Clinical Endocrinology (2017) **86**, 97–107 doi: 10.1111/cen.13140

ORIGINAL ARTICLE

Is a second recombinant human thyrotropin stimulation test useful? The value of postsurgical undetectable stimulated thyroglobulin level at the time of remnant ablation on clinical outcome

Inmaculada Prior-Sánchez*, Ana Barrera Martín*, Estefanía Moreno Ortega†, Juan A. Vallejo Casas† and María Á. Gálvez Moreno*

**Department of Endocrinology and Metabolism, Reina Sofia University Hospital, and †Department of Nuclear Medicine, Reina Sofia University Hospital, Córdoba, Spain*

Summary

Objective The management of patients with differentiated thyroid cancer (DTC) has changed in recent years, and monitoring depends on the risk of persistent/recurrent disease. The objective was to assess the prognostic value of a single stimulated thyroglobulin (Tg) measured at the time of the first radioiodine therapy (Stim-Tg1), and the utility of a second stimulated Tg measurement performed 6–12 months later (Stim-Tg2). We also examined the role of neck ultrasound (US) in the early diagnosis of recurrence.

Design This was a retrospective observational cohort study conducted in a tertiary referral hospital. Of 213 evaluated patients with DTC, 169 were finally included.

Methods Measurement of Stim-Tg1, Stim-Tg2 and neck US.

Results Stim-Tg1 was undetectable in 71 of 169 patients (42%). All of them (71/71) continued to have negative Stim-Tg2. Seventy of 71 had an excellent response to the first treatment. Sixty-eight of 71 had no evidence of disease after an average follow-up of 7·2 years. In patients with detectable Stim-Tg1 (98/169; 58%), Stim-Tg2 became negative in 40. The negative predictive value (NPV) of Stim-Tg1 was 0·96. The optimal Stim-Tg1 cut-off level for identifying persistence was 3·65 ng/ml. Recurrence was detected in 14 patients. Neck US was useful for identifying local recurrence (13/14; 92·85%).

Conclusions Stim-Tg1 is a reliable marker with a high NPV. A second stimulation test should be avoided in patients with negative Stim-Tg1. In patients with biochemical persistence, Stim-Tg2 is useful for confirming/ruling out final status. Neck US plays a valuable role in the early diagnosis of recurrence.

Correspondence: Inmaculada Prior-Sánchez, Department of Endocrinology and Metabolism, Reina Sofia University Hospital, Avenida Menéndez Pidal, s/n, 14004 Córdoba, Spain. Tel.: 693 440 573;
E-mail: priorsanchez@hotmail.com

(Received 6 April 2016; returned for revision 13 June 2016; finally revised 13 June 2016; accepted 17 June 2016)

Introduction

The classical management paradigms of patients with differentiated thyroid cancer (DTC) have changed over the last 20 years, mainly due to variations in demographic characteristics and changes in clinical presentation.[1,2] Nowadays, the incidence of DTC is steadily increasing and most newly detected cases are discovered in the asymptomatic initial stages or incidentally (e.g. during imaging studies of the neck performed for other reasons or postoperatively).[3] However, long-term follow-up is necessary in these patients because of the risk of recurrence. Recent guidelines[4–7] include an initial risk stratification based on clinical/pathological features, but a new risk assessment protocol is currently being developed, called 'ongoing risk stratification', or 'delayed risk stratification'.[8,9] This dynamic system includes response to therapy and can improve the ability to predict long-term outcomes.[9–11] It is also being considered in the new 2015 ATA guidelines.[12] Whatever the case, the management of these patients should be guided with high negative predictive value (NPV) tests that can discern between patients with an insignificant risk and those with a higher risk who need a closer follow-up. To date, reliable monitoring consists of measurement of stimulated thyroglobulin (Tg) and neck ultrasound (neck US) at 6–12 months after first treatment. At that time, most patients will appear disease-free.[4–6] The routine use of diagnostic whole-body scan (WBS) can be avoided in these patients because its sensitivity has been reported to be low.[13,14] Pre-ablative stimulated Tg levels, measured at least 3 months after surgery (Stim-Tg1), could have an important role as an early prognostic biomarker [15–19] and could be useful for selecting patients for radioactive remnant ablation (RRA).[18–20] The need to repeat stimulated Tg measurement in patients with previous stimulation thyrotropin (TSH) test negative remains controversial,

although there is some evidence in favour of avoiding a second stimulation test.[21,22] To clarify this question, we retrospectively evaluated the prognostic impact of Stim-Tg1 and assessed the clinical utility of a second stimulated Tg measurement in patients with undetectable Stim-Tg1 who had no evidence of disease. We also examined the role of neck US in the detection of recurrence.

Patients and methods

Patients

We retrospectively evaluated all patients who were diagnosed with DTC ($n = 213$) at a tertiary hospital centre in Cordoba (Spain), between January 2000 and December 2013. In all patients, initial treatment consisted of total or subsequent completion thyroidectomy, with or without lymphadenectomy, and radioactive iodine (RAI; ^{131}I) ablation therapy at a dose of 30-200 mCi (1110–7400 MBq). Mean (±SD) follow-up after surgery was 7·18 ± 3·95 years (Table 1). Patients who had positive antithyroglobulin antibodies (TgAb; $n = 13$) or microcarcinomas (tumours < 1 cm; $n = 31$) were excluded from the study. All of the diagnostic procedures were performed in accordance with the regulations of the local ethics committees and the principles of the Declaration of Helsinki.

The first stimulated Tg control (Stim-Tg1) was performed 12–16 weeks after thyroidectomy, just before RAI therapy. The second control (Stim-Tg2) was carried out 6–12 months later. To achieve TSH stimulation (TSH ≥30 mU/l), one of the following methods were employed: a) use of recombinant human thyroid-stimulating hormone (rhTSH; Thyrogen, Genzyme Therapeutics), two intramuscular injections of 0·9 mg on successive days or; b) withdrawal of L-thyroxine (L-T4) during the previous 4 weeks. Additional imaging tests consisted of neck US, postablative diagnostic ^{131}IWBS, and ^{18}FDG-PET (^{18}fluorodeoxyglucose positron emission tomography), SPECT/CT (single-photon emission computed tomography/computed tomography) or neck US along with fine-needle aspiration cytology (FNAC), performed in several patients to rule out persistent/recurrent disease. Clinical follow-up included regular physical, biochemical (serum Tg, TgAb, TSH and free T4 level measurements) and neck US examinations.

