National Institute for Health and Care Excellence

FINAL

Thyroid disease: assessment and management

[N] Imaging for fine needle aspiration

NICE guideline NG145

Diagnostic evidence review underpinning recommendations 1.9.1 to 1.9.6 in the guideline. See also evidence review O

2019

FINAL

Developed by the National Guideline Centre, hosted by the Royal College of Physicians

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and, where appropriate, their carer or guardian.

Local commissioners and providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN: 978-1-4731-3595-6

Contents

1 Imaging for Fine Needle Aspiration ... 5
1.1 Review question: Which imaging tests should be requested (for thyroid enlargement)? Which people with structural abnormalities should have a fine-needle aspiration? ... 5
1.2 Introduction ... 5
1.3 PICO table ... 5
1.4 Clinical evidence ... 6
1.4.1 Included studies ... 6
1.4.2 Excluded studies ... 6
1.4.3 Summary of clinical studies included in the evidence review ... 7
1.4.4 Quality assessment of clinical studies included in the evidence review ... 19
1.5 Economic evidence ... 23
1.5.1 Included studies ... 23
1.5.2 Excluded studies ... 23
1.5.3 Health economic modelling ... 23
1.5.4 Resource costs ... 23
1.6 Evidence statements ... 23
1.6.1 Clinical evidence statements ... 23
1.6.2 Health economic evidence statements ... 24
1.7 The committee's discussion of the evidence ... 25
1.7.1 Interpreting the evidence ... 25
1.7.2 Cost effectiveness and resource use ... 26
1.7.3 Other factors the committee took into account ... 27
References ... 28
Appendices ... 55
Appendix A: Review protocols ... 55
Appendix B: Literature search strategies ... 60
Appendix C: Clinical evidence selection ... 69
Appendix D: Clinical evidence tables ... 71
Appendix E: Coupled sensitivity and specificity forest plots and sROC curves ... 158
Appendix F: Health economic evidence selection ... 162
Appendix G: Health economic evidence tables ... 163
Appendix H: Health economic analysis ... 164
Appendix I: Excluded studies ... 165

1 Imaging for Fine Needle Aspiration

1.1 Review question: Which imaging tests should be requested (for thyroid enlargement)? Which people with structural abnormalities should have a fine-needle aspiration?

1.2 Introduction

Patients with thyroid enlargement usually present due to mass effect symptoms or cosmetic embarrassment or are identified following incidental imaging findings on investigation of other pathology. With the knowledge that the majority of thyroid disease is benign, imaging of enlargement is usually only performed if there are features of concern – for example vocal cord palsy, prior radiation or risk factors for malignancy. With incidental thyroid enlargement identified on prior cross sectional imaging, ultrasound is recommended where there is extra thyroidal extension, invasion of adjacent structures or abnormal local neck nodes.

The aim of imaging is to assess risk of malignancy, guide percutaneous sampling if it is indicated and assess the extent of glandular enlargement in patients where surgical intervention is considered. Ultrasound, performed by appropriately trained and practiced specialists is readily available, inexpensive and a sensitive modality for gland assessment. Neither computerized tomography (CT) nor Magnetic resonance imaging (MRI) are as sensitive at gland assessment (nor do they allow for real time image guided tissue sampling when desired) although CT may be used to assess mass effect on the trachea and retrosternal extension.

Tissue sampling is routinely performed when malignancy is suspected although there are a number of different assessment criteria to suggest malignant potential on ultrasound.

1.3 PICO table

For full details see the review protocol in Appendix A:.

Table 1: PICO characteristics of review question

Population	People presenting with euthyroid thyroid enlargement being investigated for possible malignancy.
Target condition	Malignancy
Index tests	Ultrasound scan CT scan MRI scan
Reference standard	Malignant status as confirmed by biopsy/subsequent development of cancer in case of false negatives that do not receive biopsy
Statistical measures [or] Outcomes	Sensitivity Specificity PPV NPV Sensitivity prioritised
Study design	Diagnostic accuracy studies Prospective studies prioritised, retrospective studies included if insufficient prospective studies identified.

1.4 Clinical evidence

1.4.1 Included studies

Forty two observational (prospective and retrospective) studies were included in the review;[6, 9, 36, 48, 67, 80, 86, 87, 92, 94, 97, 114, 116, 122, 129, 133, 138, 150, 166, 170, 175, 179, 187, 188, 190, 207, 215, 218, 231, 276, 281, 309, 323, 324, 329, 331, 332, 346, 353, 354, 363, 367]. Evidence from these studies is summarised in the clinical evidence summary below (Table 3). All studies looked at the diagnostic accuracy of ultrasound using different sonographic criteria.

Thirty-nine studies conducted in adults assessed ultrasound classified according to the British Thyroid Association (BTA, 2 studies), different version of the Kim criteria (10 studies), Society of Radiologists in Ultrasound (SRU, 3 studies), American Association of Clinical Endocrinologists/American College of Endocrinology/Associazione Medici Endocrinologi (AACE/ACE/AME, 5 studies), American Thyroid Association (ATA, 13 studies), Korean Society of Thyroid Radiology (KSThR, 3 study), different versions of the Thyroid Imaging Reporting and Data System (TIRADS, 31 studies), TIRADS combined with contrast-enhanced US (CEUS) parameter ratios (2 studies) and National Comprehensive Cancer Network (NCCN, 1 study). Of those, two studies assessed the diagnostic accuracy of gray-scale ultrasound combined with power Doppler ultrasound using the Kim criteria. One study assessed the diagnostic accuracy of ultrasound combined with elastography using the Kim criteria combined with the Rago and Asteria criteria.

Three studies assessed the diagnostic accuracy of ultrasound using Kwak's TIRADS (one study) and the ATA guidelines (three studies) in children.

See also the study selection flow chart in Appendix C:, sensitivity and specificity forest plots in Appendix E:, and study evidence tables in Appendix D:

1.4.2 Excluded studies

See the excluded studies list in Appendix I:.

1.4.3 Summary of clinical studies included in the evidence review

Table 2: Summary of studies included in the evidence review

Study	Population	Target condition	Index test	Reference standard	Comments
Ahn 2010 [6]	Patients: n=1318 mean age 46.3 years; 1398 nodules confirmed with FNAB or surgery South Korea	Thyroid cancer	Ultrasound (that should lead to FNAB) under different criteria: • Kim • Society of Radiologists in Ultrasound • American association of Clinical Endocrinologists	Surgical or cytologic findings if the patient did not undergo surgery (FNAB or surgery) Surgery was performed for 455 nodules	
Alahm 2014 [9]	Patients: n=100; mean age (SD) 41.77 (12.31) Pakistan	Thyroid cancer	Ultrasound	FNAB	
Chen 2019 [36]	Patients: n=1092; mean age (SD): 46.92 (13.59) Mean nodule size (SD): 19.63 (13.90)mm China	Thyroid cancer	Ultrasound Classified using the ACR-TI-RADS	FNAB	
Creo 2018 [48]	Children: n=112; mean age (SD): 15.5 (3.2) years USA	Thyroid cancer	US+ USD (Gray-scale US with colour Doppler) Classified using the 2015 ATA TIRADS	UG FNA Using the Bethesda System for Reporting Thyroid Cytology	Children

Study	Population	Target condition	Index test	Reference standard	Comments
Grani 2019 [80]	Patients: n=477; mean age (SD): 55.9 (13.9) years Italy	Thyroid cancer	Ultrasound Classified using the ACR TIRADS, AACE/ACE/AME, ATA, EU-TIRADS, K-TIRADS	UGFNAB Using the Italian Consensus for Thyroid Cytopathology Histological examination for nodules that had undergone surgery. FNA cytology for nodules that had not been managed surgically; using the Bethesda System	
Farihah 2018 [67]	Patients: n=91 (104 nodules) ; mean age (range): 54.7 (27-80) Malaysia	Thyroid cancer	Ultrasound Classified using the BTA Guidelines (test positive: U4-5; test negative U2-3)	UGFNAC and histopathology (for cases that were inadequate, indeterminate or suspicious of malignancy)	
Ha 2016 [86]	Patients: n=750 (902 nodules); mean age (range): 49.2 (9-81) Mean nodule size (SD; range): 1.5 cm (1.1; 0.5-10 cm)	Thyroid cancer	Ultrasound Classified using K-TIRADS	FNA or core needle biopsy (CNB) or surgery Using the Bethesda System for Cytological classification of Thyroid	Multicentre study (4 hospitals) Nodules >5mm

Study	Population	Target condition	Index test	Reference standard	Comments
	South Korea			Nodules for FNA; a six-tier pathology reporting system for CNB	
Ha 2018 [87]	Patients: n= 750 (902 nodules); mean age (range): 49.2 (9-81) years	Thyroid cancer	Ultrasound Classified using 2015 ATA, 2016 KTA/KSThR, 2017 ACR guidelines	Surgical resection (n=191/266 malignant nodules); surgery (n=36 benign nodules); FNA or core needle biopsy (n=75 malignant nodules) Using the Bethesda system	Multicentre study
Hoang 2018 [92]	Patients: n=92 (100 nodules); mean age (SD; range): 52 (14; 19-82) Mean nodule size (SD; range): 2.7cm (1.3; 0.7-5.9 cm) USA	Thyroid cancer	Ultrasound Classified using ACR-TIRADS, ATA, K-TIRADS, F-TIRADS	FNAB or surgery Using the Bethesda System for Cytological classification of Thyroid Nodules	
Hobbs 2014 [94]	Patients: n=350 (360 biopsies); mean age (range): 55 (7-91) Mean nodule size (SD): 26 mm (14) USA	Thyroid cancer	Ultrasound Classified using the SRU guidelines	FNA or surgery Using the Bethesda System for Cytological classification of Thyroid Nodules	

Study	Population	Target condition	Index test	Reference standard	Comments
Horvath 2009 [97]	1097 nodules Nodule size range: 4-60 mm Chile	Thyroid cancer	Ultrasound Classified using TI-RADS (taking BI-RADS as a model)	FNAB Classified as: benign, intermediate/suspicious (follicular lesions), or malignant according to standardised criteria (Clark et al 2005)	8 year prospective study. Malignant nodules received surgery; Benign nodules were followed up. Mean follow-up (range): 3.9 years (2.1-5.8)
Kim 2002 [114]	Patients: n=132 mean age (range) 48 (22-77); 155 nonpalpable solid nodules South Korea	Thyroid cancer	Sonography Sonographic characteristics used to classify malignancy were based on nonpublished criteria from authors' retrospective study.	Histology: FNAB and follow-up (>24 months) of 83 benign nodules; FNAB+ surgery of 44 malignant and 15 benign nodules; surgery alone on five malignant and 8 benign nodules.	
Kim 2013 [116]	Patients: n=686; mean age 49.7 ; 713 nodules South Korea	Thyroid cancer	Ultrasound (US) Need for FNAB determined by US characteristics of the ATA 2009 guidelines	FNAB	Subcentimetre nodules
Kim 2013 [122]	Patients: n=925 (1419 nodules); mean age (range): 51.87 (14-85) South Korea	Thyroid cancer	Ultrasound Classified using Kim and modified Kim criteria	UGFNA	Suggests new US-based guideline system.

Study	Population	Target condition	Index test	Reference standard	Comments
Koh 2018 [129]	Patients: n=363 (370 nodules); mean age (SD; range): 53.1 (13; 19-86) Nodule mean size (SD; range): 20.8 mm (9.8; 10-44mm) South Korea	Thyroid cancer	Ultrasound Classified using Kim, K-TIRADS, 2015 ATA	UGFNA or surgery (n=57 nodules)	
Koseoglu Atilla [133]	Patients n=2614; mean age (SD): 51.01 (13.86) Turkey	Thyroid cancer	Ultrasound Classified using the ACR TI-RADS	FNAB Interpreted using the Bethesda System for Reporting Thyroid Cytopathology	
Lauria Pantano 2018 [138]	Patients: n=946 (1169 nodules); mean age (SD; range): 56 (13.3; 16-88) Nodule media size (range): 14mm (4-56 mm) Italy	Thyroid Cancer	Ultrasound Classified according to the ATA, AACE/ACE/AME and ACR-TI-RADS by an automated algorithm	FNA Classified based on Italian Reporting System for Thyroid Cytology	
Lim-Dunham 2017 [150]	Children: n=33 (39 nodules); median age (range): benign: 16 years (8-18); malignant: 16.5 years (9-18) Median nodule size: malignant: 25.5 mm;	Thyroid cancer	Ultrasound Classified using the 2015 ATA Guidelines for Children	UGFNA or surgery (n=14 nodules)	

Study	Population	Target condition	Index test	Reference standard	Comments
	benign: 21mm USA				
Macedo 2018 [166]	Patients: n= 178 median age (range) 59 (49-66); 195 nodules Brazil	Thyroid cancer	Ultrasonography (US) Classified using modified TI-RADS (malignancy classified in the categories 4 or 5) and ATA risk assessment systems (malignancy classified in the intermediate or high suspicion risk)	Cytology (UGFNAB) Classified based on the Bethesda System for Cytological classification of Thyroid Nodules Histopathology (available for 45 cases after surgery)	Diagnostic accuracy findings reported separately for TI-RADS and ATA.
Maino 2018 [114, 170]	Patients: 340 (432 nodules), mean age (SD, range): 57 years (14.3, 16-86) Median nodules diameter: 20mm (9 - 83 mm) Italy	Thyroid cancer	Ultrasonography (US) Classified based on the ATA risk assessment and the EU-TIRADS (based on the ETA US)	US-guided FNAC Using the British Thyroid Association criteria	
Martinez-Rios 2018 [175]	Children: n=124 (123 nodules); age mean (SD, range): 13.6 (3.1, 3.3-17.7) Mean nodules size (SD, range): 27.5 (14.6mm,	Thyroid cancer	Ultrasound Classified using the ATA (high, intermediate suspicion classifications considered as probably	Histopathology/cytology or 2-year follow-up of clinical outcome for nonoperative cases	Retrospective

Study	Population	Target condition	Index test	Reference standard	Comments
	10-94 mm) Canada		malignant; low, very low suspicion and benign considered as probably malignant) and TI-RADS (4a, 4b, 4c, 5 considered as probably malignant; 2, 3 as probably benign) risk assessment systems.		
Middleton 2017 [179]	Patients: n=3315 (3822 nodules); mean age (range): 54.4 (18-97) USA	Thyroid cancer	Ultrasound Classified using TIRADS	UGFNA	Patients from six geographically diverse medical centres.
Moon 2010 [187]	Patients: n=1024 (1083 nodules); median age (range): 51 (16-83) 539 nodules ≤10mm; 544 >10mm South Korea	Thyroid cancer	Ultrasound + USD (gray-scale + power Doppler US) Classified using the Kim criteria, Kim+USD, AACE/AME	UGFNA	
Moon 2012 [188]	Patients: n=676 (703 nodules); mean age (range): 49.7 (18-79) 308 nodules > 10mm; 395 were ≤10mm; 577 nodules > 5mm; 126 nodules ≤5mm	Thyroid cancer	Ultrasound + USE (gray-scale US + elastography) Classified using the Kim criteria, Kim+USE Rago, Kim+USE Asteria	UGFNA or Surgery Surgery performed after FNA in 221 nodules (202 patients); UGFNA for two nodules in 27 patients and one nodule in 649 patients.	Solid thyroid nodules

Study	Population	Target condition	Index test	Reference standard	Comments
	South Korea				
Na 2016 [190]	Patients: n=1802 (2000 nodules); mean age (SD): 51.2 (12.2) Mean nodule size (SD, range): 20 mm (11.4, 10-100 mm) South Korea	Thyroid cancer	Ultrasound Classified using the K-TIRADS	UGFNA (Bethesda System for Reporting Thyroid Cytopathology) or CNB (diagnosed with a six-tier pathology reporting system) or surgery Surgery:690 nodules CNB: 3 nodules Repeated FNA or CNB: 381 nodules FNA or CNB and follow-up US: 926 nodules	Patients enrolled from low and high cancer volume institutions (two primary medical centres, two tertiary hospitals) Final diagnoses were determined by surgical resections in 15.5% of benign nodules, 99.3% of malignant nodules and by CNB in 0.7% of malignant nodules.
Pandya 2018 [207]	Patients: n=1947 (1947 nodules); mean age (range): 56 (26 to 86 years) Mean nodule diameter (SD): 1.7 cm (0.9cm) USA	Thyroid cancer	Ultrasound Classified based on the 2015 ATA categories of risk	UGFNA Classified according to the Bethesda System for Cytological classification of Thyroid Nodules	
Park 2016 [215]	Patients: n=592 (622 nodules); mean age (range): 49.8 (14-86) Mean nodules size	Thyroid cancer	Ultrasound Classified using the Korea Society of Thyroid radiology	UG-FNAB Classified based on the Bethesda System for Cytological	Nodules followed up for at least 2 years or that underwent surgery.

Study	Population	Target condition	Index test	Reference standard	Comments
	(range): 1.61 cm (0.6-7 cm) South Korea		(KSThR) guidelines	classification of Thyroid Nodules	
Persichetti 2018 [218]	Patients: n=789 (1100 nodules); mean age (SD): 55 (14) Mean nodule size (SD; range): 21.2mm (13.4, 6-75mm) Italy	Thyroid cancer	Ultrasound Classified using the BTA, ATA, AACE/ACE/AME systems.	UGFNA	
Rahal 2016 [231]	Patients: n=906 (n=1000 nodules) Brazil	Thyroid cancer	Ultrasound Classified using TI-RADS	UGFNA Using the Bethesda System for Cytological classification of Thyroid Nodules	
Tae 2007 [276]	Patients: n=580 (1255 nodules); mean age (SD): 47.8 (13.9) Mean nodule size (SD): 2.1 cm (1) South Korea	Thyroid cancer	Ultrasound Classified using the Kim criteria	FNAB and surgery (n=78 patients)	Palpable or non-palpable thyroid nodules
Tang 2017 [281]	Patients: n= 199; 206 nodules USA	Thyroid cancer	US Classified using the ATA risk assessment system	FNAB Using the Bethesda System for reporting thyroid cryopathology	

Study	Population	Target condition	Index test	Reference standard	Comments
				(TBSRTC)	
Weiss 2018 [309]	Patients: n=57 (61 nodules <1cm); mean age (range) 52 (19-81) Mean nodule size (range): 7.8 mm (5-9 mm) USA	Thyroid cancer	US Classified using the ACR TI-RADS risk assessment system	FNAB Using TBSRTC criteria	Subcentimeter nodules (<1 cm)
Xu 2017 [323]	Patients: n= 734 (962 nodules); mean age (SD): 46.75 (14.09) Mean nodule diameter (SD): 17.7 (12.8)mm China	Thyroid cancer	US Classified using TI-RADS (d<10mm) and 2015 ATA (d=10-20mm and d>20mm)guidelines	Surgery (n=703 nodules); >1 year follow-up (repeated cytology; n=259)	Multicentre study (eight tertiary hospitals) Diagnostic accuracy stratified by nodule diameter (d>20mm, d=10-20 mm, d<10mm) and reported separately
Xu 2018 [324]	Patients: n = 2031 (2465 nodules); mean age (SD): 47.7 (13.38) years Mean nodule size (SD): 16.63 (11.78) mm China	Thyroid cancer	US Classified based on patterns and US features of KSThR-TIRADS, ACR-TIRADS, EU-TIRADS	FNAB or surgery	Included lesions undergoing examinations from three tertiary hospitals around JiangSu Province.

Study	Population	Target condition	Index test	Reference standard	Comments
Yoon 2017 [329]	Patients: n= 4585 (4696 nodules); mean age (SD; range): 51 (11.9; 17-94) Mean nodules size (SD, range): 13.3 mm (2.7, 10-19mm) South Korea	Thyroid cancer	US Classification according to six different guidelines: SRU, NCCN, 2015 ATA, F- TI-RADS, Kim, K-TIRADS	Surgery (1072 nodules) or UGFNAB (3624 nodules) Using TBSRTC from December 2009 onwards and the following categories before that: inadequate, benign, intermediate suspected of papillary carcinoma and malignant	Thyroid nodules 1-2 cm
Yoon 2016 [331]	Patients: n=1241 (1293 nodules); mean age (SD; range): 50.8 (13.5; 18-87) Mean nodule size (SD, range): 21.5 mm (11.4, 10-113mm) South Korea	Thyroid cancer	US Classified using TIRADS (Category 3 was considered negative; categories 4a to 5 positive) and ATA (Very-low suspicion were considered negative; low-to-high suspicion positive)	(UG)FNAB (1051 nodules) or surgery (234 nodules) Using TBSRTC criteria	Nodules measured at least 10 mm
Yoon 2015 [332]	Patients: n=1257 (1309 nodules); mean age (SD; range): 50.1 (12.1; 18-83) Mean nodules size (SD;	Thyroid cancer	US US+ vascularity pattern (2-D Doppler US) Classified using Kim	UG-FNAB or surgery (347 nodules) Using TBSRTC	

Study	Population	Target condition	Index test	Reference standard	Comments
	range): 15.1mm (10.3 ; 5-66mm) South Korea		criteria		
Zhang 2018 [353]	Patients: n-162 (243 nodules); mean age (range): 54.7 (21-79) China	Thyroid cancer	US Classified using Russ TI-RADS	FNAB and pathological tests, surgery (n=82 nodules)	Nodules more than 1cm in largest diameter
Zheng 2018 [363]	Patients: n=1013 (1033 nodules); mean age (SD; range): 45.3 (13; 15-81)	Thyroid cancer	US Classified using ACR TI-RADS	FNA (n=506 nodules) or surgery (n=527 nodules)	
Zhang 2017 [354]	Patients: n=246 (319 nodules); mean age (SD; range): 46.1 (15.2; 19-74) Mean nodule size (SD; range): 11.9 mm (3.3; 2.5-46 mm) China	Thyroid cancer	US Classified using TI-RADS, TI-RADS+CEUS	FNAB (n=230 nodules) or surgery (n=89 nodules)	
Zhang 2015 [346]	Patients: n = 2921 (3980 nodules) Mean nodules diameter (SD; range): 15.7 mm (11 mm; 2.0-70.0mm) China	Thyroid cancer	US Classified using Kwak's TI-RADS	FNA (628 nodules) Surgery (partial or total thyroidectomy) performed in all nodules with benign or suspicious cytology and 55 nodules with inconclusive cytology and 10 benign nodules. Remaining	

Study	Population	Target condition	Index test	Reference standard	Comments
				737 nodules underwent surgery without FNA. Pathological diagnosis by surgery (971 nodules)	
Zhou 2018 [367]	Patients: n=161 (167 nodules); mean age (SD): 44.14 (12.01) Mean nodule size (SD):1.31 cm (0.96) China	Thyroid cancer	US Classified using conventional TI-RADS and a novel classification system using TI-RADS+ contrast-enhanced US parameter ratios	FNA or surgery Using TBSRTC	Solitary thyroid nodules

See Appendix D: for full evidence tables.

1.4.4 Quality assessment of clinical studies included in the evidence review

Table 3: Clinical evidence summary: ultrasound in adults

Index Test	Number of studies	n	Quality	Sensitivity % (95% CI)	Specificity % (95% CI)
BTA	2	1091	VERY LOW[a,b,c] due to risk of bias, serious inconsistency and serious imprecision	100% (74-100) 90% (85-95)	35% (25-45) 63% (60-67)
Kim	10	11694	VERY LOW[a,b,c]	91% (84-96)	67% (47-82)

Index Test	Number of studies	n	Quality	Sensitivity % (95% CI)	Specificity % (95% CI)
			due to risk of bias, serious inconsistency and serious imprecision		
Modified Kim	1	945	LOW[a] due to risk of bias	96% (91-98%)	85% (82-87%)
Kim + Doppler	2	2392	MODERATE[a] due to risk of bias	91% (87-94%) 91% (88-94%)	52% (49-56%) 62% (59-65%)
Kim + USE (Rago)	1	703	MODERATE[a] due to risk of bias	92% (88-95%)	65% (61-69%)
Kim + USE (Asteria)	1	703	MODERATE[a] due to risk of bias	94% (91-97%)	48% (43-52%)
SRU	3	6454	VERY LOW[a,b,c] due to risk of bias, very serious inconsistency and very serious imprecision	58% (17-92%)	51% (12-88%)
AACE/ACE/AME	5	5019	VERY LOW[b,c] due to very serious inconsistency and very serious imprecision	93% (75-98%)	51% (15-87%)
ATA	13	13786	LOW[b,c] due to serious inconsistency and serious imprecision	92% (87-95%)	50% (37-63%)
ATA (subcentimetre)	1	713	MODERATE[a] due to risk of bias	97% (94-98%)	27% (22-31%)
KSThR	3	3837	VERY LOW[b,c] due to very serious inconsistency and serious imprecision	95% (85 to 99%)	76% (20-97%)
TIRADS (ACR)	10	13249	LOW[b,c] due to serious inconsistency and serious imprecision	94% (86 to 98%)	54% (45-62%)
TIRADS (French)	7	8494	VERY LOW[a,b,c] due to risk of bias, serious	94% (87 to 98%)	53% (35-70%)

Index Test	Number of studies	n	Quality	Sensitivity % (95% CI)	Specificity % (95% CI)
			inconsistency and serious imprecision		
TIRADS (Kwak)	7	12905	VERY LOW[a,b,c] due to risk of bias, very serious inconsistency, very serious imprecision	97% (94 to 99%)	53% (27 to 78%)
TIRADS (Korean)	4	3504	VERY LOW[b,c] due to very serious inconsistency, very serious imprecision	91% (74-97%)	38% (10-76%)
TIRADS (Horvath)	2	2028	VERY LOW[a,b,c] due to risk of bias, serious inconsistency, serious imprecision	88% (85-91%) 83% (79-87%)	49% (45-52%) 73% (69-76%)
TIRADS (Zhang)	1	319	VERY LOW[a,c] due to risk of bias, serious imprecision	87% (77-93%)	91% (87-95%)
TIRADS (Zhang + CEUS)	1	319	LOW[a] due to risk of bias	97% (91-100%)	96% (93-98%)
TIRADS (Kwak + CEUS)	1	161	LOW[a,c] due to risk of bias, serious imprecision	98% (92-100%)	78% (66-87%)
NCCN	1	4696	MODERATE[a] due to risk of bias	93% (91-95%)	40% (38-41%)

The assessment of the evidence quality was conducted with emphasis on sensitivity as this was identified by the committee as the primary measure in guiding decision-making.

(a) Risk of bias was assessed using the QUADAS-2 checklist. The evidence was downgraded by 1 increment if the majority of studies were rated at high risk of bias, and downgraded by 2 increments if the majority of studies were rated at very high risk of bias.

(b) Inconsistency was assessed by inspection of the sensitivity and specificity plots. Particular attention was placed on the sensitivity threshold set by the committee as an acceptable level to recommend a test. The evidence was

- *downgraded by 1 increment if the individual study values varied across 2 areas: where values of individual studies are both above and below 50%, or both above and below the acceptable threshold 90%*
- *downgraded by 2 increments if the individual study values varied across 3 areas, where values of individual studies are above and below 50%, and also above and below the acceptable threshold 90%*

(c) Imprecision was assessed based on inspection of the credible intervals of sensitivity in the diagnostic meta-analysis or, where diagnostic meta-analysis has not been conducted, assessed according to the range of confidence intervals in the individual studies

Table 4: Clinical evidence summary: ultrasound in children

Index Test	Number of studies	n	Quality	Sensitivity % (95% CI)	Specificity % (95% CI)
ATA	3	301	VERY LOW[a,b] due to very serious inconsistency and serious imprecision	91% (69-98%)	53% (19-85%)
TIRADS (Kwak)	1	123	HIGH	100% (93-100%)	18% (10-29%)

The assessment of the evidence quality was conducted with emphasis on sensitivity as this was identified by the committee as the primary measure in guiding decision-making.

(a) Inconsistency was assessed by inspection of the sensitivity and specificity plots. Particular attention was placed on the sensitivity threshold set by the committee as an acceptable level to recommend a test. The evidence was

- *downgraded by 1 increment if the individual study values varied across 2 areas: where values of individual studies are both above and below 50%, or both above and below the acceptable threshold 90%*
- *downgraded by 2 increments if the individual study values varied across 3 areas, where values of individual studies are above and below 50%, and also above and below the acceptable threshold 90%*

(b) Imprecision was assessed based on inspection of the credible intervals of sensitivity in the diagnostic meta-analysis or, where diagnostic meta-analysis has not been conducted, assessed according to the range of confidence intervals in the individual studies

1.5 Economic evidence

1.5.1 Included studies

No relevant health economic studies were identified.

1.5.2 Excluded studies

No health economic studies that were relevant to this question were excluded due to assessment of limited applicability or methodological limitations.

See also the health economic study selection flow chart in Appendix F:.

1.5.3 Health economic modelling

This area was not prioritised for new cost-effectiveness analysis.

1.5.4 Resource costs

Relevant unit costs are provided below to aid consideration of cost effectiveness.

Table 5: UK costs of imaging tests

Imaging test	Unit costs
Ultrasound scan (USS) (a)	£53.22
Computerised Tomography (CT) (b)	£85.78
Magnetic Resonance Imaging Scan (MRI) (c)	£138.38

Source[s]: NHS reference costs 2016-17, total HRG schedule [54].
(a) Ultrasound Scan with duration of less than 20 minutes and over 20 minutes, without contrast, RD40Z, RD42Z
(b) Computerised Tomography Scan of One Area, without contrast, all age groups, RD20A, RD20B and RD20C
(c) Magnetic Resonance Imaging Scan of One Area, without contrast, all age groups, RD01A, RD01B and RD01C

1.6 Evidence statements

1.6.1 Clinical evidence statements

Thirty-four studies that evaluated ultrasound under different criteria were included in the review. Of these, two studies were conducted in children. The evidence was of very low to moderate quality for adults and low to high quality for children.

1.6.1.1 Ultrasound in adults

- **BTA:** very low quality evidence from 2 studies with 1091 participants showed that ultrasound using the BTA guidelines has a sensitivity range of 90 -100% and a specificity of 35-63%
- **Kim:** very low quality evidence from 10 studies with 11694 participants showed that ultrasound using the Kim criteria has a sensitivity of 91% and a specificity of 67%.
- **Modified Kim:** low quality evidence from 1 study with 945 participants showed that ultrasound using modified Kim criteria has a sensitivity of 96% and a specificity of 85%.
- **Kim + Doppler:** moderate quality evidence from 2 studies with 2392 participants showed that ultrasound combined with power Doppler using the Kim criteria has a sensitivity of 91% and a specificity of 52%.

- **Kim + USE (Rago):**moderate quality evidence from 1 study with 703 participants showed that ultrasound using the Kim criteria combined with elastography (USE) using the Rago criteria has a sensitivity of 92% and a specificity of 65%
- **Kim + USE (Asteria):** moderate quality evidence from 1 study with 703 participants showed that ultrasound using the Kim criteria combined with elastography (USE) using the Asteria criteria has a sensitivity of 94% and a specificity of 48%
- **SRU:** very low quality evidence from 3 studies with 6454 participants showed that ultrasound using the SRU criteria has a sensitivity of 58% and a specificity of 51%.
- **AACE/ACE/AME:** very low quality evidence from 5 studies showed that ultrasound using the AACE/ACE/AME criteria has a sensitivity of 93% and a specificity of 51%.
- **ATA:** low quality evidence from 13 studies with 13786 participants showed that ultrasound using the ATA guidelines has a sensitivity of 92% and a specificity of 50%.
- **ATA (subcentimeter):** moderate quality evidence from 1 study with 713 participants showed that for subcentimeter nodules, ultrasound using the ATA criteria has a sensitivity of 97% and a specificity of 27%.
- **KSThR:** very low quality evidence from 3 studies with 3837 participants showed that ultrasound using the KSThR criteria has a sensitivity of 95% and a specificity of 76%.
- **TIRADS (ACR):** low quality evidence from 10 studies with 13249 participants showed that ultrasound using ACR-TIRADS has a sensitivity of 94% and a specificity of 54%.
- **TIRADS (French):** very low quality evidence from 7 studies with 8494 participants showed that ultrasound using French-TIRADS has a sensitivity of 94% and a specificity of 53%.
- **TIRADS (Kwak):** very low quality evidence from 7 studies with 12905 participants showed that ultrasound using Kwak's TIRADS has a sensitivity of 97% and a specificity of 53%.
- **TIRADS (Korean):** very low quality evidence from 4 studies with 3504 participants showed that ultrasound using the Korean TIRADS has a sensitivity of 91% and a specificity of 38%.
- **TIRADS (Horvath):** very low quality evidence from 2 studies with 2028 participants showed that ultrasound using Horvath's version of the TIRADS has a sensitivity range of 83-88% and a specificity range of 49-73%.
- **TIRADS (Zhang):** very low quality evidence from 1 study with 319 participants showed that ultrasound using Zhang's version of the TIRADS has a sensitivity of 87% and a specificity of 91%.
- **TIRADS (Zhang + CEUS):** low quality evidence from 1 study with 319 participants showed that ultrasound using Zhang's TIRADS and CEUS classification has a sensitivity of 97% and a specificity of 96%.
- **TIRADS (Kwak + CEUS):** low quality evidence from 1 study with 161 participants showed that ultrasound using Kwak's TIRADS combined with CEUS classification has a sensitivity of 98% and a specificity of 78%.
- **NCCN:** moderate quality evidence from 1 study with 4696 participants showed that ultrasound using the NCCN criteria has a sensitivity of 93% and a specificity of 40%.

1.6.1.2 Ultrasound in children

- **ATA:** very low quality evidence from 3 studies with 301 participants showed that ultrasound using the ATA guidelines has a sensitivity of 91% and a specificity range of 53% in children.
- **TIRADS (Kwak):** high quality evidence from 1 study with 123 participants showed that ultrasound using Kwak's TIRADS has a sensitivity of 100% and a specificity of 18% in children.

1.6.2 Health economic evidence statements

- No relevant economic evaluations were identified.

1.7 The committee's discussion of the evidence

1.7.1 Interpreting the evidence

1.7.1.1 The diagnostic measures that matter most

The diagnostic measures of sensitivity, specificity, positive and negative predictive value of the ultrasound scan for diagnosing malignancy under different sonographic criteria were considered for this review. Sensitivity was deemed the most important measure by the committee and hence it was prioritised for decision making.

1.7.1.2 The quality of the evidence

The quality of the evidence for adults ranged from very low to moderate; the majority being of very low quality, and was downgraded due to risk of bias, inconsistency and imprecision. In children, the quality of the evidence ranged from very low to high and was downgraded for inconsistency and imprecision. No evidence was identified for the diagnostic accuracy of CT and MRI scan. Across studies, the diagnostic accuracy of ultrasound was based on histopathological confirmation that was mostly fine-needle aspiration (FNA) and/or surgery.

The committee noted that the majority of studies excluded participants whose FNA results were not definitive (i.e. included if benign or malignant result but anything else excluded). They agreed that in reality there will be a considerable number of FNA results that fall between these ends of the spectrum and specified that the appropriate management of these results is outside the scope of this guideline. However, it is unlikely to have a significant effect on the choice of optimal imaging option and ultrasound criteria.

The committee agreed that the breadth of evidence for the various ultrasound criteria was dictated by their novelty. The older criteria (for example Kim) have been available since the early 2000s whereas criteria like the BTA and some of the TIRADS have only been available for around 5 years. This inevitably impacts the number of studies available assessing their accuracy.

1.7.1.3 Benefits and harms

1.7.1.3.1 Ultrasound scan in adults

Evidence suggested that in adults, both measures of sensitivity and specificity were similarly high for the use of ultrasound under the majority of the different criteria identified. The only ultrasound criteria for which diagnostic accuracy was considerably lower compared to the other criteria were the SRU (58% sensitivity). The committee noted that the reason for this discrepancy is likely to be that the size of nodules is taken into account when assessing the likelihood of malignancy according to the SRU guidelines. They specified that nodule size is irrelevant in predicting malignancy and using size criteria can result in less sensitivity. Evidence suggested that the diagnostic accuracy of ultrasound is increased when a modified version of the Kim criteria is used compared to the conventional version (96 vs 91% sensitivity; 85 vs 67% specificity). However, the committee noted that this was shown by only one study and as the modified criteria essentially involved a raise of threshold which makes the benefit in both sensitivity and specificity counter intuitive, this was not likely to reflect a true difference in diagnostic accuracy. Gray-scale ultrasound combined with power Doppler ultrasound under the Kim criteria did not lead to increased diagnostic accuracy compared to conventional gray-scale ultrasound under the Kim criteria alone (both 91% sensitivity).

Studies using ultrasound imaging based on the TIRADS, showed that diagnostic accuracy was high for all the different versions of the guidelines identified. Evidence also showed that when using the TIRADS criteria, ultrasound combined with contrast-enhanced US parameter ratios (CEUS) may result in higher sensitivity and specificity compared to ultrasound without

CEUS. Similarly, it was evident that under the Kim criteria, ultrasound had a minor increase in diagnostic accuracy when combined with elastography (USE). The committee noted that combining elastography with ultrasound is not current practice and would require specialised equipment, training and expertise and would thus be likely to have a significant economic impact. Based on the small number of studies for CEUS and USE and the small magnitude of the benefit in diagnostic accuracy introduced, the committee agreed that the current evidence did not justify a change in current practice.

There was a lack of evidence for the diagnostic accuracy of CT and MRI. The committee noted that CT is not good at discriminating structures of the thyroid gland and that the MRI has no consistent ability to examine malignancy. There was agreement that ultrasound constitutes the only good existing imaging technique for the first assessment of thyroid enlargement. However, the committee emphasised that further imaging may be useful in other circumstances, for example CT scanning in the case of enlargement causing compression symptoms.

The committee discussed the role of incidental findings of thyroid enlargement from other imaging. They noted that these are frequent reasons for referral but in line with their experience and based on their awareness of other evidence, agreed that incidental findings rarely indicate malignancy. Despite this, further investigation is often done due to concerns around medicolegal risk. The committee agreed that in some cases incidental findings may need further investigation but that healthcare professionals should consider the overall likelihood of malignancy in a person before continuing on the investigative pathway. The committee also noted that the likelihood of malignancy will be dependent on the imaging modality. For example, incidental findings on CT scans are less concerning but rates of malignancy may be higher in incidental findings on FDG-PET scans.

1.7.1.3.2 Ultrasound scan in children

The imaging evidence identified for children in the review showed that ultrasound had high diagnostic accuracy both when using the ATA guidelines and the TIRADS proposed by Kwak. No evidence for the diagnostic accuracy of the CT and the MRI scan was identified in children. The committee agreed that similarly to adults, this was likely to be due to ultrasound being the only good existing imaging technique and that the imaging recommendations made for adults would be applicable to children as well.

1.7.2 Cost effectiveness and resource use

No health economic evidence was identified for this question. The committee was therefore not able to assess the cost effectiveness of which imaging tests should be requested (for thyroid enlargement) and which people with structural abnormalities should have a fine-needle aspiration biopsy. Unit costs for the US, CT and MRI, obtained from the NHS reference cost 2016-17, were presented to the committee. The cheapest imaging test was the US scan costing £53.22 (RD40Z, RD42Z), CT cost £85.78 (RD20A, RD20B and RD20C), and MRI was £138.38 (RD01A, RD01B and RD01C).

Although evidence was not identified to support the use of CT or MRI, the clinical evidence identified suggested that ultrasound had good diagnostic accuracy and being the cheapest option the committee recommended its use.

The committee also noted that using an established grading system that did not take into account the nodular size for referring patients to have FNAB, is likely to reduce the number of patients being referred to FNAB and therefore it is likely to be cost saving. In addition, by correctly reporting these findings, repeats could be avoided, and money saved.

Ultrasound to assess likelihood of thyroid malignancy is current practice. The most commonly used ultrasound criteria are those of the British Thyroid Association, which are in line with the recommendations made above. Therefore, overall these recommendations are not expected to have a substantial resource impact to the NHS in England.

1.7.3 Other factors the committee took into account

Although not a focus of this evidence review, the committee raised a need for US images to be recorded and stored to enable review in cases such as multi-disciplinary team meetings and referral to secondary care. They raised the importance for US reports to be explicit in terms of the criteria based on which the likelihood of malignancy was determined, including the specific nodule features examined, to facilitate clinicians in cases where re-visiting imaging is warranted. Within this framework the committee specified that confirmation that both lobes have been assessed as well as the assessment of cervical lymph nodes should also be documented in US reports.

