Thyroid nodules are very common: 4-8% of the general population have palpable nodules while 19-67% will have nodules sonographically. The vast majority of these thyroid nodules are benign with only a minority (2-12%) being malignant. The increasing use of sonography has resulted in what is often referred to as a “thyroid storm”. However, despite this imaging-induced epidemic with an increase in the incidence of thyroid cancers, often small indolent papillary carcinomas, the death rate has remained unchanged suggesting significant over-diagnosis and over-treatment.¹

Multiple sonographic features have been described and evaluated in the search for a method for reliably differentiating benign from malignant nodules. However, despite very extensive literature on the subject, there are unfortunately no sonographic features that are absolutely pathognomonic for benignity or malignancy. While there are some features that are sensitive, there is often significant overlap between benign and malignant nodules impacting on their specificity. The corollary is also the case with the most specific features, often occurring relatively infrequently, thereby impacting on their overall sensitivity and clinical usefulness.

Ultimately, fine needle aspiration cytology (FNAC) is often required to characterise nodules as being benign or malignant. The result is a large number of “unnecessary” biopsies and in the USA over-diagnosed thyroid cancer (ie no symptoms or death if untreated) accounts for 70-80% of thyroid carcinomas, often small indolent papillary carcinomas, the death rate has remained unchanged suggesting significant over-diagnosis and over-treatment.²

In 2015, the American College of Radiology Thyroid Imaging, Reporting and Data System (TI-RADS)³ committee published a white paper on a thyroid ultrasound reporting lexicon with the aim of providing a set of well-defined sonographic features that could be applied to every thyroid nodule in a structured fashion. Ultimately five categories were defined: composition, echogenicity, shape, margins, echogenic foci, and size. Terms were chosen on the basis of their consistent ability to differentiate benign from malignant nodules and based on their reproducibility among readers.

As part of the culling process for the ACR lexicon certain descriptors were excluded on the basis of their poor discriminatory ability. A number of these had formed part of previously published guidelines: vascularity (AACE/AME/ETA 2011) and clinical red flags (ATA 2009, AACE/AME/ETA 2011). Lymphadenopathy (ASRU 2005, ATA 2009, AACE/AME/ETA 2011) and growth (ASRU 2005, Korean 2011), although included in the new ACR TI-RADS white paper discussion, do not actually form part of the scoring system. There is, however, a recommendation for the biopsy of any suspicious lymph node in addition to up to any two suspicious thyroid nodules.

There have been previous “TI-RADS” publications modelled around the very successful BI-RADS for breast imaging. The original publication by Horvath (2009)⁴ uses descriptors very similar to the 2015 ACR lexicon. There were subsequent publications by Park (2009),⁵ Kwak (2011),¹² Sanchez (2014)¹ⁱ and Russ (2016).¹⁰ However, the scoring and risk stratification systems used vary between studies, which may explain their lack of generalised adoption into clinical practice.

In April 2017, the ACR published a further white paper with the aim of defining a risk stratification system for thyroid nodules (ACR TI-RADS)⁶ to guide decisions regarding FNA and follow-up. The aim is to identify nodules that have a greater likelihood of biologically significant cancer and reduce the number of unnecessary biopsies, while at the same time being simple to use, applicable to all thyroid nodules and all hospital settings while also using the ACR TI-RADS lexicon. As with PI-RADSv2 for prostate MRI the aim is not to detect all cancer but to detect all clinically significant cancer, while simultaneously minimising the need for biopsy of benign nodules.

All sonographic features are not equal, with some features being mildly, moderately and highly suspicious for malignancy. As a consequence and as with earlier versions of “TI-RADS”, but unlike some of the older guidelines, consideration and points are only given for suspicious features. No points or consideration are given to benign findings. Unlike earlier versions of “TI-RADS” there is a weighting system that gives more points for more specific and suspicious features. Nodules are scored within the five sonographic categories but otherwise the new ACR TI-RADS does not rely on grouping ultrasound features into specific patterns (figure 1 – used with ACR permission).