Staging was performed according to the instructions of the American Joint Committee on Cancer and the Union for International Cancer Control. Patients were also stratified in 3 groups for assessment of risk of recurrence according to the following criteria included in the revised 2009 ATA guidelines[4]: (1) low risk: no local or distant metastases, all macroscopic tumour had been resected, there was no tumour invasion of locoregional tissues or vascular invasion, the tumour did not have aggressive histology (e.g. tall cell, columnar cell carcinoma) or vascular invasion and, if ^{131}I was given, there was no ^{131}I uptake outside the thyroid bed on the post-therapeutic WBS; (2) intermediate risk: microscopic invasion of tumour into the perithyroidal soft tissues at initial surgery, cervical lymph node metastases or ^{131}I uptake outside the thyroid bed on the WBS performed after thyroid RRA or tumour with aggressive histology or vascular invasion; (3) high risk: macroscopic tumour invasion, incomplete tumour resection, distant metastases and thyroglobulinaemia disproportionate to what was seen on the post-treatment WBS. Finally, we also classified patients into four categories according their response to the initial therapy[8–12]: (1) excellent response: no clinical, biochemical (undetectable stimulated Tg) or structural evidence of disease (negative neck US); (2) biochemical incomplete response: abnormal Tg (suppressed Tg >1 ng/ml or stimulated Tg >10 ng/ml) in the absence of localizable disease (negative imaging); (3) structural incomplete response: persistent or newly identified locoregional or distant metastases, independent of Tg levels; (4) Indeterminate response: nonspecific biochemical or structural findings which could not be confidently classified as either benign or malignant. This included patients with faint uptake in thyroid bed on RAI scanning or stimulated Tg detectable (> 0·3 ng/ml but less than 10 ng/ml).

Patients with undetectable Stim-Tg2 (< 0·3 ng/ml), negative TgAb and no evidence of disease (on clinical examination, neck US and diagnostic ^{131}IWBS when performed) were defined as free of disease, whereas patients with detectable stimulated serum Tg and/or evidence of disease were classified as having persistent disease. Recurrence was defined as the reappearance of disease in a patient who in the previous control was free of disease.

Materials and methods

Serum Tg, TSH and TgAb were measured concurrently by radioimmunoassay (RIA), (Medipan Kit: 'SELco', Vitro). Tg assay had a lower detection limit of 0·3 ng/ml. TgAb were considered positive when above 150 IU/ml. All patients underwent neck US performed by specifically trained, experienced radiologists, using a high-resolution colour Doppler US system. Neck US was routinely conducted in all of the patients 6–12 months after the initial treatment and after that, every year during the subsequent follow-up, irrespective of the results of the rhTSH-Tg test. FNAC was performed whenever suspicious images were observed on US. The cytology examination was assessed by a cytologist experienced in thyroid cancer. Since 2009, Tg measurements in needle washout fluid (FNA-Tg) have also been obtained for confirming malignancy.

Statistical analysis

Descriptive quantitative data are expressed as mean ± standard deviation (SD), and qualitative data, as frequencies and percentages. To compare categorical variables, we analysed 2 × 2 or 2 × 3 contingency tables using Fisher's exact test or the chi-squared test. The Student's *t*-test was used to compare quantitative data. The following demographic and clinical variables were included in the univariate analysis: age at diagnosis (<45 or ≥45 years), sex, body mass index (BMI; kg/m^2), smoking habit, total thyroidectomy or subsequent completion, neck

Table 1. Baseline clinical characteristics of the DTC 169 patients and patients with Stim-Tg1 below 0·3 ng/ml

Clinical characteristic	Overall (n = 169)	Postsurgical stimulated Tg (Stim-Tg1)		P
		Undetectable (n = 71)	Detectable (n = 98)	
Age (years)	44·05 ± 16·20	42·86 ± 13·34	44·92 ± 18·01	0·534
<45	95 (56·2%)	42 (59·2%)	53 (54·1)	
≥45	74 (43·8%)	29 (40·8%)	45 (45·9%)	
Sex				
Female	129 (76·3%)	54 (76·1%)	75 (76·5%)	1·000
Male	40 (23·7%)	17 (23·9%)	23 (23·5%)	
BMI (Kg/m^2) (n = 161)	26·98 ± 5·70	27·34 ± 5·98	26·74 ± 5·52	0·334
<30	121 (75·2%)	47 (71·2%)	74 (77·9%)	
≥30	40 (24·8%)	19 (28·8%)	21 (22·1%)	
Other tumours				
No	118 (69·8%)	50 (70·4%)	68 (69·4%)	0·885
Yes	51 (30·2%)	21 (29·6%)	30 (30·6%)	
History of radiation exposure				
No	164 (97·0%)	70 (98·6%)	94 (95·9%)	0·400
Yes	5 (3·0%)	1 (1·4%)	4 (4·1%)	
Multinodular goitre				
No	112 (66·3%)	50 (70·4%)	62 (63·3%)	0·331
Yes	57 (33·7%)	21 (29·6%)	36 (36·7%)	
Completion thyroidectomy				
No	127 (75·1%)	51 (71·8%)	76 (77·6%)	0·396
Yes	42 (24·9%)	20 (28·2%)	22 (22·4%)	
Neck lymphadenectomy (n = 167)				
Yes	93 (55·7%)	41 (57·7%)	52 (54·2%)	0·753
No	74 (44·3%)	30 (42·3%)	44 (45·8%)	
Postsurgical complications				
No	124 (73·4%)	52 (73·2%)	72 (73·5%)	0·973
Yes	45 (26·6%)	19 (26·8%)	26 (26·5%)	
Histology				
Papillary	151 (89·3%)	64 (90·1%)	87 (88·8%)	1·000
Follicular	18 (10·7%)	7 (9·9%)	11 (11·2%)	
Multifocality				
No	96 (56·8%)	49 (69%)	47 (48%)	**0·008**
Yes	73 (43·2%)	22 (31%)	51 (52%)	
Histological aggressiveness criteria*				
No	128 (75·7%)	65 (91·5%)	63 (64·3%)	**<0·001**
Yes	41 (24·3%)	6 (8·5%)	35 (35·7%)	
pTNM stage				
Stage I	105 (62·1%)	54 (76·1%)	51 (52%)	**<0·001**
Stage II	33 (19·5%)	14 (19·7%)	19 (19·4%)	
Stage III	18 (10·7%)	3 (4·2%)	15 (15·3%)	
Stage IV	13 (7·7%)	0 (0·0%)	13 (13·3%)	
Risk stratification				
Low	117 (69·2%)	64 (90·1%)	53 (54·1%)	**<0·001**
Intermediate	24 (14·2%)	6 (8·5%)	18 (18·4%)	
High	28 (16·6%)	1 (1·4%)	27 (27·6%)	
Stim-Tg2				
Undetectable (<0·3 ng/ml)	111 (65·7%)	71 (100%)	40 (40·8%)	**<0·001**
Detectable (>0·3 ng/ml)	58 (34·3%)	0 (0%)	58 (59·2%)	
Method of the first TSH stimulation				
LT4 withdrawal	39 (23·1%)	14 (19·7%)	25 (25·5%)	0·378
rhTSH	130 (76·9%	57 (80·3%)	73 (74·5%)	
Method of the second TSH stimulation				
LT4 withdrawal	21 (12·4%)	9 (12·7%)	12 (12·2%)	0·952
rhTSH	148 (87·6%	62 (87·3%)	86 (87·7%)	

