

RESEARCH

Improvement in neck ultrasound report quality following the implementation of European Thyroid Association guidelines for postoperative cervical ultrasound for thyroid cancer follow-up, a prospective population study

Jiahui Wu¹, Xunyang Hu², Paula Seal^{3,4}, Parthiv Amin⁴, Brendan Diederichs^{4,5} and Ralf Paschke^{6,7}

¹Arnie Charbonneau Cancer Institute, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

²Department of Medicine, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

³EFW Radiology, Calgary, Alberta, Canada

⁴Department of Radiology, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

⁵Mayfair Radiology, Calgary, Alberta, Canada

⁶Section of Endocrinology, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

⁷Departments of Medicine, Oncology, Pathology and Laboratory Medicine, Biochemistry and Molecular Biology, and Arnie Charbonneau Cancer Institute, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

Correspondence should be addressed to R Paschke Email ralf.paschke@ucalgary.ca

Abstract

Objective: The aim of this study was to prospectively evaluate the quality of postoperative neck ultrasound (POU) for thyroid cancer patients after implementing European Thyroid Association (ETA) guideline-based POU assessment.

Methods: Our analysis involved 672 differentiated thyroid cancer patients. POU report quality was compared between the implementation radiology group (IRG), which implemented ETA guideline-based assessment in 2018, and all non-implementation radiology groups (NIRG). Differences in POU quality were evaluated before and after the implementation of guideline-based assessment. Additionally, we evaluated the ability of serum thyroglobulin (Tg) level <0.2 ng/mL or between 0.21 and 0.99 ng/mL and normal POU lesion status at 1-year follow-up to predict the absence of persistent disease or relapse at 3-year follow-up.

Results: IRG had significantly higher mean utility scores for POU reports of abnormal thyroid bed nodules compared to NIRG ($P < 0.001$). IRG's POU reports for suspicious nodules and lymph nodes were considered sufficient in 94% and 85% of cases, respectively, compared to 45% and 68% for NIRG. For patients with normal US lesion status and Tg <0.2 ng/mL or Tg 0.21–0.99 ng/mL at 1-year follow-up, the negative predictive values were 96% for both.

Conclusions: Implementation of 2013 ETA POU-reporting guidelines allowed for the provision of high-quality POU reports, which may lead to increased accuracy in assessing the response to treatment and in estimating the risk of recurrence of thyroid cancer and likely reduce unnecessary repeat POU or FNA.

Key Words

- ▶ thyroid
- ▶ thyroid cancer
- ▶ ultrasound
- ▶ ETA guideline

Introduction

Current postoperative follow-up (FU) and dynamic response to treatment (RTT) assessment of thyroid cancer primarily depends on postoperative neck ultrasound (POU) and thyroglobulin (Tg) determination (1). This FU strategy is based on the progressive improvement of sensitivity and specificity of baseline Tg determinations and progressive improvement of POU assessment. The latter began with the prospective assessment of ultrasound criteria for malignancy of cervical neck lymph nodes and the retrospective identification of suspicious ultrasound features and the prognostic importance of size stability vs growth for small thyroid bed nodules (2, 3). The retrospective review of neck ultrasounds for 90 intermediate recurrence risk (RR) thyroid cancer patients with a median FU of 10 years further demonstrated false-positive ultrasound abnormalities in 57% and structural recurrence in only 10% (4). Based on these findings, the authors suggested surveillance neck ultrasound not more frequent than 3–5 years in American Thyroid Association (ATA) intermediate RR patients who have achieved a non-stimulated Tg of <1 ng/mL without suspicious findings on the initial neck ultrasound. Similar results were reported by Jeon *et al.* and Grani *et al.* (5, 6).

In 2013, the European Thyroid Association (ETA) released POU guidelines for patients with thyroid cancer, which defined POU criteria for lymph node and thyroid bed nodule assessment (7). A retrospective assessment of the ETA guideline's ability to predict the growth and persistence of POU findings in patients with or without neck dissection and/or radioiodine-remnant ablation (RAI) found significantly lower rates of growth and persistence for ETA indeterminate lesions as compared to ETA suspicious lesions (8). With a POU reporting database comprising 3163 thyroid bed lesions, Frates *et al.* identified punctate echogenic foci to be associated with malignancy irrespective of postoperative RAI, and thyroid bed lesions smaller than 6 mm without punctate lesions to be associated with minimal malignancy risk (9).

The challenges for postoperative FU and dynamic RTT assessment with POU and Tg determination were further increased by the RR-adapted treatment de-escalation guideline update in 2016. The ATA management guidelines for adult patients with thyroid nodules and thyroid cancer includes the use of total thyroidectomy (TTX), with or without RAI, and lobectomy for the treatment of thyroid cancers

(1). However, the ATA guideline Tg criteria for the assessment of RTT and the ETA guideline criteria for POU assessment do not differentiate between the patients with or without postoperative RAI. Moreover, outcomes for lobectomy are dependent on the patient selection, and the significance of postoperative Tg determinations after lobectomy remains controversial and most likely limited (10), thus increasing the importance of POU, which is performed with varied assessment strategies.

Based on our January 2018 implementation of ETA guideline-based POU assessment with implementation radiology group (IRG), we prospectively evaluated the quality of POU in our health-care region for thyroid cancer patients after the implementation of lobectomy criteria for the treatment of thyroid cancer and implementation of standardized unanimous assessment of ATA RR. We also stratified initial recommendation for or against postoperative RAI for all new thyroid cancer patients (11) and assessed the differences in the quality of POU before and after the implementation of ETA guidelines. Furthermore, we investigated the utility of undetectable (<0.2 ng/mL) or low detectable (0.21–0.99 ng/mL) serum Tg measurements at 1-year FU in predicting the absence of persistent disease or relapse at 3-year FU, which aims to further improve RTT assessment based on the implementation of ATA guideline-based treatment de-escalation (11).

