PPV) and negative predictive values (NPV), using the optimal cut-off, were computed. Event-free survival, from the time of measurement of post-ablation sTg, was estimated with the use of Kaplan–Meier methods (with MedCalc version 13). Uni- and multivariate analyses with logistic regression were carried out using the optimal cut-off as one of the variables in the prognostication of ACEs (with IBM SPSS Statistics version 21).

Ethics

The study protocol was reviewed and approved by the ethics committee of our institution, i.e. the Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee.

Results

Demographic and Clinical Characteristics of Patients

We identified 140 patients who had received RAI for remnant ablation during the 3-year period from 2008 to 2010; among them 106 patients fulfilled the criteria for ROC analysis. Please refer to Table 1 for further information on the demographic/clinical characteristics of the patients included in the ROC analysis.

In total, 34 patients were excluded from the main ROC analysis, as mentioned above. These 34 patients were found to have similar baseline characteristics to the 106 patients included in the ROC analysis (data not shown).

Treatment Characteristics

Table S1 shows the treatments received by our patients. The median activity of RAI given for remnant ablation was 3090 MBq. 55.7% had some form of lymph node dissection in addition to total thyroidectomy. Twenty-four patients (22.6%) had two or more RAI treatments, but among them only 11 patients had extrathyroidal uptake (i.e. pathological uptake) outside the thyroid bed on the second or subsequent post-treatment whole body scan(s); these 11 cases were regarded as having ACEs. Four patients (3.8%) had further surgery, whereas 20 had external radiotherapy; among them, 19 had radiotherapy purely because of adverse factors on the initial histological report and they were not considered as having ACEs in this study. Note that the decisions to offer adjuvant radiotherapy in these 19 patients were made before RAI and were based on factors such as microscopically positive resection margins. In other words, whether to give radiotherapy or not was independent of the outcomes of RAI ablation and, hence, we have not considered these cases as having an ACE. Only one patient received chemotherapy or targeted treatments – he was given one cycle of paclitaxel and carboplatin chemotherapy for palliation (the patient declined further chemotherapy after one cycle).

Table 1
Demographic and clinical characteristics of 106 differentiated thyroid carcinoma patients included in the receiver operating characteristic (ROC) analysis

	Patients included in ROC analysis ($n = 106$)	
	No.	%
Age (mean, SD)	46.61 (13.42)	
Age (median, range)	47.56 (20.42–80.21)	
Gender		
Female	86	81.1
Male	20	18.9
Histology		
Papillary	97	91.5
Non-papillary	9	8.5
Follicular	7	6.6
Poorly differentiated	0	0.0
Mixed histology	2	1.9
Others	0	0.0
T Stage		
T1 and T2	51	48.1
T1a	14	13.2
T1b	21	19.8
T2	16	15.1
T3 and T4	55	51.9
T3	42	39.6
T4a	13	12.3
T4b	0	0.0
N Stage		
N0	48.0	45.3
N1a	19.0	17.9
N1b	39.0	36.8
Nx	0.0	0.0
M Stage		
M0	93	87.7
M1	13	12.3
AJCC prognostic groups		
Stage I and II	55	51.9
Stage I	47	44.3
Stage II	8	7.5
Stage III and IV	51	48.1
Stage III	21	19.8
Stage IVA	23	21.7
Stage IVC	7	6.6
ATA risk		
Low risk	23	21.7
Intermediate risk	58	54.7
High risk	25	23.6

Tumour, node and metastasis (TNM) staging and stage grouping are according to the seventh edition of the American Joint Committee on Cancer (AJCC) Cancer Staging Manual.
SD, standard deviation; ATA, American Thyroid Association.

Clinical Outcomes and Adverse Clinical Events

All alive patients had a minimum follow-up of 5.0 years, except one patient who was lost to follow-up after emigration at 9 months post-diagnosis; the median duration of follow-up for the eligible patients was 6.4 years, with only one disease-specific death (Table S2). 9.4% of patients were alive with evidence of disease at the last follow-up, with 8.5%

and 4.7% of patients having local-regional disease and distant metastases, respectively. Note that the clinical outcomes of the excluded patients were not statistically different from those of the eligible patients (data not shown).

Table 2 and Figure S1 show the proportion of patients developing ACEs. There were 22 ACEs (20.8%) in total, among them two patients had an ACE despite an undetectable post-ablation sTg. It is also noteworthy that three of those 22 ACEs occurred before measurement of the post-ablation sTg, with two patients having M1 disease to begin with and they never achieved complete remission.

Receiver Operating Characteristic Analysis and Optimal Cut-off Values

Post-ablation sTg was measured for 103 patients, at a median of 5.9 months (range: 4.9–11.2 months) after RAI ablation. The median TSH at the time of sTg measurement was 79.9 mIU/l (range: 31.7 to >200 mIU/l). Three patients had undetectable sTg at the time of ablation, and their post-ablation uTg (measured at 4.9–7.7 months post-ablation) was also undetectable; their post-ablation sTg was, therefore, assumed to be undetectable.