(continued)

Table 1. (continued)

Clinical characteristic	Overall (n = 169)	Postsurgical stimulated Tg (Stim-Tg1)		P
		Undetectable (n = 71)	Detectable (n = 98)	
Post-treatment WBS				
residues	157 (92·90%	71 (100%)	86 (87·8%)	**0·001**
Metastases	12 (7·10%)	0 (0%)	12 (12·2%)	
Results of first neck US				
Negative	131 (77·5%)	70 (98·6%)	61 (62·2%)	**<0·001**
Positive	38 (22·5%)	1 (1·4%)	7 (37·8%)	
Recurrence				
No	155 (91·7%)	68 (95·8%)	87 (88·8%)	0·157
Yes	14 (8·28%)	3 (4·2%)	11 (11·2%)	
Time to recurrence (years)	6·12 ± 2·19	4·53 ± 1·01	6·56 ± 2·25	0·162
Final outcome				
Disease-free	122 (72·2%)	68 (95·8%)	54 (55·1%)	**<0·001**
Persistence	47 (27·8%)	3 (4·2%)	44 (44·9%)	
Exitus				
No	158 (93·5%)	70 (98·6%)	88 (89·8%)	**0·026**
Yes	11 (6·5%)	1 (1·4%)	10 (10·2%)	
Follow-up (years)	7·18 ± 3·95	7·28 ± 4·03	7·43 ± 4·22	0·820

Bold values indicate statistically significant differences.
Data are presented as mean ±SD or frequencies (percentages). Statistical significance was obtained by χ^2-test or independent samples '*t*-' test.
*Aggressive histological criteria: presence of vascular invasion or unfavourable histology. WBS: whole-body scan.

lymphadenectomy, histology of primary tumour (papillary/follicular), multifocality, vascular invasion, aggressive histology, stage at diagnosis, risk stratification (low/intermediate/high), results of the first rhTSH stimulation test (Stim-Tg1 detectable/undetectable), results of the first neck US (negative/positive) and outcomes of postablative WBS (residues or residues plus metastases). Statistically significant variables found in the univariate analysis were entered into a multivariate logistic regression analysis to select those with independent prognostic significance on final outcome (disease-free or persistence) and to calculate the odds ratio (OR) and the 95% confidence interval (95% CI). Cook's distance was calculated to identify extreme cases. Hosmer–Lemeshow analysis was used to assess the goodness of fit of our model. We also used receiver operating characteristic (ROC) curves to determine the optimal cut-off value of Stim-Tg1 in predicting disease-free remission or disease persistence. SPSS 15·0 software was used for the statistical analysis. Observed differences were assumed to be statically significant at a level of $P \leq 0{\cdot}05$.

Results

Baseline cohort characteristics

Two hundred and thirteen consecutive patients were assessed for inclusion in this study. Forty-four were excluded (31 patients with microcarcinoma and 13 patients with positive TgAb). The clinical and histopathological characteristics of the final 169 DTC patients are summarized in Table 1. The female-to-male ratio was approximately 3:1, with a mean of 44·05 years (range:12–79 years) at the time of thyroidectomy. Mean BMI (26·98 ± 5·70 kg/m^2) was slightly higher than normal, as 24·8% of the subjects were obese (BMI > 30 kg/m^2). Only 5 patients had a history of radiation exposure and 33·7% presented previous multinodular goitre. Other tumours were associated in 51 cases, most of which were benign, for example uterine or breast tumours. Seven women had a diagnosis of breast cancer, with no temporal relation with the DTC diagnosis. Most patients underwent initial total thyroidectomy (75·1%), while the rest received completion thyroidectomy (n = 42; 24·9%). Neck lymphadenectomy was conducted in 93 patients (55·7%), depending on the suspicion of cervical lymph node metastases: central compartment neck dissection (n = 16), central and left compartment (n = 20), central and right compartment (n = 35) and radical neck dissection in 22 patients (13·2%). Postsurgical complications were found in 45 patients (26·6%): hypoparathyroidism (n = 39; permanent in 16 patients), dysphonia due to vocal cord injury (n = 11; permanent in 7 subjects). The most common histological subtype was papillary, accounting for 89·3% of all DTC. Histological aggressiveness, defined as the presence of vascular invasion or unfavourable histology (e.g. tall cell, diffuse sclerosing, solid/trabecular, insular or columnar cell variants), was found in 41 patients (24·3%). Multifocality was detected in 43·2% of cases. However, the majority of patients were AJCC pTNM Stage I or II (81·1%) with low (69·2%) or intermediate risk (14·2%) of recurrence, according to the ATA risk criteria.[4] TSH stimulation was induced in most of cases with rhTSH (76·9% in the first and 87·6% in the second control) and showed no significant differences between groups or in final disease status ($P > 0{\cdot}05$). Patients were followed for an average of 7·2 years. Minimum follow-up after surgery was

18 months. During this period, recurrence was verified in fourteen patients with a mean time to recurrence of 6·12 years (range: 3·39–10·35 years). In our series, 11 subjects died, 2 unrelated to DTC. In total, 122 individuals (72·2%) achieved disease-free status. Table 1 also illustrates the association between Stim-Tg1 (detectable or undetectable) and several clinical and prognostic variables, such as multifocality, histological aggressiveness, stage, risk of recurrence, Stim-Tg2 levels, results of the first neck US and final outcome.