Methods

Alberta Health Services is a comprehensive, integrated, single-payer public provincial health-care system with centralized laboratory, pathology, surgery, endocrinology, and oncology services. It has a single electronic medical record (EMR) system, for over 4 million inhabitants of the Calgary and Southern Alberta Healthcare regions.

In April 2017, the University of Calgary Division of Endocrinology implemented the 2015 ATA guidelines (11). Within 3 months of histologic diagnosis, thyroid cancer patients were prospectively assessed according to the 2015 ATA guidelines for their initial RR, TNM (tumor (T), nodes (N), and metastases (M)) cancer stage, and RR-dependant indication for RAI treatment and then followed for the long-term with RTT assessments. Additionally, further strategies include the introduction of ETA guideline-based POU assessment with IRG, standardized synoptic reporting

of all surgery and histopathology, and defining and adopting a detailed lobectomy proposal in collaboration with Calgary thyroid surgeons and endocrinologists to provide the criteria for offering lobectomy for the definitive treatment of thyroid cancer (11).

Study subjects

This study was approved by Health Research Ethics Board of Alberta (HREBA) – Cancer Committee (CC) (Ethics ID: HREBA.CC-16-0956). We reviewed 805 patients taken from our prospective REDCap Calgary thyroid cancer database, written consent was obtained from each patient after a full explanation of the purpose and nature of all procedures used. These patients were diagnosed with thyroid cancer and underwent thyroid

surgery between April 1, 2017, and March 1, 2023. After excluding 10 patients, whose surgeries were done in other regions, 17 patients with non-invasive follicular thyroid neoplasm with papillary-like features (NIFTP), 8 with medullary thyroid cancer, 2 with anaplastic thyroid cancer, 2 whose tumors were not resected, and 1 struma ovarii, we were left with 775 patients with differentiated thyroid cancer (DTC). We only included patients with at least one POU at their 1-year FU visit. Thus, after further excluding 69 patients with no POU because of short FU, 9 patients without available POU reports, 8 without post-RAI ultrasound, and 7 lost to FU, we included the remaining 672 DTC patients in our analysis and followed them until their last FU (Fig. 1). About 87% of 672 patients underwent thyroid surgery by 1 of 11 high-volume thyroid surgeons.

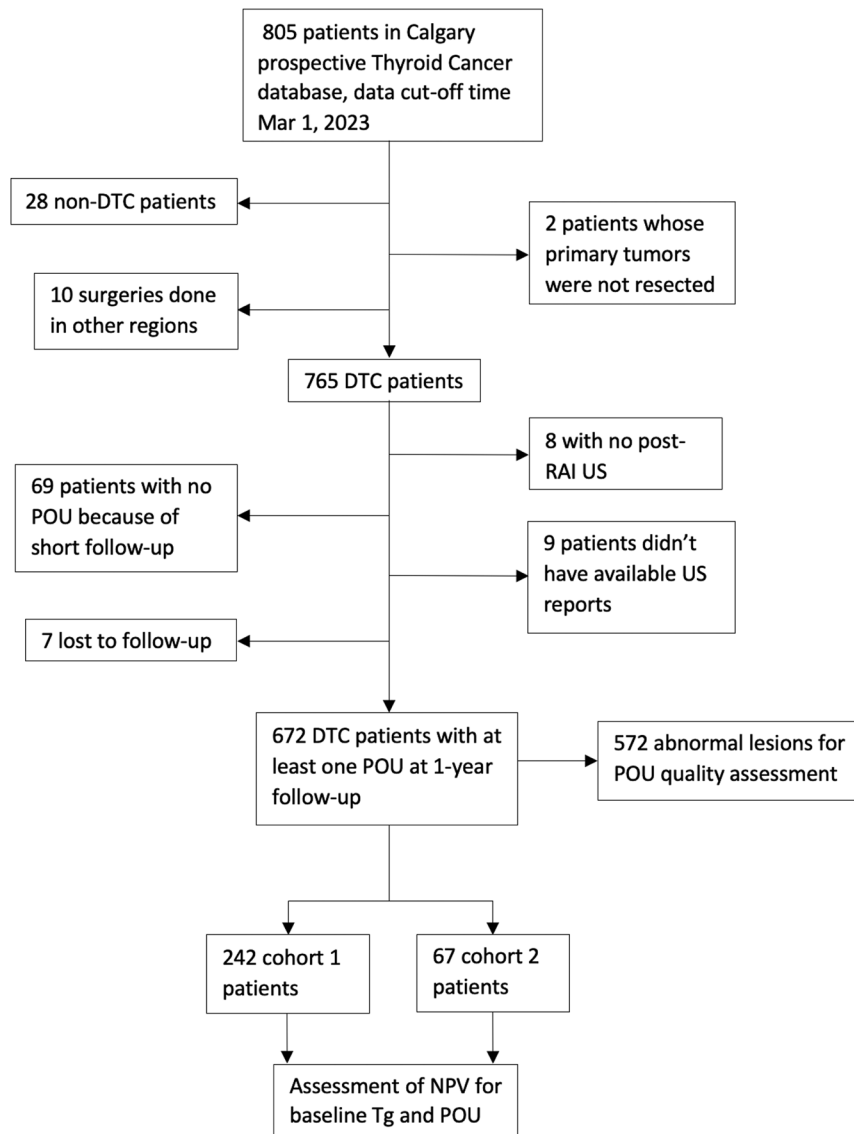


Figure 1

Patient flow chart. DTC, differentiated thyroid cancer; NPV, negative predictive value; POU, postoperative neck ultrasound; RAI, radioiodine remnant ablation; Tg, thyroglobulin; US, ultrasound.

