

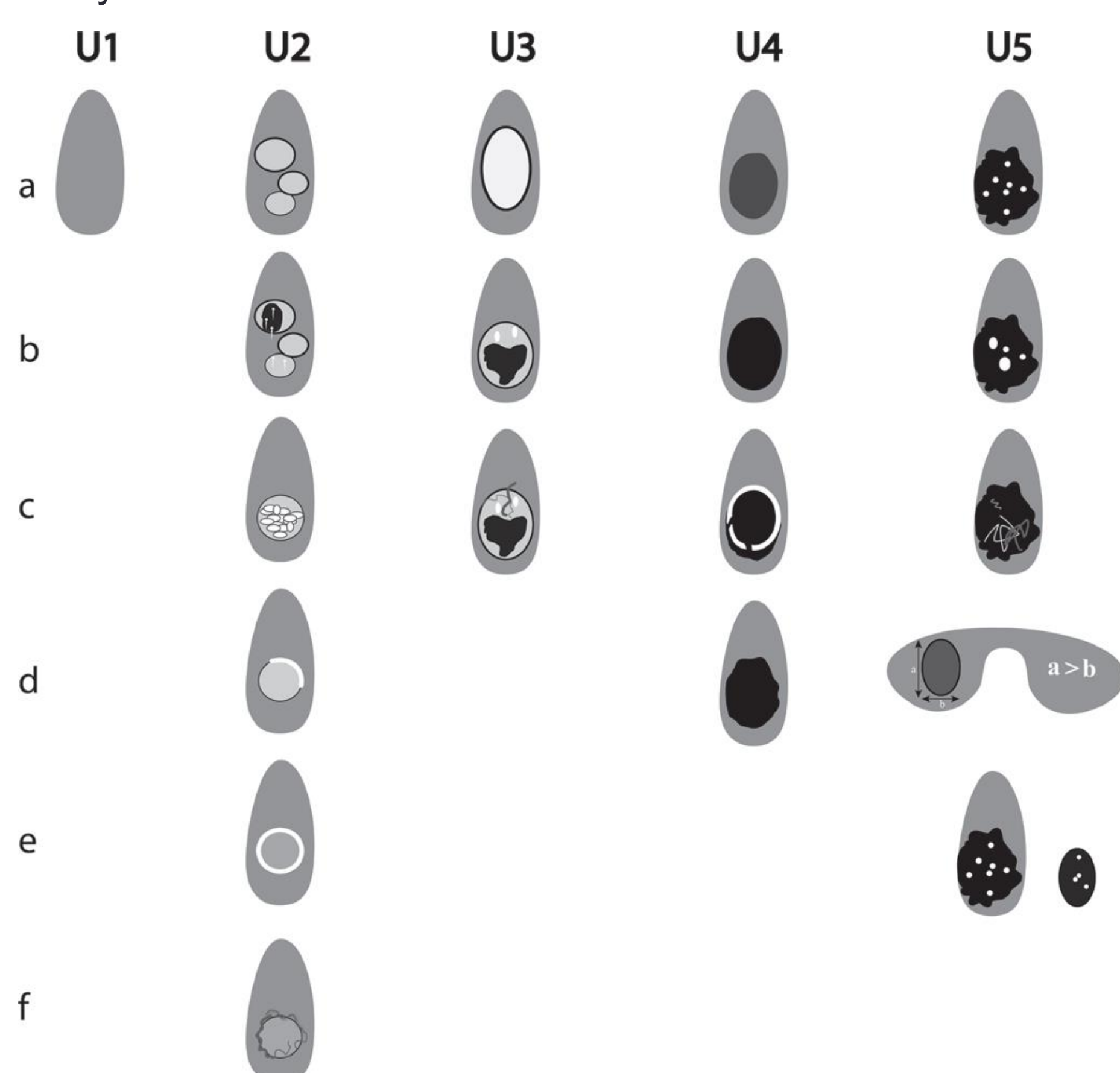
# Correlation between the British Thyroid Association ultrasound grading of thyroid nodules and histopathology specimens, with assessment of inter-rater agreement: a one-year institutional experience