Results of the first and the second TSH stimulation test (Stim-Tg1 and Stim-Tg2)

At the time of the first TSH stimulation, just before the RRA, Stim-Tg1 was undetectable in 71 of 169 patients (42%), (Fig. 1). In this group, neck US was negative in 70 of 71 patients (98·6%). The single case with positive neck US was a 43-year-old female with high-risk DTC (T4N1M0; Stage I) in whom neck US performed 12 months after thyroidectomy revealed a suspicious image. FNAC was conducted but the diagnosis was not conclusive (*indeterminate response to the first treatment*). After 3 years of follow-up, this patient continues to be closely monitored and she has not achieved the status of remission. The vast majority of patients with negative Stim-Tg1 (70/71; 98·6%) were reclassified 6–12 months after the initial RAI therapy in the group with 'excellent response to the first treatment'. An important fact which is worth highlighting is that 100% (71/71) of patients with undetectable Stim-Tg1 remained negative in the second control (undetectable Stim-Tg2). In contrast, 98 of 169 subjects (58%) had detectable Stim-Tg1. The percentage of

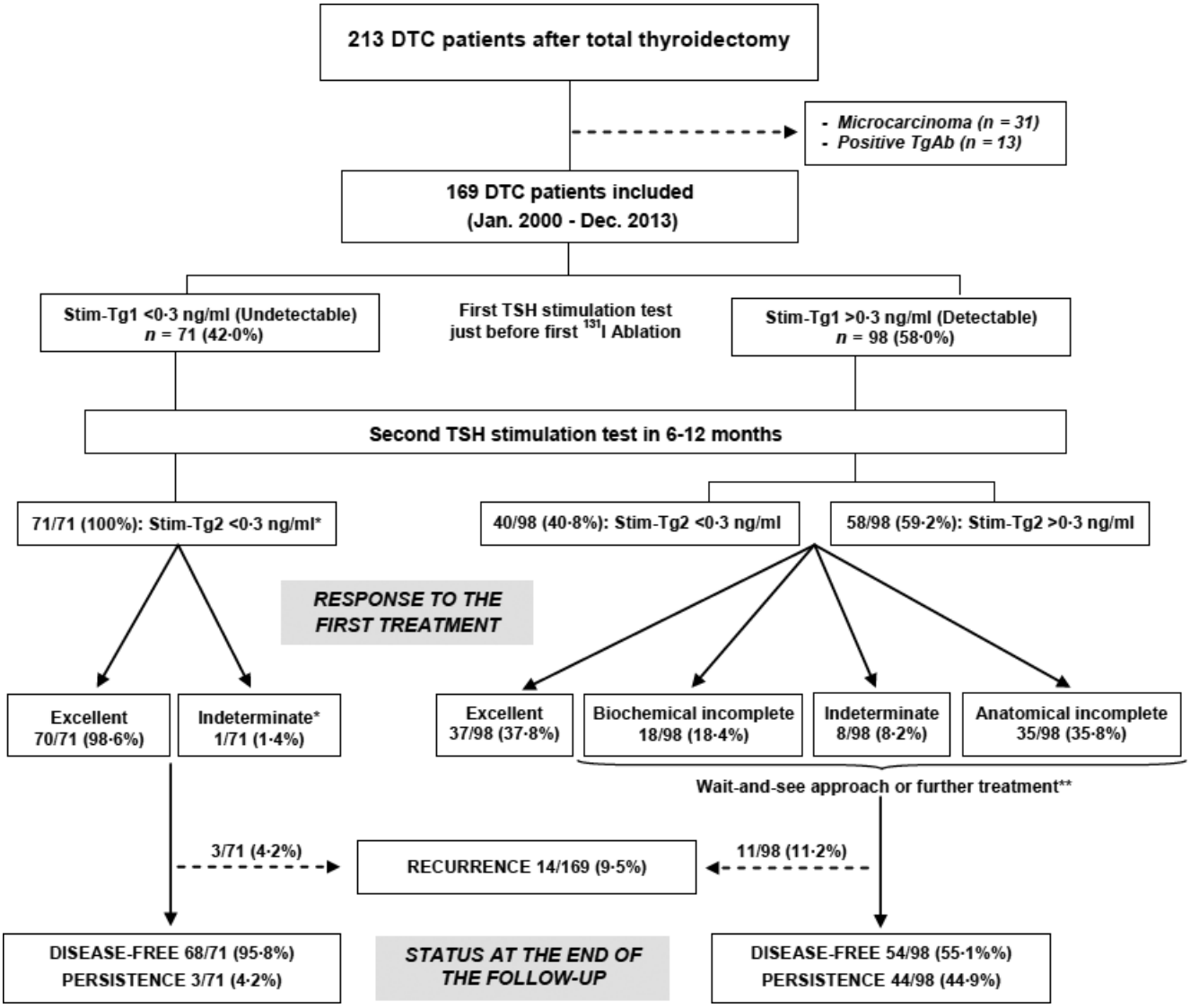

Fig. 1 Clinical outcome and follow-up of all 169 patients with DTC. *Only one patient with negative Stim-Tg1 (1·4%) had suspicious image in neck US (response indeterminate to the first treatment). FNAC was conducted to confirm malignance, but the result was not conclusive. **Further treatments were performed in patients with evidence of disease (mainly, therapeutic doses of ^{131}I or second surgeries).

patients in whom Stim-Tg2 became negative was greater in the group with lower levels of Stim-Tg1 and fell progressively in the remaining groups ($P < 0{\cdot}001$; Fig. 2).

Minimal residual 131 I uptake in the thyroid bed was detected in all patients with undetectable Stim-Tg1 (71/71). None of them had uptake outside the thyroid bed. At the second control, performed 6 to 12 months after thyroid ablation, Stim-Tg was negative in 111 patients and WBS control showed negative uptake in 98 of these patients. In 13 patients (13/111), uptake remained in the thyroid bed but no evidence of disease was found.

Patients were classified into four categories, according to response to the first treatment (thyroidectomy and RAI remnant ablation), based on Stim-Tg2 levels and the results of imaging tests, mainly neck US (Fig. 1). Overall, 107 of 169 patients (63·3%) had an excellent response (no evidence of disease by biochemical measurements and neck US). Of these subjects, 70 had undetectable Stim-Tg1 and 37 positive Stim-Tg1. A total of 10·7% (18 of 169) had an incomplete biochemical response (Stim-Tg2 > 10 ng/ml with negative imaging). In 35 patients (20·7%), anatomical disease was detected; these were included structural incomplete response group. Nine patients (5·3%) were classified with indeterminate response (Stim-Tg2 detectable but less than 10 ng/ml or nonspecific findings on imaging studies). Patients with documented disease received further treatments. Forty patients repeated ^{131}I therapy and 10 had second surgeries. In subjects with negative imaging test, an ongoing approach was applied. With respect to long-term outcomes, most patients with negative Stim-Tg1 had no evidence of disease at the end of follow-up (95·8%). Three patients in this subgroup had recurrence detected on neck US. After lymph node surgery, 2 had no evidence of disease. The third patient died at the age of 84 because of tumour dedifferentiation (Table 2). In contrast, patients with detectable Stim-Tg1 achieved disease-free status in a minor percentage (55·1%). In this group, recurrence was documented in 11 individuals.