As per ATA guidelines, the initial FU assessment occurred between 6 and 12 months after surgery and was classified as the 1-year FU, and the 3-year FU visit occurred between 30 and 36 months and was classified as the 3-year FU. During FU assessment at 1-year and 3-year post-treatment, RTT for patients who received TTX plus RAI was classified according to the 2015 ATA guidelines for adult patients with thyroid nodules and DTCs as previously reported (12). In brief, for patients with TTX only or lobectomy, the ATA guidelines do not give specific criteria for RTT. We therefore followed the RTT classification criteria reported by Momesso *et al.* (12) and the criteria are summarized in Supplementary Table 1 (see section on [supplementary materials](#) given at the end of this article). Patients who underwent TTX plus RAI vs TTX alone vs lobectomy were categorized as excellent RTT, indeterminate RTT, biochemical incomplete RTT, or structural incomplete RTT, depending on their biochemical markers and FU ultrasound.

POU evaluation

Starting from January 2018, IRG implemented the ETA guideline-based malignancy risk stratification system for all nodules reported, which categorizes nodules as ETA normal, ETA indeterminate, or ETA suspicious. IRG implemented a predetermined list of terms in a dropdown menu and required fields for each nodule characteristic and the malignancy risk classification. The purpose of these changes was to standardize and accelerate the reporting process in accordance with the ETA guideline-based POU assessment. All non-implementation radiology groups (NIRGs) continued with varied assessment strategies without ETA guideline implementation. Lesions that were not risk stratified according to the ETA guidelines were retrospectively reclassified with an ETA risk classification based on the POU report.

For IRG, the participating radiologists had between 4 and >20 years of experience in assessing postoperative thyroid cancer patients. At the time of implementation, all involved radiologists in IRG had a 2-h refresher course on the specifics of the implemented program, as well as access to written resources and live consultation with an expert for difficult cases. In the highest POU quality non-implementation radiology group (NIRG-HP), the participating radiologists had between 1 and >20 years of experience in assessing postoperative thyroid cancer patients, while the experience of respective

radiologists in other non-implementation radiology groups (NIRG-OT) is unknown.

At each FU, ultrasound lesion status was classified as normal (US-N) if all visible lesions were considered normal or no lesion mentioned; indeterminate (US-I) if there were no suspicious lesion but at least one lesion classified as indeterminate; or suspicious (US-S) when at least one lesion was classified as suspicious according to the ETA guideline (7). Serum Tg level status was classified as thyroglobulin levels-negative (Tg-N), thyroglobulin levels-indeterminate (Tg-I), or thyroglobulin levels-suspicious (Tg-S). The definition for each serum Tg level status is provided in Supplementary Table 2.

We followed the utility score (UtS) used by Hu *et al.* and Symonds *et al.* for thyroid bed nodules to evaluate the quality of neck US reports in describing thyroid bed nodules (13, 14). For each POU report, the UtS was calculated based on the number of thyroid bed lesion characteristics included in the report that were outlined in the 2013 ETA guidelines. The POU report received 1 point for every characteristic given. The UtS of POU reports ranged from 0 to 7. The characteristics are outlined in Supplementary Table 3. The UtS was evaluated for all abnormal thyroid bed nodules recorded in all POU reports in each FU. Unlike thyroid bed nodules, POU reports that identify abnormal lymph nodes typically only provide information about abnormal features, mostly suspicious features. As a result, even if the UtS is low, the report can still be considered adequate for abnormal lymph nodes. Therefore, we introduced an additional POU quality assessment criterion based on the sufficiency rate, and the quality of POU reports describing lymph nodes was evaluated by the sufficiency rate rather than UtS. A sufficient POU report is defined as including both indeterminate or suspicious lesion characteristics and the lesion classification following the abnormal lesion classification rules summarized in accordance with the ETA guidelines (Supplementary material).

Baseline Tg and POU

We further focused on two specific cohorts within our DTC patient population to evaluate the predictive ability of baseline Tg and POU: cohort 1 included those who had US-N along with Tg-N at 1-year FU and had FU at 3 years, cohort 2 included those who had US-N and Tg-I at 1-year FU and had FU at 3 years (Fig. 1). We then assessed the ultrasound status of these patients at 3-year

FU and determined the negative predictive value of their 1-year FU investigations (NPV, defined as the percentage of patients with US-N at 3-year FU, in concordance with previously published definitions) (6).

All statistical analyses were performed with the R statistical software package (15). Continuous variables were expressed as medians and ranges, while nominal variables were expressed as frequency counts and percentages. A *P*-value of <0.05 was considered statistically significant.

Results

How did the implementation of ETA POU guidelines impact POU report quality in our health-care region?

POU UtS

The assessment of POU quality involved 672 patients with DTC, of whom 299 were treated with TTX and RAI, 253 with TTX only, and 120 with lobectomy. A total of 2212 POU reports were available for the 672 patients, 1600 by IRG and 612 by NIRG, identifying 362 indeterminate lesions (270 identified by IRG and 92 by NIRG, prevalence of 17% and 15% respectively) and 210 suspicious lesions (171 identified by IRG and 39 by NIRG, prevalence of 11% and 6% respectively, Fig. 1).

The UtS of POU reports on indeterminate or suspicious thyroid bed nodules by IRG was significantly higher than NIRG (4.84 vs 3.62 for indeterminate nodules, *P* < 0.0001; 5.24 vs 3.95 for suspicious nodules, *P* < 0.001, Table 1). After the implementation of the

Table 1 Mean UtS for indeterminate/suspicious thyroid bed nodules.

Mean UtS	IRG (n = 236)	NIRG (n = 83)	P
TTX & RAI	146	30	
Indeterminate nodules	4.65	3.1	<0.0001
Suspicious nodules	5.26	3.2	<0.0005
TTX only	40	24	
Indeterminate nodules	4.06	3.14	<0.05
Suspicious nodules	4.29	5 ^a	
Lobectomy	50	29	
Indeterminate nodules	5.97	4.62	<0.0005
Suspicious nodules	5.69	4.63	<0.05
All treatment types	236	83	<0.0001
Indeterminate nodules	4.84	3.62	<0.0001
Suspicious nodules	5.24	3.95	<0.001

^aTwo nodules in this subgroup.