References

1. Abdel-Rahman HM, Abowarda MH, Abdel-Aal SM. Diffusion-weighted MRI and apparent diffusion coefficient in differentiation of benign from malignant solitary thyroid nodule. Egyptian Journal of Radiology and Nuclear Medicine. 2016; 47(4):1385-1390

2. Afifi AH, Alwafa WAHA, Aly WM, Alhammadi HAB. Diagnostic accuracy of the combined use of conventional sonography and sonoelastography in differentiating benign and malignant solitary thyroid nodules. Alexandria Journal of Medicine. 2017; 53(1):21-30

3. Aggarwal A, Aggarwal J, Awasthi S. Non invasive assessment of malignancy in solitary non palpable thyroid nodules: A comparative and complimentary evaluation of grey scale ultrasound and color doppler parameters -way towards a new scoring system? Acta Medica International. 2017; 4(1):1-6

4. Aghaghazvini L, Sharifian H, Yazdani N, Hosseiny M, Pirouzi P, Ghadiri A et al. Differentiation between benign and malignant thyroid nodules using diffusion-weighted imaging, a 3-T MRI study. Indian Journal of Radiology and Imaging. 2018; 28(4):460-464

5. Ahn HS, Lee JB, Seo M, Park SH, Choi BI. Distinguishing benign from malignant thyroid nodules using thyroid ultrasonography: utility of adding superb microvascular imaging and elastography. Radiologia Medica. 2018; 123(4):260-270

6. Ahn SS, Kim EK, Kang DR, Lim SK, Kwak JY, Kim MJ. Biopsy of thyroid nodules: comparison of three sets of guidelines. AJR American Journal of Roentgenology. 2010; 194(1):31-7

7. Akhavan A, Jafari SM, Khosravi MH, Khajehpour H, Karimi-Sari H. Reliability of fine-needle aspiration and ultrasound-based characteristics of thyroid nodules for diagnosing malignancy in Iranian patients. Diagnostic Cytopathology. 2016; 44(4):269-73

8. Al Nofal A, Gionfriddo MR, Javed A, Haydour Q, Brito JP, Prokop LJ et al. Accuracy of thyroid nodule sonography for the detection of thyroid cancer in children: systematic review and meta-analysis. Clinical Endocrinology. 2016; 84(3):423-30

9. Alam T, Khattak YJ, Beg M, Raouf A, Azeemuddin M, Khan AA. Diagnostic accuracy of ultrasonography in differentiating benign and malignant thyroid nodules using fine needle aspiration cytology as the reference standard. Asian Pacific Journal of Cancer Prevention: Apjcp. 2014; 15(22):10039-43

10. Albair Ashamallah G, M.A EL-A. Risk for malignancy of thyroid nodules: Comparative study between TIRADS and US based classification system. Egyptian Journal of Radiology and Nuclear Medicine. 2016; 47(4):1373-1384

11. Algin O, Algin E, Gokalp G, Ocakoglu G, Erdogan C, Saraydaroglu O et al. Role of duplex power Doppler ultrasound in differentiation between malignant and benign thyroid nodules. Korean Journal of Radiology. 2010; 11(6):594-602

12. Appetecchia M, Solivetti FM. The association of colour flow Doppler sonography and conventional ultrasonography improves the diagnosis of thyroid carcinoma. Hormone Research. 2006; 66(5):249-56

13. Asteria C, Giovanardi A, Pizzocaro A, Cozzaglio L, Morabito A, Somalvico F et al. US-elastography in the differential diagnosis of benign and malignant thyroid nodules. Thyroid. 2008; 18(5):523-31

14. Azizi G, Keller J, Lewis M, Puett D, Rivenbark K, Malchoff C. Performance of elastography for the evaluation of thyroid nodules: a prospective study. Thyroid. 2013; 23(6):734-40

15. Bae JM, Hahn SY, Shin JH, Ko EY. Erratum to "Inter-exam agreement and diagnostic performance of the Korean Thyroid Imaging Reporting and Data System for thyroid nodule assessment: Real-time versus static ultrasonography". European Journal of Radiology. 2018; 101:193

16. Bae JM, Hahn SY, Shin JH, Ko EY. Inter-exam agreement and diagnostic performance of the Korean thyroid imaging reporting and data system for thyroid nodule assessment: Real-time versus static ultrasonography. European Journal of Radiology. 2018; 98:14-19

17. Bhatia KS, Rasalkar DP, Lee YP, Wong KT, King AD, Yuen HY et al. Cystic change in thyroid nodules: a confounding factor for real-time qualitative thyroid ultrasound elastography. Clinical Radiology. 2011; 66(9):799-807

18. Bhatia KS, Tong CS, Cho CC, Yuen EH, Lee YY, Ahuja AT. Shear wave elastography of thyroid nodules in routine clinical practice: preliminary observations and utility for detecting malignancy. European Radiology. 2012; 22(11):2397-406

19. Bojunga J, Herrmann E, Meyer G, Weber S, Zeuzem S, Friedrich-Rust M. Real-time elastography for the differentiation of benign and malignant thyroid nodules: a meta-analysis. Thyroid. 2010; 20(10):1145-50

20. Brito JP, Gionfriddo MR, Al Nofal A, Boehmer KR, Leppin AL, Reading C et al. The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and meta-analysis. Journal of Clinical Endocrinology and Metabolism. 2014; 99(4):1253-63

21. Brophy C, Stewart J, O'Donovan N, McCarthy J, Murphy M, Sheahan P. Impact of microcalcifications on risk of malignancy in thyroid nodules with indeterminate or benign cytology. Otolaryngology - Head & Neck Surgery. 2016; 154(1):46-51

22. Cakal E, Sahin M, Unsal IO, Gungunes A, Akkaymak E, Ozkaya EC et al. Elastography in the differential diagnosis of thyroid nodules. Ultrasonic Imaging. 2015; 37(3):251-7

23. Cakir B, Aydin C, Korukluoglu B, Ozdemir D, Sisman IC, Tuzun D et al. Diagnostic value of elastosonographically determined strain index in the differential diagnosis of benign and malignant thyroid nodules. Endocrine. 2011; 39(1):89-98

24. Cam OH, Tekin M, Acar GO, Kilicaslan A. What is the Role of Diffusion Weigh Magnetic Resonance Imaging in Evaluation of Thyroid Nodules? Indian Journal of Otolaryngology & Head & Neck Surgery. 2014; 66(3):336-40

25. Camargo RY, Tomimori EK, Knobel M, Medeiros-Neto G. Preoperative assessment of thyroid nodules: role of ultrasonography and fine needle aspiration biopsy followed by cytology. Clinics (Sao Paulo, Brazil). 2007; 62(4):411-8

26. Cantisani V, Grazhdani H, Ricci P, Mortele K, Di Segni M, D'Andrea V et al. Q-elastosonography of solid thyroid nodules: assessment of diagnostic efficacy and interobserver variability in a large patient cohort. European Radiology. 2014; 24(1):143-50

27. Cantisani V, Lodise P, Di Rocco G, Grazhdani H, Giannotti D, Patrizi G et al. Diagnostic accuracy and interobserver agreement of quasistatic ultrasound elastography in the diagnosis of thyroid nodules. Ultraschall in der Medizin. 2015; 36(2):162-7

28. Cappelli C, Castellano M, Pirola I, Cumetti D, Agosti B, Gandossi E et al. The predictive value of ultrasound findings in the management of thyroid nodules. QJM:An International Journal of Medicine. 2007; 100(1):29-35

29. Cappelli C, Castellano M, Pirola I, Gandossi E, De Martino E, Cumetti D et al. Thyroid nodule shape suggests malignancy. European Journal of Endocrinology. 2006; 155(1):27-31

30. Cappelli C, Pirola I, Cumetti D, Micheletti L, Tironi A, Gandossi E et al. Is the anteroposterior and transverse diameter ratio of nonpalpable thyroid nodules a sonographic criteria for recommending fine-needle aspiration cytology? Clinical Endocrinology. 2005; 63(6):689-93

31. Cavallo A, Johnson DN, White MG, Siddiqui S, Antic T, Mathew M et al. Thyroid nodule size at ultrasound as a predictor of malignancy and final pathologic size. Thyroid. 2017; 27(5):641-650

32. Cetin N, Yucel C, Uyar Gocun P, Aladag Kurt S, Taneri F, Oktar S et al. The efficiency of ultrasound elastography in the differential diagnosis of thyroid nodules. Journal of the Belgian Society of Radiology. 2015; 98(1):20-26

33. Chandramohan A, Khurana A, Pushpa BT, Manipadam MT, Naik D, Thomas N et al. Is TIRADS a practical and accurate system for use in daily clinical practice? Indian Journal of Radiology & Imaging. 2016; 26(1):145-52

34. Chen BD, Xu HX, Zhang YF, Liu BJ, Guo LH, Li DD et al. The diagnostic performances of conventional strain elastography (SE), acoustic radiation force impulse (ARFI) imaging and point shear-wave speed (pSWS) measurement for non-calcified thyroid nodules. Clinical Hemorheology and Microcirculation. 2017; 65(3):259-273

35. Chen L, Xu J, Bao J, Huang X, Hu X, Xia Y et al. Diffusion-weighted MRI in differentiating malignant from benign thyroid nodules: a meta-analysis. BMJ Open. 2016; 6(1):e008413

36. Chen L, Zhang J, Meng L, Lai Y, Huang W. A new ultrasound nomogram for differentiating benign and malignant thyroid nodules. Clinical Endocrinology. 2019; 90(2):351-359

37. Chen M, Zhang KQ, Xu YF, Zhang SM, Cao Y, Sun WQ. Shear wave elastography and contrast-enhanced ultrasonography in the diagnosis of thyroid malignant nodules. Molecular & Clinical Oncology. 2016; 5(6):724-730

38. Chen PY, Chiou SC, Yeh HY, Chen CP, Ho C, Lin JD et al. Correlation of ultrasonography with fine needle aspiration cytology and final pathological diagnoses in patients with thyroid nodules. Chinese Journal of Radiology. 2010; 35(1):1-7

39. Chen SP, Hu YP, Chen B. Taller-than-wide sign for predicting thyroid microcarcinoma: comparison and combination of two ultrasonographic planes. Ultrasound in Medicine and Biology. 2014; 40(9):2004-11

40. Cheng P, Chen ED, Zheng HM, He QX, Li Q. Ultrasound score to select subcentimeter-sized thyroid nodules requiring ultrasound-guided fine needle aspiration biopsy in eastern China. Asian Pacific Journal of Cancer Prevention: Apjcp. 2013; 14(8):4689-92

41. Cheng SP, Lee JJ, Lin JL, Chuang SM, Chien MN, Liu CL. Characterization of thyroid nodules using the proposed thyroid imaging reporting and data system (TI-RADS). Head and Neck. 2013; 35(4):541-7

42. Chi J, Walia E, Babyn P, Wang J, Groot G, Eramian M. Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network. Journal of Digital Imaging. 2017; 30(4):477-486

43. Chiu WY, Chia NH, Wan SK, Yuen CH, Cheung MT. The investigation and management of thyroid nodules--a retrospective review of 183 cases. Annals of the Academy of Medicine, Singapore. 1998; 27(2):196-9

44. Chng CL, Tan HC, Too CW, Lim WY, Chiam PPS, Zhu L et al. Diagnostic performance of ATA, BTA and TIRADS sonographic patterns in the predication of malignancy in histologically proven thyroid nodules. Singapore Medical Journal. 2018; Epublication

45. Choi YJ, Baek JH, Baek SH, Shim WH, Lee KD, Lee HS et al. Web-based malignancy risk estimation for thyroid nodules using ultrasonography characteristics: Development and validation of a predictive model. Thyroid. 2015; 25(12):1306-12

46. Choi YJ, Baek JH, Park HS, Shim WH, Kim TY, Shong YK et al. A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of thyroid nodules on ultrasound: Initial clinical assessment. Thyroid. 2017; 27(4):546-552

47. Chong Y, Shin JH, Ko ES, Han BK. Ultrasonographic elastography of thyroid nodules: is adding strain ratio to colour mapping better? Clinical Radiology. 2013; 68(12):1241-6

48. Creo A, Alahdab F, Al Nofal A, Thomas K, Kolbe A, Pittock ST. Ultrasonography and the American Thyroid Association Ultrasound-Based Risk Stratification Tool: Utility in Pediatric and Adolescent Thyroid Nodules. Hormone Research in Paediatrics. 2018; 90(2):93-101

49. D'Souza MM, Marwaha RK, Sharma R, Jaimini A, Thomas S, Singh D et al. Prospective evaluation of solitary thyroid nodule on 18F-FDG PET/CT and high-resolution ultrasonography. Annals of Nuclear Medicine. 2010; 24(5):345-55

50. Delfim RLC, Veiga L, Vidal APA, Lopes F, Vaisman M, Teixeira P. Likelihood of malignancy in thyroid nodules according to a proposed Thyroid Imaging Reporting and Data System (TI-RADS) classification merging suspicious and benign ultrasound features. Archives of Endocrinology & Metabolism. 2017; 61(3):211-221

51. Deng J, Zhou P, Tian SM, Zhang L, Li JL, Qian Y. Comparison of diagnostic efficacy of contrast-enhanced ultrasound, acoustic radiation force impulse imaging, and their combined use in differentiating focal solid thyroid nodules. PloS One. 2014; 9(3):e90674

52. Deng S, Jiang Q, Zhu Y, Zhang Y. An analysis of the clinical value of high-frequency color Doppler ultrasound in the differential diagnosis of benign and malignant thyroid nodules. International Journal of Clinical and Experimental Medicine. 2018; 11(3):2331-2336

53. Deng XH, Tang LN, Liu SQ, Li XL, He YP, Xu HX. A proposal to stratify the intermediate-risk thyroid nodules according to the AACE/ACE/AME guidelines with ultrasound features. Scientific Reports. 2017; 7:17901

54. Department of Health. NHS reference costs 2016-17. 2017. Available from: https://www.gov.uk/government/collections/nhs-reference-costs Last accessed: 15/02/2019

55. Diao X, Zhan J, Chen L, Chen Y, Liu Y. Quantification of solid hypo-echoic thyroid nodule enhancement with contrast-enhanced ultrasound. Translational Cancer Research. 2017; 6(6):1078-1087

56. Dighe M, Kim J, Luo S, Kim Y. Utility of the ultrasound elastographic systolic thyroid stiffness index in reducing fine-needle aspirations. Journal of Ultrasound in Medicine. 2010; 29(4):565-74

57. Dighe M, Luo S, Cuevas C, Kim Y. Efficacy of thyroid ultrasound elastography in differential diagnosis of small thyroid nodules. European Journal of Radiology. 2013; 82(6):e274-80

58. Dilli A, Ayaz UY, Cakir E, Cakal E, Gultekin SS, Hekimoglu B. The efficacy of apparent diffusion coefficient value calculation in differentiation between malignant and benign thyroid nodules. Clinical Imaging. 2012; 36(4):316-322

59. Ding J, Cheng H, Ning C, Huang J, Zhang Y. Quantitative measurement for thyroid cancer characterization based on elastography. Journal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine. 2011; 30(9):1259-1266

60. Dobruch-Sobczak K, Zalewska EB, Guminska A, Slapa RZ, Mlosek K, Wareluk P et al. Diagnostic performance of shear wave elastography parameters alone and in combination with conventional b-mode ultrasound parameters for the characterization of thyroid nodules: A prospective, dual-center study. Ultrasound in Medicine and Biology. 2016; 42(12):2803-2811

61. Du YR, Ji CL, Wu Y, Gu XG. Combination of ultrasound elastography with TI-RADS in the diagnosis of small thyroid nodules (<=10 mm): A new method to increase the diagnostic performance. European Journal of Radiology. 2018; 109:33-40

62. Duan SB, Yu J, Li X, Han ZY, Zhai HY, Liang P. Diagnostic value of two-dimensional shear wave elastography in papillary thyroid microcarcinoma. OncoTargets and Therapy. 2016; 9:1311-7

63. Dy JGD, Kasala R, Yao C, Ongoco R, Mojica DJ. Thyroid imaging reporting and data system (TIRADS) in stratifying risk of thyroid malignancy at the Medical city. Journal of the ASEAN Federation of Endocrine Societies. 2017; 32(2):108-116

64. Ebeed AE, Romeih MAEH, Refat MM, Salah NM. Role of ultrasound, color doppler, elastography and micropure imaging in differentiation between benign and malignant thyroid nodules. Egyptian Journal of Radiology and Nuclear Medicine. 2017; 48(3):603-610

65. El-Hariri MA, Taha Ali TF, Tawab MA, Magid AMA, El-Shiekh AF. The clinical value of ultrasound elastography in predicting malignant thyroid nodules. Egyptian Journal of Radiology and Nuclear Medicine. 2014; 45(2):353-359

66. Elsayed NM, Elkhatib YA. Diagnostic criteria and accuracy of categorizing malignant thyroid nodules by ultrasonography and ultrasound elastography with pathologic correlation. Ultrasonic Imaging. 2016; 38(2):148-58

67. Farihah AG, Nurismah MI, Husyairi H, Shahrun Niza AS, Radhika S. Reliability of the ultrasound classification system of thyroid nodules in predicting malignancy. Medical Journal of Malaysia. 2018; 73(1):9-15

68. Fukunari N, Nagahama M, Sugino K, Mimura T, Ito K, Ito K. Clinical evaluation of color Doppler imaging for the differential diagnosis of thyroid follicular lesions. World Journal of Surgery. 2004; 28(12):1261-5

69. Gamme G, Parrington T, Wiebe E, Ghosh S, Litt B, Williams DC et al. The utility of thyroid ultrasonography in the management of thyroid nodules. Canadian Journal of Surgery. 2017; 60(2):134-139

70. Gannon AW, Langer JE, Bellah R, Ratcliffe S, Pizza J, Mostoufi-Moab S et al. Diagnostic Accuracy of Ultrasound With Color Flow Doppler in Children With Thyroid Nodules. Journal of Clinical Endocrinology and Metabolism. 2018; 103(5):1958-1965

71. Gao L, Liu R, Jiang Y, Song W, Wang Y, Liu J et al. Computer-aided system for diagnosing thyroid nodules on ultrasound: A comparison with radiologist-based clinical assessments. Head and Neck. 2018; 40(4):778-783

72. Garcia-Monco Fernandez C, Serrano-Moreno C, Donnay-Candil S, Carrero-Alvaro J. Acorrelation study between histological results and thyroid ultrasound findings. The TI-RADS classification. Endocrinologia Diabetes y Nutricion. 2018; 65(4):206-212

73. Gietka-Czernel M, Kochman M, Bujalska K, Stachlewska-Nasfeter E, Zgliczynski W. Real-time ultrasound elastography - a new tool for diagnosing thyroid nodules. Endokrynologia Polska. 2010; 61(6):652-7

74. Ginat DT, Butani D, Giampoli EJ, Patel N, Dogra V. Pearls and pitfalls of thyroid nodule sonography and fine-needle aspiration. Ultrasound Quarterly. 2010; 26(3):171-178

75. Giusti M, Orlandi D, Melle G, Massa B, Silvestri E, Minuto F et al. Is there a real diagnostic impact of elastosonography and contrast-enhanced ultrasonography in the management of thyroid nodules? Journal of Zhejiang University SCIENCE B. 2013; 14(3):195-206

76. Glogovsek M, Gaberscek S, Zorman M. A simple graphical quantitative analysis of ultrasonography images to decide when to perform fine needle aspiration biopsy in diagnosing malignancy in solid thyroid nodules? A two centres prospective study. Hellenic Journal of Nuclear Medicine. 2015; 18(1):25-30

77. Goldfarb M, Gondek S, Solorzano C, Lew JI. Surgeon-performed ultrasound can predict benignity in thyroid nodules. Surgery. 2011; 150(3):436-41

78. Goldfarb M, Gondek SS, Sanchez Y, Lew JI. Clinic-based ultrasound can predict malignancy in pediatric thyroid nodules. Thyroid. 2012; 22(8):827-31

79. Gotzberger M, Krueger S, Gartner R, Reincke M, Pichler M, Assmann G et al. Comparison of color-Doppler and qualitative and quantitative strain-elastography for differentiation of thyroid nodules in daily practice. Hormones. 2016; 15(2):197-204

80. Grani G, Lamartina L, Ascoli V, Bosco D, Biffoni M, Giacomelli L et al. Reducing the Number of Unnecessary Thyroid Biopsies While Improving Diagnostic Accuracy: Toward the "Right" TIRADS. Journal of Clinical Endocrinology and Metabolism. 2019; 104(1):95-102

81. Gu F, Han L, Yang X, Liu H, Li X, Guo K et al. Value of time-intensity curve analysis of contrast-enhanced ultrasound in the differential diagnosis of thyroid nodules. European Journal of Radiology. 2018; 105:182-187

82. Gu J, Du L, Bai M, Chen H, Jia X, Zhao J et al. Preliminary study on the diagnostic value of acoustic radiation force impulse technology for differentiating between

benign and malignant thyroid nodules. Journal of Ultrasound in Medicine. 2012; 31(5):763-71

83. Guazzaroni M, Spinelli A, Coco I, Del Giudice C, Girardi V, Simonetti G. Value of strain-ratio on thyroid real-time sonoelastography. Radiologia Medica. 2014; 119(3):149-55

84. Gul K, Ersoy R, Dirikoc A, Korukluoglu B, Ersoy PE, Aydin R et al. Ultrasonographic evaluation of thyroid nodules: comparison of ultrasonographic, cytological, and histopathological findings. Endocrine. 2009; 36(3):464-72

85. Gupta N, Norbu C, Goswami B, Chowdhury V, RaviShankar L, Gulati P et al. Role of dynamic MRI in differentiating benign from malignant follicular thyroid nodule. Auris, Nasus, Larynx. 2011; 38(6):718-723

86. Ha EJ, Moon WJ, Na DG, Lee YH, Choi N, Kim SJ et al. A multicenter prospective validation study for the Korean thyroid imaging reporting and data system in patients with thyroid nodules. Korean Journal of Radiology. 2016; 17(5):811-821

87. Ha EJ, Na DG, Moon WJ, Lee YH, Choi N. Diagnostic performance of ultrasound-based risk-stratification systems for thyroid nodules: Comparison of the 2015 American thyroid association guidelines with the 2016 Korean thyroid association/Korean society of thyroid radiology and 2017 American congress of radiology guidelines. Thyroid. 2018; 28(11):1532-1537

88. Ha SM, Ahn HS, Baek JH, Ahn HY, Chung YJ, Cho BY et al. Validation of three scoring risk-stratification models for thyroid nodules. Thyroid. 2017; 27(12):1550-1557

89. Ha SM, Kim JK, Baek JH. Detection of malignancy among suspicious thyroid nodules <1cm on ultrasound with various thyroid image reporting and data systems. Thyroid. 2017; 27(10):1307-1315

90. Hamidi C, Goya C, Hattapoglu S, Uslukaya O, Teke M, Durmaz MS et al. Acoustic Radiation Force Impulse (ARFI) imaging for the distinction between benign and malignant thyroid nodules. La Radiologia Medica. 2015; 120(6):579-583

91. He LZ, Zeng TS, Pu L, Pan SX, Xia WF, Chen LL. Thyroid Hormones, Autoantibodies, Ultrasonography, and Clinical Parameters for Predicting Thyroid Cancer. International Journal of Endocrinology Print. 2016; 2016:8215834

92. Hoang JK, Middleton WD, Farjat AE, Langer JE, Reading CC, Teefey SA et al. Reduction in thyroid nodule biopsies and improved accuracy with American College of Radiology Thyroid Imaging Reporting and Data System. Radiology. 2018; 287(1):185-193

93. Hoang JK, Middleton WD, Farjat AE, Teefey SA, Abinanti N, Boschini FJ et al. Interobserver variability of sonographic features used in the American college of radiology thyroid imaging reporting and data system. American Journal of Roentgenology. 2018; 211(1):162-167

94. Hobbs HA, Bahl M, Nelson RC, Eastwood JD, Esclamado RM, Hoang JK. Applying the Society of Radiologists in Ultrasound recommendations for fine-needle aspiration of thyroid nodules: effect on workup and malignancy detection. AJR American Journal of Roentgenology. 2014; 202(3):602-7

95. Hong Y, Liu X, Li Z, Zhang X, Chen M, Luo Z. Real-time ultrasound elastography in the differential diagnosis of benign and malignant thyroid nodules. Journal of Ultrasound in Medicine. 2009; 28(7):861-7

96. Hong YR, Wu YL, Luo ZY, Wu NB, Liu XM. Impact of nodular size on the predictive values of gray-scale, color-Doppler ultrasound, and sonoelastography for assessment of thyroid nodules. Journal of Zhejiang University SCIENCE B. 2012; 13(9):707-16

97. Horvath E, Majlis S, Rossi R, Franco C, Niedmann JP, Castro A et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. Journal of Clinical Endocrinology and Metabolism. 2009; 94(5):1748-51

98. Hu J, Xia J, Tian W. Diagnosis of thyroid cancer using computed tomography. Journal of Medical Imaging and Health Informatics. 2018; 8(5):969-972

99. Huan Q, Wang K, Lou F, Zhang L, Huang QX, Han YF et al. Epidemiological characteristics of thyroid nodules and risk factors for malignant nodules: A retrospective study from 6 304 surgical cases. Chinese Medical Journal. 2014; 127(12):2286-2292

100. Huang X, Guo LH, Xu HX, Gong XH, Liu BJ, Xu JM et al. Acoustic radiation force impulse induced strain elastography and point shear wave elastography for evaluation of thyroid nodules. International Journal of Clinical and Experimental Medicine. 2015; 8(7):10956-63

101. Hughes NM, Nae A, Barry J, Fitzgerald B, Feeley L, Sheahan P. Sonographic differences between conventional and follicular variant papillary thyroid carcinoma. European Archives of Oto-Rhino-Laryngology. 2017; 274(7):2907-2913

102. Ianni F, Campanella P, Rota CA, Prete A, Castellino L, Pontecorvi A et al. A meta-analysis-derived proposal for a clinical, ultrasonographic, and cytological scoring system to evaluate thyroid nodules: the "CUT" score. Endocrine. 2016; 52(2):313-21

103. Ishigaki S, Shimamoto K, Satake H, Sawaki A, Itoh S, Ikeda M et al. Multi-slice CT of thyroid nodules: comparison with ultrasonography. Radiation Medicine. 2004; 22(5):346-53

104. Ito Y, Amino N, Yokozawa T, Ota H, Ohshita M, Murata N et al. Ultrasonographic evaluation of thyroid nodules in 900 patients: comparison among ultrasonographic, cytological, and histological findings. Thyroid. 2007; 17(12):1269-76

105. Jiang J, Shang X, Wang H, Xu YB, Gao Y, Zhou Q. Diagnostic value of contrast-enhanced ultrasound in thyroid nodules with calcification. Kaohsiung Journal of Medical Sciences. 2015; 31(3):138-44

106. Jin XZ, Lu WW, Zhang HF, Yan YY, Gu XL. Comparative study on the diagnostic values of different ultrasound technologies for malignant thyroid nodules. Oncology Letters. 2018; 16(1):910-914

107. Jin ZQ, Lin MY, Hu WH, Li WY, Bai SJ. Gray-scale ultrasonography combined with elastography imaging for the evaluation of papillary thyroid microcarcinoma: as a prognostic clinicopathology factor. Ultrasound in Medicine and Biology. 2014; 40(8):1769-77

108. Kagoya R, Monobe H, Tojima H. Utility of elastography for differential diagnosis of benign and malignant thyroid nodules. Otolaryngology - Head & Neck Surgery. 2010; 143(2):230-4

109. Kakkos SK, Scopa CD, Chalmoukis AK, Karachalios DA, Spiliotis JD, Harkoftakis JG et al. Relative risk of cancer in sonographically detected thyroid nodules with calcifications. Journal of Clinical Ultrasound. 2000; 28(7):347-352

110. Kamran SC, Marqusee E, Kim MI, Frates MC, Ritner J, Peters H et al. Thyroid nodule size and prediction of cancer. Journal of Clinical Endocrinology and Metabolism. 2013; 98(2):564-70

111. Kathuria S, Sen J, Sen R, Singh S, Yadav RK, Sharma JP. The role of Color Doppler sonography in predicting malignancy in thyroid nodules. Ultrasound International. 2003; 9(1):8-14

112. Khamis MEM, Ismail AAA, Alaa El-deen AM, Amin MF. Additional value of qualitative strain ultrasound elastography and strain ratio in predicting thyroid malignancy. Egyptian Journal of Radiology and Nuclear Medicine. 2017; 48(4):913-920

113. Kim DW, Jung SJ, Baek HJ. Computed tomography features of benign and malignant solid thyroid nodules. Acta Radiologica. 2015; 56(10):1196-202

114. Kim EK, Park CS, Chung WY, Oh KK, Kim DI, Lee JT et al. New sonographic criteria for recommending fine-needle aspiration biopsy of nonpalpable solid nodules of the thyroid. AJR American Journal of Roentgenology. 2002; 178(3):687-91

115. Kim H, Kim JA, Son EJ, Youk JH. Quantitative assessment of shear-wave ultrasound elastography in thyroid nodules: diagnostic performance for predicting malignancy. European Radiology. 2013; 23(9):2532-7

116. Kim HG, Moon HJ, Kwak JY, Kim EK. Diagnostic accuracy of the ultrasonographic features for subcentimeter thyroid nodules suggested by the revised American Thyroid Association guidelines. Thyroid. 2013; 23(12):1583-9

117. Kim HJ, Kwak MK, Choi IH, Jin SY, Park HK, Byun DW et al. Utility of shear wave elastography to detect papillary thyroid carcinoma in thyroid nodules: efficacy of the standard deviation elasticity. Korean Journal of Internal Medicine. 2018; Epublication

118. Kim JK, Baek JH, Lee JH, Kim JL, Ha EJ, Kim TY et al. Ultrasound elastography for thyroid nodules: A reliable study? Ultrasound in Medicine and Biology. 2012; 38(9):1508-1513

119. Kim JY, Jung SL, Kim MK, Kim TJ, Byun JY. Differentiation of benign and malignant thyroid nodules based on the proportion of sponge-like areas on ultrasonography: Imaging-pathologic correlation. Ultrasonography. 2015; 34(4):304-311

120. Kim JY, Kim SY, Yang KR. Ultrasonographic criteria for fine needle aspiration of nonpalpable thyroid nodules 1-2 cm in diameter. European Journal of Radiology. 2013; 82(2):321-6

121. Kim JY, Lee CH, Kim SY, Jeon WK, Kang JH, An SK et al. Radiologic and pathologic findings of nonpalpable thyroid carcinomas detected by ultrasonography in a medical screening center. Journal of Ultrasound in Medicine. 2008; 27(2):215-23

122. Kim KM, Park JB, Kang SJ, Bae KS. Ultrasonographic guideline for thyroid nodules cytology: single institute experience. Journal of the Korean Surgical Society. 2013; 84(2):73-9

123. Kim MJ, Kim EK, Kwak JY, Park CS, Chung WY, Nam KH et al. Differentiation of thyroid nodules with macrocalcifications: role of suspicious sonographic findings. Journal of Ultrasound in Medicine. 2008; 27(8):1179-84

124. Kim SC, Kim JH, Choi SH, Yun TJ, Wi JY, Kim SA et al. Off-site evaluation of three-dimensional ultrasound for the diagnosis of thyroid nodules: comparison with two-dimensional ultrasound. European Radiology. 2016; 26(10):3353-60

125. Kim SH, Park CS, Jung SL, Kang BJ, Kim JY, Choi JJ et al. Observer variability and the performance between faculties and residents: US criteria for benign and malignant thyroid nodules. Korean Journal of Radiology. 2010; 11(2):149-55

126. Kim SY, Kim EK, Moon HJ, Yoon JH, Kwak JY. Application of Texture Analysis in the Differential Diagnosis of Benign and Malignant Thyroid Nodules: Comparison With Gray-Scale Ultrasound and Elastography. AJR American Journal of Roentgenology. 2015; 205(3):W343-51

127. Kizilkaya MC, Erozgen F, Akinci M, Kaplan R, Tuzun S, Citlak G. The predictive value of elastography in thyroid nodules and its comparison with fine-needle aspiration biopsy results. Turkish Journal of Surgery. 2014; 30(3):147-52

128. Ko MS, Jeong KS, Shong YK, Gong GY, Baek JH, Lee JH. Collapsing benign cystic nodules of the thyroid gland: sonographic differentiation from papillary thyroid carcinoma. AJNR: American Journal of Neuroradiology. 2012; 33(1):124-7

129. Koh J, Kim SY, Lee HS, Kim EK, Kwak JY, Moon HJ et al. Diagnostic performances and interobserver agreement according to observer experience: a comparison study using three guidelines for management of thyroid nodules. Acta Radiologica. 2018; 59(8):917-923

130. Koh J, Moon HJ, Park JS, Kim SJ, Kim HY, Kim EK et al. Variability in interpretation of ultrasound elastography and gray-scale ultrasound in assessing thyroid nodules. Ultrasound in Medicine and Biology. 2016; 42(1):51-9

131. Koike E, Noguchi S, Yamashita H, Murakami T, Ohshima A, Kawamoto H et al. Ultrasonographic characteristics of thyroid nodules: prediction of malignancy. Archives of Surgery. 2001; 136(3):334-7

132. Koltin D, O'Gorman CS, Murphy A, Ngan B, Daneman A, Navarro OM et al. Pediatric thyroid nodules: ultrasonographic characteristics and inter-observer variability in prediction of malignancy. Journal of Pediatric Endocrinology and Metabolism. 2016; 29(7):789-94

133. Koseoglu Atilla FD, Ozgen Saydam B, Erarslan NA, Diniz Unlu AG, Yilmaz Yasar H, Ozer M et al. Does the ACR TI-RADS scoring allow us to safely avoid unnecessary thyroid biopsy? single center analysis in a large cohort. Endocrine. 2018; 61(3):398-402

134. Kunz W, Mismar A, Wille G, Ahmad R, Materazzi G, Miccoli P. Preoperative prediction of the risk of malignancy in thyroid nodules. Acta Medica Mediterranea. 2014; 30(2):329-334

135. Kwak JY, Han KH, Yoon JH, Moon HJ, Son EJ, Park SH 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

136. Kwak JY, Jung I, Baek JH, Baek SM, Choi N, Choi YJ et al. Image reporting and characterization system for ultrasound features of thyroid nodules: multicentric Korean retrospective study. Korean Journal of Radiology. 2013; 14(1):110-7

137. Lai X, Zhang B, Jiang Y, Li J, Zhao R, Yang X et al. Sonographic and clinical features of papillary thyroid microcarcinoma less than or equal to five millimeters: A retrospective study. PloS One. 2016; 11(2):e0148567

138. Lauria Pantano A, Maddaloni E, Briganti SI, Beretta Anguissola G, Perrella E, Taffon C et al. Differences between ATA, AACE/ACE/AME and ACR TI-RADS ultrasound classifications performance in identifying cytological high-risk thyroid nodules. European Journal of Endocrinology. 2018; 178(6):595-603

139. Lee YH, Kim DW, In HS, Park JS, Kim SH, Eom JW et al. Differentiation between benign and malignant solid thyroid nodules using an US classification system. Korean Journal of Radiology. 2011; 12(5):559-67

140. Li F, Wang Y, Bai B, Wang S, Liu S. Advantages of routine ultrasound combined with contrast-enhanced ultrasound in diagnosing papillary thyroid carcinoma. Ultrasound Quarterly. 2017; 33(3):213-218

141. Li F, Zhang J, Wang Y, Liu L. Clinical value of elasticity imaging and contrast-enhanced ultrasound in the diagnosis of papillary thyroid microcarcinoma. Oncology Letters. 2015; 10(3):1371-1377

142. Li HW, Wu XW, Liu B, Liu WD, Gao N. Clinical values of of gemstone spectral CT in diagnosing thyroid disease. Journal of X-Ray Science & Technology. 2015; 23(1):45-56

143. Li J, Zhang M, Sheng L. The clinical value of real-time tissue elastography in diagnosing thyroid malignancy. International Journal of Clinical and Experimental Medicine. 2018; 11(2):824-829

144. Li JW, Chang C, Chen M, Zeng W, Gao Y, Zhou SC et al. Is Ultrasonography More Sensitive Than Computed Tomography for Identifying Calcifications in Thyroid Nodules? Journal of Ultrasound in Medicine. 2016; 35(10):2183-90

145. Li L, Wang Y, Luo DH, Zhao YF, Lin M, Guo W et al. Diagnostic value of single-source dual-energy spectral computed tomography for papillary thyroid microcarcinomas. Journal of X-Ray Science & Technology. 2017; 25(5):793-802

146. Li L, Wang Y, Zhao Y, Zou S, Lin M, Yu X et al. Evaluation with low-dose dual-phase helical computed tomography of patients with thyroid lesions. Chinese Medical Journal. 2014; 127(22):3937-43

147. Li T, Zhou P, Zhang X, Ding M, Yuchi M, Li Y. Diagnosis of thyroid nodules using virtual touch tissue quantification value and anteroposterior/transverse diameter ratio. Ultrasound in Medicine and Biology. 2015; 41(2):384-92

148. Li WB, Zhang B, Zhu QL, Jiang YX, Sun J, Yang M et al. Comparison between thin-slice 3-D volumetric ultrasound and conventional ultrasound in the differentiation of benign and malignant thyroid lesions. Ultrasound in Medicine and Biology. 2015; 41(12):3096-101

149. Liang JY, Huang XW, Hu HT, Liu YH, Zhou Q, Cao QH et al. Predicting malignancy in thyroid nodules: Radiomics Score versus 2017 American College of Radiology (ACR) Thyroid Imaging, Reporting and Data System (TI-RADS). Thyroid. 2018; 28(8):1024-1033

150. Lim-Dunham JE, Erdem Toslak I, Alsabban K, Aziz A, Martin B, Okur G et al. Ultrasound risk stratification for malignancy using the 2015 American Thyroid Association Management Guidelines for Children with Thyroid Nodules and Differentiated Thyroid Cancer. Pediatric Radiology. 2017; 47(4):429-436

151. Lim KJ, Choi CS, Yoon DY, Chang SK, Kim KK, Han H et al. Computer-aided diagnosis for the differentiation of malignant from benign thyroid nodules on ultrasonography. Academic Radiology. 2008; 15(7):853-8

152. Lin JD, Chao TC, Huang BY, Chen ST, Chang HY, Hsueh C. Thyroid cancer in the thyroid nodules evaluated by ultrasonography and fine-needle aspiration cytology. Thyroid. 2005; 15(7):708-17

153. Lingam RK, Qarib MH, Tolley NS. Evaluating thyroid nodules: predicting and selecting malignant nodules for fine-needle aspiration (FNA) cytology. Insights Into Imaging. 2013; 4(5):617-24

154. Lippolis PV, Tognini S, Materazzi G, Polini A, Mancini R, Ambrosini CE et al. Is elastography actually useful in the presurgical selection of thyroid nodules with indeterminate cytology? Journal of Clinical Endocrinology and Metabolism. 2011; 96(11):E1826-30

155. Liu BJ, Zhao CK, Xu HX, Zhang YF, Xu JM, Li DD et al. Quality measurement on shear wave speed imaging: diagnostic value in differentiation of thyroid malignancy and the associated factors. Oncotarget. 2017; 8(3):4848-4959

156. Liu BX, Xie XY, Liang JY, Zheng YL, Huang GL, Zhou LY et al. Shear wave elastography versus real-time elastography on evaluation thyroid nodules: a preliminary study. European Journal of Radiology. 2014; 83(7):1135-1143

157. Liu J, Zheng D, Li Q, Tang X, Luo Z, Yuan Z et al. A predictive model of thyroid malignancy using clinical, biochemical and sonographic parameters for patients in a multi-center setting. BMC Endocrine Disorders. 2018; 18(1):17

158. Liu MJ, Men YM, Zhang YL, Zhang YX, Liu H. Improvement of diagnostic efficiency in distinguishing the benign and malignant thyroid nodules via conventional ultrasound combined with ultrasound contrast and elastography. Oncology Letters. 2017; 14(1):867-871