We carried out ROC analysis for the whole group of eligible patients ($n = 106$), as well as separately for each ATA risk category (except for the low-risk patients as there was only one ACE observed). As shown in Figure 1, an ROC curve was plotted to determine the optimal sTg cut-off for all patients included in the ROC analysis. The area under the curve was 0.89 (95% confidence interval 0.82–0.94; $P < 0.0001$); the optimal cut-off was >1 ng/ml with sensitivity of 90.9% (95% confidence interval 70.8–98.9%) and specificity of 81.0% (95% confidence interval 70.9–88.7%). For the intermediate-risk group ($n = 58$), the optimal cut-off was still 1 ng/ml; the associated sensitivity and specificity were 85.7% (95% confidence interval 57.2–98.2%) and 79.6% (95% confidence interval 64.7–90.2%), respectively (Figure 2). Figure 3 shows the ROC curve for high-risk patients ($n = 25$); the optimal cut-off for these patients was >0.8 ng/ml, giving a sensitivity of 100% (95% confidence interval 59.0–100%) and a specificity of 72.2% (95% confidence interval 46.5–90.3%).

Prognostic Value of the Optimal Cut-off and Alternative Cut-off Levels

After establishing that the optimal cut-off was >1 ng/ml, we evaluated the PPV and NPV associated with this cut-off, as well as those for alternative cut-off levels (2, 5 and 10 ng/ml). The PPV was low with our optimal cut-off (only 55.6% for the whole ROC group), but, as expected, it became much higher when a higher cut-off value was applied. On the other hand, the NPV was excellent at 97.1% for the whole group, and it gradually dropped to 87.9% with a higher cut-off of 10 ng/ml (Table 3).

As the optimal cut-off was found to be >1 ng/ml, we carried out Kaplan–Meier estimation of the event-free survival from the time of measurement of post-ablation sTg, comparing patients having post-ablation sTg ≤1 ng/ml with those >1 ng/ml (Figure 4). The event-free survival of the two groups were significantly different; *P*-value of the log-rank test was < 0.0001.

Univariate and Multivariate Analysis

As the optimal cut-off derived from our analysis was >1 ng/ml, we dichotomised this variable to >1 ng/ml and ≤1 ng/ml, and incorporated this into our univariate analysis (Table 4). Four factors were found to be significantly associated with adverse outcomes, namely N1b disease, involvement of >5 lymph nodes, incomplete resection (R1 and R2 resection) and post-ablation sTg >1 ng/ml. These significant factors were further evaluated in multivariate analysis; only sTg level remained as a significant adverse factor ($P < 0.0001$). In other words, post-ablation sTg >1 ng/ml is an independent predictor of poor event-free survival.

Discussion

Owing to the desire to minimise exposure to radioisotopes, patients are increasingly monitored with Tg. Many previous studies have addressed the prognostic value of post-ablation sTg, interpreted with or without other imaging investigations. Many of these studies were carried out in the late 1990s and early 2000s, when the functional sensitivities of the Tg assays were about 1–3 ng/ml. The cut-off sTg values for these studies were very often the functional sensitivities of the assays used, i.e. an undetectable sTg was taken as a negative sTg. However, there were relatively few studies trying to establish the most prognostic cut-off value for post-ablation sTg with ROC analysis, which we believe is one of the best statistical methods for this purpose. The latest ATA guidelines mentiond that 'further studies are needed to refine the precise Tg value cut-off used to define what should be an excellent response to therapy in patients

Table 2
Clinical outcomes in terms of adverse clinical events (ACEs) for 106 eligible patients, stratified by American Thyroid Association risk group

	Absence of ACEs		Presence of ACEs		*P*-value
	No.	%	No.	%	
Patients included in ROC analysis ($n = 106$)					0.08
Low risk	22	26.2	1	4.5	
Intermediate risk	44	52.4	14	63.6	
High risk	18	21.4	7	31.8	

ROC, receiver operating characteristic.