IRG, implementation radiology group; NIRG, all non-implementation radiology groups; RAI, radioiodine remnant ablation; TTX, total thyroidectomy; UtS, utility score.

Table 2 Mean UtS for indeterminate/suspicious thyroid bed nodules: pre-ETA and post-ETA.

Mean UtS	IRG	NIRG	P-value
Pre-ETA	3.88 (95% CI 2.76, 5.00)	2.96 (95% CI 2.60, 3.31)	<0.05
Post-ETA	4.98 (95% CI 4.79, 5.16)	3.81 (95% CI 3.51, 4.10)	<0.0001
P-value	<0.05	<0.05	

ETA, European Thyroid Association; IRG, implementation radiology group; NIRG, all non-implementation radiology groups; UtS, utility score.

ETA guidelines by IRG, the mean UtS increased for both IRG and NIRG compared to their own pre-implementation US reports (4.98 vs 3.88 for IRG, *P* < 0.05; 3.81 vs 2.96 for NIRG, *P* < 0.05; Table 2). IRG continues to have a significantly higher mean UtS than NIRG after the implementation (4.98 vs 3.81, *P* < 0.0001, Table 2). The difference is visualized by the UtS distribution demonstrated in Fig. 2. When comparing the effect sizes of the difference between the two groups using Cohen's *d* values, we observed that the effect size was larger for the post-ETA measurements (0.82) compared to the pre-ETA measurements (0.60). This indicates that the difference in mean UtSs between the two groups is more significant after the ETA implementation than before.

POU sufficiency rate

For POU reports describing indeterminate or suspicious thyroid bed nodules, 85% and 94% of those done by IRG were classified as sufficient, while only 27% and 45% of

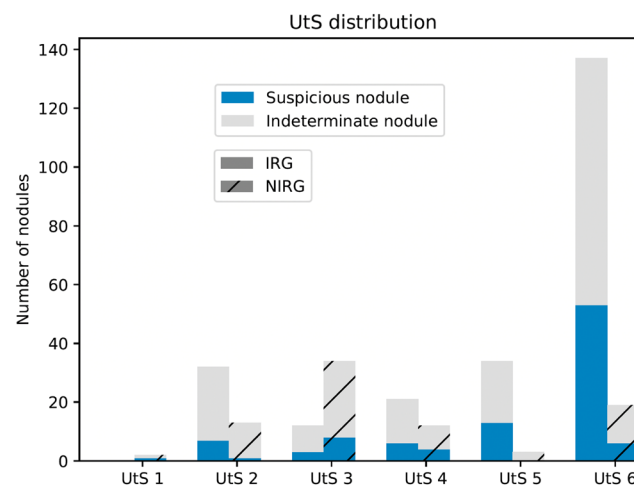


Figure 2

Distribution of indeterminate/suspicious thyroid bed nodules in each UtS category for different radiology groups. IRG, implementation radiology group; NIRG, all non-implementation radiology groups; UtS, utility score.

those done by NIRG were sufficient ($P < 0.05$). Similarly, for POU reports with indeterminate or suspicious LNs, IRG had sufficient reports in 66% and 85% of cases, respectively, whereas NIRG had sufficient reports in only 14% and 68% of cases ($P < 0.05$).

FNA and additional surgery

We further analyzed 74 abnormal lesions that were followed by either fine-needle aspiration (FNA) or additional surgery to evaluate the diagnostic accuracy of the POU reports for these lesions. Out of these 74 lesions, 36 were biopsied following a POU performed by IRG, with 14 being malignant, 12 benign, 6 indeterminate, and 4 non-diagnostic. Another 22 lesions were resected in subsequent surgeries without FNA, 20 lesions proved to be malignant and 2 were benign. Among the remaining 16 abnormal lesions, 12 were biopsied following a POU performed by NIRG, with 4 being malignant, 5 benign, 2 indeterminate, and 1 non-diagnostic. Additionally, four lesions were resected in subsequent surgeries, with all four proving to be malignant (Table 3).

The decision to perform FNA or additional surgery is a shared informed decision by the patient and the respective clinician, taking into account the assessment of the lesion size, malignancy risk classification, sonographic changes, FNA result, feasibility and risks of repeat neck surgery, and the patient's clinical status. The specific radiology group responsible for the POU does not influence the clinical decision-making process.

What is the predictive ability of baseline Tg and POU at 1-year FU in determining the absence of persistent disease or relapse at 3-year FU?

Among 269 patients with Tg-N at 1-year FU, 252 (94%) had US-N at 3-year FU POU, which corresponds to an

NPV of 94% for Tg-N at 1-year FU. For 345 patients with US-N at 1-year FU, 330 (96%) had US-N at 3-year FU POU (Table 4).

There were 242 patients in cohort 1 (96 with TTX plus RAI, 106 with TTX, and 40 with lobectomy) and 67 in cohort 2 (36 with TTX plus RAI and 31 with TTX). Details of their tumor characteristics and surgery are outlined in Table 5. About 96% of cohort 1 and 96% of cohort 2 had US-N at 3-year FU (Supplementary Tables 4 and 5). This corresponds to an NPV for US-N and Tg-N at 1-year FU of 96% and an NPV for US-N and Tg-I at 1-year FU of 96% as well. Among cohort 1, the NPV of US-N and Tg-N was 93% (89 of 96) for TTX plus RAI, 99% (105 of 106) for TTX only, and 98% (39 of 40) for lobectomy, there was no statistical difference in NPV for US-N and Tg-N between the TTX+RAI, TTX only, and Lobectomy groups ($P > 0.1$, Supplementary Table 4). Among cohort 2, the NPV of US-N and Tg-I was 94% (34 of 36) for TTX plus RAI and 97% (30 of 31) for TTX ($P > 0.1$, Supplementary Table 5). Serum Tg level and ultrasound lesion status of all 672 patients at 1-Year FU are provided in Supplementary Table 6.

