

## RESEARCH

# Improvement in thyroid ultrasound report quality with radiologists' adherence to 2015 ATA or 2017 TIRADS: a population study

X Y Hu<sup>1</sup>, J Wu<sup>2</sup>, P Seal<sup>3</sup>, S A Ghaznavi<sup>2,4</sup>, C Symonds<sup>4</sup>, S Kinnear<sup>4</sup> and R Paschke<sup>4,5</sup><sup>1</sup>Department of Medicine, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada<sup>2</sup>Arnie Charbonneau Cancer Institute, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada<sup>3</sup>EFW Radiology, Calgary, Alberta, Canada<sup>4</sup>Department of Medicine, Section of Endocrinology, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada<sup>5</sup>Departments 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, CanadaCorrespondence should be addressed to R Paschke: [ralf.paschke@ucalgary.ca](mailto:ralf.paschke@ucalgary.ca)

## Abstract

**Objectives:** There has been slow adoption of thyroid ultrasound guidelines with adherence rates as low as 30% and no population-based studies investigating adherence to guideline-based malignancy risk assessment. We therefore evaluated the impact of adherence to the 2015 ATA guidelines or 2017 ACR-TIRADS guidelines on the quality of thyroid ultrasound reports in our healthcare region.

**Methods:** We reviewed 899 thyroid ultrasound reports of patients who received fine-needle aspiration biopsy and were diagnosed with Bethesda III or IV nodules or thyroid cancer. Ultrasounds were reported by radiology group 1, group 2, or other groups, and were divided into pre-2018 (before guideline adherence) or 2018 onwards. Reports were given a utility score (0–6) based on how many relevant nodule characteristics were included.

**Results:** Group 1 had a pre-2018 utility score of 3.62 and 39.4% classification reporting rate, improving to 5.77 and 97.0% among 2018-onwards reports. Group 2 had a pre-2018 score of 2.8 and reporting rate of 11.5%, improving to 5.58 and 93.3%. Other radiology groups had a pre-2018 score of 2.49 and reporting rate of 32.2%, improving to 3.28 and 61.8%. Groups 1 and 2 had significantly higher utility scores and reporting rates in their 2018-onward reports when compared to other groups' 2018-onward reports, pre-2018 group 1 reports, and pre-2018 group 2 reports.

**Conclusions:** Dedicated adherence to published thyroid ultrasound reporting guidelines can lead to improvements in report quality. This will reduce diagnostic ambiguity and improve clinician's decision-making, leading to overall reductions in unnecessary FNA biopsy and diagnostic surgery.

## Key Words

- ▶ thyroid
- ▶ ultrasound
- ▶ thyroid imaging

## Introduction

Thyroid ultrasound (TUS) is the most sensitive and cost-effective modality for the evaluation of thyroid nodules (1). When thyroid nodules are incidentally discovered on other imaging modalities, current guidelines suggest that TUS malignancy risk assessment should be performed (2, 3).

Currently, there are multiple guidelines for the malignancy risk stratification of thyroid nodules based on TUS features, which include: the 2015 American Thyroid Association (ATA) Management Guidelines, the 2016 Korean Thyroid Association/Korean Society of Thyroid Radiology Guidelines (K-TIRADS), the 2017 European Thyroid Imaging and Reporting Data System (EU-TIRADS), the 2020 Chinese Thyroid Imaging Reporting and Data System (C-TIRADS), the 2016 AACE/ACE/AME Medical Guidelines, and the 2017 American College of Radiology Thyroid Imaging, Reporting and Data System (ACR-TIRADS). These guidelines outline and score nodule features that are suggestive of malignancy.

Using this ultrasound malignancy risk stratification information, in addition to patient characteristics, clinicians make a decision between pursuing further diagnostic testing (usually in the form of fine-needle aspiration biopsy (FNAB)), repeating ultrasound imaging, or discontinuing follow up (2, 4). In experienced hands, the false-negative rate of thyroid nodule malignancy risk stratification using ultrasound may be as low as 2.2% (5). Certain centers have noted that adherence to the use of a nodule risk stratification system leads to a substantial reduction in referrals for unnecessary FNAB, ranging from 19 to 55% reduction with ACR-TIRADS and 14% reduction with the ATA system in one study (6, 7). In addition, previous studies that prompted radiologists to use ACR-TIRADS saw a significant increase in the number of nodule features and recommendation of management in the ultrasound reports (8).

However, despite the long-term evidence and guideline-based support for their use, there has been a slow adoption of the guidelines in single-center studies, and a previous study has shown adherence rates as low as 30% for specific radiology groups or healthcare areas (7). Since the publication of these guidelines, there have been no population-based studies investigating the adherence of radiologists or other users of TUS malignancy risk assessment to guidelines. We have previously shown very low adoption rates of TUS malignancy risk stratification by radiology groups in the Calgary and Southern Alberta Healthcare regions (9). Based on these initial data, we

implemented ATA guideline-based thyroid nodule malignancy risk stratification with one of the major radiology groups in Alberta. This was soon followed by the introduction of ACR-TIRADS guideline-based thyroid nodule malignancy risk stratification by another major radiology group within the same health region.

Following these changes, our study seeks to re-evaluate the quality of TUS reports in the Calgary and Southern Alberta Healthcare regions, covering 1.5 million inhabitants, to analyze the improvement in TUS reports of patients that were selected for thyroid nodule FNA or patients with a new diagnosis of thyroid cancer. We looked at two separate prospective databases of TUS performed on patients with either thyroid nodule(s) selected for FNA or thyroid nodules with a later diagnosis of thyroid cancer. The goals of our investigation were to evaluate changes in the proportion of TUS reports with a clinically useful utility score and the proportion of TUS reports that had an ATA or ACR-TRADS guideline-based classification for thyroid nodule malignancy risk.