159. Liu YI, Kamaya A, Desser TS, Rubin DL. A bayesian network for differentiating benign from malignant thyroid nodules using sonographic and demographic features. AJR American Journal of Roentgenology. 2011; 196(5):W598-605

160. Lu C, Chang TC, Hsiao YL, Kuo MS. Ultrasonographic findings of papillary thyroid carcinoma and their relation to pathologic changes. Journal of the Formosan Medical Association. 1994; 93(11-12):933-8

161. Lu Z, Mu Y, Zhu H, Luo Y, Kong Q, Dou J et al. Clinical value of using ultrasound to assess calcification patterns in thyroid nodules. World Journal of Surgery. 2011; 35(1):122-7

162. Luo S, Kim EH, Dighe M, Kim Y. Thyroid nodule classification using ultrasound elastography via linear discriminant analysis. Ultrasonics. 2011; 51(4):425-31

163. Luo S, Lim DJ, Kim Y. Objective ultrasound elastography scoring of thyroid nodules using spatiotemporal strain information. Medical Physics. 2012; 39(3):1182-9

164. Lyshchik A, Drozd V, Demidchik Y, Reiners C. Diagnosis of thyroid cancer in children: value of gray-scale and power doppler US. Radiology. 2005; 235(2):604-13

165. Ma JJ, Ding H, Xu BH, Xu C, Song LJ, Huang BJ et al. Diagnostic performances of various gray-scale, color Doppler, and contrast-enhanced ultrasonography findings in predicting malignant thyroid nodules. Thyroid. 2014; 24(2):355-63

166. Macedo BM, Izquierdo RF, Golbert L, Meyer ELS. Reliability of Thyroid Imaging Reporting and Data System (TI-RADS), and ultrasonographic classification of the American Thyroid Association (ATA) in differentiating benign from malignant thyroid nodules. Archives of Endocrinology & Metabolism. 2018; 62(2):131-138

167. Maia FF, Matos PS, Pavin EJ, Vassallo J, Zantut-Wittmann DE. Value of ultrasound and cytological classification system to predict the malignancy of thyroid nodules with indeterminate cytology. Endocrine Pathology. 2011; 22(2):66-73

168. Maia FFR, Matos PS, Silva BP, Pallone AT, Pavin EJ, Vassallo J et al. Role of ultrasound, clinical and scintigraphyc parameters to predict malignancy in thyroid nodule. Head and Neck Oncology. 2011; 3:17

169. Maimaiti Y, Sun S, Liu Z, Zeng W, Liu CP, Wang S et al. Diagnostic accuracy of ultrasonographic features for benign and malignant thyroid nodules smaller than 10 mm. International Journal of Clinical and Experimental Medicine. 2016; 9(3):6476-6482

170. Maino F, Forleo R, Martinelli M, Fralassi N, Barbato F, Pilli T et al. Prospective Validation of ATA and ETA Sonographic Pattern Risk of Thyroid Nodules Selected for FNAC. Journal of Clinical Endocrinology and Metabolism. 2018; 103(6):2362-2368

171. Majstorov V. Ultrasonographic findings in patients with benign and malignant thyroid nodules who underwent ultrasound guided fine needle aspiration cytology. Open Access Macedonian Journal of Medical Sciences. 2015; 3(4):689-93

172. Mallikarjunappa B, Ashish SR. Ultrasound evaluation of thyroid nodules and its pathological correlation. Journal International Medical Sciences Academy. 2014; 27(1):9-11

173. Mansor M, Okasha H, Esmat S, Hashem AM, Attia KA, El-din Hussein H. Role of ultrasound elastography in prediction of malignancy in thyroid nodules. Endocrine Research. 2012; 37(2):67-77

174. Marqusee E, Benson CB, Frates MC, Doubilet PM, Larsen PR, Cibas ES et al. Usefulness of ultrasonography in the management of nodular thyroid disease. Annals of Internal Medicine. 2000; 133(9):696-700

175. Martinez-Rios C, Daneman A, Bajno L, van der Kaay DCM, Moineddin R, Wasserman JD. Utility of adult-based ultrasound malignancy risk stratifications in pediatric thyroid nodules. Pediatric Radiology. 2018; 48(1):74-84

176. Mehrotra P, McQueen A, Kolla S, Johnson SJ, Richardson DL. Does elastography reduce the need for thyroid FNAs? Clinical Endocrinology. 2013; 78(6):942-9

177. Memon I, Talpur MKB, Shaikh ZI, Talpur MO. Evaluation of ultrasonography and fine needle aspiration cytology in diagnosing thyroid nodules. Medical Forum Monthly. 2017; 28(5):69-73

178. Merino S, Arrazola J, Cardenas A, Mendoza M, De Miguel P, Fernandez C et al. Utility and interobserver agreement of ultrasound elastography in the detection of malignant thyroid nodules in clinical care. AJNR: American Journal of Neuroradiology. 2011; 32(11):2142-8

179. Middleton WD, Teefey SA, Reading CC, Langer JE, Beland MD, Szabunio MM et al. Multiinstitutional Analysis of Thyroid Nodule Risk Stratification Using the American College of Radiology Thyroid Imaging Reporting and Data System. AJR American Journal of Roentgenology. 2017; 208(6):1331-1341

180. Migda B, Migda M, Migda AM, Bierca J, Slowinska-Srzednicka J, Jakubowski W et al. Evaluation of four variants of the thyroid imaging reporting and data system (TIRADS) classification in patients with multinodular goitre - initial study. Endokrynologia Polska. 2018; 69(2):156-162

181. Migda B, Migda M, Migda MS, Slapa RZ. Use of the Kwak Thyroid Image Reporting and Data System (K-TIRADS) in differential diagnosis of thyroid nodules: systematic review and meta-analysis. European Radiology. 2018; 28(6):2380-2388

182. Mohamed RE, Abodewan KA. Diagnostic utility of real-time ultrasound elastography for prediction of malignancy in solid thyroid nodules. Egyptian Journal of Radiology and Nuclear Medicine. 2013; 44(1):33-43

183. Mohammadi A, Hajizadeh T. Evaluation of diagnostic efficacy of ultrasound scoring system to select thyroid nodules requiring fine needle aspiration biopsy. International Journal of Clinical and Experimental Medicine. 2013; 6(8):641-8

184. Mohey N, Hassan TA, Abdel-Baki S. Role of combined grey scale US and US tissue elastography in differentiating solid thyroid nodules. Egyptian Journal of Radiology and Nuclear Medicine. 2013; 44(3):505-512

185. Moon HG, Jung EJ, Park ST, Ha WS, Choi SK, Hong SC et al. Role of ultrasonography in predicting malignancy in patients with thyroid nodules. World Journal of Surgery. 2007; 31(7):1410-6

186. Moon HJ, Kwak JY, Kim EK, Kim MJ. A taller-than-wide shape in thyroid nodules in transverse and longitudinal ultrasonographic planes and the prediction of malignancy. Thyroid. 2011; 21(11):1249-53

187. Moon HJ, Kwak JY, Kim MJ, Son EJ, Kim EK. Can vascularity at power Doppler US help predict thyroid malignancy? Radiology. 2010; 255(1):260-9

188. Moon HJ, Sung JM, Kim EK, Yoon JH, Youk JH, Kwak JY. Diagnostic performance of gray-scale US and elastography in solid thyroid nodules. Radiology. 2012; 262(3):1002-13

189. Moon WJ, Jung SL, Lee JH, Na DG, Baek JH, Lee YH et al. Benign and malignant thyroid nodules: US differentiation--multicenter retrospective study. Radiology. 2008; 247(3):762-70

190. Na DG, Baek JH, Sung JY, Kim JH, Kim JK, Choi YJ et al. Thyroid imaging reporting and data system risk stratification of thyroid nodules: Categorization based on solidity and echogenicity. Thyroid. 2016; 26(4):562-72

191. Nachiappan M, Poornima V, Kumar A, Prabhu SD, Chouhan RRS. Role of ultrasonography and ultrasound elastography in the evaluation of thyroid nodules. Journal of Clinical and Diagnostic Research. 2018; 12(12):TC13-TC17

192. Nam SJ, Yoo J, Lee HS, Kim EK, Moon HJ, Yoon JH et al. Quantitative Evaluation for Differentiating Malignant and Benign Thyroid Nodules Using Histogram Analysis of Grayscale Sonograms. Journal of Ultrasound in Medicine. 2016; 35(4):775-82

193. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual [updated October 2018]. London. National Institute for Health and Care Excellence, 2014. Available from: http://www.nice.org.uk/article/PMG20/chapter/1%20Introduction%20and%20overview

194. Nemec U, Nemec SF, Novotny C, Weber M, Czerny C, Krestan CR. Quantitative evaluation of contrast-enhanced ultrasound after intravenous administration of a microbubble contrast agent for differentiation of benign and malignant thyroid nodules: assessment of diagnostic accuracy. European Radiology. 2012; 22(6):1357-65

195. Nixon IJ, Ganly I, Hann LE, Lin O, Yu C, Brandt S et al. Nomogram for predicting malignancy in thyroid nodules using clinical, biochemical, ultrasonographic, and cytologic features. Surgery. 2010; 148(6):1120-7; discussion 1127-8

196. Nixon IJ, Ganly I, Hann LE, Yu C, Palmer FL, Whitcher MM et al. Nomogram for selecting thyroid nodules for ultrasound-guided fine-needle aspiration biopsy based on a quantification of risk of malignancy. Head and Neck. 2013; 35(7):1022-5

197. Nobrega LH, Paiva FJ, Nobrega ML, Mello LE, Fonseca HA, Costa SO et al. Predicting malignant involvement in a thyroid nodule: role of ultrasonography. Endocrine Practice. 2007; 13(3):219-24

198. Noda Y, Kanematsu M, Goshima S, Kondo H, Watanabe H, Kawada H et al. MRI of the thyroid for differential diagnosis of benign thyroid nodules and papillary carcinomas. AJR American Journal of Roentgenology. 2015; 204(3):W332-5

199. Nonchev BI, Danev VH, Argatska AV, Miteva MJ, Mateva NG, Orbetzova MM. Prognostic model for ultrasound assessment of malignancy risk of thyroid nodules. Endokrinologya. 2017; 22(3):160-165

200. Okamoto T. Erratum: Test performances of three diagnostic procedures in evaluating thyroid nodules: Physical examination, ultrasonography and fine needle aspiration cytology. Endocrine Journal. 1995; 42(4):x

201. Okamoto T, Yamashita T, Harasawa A, Kanamuro T, Aiba M, Kawakami M et al. Test performances of three diagnostic procedures in evaluating thyroid nodules: physical examination, ultrasonography and fine needle aspiration cytology. Endocrine Journal. 1994; 41(3):243-7

202. Okasha H, Elkholy S, Sayed M, El-Sherbiny M, El-Hussieny R, El-Gemeie E et al. Ultrasound, endoscopic ultrasound elastography, and the strain ratio in differentiating benign from malignant lymph nodes. Arab Journal of Gastroenterology. 2018; 19(1):7-15

203. Oliveira CM, Costa RA, Patricio M, Estevao A, Graca B, Caseiro-Alves F. Sonographic Criteria Predictive of Malignant Thyroid Nodules: Which Lesions Should be Biopsied? Academic Radiology. 2018; 25(2):213-218

204. Ozel A, Erturk SM, Ercan A, Yilmaz B, Basak T, Cantisani V et al. The diagnostic efficiency of ultrasound in characterization for thyroid nodules: how many criteria are required to predict malignancy? Medical Ultrasonography. 2012; 14(1):24-8

205. Palaniappan MK, Aiyappan SK, Ranga U. Role of gray scale, color doppler and spectral doppler in differentiation between malignant and benign thyroid nodules. Journal of Clinical and Diagnostic Research. 2016; 10(8):TC01-TC06

206. Pandey NN, Pradhan GS, Manchanda A, Garg A. Diagnostic value of acoustic radiation force impulse quantification in the differentiation of benign and malignant thyroid nodules. Ultrasonic Imaging. 2017; 39(5):326-336

207. Pandya A, Caoili EM, Jawad-Makki F, Wasnik AP, Shankar PR, Bude R et al. Limitations of the 2015 ATA Guidelines for Prediction of Thyroid Cancer: A Review of 1947 Consecutive Aspirations. Journal of Clinical Endocrinology and Metabolism. 2018; 103(9):3496-3502

208. Pang T, Huang L, Deng Y, Wang T, Chen S, Gong X et al. Logistic regression analysis of conventional ultrasonography, strain elastosonography, and contrast-enhanced ultrasound characteristics for the differentiation of benign and malignant thyroid nodules. PloS One. 2017; 12(12):e0188987

209. Papini E, Guglielmi R, Bianchini A, Crescenzi A, Taccogna S, Nardi F et al. Risk of malignancy in nonpalpable thyroid nodules: predictive value of ultrasound and color-Doppler features. Journal of Clinical Endocrinology and Metabolism. 2002; 87(5):1941-6

210. Park AY, Son EJ, Han K, Youk JH, Kim JA, Park CS. Shear wave elastography of thyroid nodules for the prediction of malignancy in a large scale study. European Journal of Radiology. 2015; 84(3):407-412

211. Park JW, Kim DW, Kim D, Baek JW, Lee YJ, Baek HJ. Korean thyroid imaging reporting and data system features of follicular thyroid adenoma and carcinoma: A single-center study. Ultrasonography. 2017; 36(4):349-354

212. Park JY, Lee HJ, Jang HW, Kim HK, Yi JH, Lee W et al. A proposal for a thyroid imaging reporting and data system for ultrasound features of thyroid carcinoma. Thyroid. 2009; 19(11):1257-1264

213. Park M, Shin JH, Han BK, Ko EY, Hwang HS, Kang SS et al. Sonography of thyroid nodules with peripheral calcifications. Journal of Clinical Ultrasound. 2009; 37(6):324-8

214. Park SJ, Park SH, Choi YJ, Kim DW, Son EJ, Lee HS et al. Interobserver variability and diagnostic performance in US assessment of thyroid nodule according to size. Ultraschall in der Medizin. 2012; 33(7):E186-190

215. Park SY, Hahn SY, Shin JH, Ko EY, Oh YL. The diagnostic performance of thyroid us in each category of the bethesda system for reporting thyroid cytopathology. PloS One. 2016; 11(6):e0155898

216. Pathirana AA, Bandara KG, Faleel MA, Kuruppumullage SD, Solangarachchi N, Rupasinghe RD et al. A sonographic scoring system to assess the risk of thyroid malignancy. Ceylon Medical Journal. 2016; 61(1):32-4

217. Peccin S, de Castsro JA, Furlanetto TW, Furtado AP, Brasil BA, Czepielewski MA. Ultrasonography: is it useful in the diagnosis of cancer in thyroid nodules? Journal of Endocrinological Investigation. 2002; 25(1):39-43

218. Persichetti A, Di Stasio E, Guglielmi R, Bizzarri G, Taccogna S, Misischi I et al. Predictive Value of Malignancy of Thyroid Nodule Ultrasound Classification Systems: A Prospective Study. Journal of Clinical Endocrinology and Metabolism. 2018; 103(4):1359-1368

219. Petrone L, Mannucci E, De Feo ML, Parenti G, Biagini C, Panconesi R et al. A simple ultrasound score for the identification of candidates to fine needle aspiration of thyroid nodules. Journal of Endocrinological Investigation. 2012; 35(8):720-4

220. Phuttharak W, Somboonporn C, Hongdomnern G. Diagnostic performance of gray-scale versus combined gray-scale with colour doppler ultrasonography in the diagnosis of malignancy in thyroid nodules. Asian Pacific Journal of Cancer Prevention: Apjcp. 2009; 10(5):759-64

221. Pompili GG, Tresoldi S, Ravelli A, Primolevo A, Leo GD, Carrafiello G. Use of the ultrasound-based total malignancy score in the management of thyroid nodules. Ultrasonography. 2018; 37(4):315-322

222. Popli MB, Rastogi A, Bhalla P, Solanki Y. Utility of gray-scale ultrasound to differentiate benign from malignant thyroid nodules. Indian Journal of Radiology & Imaging. 2012; 22(1):63-8

223. Popowicz B, Klencki M, Lewinski A, Slowinska-Klencka D. The usefulness of sonographic features in selection of thyroid nodules for biopsy in relation to the nodule's size. European Journal of Endocrinology. 2009; 161(1):103-11

224. Ragazzoni F, Deandrea M, Mormile A, Ramunni MJ, Garino F, Magliona G et al. High diagnostic accuracy and interobserver reliability of real-time elastography in the

evaluation of thyroid nodules. Ultrasound in Medicine and Biology. 2012; 38(7):1154-62

225. Raggiunti B, Capone F, Franchi A, Fiore G, Filipponi S, Colagrande V et al. Ultrasoundelastography: Can it provide valid information for differentiation of benign and malignant thyroid nodules? Journal of Ultrasound. 2011; 14(3):136-41

226. Raghavendra U, Rajendra Acharya U, Gudigar A, Hong Tan J, Fujita H, Hagiwara Y et al. Fusion of spatial gray level dependency and fractal texture features for the characterization of thyroid lesions. Ultrasonics. 2017; 77:110-120

227. Rago T, Di Coscio G, Basolo F, Scutari M, Elisei R, Berti P et al. Combined clinical, thyroid ultrasound and cytological features help to predict thyroid malignancy in follicular and Hupsilonrthle cell thyroid lesions: results from a series of 505 consecutive patients. Clinical Endocrinology. 2007; 66(1):13-20

228. Rago T, Santini F, Scutari M, Pinchera A, Vitti P. Elastography: new developments in ultrasound for predicting malignancy in thyroid nodules. Journal of Clinical Endocrinology and Metabolism. 2007; 92(8):2917-22

229. Rago T, Scutari M, Loiacono V, Santini F, Tonacchera M, Torregrossa L et al. Low elasticity of thyroid nodules on ultrasound elastography is correlated with malignancy, degree of fibrosis, and high expression of Galectin-3 and Fibronectin-1. Thyroid. 2017; 27(1):103-110

230. Rago T, Vitti P, Chiovato L, Mazzeo S, De Liperi A, Miccoli P et al. Role of conventional ultrasonography and color flow-doppler sonography in predicting malignancy in 'cold' thyroid nodules. European Journal of Endocrinology. 1998; 138(1):41-6

231. Rahal AJ, Falsarella PM, Rocha RD, Lima JP, Iani MJ, Vieira FA et al. Correlation of Thyroid Imaging Reporting and Data System [TI-RADS] and fine needle aspiration: experience in 1,000 nodules. Einstein. 2016; 14(2):119-23

232. Ram N, Hafeez S, Qamar S, Hussain SZ, Asghar A, Anwar Z et al. Diagnostic validity of ultrasonography in thyroid nodules. JPMA - Journal of the Pakistan Medical Association. 2015; 65(8):875-8

233. Rao G, Rao S, Varma R, Bhagwat NM, Chadha MD, Joshi AS et al. Predicting malignancy in a solitary thyroid nodule: A prospective study on the role of color Doppler ultrasonography. Otorhinolaryngology Clinics. 2014; 6(1):9-14

234. Razavi SA, Hadduck TA, Sadigh G, Dwamena BA. Comparative effectiveness of elastographic and B-mode ultrasound criteria for diagnostic discrimination of thyroid nodules: a meta-analysis (Provisional abstract). AJR American Journal of Roentgenology. 2013; 200(6):1317-1326

235. Razek AA, Sadek AG, Kombar OR, Elmahdy TE, Nada N. Role of apparent diffusion coefficient values in differentiation between malignant and benign solitary thyroid nodules. AJNR: American Journal of Neuroradiology. 2008; 29(3):563-8

236. Refaat R, Kamel A, Elganzory M, Awad NM. Can real-time ultrasound elastography using the color score and strain ratio differentiate between benign and malignant solitary thyroid nodules? Egyptian Journal of Radiology and Nuclear Medicine. 2014; 45(1):75-87

237. Reginelli A, Urraro F, di Grezia G, Napolitano G, Maggialetti N, Cappabianca S et al. Conventional ultrasound integrated with elastosonography and B-flow imaging in the diagnosis of thyroid nodular lesions. International Journal of Surgery. 2014; 12(Suppl 1):S117-22

238. Rios A, Rodriguez JM, Torregrosa NM, Torregrosa B, Cepero A, Abellan MD et al. Evaluation of the thyroid nodule with high-resolution ultrasonography and elastography without fine needle aspiration biopsy. Medicina Clínica. 2018; 151(3):89-96

239. Rios A, Torregrosa B, Rodriguez JM, Rodriguez D, Cepero A, Abellan MD et al. Ultrasonographic risk factors of malignancy in thyroid nodules. Langenbeck's Archives of Surgery. 2016; 401(6):839-849

240. Rivo-Vazquez A, Rodriguez-Lorenzo A, Rivo-Vazquez JE, Paramo-Fernandez C, Garcia-Lorenzo F, Pardellas-Rivera H et al. The use of ultrasound elastography in the assessment of malignancy risk in thyroid nodules and multinodular goitres. Clinical Endocrinology. 2013; 79(6):887-91

241. Rosario PW, Da Silva AL, Nunes MB, Borges MAR. Risk of Malignancy in Thyroid Nodules Using the American College of Radiology Thyroid Imaging Reporting and Data System in the NIFTP Era. Hormone and Metabolic Research. 2018; 50(10):735-737

242. Rosario PW, Silva AL, Borges MA, Calsolari MR. Is Doppler ultrasound of additional value to gray-scale ultrasound in differentiating malignant and benign thyroid nodules? Archives of Endocrinology & Metabolism. 2015; 59(1):79-83

243. Russ G, Bienvenu-Perrard M, Rouxel A, Royer B, Bigorgne C. Ti-RADS score: Clinical efficiency evaluated with the Bethesda system in a one-year prospective study on 2480 nodules. European Thyroid Journal. 2011; Conference Publication:96

244. Russ G, Rouxel A, Bienvenu-Perrard M, Bigorgne C, Royer B. Do ultrasound and elastographic patterns of thyroid carcinomas vary with their size? European Thyroid Journal. 2011; Conference Publication:157

245. Sagazio A, Bucci I, Olivieri M, Di Dalmazi G, Carpenito A, Giuliani C et al. Real time elastosonography versus cytology: Diagnostic role in thyroid lesions. Endocrine Reviews Conference: 96th Annual Meeting and Expo of the Endocrine Society, ENDO. 2014; 35(SUPPL. 3)

246. Sahbaz NA, Dural AC, Akarsu C, Akbulut S, Turkay R, Baytekin HF et al. Feasibility of thyroid imaging reporting and data system (TIRADS) classification in predicting thyroid malignancy. Langenbeck's Archives of Surgery. 2017; 402(2):360-361

247. Saito OC, Chammas MC, Tochetto S, Cerri GG. Elastography of thyroid nodules: Which elastography index can predict malignancy? Ultrasound in Medicine and Biology. 2015; 41(4 Suppl):S37-S38

248. Sajjadieh HR, Sajjadieh V, Aminorroaya A, Amini M, Oveisgharan S, Reisifar M et al. Value of sonography in determining the nature of thyroid nodules: Evaluation of the sonographic characteristics. Journal of Diagnostic Medical Sonography. 2005; 21(1):38-44

249. Salehi M, Nalaini F, Izadi B, Setayeshi K, Rezaei M, Naseri SN. Gray-scale vs. color doppler ultrasound in cold thyroid nodules. Global Journal of Health Science. 2014; 7(3):147-52

250. Salmaslioglu A, Erbil Y, Dural C, Issever H, Kapran Y, Ozarmagan S et al. Predictive value of sonographic features in preoperative evaluation of malignant thyroid nodules in a multinodular goiter. World Journal of Surgery. 2008; 32(9):1948-54

251. Samulski TD, Shutty C, LiVolsi VA, Montone K, Baloch Z. The reliability of thyroid nodule ultrasound features and size to predict malignancy in fine needle aspiration

specimens: Practical utility for the evaluating pathologist. Diagnostic Cytopathology. 2015; 43(6):471-7

252. Sands NB, Karls S, Amir A, Tamilia M, Gologan O, Rochon L et al. McGill Thyroid Nodule Score (MTNS): "Rating the risk," a novel predictive scheme for cancer risk determination. Journal of Otolaryngology - Head and Neck Surgery. 2011; 40(SUPPL. 1):S1-S13

253. Sarabia JJU, Agno MN, Fabile JL, Delos Santos NC. Role of ultrasound in evaluation and differentiation of benign from malignant thyroid nodules using thyroid imaging reporting and data system in a tertiary government hospital. Journal of the American College of Surgeons. 2017; 225(Supplement 2):e94-e95

254. Schenke S, Rink T, Zimny M. TIRADS for sonographic assessment of hypofunctioning and indifferent thyroid nodules. Nuklearmedizin. 2015; 54(3):144-50

255. Schenke S, Seifert P, Zimny M, Winkens T, Binse I, Gorges R. Risk Stratification of Thyroid Nodules Using the Thyroid Imaging Reporting and Data System (TIRADS): The Omission of Thyroid Scintigraphy Increases the Rate of Falsely Suspected Lesions. Journal of Nuclear Medicine. 2019; 60(3):342-347

256. Schueller-Weidekamm C, Kaserer K, Schueller G, Scheuba C, Ringl H, Weber M et al. Can quantitative diffusion-weighted MR imaging differentiate benign and malignant cold thyroid nodules? Initial results in 25 patients. AJNR: American Journal of Neuroradiology. 2009; 30(2):417-22

257. Sebag F, Vaillant-Lombard J, Berbis J, Griset V, Henry JF, Petit P et al. Shear wave elastography: a new ultrasound imaging mode for the differential diagnosis of benign and malignant thyroid nodules. Journal of Clinical Endocrinology and Metabolism. 2010; 95(12):5281-8

258. Seo H, Na DG, Kim JH, Kim KW, Yoon JW. Ultrasound-Based Risk Stratification for Malignancy in Thyroid Nodules: A Four-Tier Categorization System. European Radiology. 2015; 25(7):2153-62

259. Seo JK, Kim YJ, Kim KG, Shin I, Shin JH, Kwak JY. Differentiation of the Follicular Neoplasm on the Gray-Scale US by Image Selection Subsampling along with the Marginal Outline Using Convolutional Neural Network. BioMed Research International. 2017; 2017:3098293

260. Seo YL, Yoon DY, Yoon SJ, Lim KJ, Yun EJ, Choi CS et al. Compressibility of thyroid masses: a sonographic sign differentiating benign from malignant lesions? AJR American Journal of Roentgenology. 2012; 198(2):434-8

261. Shankar K, Kumar MV, Kumar MVV, S K. Ultrasound of the thyroid: A preoperative tool to assess the risk of malignancy. European Journal of Cancer. 2015; 51(Suppl 3):S574

262. Shao J, Shen Y, Lu J, Wang J. Ultrasound scoring in combination with ultrasound elastography for differentiating benign and malignant thyroid nodules. Clinical Endocrinology. 2015; 83(2):254-60

263. Shi HF, Feng Q, Qiang JW, Li RK, Wang L, Yu JP. Utility of diffusion-weighted imaging in differentiating malignant from benign thyroid nodules with magnetic resonance imaging and pathologic correlation. Journal of Computer Assisted Tomography. 2013; 37(4):505-10

264. Shimura H, Haraguchi K, Hiejima Y, Fukunari N, Fujimoto Y, Katagiri M et al. Distinct diagnostic criteria for ultrasonographic examination of papillary thyroid carcinoma: a multicenter study. Thyroid. 2005; 15(3):251-8

265. Shrestha M, Crothers BA, Burch HB. The impact of thyroid nodule size on the risk of malignancy and accuracy of fine-needle aspiration: a 10-year study from a single institution. Thyroid. 2012; 22(12):1251-6

266. Shuzhen C. Comparison analysis between conventional ultrasonography and ultrasound elastography of thyroid nodules. European Journal of Radiology. 2012; 81(8):1806-11

267. Siderova MV, Hristozov K, Krasnaliev I. Differential diagnosis of thyroid nodules using strain ultrasound elastography. European Thyroid Journal. 2016; 5 (Suppl 1):159-160

268. Simon D, Hetkamp P, Farahati J, Hautzel H, Gorges R. Diagnostic assessment of dignity of thyroid nodules: The role of scintigraphy and FNAB with application of TIRADS. Langenbeck's Archives of Surgery. 2017; 402(7):1139

269. Singaporewalla RM, Hwee J, Lang TU, Desai V. Clinico-pathological correlation of thyroid nodule ultrasound and cytology using the TIRADS and Bethesda classifications. World Journal of Surgery. 2017; 41(7):1807-1811

270. Stacul F, Bertolotto M, De Gobbis F, Calderan L, Cioffi V, Romano A et al. US, colour-Doppler US and fine-needle aspiration biopsy in the diagnosis of thyroid nodules. Radiologia Medica. 2007; 112(5):751-62

271. Stoian D, Timar B, Derban M, Pantea S, Varcus F, Craina M et al. Thyroid Imaging Reporting and Data System (TI-RADS): the impact of quantitative strain elastography for better stratification of cancer risks. Medical Ultrasonography. 2015; 17(3):327-32

272. Sui X, Liu HJ, Jia HL, Fang QM. Contrast-enhanced ultrasound and real-time elastography in the differential diagnosis of malignant and benign thyroid nodules. Experimental and Therapeutic Medicine. 2016; 12(2):783-791

273. Sun J, Cai J, Wang X. Real-time ultrasound elastography for differentiation of benign and malignant thyroid nodules: a meta-analysis (Provisional abstract). Journal of Ultrasound in Medicine. 2014; 33(3):495-502

274. Swan KZ, Nielsen VE, Bibby BM, Bonnema SJ. Is the reproducibility of shear wave elastography of thyroid nodules high enough for clinical use? A methodological study. Clinical Endocrinology. 2017; 86(4):606-613

275. Szczepanek-Parulska E, Wolinski K, Stangierski A, Gurgul E, Biczysko M, Majewski P et al. Comparison of diagnostic value of conventional ultrasonography and shear wave elastography in the prediction of thyroid lesions malignancy. PloS One. 2013; 8(11):e81532

276. Tae HJ, Lim DJ, Baek KH, Park WC, Lee YS, Choi JE et al. Diagnostic value of ultrasonography to distinguish between benign and malignant lesions in the management of thyroid nodules. Thyroid. 2007; 17(5):461-6

277. Taghipour Zahir S, Binesh F, Mirouliaei M, Khajeh E, Noshad S. Malignancy risk assessment in patients with thyroid nodules using classification and regression trees. Journal of Thyroid Research. 2013; 2013:983953

278. Taha Ali TF. Solitary thyroid nodule: Diagnostic yield of combined diffusion weighted imaging and magnetic resonance spectroscopy. Egyptian Journal of Radiology and Nuclear Medicine. 2017; 48(3):593-601

279. Tahmasebi M, Dezfouli MRB, Gharibvand MM, Jahanshahi A, Nikpour N, Rahim F. Diagnostic accuracy of sonography in assessment of thyroid masses in comparison with pathology. Russian Open Medical Journal. 2016; 5(1):e0103

280. Tamsel S, Demirpolat G, Erdogan M, Nart D, Karadeniz M, Uluer H et al. Power Doppler US patterns of vascularity and spectral Doppler US parameters in predicting malignancy in thyroid nodules. Clinical Radiology. 2007; 62(3):245-51

281. Tang AL, Falciglia M, Yang H, Mark JR, Steward DL. Validation of american thyroid association ultrasound risk assessment of thyroid nodules selected for ultrasound fine-needle aspiration. Thyroid. 2017; 27(8):1077-1082

282. Tatar IG, Kurt A, Yilmaz KB, Akinci M, Kulacoglu H, Hekimoglu B. The learning curve of real time elastosonography: a preliminary study conducted for the assessment of malignancy risk in thyroid nodules. Medical Ultrasonography. 2013; 15(4):278-84

283. Tatar IG, Kurt A, Yilmaz KB, Dogan M, Hekimoglu B, Hucumenoglu S. The role of elastosonography, gray-scale and colour flow Doppler sonography in prediction of malignancy in thyroid nodules. Radiology & Oncology. 2014; 48(4):348-53

284. Tezelman S, Giles Y, Tunca F, Gok K, Poyanli A, Salmaslioglu A et al. Diagnostic value of dynamic contrast medium enhanced magnetic resonance imaging in preoperative detection of thyroid carcinoma. Archives of Surgery. 2007; 142(11):1036-41

285. Trimboli P, Guglielmi R, Monti S, Misischi I, Graziano F, Nasrollah N et al. Ultrasound sensitivity for thyroid malignancy is increased by real-time elastography: a prospective multicenter study. Journal of Clinical Endocrinology and Metabolism. 2012; 97(12):4524-30

286. Tugendsam C, Petz V, Buchinger W, Schmoll-Hauer B, Schenk IP, Rudolph K et al. Ultrasound criteria for risk stratification of thyroid nodules in the previously iodine deficient area of Austria - a single centre, retrospective analysis. Thyroid research. 2018; 11:3

287. Tunca F, Giles Y, Salmaslioglu A, Poyanli A, Yilmazbayhan D, Terzioglu T et al. The preoperative exclusion of thyroid carcinoma in multinodular goiter: Dynamic contrast-enhanced magnetic resonance imaging versus ultrasonography-guided fine-needle aspiration biopsy. Surgery. 2007; 142(6):992-1002; discussion 1002.e1-2

288. Tuzun D, Ersoy R, Kilicyazgan A, Kiyak G, Yalcin S, Cakir B. Elastosonography scoring and strain index of thyroid nodules with Hurthle cells. Minerva Endocrinologica. 2016; 41(2):157-65

289. Unluturk U, Erdogan MF, Demir O, Gullu S, Baskal N. Ultrasound elastography is not superior to grayscale ultrasound in predicting malignancy in thyroid nodules. Thyroid. 2012; 22(10):1031-8

290. Vargas-Uricoechea H, Meza-Cabrera I, Herrera-Chaparro J. Concordance between the TIRADS ultrasound criteria and the BETHESDA cytology criteria on the nontoxic thyroid nodule. Thyroid research. 2017; 10:1

291. Varverakis E, Neonakis E. Contribution of high-resolution ultrasonography in the differential diagnosis of benign from malignant thyroid nodules. Hormones. 2002; 1(1):51-6

292. Veyrieres JB, Albarel F, Lombard JV, Berbis J, Sebag F, Oliver C et al. A threshold value in Shear Wave elastography to rule out malignant thyroid nodules: a reality? European Journal of Radiology. 2012; 81(12):3965-72

293. Vidal-Casariego A, Lopez-Gonzalez L, Jimenez-Perez A, Ballesteros-Pomar MD, Kyriakos G, Urioste-Fondo A et al. Accuracy of ultrasound elastography in the diagnosis of thyroid cancer in a low-risk population. Experimental and clinical endocrinology & diabetes. 2012; 120(10):635-638

294. Vorlander C, Wolff J, Saalabian S, Lienenluke RH, Wahl RA. Real-time ultrasound elastography--a noninvasive diagnostic procedure for evaluating dominant thyroid nodules. Langenbeck's Archives of Surgery. 2010; 395(7):865-71

295. Wang D, He YP, Zhang YF, Liu BJ, Zhao CK, Fu HJ et al. The diagnostic performance of shear wave speed (SWS) imaging for thyroid nodules with elasticity modulus and SWS measurement. Oncotarget. 2017; 8(8):13387-13399

296. Wang F, Chang C, Chen M, Gao Y, Chen YL, Zhou SC et al. Does Lesion Size Affect the Value of Shear Wave Elastography for Differentiating Between Benign and Malignant Thyroid Nodules? Journal of Ultrasound in Medicine. 2018; 37(3):601-609

297. Wang H, Brylka D, Sun LN, Lin YQ, Sui GQ, Gao J. Comparison of strain ratio with elastography score system in differentiating malignant from benign thyroid nodules. Clinical Imaging. 2013; 37(1):50-5

298. Wang H, Zhao L, Xin X, Wei X, Zhang S, Li Y et al. Diagnostic value of elastosonography for thyroid microcarcinoma. Ultrasonics. 2014; 54(7):1945-9

299. Wang HL, Zhang S, Xin XJ, Zhao LH, Li CX, Mu JL et al. Application of Real-time Ultrasound Elastography in Diagnosing Benign and Malignant Thyroid Solid Nodules. Cancer Biology & Medicine. 2012; 9(2):124-7

300. Wang M, Wu WD, Chen GM, Chou SL, Dai XM, Xu JM et al. Could Tumor Size Be A Predictor for Papillary Thyroid Microcarcinoma: a Retrospective Cohort Study. Asian Pacific Journal of Cancer Prevention: Apjcp. 2015; 16(18):8625-8

301. Wang N, Xu Y, Ge C, Guo R, Guo K. Association of sonographically detected calcification with thyroid carcinoma. Head and Neck. 2006; 28(12):1077-83

302. Wang Y, Lei KR, He YP, Li XL, Ren WW, Zhao CK et al. Malignancy risk stratification of thyroid nodules: comparisons of four ultrasound Thyroid Imaging Reporting and Data Systems in surgically resected nodules. Scientific Reports. 2017; 7:11560

303. Wang Y, Nie F, Liu T, Yang D, Li Q, Li J et al. Revised value of contrast-enhanced ultrasound for solid hypo-echoic thyroid nodules graded with the thyroid imaging reporting and data system. Ultrasound in Medicine and Biology. 2018; 44(5):930-940

304. Wang Z, Zhang H, Zhang P, He L, Dong W. Diagnostic value of ultrasound-detected calcification in thyroid nodules. Annals of the Academy of Medicine, Singapore. 2014; 43(2):102-6

305. Watters DA, Ahuja AT, Evans RM, Chick W, King WW, Metreweli C et al. Role of ultrasound in the management of thyroid nodules. American Journal of Surgery. 1992; 164(6):654-7

306. Wei X, Li Y, Zhang S, Gao M. Thyroid imaging reporting and data system (TI-RADS) in the diagnostic value of thyroid nodules: a systematic review. Tumour Biology. 2014; 35(7):6769-6776

307. Wei X, Li Y, Zhang S, Gao M. Meta-analysis of thyroid imaging reporting and data system in the ultrasonographic diagnosis of 10,437 thyroid nodules. Head and Neck. 2016; 38(2):309-15

308. Wei Y, Zhou X, Liu S, Wang H, Liu L, Liu R et al. Novel and practical scoring systems for the diagnosis of thyroid nodules. PloS One. 2016; 11(9):e0163039

309. Weiss VL, Andreotti RF, Ely KA. Use of the Thyroid Imaging, Reporting, and Data System (TI-RADS) scoring system for the evaluation of subcentimeter thyroid nodules. Cancer Cytopathology. 2018; 126(8):518-524

310. Wharry LI, McCoy KL, Stang MT, Armstrong MJ, LeBeau SO, Tublin ME et al. Thyroid nodules (>=4 cm): can ultrasound and cytology reliably exclude cancer? World Journal of Surgery. 2014; 38(3):614-21

311. Witczak J, Taylor P, Chai J, Amphlett B, Soukias JM, Das G et al. Predicting malignancy in thyroid nodules: feasibility of a predictive model integrating clinical, biochemical, and ultrasound characteristics. Thyroid research. 2016; 9:4

312. Wu Q, Li J, Wu J, Li X, Chen C, Jiang N et al. Risk for malignancy of thyroid nodules based on ultrasound imaging characteristics. International Journal of Clinical and Experimental Medicine. 2016; 9(6):11817-11823

313. Wu Q, Qu Y, Zang X, Li Y, Yi X, Wang Y et al. Preliminary study of confounding factors of elastography and the application of fine-needle aspiration in thyroid nodules with indeterminate elastography. Scientific Reports. 2017; 7:18005

314. Wu Q, Wang Y, Li Y, Hu B, He ZY. Diagnostic value of contrast-enhanced ultrasound in solid thyroid nodules with and without enhancement. Endocrine. 2016; 53(2):480-8

315. Wu R, Zhu L, Li W, Tang Q, Pan F, Wu W et al. External validation of a nomogram that predicts the pathological diagnosis of thyroid nodules in a Chinese population. PloS One. 2013; 8(6):e65162

316. Xia J, Chen H, Li Q, Zhou M, Chen L, Cai Z et al. Ultrasound-based differentiation of malignant and benign thyroid Nodules: An extreme learning machine approach. Computer Methods and Programs in Biomedicine. 2017; 147:37-49