Discussion

The impact of adherence to ETA postoperative ultrasound guidelines on POU report quality

The UtS and sufficiency rate of a POU report directly relate to the reliability and the amount of information available in the report to describe any thyroid bed nodule or lymph node identified. We previously showed that when radiology groups adhered to the 2015 ATA or 2017 Thyroid Imaging Reporting and Data Systems (TIRADS) reporting guidelines for thyroid ultrasound, they significantly improved the UtS and classification reporting rate of their reports compared to other

Table 3 Abnormal lesions followed by FNA or additional surgery ($n = 74$).

	n	IRG			NIRG		
		n1	Indeterminate	Suspicious	n2	Indeterminate	Suspicious
Postoperative FNA	48	36	12	24	12	6	6
FNA malignant		14 (39%)	3 (25%)	11 (46%)	4 (33%)	2 (33%)	2 (33%)
FNA benign		12 (33%)	5 (42%)	7 (29%)	5 (42%)	2 (33%)	3 (50%)
FNA indeterminate ^a		6	3	3	2	1	1
FNA non-diagnostic		4	1	3	1	1	
Additional surgery (no FNA)	26	22	11	11	4	2	2
LN metastases –yes		20 (91%)	9 (82%)	11 (100%)	4 (100%)	2 (100%)	2 (100%)
LN metastases – no		2	2	0	0	0	0

^aFLUS (follicular lesion of undetermined significance) or SFN (suspicious for Hurthle cell neoplasm).

FNA, fine-needle aspiration; IRG, implementation radiology group; NIRG, all non-implementation radiology groups.

Table 4 All patients with 3-year FU: serum Tg level and US lesion statuses at 1-year FU and 3-year FU ($n = 441$).

	US-N at 3-year FU	NPV
TTX and RAI		
US-N at 1-year FU ($n = 153$)	141	92%
US-N at 1-year FU ($n = 95$) ^a	92	97%
Tg-N at 1-year FU ($n = 114$)	101	89%
Tg-N at 1-year FU ($n = 78$) ^a	71	91%
TTX only		
US-N at 1-year FU ($n = 145$)	143	99%
US-N at 1-year FU ($n = 144$) ^{a,b}	142	99%
Tg-N at 1-year FU ($n = 109$)	108	99%
Tg-N at 1-year FU ($n = 108$) ^{a,b}	107	99%
Lobectomy		
US-N at 1-year FU ($n = 47$)	46	98%
Tg-N at 1-year FU ($n = 47$)	44	94%
Total		
US-N at 1-year FU ($n = 345$)	330	96%
Tg-N at 1-year FU ($n = 269$)	252	94%

^aAfter excluding high RR patients;

^bone high RR patient canceled RAI.

FU, follow-up; NPV, negative predictive value; RAI, radioiodine remnant ablation; Tg, thyroglobulin; Tg-I, thyroglobulin levels-indeterminate, Tg 0.2–0.99 ng/mL for TTX and RAI and TTX only patients; Tg-N, thyroglobulin levels-negative, Tg <0.2 ng/mL for TTX and RAI and TTX only patients or Tg <30 ng/mL for lobectomy patients; Tg-S, thyroglobulin levels-suspicious, Tg ≥1 ng/mL or Tg ≥30 ng/mL for lobectomy patients; TTX, total thyroidectomy; US, ultrasound; US-I, US lesion status indeterminate, no suspicious lesion but at least one lesion classified as indeterminate; US-N, US lesion status, all visible lesions are normal, or no lesion mentioned; US-S, US lesion status suspicious, at least one lesion was classified as suspicious.

radiology groups who did not implement strict adherence (13). Similarly, our current data show that IRG had significantly higher UtS compared to NIRG when reporting indeterminate or suspicious thyroid bed nodules, both prior to and after their adherence to the ETA guidelines (Table 1). The mean UtS was significantly higher for both IRG and NIRG compared to their POU reports before and after adhering to the 2013 ETA guidelines (Table 2). This could be explained by performance variations within NIRG. Specifically, one radiology group among NIRG exhibited significantly higher mean UtS and sufficiency rates for post-ETA POU reports describing abnormal thyroid bed nodules compared to the remaining groups (Supplementary Tables 7 and 8), which will be referred to as the highest-quality non-implementation radiology group (NIRG-HP). As a result of NIRG-HP's performance, there was an overall improvement in POU quality for abnormal thyroid bed nodules among the NIRG, comprising NIRG-HP and eight other non-implementation radiology groups (NIRG-OT). However, it is worth noting that the number of POU performed by NIRG-HP was relatively small.

Considering the current reporting practice for lymph nodes, a POU report that describes an abnormal lymph node with a low UtS can still be considered clinically useful in assessing patient RTT if it provides information on abnormal features and corresponding malignancy risk classification. To facilitate the comparison of POU quality related to lymph nodes across different radiology groups, we also incorporated the sufficiency rate criterion to evaluate report quality. However, we still recommend that radiology groups further improve their POU reports describing abnormal lymph nodes by providing lymph node features and malignancy risk classifications according to the ETA guideline recommendations in the future.

POU reports done by IRG describing suspicious thyroid bed nodules have a 94% sufficiency rate (percentage of sufficient POU reports overall POU reports in this category), which is significantly higher than 45% by NIRG. The results are similar for indeterminate thyroid bed nodules (85% vs 27%), suspicious LNs (85% vs 68%), and indeterminate LNs (66% vs 14%). Additional data on NIRG-HP for POU reports describing abnormal lymph nodes showed lower sufficiency rates than IRG, particularly for indeterminate lymph nodes (Supplementary Table 8). Overall, among the four abnormal lesion categories, all radiology groups exhibited the lowest sufficiency rates for describing indeterminate lymph nodes (Supplementary Table 8). This finding highlights the need for improvement in lymph node description following the guideline-based assessment.