## Methods

Alberta Health Services is a comprehensive, integrated, single-payer healthcare system with centralized laboratory, pathology, surgery, endocrinology, and oncology services. It has a single electronic medical record (EMR) system, for over 1.5 million inhabitants of the Calgary and Southern Alberta Healthcare regions.

For this study, we reviewed 981 patients from 2 databases. Database 1 is the prospective web-based REDCap Calgary thyroid nodule database (Ethics ID: HREBA.CHC-20-0068\_REN1) and database 2 is the prospective REDCap Calgary thyroid cancer database (Ethics ID: HREBA.CC-16-0956). Database 1 included 353 patients with at least 1 indeterminate (Bethesda III/IV) nodule identified on FNA in the Calgary and Southern Alberta Healthcare region between July 31, 2020, and November 1, 2021. After excluding thyroid biopsy ultrasound reports and 2 ultrasound reports performed before 2018, there were 342 available pre-FNA US reports for these patients that suggested the presence of one or more thyroid nodules. Database 2 included 628 patients diagnosed with thyroid cancer who went on to receive thyroid surgery in the Calgary and Southern Alberta Healthcare regions between April 1, 2017, and November 1, 2021. After excluding 47 patients who were already included in database 1 and 24 patients with ultrasound reports that recorded only lymph nodes, database 2 included the remaining 557 preoperative

thyroid ultrasound reports with the presence of one or more thyroid nodules in our analysis.

We previously analyzed the routine thyroid ultrasound reports in the Calgary and Southern Alberta Healthcare regions based on two criteria: (i) whether or not an ATA or TIRADS classification was reported and (ii) the utility score of the report (9). This allowed us to assess the quality of the reports from these two radiology groups by the same methods. The classification reporting rate of a radiology group was calculated as the percentage of TUS reports that provided an ATA or TIRADS classification among all TUS reports from that group over the same time period. Each TUS report was assigned a utility score (UtS; range, 0–6), calculated based on the number of nodule characteristics provided in the report. These characteristics include size, internal content, shape, margin, echogenicity, and presence of calcifications and are the qualities evaluated in existing guidelines (2, 4). For every characteristic mentioned, the report received 1 point. Reports with a UtS of 4–6 were considered clinically useful and those with a UtS of 0–3 were considered not clinically useful. Together, the two variables (classification system used/not used, and UtS useful/not useful) generated four categories of TUS reports: clinically non-useful without a classification, clinically non-useful with a classification, clinically useful without a classification, and clinically useful with a classification.

Given that both radiology group 1 and group 2 adopted a classification system-based approach in 2018, we divided database 2 into pre-2018 and 2018 onwards to demarcate the era before and after adoption of the classification systems. Conversely, database 1 began data collection in September 2018 and did not need to be divided. For longitudinal analysis, pre-2018 TUS reports from database 2 were compared against the composite of database 1 reports and database 2 reports from 2018 onwards.

We compared the report quality of TUS performed by radiology group 1, radiology group 2, and other radiology groups in the Calgary and Southern Alberta Healthcare regions. Radiology group 1 uniformly implemented the ATA malignancy risk stratification system (which classifies nodules as ATA benign, ATA very low risk, ATA low risk, ATA intermediate risk, and ATA high risk) for all nodule reporting in January 2018 (3). They made software changes to their nodule-reporting program, which created a defined set of terms within a drop-down menu, and mandatory fields for each nodule characteristic. This was done to standardize and expedite reporting according to the ATA system. Radiology group 2 uniformly implemented the ACR-TIRADS risk stratification system (which classifies nodules as TIRAD 1–5) for all nodule reporting in November

**Table 1** Baseline characteristics of patients.

Total	899
Female	652 (73%)
Male	247 (27%)
Mean age in years (range)	53 (12–94)
Mean no. of nodules per ultrasound report	2.97
Mean size of largest nodule (cm)	2.55
Size range of largest nodule	
<1 cm	57 (6%)
1–2 cm	365 (41%)
>2 cm	457 (51%)
No size reported	20 (2%)

2018 (4). The process of implementation is similar between groups 1 and 2 and is available in the Supplementary material (see section on [supplementary materials](#) given at the end of this article). Other radiology groups used either of the two systems of their choice with variable degrees of implementation.