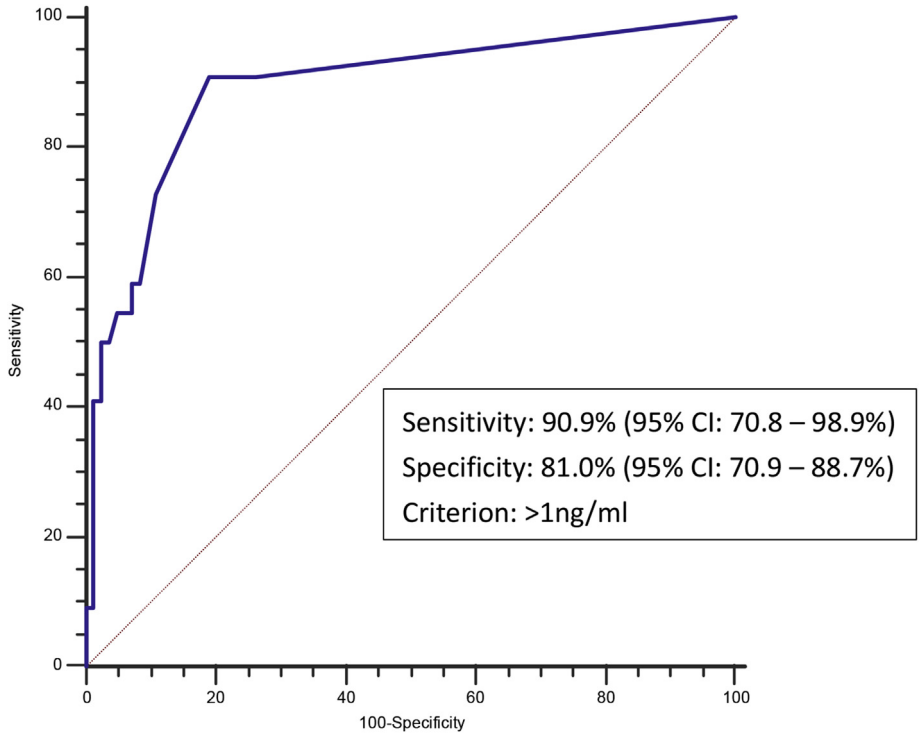

Fig 1. Receiver operating characteristic (ROC) curve for 106 differentiated thyroid carcinoma (DTC) patients included in the analysis; post-ablation stimulated thyroglobulin (sTg) level as a predictor of adverse clinical events.

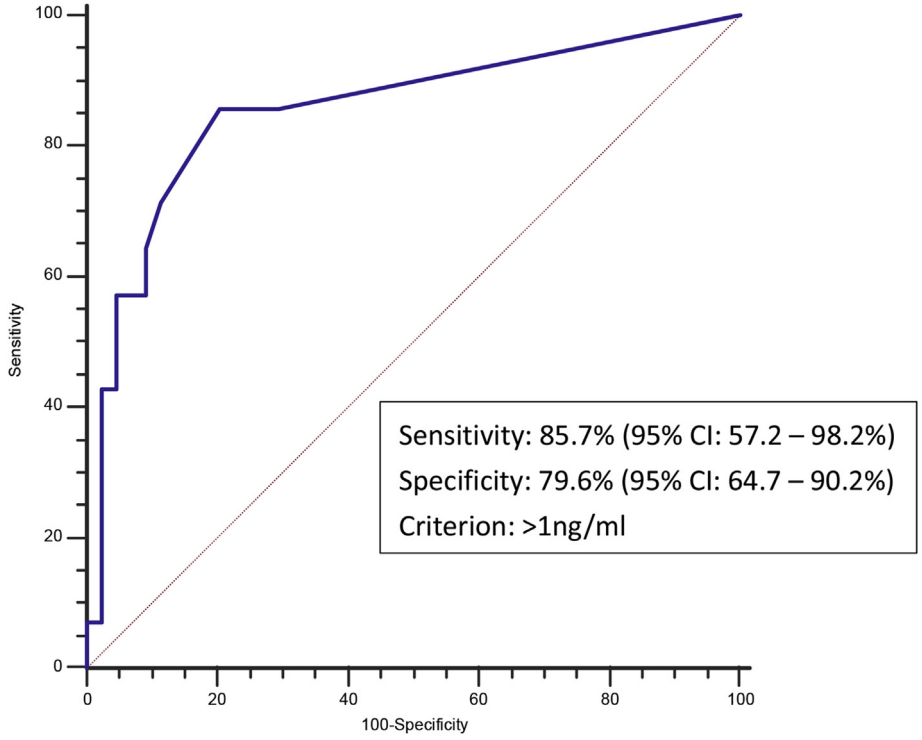

Fig 2. Receiver operating characteristic curve (ROC) curve for 58 intermediate-risk patients; post-ablation stimulated thyroglobulin (sTg) as a predictor of adverse clinical events.

treated with total thyroidectomy with or without RAI remnant ablation' [13]. The ATA panel also made a similar comment on the upper cut-off of 10 ng/ml. It is noteworthy that the functional sensitivity of the Tg assay was 0.5 ng/ml for the vast majority of patients in our cohort, and yet the optimal cut-off was 1 ng/ml; it provided strong evidence