Based on our data, IRG had significantly higher quality POU reports compared to NIRG and this difference may be explained by the implementation of the 2013 ETA reporting guidelines, and this difference most likely leads to a reduction in unnecessary repeat POU. POU is crucial in the detection of persistent disease or recurrent disease during the FU of thyroid cancer patients after their initial treatment for their thyroid cancer. A high-quality POU report provides more pertinent information, enabling clinicians to more reliably estimate RTT and future RR.

Given the limited number of FNAs and additional surgeries performed for abnormal lesions in our cohorts, particularly those following POU done by NIRG, we found no significant difference in diagnostic accuracy between IRG and NIRG. Within IRG, FNAs or surgeries conducted for suspicious lesions had higher accuracy compared to those performed for indeterminate lesions.

Table 5 Baseline characteristics of cohort 1 (US-N + Tg-N patients with 3-year FU) and cohort 2 (US-N + Tg-I patients with 3-year FU).

Serum Tg and neck lymph node status	Cohort 1: US-N + Tg-N (n = 242) TTX and RAI or TTX only: 202 (83%)			Cohort 2: US-N + Tg-I (n = 67)	
	TTX and RAI	TTX only	Lobectomy ^a	TTX and RAI	TTX only
Treatment type					
Total	96 (48%)	106 (52%)	40	36 (54%)	31 (46%)
Age at diagnosis, year (median, range)	45 (16, 77)	49 (23, 87)	43 (24, 69)	42 (18, 72)	44 (24, 71)
Females	68 (71%)	86 (81%)	31	22 (61%)	24 (77%)
Neck dissection					
Not done	46 (48%)	87 (82%)	38	14 (39%)	25 (81%)
Central compartment only	27 (28%)	17 (16%)	2	13 (36%)	4 (13%)
Central and lateral compartments	23 (24%)	2 (2%)	0	9 (25%)	2 (6%)
Tumor size, mm (median, range)	24 (3.8, 89)	15 (1, 63)	13 (1.3, 70)	26.5 (2.2, 85)	14 (1.5, 50)
Tumor foci					
Unifocal	58 (60%)	45 (42%)	32	11 (31%)	13 (42%)
Multifocal	38 (40%)	61 (58%)	8	24 (67%)	18 (58%)
Gross extrathyroidal extension	17 (18%)	1 (1%)	0	13 (36%)	3 (10%)
Microscopic extrathyroidal extension	45 (47%)	4 (4%)	0	18 (50%)	1 (3%)
Lymph node status					
Nx/N0	41 (43%)	89 (84%)	37	10 (28%)	22 (71%)
N1a	31 (32%)	16 (15%)	3	15 (42%)	8 (26%)
N1b	24 (25%)	1 (1%)	0	11 (30%)	1 (3%)
ATA recurrence risk					
Low	12 (13%)	103 (97%)	40	3 (8%)	28 (90%)
Intermediate	54 (56%)	2 (2%)	0	20 (56%)	3 (10%)
High	30 (31%)	1 (1%)	0	13 (36%)	0

^aLobectomy patients in cohort 1 were excluded for percentage calculation to be better compared with cohort 2.

ATA, American Thyroid Association; FU, follow-up; NPV, negative predictive value; RAI, radioiodine remnant ablation; RAI, radioiodine remnant ablation; Tg, thyroglobulin; Tg-I, thyroglobulin levels-indeterminate, Tg 0.2–0.99 ng/mL for TTX and RAI and TTX only patients; Tg-N, thyroglobulin levels-negative, Tg <0.2 ng/mL for TTX and RAI and TTX only patients or Tg < 30 ng/mL for lobectomy patients; Tg-S, thyroglobulin levels-suspicious, Tg ≥ 1 ng/mL or Tg ≥ 30 ng/mL for lobectomy patients; TTX, total thyroidectomy; US, ultrasound; US-I, US lesion status indeterminate, no suspicious lesion but at least one lesion classified as indeterminate; US-N, US lesion status, all visible lesions are normal, or no lesion mentioned; US-S, US lesion status suspicious, at least one lesion was classified as suspicious.

This finding is intuitive, as suspicious lesions are more likely to have suspicious features.

The 2015 ATA guidelines do not differentiate post-treatment ultrasound findings and Tg cutoffs between patients who received RAI vs no RAI, and there are limited data on abnormal findings in patients with lobectomy and Tg cutoffs (1, 16). A recent meta-analysis indicates that following total/near-TTX, using a Tg levels cutoff of 1–2.5 ng/mL might identify patients at low risk for persistent or metastatic disease (17). However, due to the limited availability of qualified patients ($n = 10$), we were unable to provide any conclusions for this cutoff. The quality of the POU report plays a crucial role in determining the need for subsequent biopsy and the diagnostic accuracy. If only POU reports of low quality are available, clinicians may have to repeat POU and Tg determination or request FNA to gather adequate information to assess the RTT and to make further decisions. POU reports with higher UtS or sufficient POU reports according to ETA guidelines would save many repeat procedures, resulting in lower patient anxiety, fewer complications, and reduced healthcare costs.

Prognostic value of baseline serum Tg and postoperative neck US at 1-year FU

It is well described that baseline Tg values early in FU have strong predictive value for US-N at 3-year FU (5, 6, 16), with Tg-N having an NPV as high as 98.8% and 98.2% for low- and intermediate-risk patients, respectively, and the NPV of Tg-I was as high as 98.2% and 94.5% for low- and intermediate-risk patients, respectively (6). Additionally, Grani *et al.* found that among low- and intermediate-RR patients with abnormal neck ultrasound at 3-year FU, 75% of 226 lesions were likely false positive and none required treatment; however, no further details were provided.

Similarly, the absence of structural disease early in FU also predicts the ongoing absence of structural disease over 5–10 years. Among 67 ATA high-risk patients with stimulated Tg <1 ng/mL before RAI, the rate of abnormal ultrasound findings is 3% after median FU of 5.6 years (18), while among those with persistently elevated or rising serum Tg, the rate is between 8% and 17% (19).