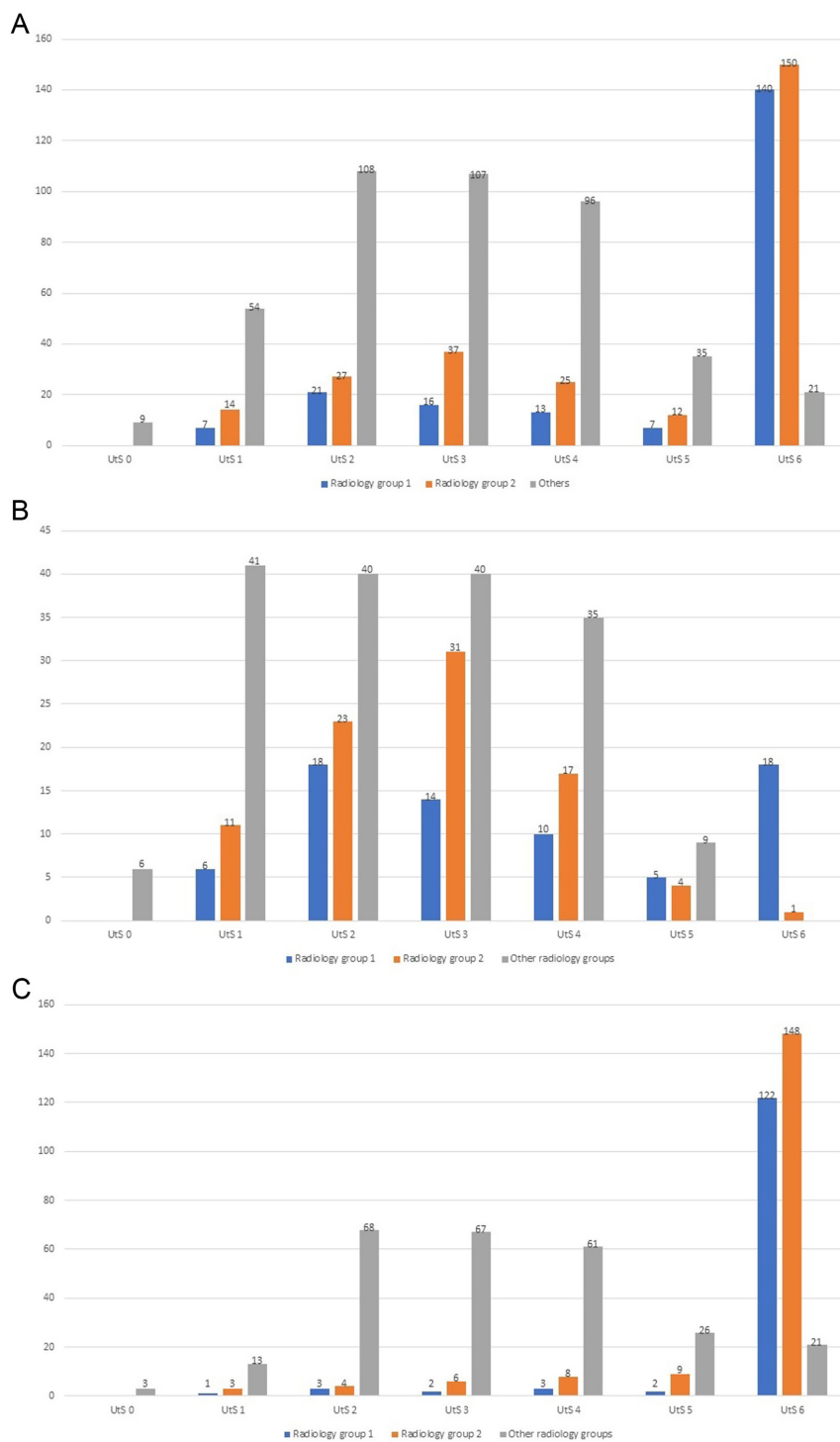
## Results

A total of 899 TUS reports were collected from the 2 databases and the respective patients' characteristics are summarized in [Table 1](#). Radiology group 1 interpreted 204 TUS, group 2 interpreted 265 TUS, and other groups collectively interpreted 430 TUS ([Table 2](#)). Across all radiology groups, there were a total of 311 (35%) US with UtS 6, 54 (6%) with UtS 5, 134 (15%) with UtS 4, 160 (18%) with UtS 3, 156 (17%) with UtS 2, 75 (8%) with UtS 1, and 9 (1%) with UtS 0 ([Fig. 1](#) and [Table 3](#)).

The mean UtS was 5.02 for radiology group 1, 4.67 for radiology group 2, and 2.97 for other radiology groups, and their UtS across different nodule sizes is shown in [Fig. 2](#) ( $P=0.01$  for group 1 vs group 2,  $P < 0.0001$  for group 1 or group 2 vs other groups). Radiology group 1 had 140 (68.6%) reports with a UtS of 6 and 160 (78.4%) with UtS of 4 or greater. Radiology group 2 had 150 (56.6%) with a UtS of 6 and 187 (70.6%) with UtS of 4 or greater. Other radiology groups had 21 (4.9%) with a UtS of 6 and 152 (35.3%) with UtS of 4 or greater ([Fig. 1](#)). The difference between the mean UtS of radiology groups 1 or 2 compared to other radiology groups was significant ( $P < 0.0001$  and

**Table 2** Number of TUS performed by each radiology group.

Total	899
US performed by radiology group 1	204 (23%)
US performed by radiology group 2	265 (29%)
US performed by others	430 (48%)



**Figure 1**

Distribution of utility score of thyroid ultrasound reports by different radiology groups. (A) Overall distribution of all reports. (B) Distribution of pre-2018 reports. (C) Distribution of 2018-onwards reports.

$P < 0.0001$ ), and the difference between the mean Uts of radiology groups 1 and 2 is also significant ( $P = 0.01$ ).

There were 351 reports that had no ATA or ACR-TIRADS classification specified. Of these, 5 had Uts 6, 13 had Uts 5, 72 had Uts 4, 100 had Uts 3, 121 had Uts 2, 39 had Uts 1, and 1 had Uts 0 (Fig. 3 and Table 4). This indicates that out

of the total number of reports in a utility score category, 1.6% with Uts 6, 24.1% with Uts 5, 53.7% with Uts 4, 62.5% with Uts 3, 77.6% with Uts 2, 52.0% with Uts 1, and 11.1% with Uts 0 had no classification (Fig. 3).