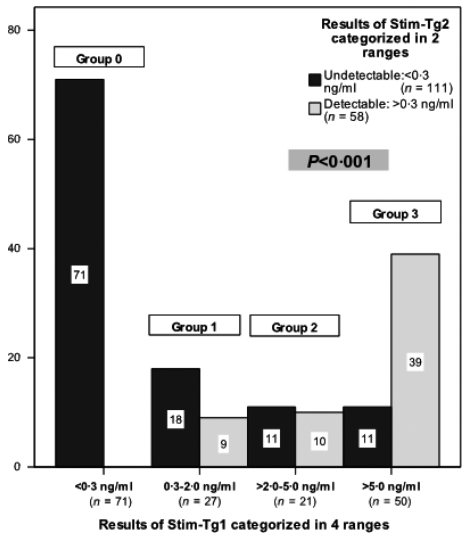

Fig. 2 Association between Stim-Tg1 and Stim-Tg2. Patients were classified into 4 groups based on the Stim-Tg1 levels. Group 0 ($n = 71$): Stim-Tg1 ≤ 0·3 ng/ml (undetectable); Group 1 ($n = 27$): Stim-Tg1 > 0·3-2 ng/ml; Group 2 ($n = 21$): Stim-Tg1 > 2-5 ng/ml; and Group 3 ($n = 50$): Stim-Tg1 > 5 ng/ml. Seventy-one of 71 patients (100%) with undetectable Stim-Tg1 remained undetectable in the second control (Stim-Tg2 < 0·3 ng/ml); χ^2-test; $P < 0{\cdot}001$.

Utility of neck US in the diagnosis of recurrence

As outlined in Table 2, 14 patients had recurrence during follow-up. In all but one (92·85%), neck US was useful for detecting local disease. Basal-Tg became detectable in only 1 female patient (number 4), 6·93 years after initial surgery, with no evidence of disease on neck US. PET was also performed and no evidence of malignant disease was found. Because the imaging tests were negative, she received no specific treatment at recurrence and a close clinical monitoring approach was adopted. Her latest basal-Tg was 0·79 ng/ml which, after the rhTSH test, increased to 1·49 ng/ml, so she has been classified as biochemical persistence. In 4 cases (numbers 1, 2, 5 and 12), recurrence was detected only by neck US, while basal-Tg remained negative. In all of these patients, FNAC was conducted confirming malignancy. In 3 subjects, Tg in washout needle aspiration biopsy was measured and the results were conclusive. Only patient number 1 is waiting for subsequent neck surgery; she has not yet been classified as in remission. The others are considered disease-free after lymph node surgery and subsequent assessment. PET was performed in 3 patients, showing distant metastases: multiple lung metastases in patient number 3; lung metastases in number 9; and mediastinal, lung and bone metastases in number 13. At the end of follow-up, none had achieved clinical remission.

ROC analysis and cut-off points of Stim-Tg1

Figure 3 shows the ROC curve for Stim-Tg1. The area under the curve (AUC) was 0·867 (95% CI: 0·804-0·930; $P < 0{\cdot}001$), indicating the excellent discrimination value of Stim-Tg1. The NPV of Stim-Tg1 was 0·96 (95% CI: 0·90–1·00), which means that 96% of patients who had undetectable Stim-Tg1 were disease-free. However, when the positive predictive value (PPV) was calculated, it decreased to 45%. The optimal cut-off point of Stim-Tg1 to maximize the rate of true positive (sensitivity) and minimize the rate of false positives (1 - specificity) was 3·65 ng/ml (sensitivity: 0·787, 95% CI: 0·706–0·867; specificity: 0·803, 95% CI: 0·725–0·881). Using this cut-off value, the positive likelihood ratio (LR) rises to 3·69.

Prognostic factors in the univariate and multivariate analysis

The prognostic impact of relevant demographic, histopathological and clinical variables [age at diagnosis (< 45 or ≥ 45 years), BMI, tobacco use, type of surgery (total thyroidectomy or completion thyroidectomy), neck lymphadenectomy (yes/no), histotype, multifocality, aggressive histological variant,

Table 2. Characteristics of fourteen DTC patients with recurrence

Patient ID/ Age/Sex/ Histology	TNM Stage	ATA Risk	Previous Therapies	Stim-Tg1 (ng/ml)	Detection method	Tumour location	Time to recurrence	Therapy at recurrence	DF
1 64 ♀ Papillary	T2NxM0	LR	TT+ RAI	<0·30	US+FNAC*	Local	3·59 year	Pending reoperation	No
2 55 ♀ Papillary	T2N1M0	LR	TT+ Central and right lateral LN dissection + RAI	<0·30	US+FNAC*	Local	4·39 year	LN surgery	Yes
3 77 ♂ Papillary	T2N1M0	IR	TT+ Central and right lateral LN dissection + RAI	<0·30	Positive-Tg+ US+PET	Local and distant	5·60 year	RAI	No (Exitus*)
4 16 ♀ Papillary	T2N1M0	LR	TT+ Central LN dissection + RAI	4·00	Positive-Tg	Neg. imaging	6·93 year	Clinical following	No
5 27 ♂ Papillary	T1N1M0	LR	CT. + Central LN dissection + RAI	11·40	US+FNAC*	Local	8·24 year	LN surgery	Yes
6 28 ♀ Papillary	T2NxM0	LR	TT+ RAI	15·20	Positive-Tg + US+FNAC	Local	4·74 year	LN surgery	Yes
7 34 ♂ Papillary	T3N1M0	HR	TT+ Bilateral LN dissection + RAI	167·57	Positive-Tg + US+FNAC	Local	3·39 year	LN surgery + RAI	No
8 33 ♀ Papillary	T2N1M0	IR	TT+ Central and left lateral LN dissection + RAI	77·90	Positive-Tg + US+FNAC*	Local	10·35 year	Pending reoperation	No
9 20 ♀ Papillary	T2N1M0	IR	TT+ Bilateral LN dissection + RAI	3·70	Positive-Tg + US+FNAC* + PET	Local and distant	10·13 year	Surgery, rosiglitazone and sorafenib	No
10 44 ♀ Papillary	T2NxM0	IR	TT+ RAI	9·90	Positive-Tg + US+FNAC	Local	4·84 year	LN surgery	Yes
11 61 ♀ Papillary	T1N1M0	IR	CT. + Central and right lateral LN dissection + RAI	5·09	Positive-Tg + US+FNAC*	Local	5·13 year	LN surgery and PEI	No
12 27 ♀ Papillary	T2N1M0	IR	TT+ Central and right lateral LN dissection + RAI	2·69	US+FNAC	Local	5·90 year	LN surgery	Yes
13 58 ♀ Papillary	T4N1M0	HR	TT+ Central and right lateral LN dissection + RAI	4·70	Positive-Tg + US+PET	Local and distant	7·09 year	External radiotherapy	No
14 65 ♀ Papillary	T4N0M0	HR	CT+ Central and left lateral LN dissection + RAI	0·45	Positive-Tg + US+FNAC	Local	5·42 year	LN surgery	No

ID, number of patient's identification; DF, disease-free, at the end of the follow-up; LR, low risk; IR: intermediate risk; HR, high risk; TT, total thyroidectomy; CT, completion thyroidectomy; LN, lymph nodes; RAI, radioactive Iodine; US, neck ultrasound; FNAC, fine-needle aspiration cytology; FNAC*, fine-needle aspiration cytology and measurement of Tg in washout needle aspiration; PET, 18fluorodeoxyglucose-positron emission tomography; PEI, percutaneous ethanol injection.