We observed similar rates in our prospective population data with the implementation of ETA POU

guidelines. Patients with US-N at 1-year FU had an NPV of 96% for having US-N at 3-year FU, while Tg-N at 1-year FU had an NPV of 94% for US-N at 3-year FU (Table 4). However, we found that the NPVs for TTX plus RAI group were lower for both US-N and Tg-N compared to TTX only and lobectomy. For US-N, the NPV was 92% for TTX and RAI, 99% for TTX only, and 98% for lobectomy, respectively, and for Tg-N, the NPV was 89%, 99%, and 94%, respectively (Table 4). This may be due to a higher percentage of high-risk patients in the TTX and RAI group. After excluding all high-risk patients, the NPV for US-N at 1-year FU was 97% and for Tg-N at 1-year FU was 91%, which are similar to the NPVs for TTX only and lobectomy groups. For low- and intermediate-risk patients, the treatment modality did not affect the NPV of Tg-N or US-N. Finally, the combination of US-N and Tg-N (cohort 1) at 1-year FU had an NPV of 96%, which was not significantly different from using Tg-N alone (94%, $P=0.1$) or US-N alone (96%, $P>0.1$).

It is noted in a retrospective study of 619 patients with papillary thyroid cancer (PTC) treated with lobectomy that postoperative serum Tg levels or ultrasound findings do not correlate with disease recurrence (20), but other studies have found that Tg levels >30 ng/mL still indicate a high probability of recurrent structural disease (12). In our study, among 40 lobectomy patients who had US-N and serum Tg <30 ng/mL at 1-year FU, only one developed indeterminate ultrasound findings at 3-year FU (Supplementary Table 4). In this solitary patient, the nodule is stable and less than 1 cm without suspicious features, indicating a 0% recurrence rate at 3-year FU among lobectomy patients with excellent RTT at 1-year FU. However, relapses after lobectomy may occur very late (21).

Overall, our prospective data indicate that for patients treated with TTX, with or without RAI, a normal serum Tg at 1-year FU is highly predictive of no structural disease at 3-year FU and the addition of a neck ultrasound does not significantly increase the predictive value.

Strengths of our data include collection from a large population and in a prospective manner over a FU period of 3 years. We also included a large number of POU reports and many indeterminate and suspicious lesions in the analysis.

Our limitations include the low number of FNAs and subsequent surgeries performed after POU, particularly by NIRG. Our FU was restricted to 3 years, whereas

the peak for thyroid cancer RR has been reported at 5 years (22, 23, 24). There are also potential confounders, such as variations in fellowship training, technologist expertise or experience, and potential differences in the referred patient populations for each radiology group. These unmeasured factors could introduce confounding effects and impact the comparison between groups.

Conclusion

The implementation of the 2013 ETA guideline-based POU assessment allows radiology groups to provide high-quality POU reports to clinicians, which may lead to increased accuracy in assessing RTT and estimating the risk of recurrence of thyroid cancer and likely the reduction of unnecessary repeat ultrasound or FNA.

Supplementary materials

This is linked to the online version of the paper at <https://doi.org/10.1530/ETJ-23-0110>.

Declaration of interest

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

Funding

This work was supported by the University of Calgary PhD/Doctorate Student Funding and the University of Calgary Faculty of Graduate Studies Doctoral Entrance Scholarship.

References

- 1 Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, Pacini F, Randolph GW, Sawka AM, Schlumberger M, 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** 1–133. (<https://doi.org/10.1089/thy.2015.0020>)
- 2 Leboulleux S, Girard E, Rose M, Travagli JP, Sabbah N, Caillou B, Hartl DM, Lassau N, Baudin E & Schlumberger M. Ultrasound criteria of malignancy for cervical lymph nodes in patients followed up for differentiated thyroid cancer. *Journal of Clinical Endocrinology and Metabolism* 2007 **92** 3590–3594. (<https://doi.org/10.1210/jc.2007-0444>)
- 3 Randolph GW, Duh QY, Heller KS, LiVolsi VA, Mandel SJ, Steward DL, Tufano RP, Tuttle RM & American Thyroid Association Surgical Affairs Committee's Taskforce on Thyroid Cancer Nodal Surgery. The prognostic significance of nodal metastases from papillary