Olivia Francies, Susan Jawad, Simon Morley, Sally Daniels, Sofia Otero



## BACKGROUND

The British Thyroid Association (BTA) guidelines 2014 introduced an ultrasound (US) classification for thyroid nodules: U1 – U5



**U1. Normal thyroid**

**U2. Benign:**  
(a) halo, iso-/mildly hyper-echoic  
(b) cystic +/- ring down sign (colloid)  
(c) micro-cystic/spongiform  
(d & e) peripheral egg shell calcification  
(f) peripheral vascularity

**U3. Indeterminate/Equivocal:**  
(a) hyper-echoic, solid, halo  
(b) ?hypo-echoic, equivocal echogenic foci  
(c) mixed/central vascularity

**U4. Suspicious:**  
(a) solid, hypo-echoic (cf thyroid)  
(b) solid, very hypo-echoic (cf strap muscle)  
(c) disrupted peripheral calcification, hypo-echoic  
(d) lobulated outline

**U5. Malignant:**  
(a) solid, hypo-echoic, lobulated/irregular outline, micro-calcification  
(b) solid, hypo-echoic, lobulated/irregular outline, globular calcification  
(c) intra-nodular vascularity  
(d) shape (taller >wide) (AP>TR)  
(e) characteristic associated lymphadenopathy

## RESULTS

- Total thyroidectomy specimens 142
- Total number with US at UCLH 126 (+1 external)
- 6 samples excluded for Graves' disease
- Total number of samples included **121**

1) Malignancy rate for U grade (as given on the original report)