Within database 2, the classification reporting rate in the pre-2018 era was 28.3% (93 reports), compared

**Table 3** Overall thyroid ultrasound reports characteristics and utility scores.

	n	%
TUS performed by radiology group 1	204	
Nodule characteristic reported		
Size	204	100
Internal content	168	82
Echogenicity	167	82
Calcification	184	90
Margin	150	74
Shape	150	74
Utility score (UtS)		
0	0	0
1	7	3
2	21	10
3	16	8
4	13	6
5	7	3
6	140	69
≥3	176	86
≥4	160	78
Mean UtS by nodule size (0–6)		
<1 cm	4.55	
1–2 cm	5.04	
>2 cm	5.11	
All nodules	5.02	
US performed by radiology group 2	265	
Nodule characteristic reported		
Size	265	100
Internal content	199	75
Echogenicity	195	73
Calcification	213	80
Margin	189	71
Shape	174	65
Utility score (UtS)		
0	0	0
1	14	5
2	27	10
3	37	14
4	25	9
5	12	5
6	150	56
≥3	224	85
≥4	187	71
Mean UtS by nodule size (0–6)		
<1 cm	4.00	
1–2 cm	4.70	
>2 cm	4.70	
All nodules	4.67	
US performed by other radiology groups	430	
Nodule characteristic reported		
Size	410	95
Internal content	195	45
Echogenicity	278	65
Calcification	265	61
Margin	138	32
Shape	72	17
Utility score (UtS)		
0	9	2
1	54	13
2	108	25
3	107	25
4	96	22
5	35	8
6	21	5
≥3	259	60
≥4	152	35
Mean UtS by nodule size (0–6)		
<1 cm	3.00	
1–2 cm	3.26	
>2 cm	2.89	
All nodules	2.97	

to 61.4% (140 reports) in the post-2018 era ( $P < 0.0001$ ). Moreover, the UtS in the pre-2018 era was 2.82, compared to 3.86 in the post-2018 era ( $P < 0.0001$ ). When further divided by radiology groups, radiology group 1 had a pre-2018 mean UtS of 3.62 and 39.4% classification reporting rate compared to a 2018-onwards mean UtS of 5.31 ( $P < 0.0001$ ) and a classification reporting rate of 91.1% ( $P < 0.0001$ ). Radiology group 2 had a pre-2018 mean UtS of 2.8 and a classification reporting rate of 11.5% compared to a 2018-onwards mean UtS of 4.53 ( $P < 0.0001$ ) and a classification reporting rate of 80.0% ( $P < 0.0001$ ). Other radiology groups had a pre-2018 mean UtS of 2.49 and classification reporting rate of 32.2% compared to a 2018-onwards mean UtS of 3.17 ( $P < 0.0001$ ) and classification reporting rate of 45.7% ( $P = 0.008$ ) (Figs 1A, B, 4, 5 and Table 5).

After combining the thyroid nodule patient database with the 2018-onwards thyroid cancer database, there was a total of 570 TUS reports with a mean UtS of 4.58 and a classification reporting rate of 79.8%. In this composite data set, groups 1 and 2 achieved a mean UtS of 5.77 and 5.58, respectively ( $P = 0.048$ ), and other radiology groups had a mean UtS of 3.28 ( $P < 0.0001$  compared to either group 1 or 2). Groups 1 and 2 had a similar ATA or TIRADS classification reporting rate of 97.0 and 93.3%, respectively ( $P = 0.060$ ), while other groups had a significantly lower classification reporting rate of 61.8% ( $P < 0.0001$  compared to either group 1 or 2) (Figs 1B, 4, 5 and Table 6).

The distribution of TUS report quality according to our four categories of TUS reports is shown in Tables 7, 8 and 9, as well as Fig. 6.

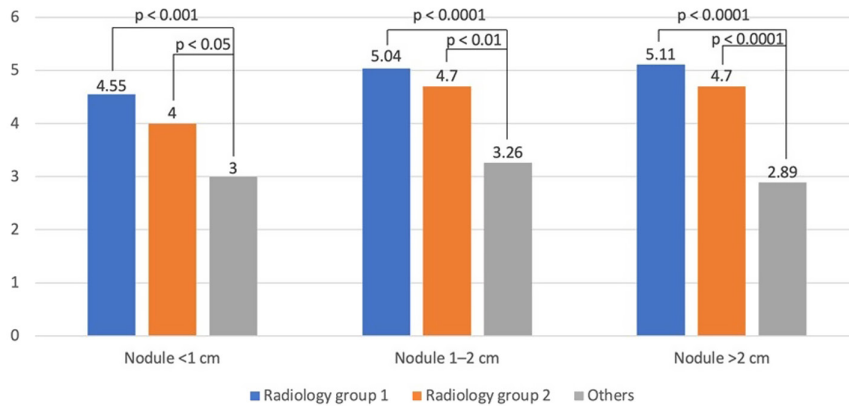
## Discussion

### The UtS and its impact on clinical decision-making

TUS is a key diagnostic tool when deciding on further workup and management for a patient with a thyroid nodule. We previously showed that endocrinologists specializing in the management of thyroid nodules require a TUS with a minimal acceptable UtS of 4 to allow estimation of malignancy-risk category using existing guideline algorithms (ATA, ACR-TIRADS, and other international guidelines) (9).