317. Xing P, Chen Q, Yang ZW, Liu CB, Wu CJ. Combination of conventional ultrasound and tissue quantification using acoustic radiation force impulse technology for differential diagnosis of small thyroid nodules. International Journal of Clinical and Experimental Medicine. 2016; 9(5):8288-8295

318. Xing P, Wu L, Zhang C, Li S, Liu C, Wu C. Differentiation of benign from malignant thyroid lesions: calculation of the strain ratio on thyroid sonoelastography. Journal of Ultrasound in Medicine. 2011; 30(5):663-9

319. Xu JM, Xu HX, Xu XH, Liu C, Zhang YF, Guo LH et al. Solid hypo-echoic thyroid nodules on ultrasound: the diagnostic value of acoustic radiation force impulse elastography. Ultrasound in Medicine and Biology. 2014; 40(9):2020-30

320. Xu JM, Xu HX, Zhang YF, Guo LH, Liu LN, Bo XW et al. Virtual Touch Tissue Imaging for Differential Diagnosis of Thyroid Nodules: Additional Value of the Area Ratio. Journal of Ultrasound in Medicine. 2016; 35(5):917-26

321. Xu JM, Xu XH, Xu HX, Zhang YF, Zhang J, Guo LH et al. Conventional US, US elasticity imaging, and acoustic radiation force impulse imaging for prediction of malignancy in thyroid nodules. Radiology. 2014; 272(2):577-86

322. Xu SY, Zhan WW, Wang WH. Evaluation of Thyroid Nodules by a Scoring and Categorizing Method Based on Sonographic Features. Journal of Ultrasound in Medicine. 2015; 34(12):2179-85

323. Xu T, Gu JY, Ye XH, Xu SH, Wu Y, Shao XY et al. Thyroid nodule sizes influence the diagnostic performance of TIRADS and ultrasound patterns of 2015 ATA guidelines: a multicenter retrospective study. Scientific Reports. 2017; 7:43183

324. Xu T, Wu Y, Wu RX, Zhang YZ, Gu JY, Ye XH et al. Validation and comparison of three newly-released Thyroid Imaging Reporting and Data Systems for cancer risk determination. Endocrine. 2018; Epublication

325. Xue E, Zheng M, Zhang S, Huang L, Qian Q, Huang Y. Ultrasonography-Based Classification and Reporting System for the Malignant Risk of Thyroid Nodules. Journal of Nippon Medical School. 2017; 84(3):118-124

326. Xue J, Cao XL, Shi L, Lin CH, Wang J, Wang L. The diagnostic value of combination of TI-RADS and ultrasound elastography in the differentiation of benign and malignant thyroid nodules. Clinical Imaging. 2016; 40(5):913-6

327. Yang J, Song Y, Wei W, Ruan L, Ai H. Comparison of the effectiveness of ultrasound elastography with that of conventional ultrasound for differential diagnosis of thyroid lesions with suspicious ultrasound features. Oncology Letters. 2017; 14(3):3515-3521

328. Yerli H, Yilmaz T. Diastolic sonoelastographic strain index for the management of thyroid nodules. Erciyes Medical Journal. 2017; 39(2):48-53

329. Yoon JH, Han K, Kim EK, Moon HJ, Kwak JY. Diagnosis and management of small thyroid nodules: A comparative study with six guidelines for thyroid nodules. Radiology. 2017; 283(2):560-569

330. Yoon JH, Kim EK, Kwak JY, Park VY, Moon HJ. Application of various additional imaging techniques for thyroid ultrasound: Direct comparison of combined various elastography and Doppler parameters to gray-scale ultrasound in differential diagnosis of thyroid nodules. Ultrasound in Medicine and Biology. 2018; 44(8):1679-1686

331. Yoon JH, Lee HS, Kim EK, Moon HJ, Kwak JY. Malignancy risk stratification of thyroid nodules: Comparison between the thyroid Imaging Reporting and Data System and the 2014 American Thyroid Association management guidelines. Radiology. 2016; 278(3):917-924

332. Yoon JH, Shin HJ, Kim EK, Moon HJ, Roh YH, Kwak JY. Quantitative evaluation of vascularity using 2-d power doppler ultrasonography may not identify malignancy of the thyroid. Ultrasound in Medicine and Biology. 2015; 41(11):2873-83

333. Yoon JH, Yoo J, Kim EK, Moon HJ, Lee HS, Seo JY et al. Real-time elastography in the evaluation of diffuse thyroid disease: a study based on elastography histogram parameters. Ultrasound in Medicine and Biology. 2014; 40(9):2012-9

334. Yu Q, Jiang T, Zhou A, Zhang L, Zhang C, Xu P. Computer-aided diagnosis of malignant or benign thyroid nodes based on ultrasound images. European Archives of Oto-Rhino-Laryngology. 2017; 274(7):2891-2897

335. Yuan Y, Yue XH, Tao XF. The diagnostic value of dynamic contrast-enhanced MRI for thyroid tumors. European Journal of Radiology. 2012; 81(11):3313-8

336. Yuan Z, Quan J, Yunxiao Z, Jian C, Zhu H. Contrast-enhanced ultrasound in the diagnosis of solitary thyroid nodules. Journal of Cancer Research and Therapeutics. 2015; 11(1):41-5

337. Yunus M, Ahmed Z. Significance of ultrasound features in predicting malignant solid thyroid nodules: need for fine-needle aspiration. JPMA - Journal of the Pakistan Medical Association. 2010; 60(10):848-53

338. Zayadeen AR, Abu-Yousef M, Berbaum K. Retrospective evaluation of ultrasound features of thyroid nodules to assess malignancy risk: A step toward tirads. American Journal of Roentgenology. 2016; 207(3):460-469

339. Zhan J, Diao XH, Chen L, Jin JM, Chen Y. Role of contrast-enhanced ultrasound in diagnosis of thyroid nodules in acoustic radiation force impulse "gray zone". Ultrasound in Medicine and Biology. 2017; 43(6):1179-1186

340. Zhang B, Jiang YX, Liu JB, Yang M, Dai Q, Zhu QL et al. Utility of contrast-enhanced ultrasound for evaluation of thyroid nodules. Thyroid. 2010; 20(1):51-7

341. Zhang F, Zhao X, Han R, Du M, Li P, Ji X. Comparison of Acoustic Radiation Force Impulse Imaging and Strain Elastography in Differentiating Malignant From Benign Thyroid Nodules. Journal of Ultrasound in Medicine. 2017; 36(12):2533-2543

342. Zhang FJ, Han RL. The value of acoustic radiation force impulse (ARFI) in the differential diagnosis of thyroid nodules. European Journal of Radiology. 2013; 82(11):e686-90

343. Zhang FJ, Han RL, Zhao XM. The value of virtual touch tissue image (VTI) and virtual touch tissue quantification (VTQ) in the differential diagnosis of thyroid nodules. European Journal of Radiology. 2014; 83(11):2033-40

344. Zhang H, Gu J, Xing J, Bai M, Gao F, Du L. Nodule size is a key factor for differentiating benign and malignant thyroid nodules using virtual touch tissue quantification and conventional sonography. International Journal of Clinical and Experimental Medicine. 2017; 10(5):7825-7833

345. Zhang H, Shi Q, Gu J, Jiang L, Bai M, Liu L et al. Combined value of Virtual Touch tissue quantification and conventional sonographic features for differentiating benign and malignant thyroid nodules smaller than 10 mm. Journal of Ultrasound in Medicine. 2014; 33(2):257-64

346. Zhang J, Liu BJ, Xu HX, Xu JM, Zhang YF, Liu C et al. Prospective validation of an ultrasound-based thyroid imaging reporting and data system (TI-RADS) on 3980 thyroid nodules. International Journal of Clinical and Experimental Medicine. 2015; 8(4):5911-7

347. Zhang S, Zheng XQ, Wei X, Zhao J, Zhang Y, Gao M. Evaluation of ultrasound application in diagnosis and clinical staging of thyroid cancers. Indian Journal of Cancer. 2014; 51(3):193-199

348. Zhang T, Li F, Mu J, Liu J, Zhang S. Multivariate evaluation of Thyroid Imaging Reporting and Data System (TI-RADS) in diagnosis malignant thyroid nodule: application to PCA and PLS-DA analysis. International Journal of Clinical Oncology. 2017; 22(3):448-454

349. Zhang XL, Qian LX. Ultrasonic features of papillary thyroid microcarcinoma and non-microcarcinoma. Experimental and Therapeutic Medicine. 2014; 8(4):1335-1339

350. Zhang XY, Zhang B, Jiang YX, Yang X, Zhao RN, Lai XJ. Diagnostic Value of Elastography for Thyroid Nodules in Hashimoto's Thyroiditis. Chung-Kuo i Hsueh Ko Hsueh Yuan Hsueh Pao Acta Academiae Medicinae Sinicae. 2018; 40(1):59-66

351. Zhang Y, Luo YK, Tang J, Li M, Wang ZL, Wen Q. Clinical Value of Ultrasonography in Diagnosing Diffuse Thyroid Diseases Accompanied with Suspicious Nodules. Chung-Kuo i Hsueh Ko Hsueh Yuan Hsueh Pao Acta Academiae Medicinae Sinicae. 2015; 37(3):290-3

352. Zhang Y, Luo YK, Zhang MB, Li J, Li J, Tang J. Diagnostic accuracy of contrast-enhanced ultrasound enhancement patterns for thyroid nodules. Medical Science Monitor. 2016; 22:4755-4764

353. Zhang Y, Meng F, Hong L, Chu L. A Risk Score Model for Evaluation and Management of Patients with Thyroid Nodules. Hormone and Metabolic Research. 2018; 50(7):543-550

354. Zhang Y, Zhou P, Tian SM, Zhao YF, Li JL, Li L. Usefulness of combined use of contrast-enhanced ultrasound and TI-RADS classification for the differentiation of benign from malignant lesions of thyroid nodules. European Radiology. 2017; 27(4):1527-1536

355. Zhang YF, He Y, Xu HX, Xu XH, Liu C, Guo LH et al. Virtual touch tissue imaging on acoustic radiation force impulse elastography: a new technique for differential diagnosis between benign and malignant thyroid nodules. Journal of Ultrasound in Medicine. 2014; 33(4):585-95

356. Zhang YF, Xu HX, He Y, Liu C, Guo LH, Liu LN et al. Virtual touch tissue quantification of acoustic radiation force impulse: a new ultrasound elastic imaging in the diagnosis of thyroid nodules. PloS One. 2012; 7(11):e49094

357. Zhang YF, Xu HX, Xu JM, Liu C, Guo LH, Liu LN et al. Acoustic radiation force impulse elastography in the diagnosis of thyroid nodules: Useful or not useful? Ultrasound in Medicine and Biology. 2015; 41(10):2581-93

358. Zhang YZ, Xu T, Cui D, Li X, Yao Q, Gong HY et al. Value of TIRADS, BSRTC and FNA-BRAF V600E mutation analysis in differentiating high-risk thyroid nodules. Scientific Reports. 2015; 5:16927

359. Zhang YZ, Xu T, Gong HY, Li CY, Ye XH, Lin HJ et al. Application of high-resolution ultrasound, real-time elastography, and contrast-enhanced ultrasound in differentiating solid thyroid nodules. Medicine. 2016; 95(45):e5329

360. Zhao CK, Chen SG, Alizad A, He YP, Wang Q, Wang D et al. Three-Dimensional Shear Wave Elastography for Differentiating Benign From Malignant Thyroid Nodules. Journal of Ultrasound in Medicine. 2018; 37(7):1777-1788

361. Zhao H, Liu X, Lei B, Cheng P, Li J, Wu Y et al. Diagnostic performance of thyroid imaging reporting and data system (TI-RADS) alone and in combination with contrast-enhanced ultrasonography for the characterization of thyroid nodules. Clinical Hemorheology and Microcirculation. 2018; Epublication

362. Zheng B, Tublin ME, Klym AH, Gur D. Classification of thyroid nodules using a resonance-frequency-based electrical impedance spectroscopy: a preliminary assessment. Thyroid. 2013; 23(7):854-62

363. Zheng Y, Xu S, Kang H, Zhan W. A single-center retrospective validation study of the american college of radiology thyroid imaging reporting and data system. Ultrasound Quarterly. 2018; 34(2):77-83

364. Zhou H, Yue WW, Du LY, Xu JM, Liu BJ, Li XL et al. A modified thyroid imaging reporting and data system (mti-rads) for thyroid nodules in coexisting Hashimoto's thyroiditis. Scientific Reports. 2016; 6:26410

365. Zhou H, Zhou XL, Xu HX, Li DD, Liu BJ, Zhang YF et al. Virtual Touch Tissue Imaging and Quantification in the Evaluation of Thyroid Nodules. Journal of Ultrasound in Medicine. 2017; 36(2):251-260

366. Zhou JQ, Zhou C, Zhan WW, Zhou W, Dong YJ. Maximal, minimal, and mean pulsed Doppler parameters: which should be utilized in the diagnosis of thyroid nodules? Clinical Radiology. 2014; 69(12):e477-84

367. Zhou X, Zhou P, Hu Z, Tian SM, Zhao Y, Liu W et al. Diagnostic Efficiency of Quantitative Contrast-Enhanced Ultrasound Indicators for Discriminating Benign From Malignant Solid Thyroid Nodules. Journal of Ultrasound in Medicine. 2018; 37(2):425-437

368. Zhu LC, Ye YL, Luo WH, Su M, Wei HP, Zhang XB et al. A model to discriminate malignant from benign thyroid nodules using artificial neural network. PloS One. 2013; 8(12):e82211

Appendices

Appendix A: Review protocols

Table 6:

ID	Field	Content
I	Review question	Which imaging tests should be requested (for thyroid enlargement)? Which people with structural thyroid abnormalities should have a fine-needle aspiration biopsy?
II	Type of review question	Diagnostic accuracy A review of health economic evidence related to the same review question was conducted in parallel with this review. For details see the health economic review protocol for this NICE guideline.
III	Objective of the review	Determine which imaging tests are most accurate and therefore appropriate for people with thyroid enlargement
IV	Eligibility criteria – population / disease / condition / issue / domain	• People presenting with euthyroid thyroid enlargement being investigated for possible malignancy
V	Eligibility criteria – intervention(s) / exposure(s) / prognostic factor(s)	• Ultrasound scan • CT scan • MRI scan
VI	Eligibility criteria – comparator(s) / control or reference (gold) standard	• Reference standard will be malignant status as confirmed by biopsy/subsequent development of cancer in case of false negatives that do not receive biopsy
VII	Outcomes and prioritisation	• Sensitivity, specificity, PPV, NPV of tests for diagnosing thyroid cancer Sensitivity prioritised
VIII	Eligibility criteria – study design	• Diagnostic accuracy studies • Prospective studies prioritised, retrospective studies included if insufficient prospective studies identified
IX	Other inclusion exclusion criteria	• Excluding two gate study design • Excluding studies that only assess results of those who go on to have surgery as not a representative population • Studies assessing ultrasound only included if full criteria used (as opposed to accuracy of single feature) • Studies assessing variants of ultrasound (for example elastography) only included if combined with conventional criteria
X	Proposed sensitivity / subgroup analysis, or	**Stratifications** • Criteria used (for example Kim, TIRADS, AACE, ATA, BTA for US) • CT with contrast vs CT without contrast

	meta-regression	
XI	Selection process – duplicate screening / selection / analysis	• A sample of at least 10% of the abstract lists were double-sifted by a senior research fellow and discrepancies rectified, with committee input where consensus could not be reached, for more information please see the separate Methods report for this guideline.
XII	Data management (software)	• EndNote was used for reference management, sifting, citations and bibliographies. • Pair forest plots were constructed using Cochrane Review Manager (RevMan5). • WinBUGS was used for diagnostic meta-analysis
XIII	Information sources – databases and dates	• Medline, Embase and the Cochrane library
XIV	Identify if an update	Not an update
XV	Author contacts	https://www.nice.org.uk/guidance/indevelopment/gid-ng10074
XVI	Highlight if amendment to previous protocol	Not an amendment
XVII	Search strategy – for one database	For details please see Appendix B:
XVIII	Data collection process – forms / duplicate	A standardised evidence table format will be used and published as Appendix D: of the evidence report.
XIX	Data items – define all variables to be collected	For details please see evidence tables in Appendix D: (clinical evidence tables) or Appendix G: (health economic evidence tables).
XX	Methods for assessing bias at outcome / study level	QUADAS-2 checklists were used to critically appraise individual studies. The risk of bias across all available evidence was evaluated for each index test using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group http://www.gradeworkinggroup.org/
XXI	Criteria for quantitative synthesis	For details please see section 6.4 of Developing NICE guidelines: the manual.
XXII	Methods for quantitative analysis – combining studies and exploring (in)consistency	For details please see the separate Methods report for this guideline.
XXIII	Meta-bias assessment – publication bias, selective	For details please see section 6.2 of Developing NICE guidelines: the manual.

	reporting bias	
XXIV	Confidence in cumulative evidence	For details please see sections 6.4 and 9.1 of Developing NICE guidelines: the manual.
XXV	Rationale / context – what is known	For details please see the introduction to the evidence review.
XXVI	Describe contributions of authors and guarantor	A multidisciplinary committee developed the evidence review. The committee was convened by the National Guideline Centre (NGC) and chaired by Sarah Fishburn in line with section 3 of Developing NICE guidelines: the manual. Staff from NGC undertook systematic literature searches, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate, and drafted the evidence review in collaboration with the committee. For details please see Developing NICE guidelines: the manual.
XXVII	Sources of funding / support	NGC is funded by NICE and hosted by the Royal College of Physicians.
XXVIII	Name of sponsor	NGC is funded by NICE and hosted by the Royal College of Physicians.
XXIX	Roles of sponsor	NICE funds NGC to develop guidelines for those working in the NHS, public health and social care in England.
XXX	PROSPERO registration number	Not registered

Table 7: Health economic review protocol

Review question	All questions – health economic evidence
Objectives	To identify health economic studies relevant to any of the review questions.
Search criteria	• Populations, interventions and comparators must be as specified in the clinical review protocol above. • Studies must be of a relevant health economic study design (cost–utility analysis, cost-effectiveness analysis, cost–benefit analysis, cost–consequences analysis, comparative cost analysis). • Studies must not be a letter, editorial or commentary, or a review of health economic evaluations. (Recent reviews will be ordered although not reviewed. The bibliographies will be checked for relevant studies, which will then be ordered.) • Unpublished reports will not be considered unless submitted as part of a call for evidence. • Studies must be in English.
Search strategy	A health economic study search will be undertaken using population-specific terms and a health economic study filter – see Appendix B: below.
Review strategy	Studies not meeting any of the search criteria above will be excluded. Studies published before 2003, abstract-only studies and studies from non-OECD countries or the USA will also be excluded. Each remaining study will be assessed for applicability and methodological limitations using the NICE economic evaluation checklist which can be found in appendix H of Developing NICE guidelines: the manual (2014).[193] **Inclusion and exclusion criteria** • If a study is rated as both 'Directly applicable' and with 'Minor limitations' then it will be included in the guideline. A health economic evidence table will be completed, and it will be included in the health economic evidence profile. • If a study is rated as either 'Not applicable' or with 'Very serious limitations' then it will usually be excluded from the guideline. If it is excluded, then a health economic evidence table will not be completed, and it will not be included in the health economic evidence profile. • If a study is rated as 'Partially applicable', with 'Potentially serious limitations' or both then there is discretion over whether it should be included. **Where there is discretion** The health economist will make a decision based on the relative applicability and quality of the available evidence for that question, in discussion with the guideline committee if required. The ultimate aim is to include health economic studies that are helpful for decision-making in the context of the guideline and the current NHS setting. If several studies are considered of sufficiently high applicability and methodological quality that they could all be included, then the health economist, in discussion with the committee if required, may decide to include only the most applicable studies and to selectively exclude the remaining studies. All studies excluded on the basis of applicability or methodological limitations will be listed with explanation in the excluded health economic studies appendix below. The health economist will be guided by the following hierarchies. *Setting:* • UK NHS (most applicable). • OECD countries with predominantly public health insurance systems (for example, France, Germany, Sweden). • OECD countries with predominantly private health insurance systems (for example, Switzerland).

	• Studies set in non-OECD countries or in the USA will be excluded before being assessed for applicability and methodological limitations. *Health economic study type:* • Cost–utility analysis (most applicable). • Other type of full economic evaluation (cost–benefit analysis, cost-effectiveness analysis, cost–consequences analysis). • Comparative cost analysis. • Non-comparative cost analyses including cost-of-illness studies will be excluded before being assessed for applicability and methodological limitations. *Year of analysis:* • The more recent the study, the more applicable it will be. • Studies published in 2003 or later but that depend on unit costs and resource data entirely or predominantly from before 2003 will be rated as 'Not applicable'. • Studies published before 2003 will be excluded before being assessed for applicability and methodological limitations. *Quality and relevance of effectiveness data used in the health economic analysis:* • The more closely the clinical effectiveness data used in the health economic analysis match with the outcomes of the studies included in the clinical review the more useful the analysis will be for decision-making in the guideline.

Appendix B: Literature search strategies

The literature searches for this review are detailed below and complied with the methodology outlined in Developing NICE guidelines: the manual 2014, updated 2018 https://www.nice.org.uk/guidance/pmg20/resources/developing-nice-guidelines-the-manual-pdf-72286708700869

For more detailed information, please see the Methodology Review.]

B.1 Clinical search literature search strategy

Searches were constructed using a PICO framework where population (P) terms were combined with Intervention (I) and in some cases Comparison (C) terms. Outcomes (O) are rarely used in search strategies for interventions as these concepts may not be well described in title, abstract or indexes and therefore difficult to retrieve. Search filters were applied to the search where appropriate.

Table 8: Database date parameters and filters used

Database	Dates searched	Search filter used
Medline (OVID)	1946 – 07 January 2019	Exclusions Randomised controlled trials Systematic review studies Observational studies Diagnostic tests studies
Embase (OVID)	1974 – 07 January 2019	Exclusions Randomised controlled trials Systematic review studies Observational studies Diagnostic tests studies
The Cochrane Library (Wiley)	Cochrane Reviews to 2019 Issue 1 or 12 CENTRAL to 2019 Issue 1 or 12 DARE, and NHSEED to 2015 Issue 2 of 4 HTA to 2016 Issue 2 of 4	None

Medline (Ovid) search terms

1.	exp thyroid diseases/
2.	hyperthyroid*.ti,ab.
3.	hypothyroid*.ti,ab.
4.	thyrotoxicosis.ti,ab.
5.	(thyroid adj3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*)).ti,ab.
6.	or/1-5
7.	letter/
8.	editorial/
9.	news/
10.	exp historical article/
11.	Anecdotes as Topic/
12.	comment/

13.	case report/
14.	(letter or comment*).ti.
15.	or/7-14
16.	randomized controlled trial/ or random*.ti,ab.
17.	15 not 16
18.	animals/ not humans/
19.	exp Animals, Laboratory/
20.	exp Animal Experimentation/
21.	exp Models, Animal/
22.	exp Rodentia/
23.	(rat or rats or mouse or mice).ti.
24.	or/17-23
25.	6 not 24
26.	Ultrasonography/
27.	Magnetic Resonance Imaging/
28.	Tomography, X-Ray Computed/
29.	(ultrasonograph* or ultrasound* or ultra sound or sonograph* or sonogram* or echograph* or echotomograph* or doppler).ti,ab.
30.	magnetic resonance.ti,ab.
31.	(MR or MRI).ti,ab.
32.	(diffusion weighted imag* or DWI).ti,ab.
33.	(computed adj3 tomography).ti,ab.
34.	(CT or CAT).ti,ab.
35.	or/26-34
36.	25 and 35
37.	limit 36 to English language
38.	randomized controlled trial.pt.
39.	controlled clinical trial.pt.
40.	randomi#ed.ti,ab.
41.	placebo.ab.
42.	randomly.ti,ab.
43.	Clinical Trials as topic.sh.
44.	trial.ti.
45.	or/38-44
46.	exp "sensitivity and specificity"/
47.	(sensitivity or specificity).ti,ab.
48.	((pre test or pretest or post test) adj probability).ti,ab.
49.	(predictive value* or PPV or NPV).ti,ab.
50.	likelihood ratio*.ti,ab.
51.	likelihood function/
52.	((area under adj4 curve) or AUC).ti,ab.
53.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
54.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness)).ti,ab.
55.	gold standard.ab.
56.	or/46-55

57.	Epidemiologic studies/
58.	Observational study/
59.	exp Cohort studies/
60.	(cohort adj (study or studies or analys* or data)).ti,ab.
61.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
62.	((longitudinal or retrospective or prospective or cross sectional) and (study or studies or review or analys* or cohort* or data)).ti,ab.
63.	Controlled Before-After Studies/
64.	Historically Controlled Study/
65.	Interrupted Time Series Analysis/
66.	(before adj2 after adj2 (study or studies or data)).ti,ab.
67.	or/57-66
68.	exp case control study/
69.	case control*.ti,ab.
70.	or/68-69
71.	67 or 70
72.	Cross-sectional studies/
73.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
74.	or/72-73
75.	67 or 74
76.	67 or 70 or 74
77.	Meta-Analysis/
78.	exp Meta-Analysis as Topic/
79.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
80.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
81.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
82.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
83.	(search* adj4 literature).ab.
84.	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
85.	cochrane.jw.
86.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
87.	or/77-86
88.	37 and (45 or 56 or 87 or 76)

Embase (Ovid) search terms

1.	exp thyroid disease/
2.	hyperthyroid*.ti,ab.
3.	hypothyroid*.ti,ab.
4.	thyrotoxicosis.ti,ab.
5.	(thyroid adj3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*)).ti,ab.
6.	or/1-5
7.	letter.pt. or letter/

8.	note.pt.
9.	editorial.pt.
10.	case report/ or case study/
11.	(letter or comment*).ti.
12.	or/7-11
13.	randomized controlled trial/ or random*.ti,ab.
14.	12 not 13
15.	animal/ not human/
16.	nonhuman/
17.	exp Animal Experiment/
18.	exp Experimental Animal/
19.	animal model/
20.	exp Rodent/
21.	(rat or rats or mouse or mice).ti.
22.	or/14-21
23.	6 not 22
24.	limit 23 to English language
25.	echography/
26.	nuclear magnetic resonance imaging/
27.	computer assisted tomography/
28.	(ultrasonograph* or ultrasound* or ultra sound or sonograph* or sonogram* or echograph* or echotomograph* or doppler).ti,ab.
29.	magnetic resonance.ti,ab.
30.	(MR or MRI).ti,ab.
31.	(diffusion weighted imag* or DWI).ti,ab.
32.	(computed adj3 tomography).ti,ab.
33.	(CT or CAT).ti,ab.
34.	or/25-33
35.	24 and 34
36.	exp "sensitivity and specificity"/
37.	(sensitivity or specificity).ti,ab.
38.	((pre test or pretest or post test) adj probability).ti,ab.
39.	(predictive value* or PPV or NPV).ti,ab.
40.	likelihood ratio*.ti,ab.
41.	((area under adj4 curve) or AUC).ti,ab.
42.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
43.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness)).ti,ab.
44.	diagnostic accuracy/
45.	diagnostic test accuracy study/
46.	gold standard.ab.
47.	or/36-46
48.	random*.ti,ab.
49.	factorial*.ti,ab.
50.	(crossover* or cross over*).ti,ab.
51.	((doubl* or singl*) adj blind*).ti,ab.

52.	(assign* or allocat* or volunteer* or placebo*).ti,ab.
53.	crossover procedure/
54.	single blind procedure/
55.	randomized controlled trial/
56.	double blind procedure/
57.	or/48-56
58.	systematic review/
59.	meta-analysis/
60.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
61.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
62.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
63.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
64.	(search* adj4 literature).ab.
65.	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
66.	cochrane.jw.
67.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
68.	or/58-67
69.	Clinical study/
70.	Observational study/
71.	family study/
72.	longitudinal study/
73.	retrospective study/
74.	prospective study/
75.	cohort analysis/
76.	follow-up/
77.	cohort*.ti,ab.
78.	76 and 77
79.	(cohort adj (study or studies or analys* or data)).ti,ab.
80.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
81.	((longitudinal or retrospective or prospective or cross sectional) and (study or studies or review or analys* or cohort* or data)).ti,ab.
82.	(before adj2 after adj2 (study or studies or data)).ti,ab.
83.	or/69-75,78-82
84.	exp case control study/
85.	case control*.ti,ab.
86.	or/84-85
87.	83 or 86
88.	cross-sectional study/
89.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
90.	or/88-89
91.	83 or 90
92.	83 or 86 or 90
93.	35 and (47 or 57 or 68 or 92)

Cochrane Library (Wiley) search terms

#1.	MeSH descriptor: [Thyroid Diseases] explode all trees
#2.	hyperthyroid*:ti,ab
#3.	hypothyroid*:ti,ab
#4.	thyrotoxicosis:ti,ab
#5.	(thyroid near/3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*)):ti,ab
#6.	(or #1-#5)
#7.	MeSH descriptor: [Ultrasonography] explode all trees
#8.	MeSH descriptor: [Magnetic Resonance Imaging] explode all trees
#9.	MeSH descriptor: [Tomography, X-Ray Computed] explode all trees
#10.	(ultrasonograph* or ultrasound* or ultra sound or sonograph* or sonogram* or echograph* or echotomograph* or doppler):ti,ab
#11.	magnetic resonance:ti,ab
#12.	(MR or MRI):ti,ab
#13.	(diffusion weighted imag* or DWI):ti,ab
#14.	(computed near/3 tomography):ti,ab
#15.	(CT or CAT):ti,ab
#16.	(or #7-#15)
#17.	#6 and #16

B.2 Health Economics literature search strategy

Health economic evidence was identified by conducting a broad search relating to a thyroid disease population in NHS Economic Evaluation Database (NHS EED – this ceased to be updated after March 2015) and the Health Technology Assessment database (HTA) with no date restrictions. NHS EED and HTA databases are hosted by the Centre for Research and Dissemination (CRD). Additional searches were run on Medline and Embase for health economics, economic modelling and quality of life studies.

Table 9: Database date parameters and filters used

Database	Dates searched	Search filter used
Medline	2014 – 07 January 2019	Exclusions Health economics studies Health economics modelling studies Quality of life studies
Embase	2014 – 07 January 2019	Exclusions Health economics studies Health economics modelling studies Quality of life studies
Centre for Research and Dissemination (CRD)	HTA - Inception – 07 January 2019 NHSEED - Inception to March 2015	None

Medline (Ovid) search terms

1.	exp thyroid diseases/
2.	hyperthyroid*.ti,ab.

3.	hypothyroid*.ti,ab.
4.	thyrotoxicosis.ti,ab.
5.	(thyroid adj3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*)).ti,ab.
6.	or/1-5
7.	letter/
8.	editorial/
9.	news/
10.	exp historical article/
11.	Anecdotes as Topic/
12.	comment/
13.	case report/
14.	(letter or comment*).ti.
15.	or/7-14
16.	randomized controlled trial/ or random*.ti,ab.
17.	15 not 16
18.	animals/ not humans/
19.	exp Animals, Laboratory/
20.	exp Animal Experimentation/
21.	exp Models, Animal/
22.	exp Rodentia/
23.	(rat or rats or mouse or mice).ti.
24.	or/17-23
25.	6 not 24
26.	limit 25 to English language
27.	Economics/
28.	Value of life/
29.	exp "Costs and Cost Analysis"/
30.	exp Economics, Hospital/
31.	exp Economics, Medical/
32.	Economics, Nursing/
33.	Economics, Pharmaceutical/
34.	exp "Fees and Charges"/
35.	exp Budgets/
36.	budget*.ti,ab.
37.	cost*.ti.
38.	(economic* or pharmaco?economic*).ti.
39.	(price* or pricing*).ti,ab.
40.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
41.	(financ* or fee or fees).ti,ab.
42.	(value adj2 (money or monetary)).ti,ab.
43.	or/27-42
44.	exp models, economic/
45.	*Models, Theoretical/

46.	*Models, Organizational/
47.	markov chains/
48.	monte carlo method/
49.	exp Decision Theory/
50.	(markov* or monte carlo).ti,ab.
51.	econom* model*.ti,ab.
52.	(decision* adj2 (tree* or analy* or model*)).ti,ab.
53.	or/44-52
54.	quality-adjusted life years/
55.	sickness impact profile/
56.	(quality adj2 (wellbeing or well being)).ti,ab.
57.	sickness impact profile.ti,ab.
58.	disability adjusted life.ti,ab.
59.	(qal* or qtime* or qwb* or daly*).ti,ab.
60.	(euroqol* or eq5d* or eq 5*).ti,ab.
61.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
62.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
63.	(hui or hui1 or hui2 or hui3).ti,ab.
64.	(health* year* equivalent* or hye or hyes).ti,ab.
65.	discrete choice*.ti,ab.
66.	rosser.ti,ab.
67.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.
68.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.
69.	(sf20 or sf 20 or short form 20 or shortform 20 or shortform20).ti,ab.
70.	(sf12* or sf 12* or short form 12* or shortform 12* or shortform12*).ti,ab.
71.	(sf8* or sf 8* or short form 8* or shortform 8* or shortform8*).ti,ab.
72.	(sf6* or sf 6* or short form 6* or shortform 6* or shortform6*).ti,ab.
73.	or/54-72
74.	26 and (43 or 53 or 73)

Embase (Ovid) search terms

1.	exp thyroid diseases/
2.	hyperthyroid*.ti,ab.
3.	hypothyroid*.ti,ab.
4.	thyrotoxicosis*.ti,ab.
5.	(thyroid adj3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*)).ti,ab.
6.	or/1-5
7.	letter.pt. or letter/
8.	note.pt.
9.	editorial.pt.
10.	case report/ or case study/
11.	(letter or comment*).ti.
12.	or/7-11

13.	randomized controlled trial/ or random*.ti,ab.
14.	12 not 13
15.	animal/ not human/
16.	nonhuman/
17.	exp Animal Experiment/
18.	exp Experimental Animal/
19.	animal model/
20.	exp Rodent/
21.	(rat or rats or mouse or mice).ti.
22.	or/14-21
23.	6 not 22
24.	limit 23 to English language
25.	health economics/
26.	exp economic evaluation/
27.	exp health care cost/
28.	exp fee/
29.	budget/
30.	funding/
31.	budget*.ti,ab.
32.	cost*.ti.
33.	(economic* or pharmaco?economic*).ti.
34.	(price* or pricing*).ti,ab.
35.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
36.	(financ* or fee or fees).ti,ab.
37.	(value adj2 (money or monetary)).ti,ab.
38.	or/25-37
39.	statistical model/
40.	exp economic aspect/
41.	39 and 40
42.	*theoretical model/
43.	*nonbiological model/
44.	stochastic model/
45.	decision theory/
46.	decision tree/
47.	monte carlo method/
48.	(markov* or monte carlo).ti,ab.
49.	econom* model*.ti,ab.
50.	(decision* adj2 (tree* or analy* or model*)).ti,ab.
51.	or/41-50
52.	quality adjusted life year/

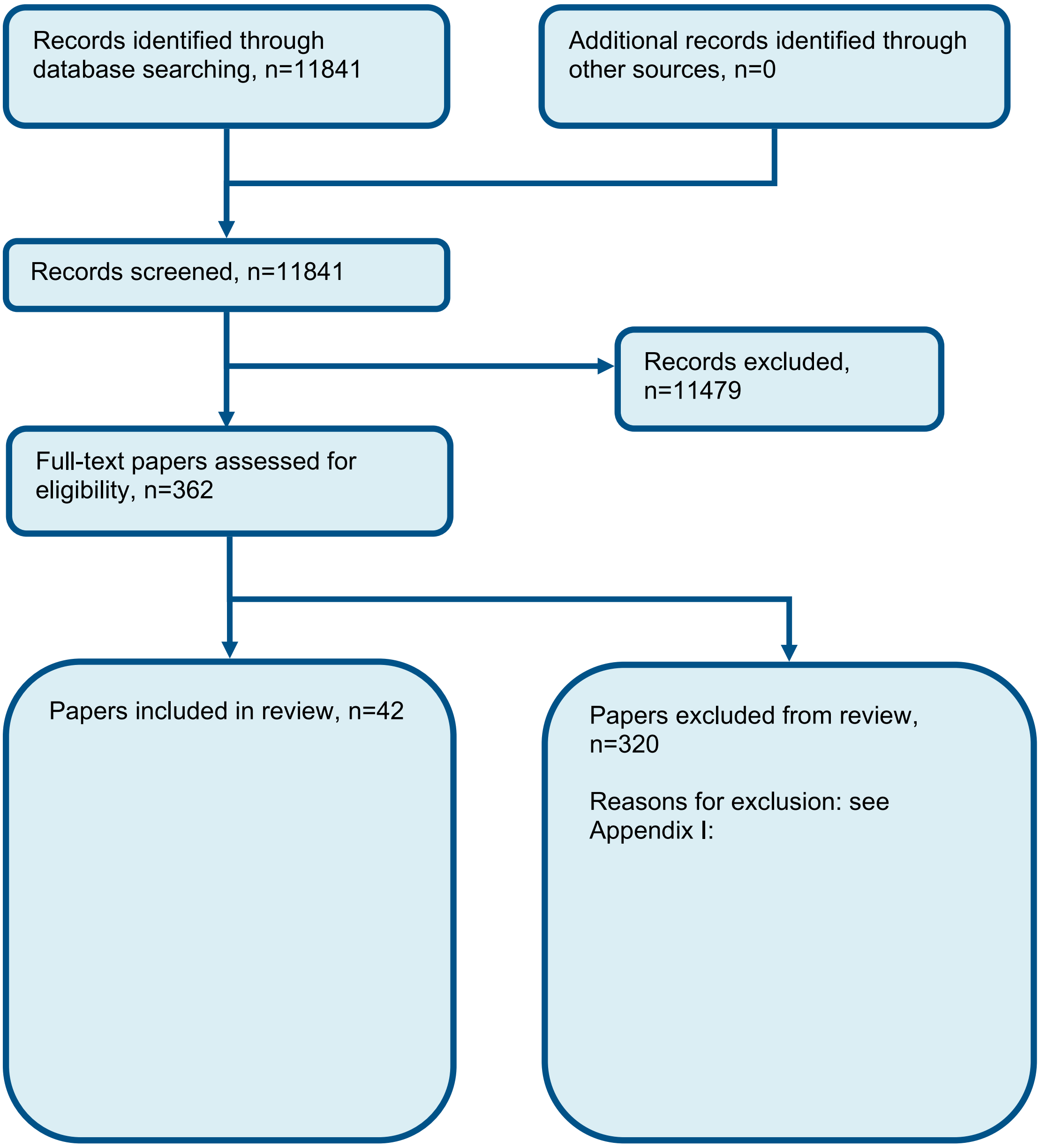
53.	"quality of life index"/
54.	short form 12/ or short form 20/ or short form 36/ or short form 8/
55.	sickness impact profile/
56.	(quality adj2 (wellbeing or well being)).ti,ab.
57.	sickness impact profile.ti,ab.
58.	disability adjusted life.ti,ab.
59.	(qal* or qtime* or qwb* or daly*).ti,ab.
60.	(euroqol* or eq5d* or eq 5*).ti,ab.
61.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
62.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
63.	(hui or hui1 or hui2 or hui3).ti,ab.
64.	(health* year* equivalent* or hye or hyes).ti,ab.
65.	discrete choice*.ti,ab.
66.	rosser.ti,ab.
67.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.
68.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.
69.	(sf20 or sf 20 or short form 20 or shortform 20 or shortform20).ti,ab.
70.	(sf12* or sf 12* or short form 12* or shortform 12* or shortform12*).ti,ab.
71.	(sf8* or sf 8* or short form 8* or shortform 8* or shortform8*).ti,ab.
72.	(sf6* or sf 6* or short form 6* or shortform 6* or shortform6*).ti,ab.
73.	or/52-72
74.	24 and (38 or 51 or 73)

NHS EED and HTA (CRD) search terms

#1.	MeSH DESCRIPTOR Thyroid Diseases EXPLODE ALL TREES
#2.	hyperthyroid*
#3.	hypothyroid*
#4.	thyrotoxicosis*
#5.	(thyroid adj3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*))
#6.	#1 OR #2 OR #3 OR #4 or #5

Appendix C: Clinical evidence selection

Figure 1: Flow chart of clinical study selection for the review of imaging and who to FNAB



Appendix D: Clinical evidence tables

Reference	Ahn 2010 [6]
Study type	Retrospective review
Study methodology	Data source: patients biopsied under ultrasound guidance from September 2002 through July 2004 at the Institute of Radiological Science at Yosnei University Recruitment: unclear
Number of patients	n = 1318 (1398 nodules)
Patient characteristics	Age, mean (range): 46.3 (9-82) Gender (male to female ratio): 101:1217 Ethnicity: not specified Setting: Department of radiology and Research Institute of Radiological Science, Yosnei University, College of Medicine. Country: South Korea Inclusion criteria: Ultrasound was performed on the largest of nodules with similar ultrasound features but on each nodule when multiple nodules had several different ultrasound features. Nodules with benign (n=1016) or malignant (n=244) cytologic findings were included. Exclusion criteria: 128 of 161 nodules with nondiagnostic cytology, 25 of 52 nodules with cytologic findings of follicular neoplasm and 32 of 110 nodules suspicious for papillary carcinoma were excluded.
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound</u> Ultrasound was performed with a 7-to 12- MHz transducer prospectively by one experienced radiologist who described the sonographic characteristics of thyroid nodules with respect to size, multiplicity, composition, echogenicity, margin, calcification, shape, and abnormal cervical lymph nodes. All images were sent to the local PACS for review. Size was measured at the maximum dimension. Substantial growth was retrospectively assessed in 287 nodules examined with ultrasound at least 6 months before FNAB.