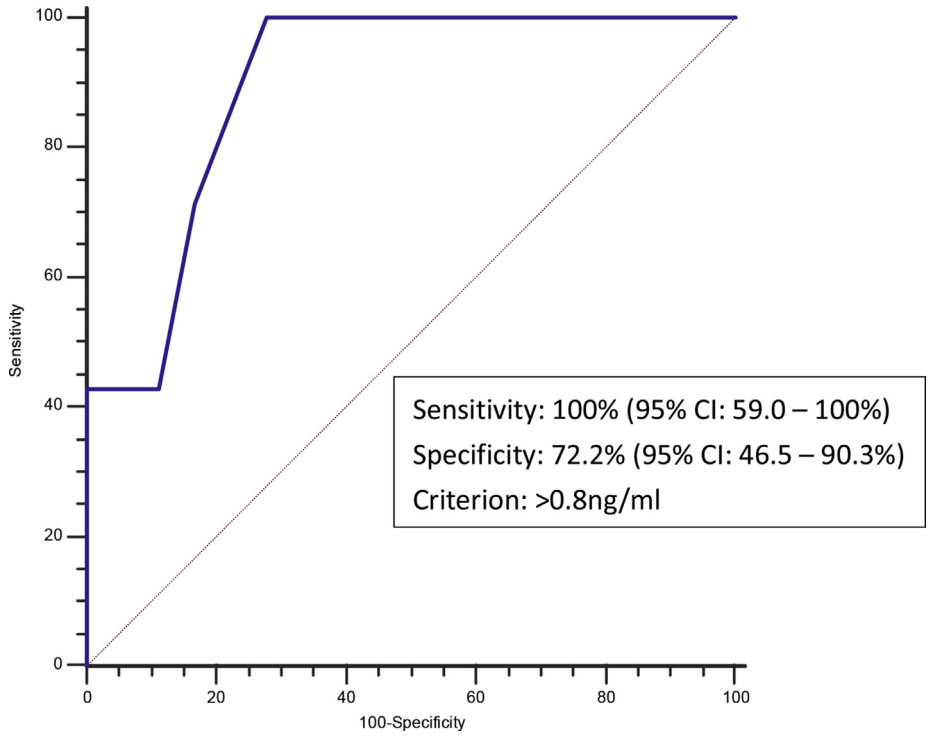

Fig 3. Receiver operating characteristic (ROC) curve for 25 high-risk patients; post-ablation stimulated thyroglobulin (sTg) as a predictor of adverse clinical events.

Table 3
The predictive values of different post-ablation stimulated thyroglobulin (sTg) cut-off levels, stratified by American Thyroid Association risk group

		sTg cut-off values (ng/ml)				
		>0.8 (Cut-off for high risk)	>1 (Optimal cut-off)	>2	>5	>10
106 patients included in the ROC analysis						
Whole group	No. with ACEs		20	16	12	11
($n = 106$)	% with ACEs		18.9	15.1	11.3	10.4
	PPV		55.6	64.0	66.7	73.3
	NPV		97.1	92.6	88.6	87.9
Low risk	No. with ACEs		1	1	1	1
($n = 23$)	% with ACEs		4.3	4.3	4.3	4.3
	PPV		25.0	50.0	50.0	100.0
	NPV		100.0	100.0	100.0	100.0
Intermediate risk	No. with ACEs		12	10	8	7
($n = 58$)	% with ACEs		20.7	17.2	13.8	12.1
	PPV		57.1	66.7	66.7	77.8
	NPV		94.6	90.7	87.0	85.7
High risk	No. with ACEs	7	7	5	3	3
($n = 25$)	% with ACEs	28.0	28.0	20.0	12.0	12.0
	PPV	58.3	58.3	62.5	75.0	60.0
	NPV	100.0	100.0	88.2	81.0	80.0

ROC, receiver operating characteristic; ACE, adverse clinical event; PPV, positive predictive value; NPV, negative predictive value.

that using an undetectable sTg level as the cut-off may not be appropriate.

We have identified seven previous studies that carried out ROC analysis to determine the optimal sTg cut-off post-ablation; the results are summarised in Table 5 [14–20]. The optimal cut-off values ranged from 0.3 to 10 ng/ml. In view of the relative sparsity of data, the significant heterogeneity in the design of previous ROC studies and the large discrepancy between results derived from the latter, we think it is worthwhile carrying out a similar analysis. In the current