Our databases only include patients who were selected for FNAs, and 557 (62.0%) of whom were diagnosed with thyroid cancer and thus are more likely to exhibit alarming clinical or radiographic features. Prior to 2018, even among



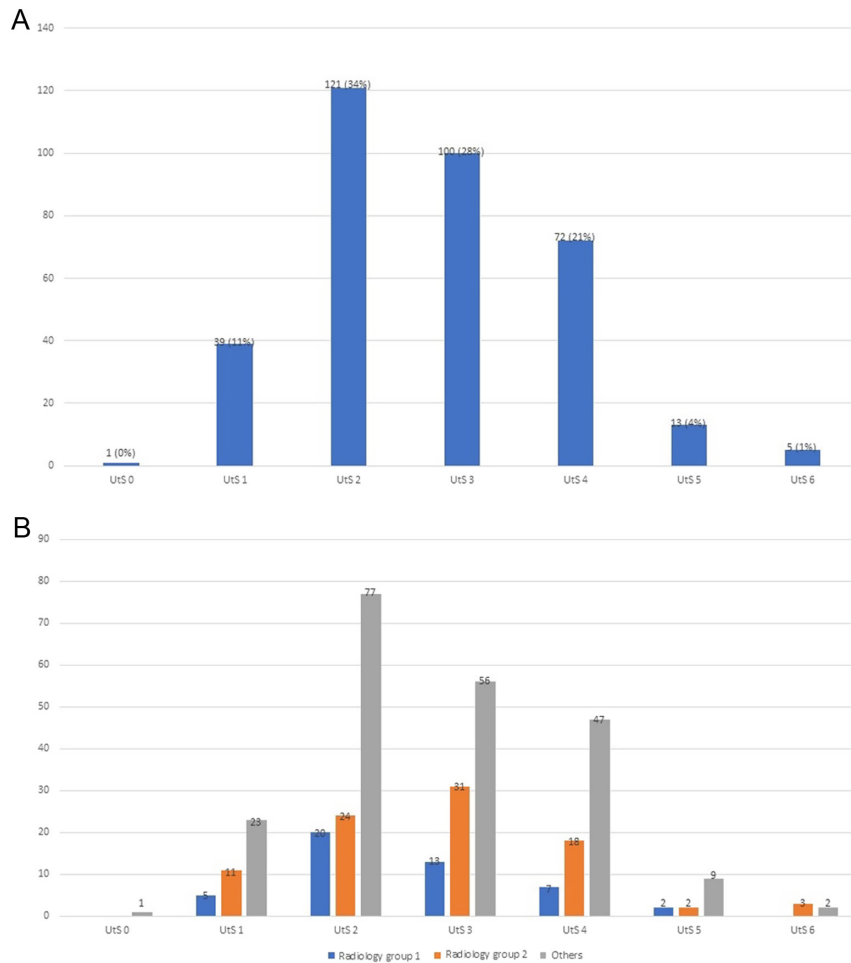


**Figure 2**  
Mean UtS of TUS reports by different radiology groups, divided by the size of the largest nodule detected on the thyroid ultrasound.

our patients diagnosed with thyroid cancer, the mean UtS across all radiology groups was 2.82 and none of the groups individually reached a mean UtS above the threshold of being clinically useful. Additionally, 71.7% of the TUS performed pre-2018 did not include an ATA or TIRADS classification, so clinicians were unable to classify these nodules themselves. We found that as the UtS decreased from 6 to 2, the classification reporting rate also decreased

from 98.3 to 22.4%, which results in TUS reports that do not guide clinical decision-making but rather leave clinicians with diagnostic ambiguity, which may result in unnecessary FNA biopsy, molecular testing, or thyroid surgery.

Previous studies found that when a thyroid nodule FNA result is indeterminate, the ATA and ACR-TIRADS risk classification of the nodule can be very useful in estimating



**Figure 3**  
Thyroid ultrasound reports from both databases without a classification reported. (A) Across different utility scores. (B) By utility scores and by different radiology groups.

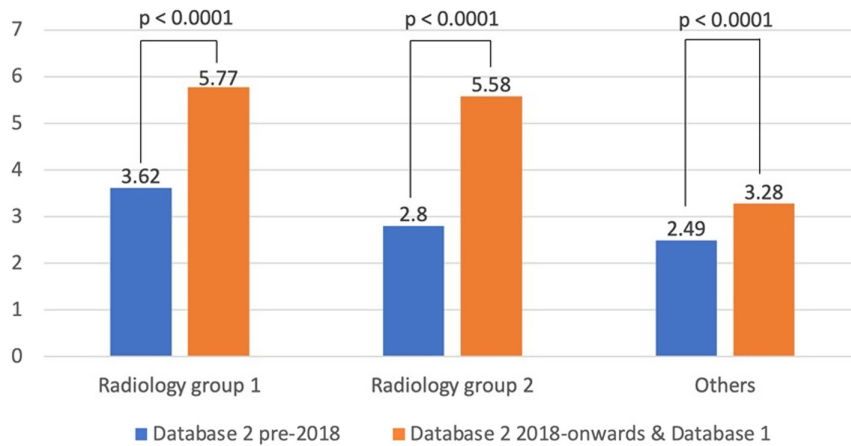
**Table 4** Number and percentage of reports without an ATA or TIRADS classification, divided by UtS and different radiology company.