Reference	Ahn 2010 [6]				
	Reference standard: Surgery or cytology if no surgery Surgery was performed for 455 nodules (of 1583), including 33 with nondiagnostic cytologic findings, 111 benign nodules, 27 follicular neoplasms, 78 nodules suspicious for papillary carcinoma and 206 malignant nodules. Ultrasound-guided FNAB was performed by one experienced radiologist using a 23-gauge needle attached to a 20-mL disposable plastic syringe and aspirator. Each lesion was aspirated at least twice. The cytopathologist was not on site during the biopsy. Time between measurement of index test and reference standard: unclear, FNAB was performed after surgery.				
2×2 table	**Kim**	Reference standard +	Reference standard −	Total	Notes: Final diagnosis was based on surgical pathologic findings or on cytologic findings if the patient did not undergo surgery
	Index test +	303	205	508	
	Index test −	24	866	890	
	Total	327	1071	1398	
2×2 table	**SRU**	Reference standard +	Reference standard −	Total	Notes: Final diagnosis was based on surgical pathologic findings or on cytologic findings if the patient did not undergo surgery
	Index test +	116	489	605	
	Index test −	211	582	793	
	Total	327	1071	1398	
2×2 table	**AACE**	Reference standard +	Reference standard −	Total	Notes: Final diagnosis was based on surgical pathologic findings or on cytologic findings if the patient did not undergo surgery
	Index test +	259	98	357	
	Index test −	68	973	1041	
	Total	327	1071	1398	
Statistical measures	Index text: Ultrasound (Kim criteria) Sensitivity: 92.7% Specificity: 80.9% PPV: 59.6% NPV: 97.3% AUC: 0.868 Index text Ultrasound (Society of radiologists in ultrasound criteria) Sensitivity: 35.5% Specificity: 54.3% PPV: 19.2% NPV: 80.8%				

Reference	Ahn 2010 [6]
	AUC: 0.551 <u>Index text Ultrasound (AACE)</u> Sensitivity: 79.2% Specificity: 90.8% PPV: 72.3% NPV: 93.5% AUC: 0.850
Source of funding	Not stated
Limitations	Risk of bias: serious; high risk of bias in patient selection; flow and timing Indirectness: none
Comments	

Reference	Alam 2014 [9]
Study type	Cross-sectional prospective
Study methodology	Data source: patients referred to radiology department for thyroid ultrasound followed by FNAB from December 2010 to December 2012 Recruitment: non-probability consecutive sampling
Number of patients	n = 100
Patient characteristics	Age, mean (SD): 41.77 (12.31) Gender (male to female ratio): 24:76 Ethnicity: not specified Setting: Department of Radiology, Aga Khan University Hospital, Karachi (AKUH) Country: Pakistan Inclusion criteria: patients with palpable thyroid nodules diagnosed by primary physician in clinical examination, referred to radiology department of AKUH for thyroid ultrasound followed by fine-needle aspiration cytology of thyroid nodules

<table>
<tr><th>Reference</th><th>Alam 2014 [9]</th></tr>
<tr><td></td><td>Exclusion criteria: proven thyroid malignancy, US or FNAC conducted outside the study institution.</td></tr>
<tr><td>Target condition(s)</td><td>Thyroid cancer</td></tr>
<tr><td>Index test(s) and reference standard</td><td><u>Index test: Ultrasound</u>
All ultrasounds were performed by a single radiologist on Nemio XG ultrasound machine equipped with 3.5-5 MHz Curvilinear and 7.5-15 MHz Linear probe. Transverse and longitudinal images were taken and send to the Picture and Archiving System (PACS) for later review

A nodule was considered positive or malignant if one or more than one of the following sonographic features were found: micro calcification defined as punctuate (less than 2mm) hyper echoic foci either with or without acoustic shadows; micro-lobulation was characterized as presence of many small lobules on surface of a nodule or irregular margins; marked hypo echogenicity demarcated as decreased echogenicity compared with surrounding neck muscle; shape characterised as taller than wider.

A nodule was categorised as negative (malignancy not found) if none of the above feature was seen.

<u>Reference standard: Fine-needle aspiration cytology</u>
FNAC followed all ultrasounds; conducted by a single consultant radiologist with more than 5 years of experience in performing the procedure. FNAC specimen was analysed by cryopathologist with 5 years of experience who was blinded to US diagnosis. FNAC diagnosis of malignancy was acquired from medical record system.

Time between measurement of index test and reference standard:</td></tr>
<tr><td>2×2 table</td><td><table>
<tr><th></th><th>Reference standard +</th><th>Reference standard −</th><th>Total</th></tr>
<tr><td>Index test +</td><td>22</td><td>16</td><td>38</td></tr>
<tr><td>Index test −</td><td>2</td><td>60</td><td>62</td></tr>
<tr><td>Total</td><td>24</td><td>76</td><td>100</td></tr>
</table></td></tr>
<tr><td>Statistical measures</td><td><u>Index text Ultrasound</u>
Sensitivity : 91.7%
Specificity: 78.94%
PPV: 57.9%
NPV: 96.8%

Overall accuracy: 82%</td></tr>
<tr><td>Source of funding</td><td>Not stated</td></tr>
</table>

Reference	Alam 2014 [9]
Limitations	Risk of bias: none Indirectness: none
Comments	

Reference	Chen 2019 [36]
Study type	Retrospective
Study methodology	Data source: patients with thyroid nodules seen at Guangdong Province Hospital of Chinese Medicine from January 2014 to September 2012 Recruitment: not specified
Number of patients	n = 1092
Patient characteristics	Age, mean (SD): 46.92 (13.59) Gender (male to female ratio): 240:825 Ethnicity: Chinese Setting: Guangdong Province Hospital of Chinese Medicine Country: China Inclusion criteria: a single round or oval nodule with a diameter of 3-93 mm on ultrasound; complete clinical data and thyroid ultrasound imaging data; pathological confirmation of the status of all nodules Exclusion criteria: multiple enlarged neck lymph nodes on ultrasound; findings of inflammation on imaging and distant metastasis identified on auxiliary examination
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound</u> A GE LOGIQ E9 ultrasound system with a linear array probe was used to acquire ultrasound images in the frequency range of 6-15 MHz. Thyroid glands and the surrounding area were scanned while patients were in the supine position with the neck fully exposed. The size, shape, internal structure, echogenicity, features of the border and presence of calcifications were carefully observed and recorded.

<table>
<tr><td>Reference</td><td>Chen 2019 [36]</td></tr>
<tr><td></td><td>ACR-TIRADS classification, based on ultrasound indicators including the internal structure, echogenicity, morphology, boundary features and focal echogenicity of the nodules was applied. Scored for each indicator were determined according to the ACR TI-RADS guidelines, and the sum of scores for each nodule was calculated to determine the TI-RADS level for the respective nodule. Ultrasound images were independently reviewed by two doctors. When doctors' opinions differed, the decision was made by senior doctors.

<u>Reference standard: Fine-needle aspiration cytology (and occasionally Surgery)</u>
<u>Pathology of all thyroid cases included in the study was confirmed by fine-needle aspiration biopsy. Patients were divided into benign and malignant thyroid nodules groups according to cytological results. Surgery was performed in these patients according to the ATA guideline</u>

Time between measurement of index test and reference standard: not specified</td></tr>
<tr><td>2×2 table</td><td><table>
<tr><td>ACR-TIRADS</td><td>Reference standard +</td><td>Reference standard −</td><td>Total</td></tr>
<tr><td>Index test +</td><td>385</td><td>313</td><td>698</td></tr>
<tr><td>Index test −</td><td>10</td><td>384</td><td>394</td></tr>
<tr><td>Total</td><td>395</td><td>697</td><td>1092</td></tr>
</table></td></tr>
<tr><td>Statistical measures</td><td><u>Index text Ultrasound (ACR-TIRADS)</u>
Sensitivity : 96%
Specificity: 53 %</td></tr>
<tr><td>Source of funding</td><td>Department development foundation of Guangdong Province Hospital of Chinese Medicine, Grant/Award number; 2017-01</td></tr>
<tr><td>Limitations</td><td>Risk of bias: serious risk due to potential bias in the interpretation of index test and reference standard results
Indirectness: none</td></tr>
<tr><td>Comments</td><td>Diagnostic accuracy of ACR-TIRADS</td></tr>
</table>

Reference	**Creo 2018 [48]**
Study type	Retrospective
Study methodology	Data source: Paediatric patients (≤21 years old) presenting at tertiary centre with a thyroid nodule between 1996 and 2015 Recruitment: not specified

Reference	**Creo 2018** [48]
Number of patients	n = 112 (145 thyroid nodules)
Patient characteristics	Age, mean (SD): 15.5 (3.2) Gender (male to female ratio): 16:96 Ethnicity: not specified Setting: Division of Paediatric Endocrinology and Metabolism Country: USA Inclusion criteria: patients <21 years of age, initial US performed at Mayo Clinic followed by either: 1) histopathology results after thyroidectomy, 2) FNA biopsy cytology results with a follow-up FNA performed at the institution ≥1 year after initial biopsy, 3)US FNA biopsy cytology results with a stable follow-up US performed at the institution ≥1 year after initial biopsy, 4) stable follow-up US performed at the institution ≥1 year after initial US; 2 largest nodules in patients with more than 1 nodule. Exclusion criteria: patients with a genetic syndrome known to increase thyroid cancer risk, patients with history of radiation exposure.
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound + USD</u> Diagnostic gray-scale US with colour Doppler was obtained using high-frequency linear array transducers. Both cine and still imaging were recorded using longitudinal and transverse views. All images were reviewed on the same imaging system by 2 paediatric radiologists with a combined experience of 27 years after paediatric radiology fellowship training. The radiologists described specific nodule features based upon the TIRADS description for reporting thyroid nodule features. After radiologists recorded the features, an independent reviewer assigned each nodule a level of suspicion for malignancy based on the 2015 ATA Adult risk Classification Guidelines. Radiologists were simply asked to provide their overall impression and were given the descriptive choices of benign, indeterminate, or malignant, which was informed by the presence of absence of calcifications, the type of margins, as well as the size and composition of nodules. <u>Reference standard: Cytology and Histology</u> FNA was performed by institutional radiologists by free-hand technique with US guidance. Cytology results were reported using the Bethesda System for Reporting Thyroid Cytology. This includes (I) nondiagnostic,(II) benign, (III) atypia of undetermined significance, (IV) suspicious for follicular neoplasm, (V) suspicious for malignancy, and (VI) malignant categories. In a child with concerning cytology results who underwent thyroidectomy, appropriate follow-up with repeat FNA or repeat US≥ 1 year was used to ensure the nodule was accurately classified as benign.

Reference	Creo 2018 [48]
	Time between measurement of index test and reference standard: not specified
2×2 table	

2015 ATA	Reference standard +	Reference standard −	Total
Index test +	46	63	109
Index test −	4	32	36
Total	50	95	145

Statistical measures	Index text Ultrasound (2015 ATA) Sensitivity : 92% Specificity: 32%
Source of funding	Not specified
Limitations	Risk of bias: none Indirectness: none
Comments	Diagnostic accuracy of 2015 ATA TIRADS

Reference	Grani 2019 [80]
Study type	Retrospective
Study methodology	Data source: patients referred for FNA cytology of a thyroid nodule at the Thyroid cancer Unit of a large academic referral centre between 1 November 2015 and 30 May 2018 Recruitment: prospective
Number of patients	n = 477 (502 thyroid nodules)
Patient characteristics	Age, mean (SD): 55.9 (13.9) Gender (male to female ratio): 119:358 Ethnicity: not specified Setting: Thyroid Cancer Unit of academic referral centre (Sapienza, University of Rome)

Reference	**Grani 2019 [80]**				
	Country: Italy Inclusion criteria: all patients consecutively referred to the unit for FNA cytology of a thyroid nodule between 1 November 2015 and 30 May 2018 Exclusion criteria: subcentimeter nodules, nodules with an inconclusive reference standard diagnosis were excluded				
Target condition(s)	Thyroid cancer				
Index test(s) and reference standard	Index test: Ultrasound Each nodule was carefully examined with a HI VISION Avius ultrasound system and a 13-MHz linear-array transducer. Two clinicians experienced in thyroid sonography recorded their consensus judgment of the sonographic features of each nodule on a standardized rating form, internally developed and based on published recommendations, based on nodule diameter, margin, structure/composition, echogenicity, calcification, other hypoechoic foci, suspected extra thyroidal extension, as well as location of the solid component for mixed-content nodules. For each nodule, the consensus rating of each ultrasound feature was used to classify the risk of malignancy according to the following risk stratification criteria: AACE/ACE/AME, the ACR-TIRADS, the ATA, the EU-TIRADS, and the K-TIRADS. Reference standard: UGFNAB/ Histology Biopsies were conducted under ultrasound-guidance by clinicians (endocrinologists trained in thyroid sonography using 23- to 25-gauge needles, using the non-aspiration technique in most cases. Direct smears of each specimen were analysed by experienced thyroid cytopathologists and classified according to criteria published in the Italian Consensus for Thyroid Cytopathology. When surgery had been performed, the reference standard diagnosis was based on histological examinations of the respected nodule. When the nodule had been managed non-surgically the reference standard was FNA cytology: nodules were considered malignant when they had been classified as TIR4 or TIR5 (suspected malignancy or malignancy, corresponding to the Bethesda classes V and VI) and benign when they had been classified as TIR 2, corresponding to Bethesda class II. Time between measurement of index test and reference standard: not specified				
2×2 table	ACR TIRADS	Reference standard +	Reference standard −	Total	In 34 malignant cases, the diagnosis was based on histological findings while the remaining 2 were classified cytologically as TIR4/ Bethesda V.
	Index test +	30	204	234	
	Index test −	**6**	262	**268**	
	Total	36	466	502	
2×2 table	AACE/ACE/AME	Reference standard +	Reference standard −	Total	In 34 malignant cases, the diagnosis was based on histological findings while the
	Index test +	31	296	327	

Reference	Grani 2019 [80]				
	Index test −	5	170	175	remaining 2 were classified cytologically as TIR4/ Bethesda V.
	Total	36	466	502	
2×2 table	ATA	Reference standard +	Reference standard −	Total	In 34 malignant cases, the diagnosis was based on histological findings while the remaining 2 were classified cytologically as TIR4/ Bethesda V. Excluding 90 not classifiable nodules.
	Index test +	27	255	282	
	Index test −	9	211	220	
	Total	36	466	502	
2×2 table	EU-TIRADS	Reference standard +	Reference standard −	Total	In 34 malignant cases, the diagnosis was based on histological findings while the remaining 2 were classified cytologically as TIR4/ Bethesda V.
	Index test +	31	317	348	
	Index test −	5	149	154	
	Total	36	466	502	
2×2 table	K-TIRADS	Reference standard +	Reference standard −	Total	In 34 malignant cases, the diagnosis was based on histological findings while the remaining 2 were classified cytologically as TIR4/ Bethesda V.
	Index test +	33	383	416	
	Index test −	3	83	86	
	Total	36	466	502	
Statistical measures	Index text Ultrasound (ACR TIRADS) Sensitivity : 83.3 % Specificity: 56.2% Index text Ultrasound (AACE/ACE/AME) Sensitivity : 86.1 % Specificity: 36.5 % Index text Ultrasound (ATA) Sensitivity : 75 % Specificity: 45.3% Index text Ultrasound (EU-TIRADS) Sensitivity : 86.1 % Specificity: 32% Index text Ultrasound (K-TIRADS)				

Reference	Grani 2019 [80]
	Sensitivity : 91.7 % Specificity: 17.8%
Source of funding	Not specified
Limitations	Risk of bias: serious due to potential risk of bias in the interpretation of the reference standard; flow and timing. Indirectness: none
Comments	Diagnostic accuracy of the ACR TIRADS, AACE/ACE/AME, ATA, EU-TIRADS, K-TIRADS

Reference	Farihah 2018 [67]
Study type	Cross-sectional retrospective
Study methodology	Data source: patients who underwent US-guided FNAC for US-detected focal thyroid nodules from January 2014 to May 2016, with available pathology results Recruitment: not specified
Number of patients	n = 91 (104 nodules)
Patient characteristics	Age, mean (range): 54.7 (27-80) Gender (male to female ratio): 21:83 Ethnicity: 51(49%) Malay, 25 (33.7%) Chinese, 13 (12.5%) Indian, 5 (4.8%) other races. Setting: Radiology Department of Universiti Kebangsaan Malaysia Medical Centre (UKMMC) Country: Malaysia Inclusion criteria: nodules with benign or malignant results at cytology or histology examination; patients who underwent thyroid surgery after specimens from cytology examination were classified as suspicious for thyroid carcinoma, indeterminate, or inadequate. Exclusion criteria: patients who had nodules cytologically diagnosed as suspicious for thyroid carcinoma, indeterminate or inadequate but

<table>
<tr><th>Reference</th><th>Farihah 2018 [67]</th></tr>
<tr><td></td><td>did not undergo surgery; patients with previous history of total or partial thyroidectomy, with or without radioiodine ablation.</td></tr>
<tr><td>Target condition(s)</td><td>Thyroid cancer</td></tr>
<tr><td>Index test(s) and reference standard</td><td><u>Index test: Ultrasound</u>
All available US scans of the thyroid gland and neck areas were performed using a linear-array transducer (5-12 MHz) on ultrasound scanners HD11/ HD11 XE/ iU22 Phillips Medical Systems or Toshiba Xario200 using an optimized gain.
The radiologist, using Osirix workstation or Medweb, reviewed all images. All thyroid nodules were characterised according to the relevant nodule size, composition, cystic component, echogenicity, margins, evidence of calcifications, taller than wide, halo, colour flow and lymphadenopathy.
Nodules were given a U1-U5 score based on the features described by the BTA Guidelines i.e. normal (U1), benign (U2), equivocal/indeterminate (U3), suspicious (U4) and malignant (U5)

U2 and U3 were classified as negative; U4 and U5 as positive

<u>Reference standard: US-guided Fine-needle aspiration cytology and histopathology</u>
US-guided FNAC was performed in either the thyroid nodule with suspicious US features or the largest thyroid nodule if no suspicious US features were detected. US-guided FNAC was performed with a 23-gauge needle attached to a 10 ml disposable plastic syringe. Cytopathology reports were classified as benign, indeterminate, suspicious of malignancy, malignant or inadequate. Histopathology reports were obtained for cases that were cytologically reported as inadequate, indeterminate or suspicious of malignancy.

Time between measurement of index test and reference standard: not specified</td></tr>
<tr><td>2×2 table</td><td><table>
<tr><td></td><td>Reference standard +</td><td>Reference standard −</td><td>Total</td><td rowspan="4">Using BTA recommendations to biopsy U3 upwards</td></tr>
<tr><td>Index test +</td><td>12</td><td>60</td><td>72</td></tr>
<tr><td>Index test −</td><td>0</td><td>32</td><td>32</td></tr>
<tr><td>Total</td><td>12</td><td>92</td><td>104</td></tr>
</table></td></tr>
<tr><td>Statistical measures</td><td><u>Index text Ultrasound</u>
Sensitivity : 100%
Specificity: 35%</td></tr>
<tr><td>Source of funding</td><td>Not stated</td></tr>
</table>

Reference	**Farihah 2018** [67]
Limitations	Risk of bias: very serious due to patient selection; risk of bias in the interpretation of the index test and reference standard Indirectness: none
Comments	Diagnostic accuracy of BTA guidelines

Reference	**Ha 2016** [86]
Study type	Prospective multicentre
Study methodology	Data source: patient data collected from four different hospitals from June 2013 to May 2015 Recruitment: consecutive
Number of patients	n = 750 (902 nodules)
Patient characteristics	Age, mean (range): 49.2 (9-81) Gender (male to female ratio): 156:594 Ethnicity: not specified Setting: four different hospitals Country: South Korea Inclusion criteria: nodules >5mm in patients from four different hospitals who had undergone thyroid US from June 2013 to May 2015 Exclusion criteria: nodules with no final diagnosis obtained (n=198); entirely calcified nodules with US characteristics that could not be analysed (n=9)
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound</u> All US examinations were performed with a 10-16 MHz linear probe and a real-time US system, by five board-certified radiologists, in four different hospitals specialising in thyroid imaging. Nodules were classified according to K-TIRADS. Malignancy risk was stratified into the 5 categories of K-TIRADS according to US patterns by combining solidity, echogenicity, and suspicious US features as follows: 1=normal; 2=benign; 3= low suspicion; 4=intermediate suspicion; 5=high suspicion

<table>
<tr><td>Reference</td><td colspan="5">Ha 2016 [86]</td></tr>
<tr><td></td><td colspan="5"><u>Reference standard: US-guided Fine-needle aspiration or Core needle biopsies (CNBs) or surgery</u>
US-guided FNAs or CNBs were performed by the same radiologists who performed the thyroid US. US-guided FNAs were performed with 23-gauge needles and a combination of capillary and aspiration FNA techniques. CNB was performed using a disposable 18-gauge, single -or double-action spring-activated needle. FNA was usually performed for thyroid nodules > 1 cm, with exception of pure cystic nodules, partially cystic nodules with comet-tail artifacts, and spongiform nodules that usually underwent FNA for therapeutic cyst aspiration, ethanol or radiofrequency ablation therapy, or nodule size of >2cm in case of spongiform nodule. FNA was performed for thyroid nodules <1 cm in case of suspicious US features, or for decisions on surgical planning.
The interpretation of FNA was based on the Bethesda system for reporting thyroid cytopathology and CNB results were diagnosed with a six-tier pathology reporting system

Time between measurement of index test and reference standard: not specified</td></tr>
<tr><td rowspan="4">2×2 table</td><td></td><td>Reference standard +</td><td>Reference standard −</td><td>Total</td><td rowspan="4">FNA or CNB biopsy on 409 nodules (n=75 malignant, n=334 benign) ; repeated FNA or CNB biopsy on 256 nodules (benign); Surgery on 237 nodules (n=191 malignant, n=46 benign)</td></tr>
<tr><td>Index test +</td><td>254</td><td>263</td><td>517</td></tr>
<tr><td>Index test −</td><td>12</td><td>373</td><td>385</td></tr>
<tr><td>Total</td><td>266</td><td>636</td><td>902</td></tr>
<tr><td>Statistical measures</td><td colspan="5"><u>Index text Ultrasound</u>
Sensitivity : 95.5%
Specificity: 58.6%
PPV: 44.5%
NPV: 96.9%

Overall accuracy: 69.5%</td></tr>
<tr><td>Source of funding</td><td colspan="5">Not stated</td></tr>
<tr><td>Limitations</td><td colspan="5">Risk of bias: none
Indirectness: none</td></tr>
<tr><td>Comments</td><td colspan="5">Diagnostic accuracy of K-TIRADS guidelines</td></tr>
</table>

Reference	**Ha 2018 [87]**
Study type	Retrospective multicentre
Study	Data source: patient data collected from four different hospitals from June 2013 to May 2015

Reference	**Ha 2018** [87]
methodology	Recruitment: consecutive
Number of patients	n = 750 (902 nodules)
Patient characteristics	Age, mean (range): 49.2 (9-81) Gender (male to female ratio): 156:594 Ethnicity: not specified Setting: four different hospitals (one primary medical centre and three tertiary hospitals) Country: South Korea Inclusion criteria: nodules >5mm in patients from four different hospitals who had undergone thyroid US from June 2013 to May 2015 Exclusion criteria: nodules with no final diagnosis obtained (n=198); entirely calcified nodules with US characteristics that could not be analysed (n=9)
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound</u> All US examinations were performed with a 10-16 MHz linear probe and a real-time US system, by five board-certified radiologists, in four different hospitals specialising in thyroid imaging (with 8-20 years of clinical experience with thyroid US). Nodules were classified according to Malignancy risk was stratified into different categories for the different criteria used based on US patterns by combining solidity, echogenicity, calcification as follows: high, intermediate, low, very low suspicion, benign or not specified for the ATA 2015 guidelines; highly, moderately, mildly, not suspicious or benign for the ACR 2017 guidelines; high, intermediate, low suspicion or benign for the KTA/KAThR 2016 guidelines <u>Reference standard:</u> Final diagnoses were determined via surgical resection in 191 of 266 malignant nodules, 36 benign nodules were confirmed by surgery, 75 malignant nodules were diagnosed via FNA or core-needle biopsy. Final diagnosis was determined by the cytopathologic results of on the Bethesda system and surgical findings.

Reference	Ha 2018 [87]				
	Time between measurement of index test and reference standard: not specified				
2×2 table	**ATA**	Reference standard +	Reference standard −	Total	Calculated considering ATA 2015 categories of high, intermediate as malignant; low suspicion, very low suspicion, benign as benign and excluding 'not specified' nodules not meeting criteria for any pattern of malignancy
	Index test +	247	202	449	
	Index test −	12	372	384	
	Total	259	574	833	
	ACR	Reference standard +	Reference standard −	Total	Calculated considering ACR 2017 categories of highly, moderately suspicious as malignant; mildly not suspicious and benign as benign.
	Index test +	255	297	552	
	Index test −	11	339	350	
	Total	266	636	902	
	KTA/KSThR	Reference standard +	Reference standard −	Total	Calculated considering KTA/KSThR 2016 categories of high and intermediate suspicion as malignant; low suspicion and benign as benign.
	Index test +	254	263	517	
	Index test −	12	373	385	
	Total	266	636	902	
Statistical measures	Index text Ultrasound (ATA) 1 Sensitivity : 95.4% Specificity: 64.8% Index text Ultrasound (ACR) Sensitivity : 95.8% Specificity: 53.3% Index text Ultrasound (KTA/KSThR) Sensitivity : 95.5% Specificity: 58.6%				
Source of funding	Not stated				
Limitations	Risk of bias: none Indirectness: none				
Comments	Diagnostic accuracy of ATA, KTA/KSThR, ACR guidelines				

Reference	Hoang 2018 [92]
Study type	Retrospective
Study methodology	Data source: patients undergoing FNAB with definitive cytology results or surgical resection from April 2009 to May 2010 Recruitment: consecutive
Number of patients	n = 92 (100 nodules)
Patient characteristics	Age, mean (SD; range): 52 (14; 19-82) Gender (male to female ratio): Ethnicity: not specified Setting: unspecified institution Country: USA Inclusion criteria: patients undergoing FNAB with definitive cytology results or surgical resection from April 2009 to May 2010 at a single institution Exclusion criteria: absence of a dedicated video clip of the biopsied nodule
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound</u> The US examinations were performed by using a variety of commercially available units equipped with 5-15-MHz linear array transducers. In all cases, images of the biopsied nodules were obtained in transverse and longitudinal planes. Video clips of the biopsied nodules were obtained in at least one plane. 11 radiologists from nine different institutions evaluated the nodules on the ACR portal. Readers were blinded to the pathology results. Three expert readers, that were on the ACR TI-RADS committee and had between 26 and 34 years of post-training experience, interpreted the sonograms independently and their consensus was used as the truth for the nodule imaging features. The other eight radiologists were test readers who had no knowledge of ACR TIRADS. All reported thyroid US in their clinical practice. All radiologists assessed the nodules for the five feature categories in the ACR TI-RADS lexicon (composition, echogenicity, shape, margin, and echogenic foci) after reviewing two to four static US images and one or two video images of the same nodule. Test readers also assigned a malignancy risk that matched the five risk stratification levels used in the ACR TI-RADS guidelines (highly suspicious, moderately suspicious, mildly suspicious, not suspicious or benign). Expert and test readers' feature assignments for nodules and maximum nodules size were then used to retrospectively assign an ACR TI-RADS risk stratification level and biopsy

Reference	Hoang 2018 [92]				
	recommendation. ATA and Korean and French TI-RADS guidelines were retrospectively applied. Reference standard: Cytology and Pathology Time between measurement of index test and reference standard: not specified				
2×2 table	**ACR-TIRADS**	Reference standard +	Reference standard −	Total	
	Index test +	14	48	62	
	Index test −	1	37	38	
	Total	15	85	100	
2×2 table	**ATA**	Reference standard +	Reference standard −	Total	
	Index test +	13	70	83	
	Index test −	2	15	17	
	Total	15	85	100	
2×2 table	**F-TIRADS**	Reference standard +	Reference standard −	Total	
	Index test +	13	57	70	
	Index test −	2	28	30	
	Total	15	85	100	
2×2 table	**K-TIRADS**	Reference standard +	Reference standard −	Total	
	Index test +	13	71	84	
	Index test −	2	14	16	
	Total	15	85	100	
Statistical measures	Index text Ultrasound (ACR-TIRADS) Sensitivity: 92% Specificity: 44% Accuracy: 52% Index text Ultrasound (ATA) Sensitivity: 87% Specificity: 18%				

Reference	Hoang 2018 [92]
	Accuracy: 28% <u>Index text Ultrasound (F-TIRADS)</u> Sensitivity: 87% Specificity: 33% Accuracy: 41% <u>Index text Ultrasound (K-TIRADS)</u> Sensitivity: 87% Specificity: 16% Accuracy: 27%
Source of funding	Not specified
Limitations	Risk of bias: serious risk due to reference standard; flow and timing Indirectness: none
Comments	Diagnostic performance of ATA, ACR-TIRADS, K-TIRADS, F-TIRADS

Reference	Hobbs 2014 [94]
Study type	Retrospective
Study methodology	Data source: 400 consecutive records of US-guided FNA encounters through the department of radiology from July 2010 to June 2011 Recruitment: consecutive
Number of patients	n = 350 (360 biopsy encounters)
Patient characteristics	Age, mean (range): 55 (7-91) Gender (male to female ratio): 60:290 Ethnicity: not specified Setting: Department of Radiology, Division of Neuroradiology, Duke University Medical Centre, Durham Country: USA

Reference	Hobbs 2014 [94]
	Inclusion criteria: US-guided FNA encounters through the department of radiology , defined as presentation to the department of radiology on a given date for FNA of one or more thyroid nodules during a 12-month period from July 2010 to June 2011. Exclusion criteria: patients without definitive pathology results
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	Index test: Ultrasound Diagnostic ultrasound images of the thyroid nodules were obtained before the biopsy using a 12 MHz transducer. Thyroid nodules were measured on the ultrasound unit by the technologist or radiologist at the time of imaging and were documented in the examination report. These sizes were used, and nodules were not measured retrospectively. A board-certified radiologist (7 years of experience) reviewed ultrasound images on PACS workstation for findings according to the SRU recommendations which were met if the biopsied nodule had any of the following characteristics: size 10 mm or larger with microcalcifications, size 15 mm or larger with solid composition or coarse calcifications, size 20 mm or larger with mixed solid-cystic composition, or substantial growth since the prior ultrasound. Biopsy encounters were categorised on the basis of sonographic findings as meeting the SRU recommendations for biopsy, referred to as 'SRU-positive' or not ('SRU-negative'). Reference standard: US-guided Fine-needle aspiration cytopathology (n=253 patients) or surgery (n=87 patients) FNA cytopathology was characterised by the Bethesda class categories. FNAs included Bethesda class II or VI cytopathologic results or final surgical pathology (n=360 biopsy encounters). 40 patients were excluded because FNA cytopathologic results revealed Bethesda class I, III, IV or V without repeat FNA or surgery for definitive pathology results. Time between measurement of index test and reference standard: one day for almost all patients
2×2 table	(see table below)
Statistical measures	Index text Ultrasound Sensitivity : 83% Specificity: 25% PPV: 8.76% NPV: 94.2%
Source of	Not stated

2×2 table:

	Reference standard +	Reference standard −	Total
Index test +	24	250	274
Index test −	5	81	86
Total	29	331	360

Reference	Hobbs 2014 [94]
funding	
Limitations	Risk of bias: Serious due to risk of bias in the interpretation of the index test results. Indirectness: none
Comments	Diagnostic accuracy of SRU guidelines

Reference	Horvath 2009 [97]
Study type	Prospective
Study methodology	Data source: 1959 thyroid nodules submitted for FNAB Recruitment: not specified
Number of patients	n = 1097 nodules
Patient characteristics	Age, mean (range): not specified Gender (male to female ratio): not specified Ethnicity: not specified Setting: not specified Country: Chile Inclusion criteria: not specified Exclusion criteria: not specified
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	Index test: Ultrasound Us equipment used was the ATL HDI 5000 and the Philips IU22 with a 5-12 and 5 to 17-MHz probe and colour Doppler. Nodules were classified based on the TI-RADS categories as follows: TIRADS 2: benign findings; TIRADS 3: probably benign; TIRADS 4A: undetermined; TIRADS 4B: suspicious; TIRADS 5: consistent with malignancy; TIRADS 6: malignant Reference standard: Fine-needle aspiration biopsy FNAB was performed by five specialising radiologists, under US guidance using a 19 or 21-gauge needle attached to a 10-cc syringe. Two experience pathologists read all the samples. The histological result of the FNAB was classified as either benign,

<table>
<tr><th>Reference</th><th>Horvath 2009 [97]</th></tr>
<tr><td></td><td>indeterminate/suspicious (follicular lesions) or malignant, according to standard pathological criteria. For the TI-RADS evaluation, two groups were considered: benign and non-benign (including malignant and follicular lesions).

All nodules with malignant FNAB results were submitted to surgery. Benign lesions were followed-up.

Time between measurement of index test and reference standard: not specified</td></tr>
<tr><td>2×2 table</td><td><table><tr><td></td><td>Reference standard +</td><td>Reference standard −</td><td>Total</td></tr><tr><td>Index test +</td><td>349</td><td>360</td><td>709</td></tr><tr><td>Index test −</td><td>46</td><td>342</td><td>389</td></tr><tr><td>Total</td><td>394</td><td>703</td><td>1097</td></tr></table></td></tr>
<tr><td>Statistical measures</td><td><u>Index text Ultrasound</u>
Sensitivity : 88%
Specificity: 49%
PPV: 49%
NPV: 88%
Accuracy: 94%</td></tr>
<tr><td>Source of funding</td><td>Not stated</td></tr>
<tr><td>Limitations</td><td>Risk of bias: Very serious due to patient selection; flow and timing
Indirectness: none</td></tr>
<tr><td>Comments</td><td>Diagnostic accuracy of TI-RADS guidelines, using BI-RADS as a model</td></tr>
</table>

Reference	**Kim 2002 [114]**
Study type	Prospective
Study methodology	Data source: patients undergoing sonography of the thyroid for non-thyroid indications between December 1997 and May 1998 Recruitment: unclear, patients with solid nodules.
Number of patients	n = 132 (155 nodules)

Reference	Kim 2002 [114]
Patient characteristics	Age, mean (range): 48 (20-77) Gender (male to female ratio): 12:120 Ethnicity: Not specified Setting: Department of Diagnostic radiology, Severance Hospital, Yosnei University College of Medicine Country: Korea Inclusion criteria: solid nonpalpable thyroid nodules Exclusion criteria: not specified; excluded cystic nodules, nodules with mixed cystic and solid portions.
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Sonography</u> Performed by one radiologist with an HDI 3000 scanner using electronically focused near-field probes with a bandwidth of 7-12 MHz. Nodules were classified as positive (malignant) if one of the following sonographic features was present: micro calcifications, an irregular or microlobulated margin, marked hypoechogenicity, a shape that is taller than it is wide. If a nodule had no suspicious features was classified as negative (benign) . <u>Reference standard: Fine-needle aspiration biopsy (with or without surgery or surgery alone)</u> All solid nodules were aspirated in patients with two or more solid nodules. Further details of the FNAB were not specified. Time between measurement of index test and reference standard: not specified

2×2 table		Reference standard +	Reference standard −	Total	
	Index test +	46	36	82	Reference standard was: FNAB and follow-up (>24 months) of 83 benign nodules; follow up by FNAB and surgery on 44 malignant and 15 benign lesions; surgery alone on five malignant and eight benign lesions.
	Index test −	3	70	73	
	Total	49	106	155	

Reference	**Kim 2002** [114]
Statistical measures	Index text Sonography: Sensitivity : 93.8% Specificity: 66% PPV: 56.1% NPV: 95.9% Overall accuracy: 74.8%
Source of funding	Not stated
Limitations	Risk of bias: serious risk of bias due to potential bias in patient selection, interpretation of the index test and/or the reference standard Indirectness: none
Comments	

Reference	**Kim 2013** [114]
Study type	Prospective (review of retrospective data)
Study methodology	Data source: patients biopsied under ultrasound guidance from September 2007 to March 2008 Recruitment: unclear, patients meeting inclusion criteria
Number of patients	n = 686 (713 nodules)
Patient characteristics	Age, mean (range): 49.7 Gender (male to female ratio): 87:599 Ethnicity: Not specified Setting: Department of radiology, Research Institute of Radiological Science, Yosnei University College of Medicine Country: South Korea

<table>
<tr><th>Reference</th><th>Kim 2013 [114]</th></tr>
<tr><td></td><td>Inclusion criteria: nodules 6-10 mm biopsied under ultrasound guidance that were operated on for nondiagnostic, indeterminate, malignant or suspicious cytological results and that were operated on or showed no interval change for at least 1 year of follow-up for benign cytology.

Exclusion criteria: nodules with insufficient cytological results for deciding whether benign or malignant</td></tr>
<tr><td>Target condition(s)</td><td>Thyroid cancer</td></tr>
<tr><td>Index test(s) and reference standard</td><td><u>Index test: Ultrasound (US)</u>
US images were obtained using 5-12 MHz linear transducers (HDI 5000 and IU-22, respectively). Real-time ultrasound was performed by seven radiologists (four faculty members with 5-13 years of experience and three fellows).

US features of all thyroid nodules that underwent UGFNAB were prospectively recorded according to internal component, echogenicity, margin, calcification, shape and vascularity at the time of the FNAB.

<u>Reference standard: UG-FNAB</u>
Performed with a 23 gauge needle attached to either a 2mL or 20 mL disposable plastic syringe. Aspiration was done at least twice in each nodule and aspirated material was expelled onto glass slides and smeared.