*The single case of disease's progression with postsurgical negative thyroglobulin (Stim-Tg1 < 0·3 ng/ml) occurred in a 77-year-old male because of primary tumour dedifferentiation. He had an aggressive subtype and he presented a local and distant recurrence after 6 years of follow-up. PET showed multiple lung metastases and he died at the age of 84.

ATA initial risk stratification (low, intermediate or high), pTNM Stage, existence of metastases, recurrence, Stim-Tg1 (undetectable or detectable), results of the first neck US and postsurgical WBS] was studied by univariate and multivariate analysis (Table 3). One patient had to be excluded from the multivariate analysis because his Cook's distance was greater than 1 (1·31). This subject had an extreme influence on our final model, preventing proper interpretation. As highlighted in Table 3, only the initial stratification risk (OR: 8·63, 95% CI: 2·61–28·52; $P < 0·001$), the results of neck US (OR: 4·54, 95% CI: 1·15–17·88; $P = 0·031$) and the first stimulated Tg values divided in two groups (Stim-Tg, undetectable/detectable) (OR: 5·61, 95% CI: 1·10–28·74; $P = 0·038$) were found to be independent prognostic variables on final outcome (disease-free/persistence).

Discussion

In a retrospective cohort of 169 DTC patients stratified as low, intermediate and high risk on the basis of pathological features,

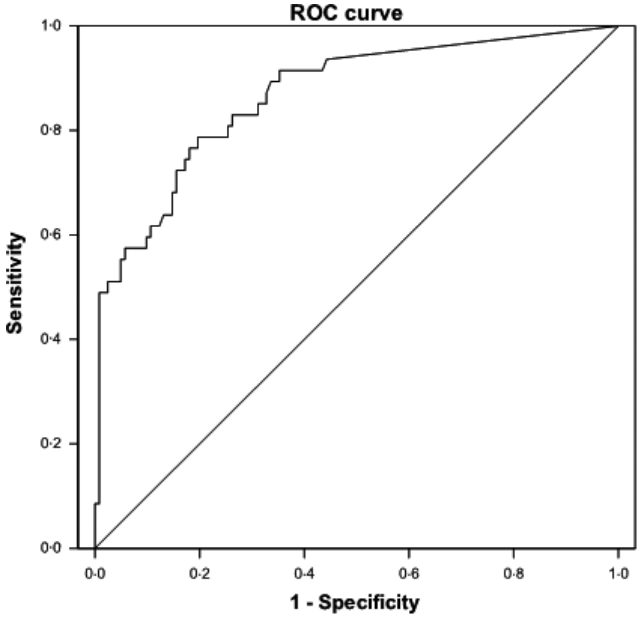

Fig. 3 ROC curve for Stim-Tg1. Area under the curve (AUC): 0·867; asymptotic 95% confidence interval: 0·804–0·930; $P < 0·001$.

we observed that 71 (42%) had undetectable postsurgical stimulated Tg, just at the time of the first RRA (Stim-Tg1). In all of these patients, stimulated Tg remained negative at the second stimulation control performed 6–12 months later (Stim-Tg2) and furthermore, 98·6% (70/71) had an '*excellent response to the first treatment*', according to the newly proposed dynamic risk assessment.[12]

Several robust retrospective reports[13,22–24] have shown that a single stimulated Tg measurement associated with a neck US 1 year after initial treatment is sufficient for selecting most patients who are free of disease (nearly 80%). Judging from our results (high correlation between Stim-Tg1 and Stim-Tg2), final outcome could be predicted earlier in patients with undetectable Stim-Tg1 and negative US findings. We monitored 71 patients over a mean follow-up of 7·2 years, concluding that all but three finally achieved complete remission. This means that Stim-Tg1 had a high NPV (0·96), in line with other reports[15–19] and a recent meta-analysis.[25] However, this meta-analysis concluded that the PPV of pre-ablation stimulated Tg is poor (almost 47%). Our PPV percentage is comparable (45%). Therefore, the value of Stim-Tg1 may only be as a negative predictor of persistent disease.

It is worth stressing that a significant number of patients of our cohort had undetectable Stim-Tg1 (42%). Similar results were obtained by other authors.[17–19] Nascimento *et al.*[17] reported postoperative stimulated Tg <1·0 ng/ml prior to RRA in 30·7%. Vaisman *et al.*[18] reported this finding in 56·7% of their series. Orlov *et al.*[19] reported a greater proportion (65%) in a prospective study of 129 low- and intermediate-risk DTC patients. These outcomes may be directly related to changes in clinical and demographic characteristics of newly diagnosed cases and also to advances in US techniques.[1–3] Furthermore, it may reflect the experience and qualification of the surgeons. However, the current trend is to be a bit more conservatism during primary thyroid surgery. The new 2015 ATA guideline[12] recommends that the initial surgical procedure can be a unilateral procedure (lobectomy), for patients with thyroid cancer >1 cm and <4 cm without extrathyroidal extension, and without clinical evidence of any lymph node metastases (recommendation 35); but this operative approach may make the follow-up more difficult, because either basal or stimulated Tg cannot be employed as reliable monitoring tool. In this regard, clinicians loosed the help of this splendid marker in assessing the persistence of disease after surgery and detecting future disease recurrence during the follow-up.

On the other hand, in our series, the post-treatment WBS revealed slight uptake in the thyroid bed but showed no pathological uptake in any patient, indicating that there would be little need for a WBS especially in the presence of a Stim-Tg <0·3 ng/ml. Several studies[13,14] have concluded that diagnostic 131 I WBS may be avoided in patients with undetectable Stim-Tg levels because it does not correlate with results of Tg determination and its accuracy has turned out to be low. In addition, the new ATA guideline[12] includes that 'After the first post-treatment WBS performed following RAI remnant ablation or adjuvant therapy, low-risk and intermediate-risk patients (lower risk features) with undetectable Tg on thyroid hormone with negative antithyroglobulin antibodies and a negative US (excellent response to therapy) do not require routine diagnostic WBS during follow-up' (Recommendation 66).