	UtS 0	UtS 1	UtS 2	UtS 3	UtS 4	UtS 5	UtS 6
Radiology group 1	0	5 (71%)	20 (95%)	13 (81%)	7 (54%)	2 (29%)	0
Radiology group 2	0	11 (79%)	24 (89%)	31 (84%)	18 (72%)	2 (17%)	3 (2%)
Other radiology groups	1 (1%)	23 (43%)	77 (71%)	56 (52%)	47 (49%)	9 (26%)	2 (9%)

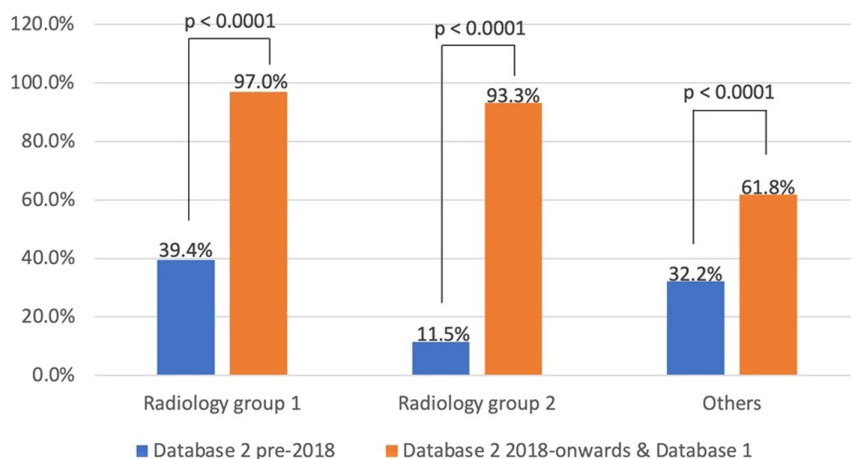
the risk of malignancy (10, 11, 12, 13). Depending on the Bethesda category of the nodule, the negative predictive value of the TUS is as high as 94–100% for nodules classified as very low suspicion or TIRADS 1/2. Similarly, for nodules classified as high suspicion or TIRADS 5, the positive predictive value is as high as 63–100% with an odds ratio of 9.8–19.4 (10, 11, 12, 13). Studies like these further illustrate the value of high-quality TUS reports. Additionally, with the predictive values reported above, it is feasible that the combination of TUS, molecular analysis, and cytology can be used to confidently rule out malignancy in nodules that would otherwise require diagnostic surgery.

### Impact of adherence to TUS guidelines on TUS report quality

Our data demonstrate the improvement in TUS report quality when guidelines are adhered to. In 2018, when radiology groups 1 and 2 began adhering to published guidelines, they both experienced a drastic increase in the mean UtS of their TUS reports and had much higher classification reporting rates. Groups 1 and 2 both reached a mean UtS far above the minimally acceptable score of 4 for estimation of malignancy risk, while other groups remained below that threshold. Additionally, the classification reporting rate also reached above 90% for both groups, while other groups remained low at 61.8%.



**Figure 4**  
Mean UtS of TUS reports pre-2018 vs 2018-onwards composite database, by radiology group.



**Figure 5**  
Classification reporting rate pre-2018 vs 2018-onwards composite database, by radiology group.

**Table 5** Mean utility score and percentage of reports with an ATA or TIRADS classification for the pre-2018 database.

	Mean UtS	Classification reporting rate	Number of reports
Radiology group 1	3.62	39.4%	71
Radiology group 2	2.8	11.5%	87
Other radiology groups	2.49	32.2%	171

Undoubtedly, the dedicated adherence to ATA or TIRADS guidelines has substantially improved the TUS report quality of both groups.

In addition to outperforming other radiology groups when comparing 2018-onwards TUS reports, groups 1 and 2 also had significantly higher quality reports when compared to their own pre-2018 reports. Groups 1 and 2 increased their mean UtS from 3.62 to 5.77 ( $P < 0.0001$ ) and 2.80 to 5.58 ( $P < 0.0001$ ) respectively, and their classification reporting rates increased from 39.4 to 97.0% and 11.5 to 93.3%, respectively. These are very impressive advances, especially for group 2, which had very low-quality reports pre-2018. The borderline statistically significant difference in UtS between groups 1 and 2 in terms of their mean UtS in the 2018-onwards composite database is likely because group 2 began their adherence to guidelines in November of 2018, 10 months later than group 1, and the 2018-onwards composite database captured all scans from that year. As a result, the composite database contains 10 months of TUS reports from radiology group 2 prior to their adherence to the TIRADS reporting guidelines. However, the difference

**Table 6** Mean utility score and percentage of reports with an ATA or TIRADS classification for the 2018-onwards composite database.

	Mean UtS	Classification reporting rate	Number of reports
Radiology group 1	5.77	97.0%	133
Radiology group 2	5.58	93.3%	178
Other radiology groups	3.28	61.8%	259

is not clinically significant, as both groups have reached a high standard of TUS report quality. In contrast, other radiology groups that did not adopt adherence to ATA or TI-RADS guidelines only improved their mean UtS from 2.49 to 3.28 and their classification reporting rate from 32.2 to 61.8%. Although this is a statistically significant increase in both values, their reports still fall short of the minimally useful UtS. Additionally, groups 1 and 2 had a non-significant difference in their ATA or TIRADS classification reporting rate in their 2018-onwards TUS reports.

In a previous baseline study, which looked at the reports of 1930 diagnostic TUS for thyroid nodules, performed between January 1, 2012, and December 31, 2012, in a similar population of patients as our study, we found a mean UtS of 2.57 and at least 86% of the TUS reports did not provide sufficient information for adequate malignancy risk stratification (9). These findings corroborate well with our data and show that there was very little change in the quality of TUS reports between 2012 and 2017 in

**Table 7** Overall quality of thyroid ultrasound reports divided by different radiology groups. Clinically useful reports defined as reports with a utility score of 4, 5, or 6 and non-useful reports defined as reports with a utility score of 0, 1, 2, or 3.