Time between measurement of index test and reference standard: not specified</td></tr>
<tr><td>2×2 table</td><td><table>
<tr><th>ATA</th><th>Reference standard +</th><th>Reference standard −</th><th>Total</th></tr>
<tr><td>Index test +</td><td>286</td><td>306</td><td>592</td></tr>
<tr><td>Index test −</td><td>10</td><td>111</td><td>121</td></tr>
<tr><td>Total</td><td>296</td><td>417</td><td>713</td></tr>
</table></td></tr>
<tr><td>Statistical measures</td><td><u>Index text: Ultrasound ATA</u>
Sensitivity : 96.6%
Specificity: 26.6%
PPV: 48.3%
NPV: 91.7%
AUC:0.616%</td></tr>
<tr><td>Source of funding</td><td>Not stated</td></tr>
<tr><td>Limitations</td><td>Risk of bias: serious risk of bias due to potential risk in the conduct or interpretation of the index test and/or reference standard</td></tr>
</table>

Reference	**Kim 2013 [114]**
	Indirectness: none
Comments	Diagnostic accuracy of ultrasonographic features of the ATA 2009 guidelines

Reference	**Kim 2013 [114]**
Study type	Retrospective
Study methodology	Data source: patients having undergone US and US-guided FNA between March 2010 and July 2011 Recruitment: unclear
Number of patients	n = 925 (1419 nodules)
Patient characteristics	Age, mean (range): 51.87 (14-85) Gender (male to female ratio): 104:821 Ethnicity: Not specified Setting: Department of Surgery, Wonju Christian Hospital Country: South Korea Inclusion criteria: patients having undergone US and US-guided FNA between March 2010 and July 2011 at the Department of Surgery, Wonju Christian Hospital Exclusion criteria: not specified
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound (US)</u> All neck ultrasounds were performed by a surgeon under the supervision of three experienced endocrine surgeons using high frequency linear array transducers 7.5-13 MHz. Nodules were classified according to new US guidelines which were established via discussions among experienced physicians who participated in the study. Each nodule was classified by standard US characteristics: suspicious for malignancy, intermediate, probably benign. <u>Reference standard: UG-FNAB</u>

Reference	Kim 2013 [114]			
	FNA was performed by the same single surgeon following US evaluation. Benign cytological results were defined by the Bethesda classification system including histopathology consistent with benign follicular nodule, Hashimoto thyroiditis, and subacute thyroiditis. The intermediate category included results consistent with atypical cells of undetermined significance, follicular neoplasms, and suspicion of malignancies. Malignant category was defined as all histopathology positive for malignancy Time between measurement of index test and reference standard: not specified			
2×2 table		Reference standard +	Reference standard −	Total
	Index test +	147	354	501
	Index test −	0	127	127
	Total	147	481	628
2×2 table	**New guidelines**	Reference standard +	Reference standard −	Total
	Index test +	142	121	263
	Index test −	6	676	682
	Total	148	797	945
Statistical measures	<u>Index text: Ultrasound (current guidelines)</u> Sensitivity : 99.3% Specificity: 62.6% PPV: 25% NPV: 99.8% Accuracy: 24.1% <u>Index text: Ultrasound (new guidelines)</u> Sensitivity : 96% Specificity: 86.7% PPV: 47.7% NPV: 99.4% Accuracy: 66%			
Source of funding	Not stated			
Limitations	Risk of bias: very serious due to risk of bias in patient selection, in the conduct or interpretation of the index test, flow and timing			

Reference	Kim 2013 [114]
	Indirectness: none
Comments	Diagnostic accuracy of US features of current and new guidelines

Reference	Koh 2018 [114]
Study type	Retrospective
Study methodology	Data source: thyroid nodules with benign or malignant diagnosis confirmed by surgery or US-guided FNA between November 2013 to July 2014 Recruitment: consecutive
Number of patients	n = 363 (370 nodules)
Patient characteristics	Age, mean (SD; range): 53.1 (13; 19-86) Gender (male to female ratio): 65:298 Ethnicity: Not specified Setting: Department of Radiology, Severance Hospital, Research Institute of Radiological Science, Yosnei University, College of Medicine Country: South Korea Inclusion criteria: nodules ≥10 mm in size, proven to be benign or malignant by surgery or diagnosed as benign or malignant on US-FNA wither on initial aspiration or repeat US-FNA after initial non-diagnostic or indeterminate cytology results. Exclusion criteria: Symptomatic thyroid cysts that were aspirated for symptom relief
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound (US)</u> Gray-scale US was performed with a 5-12 MHz linear probe by 14 board-certified radiologists with 1-19 years of experience in thyroid imaging (four staff radiologists, 10 fellows), including four study observers. One radiologist captured transverse and longitudinal images of each thyroid nodule from the picture PACS. Four observers with 19, 15, two and one years of experience in thyroid imaging, independently reviewed the images and filled out data interpretation forms. All four observers were blind to the clinical information of the patient or cytologic results during the image review. After assessing US features, final assessment of nodules was based on the Kim criteria, TI-RADS by Kwak et al, and the 2015 ATA

Reference	Koh 2018 [114]			
	guidelines. Test positive: Suspicious malignant for Kim; categories 4 and 5 for TI-RADS; low, intermediate and high suspicion for the 2015 ATA guidelines. <u>Reference standard: UG-FNAB or surgery (n=57)</u> US-fine needle aspiration cytology either on initial aspiration or repeat US-FNA after initial non-diagnostic or indeterminate cytology results. Time between measurement of index test and reference standard: not specified			
2×2 table	**Kim**	Reference standard +	Reference standard −	Total
	Index test +	158	303	461
	Index test −	54	965	1019
	Total	212	1268	1480
2×2 table	**K-TIRADS**	Reference standard +	Reference standard −	Total
	Index test +	193	759	952
	Index test −	19	509	528
	Total	212	1268	1480
2×2 table	**2015 ATA**	Reference standard +	Reference standard −	Total
	Index test +	197	999	1196
	Index test −	15	269	284
	Total	212	1268	1480
Statistical measures	<u>Index text: Ultrasound (Kim)</u> Sensitivity : 74.5% Specificity: 76.1% PPV: 34.3% NPV: 94.7% AUC:0.753 Accuracy: 75.9% <u>Index text: Ultrasound (Kwak-TIRADS)</u> Sensitivity : 91% Specificity: 40.1% PPV: 20.3%			

Reference	Koh 2018 [114]
	NPV: 96.4% AUC:0.809 Accuracy: 47.4% Index text: Ultrasound (2015 ATA) Sensitivity : 92.9% Specificity: 21.2% PPV: 16.5 % NPV: 94.7% AUC:0.804 Accuracy: 31.5%
Source of funding	No funding
Limitations	Risk of bias: none Indirectness: none
Comments	Diagnostic accuracy of US using Kim, K-TIRADS, 2015 ATA guidelines

Reference	Koseoglu Atilla 2018 [133]
Study type	Retrospective
Study methodology	Data source: patients with thyroid nodules who underwent FNA between 2010 and 2014 in Tepecik Training and Research Hospital Recruitment: consecutive
Number of patients	n = 2847 patients; 2614 finally included
Patient characteristics	Age, mean (SD): 51.01 (13.86) Gender (male to female ratio): 2263/351 Ethnicity: not specified Setting: Tepecick Training and Research Hospital

Reference	Koseoglu Atilla 2018 [133]
	Country: Turkey Inclusion criteria: consecutive patients with thyroid nodules undergoing FNA between 2010 and 2014; i.e. patients with solid nodules ≥1cm, or with mixed cystic nodules ≥1.5-2cm and songiform nodules ≥2cm and patients with high risk history who had nodules ≥5mm Exclusion criteria: patients with non-diagnostic FNABs
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound</u> US was performed by using high-spatial resolution US machines equipped with a 5.5-12.5 MHz linear probe. Nodules were classified according to the ACR TI-RADS guideline based on composition, echogenicity, shape, and margin characteristics of the nodules as bening (TR1), not suspicious (TR2), mildly suspicious (TR3), moderately suspicious (TR4) and highly suspicious (TR5). <u>Reference standard: US-guided FNA</u> FNAB was performed according was performed according to the 2009 ATA guideline. Cytopathological interpretation of FNAB samples was done using the Bethesda System for reporting Thyroid Cytopathology. Time between measurement of index test and reference standard: not specified

2x2 table	**ACR TI-RADS**	Reference standard +	Reference standard −	Total	Patients with non-diagnostic FNABs (Bethesda I) were excluded (n=233)
	Index test +	79	880	959	
	Index test −	22	1633	1655	
	Total	101	2513	2614	

Statistical measures	<u>Index text Ultrasound (ACR-TIRADS)</u> Sensitivity : 78.22% Specificity: 65%
Source of	Not specified

Reference	**Koseoglu Atilla 2018** [133]
funding	
Limitations	Risk of bias: none Indirectness: none
Comments	Diagnostic accuracy of US using ACR-TI-RADS

Reference	**Lauria Pantano 2018** [138]
Study type	Retrospective (cross-sectional)
Study methodology	Data source: nodules undergoing FNA from January 2015 to May 2016 Recruitment: not specified
Number of patients	n = 946 (1169 nodules)
Patient characteristics	Age, mean (SD): 56(13.3) Gender (male to female ratio): 199:946 Ethnicity: not specified Setting: Unit of Endocrinology and Diabetes of the Campus Bio-Medico University Country: Italy Inclusion criteria: All nodules undergoing FNA from January 2015 to May 2016 Exclusion criteria: nodules with TIR1 (non-diagnostic cytology)
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound</u> US of the thyroid gland and neck area was performed by experienced physicians at a frequency range of 10-12 MHz on a MyLab 50. Nodules were then classified according to the ATA, AACE/ACE/AME US and ACR TI-RADS risk stratification by an automated algorithm. Based on the description retrieved from medical records, a yes or no answer to each of the following features were input for each nodule into a Microsoft excel worksheet: purely cystic, more than 50% cystic, eccentric solid area, spongiform, spongiform with internal vascularisation, mixed cystic and solid, solid hypoechoic, solid marked (or very hypoechoic), solid isoechoic, hyperechoic,

Reference	Lauria Pantano 2018 [138]				
	macrocalcifications, microcalcifications, internal hyperechoic spots, calcified rim, irregular margins, taller than wide shape, rim calcifications with small extrusive soft tissue component, evidence of extrathyroidal extension/ suspicious nodes. Then, by using a pre-specified coding developed according to the above-mentioned guidelines, the software combined all the yes or no answers and automatically assigned one ATA, one AACE/ACE/AME and one ACR TI-RADS category to each nodule Reference standard: US-guided FNA US-guided FNA was performed by experienced physicians. All FNAs were performed based on an impartial clinical indication, independent from the study. FNA was performed by free-hand technique under US guidance, using a 23- or 25-gauge needle. Cytology specimens were evaluated by expert cytopathologists conforming to the Italian Reporting System for Thyroid Cytology as follows: TIR1 (non-diagnostic), TIR1C (nondiagnostic cystic), TIR2 (non-malignant/benign), TIR3a (low-risk indeterminate lesion), TIR3b (high-risk indeterminate lesion), TIR4 (suspicious of malignancy) or TIR 5 (malignant). TIR1 nodules were excluded from the study. Nodules with TIR1c cytology were considered clinically non-malignant/benign. TIR3b, TIR4 and TIR5 were classified as cytologically high risk of malignancy. TIR1c, TIR2 and TIR3a were considered cytologically benign. Time between measurement of index test and reference standard: not specified				
2×2 table	**ATA**	Reference standard +	Reference standard −	Total	N=54 nodules did not match the ATA sonographic patterns and were categorised as ‘ATA unclassified’, n=9 of those were cytologically high risk
	Index test +	87	525	612	
	Index test −	17	394	411	
	Total	104	919	1023	
2x2 table	**AACE/ACE/AME**	Reference standard +	Reference standard −	Total	N=28 did not match the AACE/ACE/AME categories and were categorised as ‘AACE/ACE/AME unclassified’ ; of these n=1 was cytologically high risk
	Index test +	109	786	895	
	Index test −	3	151	154	
	Total	112	937	995	
2x2 table	**ACR TI-RADS**	Reference standard +	Reference standard −	Total	
	Index test +	93	525	618	
	Index test −	20	439	459	
	Total	113	964	1077	

Reference	Lauria Pantano 2018 [138]
Statistical measures	<u>Index text Ultrasound (ATA)</u> Sensitivity : 83.7% Specificity: 42.9% PPV: 14.2% NPV: 95.9% <u>Index text Ultrasound (AACE/ACE/AME)</u> Sensitivity : 97.3% Specificity: 16.1% PPV: 12.2% NPV: 98.1% <u>Index text Ultrasound (ACR-TIRADS)</u> Sensitivity : 82.3 % Specificity: 45.5% PPV: 15% NPV: 96.6%
Source of funding	No funding
Limitations	Risk of bias: none Indirectness: none
Comments	Diagnostic accuracy of US using ATA, AACE/ACE/AME, ACR-TI-RADS

Reference	Lim-Dunham 2017 [150]
Study type	Retrospective study
Study methodology	Data source: paediatric patients who underwent US fine-needle aspiration biopsy Recruitment: consecutive
Number of patients	n = 33 (39 nodules)
Patient characteristics	Age, median (range): Benign nodules 16 (8-18); malignant 16.5 (9-18) Gender (male to female ratio): 5:28

Reference	Lim-Dunham 2017 [150]
	Ethnicity: not specified Setting: Department of Radiology, Loyola University Chicago Stritch School of Medicine Country: USA Inclusion criteria: patients ages 18 years and younger who were referred to the radiology department for US-FNAB of one or more thyroid nodules at authors' medical centre between 1996 and 2016 Exclusion criteria: lack of preliminary US images (n=29), uncertainty in correlating the identity of the nodule on US with pathology (n= 3) and poor US image quality (n=14).
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound</u> All individuals underwent diagnostic gray-scale US with colour Doppler using high frequency (8-15 MHz) linear array transducers. Based on US features, each nodule was assigned a level of suspicion of malignancy based on the 2015 ATA management guidelines: benign very low suspicion, low suspicion intermediate suspicion, high suspicion. <u>Reference standard: UG-FNAB or surgery (n=14)</u> Two board-certified paediatric radiologists each with more than 10 years' experience performed the FNAB procedures in the radiology department by free-hand technique with US guidance using a 25-gauge needle. Nodules less than 5 mm or located adjacent to the common carotid artery or internal jugular vein were not considered for UG-FNAB. Between two and eight samples were taken from the solid component of each nodule. A staff pathologist was present during the procedure to verify diagnostic adequacy of the sample. A decision to proceed with surgical thyroidectomy was made by the endocrine surgeon. If a patient did not undergo surgery, the cytopathology from the UG-FNAB was used to classify nodules. Nodules were classified according to the Bethesda System for reporting Thyroid Cytopathology as follows: Class I, nondiagnostic; Class II benign; Class III, atypia or follicular lesion of undetermined significance; Class IV, follicular neoplasm/suspicion for a follicular neoplasm; Class V, suspicious for malignancy; and Class VI, malignant. Bethesda Class II and III were considered benign and Class IV, V and VI were considered malignant. Time between measurement of index test and reference standard: not specified

2×2 table

	Reference standard +	Reference standard −	Total	Notes: 14 nodules were classified based on surgical pathology (n=2 benign, n=12 malignant)
Index test +	12	9	21	
Index test −	0	12	12	

Reference	**Lim-Dunham 2017** [150]				
	Total	12	21	33	Analysis included each patient's largest nodule observation.
Statistical measures	Index text Ultrasound Sensitivity : 100% Specificity: 57.1% PPV: 57.1% NPV: 100%				
Source of funding	Not specified				
Limitations	Risk of bias: none Indirectness: none				
Comments	Diagnostic accuracy of US in children using the 2015 ATA guidelines				

Reference	**Macedo 2018** [114]
Study type	Prospective
Study methodology	Data source: patients with thyroid nodules attending tertiary university-based hospital between July 2014 to August 2015 Recruitment: consecutive unselected patients
Number of patients	n = 178 (195 nodules)
Patient characteristics	Age, median (range): 59 (49-66) Gender (male to female ratio): 9:169 Ethnicity: Not specified Setting: Endocrinology Division, Santa Casa de Misericordia de Porto Alegre (tertiary, university-based hospital) Country: Brazil (Southern iodine-replete area) Inclusion criteria: unselected patients with thyroid nodules attending hospital between July 2014 and August 2015

Reference	Macedo 2018 [114]
	Exclusion criteria: Patients with known thyroid cancer and/or purely cystic nodules
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	Index test: Ultrasound (US) (using TI-RADS & ATA) Thyroid Ultrasound Conventional B-mode and Doppler images of the neck and thyroid gland were obtained by ultrasound machine (ACUSON S2000, Siemens and ACUSON Antares, Siemens HealthCare, Erlangen, Germany) using a high-frequency probe (12 MHz). All US examinations were performed by the same radiologist (RFI) who has more than 10 years of experience in thyroid ultrasound. All images were examined on real-time two-dimensional gray-scale and Doppler imaging. Findings that were considered in favour of malignancy were hypoechoic or markedly hypoechoic in echogenicity; irregular, micro lobulated, or ill-defined margins; presence of micro calcification; round shape and the presence of lymphadenopathy. Prospective evaluation using the modified Russ classification was performed. Each nodule was classified into a TI-RADS category (2, 3, 4 and 5) based on US features. Benign patterns: category 3 or 2; Suspect patterns: category 5 or 4. Posteriorly, the same radiologist (RFI), blind about pathological results, scored all evaluated nodules based on new ATA thyroid nodule guideline. Based on the number of features suspicious for malignancy four different sonographic patterns were considered: ‘very low suspicion’; ‘low suspicion’; ‘intermediate’ and ‘high suspicion’. Benign patterns: low risk and very low risk category; Suspect patterns: high risk and intermediate risk category. Reference standard: FNA, cytology, histology All 195 nodules were submitted to FNA performed by using a capillary US-guided technique with a 23-gauge needle attached to a 10 mL disposable plastic syringe. Only one needle pass was made per lesion in most cases. Cytology smears were prepared on four to six slides. One cytopathologist from the institution with vast experience in thyroid pathology interpreted the smears. A thyroid FNA specimen was considered satisfactory if at least 6 groups of follicular cells were present, and each group comprised at least 10 cells. The Bethesda System for Cytological classification of Thyroid Nodules was used to interpret smears as: 1) non-diagnostic or unsatisfactory,2) benign, 3) atypia of undetermined significance, 4) follicular neoplasm or suspicious for a follicular neoplasm, 5) suspicious for malignancy and 6) malignant. Histology was available for 45 cases: Surgery was indicated based on cytopathological results (Bethesda 4,5 and 6), or when the nodule was benign (Bethesda 2) but larger than 2-3 cm and causing compressive symptoms. Anatomopathological examinations of tissue samples obtained at thyroidectomy were carried out according to the World Health organization Guidelines and the pathology reports pertaining to these samples were considered identical to the gold standard for the diagnosis of thyroid cancer.

Reference	**Macedo 2018** [114]				
	Time between measurement of index test and reference standard: not specified				
2×2 table	TIRADs	Reference standard +	Reference standard −	Total	Notes: Only Bethesda categories 2 and 6 were used (n=138) to compare TI-RADS and ATA score with cytological results.
	Index test +	5	51	56	
	Index test −	0	82	82	
	Total	5	133	138	
2×2 table	ATA	Reference standard +	Reference standard −	Total	Notes: Only Bethesda categories 2 and 6 were used (n=138) to compare TI-RADS and ATA score with cytological results.
	Index test +	5	33	38	
	Index test −	0	100	100	
	Total	5	133	138	
Statistical measures	Index text: Ultrasound (TI-RADS) Sensitivity : 100% Specificity: 61.6% NPV: 100% Accuracy: 63% Index text: Ultrasound (ATA) Sensitivity : 100% Specificity: 75 % NPV: 100% Accuracy: 76%				
Source of funding	Not stated				
Limitations	Risk of bias: serious risk of bias due to flow and timing, potential bias in the interpretation of the reference standard Indirectness: none				
Comments	Diagnostic accuracy of ultrasonography using ATA and TI-RADS risk stratification.				

Reference	**Maino 2018** [114, 170]
Study type	Prospective
Study methodology	Data source: patients with nodules submitted to FNAC from November 2016 to June 2017 Recruitment: not specified

Reference	**Maino 2018 [114, 170]**
Number of patients	n = 340 (432 nodules)
Patient characteristics	Age, mean (SD, range): 57 (14.3, 16-86) Gender (male to female ratio): 77:263 Ethnicity: Not specified Setting: Department of Medical, Surgical and Neurological Sciences, University of Sienna Country: Italy Inclusion criteria: all nodules submitted to FNAC for diagnostic purposes Exclusion criteria: not specified; nodules with non-diagnostic cytology were finally excluded from analysis
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	Index test: Ultrasound (US) Neck US was performed by the same experienced endocrinologist of our staff using a high-resolution US colour Doppler apparatus with a 7.5 MHz linear transducer. US features of each thyroid nodule were described and recorded in the database by the endocrinologist who performed the examination and nodules were stratified using sonographic patterns as described and published in the 2015 ATA guidelines into: benign, very low suspicion, low suspicion, intermediate suspicion and high suspicion categories and as described in the ETA US risk stratification system into: EU-TIRADS 2 (benign), EU-TIRADS 3 (low risk), EU-TIRADS 4 (intermediate risk) and EU-TIRADS 5 (high risk) Reference standard: FNA, cytology, histology US-guided FNAC was performed for at least two separate passes for each thyroid nodule by using a 23/25-gauge needle. Material was air dried, trained with May-Grunwald Giemsa and interpreted by the same experienced cytologist. Cytology reports from US-guided FNAC of thyroid nodules were based on the five categories according to the criteria of the British Thyroid Association (Thy 1: nondiagnostic; Thy2: benign, Thy 3: undetermined significance; Thy 4: suspicious for malignancy; and Thy 5: malignant) All patients with Thy4/Thy5 cytologies were send to surgery; in Thy2 only those with compressive symptoms were send to surgery and the remaining were observed by annual follow-up . Time between measurement of index test and reference standard: not specified

Reference	**Maino 2018** [114, 170]				
2×2 table	ATA	Reference standard +	Reference standard −	Total	Notes: 381 nodules finally included, excluding Thy 1 nodules: nondiagnostic ; 2x2 calculated excluding Thy3 nodules with undetermined significance (n=31)
	Index test +	11	64	75	
	Index test −	3	272	275	
	Total	14	336	350	
2×2 table	EU TIRADs	Reference standard +	Reference standard −	Total	Notes: 381 nodules finally included, excluding Thy 1 nodules: nondiagnostic; 2x2 calculated excluding Thy 3 nodules with undetermined significance (n=31)
	Index test +	11	66	77	
	Index test −	3	270	273	
	Total	14	336	350	
Statistical measures	<u>Index text: Ultrasound (ATA)</u> Sensitivity : 78.6% Specificity: 80.9% <u>Index text: Ultrasound (EU-TI-RADS)</u> Sensitivity : 78.6% Specificity: 80.4%				
Source of funding	Ministero Italiano dell'Universita e Ricerca				
Limitations	Risk of bias: serious risk due to patient selection, potential bias in the interpretation of index test results, flow and timing Indirectness: none				
Comments	Diagnostic accuracy of ultrasonography using ATA and EU TI-RADS risk stratification.				

Reference	**Martinez-Rios 2018** [175]
Study type	Retrospective cohort
Study methodology	Data source: children referred to Hospital for Sick Children, Toronto with US and clinical data from January 1992 to October 2015 Recruitment: not specified, children referred to hospital for the evaluation of thyroid nodules
Number of patients	n = 124 (125 nodules)

Reference	Martinez-Rios 2018 [175]
Patient characteristics	Age, mean (SD, range): 13.6 (3.1, 3.3-17.7) Gender (male to female ratio): 40: 84 Ethnicity: Not specified Setting: Hospital of Sick Children, Toronto Country: Canada Inclusion criteria: patients younger than 18 years; benign or malignant thyroid nodules with confirmed histology and or cytology or no histology available but a minimum 2-year follow-up with clinical sonographic stability of the nodule; thyroid nodules measuring more than 10 mm. Exclusion criteria: poor image quality/no US imaging available; previous exposure to irradiation; previous oncological conditions; known family history of RET, DICERI or PTEN gene mutations.
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound (US) (using TI-RADS & ATA)</u> US findings of a combination of gray-scale and colour Doppler US images of the thyroid gland and the bilateral cervical lymph nodes compartments were analysed. Imaging was performed with iU22 and Alpio ultrasound equipment. US examinations were performed according to the standards protocols for thyroid gland/neck protocols of the research department. All examinations were performed with high-frequency linear-array transducers in longitudinal and transverse planes. The entire US examination was reviewed and background echogenicity of the thyroid gland, number of thyroid nodules, nodule location within the gland, and size of nodules in mm. US data were reviewed by three radiologists using the ATA and TI-RADS methods. US studies were initially scored by a consensus of two paediatric radiologists (each with 2 years' experience) and then a score by an independent paediatric radiologist (with 37 years' experience) was obtained. Readers were blinded to final diagnoses and clinical data. For the purposes of assigning test characteristics, when assessing the ATA method: high and intermediate suspicion classifications were considered as probably malignant; low suspicion, very low suspicion and benign were considered as probably benign. For TI-RADS: categories 4a, 4b, 4c and 5 were considered as probably malignant; categories 2 and 3 were considered as probably benign. <u>Reference standard: histopathology/cytology or 2-year follow-up of clinical outcome for non-operative cases</u> The reference standard was surgical histopathology or cytology or at least 2 years' clinical follow-up without evolution of malignant features

Reference	Martinez-Rios 2018 [175]				
	Time between measurement of index test and reference standard: not specified				
2×2 table	**TIRADs**	Reference standard +	Reference standard −	Total	Notes: excluded 1 histologically indeterminate nodule
	Index test +	52	58	110	
	Index test −	0	13	13	
	Total	52	71	123	
2×2 table	**ATA**	Reference standard +	Reference standard −	Total	Notes: excluded 1 histologically indeterminate nodule
	Index test +	45	22	67	
	Index test −	7	49	56	
	Total	52	71	123	
Statistical measures	Index text: Ultrasound (TI-RADS) Sensitivity : 100% Specificity: 18.3% PPV: 47.3% NPV: 100% Index text: Ultrasound (ATA) Sensitivity :86.5 % Specificity: 69% PPV:67.2 % NPV: 87.5%				
Source of funding	Not stated				
Limitations	Risk of bias: none Indirectness: none				
Comments	Diagnostic accuracy of ultrasonography using ATA and TI-RADS risk stratification.				

Reference	Middleton [179]
Study type	Prospective
Study methodology	Data source: patients who had undergone US and US-guided FNA of a focal nodule between August 2006 and May 2010 Recruitment: not specified

Reference	Middleton [179]
Number of patients	n = 3315 (3822 nodules)
Patient characteristics	Age, mean (SD): 54.4(18-97) Gender (male to female ratio): 766:3056 Ethnicity: not specified Setting: Mallinckordt Institute of Radiology, Washington University St Louis; Department of Diagnostic radiology, Mayo Clinic, Rochester; Department of Radiology, The Parelman school of medicine at the University of Pennsylvania; department of Diagnostic imaging, Rhode island hospital, Brown University; Department of Radiology, University of Kentucky College of Medicine; Department of radiology, Stanford University Medical Centre Country: USA Inclusion criteria: All patients 18 years or older who had undergone diagnostic thyroid ultrasound examinations and US-guided FNA of a focal nodule between August 2006 and May 2010. Exclusion criteria: non-diagnostic findings by FNA, surgical histologic analysis or both (n=173), or results that were indeterminate or suspicious for malignancy with no subsequent definitive diagnosis (n=227).
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound</u> Images of the biopsied nodules were obtained using a variety of commercially available ultrasound units, with specific attention prospectively directed to nodule characteristics (e.g. composition, echogenicity, margins, echogenic foci) similar to those used int the ACR lexicon to describe thyroid nodules. The sonographic images and cine clips of thyroid nodules were saved and sent to a central reading site. Nodules were analysed at the central study site by two radiologists who had access to the original ultrasound report but had no knowledge of the findings of cytologic analysis. Points were assigned to each nodule for the separate categories of composition, echogenicity, margins, and echogenic foci on the basis of the TIRADS guidelines. Nodule shape (i.e. taller than wide) was included in TIRADS but not in the present analysis. The sum of the points in each category determined the TIRADS level assigned to each nodule, with TR1 indicating 0 points; TR2, 2 points, TR3, 3 points, TR4, 4-6 points; TR5 7 or more points. <u>Reference standard: US-guided Fine-needle aspiration</u> A total of one to three nodules were biopsied for each patient. The procedure used for specimen procurement was left to the discretion of the physician performing the FNA. The physician was free to perform the number of needle passes deemed appropriate at their institution.

Reference	**Middleton** [179]				
	Cytopathologic interpretations from each institution were used to distinguish between benign and malignant nodules. The results of the FNA were divided into five categories: malignant, suspicious for malignancy, indeterminate, benign and nondiagnostic. Nodules for which results were suspicious for malignancy, indeterminate or nondiagnostic were excluded from the study unless they were followed by diagnostic FNA or surgical resection that provided histologic confirmation of malignancy or benignancy. Time between measurement of index test and reference standard: not specified				
2×2 table		Reference standard +	Reference standard −	Total	Notes: 303 malignant nodules were diagnosed on the basis of cytologic analysis, were resected and had histologically confirmed diagnosis.
	Index test +	297	1488	1785	
	Index test −	55	1582	1637	
	Total	352	3070	3422	
Statistical measures	<u>Index text Ultrasound (ACR-TIRADS)</u> Sensitivity : 84.4% Specificity: 51.5% PPV: 16.6% NPV: 96.6 %				
Source of funding	Not specified				
Limitations	Risk of bias: none Indirectness: none				
Comments	Diagnostic accuracy of US using TIRADS classification				

Reference	**Moon 2010** [187]
Study type	Retrospective
Study methodology	Data source: patients that underwent US and US-guided FNAB from June 2007 to August 2007 Recruitment: consecutive
Number of	n = 1024 (1083 nodules)

Reference	Moon 2010 [187]
patients	
Patient characteristics	Age, median (range): 50(16-83) Gender (male to female ratio): 138:886 Ethnicity: not specified Setting: Severance Hospital (reference centre) Country: South Korea Inclusion criteria: nodules with benign or malignant results at cytologic evaluation, or with thyroid surgery performed after cytologic results suggestive of papillary thyroid carcinoma, indeterminate results, or with benign or malignant results at cytologic examination or with surgery in case of indeterminate results or inadequate cytologic results. Exclusion criteria: not specified
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: gray-scale Ultrasound, elastography</u> All gray-scale and power Doppler US examinations were performed by using a 5-12 MHz linear probe. Power Doppler examinations were performed by using the standard equipment settings for thyroid glands. US examinations were performed by one of five radiologists with 7 to 13 years of experience. US features of all thyroid nodules that underwent US-guided FNA were prospectively recorded according to the internal component, echogenicity, margin, calcifications, shape, and final assessment at the time of FNA by the radiologists who performed the US examination and US-guided FNAB. Vascularity was determined at power Doppler US. Three types of vascularity were identified: type 1, no vascularity; type 2, peripheral vascularity; type 3, intranodular vascularity. Suspicious malignant gray-scale US features were classified by using criteria of marked hypoechogenicity, non-circumscribed margin, microcalcifications and taller than wide shape. When thyroid nodules showed one or more of these suspicious malignant features, they were classified as suspicious. When thyroid nodules showed none of these suspicious features, they were classified as probably benign. To compare the diagnostic performance of the combination of only gray-scale US features and the combination of gray-scale and power Doppler US features, six criteria were assigned as follows: criterion 1, any single suspicious gray-scale US-feature; criterion 2, addition of peripheral and intranodular vascularities as one of suspicious features to criterion 1; criterion 3, addition of peripheral vascularity as a suspicious feature to criterion 1; criterion 4, addition of intranodular vascularity as a suspicious feature to criterion 1; criterion 5, addition of no vascularity as a suspicious feature to criterion1; and criterion 6, AACE and AME guidelines-all hypoechoic nodules with at least one of the following additional US features: irregular margins, intranodular vascular spots, taller-than-wide shape, or microcalcifications. <u>Reference standard: US-guided Fine-needle biopsy</u>

Reference	Moon 2010 [187]				
	US-guided FNABs were performed by the same radiologist who performed US examinations, by using a 5-12 MHz linear probe. US-guided FNAB was performed on either thyroid nodules with suspicious US features or the largest thyroid nodules without suspicious US features. It was not performed on entirely cystic nodules. US-guided FNAB was performed with a 23-gauge needle and a 2-mL disposable plastic syringe. Time between measurement of index test and reference standard: not specified				
2×2 table	**Kim**	Reference standard +	Reference standard −	Total	
	Index test +	227	115	342	
	Index test −	42	699	741	
	Total	269	814	1083	
2x2	**Kim+USD**	Reference standard +	Reference standard −	Total	
	Index test +	245	387	632	
	Index test −	24	427	451	
	Total	269	814	1083	
2x2	**AACE/AME**	Reference standard +	Reference standard −	Total	
	Index test +	220	168	388	
	Index test −	49	646	695	
	Total	269	814	1083	
Statistical measures	<u>Index text Ultrasound (Kim)</u> Sensitivity : 84.4% Specificity: 85.9% PPV: 66.4% NPV: 94.3 % <u>Index text Ultrasound (Kim+ USD)</u> Sensitivity : 91.1% Specificity: 52.5% PPV: 38.8% NPV: 94.7% <u>Index text Ultrasound (AACE/AME)</u> Sensitivity : 81.8 %				

Reference	Moon 2010 [187]
	Specificity: 79.4 % PPV: 56.7% NPV: 92.9%
Source of funding	Not specified
Limitations	Risk of bias: serious due to risk of bias in the interpretation of the index test and reference standard test results Indirectness: none
Comments	Diagnostic accuracy of gray-scale US and power Doppler US (USD) using Kim, Kim +USD Rago, AACE/AME classification

Reference	Moon 2012 [188]
Study type	Retrospective
Study methodology	Data source: thyroid nodules imaged at gray-scale US, elastography and US-guided FNA from June to November 2009 Recruitment: not specified
Number of patients	n = 676 (703 nodules)
Patient characteristics	Age, mean (range): 49.7(18-79) Gender (male to female ratio): 120:556 Ethnicity: not specified Setting: Department of Radiology, Yosnei University College of Medicine Country: South Korea Inclusion criteria: nodules with benign or malignant results at cytologic evaluation, with thyroid surgery performed after obtaining cytologic results suspicious for papillary thyroid carcinoma or indeterminate results, or with benign or malignant results at follow-up US-guided FNA or thyroid surgery after cytologic results of inadequate specimen.

Reference	Moon 2012 [188]				
	Exclusion criteria: nodules containing cystic components (n=101), unsuccessful elastography (n=17), nodules suspicious for papillary thyroid carcinoma or with indeterminate or inadequate results (n=43) at cytologic evaluation that had not undergone surgery or repeat US-guided FNA				
Target condition(s)	Thyroid cancer				
Index test(s) and reference standard	Index test: gray-scale Ultrasound, elastography All gray-scale US images were obtained by using a 6-14 MHz linear array transducer. Real time gray-scale US was performed by one of eight radiologists with 1 to 15 years of experience. Gray-scale US features of thyroid nodules that underwent US-guided FNA were prospectively recorded according to the internal component, echogenicity, margin, calcifications, shape, and final assessment at the time of FNA by the radiologists who performed the US examination and FNA. Suspicious malignant gray-scale US features included marked hypoechogenicity, poorly defined margin, microcalcifications and taller than wide shape. When thyroid nodules showed one or more of these suspicious malignant features, they were assessed as suspicious. When thyroid nodules showed no suspicious features, they were assessed as probably benign. After gray-scale examination, elastography was routinely performed by the same radiologists who performed gray-scale US, in thyroid nodules detected at gray-scale US and targeted for US-guided FNA by using the same US machine and probe, using a free-hand technique. Elastography images were classified according to the scores by Rago et al and Asteria et al. According to Rago elasography score were classified on a scale from 1 to 5; score of 4 and 4 were classified as suspicious for malignancy. According to Asteria et al elastography scores were classified on a scale from 1 to 4; nodules with score of 3 and 4 were classified as suspicious for malignancy. Reference standard: US-guided Fine-needle biopsy US-guided FNA biopsy was performed by the same radiologist who performed gray-scale US and elastography, by using a 23-gauge needle and a 2-mL disposable plastic syringe with a freehand technique. Time between measurement of index test and reference standard: not specified				
2×2 table	**Kim**	Reference standard +	Reference standard −	Total	Surgery performed after FNA in 221 nodules (202 patients); UGFNA for two nodules in 27 patients and one nodule in 649 patients.
	Index test +	199	162	361	
	Index test −	18	324	342	
	Total	217	486	703	
2x2	**Kim+USE Rago**	Reference standard +	Reference standard −	Total	Surgery performed after FNA in 221 nodules (202 patients); UGFNA for two nodules in 27 patients and one nodule in 649 patients.
	Index test +	200	170	370	
	Index test −	17	316	333	
	Total	217	486	703	
2x2	**Kim+USE Asteria**	Reference standard +	Reference standard −	Total	Surgery performed after FNA in 221

Reference	Moon 2012 [188]				
	Index test +	205	255	460	nodules (202 patients); UGFNA for two nodules in 27 patients and one nodule in 649 patients.
	Index test −	12	231	243	
	Total	217	486	703	
Statistical measures	Index text Ultrasound (Kim) Sensitivity : 91.7% Specificity: 66.7 % PPV: 55.1% NPV: 94.7 % Accuracy: 74.4% Index text Ultrasound (Kim+ USE Rago) Sensitivity : 92.2% Specificity: 65% PPV: 54.1% NPV: 94.9% Accuracy: 73.4% Index text Ultrasound (Kim+ USE Asteria) Sensitivity : 94.5% Specificity: 47.5% PPV: 44.6% NPV: 95.1% Accuracy: 62%				
Source of funding	Not specified				
Limitations	Risk of bias: serious due to risk of bias in the interpretation of the index test and reference standard test results Indirectness: none				
Comments	Diagnostic accuracy of gray-scale US and elastography (USE) using Kim, Kim +USE Rago, Kim+USE Asteria classification				

Reference	Na 2016 [190]
Study type	Retrospective
Study methodology	Data source: patients with thyroid nodules with final diagnosis who had FNA or core needle biopsy (CNB) at low and high cancer volume institutions, from January 2010 to May 2011

Reference	Na 2016 [190]
	Recruitment: consecutive enrolment of predetermined number of 2000 nodules (1000 from each low and high cancer volume institutions)
Number of patients	n = 1802 (2000 nodules)
Patient characteristics	Age, mean (SD): 51.2 (12.2) Gender (male to female ratio): 415:1387 Ethnicity: not specified Setting: low and high cancer volume institutions (two primary medical centres, two tertiary hospitals) Country: South Korea Inclusion criteria: patients enrolled from low and high cancer volume institutions from January 2010 to May 2011, with thyroid nodules (≥1cm) with final diagnosis, who had undergone FNA or CNB. Exclusion criteria: no final diagnosis (n=1242), entirely calcified nodules for which US characteristics could not be analysed (n=14)
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound</u> A high resolution US scan using a 10-12MHz or 5-14MHz linear-array transducer was employed. US images were retrospectively reviewed by one of three experienced radiologists with 19, 16 and 12 years of experience. All reviewers with no knowledge of FNA results or final diagnosis assessed the following US features of thyroid nodules: internal content, echogenicity, margin, shape, calcification, nodule vascularity, spongiform appearance, and comet-tail artefact. Colour Doppler US images were available in 1295 nodules. Risk stratification of nodules was according to K-TIRADS and was based on solidity and echogenicity. Nodules were classified into 5 categories: 1. no nodule, 2. Benign, 3. Low suspicion, 4.intermediate suspicion, 5. High suspicion. <u>Reference standard: US-guided Fine-needle aspiration or Core-needle biopsy</u> FNA was performed with a conventional method and at least two samplings were performed for each nodule. CNB was performed using a disposable 18-gauge, single-or-double action spring-activated needle. The interpretation of FNA was based on the Bethesda System for Reporting Thyroid Cytopathology, and CNB results were diagnosed with a six-tier pathology reporting system. In case of a nondiagnostic result from the initial FNA, the results of repeated FNA or CNB were used. Final diagnoses were determined by surgical resections in 239/1546 (15.5%) benign nodules, 451/454 (99.3%) malignant nodules and by CNB in three cases (0.7%)

Reference	**Na 2016 [190]**
	Time between measurement of index test and reference standard: not specified

2×2 table		Reference standard +	Reference standard −	Total	
	Index test +	367	462	829	Surgery:690 nodules CNB: 3 nodules Repeated FNA or CNB: 381 nodules FNA or CNB and follow-up US: 926 nodules
	Index test −	87	1084	1171	
	Total	454	1546	2000	

Statistical measures	<u>Index text Ultrasound (Korean-TIRADS)</u> Sensitivity : 80.8% Specificity:70.6 % PPV: 44.6% NPV: 92.6 % Accuracy: 72.9%
Source of funding	Not specified
Limitations	Risk of bias: none Indirectness: none
Comments	Diagnostic accuracy of K-TIRADS US classification

Reference	**Pandya 2018 [207]**
Study type	Retrospective
Study methodology	Data source: subjects undergoing first-time FNA of a thyroid nodule between October 2009 and February 2016, identified via the electronic medical record system and Department of Radiology records. Recruitment: consecutive
Number of patients	n = 1947
Patient characteristics	Age, mean (range): 56 (26-86) Gender (male to female ratio): 475:1472

Reference	Pandya 2018 [207]
	Ethnicity: not specified Setting: Department of Radiology, University of Michigan Health Systems Country: USA Inclusion criteria: subjects undergoing first-time FNA of a thyroid nodule between October 2009 and February 2016, identified via the electronic medical record system and Department of Radiology records, for patients that had undergone repeat procedural visits of FNA of a thyroid nodule only the most recent procedure was included Exclusion criteria:
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	Index test: Ultrasound The diagnostic thyroid ultrasound was performed with electronically focused linear transducers ranging in frequency from 6 to 15 mHz. One radiologist, with 9 years' experience, retrospectively reviewed the diagnostic thyroid ultrasound images on a picture archiving and communication workstation, determined whether each nodule had microcalcifications, assigned each nodule one of 14 morphologic descriptors according to the 2015 ATA guidelines, and placed each nodule into one of five 2015 ATA categories of risk (ATA 1, 2, 3, 4, 5), based on echogenicity, margins, shape, cystic nature and presence of microcalcifications. Reference standard: US-guided Fine-needle aspiration US-FNA was performed by a member of the cross-sectional interventional service of the radiology department at the University of Michigan. The diagnostic FNA was performed with electronically focused linear transducers ranging in frequency from 6 to 15 mHz. Aspirations were performed with a series of 25-gauge needles and free-hand technique under direct sonographic visualization. The needle was inserted into the targeted nodules, and aspirations were performed with a capillary method. Varying areas of the nodule were sampled in each pass. A minimum of six passes were performed unless a cytopathologist was present. In these latter cases, cellular adequacy was obtained. The maximum number of passes was 12. All thyroid FNAS were interpreted according to the Bethesda System for Reporting Thyroid Cytopathology in the categories as follows. Nondiagnostic, benign, atypia of undetermined significance, suspicious for malignancy and malignancy. For subjects whose initial FNA results were inconclusive (i.e. nondiagnostic, atypia or follicular lesion of undetermined significance, or suspicious for neoplasm) the electronic medical record was reviewed to determine whether a subsequent targeted FNA or surgery was performed to enable a more definitive diagnosis within a year of the initial FNA. In such cases that final diagnosis was recorded. In cases where no definitive diagnosis was obtained, the initial cytopathology was considered the final result.