- thyroid carcinoma can be stratified based on the size and number of metastatic lymph nodes, as well as the presence of extranodal extension. *Thyroid* 2012 **22** 1144–1152. (<https://doi.org/10.1089/thy.2012.0043>)
- 4 Peiling Yang S, Bach AM, Tuttle RM & Fish SA. Frequent screening with serial neck ultrasound is more likely to identify false-positive abnormalities than clinically significant disease in the surveillance of intermediate risk papillary thyroid cancer patients without suspicious findings on follow-up ultrasound evaluation. *Journal of Clinical Endocrinology and Metabolism* 2015 **100** 1561–1567. (<https://doi.org/10.1210/jc.2014-3651>)
 - 5 Jeon MJ, Kim M, Park S, Oh HS, Kim TY, Kim WB, Shong YK & Kim WG. A follow-up strategy for patients with an excellent response to initial therapy for differentiated thyroid carcinoma: less is better. *Thyroid* 2018 **28** 187–192. (<https://doi.org/10.1089/thy.2017.0130>)
 - 6 Grani G, Ramundo V, Falcone R, Lamartina L, Montesano T, Biffoni M, Giacomelli L, Sponziello M, Verrienti A, Schlumberger M, et al. Thyroid cancer patients with no evidence of disease: the need for repeat neck ultrasound. *Journal of Clinical Endocrinology and Metabolism* 2019 **104** 4981–4989. (<https://doi.org/10.1210/jc.2019-00962>)
 - 7 Leenhardt L, Erdogan MF, Hegedus L, Mandel SJ, Paschke R, Rago T & Russ G. 2013 European Thyroid Association guidelines for cervical ultrasound scan and ultrasound-guided techniques in the postoperative management of patients with thyroid cancer. *European Thyroid Journal* 2013 **2** 147–159. (<https://doi.org/10.1159/000354537>)
 - 8 Lamartina L, Grani G, Biffoni M, Giacomelli L, Costante G, Lupo S, Maranghi M, Plasmati K, Sponziello M, Trulli F, et al. Risk stratification of neck lesions detected sonographically during the follow-up of differentiated thyroid cancer. *Journal of Clinical Endocrinology and Metabolism* 2016 **101** 3036–3044. (<https://doi.org/10.1210/jc.2016-1440>)
 - 9 Frates MC, Parziale MP, Alexander EK, Barletta JA & Benson CB. Role of sonographic characteristics of thyroid bed lesions identified following thyroidectomy in the diagnosis or exclusion of recurrent cancer. *Radiology* 2021 **299** 374–380. (<https://doi.org/10.1148/radiol.202101596>)
 - 10 Park S, Jeon MJ, Oh HS, Lee YM, Sung TY, Han M, Han JM, Kim TY, Chung KW, Kim WB, et al. Changes in serum thyroglobulin levels after lobectomy in patients with low-risk papillary thyroid cancer. *Thyroid* 2018 **28** 997–1003. (<https://doi.org/10.1089/thy.2018.0046>)
 - 11 Wu J, Hu XY, Ghaznavi S, Kinnear S, Symonds CJ, Grundy P, Parkins VM, Sharma P, Lamb D, Khalil M, et al. The prospective implementation of the 2015 ATA guidelines and modified ATA recurrence risk stratification system for treatment of differentiated thyroid cancer in a Canadian tertiary care referral setting. *Thyroid* 2022 **32** 1509–1518. (<https://doi.org/10.1089/thy.2022.0055>)
 - 12 Momesso DP, Vaisman F, Yang SP, Bulzico DA, Corbo R, Vaisman M & Tuttle RM. Dynamic risk stratification in patients with differentiated thyroid cancer treated without radioactive iodine. *Journal of Clinical Endocrinology and Metabolism* 2016 **101** 2692–2700. (<https://doi.org/10.1210/jc.2015-4290>)
 - 13 Hu XY, Wu J, Seal P, Ghaznavi SA, Symonds C, Kinnear S & Paschke R. Improvement in thyroid ultrasound report quality with radiologists' adherence to 2015 ATA or 2017 TIRADS: a population study. *European Thyroid Journal* 2022 **11**. (<https://doi.org/10.1530/ETJ-22-0035>)
 - 14 Symonds CJ, Seal P, Ghaznavi S, Cheung WY & Paschke R. Thyroid nodule ultrasound reports in routine clinical practice provide insufficient information to estimate risk of malignancy. *Endocrine* 2018 **61** 303–307. (<https://doi.org/10.1007/s12020-018-1634-0>)
 - 15 R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2016.
 - 16 Pitoia F, Abelleira E & Cross G. Thyroglobulin levels measured at the time of remnant ablation to predict response to treatment in differentiated thyroid cancer after thyroid hormone withdrawal or recombinant human TSH. *Endocrine* 2017 **55** 200–208. (<https://doi.org/10.1007/s12020-016-1104-5>)
 - 17 Chou R, Dana T, Brent GA, Goldner W, Haymart M, Leung AM, Ringel MD & Sosa JA. Serum thyroglobulin measurement following surgery without radioactive iodine for differentiated thyroid cancer: a systematic review. *Thyroid* 2022 **32** 613–639. (<https://doi.org/10.1089/thy.2021.0666>)
 - 18 Tian T, Kou Y, Huang R & Liu B. Prognosis of high-risk papillary thyroid cancer patients with pre-ablation stimulated Tg < 1 ng/ml. *Endocrine Practice* 2019 **25** 220–225. (<https://doi.org/10.4158/EP-2018-0436>)
 - 19 Pitoia F, Abelleira E, Tala H, Bueno F, Urciuoli C & Cross G. Biochemical persistence in thyroid cancer: is there anything to worry about? *Endocrine* 2014 **46** 532–537. (<https://doi.org/10.1007/s12020-013-0097-6>)
 - 20 Cho JW, Lee YM, Lee YH, Hong SJ & Yoon JH. Dynamic risk stratification system in post-lobectomy low-risk and intermediate-risk papillary thyroid carcinoma patients. *Clinical Endocrinology* 2018 **89** 100–109. (<https://doi.org/10.1111/cen.13721>)
 - 21 Bosset M, Bonjour M, Castellnou S, Hafdi-Nejjari Z, Bournaud-Salinas C, Decaussin-Petrucci M, Lifante JC, Perrin A, Peix JL, Moulin P, et al. Long-term outcome of lobectomy for thyroid cancer. *European Thyroid Journal* 2021 **10** 486–494. (<https://doi.org/10.1159/000510620>)
 - 22 Tuttle RM, Tala H, Shah J, Leboeuf R, Ghossein R, Gonen M, Brokhin M, Omry G, Fagin JA & Shaha A. 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** 1341–1349. (<https://doi.org/10.1089/thy.2010.0178>)
 - 23 Vaisman F, Momesso D, Bulzico DA, Pessoa CH, Dias F, Corbo R, Vaisman M & Tuttle RM. Spontaneous remission in thyroid cancer patients after biochemical incomplete response to initial therapy. *Clinical Endocrinology* 2012 **77** 132–138. (<https://doi.org/10.1111/j.1365-2265.2012.04342.x>)
 - 24 Pitoia F, Bueno F, Urciuoli C, Abelleira E, Cross G & Tuttle RM. Outcomes of patients with differentiated thyroid cancer risk-stratified according to the American Thyroid Association and Latin American Thyroid Society risk of recurrence classification systems. *Thyroid* 2013 **23** 1401–1407. (<https://doi.org/10.1089/thy.2013.0011>)

Received 31 May 2023

Accepted 12 July 2023

Available online 13 July 2023

Version of Record published 11 August 2023