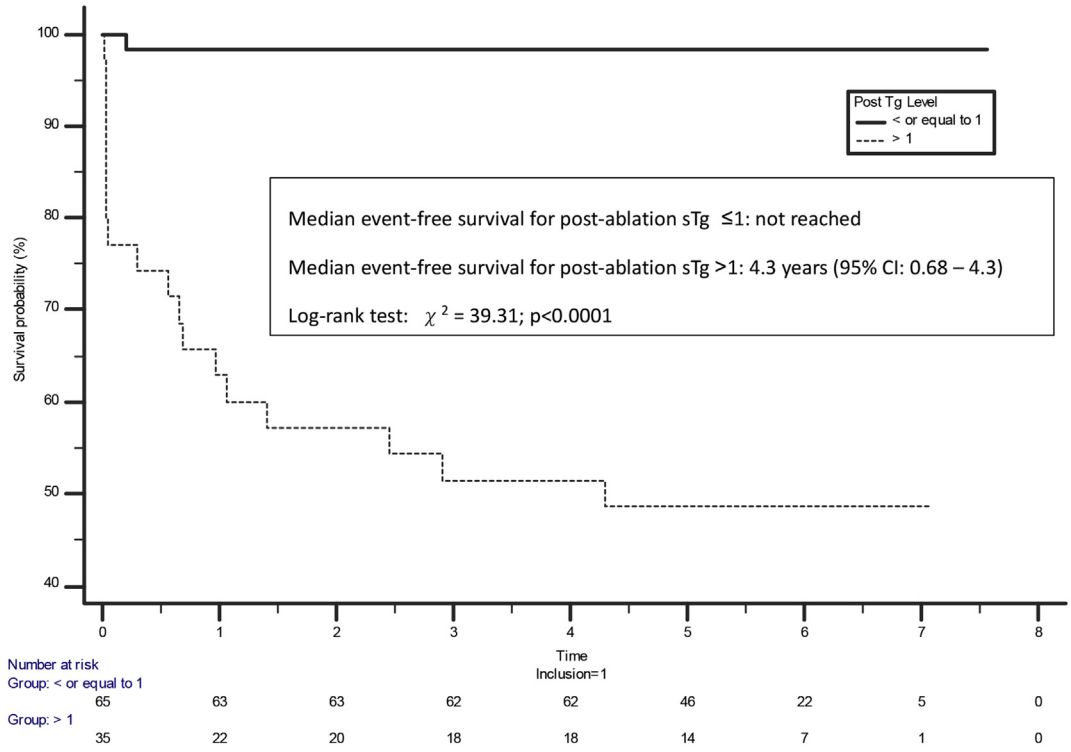

Fig 4. Kaplan–Meier estimation of the event-free survival from the time of post-ablation stimulated thyroglobulin (sTg) for 100 patients*, stratified by low (≤1 ng/ml) and high (>1 ng/ml) post-ablation sTg levels. *P*-value of log-rank test is <0.0001. *Three patients were excluded from this graph as they had been found to have persistent disease prior to the measurement of post-ablation sTg. Another three patients were excluded as they had undetectable pre-ablation sTg and hence only unstimulated Tg was checked after ablation.

study, we found that a post-ablation sTg of 1 ng/ml has the optimal balance between sensitivity and specificity, and this optimal cut-off is similar to those derived by Schlumberger *et al.* [15] (0.97 and 1.2) and Gadawska-Juszczyk and Kowalska [20] (≤1.16). In our cohort of patients ($n = 106$), the proportion of patients developing an ACE was 20.8%. This is similar to previously reported rates of having persistent/recurrent disease in DTC (10–20%). Therefore, the PPVs and NPVs in the present study may well be observed in other patient populations when this optimal cut-off is applied.

In dynamic risk stratification, the lower cut-off (i.e. the cut-off value for excellent response) should ideally have a high sensitivity and a high NPV. This is because those labelled as having an excellent response would have TSH suppression relaxed and follow-up frequency would be reduced (e.g. only having uTg measured annually) [4]. This would have obvious clinical implications if the sensitivity and NPV of the lower cut-off are not high enough. The sensitivity of our optimal cut-off of 1 ng/ml is 90.9%, which is excellent, and the NPV is very high at 97.1%, which is in keeping with many previous studies looking at the NPV of undetectable sTg at 6–18 months post-ablation. This cut-off value is the same as the lower cut-off recommended in the British Thyroid Association (BTA) and ATA guidelines, and our results could be viewed as a validation of their recommendations [4,13].

On the other hand, those labelled as having a biochemical incomplete response would have TSH suppressed to <0.1 mIU/l (BTA) or 0.1–0.5 mIU/l (ATA) and may well be given another high-activity RAI treatment. Hence, it is of vital importance that the higher cut-off should have a high specificity and, in particular, a high PPV. In our study, the PPV using a cut-off of 1 ng/ml is rather low at 55.6%. If we use a higher value of 10 ng/ml as the cut-off for biochemical incomplete response, the PPV would be much higher at 73.3%. In other words, we would only over-treat 26.7% of patients if we gave empirical therapeutic RAI to those with sTg > 10 ng/ml, and this figure would be 44.4% if a lower threshold of 1 ng/ml is used (assuming those with recurrent/persistent disease have iodine-avid disease).

We stratified our patients according to their ATA risk category and carried out ROC analysis separately for intermediate- and high-risk patients. Although the number of patients in each group was small, the most prognostic sTg value was consistent among patients with dissimilar baseline risks. Quite a lot of previous studies focused specifically on low-risk cases, such as those with undetectable (or low) uTg levels and no clinical evidence of disease after initial treatment [8,21,22]. Further studies are therefore warranted, especially for high-risk cases.

Some previous studies have shown that post-ablation sTg is an independent predictor of recurrence. Kim *et al.* [23] showed that a high (≥1 ng/ml) post-ablation sTg was an independent predictor of poor recurrence-free survival. Pelttari *et al.* [24] found that post-ablative detectable sTg concentrations (functional sensitivity was 3 ng/ml then 1

Table 4
Univariate and multivariate analyses of 106 patients included in the receiver operating characteristic analysis

	Univariate analysis			Multivariate analysis*		
	No.	Odds ratio (95% confidence interval)	*P*-value	No.	Odds ratio (95% confidence interval)	*P*-value
Age†						
<45 years	43	1.00		43	1.00	
≥45 years	63	0.49 (0.19–1.26)	0.14	62	0.44 (0.10–1.95)	0.28
Gender						
Female	86	1.00		86	1.00	
Male	20	0.62 (0.17–2.35)	0.48	19	0.18 (0.02–1.31)	0.09
T stage						
T1 and T2	51	1.00		50	1.00	
T3 and T4	55	1.14 (0.45–2.94)	0.78	55	1.13 (0.28–4.66)	0.86
N stage						
N0	48	1.00		47	1.00	
N1a	19	4.00 (0.80–19.95)	0.09	19	1.65 (0.23–12.02)	0.62
N1b	39	9.38 (2.47–35.62)	**0.001**	39	2.55 (0.37–17.70)	0.34
No. of positive lymph nodes						
≤5	89	1.00		88	1.00	
>5	17	9.17 (2.93–28.70)	**<0.0001**	17	2.04 (0.36–11.69)	0.43
M stage						
M0	93	1.00		93	1.00	
M1	13	2.79 (0.81–9.61)	0.10	12	2.68 (0.36–20.18)	0.34
Post-ablation sTg						
≤1	69	1.00		68	1.00	
>1	37	39.41 (8.38–185.30)	**<0.0001**	37	35.86 (6.05–212.45)	**<0.0001**
Resection margins						
R0	76	1.00		75	1.00	
R1 and R2	30	2.57 (1.00–7.09)	**0.049**	30	2.14 (0.48–9.51)	0.32
Angioinvasion‡						
Absent	86	1		86	1	
Present	19	2.02 (0.67–6.12)	0.21	19	2.44 (0.38–15.72)	0.35

TNM, tumour, node and metastasis; sTg, stimulated thyroglobulin (unit: ng/ml); R0, complete resection with clear margins; R1, incomplete resection with microscopic positive resection margins; R2, incomplete resection with macroscopic positive resection margins.
Bold represents Significant *P* values.