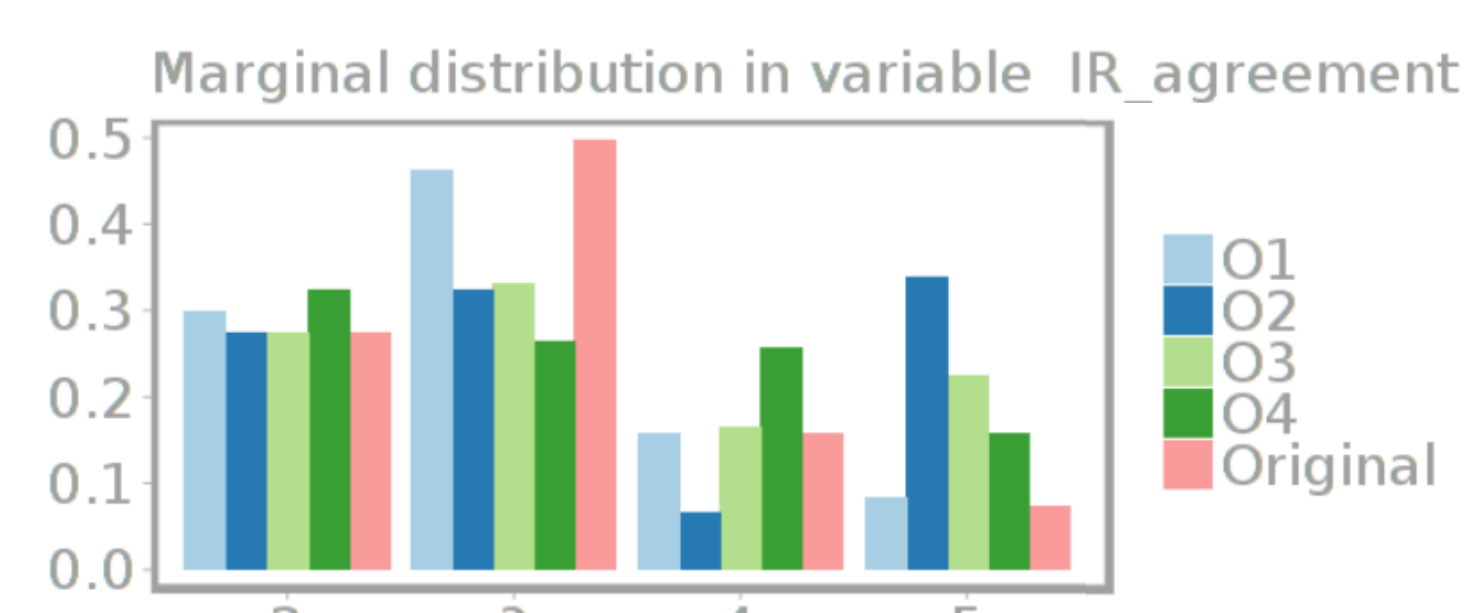
U2	3%
U3	40%
U4	79%
U5	100%

2) Sensitivity and specificity for detection of malignant nodules

	Sensitivity (%)	Specificity (%)
Original report	98.0	44.4
Rater 1	95.9	47.2
Rater 2	95.9	43.0
Rater 3	95.8	43.0
Rater 4	91.8	48.6
	<b>High</b>	<b>Moderate</b>

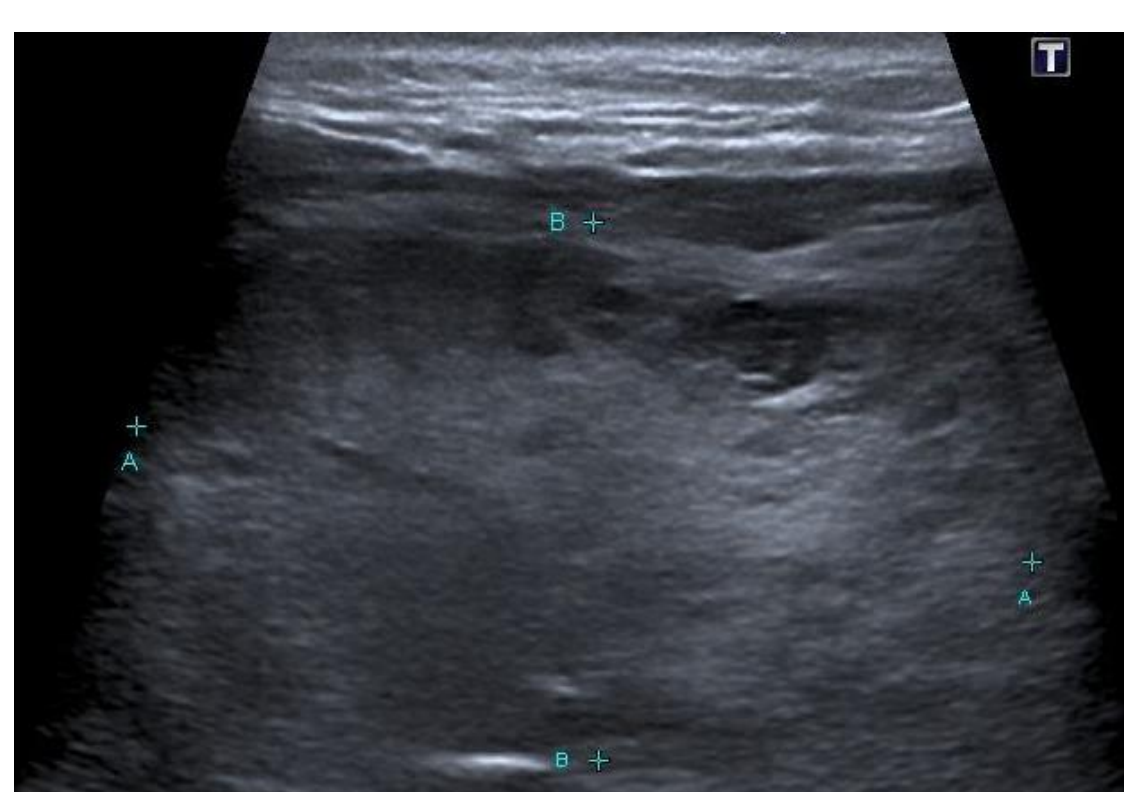
3) Inter-rater agreement: Fleiss kappa score of 0.51 = **Moderate**

Data	
5 raters and 121 cases	
1 variable with 605 decisions in total	
1 missing data	
1: IR_agreement	
Fleiss	Krippendorff
A_obs = 0.647	D_obs = 0.353
A_exp = 0.282	D_exp = 0.72
Kappa = 0.509	Alpha = 0.509



The graph demonstrates good inter-rater agreement for U2 nodules but greater variability for U3-5 nodules.