Figure 2 illustrates how the new ACR TI-RADS scoring system is applied clinically. Each thyroid nodule, up to a maximum of four per patient, is scored based on the five ACR lexicon sonographic categories: Composition, echogenicity, shape, margin and echogenic foci (figure 1). The points, from the weighted scoring system, are summed to give an ACR TI-RADS score (score of 10 in this case) and the nodule assigned a benign (TR1) or non-suspicious (TR2) category.

Dr Declan Sheppard
Consultant radiologist, University Hospital Galway
Honorary senior lecturer, National University of Ireland, Galway
dclan.sheppard@hse.ie

The new ACR TI-RADS: A step towards a risk stratified approach to thyroid FNA?
which do not require FNAC or follow-up, or alternatively a mildly (TR3), moderately (TR4) or highly suspicious (TR5) as in this case based on a score ≥7) category. The decision regarding FNAC (TR3 ≥2.5cm, TR4 ≥1.5cm, TR5 ≥1cm) (Yes in this case based on a TR5 category and >1cm) and follow-up (TR3 ≥1.5cm, TR4 ≥1cm, TR5 ≥0.5cm) is based on the maximum diameter. Unlike some of the earlier versions of TI-RADS there are no subcategories (eg no T4a or b).

The differing size cut-offs are based on the fact that prognosis is inversely related to size, with larger lesions doing worse. As a consequence, larger more benign appearing nodules may need to be biopsied while smaller more suspicious lesions may not. Additionally, the system has higher size thresholds for suggesting FNA than the ATA and Korean Society of Thyroid radiology guidelines in mild and moderately suspicious nodules in the hope of cutting down on the number of unnecessary biopsies. The ACR TI-RADS guidelines, in keeping with prior guidelines, would not recommend the biopsy of sub-centimetre nodules even if otherwise suspicious. The ACR TI-RADS committee recommends a maximum of two biopsies per patient, which should be based on the sonographically most suspicious nodules (ie those with the highest TR scores) rather than based on size, and consequently discourages the use of the term “dominant nodule” as it implies that size is more important than suspicious sonographic appearances.

While an increasing score reflects an increasing risk of malignancy the underlying low prevalence of malignancy and the low pre-test probability would suggest that rather than defining the probability of malignancy the aim should be to identify those cases that warrant biopsy.

There is a slightly lower size threshold for follow-up, which aligns with the trend for observation ("watchful waiting") in small, even suspicious nodules where early surgery offers no additional survival benefit. The ACR TI-RADS committee recommends scanning intervals of not less than one year: TR3 (one, three and five years); TR4 (one, two, three and five years) and TR5 (one, two, three, four and five years) with discharge at five years if stable over the full five-year period.

A benefit of the new ACR TI-RADS system and lexicon is that it encourages a systematic assessment of a thyroid nodule and facilitates the use of structured reporting templates. The subsequent points-based system, as per BI-RADS, allows for easier review and comparison of cases and ultimately will facilitate research.

There have already been some initial publications around the use of the new guidelines with promising early results.13,17

In conclusion, the new ACR TI-RADS proposes a risk stratified approach to thyroid nodule FNA and follow-up. Our initial experience17 has found that it is initially slower to use and may not result in the reduced number of biopsies originally envisaged.

We reprogrammed a locally developed 'app' that allows for simultaneous scoring of multiple guidelines8-12 based around the 2015 ACR TI-RADS lexicon9 to additionally and simultaneously score the new ACR TI-RADS13 but also a locally developed slimmed down version "TI-RADS lite". Our results with this app-based approach have been very encouraging with regard to ease of use and accuracy and our TI-RADS lite (unpublished data) does appear to successfully reduce the number of unnecessary biopsies while maintaining high sensitivities.

References

Figure 1

Figure 2