* Factors included in the multivariate analysis were: age, gender, T stage, N stage, number of positive lymph nodes, M stage, post-ablation sTg levels, completeness of resection and angioinvasion. Please note that 105 patients were included in the multivariate analysis.

† We are aware of the recent change in age cut-off for staging from 45 to 55 years in the eighth edition of the American Joint Committee on Cancer (AJCC) Cancer Staging Manual. As a result, we carried out an additional analysis using this new age cut-off of 55 years and found that the *P*-value was not significant in either the univariate or multivariate analysis (please refer to Table S3).

‡ Data for one patient missing.

Table 5
Summary of published studies using receiver operating characteristic analysis to determine the optimal cut-off of post-ablation stimulated thyroglobulin (sTg) [14–20]

Reference	sTg cut-off (ng/ml)	Method of TSH stimulation (THW versus rhTSH)	Time of sTg measurement after initial treatment	Functional sensitivity of Tg assay (ng/ml)	All patients had RAI ablation?
[14]	≥10	THW	6 months after initial surgery	0.3 then 0.05	Yes
[15]	0.97 and 1.2	Both THW and rhTSH	9–12 months after initiation of thyroxine (i.e. ablation)	0.02 and 0.11	Yes
[16]	≥2.5	rhTSH	Median of 3.3 years from initial surgery	0.5, then ≤0.3, then 0.2	Yes
[17]	≥0.3	Majority had THW	9–12 months after primary treatment	<0.1	No, 87% had RAI
[18]	>1.4	Both THW and rhTSH	9–12 months after RAI ablation	0.11	Yes
[19]	>0.6	Both THW and rhTSH	6–10 months after RAI ablation	0.4	Yes
[20]	≤1.16 (THW) and ≤1.24 (rhTSH)	Both THW and rhTSH	6–9 months post RAI ablation	n/a, but analytical sensitivity <0.2	Yes

TSH, thyroid-stimulating hormone; THW, thyroid hormone withdrawal; rhTSH, recombinant TSH; RAI, radioactive iodine.

ng/ml) and local infiltration at primary surgery were the only independent predictors of recurrence. In the current study, we also carried out multivariate analysis with post-ablation sTg >1 ng/ml as one of the prognostic factors. It turned out to be the only factor that could independently predict poor event-free survival. Although we do not think that this is an essential quality of a prognostic marker, our results provide further evidence that post-ablation sTg has a high predictive value in dynamic risk stratification.

One very important issue in the design of our study was deciding on what constitutes an ACE. In numerous previous studies, patients with sTg > 2 (or detectable sTg) were regarded as having persistent or recurrent disease. In the current study, we were trying to establish the value of post-ablation sTg in the prognostication of ACEs and using sTg itself as one of the parameters of outcome may not be appropriate.

In dynamic risk stratification, what we are really interested in, and what we are trying to predict with post-ablation sTg, is the risk of failure of initial therapy [1]. In our study, those who had external radiotherapy based on unfavourable factors on the initial histopathology report (e.g. microscopic positive resection margin) were not considered as having an ACE, as the external radiotherapy was part of the initial therapy in these cases. Similarly, those who had further RAI treatments post-ablation were not necessarily failing initial therapy. In our analysis, those who had uptake in the thyroid bed only (i.e. without pathological extra-thyroidal accumulation of RAI) on the second and subsequent high-activity RAI scans were regarded as having unsuccessful ablation initially, as the thyroid bed uptake could well be due to normal thyroid remnants (this is generally regarded as a more likely reason for the thyroid-bed uptake in these cases). It is very challenging to determine the exact nature of the thyroid bed uptake in these cases, which is a common problem encountered in studies of this kind. Eustatia-Rutten *et al.* [25] tackled this problem in their meta-analysis by performing three separate analyses. It was found that the specificity of post-ablation sTg would decrease to 0.895 from 0.947 if thyroid bed activity was considered benign, which was exactly the assumption we made in the current study. This may, to some extent, explain the relatively low specificity for the optimal cut-off in our study.