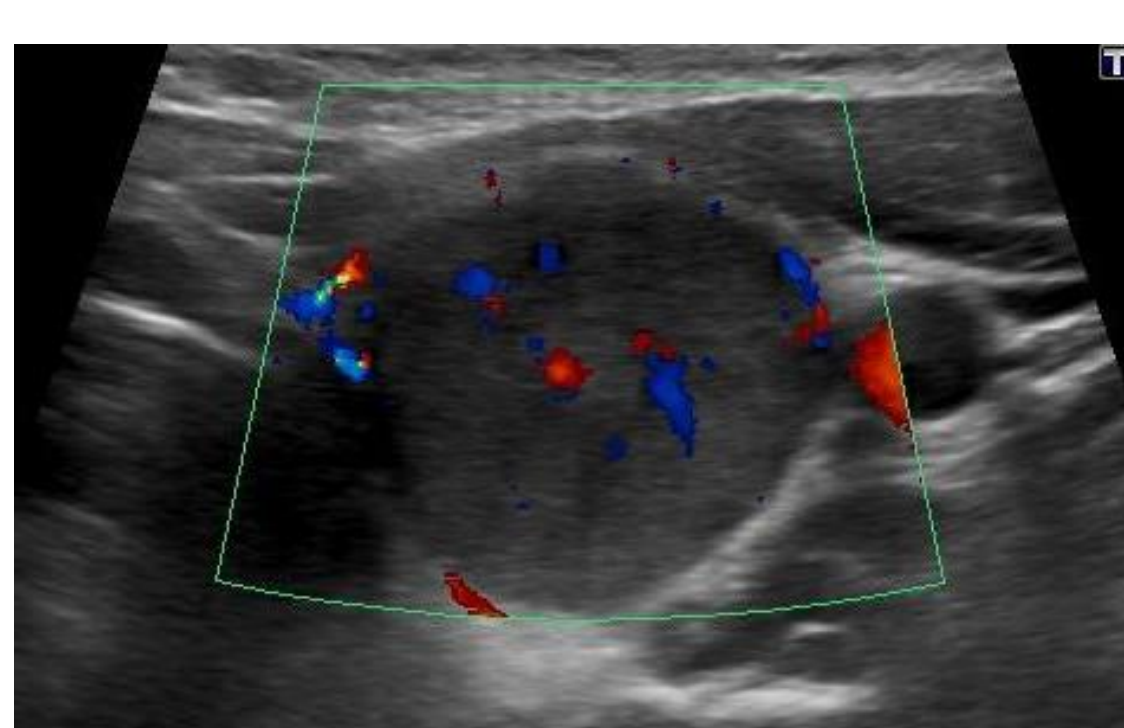
### Example cases



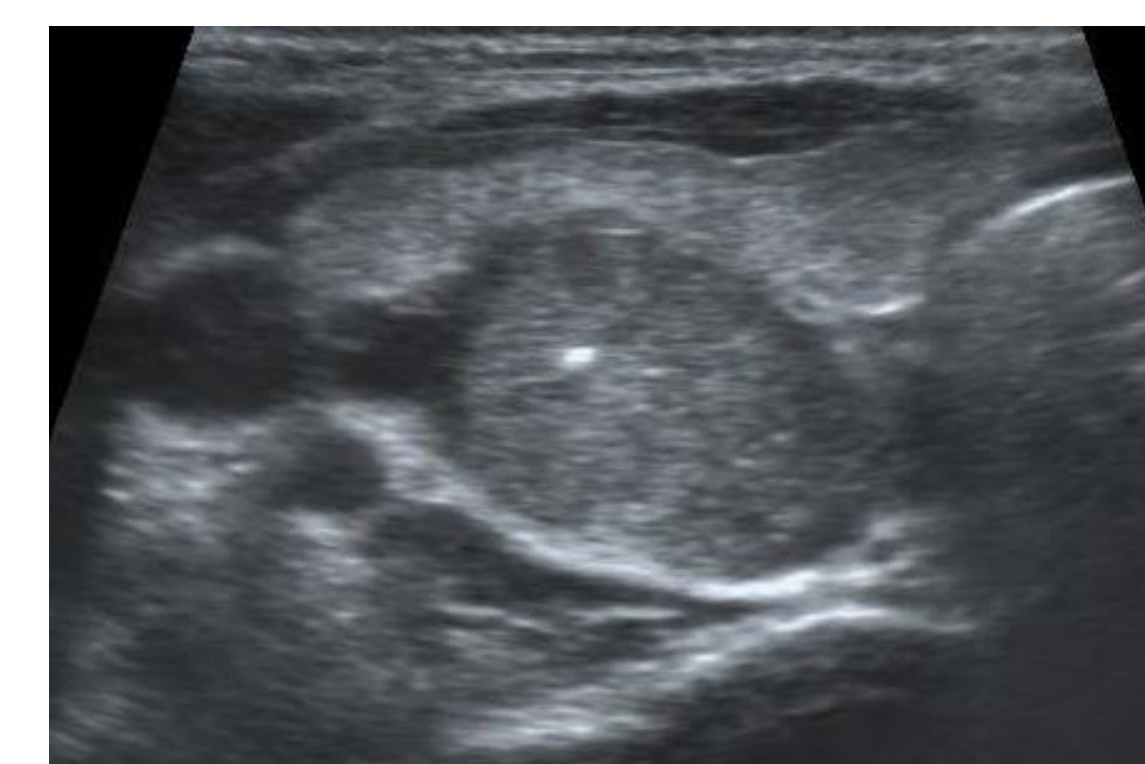
Isoechoic nodule with micro-cystic change  
**Uniformly graded U2**  
= benign multinodular goitre