Reference	**Pandya 2018 [207]**				
	Time between measurement of index test and reference standard: not specified				
2×2 table		Reference standard +	Reference standard −	Total	Nodules identified as indeterminate by US were treated as benign
	Index test +	85	546	631	
	Index test −	13	706	719	
	Total	98	1252	1350	
Statistical measures	<u>Index text Ultrasound</u> Sensitivity : 86.7% Specificity: 56.4%				
Source of funding	Not specified				
Limitations	Risk of bias: none Indirectness: none				
Comments					

Reference	**Park 2016 [215]**
Study type	Retrospective
Study methodology	Data source: thyroid nodules assessed by US-guided FNA between August and October 2010 at tertiary referral centre Recruitment: not specified
Number of patients	n = 592 (622 nodules)
Patient characteristics	Age, mean (range): 49.8 (14-86) Gender (male to female ratio): 119:473

Reference	Park 2016 [215]
	Ethnicity: not specified Setting: tertiary referral centre Country: South Korea Inclusion criteria: thyroid nodules assessed by US-guided FNA between August and October 2010 at tertiary referral centre Exclusion criteria: nonthyroidal lesions, nodules smaller than 0.5cm, and nodules with no acceptable follow-up or operation
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound</u> Thyroid US was performed at a frequency range of 7 to 15 MHz on an iU22 by one of 7 radiologists. All radiologists had 1 to 11 years of experience in thyroid imaging. The US features were prospectively analysed by the radiologist who performed the US examination. All nodules were classified into one of three categories: benign, intermediate, malignant according to the KSThR guidelines. Taking into account internal components, echogenicity, margin, calcification, shape, and orientation of the thyroid nodule, based on the KSThR nodules were classified as follows: Probable benign, indeterminate or suspicious malignant. <u>Reference standard: US-guided Fine-needle aspiration</u> US-FNA was performed by one of the seven trained radiologists who conducted the US examinations. US-FNA was performed manually with a 23-gauge needle attached to a 2-mL disposable syringe. On average 1-2 passes were performed for each nodule. One of six cytopathologists interpreted the FNA specimens. All cases were reported using a six-tiered diagnostic system according to the Bethesda System for Reporting Thyroid Cytopathology. Nodules were considered benign if they met at least one of the following conditions: 1. They were pathologically confirmed as benign by thyroidectomy or core needle biopsy; 2. Had US-follow up of at least 2 years with either no interval change or a decrease in size after an initial benign cytology finding; and 3. Had benign cytology in more than two FNAs. Nodules were malignant if they were confirmed as malignant thyroid carcinoma by two serial FNAs or by thyroidectomy. Time between measurement of index test and reference standard: not specified

2×2 table		Reference standard +	Reference standard −	Total	Nodules identified as indeterminate by US were treated as benign
	Index test +	140	16	156	
	Index test −	11	303	314	
	Total	151	319	470	

Reference	Park 2016 [215]
Statistical measures	Index text Ultrasound Sensitivity : 93% Specificity: 95%
Source of funding	Samsung Medical Centre
Limitations	Risk of bias: serious due to risk of bias in the interpretation of the index test results Indirectness: none
Comments	Diagnostic accuracy of KSThR US classification

Reference	Persichetti 2018 [218]
Study type	Prospective
Study methodology	Data source: patients referred for US-guided FNA from January to September 2016 Recruitment: consecutive
Number of patients	n = 789 (1100 nodules)
Patient characteristics	Age, mean (SD): 55 (14) Gender (male to female ratio): 181:608 Ethnicity: white Setting: Regina Apostolorum Thyroid Centre Country: Italy Inclusion criteria: nodules from patients referred for US-guided FNA at the Regina Apostolorum Thyroid Centre from seven endocrine clinics from January to September 2016

Reference	Persichetti 2018 [218]				
	Exclusion criteria: patients with class I of the Bethesda System for Reporting Thyroid Cytopathology or an incomplete assessment i.e. patients who after the first cytological evaluation did not repeat a second FNA or who did not undergo surgery and were lost at follow-up				
Target condition(s)	Thyroid cancer				
Index test(s) and reference standard	<u>Index test: Ultrasound</u> All sonographic examinations were performed with two identical state-of-the art US machines equipped with a 5-to 15-MHz linear transducer and with colour Doppler, power Doppler, and elastography software. US images were independently evaluated by four examiners for the assignment of the malignancy risk according to the ATA, BTA, AACE/ACE/AME guidelines on the basis of US features described in their classification systems. The operators had specific experience in endocrine neck US examination for a time that ranged from 9 to 21 years. Based on the ATA, nodules were classified as: benign, low suspicion, intermediate suspicion or high suspicion Based on the BTA, nodules were classified as: U1 Normal, U2 benign, U3 intermediate, U4 suspicious and U5 Malignant. Based on the AACE/ACE/AME, nodules were classified as: low risk, intermediate risk and high risk To compare the diagnostic accuracy of the different classification systems, three major US categories were generated pooling together the classes characterised by a similar estimated risk of cancer. The comparison of the high-risk US classes vs low-intermediate risk categories and of the high-and intermediate-risk vs the low-risk categories was performed. <u>Reference standard: US-guided Fine-needle aspiration</u> FNA was performed under US guidance according to the US procedure described previously, and cytological samples were classified in six diagnostic categories according to the TBSRTC. To decrease the risk of false-negative results, patients had a second FNA 6 to 8 months after the first cytological assessment. Confirmed class III cytology nodules with positive immunocytochemical and clinical features or any suspicious finding were submitted to surgery. Class IV, V and VI nodules were committed to surgical treatment. Time between measurement of index test and reference standard: not specified				
2×2 table	**BTA**	Reference standard +	Reference standard −	Total	987 nodules were included in the analysis; n=39 patients with incomplete assessment and n=74 with Bethesda class I were excluded U2 as benign, U3/4/5 as malignant
	Index test +	141	304	445	
	Index test −	15	527	542	
	Total	156	831	987	
2×2 table	**ATA**	Reference standard +	Reference standard −	Total	987 nodules were included in the analysis; n=39 patients with incomplete assessment and n=74 with Bethesda class I were excluded ATA: high/intermediate vs benign, very low and
	Index test +	145	399	544	
	Index test −	11	432	443	
	Total	156	831	987	

Reference	**Persichetti 2018 [218]**				
					low suspicion
2×2 table	**AACE/ACE/AME**	Reference standard +	Reference standard −	Total	987 nodules were included in the analysis; n=39 patients with incomplete assessment and n=74 with Bethesda class I were excluded
	Index test +	154	653	807	
	Index test −	2	178	180	
	Total	156	831	987	High vs low-intermediate risk for AACE/ACE/AME: high/intermediate-risk vs low- risk
Statistical measures	<u>Index text Ultrasound (BTA)</u> Sensitivity : 90% Specificity: 63% <u>Index text Ultrasound (ATA)</u> Sensitivity : 93% Specificity: 52% <u>Index text Ultrasound (AACE/ACE/AME)</u> Sensitivity : 99% Specificity: 21%				
Source of funding	Not stated				
Limitations	Risk of bias: none Indirectness: none				
Comments	Diagnostic accuracy of BTA, ATA, AACE/ACE/AME US classification				

Reference	**Rahal 2016 [231]**
Study type	Retrospective
Study methodology	Data source: patients with thyroid nodules undergoing US scan of thyroid gland and neck area and US-guided FNA from November 2011 to February 2014 Recruitment: prospective; not specified.
Number of	n = 906 (1000 nodules)

Reference	Rahal 2016 [231]				
patients					
Patient characteristics	Age, mean (SD): not specified Gender (male to female ratio): not specified Ethnicity: not specified Setting: Hospital Israelita Albert Einstein Country: Brazil Inclusion criteria: thyroid nodules in patients who underwent sonographic evaluation, followed by fine needle aspiration. Exclusion criteria: not specified; nodules with a non-diagnostic or inadequate Bethesda classification were excluded from analysis				
Target condition(s)	Thyroid cancer				
Index test(s) and reference standard	Index test: Ultrasound (TI-RADS) US scan of thyroid gland and neck area was performed by experienced physicians, using the ATL HDI 5000, IU 22 Philips, Aplio 500 Platinum and My Lab 75 and the acquired images stored in the PACS System. Nodules were classified according to TI-RADS system as follows: 1 negative finding, 2 Benign, 3 probably benign, 4A low suspicion, 4B intermediate suspicion, 4C moderate suspicion, 5 High suspicion and 6 known proved malignancy. The US features associated to higher malignancy risks were irregular margins, hipoechogenicity, marked hypoechogenicity, morphology taller than wide and microcalcifications. Reference standard: US-guided FNA FNA was performed by freehand technique under US guidance, using a 23-gauge needle attached to a 20cc syringe. Experienced pathologists evaluated all samples according to Bethesda system: I non-diagnostic or inadequate, II benign, III atypia/follicular lesion of undetermined significance, IV follicular neoplasm or suspicious for follicular neoplasm, V suspicious of malignancy, VI malignant. Nodules classified as IV, V and VI were considered suspicious for malignancy. Time between measurement of index test and reference standard: not specified				
2×2 table		Reference standard +	Reference standard −	Total	There were 976 nodules included, 24 were classified as Bethesda I and excluded
	Index test +	114	274	388	
	Index test −	9	579	588	
	Total	123	853	976	

Reference	Rahal 2016 [231]
Statistical measures	Index text Ultrasound (TI-RADS) Sensitivity: 92.7% Specificity: 67.9% PPV: 29.4% NPV:98.5 %
Source of funding	Not specified
Limitations	Risk of bias: serious due to risk of bias in patient selection Indirectness: none
Comments	Diagnostic performance of TI-RADS.

Reference	Tae 2007 [276]
Study type	Prospective study
Study methodology	Data source: 1170 patients who underwent thyroid ultrasonography between January 2003 and January 2005 Recruitment: not specified
Number of patients	n = 580 (1255 nodules)
Patient characteristics	Age, mean (SD): 47.8 (13.9) Gender (male to female ratio): 77:503 Ethnicity: not specified Setting: St Mary's Hospital, Seoul, Republic of Korea Country: South Korea Inclusion criteria: patients who underwent thyroid ultrasonography at St Mary's Hospital, Seoul, Republic of Korea between January 2003 and January 2005;

Reference	Tae 2007 [276]
	Exclusion criteria: not specified; patients with unsatisfactory specimen (n=38) were excluded from analysis
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	Index test: Ultrasound All thyroid ultrasonography was performed by one radiologist, using an HDI 5000 ultrasound scanner equipped with a 5-12 MHz linear-array transducer. Nodules were classified based on the Kim criteria. If a single feature suggestive of malignancy was present, the nodule was classified as category 3, if the nodule showed no suspicious features it was classified as category 2. Anechoic, cystic lesions were classified as category 1. Category 3 was classified as malignant, categories 1, 2 as benign. Reference standard: FNAB or surgery (n=78 patients) FNAs were performed using 22-gauge needles. Palpable, single or dominant nodules >1cm nodules were aspirated by palpation (n=412). Aspiration was performed by sonographic guidance if the nodule was nonpalpable or cystic with a solid portion (n=168). The results of aspiration cytology were categorised as benign, suspicious of malignancy, malignant, and nondiagnostic. A cytology suspicious of follicular or Hurthle cell neoplasm or uncertain findings that could not rule out malignancy were included in a 'suspicious of malignancy' category. Time between measurement of index test and reference standard: not specified
2×2 table	See 2×2 table below
Statistical measures	Index text Ultrasound Sensitivity: 87% Specificity: 86.5% PPV: 48.4% NPV: 97.8% Accuracy: 86.5%
Source of funding	Not stated
Limitations	Risk of bias: serious due high risk of bias in the interpretation of the index test, reference standard results, flow and timing Indirectness: none

2×2 table

	Reference standard +	Reference standard −	Total
Index test +	60	64	124
Index test −	9	409	418
Total	69	473	542

Notes: 78 cases diagnosed as malignant by ultrasonography received surgical treatment.

Patients with unsatisfactory specimen (n=38) were excluded from analysis

Reference	Tae 2007 [276]
Comments	Diagnostic accuracy of US using the Kim criteria

Reference	Tang 2017 [281]
Study type	Prospective study
Study methodology	Data source: patients with thyroid nodules consenting to UGFNA Recruitment: consecutive patients meeting inclusion criteria from March 2015 to May 2016
Number of patients	n = 199 (206 nodules)
Patient characteristics	Age, mean (SD): not specified Gender (male to female ratio): 54:157 Ethnicity: not specified Setting: Department of Pathology and Laboratory Medicine, University of Cincinnati College of Medicine Country: USA Inclusion criteria: having a dominant or suspicious nodule seen on office US and been recommended for UGFNA. Exclusion criteria: patients with known thyroid malignancy or previous benign biopsy and patients who do not meet criteria for biopsy
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	Index test: Ultrasound Patients underwent an office US using a high resolution machine. Real-time US was performed by senior author, and nodules were stratified using sonographic patterns as described and published in the 2015 ATA guidelines. Nodules were classified into the best fit category of high, intermediate, low, very low suspicion or benign based on specific sonographic patterns.

Reference	Tang 2017 [281]
	Reference standard: UG-FNAB or surgery UGFNAB was performed by the same clinician using three to four separate passes with a 22-to 25-gauge needle utilizing capillary and aspiration techniques. Cytology results were reported based upon the Bethesda System for Reporting Thyroid Cytopathology i.e. benign, atypia of undetermined significance/ follicular lesion of undetermined significance, follicular neoplasm, suspicious for malignancy, malignant and non-diagnostic, reported by trained cytopathologists 64 patients with cytology deemed malignant, indeterminate or benign with large nodules underwent surgical excision with subsequent permanent final histological diagnosis. The index nodules undergoing US-FNA were assessed as benign or malignant. 65 nodules were surgically removed and used for analysis. Time between measurement of index test and reference standard: not specified
2×2 table	(see below)
Statistical measures	Index text: Ultrasound (ATA) Sensitivity : 91.7% Specificity: 74.1% PPV: 26.8% NPV: 98.9%
Source of funding	Not specified
Limitations	Risk of bias: none Indirectness: none
Comments	Diagnostic accuracy of US using the 2015 ATA guidelines

2×2 table:

	Reference standard +	Reference standard −	Total
Index test +	11	30	41
Index test −	1	86	87
Total	12	116	128

Reference	Weiss 2018 [309]
Study type	Retrospective chart review

Reference	Weiss 2018 [309]
Study methodology	Data source: patients with thyroid nodules consenting to UGFNA Recruitment: consecutive thyroid FNAs during an 18 month period (2016-2017)
Number of patients	n = 1157 (1491 nodules); US in n=57 (61 sub-nodules <1cm)
Patient characteristics	Age, mean (range): 52 (19-81) Gender (male to female ratio): 5:42 Ethnicity: not specified Setting: Department of Pathology, Microbiology and Immunology, Vanderbilt University School of Medicine Country: USA Inclusion criteria: sub-centimeter nodules identified by radiographic information. Ultrasound studies obtained before FNAs included. Exclusion criteria: not specified. Further population details: Patients with nodules <1 cm identified through radiographic information; biopsied because of: concomitant larger companion nodule (44%); personal history of cancer (19%); family history of cancer (9%) or suspicious sonogram that included calcification and/ or irregular contours (16%); unclear reason (14%). 40% of patients who had sub-centimeter nodules were under the care of an endocrine specialist.
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	Index test: Ultrasound (TI-RADS) 51 ultrasound studies were reviewed by a blinded board-certified radiologist subspecialising in thyroid ultrasonography, with 30 years' experience of performing thyroid ultrasound. High TI-RAD score included TR4 and TR5, intermediate TI-RAD score was TR3; low TI-RAD score included TR1 and TR2. Reference standard: FNAB All nodules were interpreted using the TBSRTC criteria.

<table>
<tr><th>Reference</th><th>Weiss 2018 [309]</th></tr>
<tr><td></td><td>Time between measurement of index test and reference standard: not specified</td></tr>
<tr><td>2×2 table</td><td><table><tr><td></td><td>Reference standard +</td><td>Reference standard −</td><td>Total</td></tr><tr><td>Index test +</td><td>5</td><td>9</td><td>14</td></tr><tr><td>Index test −</td><td>0</td><td>28</td><td>28</td></tr><tr><td>Total</td><td>5</td><td>37</td><td>42</td></tr></table>Radiographic information for 61 sub-centimeter nodules from 51 patients

(Ultrasound obtained before FNAB was available for 51 nodules)</td></tr>
<tr><td>Statistical measures</td><td><u>Index text Ultrasound</u>
Sensitivity : 100%
Specificity: 75.7%
PPV: 35.7%
NPV: 100%</td></tr>
<tr><td>Source of funding</td><td>Not specified</td></tr>
<tr><td>Limitations</td><td>Risk of bias: none
Indirectness: none</td></tr>
<tr><td>Comments</td><td>Risk of malignancy in sub-centimeter nodules using the ACR TI-RADS scoring system</td></tr>
</table>

Reference	Xu 2017 [323]
Study type	Retrospective (multicentre)
Study methodology	Data source: patient data collected from eight tertiary hospitals from January 6,2014 to December 20,2014 Recruitment: consecutive
Number of patients	n = 734 (962 nodules)
Patient characteristics	Age, mean (SD): 46.8 (14.09) Gender (male to female ratio): 156:578 Ethnicity: not specified

Reference	Xu 2017 [323]
	Setting: eight tertiary hospitals around Jiangsu province Country: China Inclusion criteria: patients who underwent thyroid surgery regardless of cytologic results, patients who underwent FNAB at least two times within a 1-year interval for benign thyroid lesions, patients who had benign results on cytology and showed no change or decreased size at follow-up US for at least a year. Exclusion criteria: Inadequate data of HRUS, FNAC or postoperative pathology; BSRTC I, III, IV; BSRTC II without repeated FNAC or follow-up US; BSRTC II with follow-up FNAC or US, but follow-up interval no more than one year.
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	Index test: Ultrasound (TI-RADS; 2015 ATA) All US images were obtained by using a 4-13 MHz linear array transducer. The scanning protocol in all cases included both transverse and longitudinal real-time imaging of the thyroid nodules. The features used in the analysis of thyroid nodules included size, composition, echogenicity of solid portion, orientation, shape, margin, and calcifications. All US patterns were diagnosed by a radiologist with 10 years of experience in thyroid imaging. 931 patterns were categorised based on the TI-RADS classification (2,3, 4A, 4B, 5) 906 patterns were categorised based on the ATA ultrasound patterns (benign, very low suspicion, low suspicion, intermediate suspicion, high suspicion. Reference standard: Histopathology (surgery, n=703)/ follow-up (n=259) Time between measurement of index test and reference standard: not specified

2×2 table	**TI-RADS**	Reference standard +	Reference standard −	Total	
	Index test +	301	156	363	Histopathological confirmation available for 703 nodules (375 malignant and 328 benign); 259 nodules regarded as benign due to repeated benign cytology or follow-up ultrasound after the first benign cytology 31 nodules could not be categorised by TI-RADS
	Index test −	62	412	568	
	Total	363	568	931	

Reference	**Xu 2017** [323]				
2×2 table	**ATA**	Reference standard +	Reference standard −	Total	Histopathological confirmation available for 703 nodules (375 malignant and 328 benign); 259 nodules regarded as benign due to repeated benign cytology or follow-up ultrasound after the first benign cytology 56 nodules could not be categorised by ATA
	Index test +	336	321	657	
	Index test −	23	226	249	
	Total	359	547	906	
Statistical measures	<u>Index text Ultrasound (TI-RADS)</u> Sensitivity: 83.2% Specificity: 71.5% AUC: 0.826 <u>Index text Ultrasound (2015 ATA)</u> Sensitivity: 94% Specificity: 41%				
Source of funding	Not specified				
Limitations	Risk of bias: none Indirectness: none				
Comments	Diagnostic performance of TI-RADS and 2015 ATA scoring systems based on nodule size.				

Reference	**Xu 2018** [324]
Study type	Retrospective
Study methodology	Data source: 3210 lesions that underwent thyroid US examination and FNA and/or surgery between January 2014 to October 2017 Recruitment: consecutive
Number of patients	n = 2031 (2465 nodules)
Patient characteristics	Age, mean (SD): 47.7 (13.38) Gender (male to female ratio): 415: 1616

Reference	Xu 2018 [324]
	Ethnicity: not specified Setting: three tertiary hospitals around JiangSu Province Country: China Inclusion criteria: nodules with definite histopathology results, nodules with complete Bethesda system for reporting thyroid cytopathology results (BSRTC) Exclusion criteria: nodules without postoperative pathology except for BSRTC II cytology results, nodules of BSRTC II cytology whose US follow-up interval less than one year or during which increase in size (defined as more than 50% change in volume or a 20% increase in at least two nodule dimensions with a minimal increase of 2mm in solid nodules or in the solid portion of mixed-cystic solid nodule) or change in US features
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	Index test: Ultrasound (TI-RADS;) All US images were obtained by using a 4-13 MHz linear array transducer. The scanning protocol in all cases included both transverse and longitudinal real-time imaging of the thyroid nodules. Designated radiologists from three centres were asked to assess the thyroid nodules using one set of standards according to published literature. The features used in the analysis of thyroid nodules included size, composition, echogenicity of solid portion, echotexture, vascularity, shape, margin and calcification. One specialist from each centre extracted US features based on static US patterns and description of features and then input these features into database. One experience radiologist in thyroid imaging did all classifications according to the database. All nodules were scored based on patterns and US features of KSThR-TIRADS as followed. Category 2 Benign, category 3 low suspicion, category 4 intermediate suspicion, category 5 high suspicion. All nodules were scored based on ACR-TI-RADS: TR1, TR2, TR3, TR4, TR5 All nodules were scored based on patterns and US-features of EU-TIRADS as follows: EU-TIRADS 2, 3, 4, 5 Reference standard: FNA and or surgery Among nodules, 505 benign nodules and 1005 malignant nodules were confirmed by histopathology; the remaining 955 benign lesions were diagnosed based on the benign cytology and follow-up ultrasound. Time between measurement of index test and reference standard: not specified

Reference	Xu 2018 [324]			
2×2 table	**KSThR- TI-RADS**	Reference standard +	Reference standard −	Total
	Index test +	966	671	1637
	Index test −	39	789	828
	Total	1005	1460	2465
2×2 table	**ACR-TIRADS**	Reference standard +	Reference standard −	Total
	Index test +	971	687	1658
	Index test −	34	773	807
	Total	1005	1460	2465
	EU-TIRADS	Reference standard +	Reference standard −	Total
	Index test +	986	810	1796
	Index test −	19	650	669
	Total	1005	1460	2465
Statistical measures	Index text Ultrasound (KSThR TI-RADS) Sensitivity: 96.1% Specificity: 54% Index text Ultrasound (ACR- TI-RADS) Sensitivity: 96.6% Specificity: 52.9% Index text Ultrasound (EU- TI-RADS) Sensitivity: 98.1% Specificity: 44.5%			
Source of funding	Not specified			
Limitations	Risk of bias: none Indirectness: none			
Comments	Diagnostic performance of TI-RADS			

Reference	**Yoon 2017** [329]
Study type	Retrospective
Study methodology	Data source: patient data collected from March 2007 to February 2010 Recruitment: not specified
Number of patients	n = 4585 (4696 nodules)
Patient characteristics	Age, mean (SD; range): 51 (11.9; 17-94) Gender (male to female ratio): 3836:749 Ethnicity: not specified Setting: tertiary referral centre Country: Korea Inclusion criteria: patients who underwent US-guided FNA for diagnosis of thyroid nodules at a tertiary referral centre from March 2007 to February 2010. Exclusion criteria: lack of follow-up after results of initial nondiagnostic results, atypia or follicular lesion of undetermined significance, follicular neoplasm or suspicion of follicular neoplasm, or suspicion of malignancy
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound</u> Real-time US examinations of both thyroid glands and the cervical regions were performed by using a 6-13-MHz or 5-12-MHz linear transducer. Examinations were performed by one of 14 radiologists (four faculty and 10 fellows) with 1-12 years of experience in thyroid imaging. US features of the thyroid nodules that underwent US-guided FNA were prospectively re-corded by each radiologist who had performed the US and /or US-guided FNA according to composition, echogenicity, margin, calcifications and shape. Solid, hypoechogenicity or marked hypoechogenicity, microlobulated or irregular margins, presence of microcalcifications and nonparallel shape were considered to be US features suspicious for malignancy. <u>Reference standard: Histopathology: surgery(n=1072), initial UGFNA (n=3443), repeat UGFNA (n=181)</u> US-guided FNA was performed on nodules that showed US features that were suspicious of malignancy or on the largest nodule when none of the multiple thyroid nodules manifested with US features suspicious for malignancy. The decision to perform FNA was at the discretion of the interpreting radiologist who used the aforementioned criteria. Examinations were performed by one of 14 radiologists

Reference	Yoon 2017 [329]			
	(four faculty and 10 fellows) with 1-12 years of experience in thyroid imaging, by using a 23-gauge needle attached to a 2-mL disposable syringe either by using an aspirator or the freehand technique, depending on the performer's preference. Each nodule was aspirated at least twice and local anaesthesia was not routinely applied. Aspirated material was expelled on glass slides, which were immediately placed in 95% alcohol for Papanicolaou staining. One of eight cytopathologists reviewed the slides. Until 2009 cytology reports were categorised into: inadequate, benign, intermediate suspected of papillary carcinoma and malignant; From December 2009 onwards the 6 categories of the Bethesda System have been used to report results from thyroid cytologic analysis. Total or near-total thyroidectomy was performed in patients over the age of 45 years, who had multiple tumors, with the presence of extrathyroidal extension or lymph node (LN) metastasis on either pre-or intraoperative findings. Hemithyroidectomy was performed in patients under the age of 45 years, without finding of multiple tumors, extrathyroidal extension, or LN metastasis. Time between measurement of index test and reference standard: not specified			
2×2 table	**SRU**	Reference standard +	Reference standard −	Total
	Index test +	564	921	1485
	Index test −	480	2731	3211
	Total	1044	3652	4696
2×2 table	**NCCN**	Reference standard +	Reference standard −	Total
	Index test +	973	2200	3173
	Index test −	71	1452	1523
	Total	1044	3652	4696
2×2 table	**ATA**	Reference standard +	Reference standard −	Total
	Index test +	999	2165	3164
	Index test −	45	1487	1532
	Total	1044	3652	4696
2×2 table	**F-TIRADS**	Reference standard +	Reference standard −	Total
	Index test +	994	1754	2748
	Index test −	50	1898	1948
	Total	1044	3652	4696
2×2 table	**Kim**	Reference standard +	Reference standard −	Total
	Index test +	908	616	1524
	Index test −	136	3036	3172

Reference	**Yoon 2017** [329]			
	Total	1044	3652	4696
2×2 table	**K-TIRADS**	Reference standard +	Reference standard −	Total
	Index test +	1031	2719	3750
	Index test −	13	933	946
	Total	1044	3652	4696
Statistical measures	<u>Index text Ultrasound (SRU)</u> Sensitivity: 54% Specificity: 74.8% PPV: 38% NPV: 85.1% Accuracy: 70.2% <u>Index text Ultrasound (NCCN)</u> Sensitivity: 93.2% Specificity: 39.8% PPV: 30.7% NPV: 95.3% Accuracy: 51.6% <u>Index text Ultrasound (ATA)</u> Sensitivity: 95.7% Specificity: 40.7% PPV: 31.6% NPV: 97.1% Accuracy: 52.9% <u>Index text Ultrasound (F-TIRADS)</u> Sensitivity: 95.2% Specificity: 52% PPV: 36.2% NPV: 97.4% Accuracy: 61.6% <u>Index text Ultrasound (Kim)</u>			

Reference	Yoon 2017 [329]
	Sensitivity: 87.0% Specificity: 83.1% PPV: 59.6% NPV: 95.7% Accuracy: 84% <u>Index text Ultrasound (K-TIRADS)</u> Sensitivity: 98.8% Specificity: 25.6% PPV: 27.5% NPV: 98.6% Accuracy: 41.8%
Source of funding	Not specified
Limitations	Risk of bias: serious risk due to potential bias in the interpretation of the index test results; flow and timing Indirectness: none
Comments	Diagnostic performance of SRU, NCCN, 2015 ATA, F- TI-RADS, Kim, K-TIRADS

Reference	Yoon 2016 [331]
Study type	Retrospective
Study methodology	Data source: patient data collected from November 2013 to July 2014 at a tertiary referral centre Recruitment: not specified
Number of patients	n = 1241 (1293 nodules)
Patient characteristics	Age, mean (SD; range): 50.8 (13.5; 18-87) Gender (male to female ratio): 257:1036 Ethnicity: not specified Setting: tertiary referral centre

Reference	Yoon 2016 [331]
	Country: Korea Inclusion criteria: nodules of patients who underwent US-guided FNA for diagnostic purposes at a tertiary referral centre from November 2013 to July 2014; nodules were included if they had: undergone surgery, definitive diagnostic cytologic findings of benignity or malignancy at US-guided FNA, or inconclusive cytologic findings at initial US-guided FNA but definitive cytologic findings of benignity or malignancy at follow-up US-guided FNA. Exclusion criteria: aspiration of cysts for symptom relief of typically benign thyroid cysts or for diagnosis of perithyroidal lesions such as parathyroid cysts, thyroglossal duct cysts, or other cystic masses arising in the cervical region (n=21); maximal diameter less than 10 mm (n=913); non-mass forming lesions (n=6); and inadequate follow-up (n=353) because nodules were lost to follow-up after inconclusive diagnostic cytologic findings.
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound</u> US examinations were performed by using a 6-13-MHz linear array transducer. Real-time US and subsequent US-guided FNA were performed by one of 10 radiologists (four faculty and six fellows) with 1-15 years of experience in thyroid imaging. US features of each thyroid nodule that were described and re-corded by one of the 10 radiologists who performed the examinations according to composition, echogenicity, margin, calcifications and shape. Marked hypoechogenicity, non-circumscribed margins, microcalcifications or mixed calcifications and nonparallel shape were considered to be US features suspicious for malignancy on the basis of published criteria. Nodules were retrospectively classified according to the 2014 ATA guidelines, by one radiologist with 7 years of experience in thyroid imaging, as showing high, intermediate, low, or very low suspicion of malignancy. For TIRADS, nodules were classified on the basis of the number of suspicious US features present as follows: solidity, hypoechogenicity or marked hypoechogenicity, microlobulated to irregular margin, microcalcifications or mixed calcifications, and nonparallel shape. Thyroid nodules without any suspicious features were classified as TIRADS category 3. Nodules showing one, two, three or four, or five suspicious US features were classified as category 4a, 4b, 4c, or 5 respectively. <u>Reference standard: UGFNA (n=1051) and surgery (n=242)</u> Ultrasound guided fine needle aspiration was performed in nodules measuring more than 5 mm in maximum diameter, nodules with at least one suspicious US feature, or the largest mass when none of the multiple thyroid nodule detected at US showed any suspicious US features. UGFNA was performed at least twice for each thyroid nodule using a 23-gauge needle attached to a 2-mLsyringe without an aspirator. Local anaesthesia was not routinely applied. Aspirated material was expelled on glass slides, which were immediately placed in 95% alcohol for Papanicolaou staining. Cytopathologists were not present during procedures. One of five cytopathologists interpreted the slides and cytology reports were based on the 6 categories of the Bethesda System for Reporting Thyroid Cytopathology. Time between measurement of index test and reference standard: not specified

Reference	**Yoon 2016** [331]				
2×2 table	**ATA (2014)**	Reference standard +	Reference standard −	Total	Very-low suspicion nodules were considered negative and low-to-high suspicion as positive. 44 of the 1293 nodules did not meet the criteria for any pattern and were classified as not specified. 242 nodules (18.7%) underwent surgery and 1051 (81.3%) was diagnosed on the basis of cytologic findings and follow-up US.
	Index test +	223	663	886	
	Index test −	11	396	407	
	Total	234	1059	1293	
2×2 table	**TIRADS**	Reference standard +	Reference standard −	Total	Category 3 was considered as negative and categories 4a to 5 as positive. 242 nodules (18.7%) underwent surgery and 1051 (81.3%) was diagnosed on the basis of cytologic findings and follow-up US.
	Index test +	228	749	977	
	Index test −	6	310	316	
	Total	234	1059	1293	
Statistical measures	<u>Index text Ultrasound (ATA)</u> Sensitivity: 95.3% Specificity: 37.4% PPV: 25.2% NPV: 97.3% Accuracy: 47.9% <u>Index text Ultrasound (TIRADS)</u> Sensitivity: 97.4% Specificity: 29.3% PPV: 23.3% NPV: 98.1% Accuracy: 41.6%				
Source of funding	Not specified				
Limitations	Risk of bias: serious risk of bias due to potential bias in the interpretation of the index test results; flow and timing Indirectness: none				
Comments	Diagnostic performance of 2014 ATA and TI-RADS				

Reference	**Yoon 2015** [332]

Reference	**Yoon 2015 [332]**
Study type	Retrospective
Study methodology	Data source: data of patients undergoing US-FNA at tertiary referral centre collected from December 2010 to July 2011 Recruitment: not specified
Number of patients	n = 1257 (1309 nodules)
Patient characteristics	Age, mean (SD; range): 50.1 (12.1; 18-83) Gender (male to female ratio): 192: 1065 Ethnicity: not specified Setting: tertiary referral centre Country: South Korea Inclusion criteria: thyroid nodules with diagnosis confirmed by surgery after inadequate AUS/FLUS, FN or suspicion of malignancy results on cytology, or nodules definitively diagnosed as benign or malignant nodules on US-FNA cytology. Exclusion criteria: nodules with inadequate cytology that had not been followed with either US-FNA or US examinations (n=227) and nodules diagnosed as atypia of undetermined significance/follicular lesions of undetermined significance (n=84), follicular neoplasm (n=9) or suspicious for malignancy (n=19) on cytology that had not been followed by US-FNA or surgery.
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound</u> US examinations were performed by using a 5-12-MHz linear array transducer. Real-time US was performed by one of 12 radiologists (four faculty and eight fellows) with 1-15 years of experience in thyroid imaging. US features of each thyroid nodule that were prospectively re-corded by one of the radiologists who performed the US examinations and subsequent US-FNA. Each nodule was described according to tumour composition, echogenicity, margin, calcifications and shape. Marked hypoechogenicity, non-circumscribed margins, microcalcifications or mixed calcifications and nonparallel shape were considered to be malignant features based on the Kim criteria. The final assessment was ‘probably benign’ when none of the aforementioned suspicious US features were present and ‘suspicious malignant’ when one or more of the malignant features was present in a thyroid nodule. <u>Index test: Ultrasound+ vascularity pattern</u> Vascularity was evaluated on 2-D Doppler US images acquired during US examinations. The same US scanner setting and the same 2-D

Reference	**Yoon 2015 [332]**				
	power Doppler colour map were used throughout the study to minimise the effect of machine settings on data acquisition. Vascularity was classified into three patterns: no vascularity, peripheral vascularity, intra-nodular vascularity <u>Reference standard: UGFNA (n=962) or surgery (n=347)</u> Ultrasound guided fine needle aspiration was performed on nodules with suspicious US features or on the largest mass when none of the multiple thyroid nodule detected had suspicious US features. UGFNA was performed with a freehand technique by the same radiologist who had performed the US examinations; 23-gauge needle attached to a 2-mL disposable plastic syringe without an aspirator were used. Each nodule was aspirated at least twice. Samples obtained were expelled on glass slides, which were smeared and immediately placed in 95% alcohol for Papanicolaou staining. Cytopathologists were not present during the US-FNA procedure. One of five experienced cytopathologists reviewed the slides and cytology reports were based on the 6 categories of the Bethesda System for Reporting Thyroid Cytopathology. Time between measurement of index test and reference standard: not specified				
2×2 table	**Kim**	Reference standard +	Reference standard −	Total	Surgery (n=347): benign (n=19), malignant (n=328) FNA (n=962): benign (n=910), malignant (n=52)
	Index test +	340	238	578	
	Index test −	42	689	731	
	Total	382	927	1309	
2×2 table	**Kim+USD**	Reference standard +	Reference standard −	Total	Surgery (n=347): benign (n=19), malignant (n=328) FNA (n=962): benign (n=910), malignant (n=52)
	Index test +	349	351	700	
	Index test −	33	576	609	
	Total	382	927	1309	
Statistical measures	<u>Index text Ultrasound (Kim)</u> Sensitivity: 89% Specificity: 74.3% PPV: 58.8% NPV: 94.3% AUC: 0.821% <u>Index text Ultrasound (Kim + USD)</u> Sensitivity: 91.4% Specificity: 62.1% PPV: 49.9% NPV: 94.6% Accuracy: 0.766%				

Reference	**Yoon 2015** [332]
Source of funding	Not specified
Limitations	Risk of bias: high due to potential risk of bias in the interpretation of the index test and reference standard results; flow and timing Indirectness: none
Comments	Diagnostic performance of US using the Kim criteria and Kim +USD

Reference	**Zhang 2018** [353]
Study type	Prospective
Study methodology	Data source: patients with thyroid nodules more than 1cm in diameter from July 2011 to October 2017 Recruitment: not specified.
Number of patients	n = 162 (243 nodules)
Patient characteristics	Age, mean (range): 54.7 (21-79) Gender (male to female ratio): 41: 121 Ethnicity: not specified Setting: Nanjing integrated traditional Chinese and western medicine hospital, Nanjing University of Chinese medicine Country: China Inclusion criteria: Patients with thyroid nodules more than 1cm in largest diameter, patients agreed to surgery if FNAB results are malignant, suspicious for malignancy and indeterminate follicular lesions, patients agreed to initial US-guided FNAB and US follow-up (>12 months after US-guided FNAB) for benign thyroid lesions (except for adenomas); and patients agreed to US-guided FNAB for benign thyroid lesions at least twice within one-year interval. Exclusion criteria: not specified
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound (TI-RADS)</u> Us evaluation was performed by clinically experienced radiologist with 18 years of thyroid US experience or by residents and fellows under his supervision .