Report type	Radiology group 1	Radiology group 2	Other radiology groups
Clinically non-useful reports; classification reported	6 (3%)	12 (4%)	121 (28%)
Clinically non-useful reports; classification not reported	38 (19%)	66 (25%)	157 (37%)
Clinically useful reports; classification not reported	9 (4%)	23 (9%)	58 (13%)
Clinically useful reports; classification reported	151 (74%)	164 (62%)	94 (22%)
Sum	204 (100%)	265 (100%)	430 (100%)

**Table 8** Quality of thyroid ultrasound reports from pre-2018 database 2, divided by different radiology groups. Clinically useful reports defined as reports with a utility score of 4, 5, or 6 and non-useful reports defined as reports with a utility score of 0, 1, 2, or 3.

Report type	Radiology group 1	Radiology group 2	Other radiology groups
Clinically non-useful reports; classification not reported	35 (50%)	60 (69%)	88 (52%)
Clinically non-useful reports; classification reported	3 (4%)	5 (6%)	39 (23%)
Clinically useful reports; classification not reported	8 (11%)	17 (19%)	28 (16%)
Clinically useful reports; classification reported	25 (35%)	5 (6%)	16 (9%)
Sum	71 (100%)	87 (100%)	171



**Table 9** Quality of thyroid ultrasound reports for the composite of the 2018-onwards database, divided by different radiology groups. Clinically useful reports defined as reports with a utility score of 4, 5, or 6 and non-useful reports defined as reports with a utility score of 0, 1, 2, or 3.

Report type	Radiology group 1	Radiology group 2	Other radiology groups
Clinically non-useful reports; classification not reported	3 (2%)	6 (4%)	69 (27%)
Clinically non-useful reports; classification reported	3 (2%)	7 (4%)	82 (32%)
Clinically useful reports; classification not reported	1 (1%)	6 (3%)	30 (11%)
Clinically useful reports; classification reported	126 (95%)	159 (89%)	78 (30%)
Sum	133 (100%)	178 (100%)	259

our population, despite the ATA guidelines having been published in 2015. And for radiology groups that did not adhere to guidelines, their degree of improvement from Symonds *et al.* 2012 data to our data up to November 2021 is disappointing at best. We are hopeful that with the publication of these data, other radiology groups, who still perform 45% of TUS in our health care region, will follow the exemplary lead of our two local groups, either by competition or regional health authority quality reviews, to also adopt strict adherence to TUS reporting guidelines.

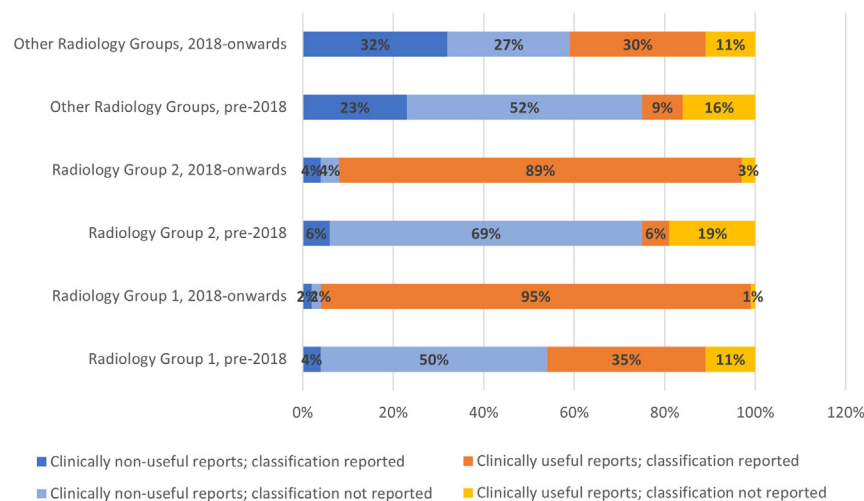
It is noted that the ATA guideline is unable to classify every thyroid nodule, most commonly due to the presence of non-high-suspicion calcifications. These are most often found in nodules that were Bethesda I or II on biopsy (14). Our databases had one report of a non-classifiable nodule, likely due to the small sample size of reports that used the ATA guidelines and their low prevalence among Bethesda III, IV, V, and VI nodules.

A potential limitation is that the ACR-TIRADS guideline has been mainly tested for papillary thyroid cancers and may thus not be accurate for other forms of thyroid cancer (15). However, in our cohorts, non-PTC thyroid cancers make up a small percentage of thyroid malignancies and likely did not affect our data significantly. Another

limitation is that our TUS reports were not retrospectively analyzed to ensure the accuracy of the TUS reports. Since our databases were selected for patients with high-risk features, we had few TUS reports of nodules that were Bethesda I or II and are unable to comment on the effect of the application of the ATA or ACR-TIRADS guidelines to these nodules.

### Conclusion

Our paper demonstrates that dedicated effort towards adhering to published TUS reporting guidelines can lead to significant improvements in the quality of TUS reports and that without such efforts, TUS report quality would continue to fall short of the minimally acceptable standards even years after the release of TUS reporting guidelines. We are optimistic that other radiology groups will begin to follow suit in improving their TUS report quality, as this will reduce diagnostic ambiguity and improve clinician’s decision-making capacity, which will lead to an overall reduction in unnecessary FNA biopsy and diagnostic surgery.



**Figure 6** Pre-2018 vs 2018-onwards distribution of thyroid ultrasound reports by category of clinical usefulness and classification reporting, by different radiology groups.