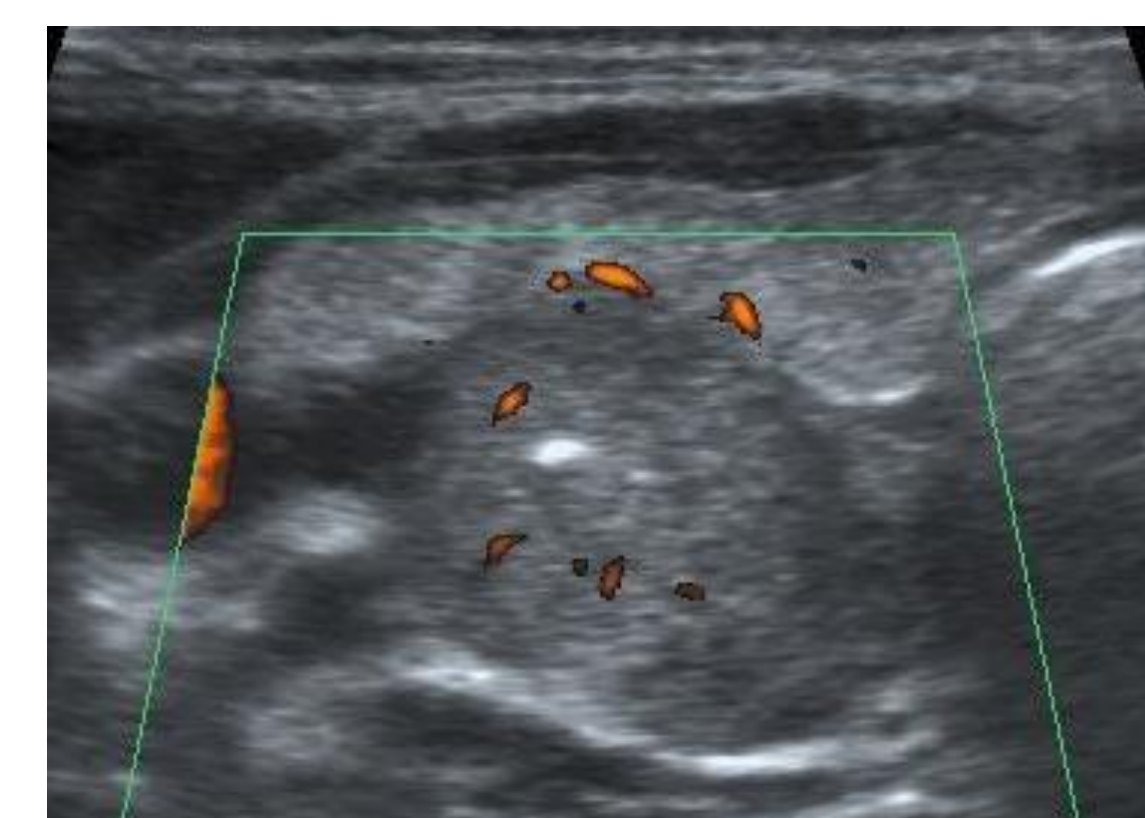
Solid, hypoechoic nodule with lobulated outline  
**Graded U3 & U4**  
= follicular variant of papillary cell carcinoma