With regard to clinical monitoring, recurrence in our study cohort was only found in fourteen patients (8·28%), mainly limited to locoregional disease. All cases but one (13 of 14) were detected by routine neck US. In nine, basal serum Tg also became detectable; therefore, in 4 patients, recurrence was revealed by neck US only, while serum Tg remained undetectable. These facts highlight the valuable role played by US in diagnosing lymph node metastases and local recurrence, which was also demonstrated in recent reports.[19–22,26] Conversely, in spite of experienced surgeons and use of 131 I remnant ablation, this series demonstrated that cancer recurrences may occur, even in patients catalogued as ATA low/intermediate risk. In our cohort, most of the patients who suffered from recurrent/persistent disease were low ($n = 5$) or intermediate ($n = 6$) risk. One possible explanation could be that this initial stratification is mainly based on pathological features without taking account the response to therapy. We considered that the proposed dynamic system 'delayed risk stratification' can improve the diagnostic accuracy to predict long-term status.[9–11] However, because of recurrence may happen, patients with DTC must have ongoing follow-up, although the intensity of the follow-up might be adjusted periodically based on biochemical or clinical outcomes.

The present study had several limitations: firstly, those inherent to retrospective nature. Secondly, the time to follow-up is relatively short (7·2 years on average) and recurrence has been reported up to 20 years after initial therapy. Thirdly, several patients were lost to follow-up during surveillance. Seven moved to other locations and continued monitoring in other hospitals. Nevertheless, all dropouts had a minimum observation period of

Table 3. Multivariate analysis and prognostic factors on final outcome

	Disease-free 122 (72·2%)	Persistence 47 (27·8%)				
	n (%)		OR (95%CI)*	*P*	Adjusted OR** (95%CI)	*P*
Age (years)						
<45	75 (78·9%)	20 (21·1%)	2·126 (1·072–4·213)	*0·031*		NS
≥45	47 (63·5%)	27 (36·5%)				
Sex						
Female	96 (74·4%)	33 (25·6%)	1·629 (0·758–3·499)	0·211		NS
Male	26 (65%)	14 (35%)				
BMI (Kg/m^2)						
Min–Max	16·55–43·50	18·82–47·26	1·079 (1·016–1·145)	*0·013*		NS
Mean (SD)	26·25 (5·09)	28·82 (6·72)				
Tobacco (*n* = 166)						
No	90 (74·4%)	31 (25·6%)	1·584 (0·760–3·302)	0·220		NS
Yes	29 (64·4%)	16 (35·6%)				
Total thyroidectomy						
Yes	86 (67·7%)	41 (32·3%)	0·346 (0·135–0·886)	0·027		NS
No	36 (85·7%)	6 (14·3%)				
Neck lymphadenectomy (*n* = 167)						
Yes	60 (64·5%)	33 (35·5%)	0·346 (0·163–0·733)	0·006		NS
No	62 (83·8%)	12 (16·2%)				
Histotype						
Papillary	108 (71·5%)	43 (28·5%)	0·711 (0·222–2·282)	0·566		NS
Follicular	14 (77·8%)	4 (22·2%)				
Multifocality						
No	78 (81·3%)	18 18·8%)	2·922 (1·457–5·862)	*0·003*		NS
Yes	44 (60·3%)	29 (39·7%)				
Vascular invasion						
No	117 (81·3%)	27 (18·8%)	21·667 (6·845–68·578)	*<0·001*		
Yes	5 (20·0%)	20 (80·0%)				
Aggressive variant						
No	117 (78·5%)	32 (21·5%)	10·875 (3·674–32·189)	*<0·001*		NS
Yes	5 (25%)	15 (75·0%)				
Risk stratification						
(Reference) Low	109 (93·2%)	8 (6·8%)	13·625 (4·650–39·922)	<0·001	8·628 (2·610–28·518)	<0·001
Intermediate	12 (50·0%)	12 (50·0%)				
High	1 (3·6%)	27 (96·4%)				
pTNM Stage						
I–II	114 (83·2%)	23 (16·8%)	8·14 (5·890–36·881)	*<0·001*		NS
III–IV	8 (25%)	24 (75%)				
Metastases						
No	121 (83·4%)	21 (16·6%)		0·998		NS
Yes	1 (4·2%)	23 (95·8%)				
Recurrence						
No	117 (75·5%)	38 (24·5%)	5·495 (1·735–17·405)	*0·004*		NS
Yes	5 (35·7%)	9 (64·3%)				
Stim-Tg1						
Undetectable	68 (95·8%)	3 (4·2%)	18·818 (5·537–63·957)	*<0·001*	5·614 (1·097–28·739)	0·038
Detectable	54 (55·1%)	44 (44·9%)				
Neck US						
Negative	114 (87·0%)	17 (13·0%)	28·739 (10·919–75·641)	*<0·001*	4·534 (1·150–17·883)	0·031
Positive	8 (21·1%)	30 (78·9%)				
Postsurgical WBS						
Residues	121 (77·1%)	36 (22·9%)		0·999		NS
Metastases	1 (8·3%)	11 (91·7%)				

BMI, body mass index; Tg, thyroglobulin; US, ultrasound; WBS, whole-body scan; NS, not significant.
*Crude odd ratio obtained by univariate logistic regression.
**Adjusted odds ratio by multivariate logistic regression. Likelihood ratio test: 76·498; $P < 0·001$; GL: 4.
Goodness of fit test: Hosmer–Lemeshow analysis C = 0·188; GL = 3; $P = 0·980$. Area under the curve ROC: 0·945 (95% CI = 0·904–0·987); $P < 0·001$.
Italic values indicate statistically significant differences.

3 years (average of 5·28 years), and none of them showed evidence of disease at the end of their follow-up. Moreover, in our hospital, DTC clinical management protocol is in place, so all patients received the same follow-up, providing homogeneity and consistency among the study population. We also consider that our study represents an interesting and useful summary of more than a decade of experience with patients with thyroid cancer.

In conclusion, our results suggest that undetectable Stim-Tg1 determined at least 3 months after total thyroidectomy has a prognostic value for predicting final outcome. We consider that in patients with undetectable Stim-Tg1, a second stimulation test lacks clinical utility and cost-effectiveness, as long as basal-Tg remains undetectable and no evidence of disease is found in routine neck US. Notwithstanding, in patients who have biochemical (detectable Stim-Tg1) or clinical evidence of disease in the imaging scan, Stim-Tg2 is useful for confirming or ruling out disease persistence. Nevertheless, this proposal requires a total thyroidectomy as the initial surgical strategy.

Acknowledgement

We gratefully acknowledge the assistance of Statistics Department and the translation service.