### Supplementary materials

This is linked to the online version of the paper at <https://doi.org/10.1530/ETJ-22-0035>.

### 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

For this study was obtained from: Cardiometabolic fund University of Calgary; Sanofi; EFW Radiology and a Canadian Society of Endocrinology and Metabolism/Thyroid Foundation of Canada Research Award.

### Acknowledgements

The authors thank Dr Brendan Diederichs and Dr Chris Cheung from Mayfair for providing us with the details for their ACR-TIRADS guideline-based implementation of thyroid nodule ultrasound malignancy risk stratification.

## References

- 1 Hegedüs L. Thyroid ultrasound. *Endocrinology and Metabolism Clinics of North America* 2001 **30** 339–360. ([https://doi.org/10.1016/s0889-8529\(05\)70190-0](https://doi.org/10.1016/s0889-8529(05)70190-0))
- 2 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>)
- 3 Gharib H, Papini E, Garber J, Duick D, Harrell R, Hegedus L, Paschke R, Valcavi R, Paolo V & American Association of Clinical Endocrinologists. American College of Endocrinology, and Associazione Medici Endocrinologi Medical Guidelines for clinical practice for the diagnosis and management of thyroid nodules – 2016 update appendix. *Endocrine Practice* 2016 **22** 1–60. (<https://doi.org/10.4158/ep161208.g>)
- 4 Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, Cronan JJ, Beland MD, Desser TS, Frates MC, *et al.* ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee. *Journal of the American College of Radiology* 2017 **14** 587–595. (<https://doi.org/10.1016/j.jacr.2017.01.046>)
- 5 Pandya A, Caoili EM, Jawad-Makki F, Wasnik AP, Shankar PR, Bude R, Haymart MR & Davenport MS. Retrospective cohort study of 1947 thyroid nodules: a comparison of the 2017 American College of Radiology TI-RADS and the 2015 American Thyroid Association classifications. *American Journal of Roentgenology* 2020 **214** 900–906. (<https://doi.org/10.2214/AJR.19.21904>)
- 6 Thejeel B, Rahimi B, Seidler M, Al-Agha R & Fung C. Evaluation of thyroid ultrasound report quality and assessing effect of adherence to risk stratification criteria on referral for thyroid nodule biopsy. *Canadian Association of Radiologists Journal* 2021 **72** 234–241. (<https://doi.org/10.1177/0846537119900634>)
- 7 Maniuk T, Kielar AZ, O'Sullivan JP, El-Khodary M, Lochnan H, Purgina B & Odell MJ. Effect of implementing community of practice modified thyroid imaging reporting and data system on reporting adherence and number of thyroid biopsies. *Academic Radiology* 2018 **25** 915–924. (<https://doi.org/10.1016/j.acra.2017.12.009>)
- 8 Griffin AS, Mitsky J, Rawal U, Bronner AJ, Tessler FN & Hoang JK. Improved quality of thyroid ultrasound reports after implementation of the ACR thyroid imaging reporting and data system nodule lexicon and risk stratification system. *Journal of the American College of Radiology* 2018 **15** 743–748. (<https://doi.org/10.1016/j.jacr.2018.01.024>)
- 9 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>)
- 10 Trimboli P, Deandrea M, Mormile A, Ceriani L, Garino F, Limone P, Garino F, Limone P & Giovannella L. American Thyroid Association ultrasound system for the initial assessment of thyroid nodules: use in stratifying the risk of malignancy of indeterminate lesions. *Head and Neck* 2017 **40** 722–727. (<https://doi.org/10.1002/hed.25038>)
- 11 Barbosa TLM, Junior COM, Graf H, Cavalvanti T, Trippia MA, da Silveira Ugino RT, de Oliveira GL, Granella VH & de Carvalho GA. ACR TI-RADS and ATA US scores are helpful for the management of thyroid nodules with indeterminate cytology. *BMC Endocrine Disorders* 2019 **19** 112. (<https://doi.org/10.1186/s12902-019-0429-5>)
- 12 Grani G, Lamartina L, Ascoli V, Bosco D, Nardi F, D'Ambrosio F, Rubini A, Giacomelli L, Biffoni M, Filetti S, *et al.* Ultrasonography scoring systems can rule out malignancy in cytologically indeterminate thyroid nodules. *Endocrine* 2017 **57** 256–261. (<https://doi.org/10.1007/s12020-016-1148-6>)
- 13 Ahmadi S, Herbst R, Oyekunle T, Jiang X, Strickland K, Roman S & Sosa JA. Using the ATA and ACR ti-rads sonographic classifications as adjunctive predictors of malignancy for indeterminate thyroid nodules. *Endocrine Practice* 2019 **25** 908–917. (<https://doi.org/10.4158/EP-2018-0559>)
- 14 Kobaly K, Kim CS, Langer JE & Mandel SJ. Macrocalcifications do not alter malignancy risk within the American Thyroid Association sonographic pattern system when present in non-high suspicion thyroid nodules. *Thyroid* 2021 **31** 1542–1548. (<https://doi.org/10.1089/thy.2021.0140>)
- 15 Trimboli P, Castellana M, Piccardo A, Romanelli F, Grani G, Giovannella L & Durante C. The ultrasound risk stratification systems for thyroid nodule have been evaluated against papillary carcinoma. A meta-analysis. *Reviews in Endocrine and Metabolic Disorders* 2021 **22** 453–460. (<https://doi.org/10.1007/s11154-020-09592-3>)

Received in final form 2 May 2022

Accepted 5 May 2022

Accepted Manuscript published online 6 May 2022