Solid, hypoechoic nodule with intra-nodular vascularity  
**Graded U3 - U5**  
= minimally invasive follicular cell carcinoma



Solid, hypoechoic nodule with an irregular outline, internal vascularity and containing echogenic foci consistent with calcification  
**Graded U3 - U5**  
= papillary thyroid carcinoma



Solid, hypoechoic nodule  
**Graded U3 & U4**  
= follicular adenoma



## CONCLUSIONS

Malignancy rates for each U category are in line with previous scoring systems

- High sensitivity
  - Only moderate specificity
- Moderate inter-rater agreement overall
- Inter-rater agreement for U2 is good
  - Inter-rater agreement for other U grades is variable

## RECOMMENDATIONS

- Use a simplified scoring system for thyroid nodules of benign, indeterminate and suspicious for malignancy
- Calculate the sensitivity and specificity and inter-rater agreement for this simplified scoring system
- If this is equivalent or better than the current U1 – U5 system, consider implementing this scoring system
- Repeat the audit after 6 months
- Disseminate the results directly to the head and neck radiology team

## REFERENCES

- Perros P, Colley S, Boelaert K et al. British Thyroid Association Guidelines for the Management of Thyroid Cancer. Third edition. Clinical Endocrinology. 2014; 81(Suppl 1): 1-122. Chapter 7 p14-18. John Wiley & Sons Ltd. Wiley Blackwell
- Lee YH et al. Differentiation between benign and malignant thyroid solid nodules using an ultrasound classification system. Korean Journal of Radiology. 2011; 12: 559-567
- Kwak JY et al. Thyroid Imaging Reporting and Data System for US Features of Nodules: A Step in Establishing Better Stratification of Cancer Risk. Radiology. 2011; 260(3): 892-9. doi: 10.1148/radiol.11110206

## AIMS

- Calculate our institutional rate of malignancy for each U category
- Calculate the sensitivity and specificity for the detection of malignant thyroid nodules
- Calculate inter-rater agreement for the US classification (U2 – U5) of thyroid nodules

## METHODOLOGY

- All total and hemi-thyroidectomies for a 12-month period between 2015–2016 were considered
  - Samples excluded: completion thyroidectomies for malignancy and diffuse non-nodular disease (Graves')
- The U grade in the original US report was correlated with the final histopathology
  - For multinodular thyroids, the highest U grade given in the report was used for analysis
  - Incidental micro-papillary carcinomas (where these nodules were not seen or classified at US) were not included in the data
- The malignancy rate for each U category was calculated from the original report
- 4 raters (3 Consultant Radiologists and 1 Sonographer all with an interest in head & neck imaging) retrospectively assessed the stored US images and allocated a U grade for the most concerning nodule(s)
- Sensitivity and specificity for malignancy for the original report and each of the 4 raters was calculated
- Inter-rater agreement was calculated for the original report and each of the 4 raters using a Fleiss Kappa score