Our study had a few limitations. First, it is difficult to establish the exact nature of the thyroid bed uptake, as mentioned above. Second, we excluded patients receiving rhTSH injections from our main ROC analysis. The method of TSH stimulation in dynamic risk stratification was not clearly defined in the BTA and ATA guidelines, and there was also heterogeneity in this aspect, both between studies and within some studies. Haugen *et al.* [26] showed that sTg with thyroxine withdrawal (THW-Tg) was comparable with sTg obtained with rhTSH (rhTSH-Tg) in terms of detecting disease/tissue limited to the thyroid bed and distant metastases, provided sTg ≥ 2 was used as the cut-off. However, there is also ample evidence that rhTSH-Tg is generally lower than THW-Tg [20,27,28]. Kowalska *et al.* [28] showed that rhTSH-Tg was significantly lower than THW-Tg, and if cut-off values of 2 ng/ml and 10 ng/ml were used for THW-Tg, the corresponding values for rhTSH-Tg should be 0.6 and 2.3 ng/ml, respectively. Gadawska-Juszczyk and Kowalska [20] found, using ROC analysis, that the optimal cut-off for THW-Tg in predicting remission was 1.16 ng/ml, whereas that for rhTSH-Tg was 1.24 ng/ml. The use of rhTSH in ablation and in follow-up diagnostic studies is definitely on the rise; whether our results are applicable to these patients is uncertain. In Hong Kong, patients have to pay for their rhTSH if it is for diagnostic purposes (e.g. for measuring sTg). Hence, the vast majority of our patients would still have their post-ablation sTg evaluated under endogenous TSH stimulation.

It is noteworthy that the sTg cut-off values could be dependent on the time point at which sTg is measured. Heemstra *et al.* [14] showed that the sensitivity, specificity, PPV and NPV for the same cut-off were different at different time points, and they established that the highest diagnostic accuracies of serum Tg for tumour presence were found during TSH stimulation, at 6 months after initial therapy. The patients in our study also had post-ablation sTg measured at 6 months after RAI, and our results may not be applicable to patients whose sTg is checked at a time point very different from that of our study (e.g. 3 years after ablation). A different study would have to be carried out to determine the optimal sTg cut-off for such patients. The same is true for other patient groups who were excluded from our study, including those who have only had lobectomy, and those who have not undergone RAI ablation after total thyroidectomy. The picture gets even more complicated with the advent of second-generation ultra-sensitive assays with functional sensitivities of around 0.1–0.2 ng/ml; further studies are again needed.

Conclusion

Based on ROC analysis of sensitivities and specificities, our data showed that a post-ablation sTg value of 1 ng/ml is the optimal cut-off in prognostication of ACEs in DTC. This is consistent with the latest recommendations by the BTA and ATA on dynamic risk stratification.

Conflict of Interest

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clon.2018.10.009.

References

[1] Tuttle RM. Risk-adapted management of thyroid cancer. *Endocr Pract* 2008;14(6):764–774.
[2] American Thyroid Association (ATA) Guidelines Taskforce on Thyroid Nodules and Differentiated Thyroid CancerCooper DS,