Declaration of interest

No competing financial interests exist.

Funding

This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

References

1 Durante, C., Costante, G. & Filetti, S. (2013) Differentiated thyroid carcinoma: defining new paradigms for postoperative management. *Endocrine Related Cancer*, **20**, 141–154.
2 Pacini, F. & Castagna, M.G. (2012) New insight in the follow-up strategies of differentiated thyroid cancer. *Journal of Endocrinological Investigation*, **35**, 36–39.
3 Hall, S.F., Walker, H., Siemens, R. *et al.* (2009) Increasing detection and increasing incidence in thyroid cancer. *World Journal of Surgery*, **33**, 2567–2571.
4 Cooper, D.S., Doherty, G.M., Haugen, B.R. *et al.* (2009) Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated 428 thyroid cancer. *Thyroid*, **19**, 1167–1214.
5 Pacini, F., Schlumberger, M., Dralle, H. *et al.* (2006) WiersingaW, European Thyroid Cancer Taskforce. European consensus for the management of patients with differentiated thyroid carcinoma of the follicular epithelium. *European Journal of Endocrinology*, **154**, 787–803.
6 Pacini, F., Castagna, M.G. & Brilli, L. (2012) Pentheroudakis G; ESMO Guidelines Working Group. Thyroid cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, **23**, 110–119.
7 Perros, P., Boelaert, K., Colley, S. *et al.* British Thyroid Association. (2014) Guidelines for the management of thyroid cancer. *Clinical Endocrinology (Oxford)*, **81**, 1–122.
8 Tuttle, R.M., Tala, H., Shah, J. *et al.* (2010) 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*, **443**, 1341–1349.
9 Castagna, M.G., Maino, F., Cipri, C. *et al.* (2011) Delayed risk stratification, to include the response to initial treatment (surgery and radioiodine ablation), has better outcome predictivity in differentiated thyroid cancer patients. *European Journal of Endocrinology*, **165**, 441–446.
10 Hong, C.M., Lee, W.K., Jeong, S.Y. *et al.* (2014) Superiority of delayed risk stratification in differentiated thyroid cancer after total thyroidectomy and radioactive iodine ablation. *Nuclear Medicine Communications*, **35**, 1119–1126.
11 Jeon, M.J., Kim, W.G., Park, W.R. *et al.* (2013) Modified dynamic risk stratification for predicting recurrence using the response to initial therapy in patients with differentiated thyroid carcinoma. *European Journal of Endocrinology*, **170**, 23–30.
12 Haugen, B.R., Alexander, E.K., Bible, K.C. *et al.* (2016) 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*, **26**, 1–133.
13 Pacini, F., Capezzone, M., Elisei, R. *et al.* (2002) Diagnostic 131-iodine whole-body scan may be avoided in thyroid cancer patients who have undetectable stimulated serum Tg levels after initial treatment. *Journal of Clinical Endocrinology and Metabolism*, **87**, 1499–1501.
14 Torlontano, M., Crocetti, U., D'Aloiso, L. *et al.* (2003) Serum thyroglobulin and 131I whole body scan after recombinant human TSH stimulation in the follow-up of low-risk patients with differentiated thyroid cancer. *European Journal of Endocrinology*, **148**, 19–24.
15 Kim, T.Y., Kim, W.B., Kim, E.S. *et al.* (2005) Serum thyroglobulin levels at the time of 131I remnant ablation just after thyroidectomy are useful for early prediction of clinical recurrence in low-risk patients with differentiated thyroid carcinoma. *Journal of Clinical Endocrinology and Metabolism*, **90**, 1440–1445.
16 Heemstra, K.A., Liu, Y.Y., Stokkel, M. *et al.* (2007) Serum thyroglobulin concentrations predict disease-free remission and death in differentiated thyroid carcinoma. *Clinical Endocrinology (Oxford)*, **66**, 58–64.
17 Nascimento, C., Borget, I., Al Ghuzlan, A. *et al.* (2011) Persistent disease and recurrence in differentiated thyroid cancer patients with undetectable postoperative stimulated thyroglobulin level. *Endocrine Related Cancer*, **18**, 29–40.
18 Vaisman, A., Orlov, S., Yip, J. *et al.* (2010) Application of post-surgical stimulated thyroglobulin for radioiodine remnant ablation selection in low-risk papillary thyroid carcinoma. *Head and Neck*, **32**, 689–698.
19 Orlov, S., Salari, F., Kashat, L. *et al.* (2015) Post-operative stimulated thyroglobulin and neck ultrasound as personalized criteria for risk stratification and radioactive iodine selection in low- and intermediate-risk papillary thyroid cancer. *Endocrine*, **50**, 130–137.
20 Rosario, P.W., Xavier, A.C. & Calsolari, M.R. (2011) Value of postoperative thyroglobulin and ultrasonography for the

indication of ablation and 131I activity in patients with thyroid cancer and low risk of recurrence. *Thyroid*, **21**, 49–53.

21 Castagna, M.G., Brilli, L., Pilli, T. *et al.* (2008) 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. *Journal of Clinical Endocrinology and Metabolism*, **93**, 76–81.

22 Crocetti, U., Durante, C., Attard, M. *et al.* (2008) Predictive value of recombinant human TSH stimulation and neck ultrasonography in differentiated thyroid cancer patients. *Thyroid*, **18**, 1049–1053.

23 Toubeau, M., Touzery, C., Arveux, P. *et al.* (2004) Predictive value for disease progression of serum thyroglobulin levels measured in the postoperative period and after (131)I ablation therapy in patients with differentiated thyroid cancer. *Journal of Nuclear Medicine*, **45**, 988–994.

24 Baudin, E., Do Cao, C., Cailleux, A.F. *et al.* (2003) Positive predictive value of serum thyroglobulin levels, measured during the first year of follow-up after thyroid hormone withdrawal, in thyroid cancer patients. *Journal of Clinical Endocrinology and Metabolism*, **88**, 1107–1111.

25 Webb, R.C., Howard, R.S., Stojadinovic, A. *et al.* (2012) The utility of serum thyroglobulin measurement at the time of remnant ablation for predicting disease free status in patients with differentiated thyroid cancer: a metaanalysis involving 3947 patients. *Journal of Clinical Endocrinology and Metabolism*, **97**, 2754–2763.

26 Chindris, A.M., Diehl, N.N., Crook, J.E. *et al.* (2012) Undetectable sensitive serum thyroglobulin (<0.1 ng/ml) in 163 patients with follicular cell-derived thyroid cancer: results of rhTSH stimulation and neck ultrasonography and long-term biochemical and clinical follow-up. *Journal of Clinical Endocrinology and Metabolism*, **97**, 2714–2723.