Doherty GM, Haugen BR, Kloos RT, Lee SL, *et al*. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* 2009;19(11):1167–1214.
[3] Tuttle RM, Tala H, Shah J, Leboeuf R, Ghossein R, Gonen M, *et al*. Estimating risk of recurrence in differentiated thyroid cancer after total thyroidectomy and radioactive iodine remnant ablation: using response to therapy variables to modify the initial risk estimates predicted by the new American Thyroid Association staging system. *Thyroid* 2010; 20(12):1341–1349.
[4] Perros P, Boelaert K, Colley S, Evans C, Evans RM, Gerrard BG, *et al*, British Thyroid Association. Guidelines for the management of thyroid cancer. *Clin Endocrinol* 2014;81(Suppl. 1): 1–122.
[5] Vaisman F, Momesso D, Bulzico DA, Pessoa CH, Dias F, Corbo R, *et al*. Spontaneous remission in thyroid cancer patients after biochemical incomplete response to initial therapy. *Clin Endocrinol* 2012;77(1):132–138.
[6] Schlumberger M, Mancusi F, Baudin E, Pacini F. 131I therapy for elevated thyroglobulin levels. *Thyroid* 1997;7(2):273–276.
[7] Cailleux AF, Baudin E, Travagli JP, Ricard M, Schlumberger M. Is diagnostic iodine-131 scanning useful after total thyroid ablation for differentiated thyroid cancer? *J Clin Endocrinol Metab* 2000;85(1):175–178.
[8] Mazzaferri EL, Robbins RJ, Spencer CA, Braverman LE, Pacini F, Wartofsky L, *et al*. A consensus report of the role of serum thyroglobulin as a monitoring method for low-risk patients with papillary thyroid carcinoma. *J Clin Endocrinol Metab* 2003;88(4):1433–1441.
[9] Baudin E, Do Cao C, Cailleux AF, Leboulleux S, Travagli JP, Schlumberger M. Positive predictive value of serum thyroglobulin levels, measured during the first year of follow-up after thyroid hormone withdrawal, in thyroid cancer patients. *J Clin Endocrinol Metab* 2003;88(3):1107–1111.
[10] Kim MH, Ko SH, Bae JS, Lim DJ, Baek KH, Lee JM, *et al*. Combination of initial stimulation thyroglobulins and staging system by revised ATA guidelines can elaborately discriminate prognosis of patients with differentiated thyroid carcinoma after high-dose remnant ablation. *Clin Nucl Med* 2012; 37(11):1069–1074.
[11] Castagna MG, Brilli L, Pilli T, Montanaro A, Cipri C, Fioravanti C, *et al*. Limited value of repeat recombinant human thyrotropin (rhTSH)-stimulated thyroglobulin testing in differentiated thyroid carcinoma patients with previous negative rhTSH-stimulated thyroglobulin and undetectable basal serum thyroglobulin levels. *J Clin Endocrinol Metab* 2008;93(1):76–81.
[12] Nascimento C, Borget I, Al Ghuzlan A, Deandreis D, Chami L, Travagli JP, *et al*. Persistent disease and recurrence in differentiated thyroid cancer patients with undetectable postoperative stimulated thyroglobulin level. *Endocr Relat Cancer* 2011;18(2):R29–R40.
[13] Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, *et al*. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid* 2016;26(1):1–133.
[14] Heemstra KA, Liu YY, Stokkel M, Kievit J, Corssmit E, Pereira AM, *et al*. Serum thyroglobulin concentrations predict disease-free remission and death in differentiated thyroid carcinoma. *Clin Endocrinol* 2007;66(1):58–64.
[15] Schlumberger M, Hitzel A, Toubert ME, Corone C, Troalen F, Schlageter MH, *et al*. Comparison of seven serum thyroglobulin assays in the follow-up of papillary and follicular thyroid cancer patients. *J Clin Endocrinol Metab* 2007; 92(7):2487–2495.
[16] Kloos RT. Thyroid cancer recurrence in patients clinically free of disease with undetectable or very low serum thyroglobulin values. *J Clin Endocrinol Metab* 2010;95(12):5241–5248.
[17] Soyluk O, Boztepe H, Aral F, Alagol F, Özbey NC. Papillary thyroid carcinoma patients assessed to be at low or intermediary risk after primary treatment are at greater risk of long term recurrence if they are thyroglobulin antibody positive or do not have distinctly low thyroglobulin at initial assessment. *Thyroid* 2011;21(12):1301–1308.
[18] Brassard M, Borget I, Edet-Sanson A, Giraudet AL, Mundler O, Toubeau M, *et al*, THYRDIAG Working Group. Long-term follow-up of patients with papillary and follicular thyroid cancer: a prospective study on 715 patients. *J Clin Endocrinol Metab* 2011;96(5):1352–1359.
[19] Aldawish M, Jha N, McEwan AJ, Severin D, Ghosh S, Morrish DW. Low but measurable stimulated serum thyroglobulin levels <2 μg/L frequently predict incomplete response in differentiated thyroid cancer patients. *Endocr Res* 2014;39(4):157–163.
[20] Gadawska-Juszczyk K, Kowalska A. Comparison of the usefulness of post-ablative and post-operative thyroglobulin concentration measuring in prognostic assessment of patients with differentiated thyroid cancer. *Endokrynol Pol* 2015; 66(6):486–494.
[21] Giovanella L, Ceriani L, Ghelfo A, Keller F, Sacchi A, Maffioli M, *et al*. Thyroglobulin assay during thyroxine treatment in low-risk differentiated thyroid cancer management: comparison with recombinant human thyrotropin-stimulated assay and imaging procedures. *Clin Chem Lab Med* 2006;44(5):648–652.
[22] Kloos RT, Mazzaferri EL. A single recombinant human thyrotropin-stimulated serum thyroglobulin measurement predicts differentiated thyroid carcinoma metastases three to five years later. *J Clin Endocrinol Metab* 2005;90(9): 5047–5057.
[23] Kim JW, Roh JL, Gong G, Cho KJ, Choi SH, Nam SY, *et al*. Treatment outcomes and risk factors for recurrence after definitive surgery of locally invasive well-differentiated papillary thyroid carcinoma. *Thyroid* 2016;26(2):262–270.
[24] Pelttari H, Välimäki MJ, Löyttyniemi E, Schalin-Jäntti C. Post-ablative serum thyroglobulin is an independent predictor of recurrence in low-risk differentiated thyroid carcinoma: a 16-year follow-up study. *Eur J Endocrinol* 2010;163(5):757–763.
[25] Eustatia-Rutten CF, Smit JW, Romijn JA, van der Kleij-Corssmit EP, Pereira AM, Stokkel MP, *et al*. Diagnostic value of serum thyroglobulin measurements in the follow-up of differentiated thyroid carcinoma, a structured meta-analysis. *Clin Endocrinol* 2004;61(1):61–74.
[26] Haugen BR, Pacini F, Reiners C, Schlumberger M, Ladenson PW, Sherman SI, *et al*. A comparison of recombinant human thyrotropin and thyroid hormone withdrawal for the detection of thyroid remnant or cancer. *J Clin Endocrinol Metab* 1999;84(11):3877–3885.
[27] Pacini F, Molinaro E, Lippi F, Castagna MG, Agate L, Ceccarelli C, *et al*. Prediction of disease status by recombinant human TSH-stimulated serum Tg in the postsurgical follow-up of differentiated thyroid carcinoma. *J Clin Endocrinol Metab* 2001;86(12):5686–5690.
[28] Kowalska A, Pałyga I, Gąsior-Perczak D, Walczyk A, Trybek T, Słuszniak A, *et al*. The cut-off level of recombinant human TSH-stimulated thyroglobulin in the follow-up of patients with differentiated thyroid cancer. *PLoS One* 2015;10(7): e0133852.