Reference	**Zhang 2018 [353]**
	US findings were classified according to TIRADS system as described by Russ et al. into the following categories: 1= normal thyroid tissue without any nodular aspect; 2=simple cyst, spongiform nodules, 'white knight', isolated macrocalcification, nodular hyperplasia; 3= no signs of high suspicion, isoechoic or hyperechoic, partial in capsulated; 4a= no signs of high suspicion, mildly isoechoic, encapsulated nodule; 4B= irregular shape, taller than wide, irregular borders, microcalcifications, markedly hypoechoic, high stiffness with elastography, 1 or 2 signs and no lymph node metastasis; 5= irregular shape, taller than wide, irregular borders, microcalcifications, markedly hypoechoic, high stiffness with elastography: strain ratio >4, 3 to 5 signs and/or lymph node metastasis. TIRADS categories were interpreted as follows: category 1=normal thyroid findings; category 2= constantly benign aspects, category 3=very probably benign, category 4A= undetermined, 4B=suspicious and 5= highly suspicious. <u>Reference standard: pathological examination or FNAB</u> US-guided FNAB was performed by LQ.H, with 16 years of pathological diagnosis experience, routinely using a 23- gauge needle. A 21-gauge needle was chosen when a nodule had a large cystic portion and for second-needle passage when the first FNA failed due to severe nodule stiffness. Direct smears were made, immediately fixed with alcohol after FNA and stained with Papanicolaou stain. The adequacy of the specimens was assessed using visual inspection, classified into two groups: insufficient (fewer than six particles) or sufficient (more than 6 visible particles). Additional FNA procedures were performed when the lesion was considered inaccurately targeted in the case of small nodules or when an insufficient specimen was suspected by visual inspection. US-guided FNAB were performed at the hospital. Pathology results were obtained after surgery if FNAB results were malignant, suspicious for malignancy and indeterminate follicular lesions. For malignant nodules the pathological diagnosis was confirmed by surgery. A final diagnosis of benign nodule was made when one of the following parameters were met: repeated FNA confirmed at least twice; surgical specimen; and benign cytology findings on the FNA in confirmed with a stable size or reduced size during follow-up US (>12 months). If the nodule was surgically resected, the FNAB diagnosis was then compared to the surgical pathology diagnosis to evaluate concordance. Time between measurement of index test and reference standard: not specified

2×2 table		Reference standard +	Reference standard −	Total	Resection was performed on 82 nodules
	Index test +	64	66	130	
	Index test −	3	110	113	
	Total	67	176	243	

Reference	**Zhang 2018** [353]
Statistical measures	Index text Ultrasound (F-TIRADS) Sensitivity: 92.5% Specificity: 68.2% PPV: 52.5% NPV: 96% Accuracy:74.9%
Source of funding	Not specified
Limitations	Risk of bias: very serious due to high risk of bias in patient selection; index test, reference standard, flow and timing Indirectness: none
Comments	Diagnostic performance of F- TIRADS.

Reference	**Zhang 2017** [354]
Study type	Retrospective
Study methodology	Data source: patients with thyroid nodules who had received conventional US and CEUS examinations between December 2012 and December 2014 Recruitment: retrospective; not specified.
Number of patients	n=246 (319 nodules)
Patient characteristics	Age, mean (SD; range): 46.1 (15.2; 19-74) Gender (male to female ratio): 85: 161 Ethnicity: not specified Setting: Department of Ultrasound, The Third Xiangya Hospital, Central South University Country: China Inclusion criteria: Patients who had received conventional US and CEUS examinations and postoperative pathological diagnoses or FNABs between December 2012 and December 2014

Reference	Zhang 2017 [354]
	Exclusion criteria: not specified
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound (TI-RADS; TI-RADS+CEUS)</u> Diagnosis was performed with a S2000 colour Doppler US system equipped with an 14L5 transducer for conventional US and equipped with an 9 L4 transducer for conventional transducer for CEUS. Every section of the thyroid was scanned. TI-RADS were used to evaluate and classify every nodule. The CPS technique and SonoVue contrast agent were used. A 20-G needle was inserted into the patients' peripheral veins to establish intravenous access. All examinations were performed by an experienced radiologist with more than 10 years' experience in US diagnosis and more than 1 years' experience of performing CEUS of thyroid nodules. US imaging data were analysed by two other experienced radiologists who performed blind independent analyses of the TI-RADS and CEUS images to retrospectively analyse the nature of the thyroid nodules. The 4a, 4b thyroid nodules which were categorised by TI-RADS and a combination of TI-RADS and CEUS were studied retrospectively. TI-RADS classification: score 1: normal thyroid; score 2: no malignant sign, benign lesions; score 3: one malignant sign, high probability of benignity; score 4a: two malignant signs, possible benignity; score 4b: three malignant signs, high probability of malignancy; score 5: four to five malignant signs, highly suggestive of malignancy. Scores 1-4a diagnosed as benign; scores 4b-5 diagnosed as malignant. CEUS classification: circular enhancement; high enhancement; equal enhancement; low enhancement. High, circular or equal enhancement diagnosed as benign; low enhancement diagnosed as malignant. <u>Reference standard: FNAB (230 nodules) or surgery (89 nodules):</u> Time between measurement of index test and reference standard: not specified

2x2	K TIRADS	Reference standard +	Reference standard −	Total
	Index test +	75	176	251
	Index test −	0	68	68
	Total	75	244	319

2×2 table	ZhangTI-RADS	Reference standard +	Reference standard −	Total
	Index test +	65	21	86
	Index test −	10	223	233
	Total	75	244	319

Reference	Zhang 2017 [354]			
2x2 table	TIRADS+CEUS	Reference standard +	Reference standard −	Total
	Index test +	73	10	83
	Index test −	2	234	236
	Total	75	244	319
Statistical measures	Index text Ultrasound (K TI-RADS) Sensitivity: 96% Specificity: 67.6% PPV: 47.7% NPV: 98.2% Accuracy:73.8% Index text Ultrasound (Zhang TI-RADS) Sensitivity: 86.7% Specificity: 91.4% PPV: 75.6% NPV: 95.7% Accuracy:90.3% Index text Ultrasound (TI-RADS+CEUS) Sensitivity: 97.3% Specificity: 95.5% PPV: 88% NPV: 99.1% Accuracy:96.0%			
Source of funding	No funding			
Limitations	Risk of bias: serious risk due to patient selection, index test Indirectness: none			
Comments	Diagnostic performance of TI-RADS; TI-RADS+ CEUS.			

Reference	Zhang 2015 [346]

Reference	**Zhang 2015** [346]
Study type	Prospective
Study methodology	Data source: patients with thyroid nodules from October 2011 to June 2013 Recruitment: prospective; not specified.
Number of patients	n = 2921 (3980 nodules)
Patient characteristics	Age, mean (SD): 51.6 (11.6, 16-78) Gender (male to female ratio): 951: 1970 Ethnicity: not specified Setting: not specified/ Department of Medical Ultrasound, Shanghai Tenth People's Hospital? Country: China Inclusion criteria: Patients with thyroid nodules Exclusion criteria: loss at follow-up, less than 12 month follow-up for benign nodules, no cytology/pathology results with TI-RADS category 4 and 5, increase in size on follow-up US without further cytopathological evaluation. Nodule diameter ranged from 2.0 mm to 70.0 mm; mean (SD) 15.7 (11) mm.
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	<u>Index test: Ultrasound (TI-RADS)</u> Us scanning was performed with an S2000 US system, using a 4-9 MHz linear-array transducer and a Logiq E9 US system using a 9-15 MHz linear-array transducer. All the image analysis and TI-RADS classification was performed by two board-certified investigators with consensus who were blind to the final results. TI-RADS category 2 and 3 were regarded as 'test negative'; TI-RADS category 4 and 5 as 'test positive'. Therefore, benign lesions classified as 2 and 3 were regarded as true negative and non-benign lesions classified as 4 or 5 as true positive <u>Reference standard: pathological examination or FNA cytology</u>

Reference	**Zhang 2015** [346]
	UGFNA was performed under sterile conditions. Three to four passes were made for each nodule using a 23-gauge needle. On-site accuracy was not performed in this study. Samples were submitted for cytology. Time between measurement of index test and reference standard: not specified

2×2 table		Reference standard +	Reference standard −	Total	
	Index test +	222	339	561	Final diagnosis based on FNA (n-628 nodules); Surgery (partial or total thyroidectomy) performed in all nodules with benign or suspicious cytology and 55 nodules with inconclusive cytology and 10 benign nodules. Remaining 737 nodules underwent surgery without FNA. 971 had pathological results.
	Index test −	6	3413	3419	
	Total	228	3752	3980	

Statistical measures	<u>Index text Ultrasound (TI-RADS)</u> Sensitivity: 97% Specificity: 90% PPV: 40% NPV: 99% Accuracy:91%
Source of funding	Shanghai Hospital Development Centre; Shanghai Human Resource and Social security Bureau; National Natural Science foundation of China
Limitations	Risk of bias: serious risk due to flow and timing. Indirectness: serious due to indirect reference standard for some cases
Comments	Diagnostic performance of Kwak's TI-RADS.

Reference	**Zheng 2018** [363]
Study type	Retrospective
Study methodology	Data source: patients who had undergone sonography and had thyroid surgery or FNA from January 2015 to December 2016 Recruitment: retrospective; not specified.
Number of patients	n = 1163 (21189 nodules)
Patient	Age, mean (SD; range): 45.3 (13; 15-81)

Reference	Zheng 2018 [363]				
characteristics	Gender (male to female ratio): 308: 725 Ethnicity: not specified Setting: Department of Ultrasound, Rui Jin Hospital, School of Medicine, Shanghai Jiao Tong University Country: China Inclusion criteria: Patients who had undergone sonography and had thyroid surgery or FNA at Rui Jin Hospital from January 2015 to December 2016 Exclusion criteria: nodules without final histopathological or cytological results (n=27); with final histopathological or cytological results but not refer to the suspicious lesions in US (n=16); with typically benign US features (n=6); inadequate sonographic data acquisition (n=107)				
Target condition(s)	Thyroid cancer				
Index test(s) and reference standard	<u>Index test: Ultrasound (TI-RADS)</u> Conventional Us was performed with a 5 to 12 MHz linear array transducer. During scanning, patients lay on the bed in supine position with slight flexion of the head to fully expose the front area of the neck. Images of nodules were acquired by carefully scanning the thyroid and adjacent tissues both transversely and longitudinally. If multiple nodules were present, every suspicious one would be focused on. Ultrasound examination and image acquisition are performed by radiologists with more than 5 years of experience. Two reviewers with more than 5 years of experience in thyroid US independently performed retrospective analysis of ultrasonic images of the surgical nodules without knowing pathological or cytological results and other clinical information. Discordance was solved by another reviewer with more than 10 years of experience in thyroid US. When assessing nodules, 2 reviewers selected 1 feature from the first 4 categories of the e ACR TI-RADS: composition, echogenicity, shape, margins and all the features that apply from echogenic foci category. The sum of the points determined by TI-RADS level, with TR1 indicating 0 points, TR2, 2 points; TR3, 3 points; TR4 4 to 6 points and TR5, 7 or more points. <u>Reference standard: surgery (n=527) or FNA (n=506)</u> Time between measurement of index test and reference standard: not specified				
2×2 table		Reference standard +	Reference standard −	Total	Nodules with typically benign US features and

Reference	**Zheng 2018** [363]				
	Index test +	307	410	715	without any suspicious features were excluded
	Index test −	1	315	318	
	Total	308	725	1033	
Statistical measures	Index text Ultrasound (ACR-TIRADS) Sensitivity: 99% Specificity: 43.4% PPV: 42.7% NPV: 99.1% Accuracy:60%				
Source of funding	Not specified				
Limitations	Risk of bias: serious risk due to high risk of bias in the conduct of the reference standard; flow and timing Indirectness: none				
Comments	Diagnostic performance of ACR TIRADS.				

Reference	**Zhou 2018** [367]
Study type	Prospective
Study methodology	Data source: patients with thyroid nodules from July to September 2016 Recruitment: not specified.
Number of patients	n = 161 (167 nodules)
Patient characteristics	Age, mean (SD): 44.14 (12.01) Gender (male to female ratio): 43: 118 Ethnicity: not specified Setting: Department of Ultrasound of the Third Xiangya Hospital Country: China

Reference	Zhou 2018 [367]
	Inclusion criteria: Patients with solid or mainly solid thyroid nodules, with at least 1 of the suspicious features (solid component, hypoechogenicity or marked hypoechogenicity, irregular margins, microcalcifications, and a taller than wide shape) on US imaging. Exclusion criteria: dominantly cystic nodules, pregnancy, suspicious thyroid nodules that were eggshell calcified.
Target condition(s)	Thyroid cancer
Index test(s) and reference standard	Index test: Ultrasound (TI-RADS) All Us examinations were performed with commercially available scanners equipped with L12-3E transducer for both conventional and contrast-enhanced US with ultra-wideband nonlinear contrast imaging. The following information was gathered with conventional US: echo (hypo, iso, or hyper level), composition (solitary or mixed), taller-than- wide shape, nodule margin and calcifications. After conventional US, the transducer was switched to the contrast-enhanced US mode. Images were quantitative analysed with Contrast Imaging QA software. The Region of Interest (ROI) was set in the most evident enhanced region and the same ROI area was copied in perinodule thyroid tissues and served as a control. A time-intensity curve and all of the quantitative parameters were generated to show the contrast-enhanced US performance as follows: 1. Peak intensity; 2.ascend slope-compared to perinodule tissue; 3.descent slope- compared to perinodule tissue ; 4.time to peak- compared to perinodule tissue; 5. Time from peak to one-half- compared to perinodule tissue; 6. AUC - compared to perinodule tissue. The ratios of nodule and perinodule values were adopted to evaluate the thyroid nodules. Ultrasound examinations was performed by a single experienced examiner, and the quantitative analysis of contrast-enhanced US was performed by trained sonographers, blinded to clinical data and other imaging findings. Nodules were classified based on the TI-RADS classification as levels: 3, 4a, 4b and 5. TI-RADS 4a nodules were supposed to be benign. Reference standard: FNA or surgery A cytological analysis was done based on the Bethesda classification system. Time between measurement of index test and reference standard: not specified

2×2 table	**TI-RADS**	Reference standard +	Reference standard −	Total	Results for 161 patients with solid thyroid nodules.
	Index test +	91	15	106	
	Index test −	2	53	55	
	Total	93	68	161	
2×2 table	**TI-RADS+contrast-enhanced US parameter ratios**	Reference standard +	Reference standard −	Total	Results for 161 patients with solid thyroid nodules.

Reference	Zhou 2018 [367]				
	Index test +	91	15	106	
	Index test −	2	53	55	
	Total	93	68	161	
Statistical measures	<u>Index text Ultrasound (TI-RADS)</u> Sensitivity: 98% Specificity: 78% <u>Index text Ultrasound (TI-RADS+ contrast-enhanced US parameter ratios)</u> Sensitivity: 98% Specificity: 78%				
Source of funding	Not stated				
Limitations	Risk of bias: serious due to risk of bias in reference bias and flow and timing Indirectness: none				
Comments	Diagnostic performance of TI-RADS and TI-RADS+ contrast-enhanced US parameter ratios				

Appendix E: Coupled sensitivity and specificity forest plots and sROC curves

E.1 Coupled sensitivity and specificity forest plots

Figure 2: BTA

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Farihah 2018	12	60	0	32	1.00 [0.74, 1.00]	0.35 [0.25, 0.45]
Persichetti 2018	141	304	15	527	0.90 [0.85, 0.95]	0.63 [0.60, 0.67]

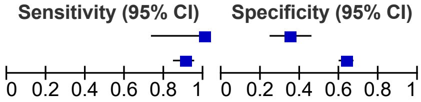


Figure 3: Kim

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ahn 2010	303	205	24	866	0.93 [0.89, 0.95]	0.81 [0.78, 0.83]
Alam 2014	22	16	2	60	0.92 [0.73, 0.99]	0.79 [0.68, 0.87]
Kim 2002	46	36	3	70	0.94 [0.83, 0.99]	0.66 [0.56, 0.75]
Kim 2013	147	354	0	127	1.00 [0.98, 1.00]	0.26 [0.23, 0.31]
Koh 2018	158	303	54	965	0.75 [0.68, 0.80]	0.76 [0.74, 0.78]
Moon 2010	227	115	42	699	0.84 [0.79, 0.89]	0.86 [0.83, 0.88]
Moon 2012	199	162	18	324	0.92 [0.87, 0.95]	0.67 [0.62, 0.71]
Tae 2007	60	64	9	9	0.87 [0.77, 0.94]	0.12 [0.06, 0.22]
Yoon 2015	340	238	42	689	0.89 [0.85, 0.92]	0.74 [0.71, 0.77]
Yoon 2017	908	616	136	3036	0.87 [0.85, 0.89]	0.83 [0.82, 0.84]

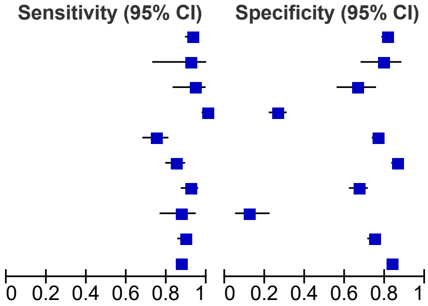


Figure 4: Modified Kim

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Kim 2013	142	121	6	676	0.96 [0.91, 0.98]	0.85 [0.82, 0.87]

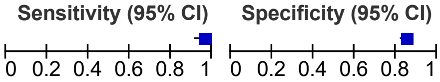


Figure 5: Kim + Doppler

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Moon 2010	245	387	24	427	0.91 [0.87, 0.94]	0.52 [0.49, 0.56]
Yoon 2015	349	351	33	576	0.91 [0.88, 0.94]	0.62 [0.59, 0.65]

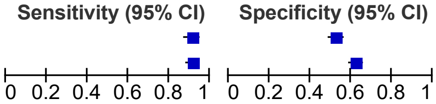


Figure 6: Kim + USE (Rago)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Moon 2012	200	170	17	316	0.92 [0.88, 0.95]	0.65 [0.61, 0.69]

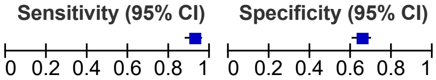


Figure 7: Kim + USE (Asteria)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Moon 2012	205	255	12	231	0.94 [0.91, 0.97]	0.48 [0.43, 0.52]

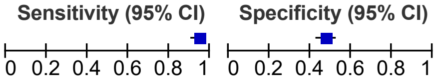


Figure 8: SRU

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ahn 2010	116	489	211	582	0.35 [0.30, 0.41]	0.54 [0.51, 0.57]
Hobbs 2014	24	250	5	81	0.83 [0.64, 0.94]	0.24 [0.20, 0.29]
Yoon 2017	564	921	480	2731	0.54 [0.51, 0.57]	0.75 [0.73, 0.76]

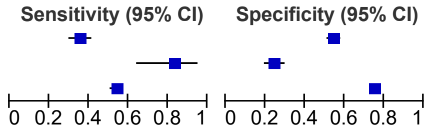


Figure 9: AACE/ACE/AME

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ahn 2010	259	98	68	973	0.79 [0.74, 0.83]	0.91 [0.89, 0.93]
Grani 2018	31	296	5	170	0.86 [0.71, 0.95]	0.36 [0.32, 0.41]
Lauria Pantano 2018	109	786	3	151	0.97 [0.92, 0.99]	0.16 [0.14, 0.19]
Moon 2010	220	168	49	646	0.82 [0.77, 0.86]	0.79 [0.76, 0.82]
Persichetti 2018	154	653	2	178	0.99 [0.95, 1.00]	0.21 [0.19, 0.24]

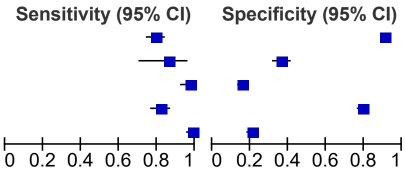


Figure 10: ATA

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Grani 2018	27	255	9	211	0.75 [0.58, 0.88]	0.45 [0.41, 0.50]
Ha 2018	247	202	12	372	0.95 [0.92, 0.98]	0.65 [0.61, 0.69]
Hoang 2018	13	70	2	15	0.87 [0.60, 0.98]	0.18 [0.10, 0.27]
Koh 2018	197	999	15	269	0.93 [0.89, 0.96]	0.21 [0.19, 0.24]
Lauria Pantano 2018	87	525	17	394	0.84 [0.75, 0.90]	0.43 [0.40, 0.46]
Macedo 2018	5	33	0	100	1.00 [0.48, 1.00]	0.75 [0.67, 0.82]
Maino 2018	11	64	3	272	0.79 [0.49, 0.95]	0.81 [0.76, 0.85]
Pandya 2018	85	546	13	706	0.87 [0.78, 0.93]	0.56 [0.54, 0.59]
Persichetti 2018	145	399	11	432	0.93 [0.88, 0.96]	0.52 [0.49, 0.55]
Tang 2017	11	30	1	86	0.92 [0.62, 1.00]	0.74 [0.65, 0.82]
Xu 2017	336	321	23	226	0.94 [0.91, 0.96]	0.41 [0.37, 0.46]
Yoon 2016	223	663	11	396	0.95 [0.92, 0.98]	0.37 [0.34, 0.40]
Yoon 2017	999	2165	45	1487	0.96 [0.94, 0.97]	0.41 [0.39, 0.42]

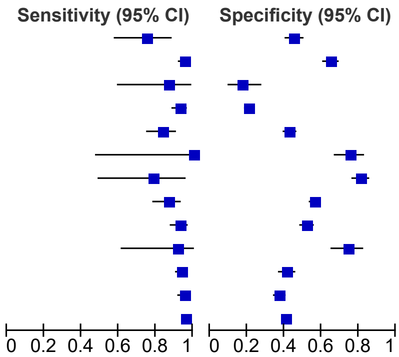


Figure 11: ATA (subcentimetre)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Kim 2013	286	306	10	111	0.97 [0.94, 0.98]	0.27 [0.22, 0.31]

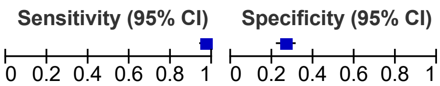


Figure 12: KSThR

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ha 2018	254	263	12	373	0.95 [0.92, 0.98]	0.59 [0.55, 0.63]
Park 2016	140	16	11	303	0.93 [0.87, 0.96]	0.95 [0.92, 0.97]
Xu 2018	966	671	39	789	0.96 [0.95, 0.97]	0.54 [0.51, 0.57]

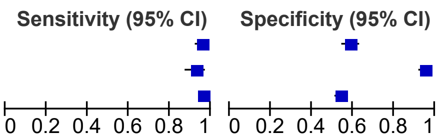


Figure 13: TIRADS (ACR)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Atilla 2018	79	880	22	1633	0.78 [0.69, 0.86]	0.65 [0.63, 0.67]
Chen 2018	385	313	10	384	0.97 [0.95, 0.99]	0.55 [0.51, 0.59]
Grani 2018	30	204	6	262	0.83 [0.67, 0.94]	0.56 [0.52, 0.61]
Ha 2018	255	297	11	339	0.96 [0.93, 0.98]	0.53 [0.49, 0.57]
Hoang 2018	14	48	1	37	0.93 [0.68, 1.00]	0.44 [0.33, 0.55]
Lauria Pantano 2018	93	525	20	439	0.82 [0.74, 0.89]	0.46 [0.42, 0.49]
Middleton 2017	297	1488	55	1582	0.84 [0.80, 0.88]	0.52 [0.50, 0.53]
Weiss 2018	5	9	0	28	1.00 [0.48, 1.00]	0.76 [0.59, 0.88]
Xu 2018	971	687	34	773	0.97 [0.95, 0.98]	0.53 [0.50, 0.56]
Zheng 2018	307	410	1	315	1.00 [0.98, 1.00]	0.43 [0.40, 0.47]

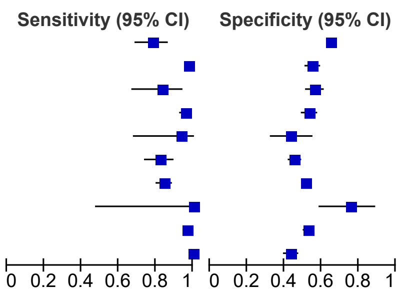


Figure 14: TIRADS (French)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Grani 2018	31	317	5	149	0.86 [0.71, 0.95]	0.32 [0.28, 0.36]
Hoang 2018	13	57	2	28	0.87 [0.60, 0.98]	0.33 [0.23, 0.44]
Macedo 2018	5	51	0	82	1.00 [0.48, 1.00]	0.62 [0.53, 0.70]
Maino 2018	11	66	3	270	0.79 [0.49, 0.95]	0.80 [0.76, 0.84]
Xu 2018	986	810	19	650	0.98 [0.97, 0.99]	0.45 [0.42, 0.47]
Yoon 2017	994	1754	50	1898	0.95 [0.94, 0.96]	0.52 [0.50, 0.54]
Zhang 2018	64	66	3	110	0.96 [0.87, 0.99]	0.63 [0.55, 0.70]

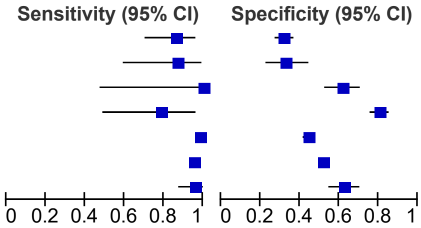


Figure 15: TIRADS (Kwak)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Koh 2018	193	759	19	509	0.91 [0.86, 0.95]	0.40 [0.37, 0.43]
Rahal 2016	114	274	9	579	0.93 [0.87, 0.97]	0.68 [0.65, 0.71]
Yoon 2016	228	749	6	310	0.97 [0.95, 0.99]	0.29 [0.27, 0.32]
Yoon 2017	1031	2719	13	933	0.99 [0.98, 0.99]	0.26 [0.24, 0.27]
Zhang 2015	222	339	6	3413	0.97 [0.94, 0.99]	0.91 [0.90, 0.92]
Zhang 2018	75	176	0	68	1.00 [0.95, 1.00]	0.28 [0.22, 0.34]
Zhou 2018	91	15	2	53	0.98 [0.92, 1.00]	0.78 [0.66, 0.87]

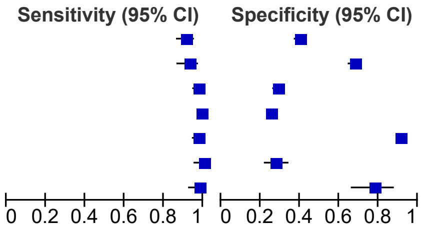


Figure 16: TIRADS (Korean)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Grani 2018	33	383	3	83	0.92 [0.78, 0.98]	0.18 [0.14, 0.22]
Ha 2016	254	263	12	373	0.95 [0.92, 0.98]	0.59 [0.55, 0.63]
Hoang 2018	13	71	2	14	0.87 [0.60, 0.98]	0.16 [0.09, 0.26]
Na 2016	367	462	87	1084	0.81 [0.77, 0.84]	0.70 [0.68, 0.72]

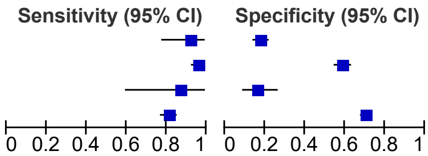


Figure 17: TIRADS (Horvarth)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Horvath 2009	349	360	46	342	0.88 [0.85, 0.91]	0.49 [0.45, 0.52]
Xu 2017	301	156	62	412	0.83 [0.79, 0.87]	0.73 [0.69, 0.76]

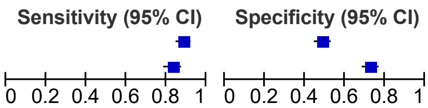


Figure 18: TIRADS (Zhang)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Zhang 2018	65	21	10	223	0.87 [0.77, 0.93]	0.91 [0.87, 0.95]

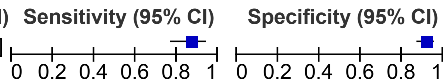


Figure 19: TIRADS (Zhang + CEUS)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Zhang 2018	73	10	2	234	0.97 [0.91, 1.00]	0.96 [0.93, 0.98]

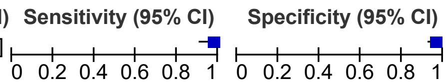


Figure 20: TIRADS (Kwak + CEUS)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Zhou 2018	91	15	2	53	0.98 [0.92, 1.00]	0.78 [0.66, 0.87]

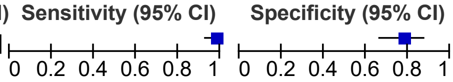


Figure 21: NCCN

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Yoon 2017	973	2200	71	1452	0.93 [0.91, 0.95]	0.40 [0.38, 0.41]

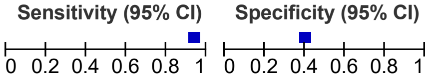


Figure 23: Children - ATA

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Creo 2018	46	63	4	32	0.92 [0.81, 0.98]	0.34 [0.24, 0.44]
Lim-Dunham 2017	12	9	0	12	1.00 [0.74, 1.00]	0.57 [0.34, 0.78]
Martinez-Rios 2018	45	22	7	49	0.87 [0.74, 0.94]	0.69 [0.57, 0.79]

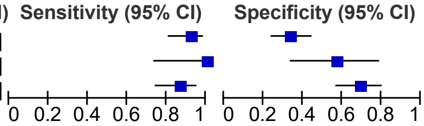
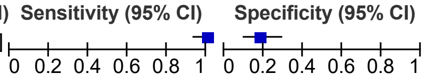


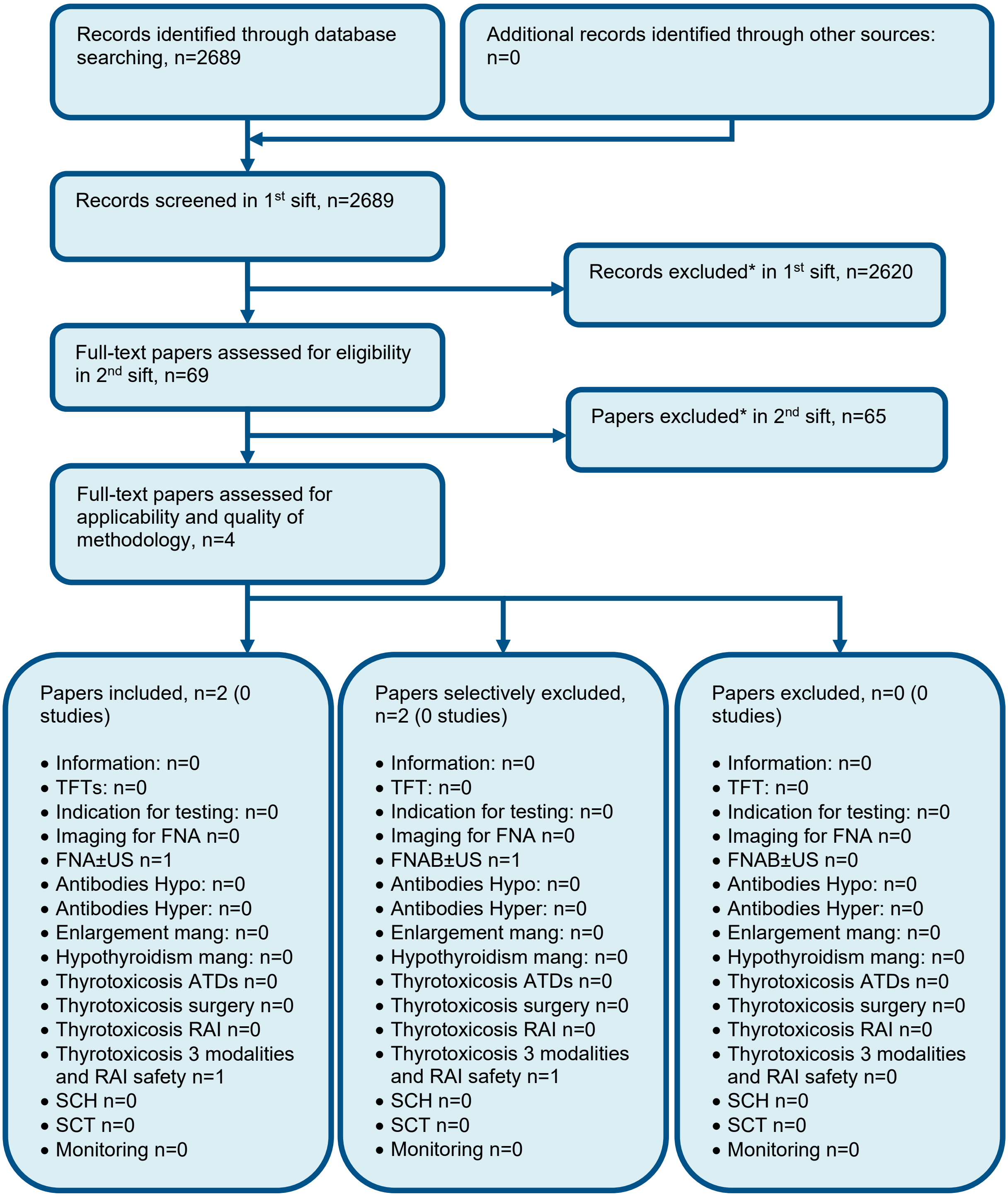
Figure 24: Children - TIRADS (Kwak)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Martinez-Rios 2018	52	58	0	13	1.00 [0.93, 1.00]	0.18 [0.10, 0.29]



Appendix F: Health economic evidence selection

Figure 22: Flow chart of health economic study selection for the guideline



* Non-relevant population, intervention, comparison, design or setting; non-English language
TFT; thyroid function test, FNA; fine-needle aspiration, US; ultrasound, RAI; radioactive iodine, ATDs; antithyroid drugs, Mang; management, SCH; Subclinical hypothyroidism, SCT; Subclinical thyrotoxicosis.

Appendix G: Health economic evidence tables

None

Appendix H: Health economic analysis

None

Appendix I: Excluded studies

I.1 Excluded clinical studies

Table 10: Studies excluded from the clinical review

Reference	Exclusion reason
Abdel-Rahman 2016[1]	Incorrect population
Afifi 2017[2]	USE not combined with US criteria
Aggarwal 2017[3]	No usable outcomes
Aghaghazvini 2018[4]	Incorrect population
Ahn 2018[5]	Inappropriate test
Akhavan 2016[7]	US no criteria
Al Nofal 2016[8]	SR, references checked
Albair Ashamallah 2016[10]	Incorrect population
Algin 2010[11]	USD not combined with US criteria
Appetecchia 2006[12]	No usable outcomes
Asteria 2008[13]	Inappropriate test
Azizi 2013[14]	Inappropriate tests
Bae 2018[15]	Erratum
Bae 2018[16]	Inappropriate study design
Bhatia 2011[17]	Inappropriate test
Bhatia 2012[18]	No usable outcomes
Bojunga 2010[19]	SR, references checked
Brito 2014[20]	SR, references checked
Brophy 2016[21]	No usable outcomes
Cakal 2015[22]	USE not combined with US criteria
Cakir 2011[23]	Inappropriate study design
Cam 2014[24]	No criteria used
Camargo 2007[25]	Inappropriate study design
Cantisani 2014[26]	No usable outcomes
Cantisani 2015[27]	Inappropriate study design
Cappelli 2005[30]	Inappropriate tests
Cappelli 2006[29]	Inappropriate study design
Cappelli 2007[28]	Inappropriate study design
Cavallo 2017[31]	Inappropriate tests
Cetin 2015[32]	Incorrect population
Chandramohan 2016[33]	Inappropriate study design
Chen 2010[38]	Inappropriate study design
Chen 2014[39]	Inappropriate tests
Chen 2016[35]	SR, references checked
Chen 2016[37]	Inappropriate study design
Chen 2017[34]	Inappropriate tests
Cheng 2013[40]	US no criteria

Reference	Exclusion reason
Cheng 2013[41]	Inappropriate population
Chi 2017[42]	Inappropriate test
Chiu 1998[43]	Inappropriate tests
Chng 2018[44]	Surgery only
Choi 2015[45]	US no criteria
Choi 2017[46]	Inappropriate test
Chong 2013[47]	Inappropriate test
Delfim 2017[50]	Two gate study design
Deng 2014[51]	Inappropriate study design
Deng 2017[53]	Inappropriate population
Deng 2018[52]	Inappropriate tests, no combination
Diao 2017[55]	Inappropriate population
Dighe 2010[56]	Inappropriate study design
Dighe 2013[57]	Inappropriate tests
Dilli 2012[58]	No usable outcomes
Ding 2011[59]	Inappropriate study design
Dobruch-Sobczak 2016[60]	No usable outcomes
D'Souza 2010[49]	Inappropriate test
Du 2018[61]	Inappropriate population
Duan 2016[62]	Inappropriate population
Dy 2017[63]	Inappropriate study design
Ebeed 2017[64]	Inappropriate population
El-Hariri 2014[65]	Inappropriate test
Elsayed 2016[66]	Inappropriate test
Fukunari 2004[68]	Inappropriate test
Gamme 2017[69]	Inappropriate study design
Gannon 2018[70]	Inappropriate test
Gao 2018[71]	Inappropriate population
Garcia-Monco Fernandez 2018[72]	Inappropriate population
Gietka-Czernel 2010[73]	USE no combination
Ginat 2010[74]	US no criteria
Giusti 2013[75]	Inappropriate population
Glogovsek 2015[76]	No usable outcomes
Goldfarb 2011[77]	Inappropriate population
Goldfarb 2012[78]	Inappropriate population
Gotzberger 2016[79]	Inappropriate study design
Gu 2012[82]	Inappropriate study design
Gu 2018[81]	Inappropriate tests
Guazzaroni 2014[83]	Inappropriate tests
Gul 2009[84]	Inappropriate tests
Gupta 2011[85]	Inappropriate population
Ha 2017[88]	No usable outcomes
Ha 2017[89]	No usable outcomes
Hamidi 2015[90]	Inappropriate tests
He 2016[91]	Inappropriate study design

Reference	Exclusion reason
Hoang 2018[93]	No usable outcomes
Hong 2009[95]	Inappropriate population
Hong 2012[96]	Inappropriate population
Hu 2018[98]	Inappropriate population
Huan 2014[99]	Inappropriate population
Huang 2015[100]	Inappropriate study design
Hughes 2017[101]	Inappropriate population
Ianni 2016[102]	Inappropriate study design
Ishigaki 2004[103]	Inappropriate study design
Ito 2007[104]	Inappropriate tests
Jiang 2015[105]	Inappropriate study design
Jin 2014[107]	Inappropriate population
Jin 2018[106]	Inappropriate tests
Kagoya 2010[108]	Inappropriate tests
Kakkos 2000[109]	Inappropriate tests
Kamran 2013[110]	Inappropriate tests
Kathuria 2003[111]	USD no combination
Khamis 2017[112]	Inappropriate tests
Kim 2008[121]	Inappropriate population
Kim 2008[123]	Inappropriate study design
Kim 2010[125]	Inappropriate study design
Kim 2012[118]	USE no combination
Kim 2013[115]	USE not validated criteria
Kim 2013[120]	Inappropriate tests
Kim 2015[113]	Inappropriate study design
Kim 2015[119]	Inappropriate population
Kim 2015[126]	USE no established criteria
Kim 2016[124]	Inappropriate population
Kim 2018[117]	Inappropriate study design
Kizilkaya 2014[127]	Inappropriate population
Ko 2012[128]	Inappropriate population
Koh 2016[130]	Inappropriate tests
Koike 2001[131]	Inappropriate tests
Koltin 2016[132]	Inappropriate tests
Kunz 2014[134]	Inappropriate population
Kwak 2011[135]	No usable outcomes
Kwak 2013[136]	No usable outcomes
Lai 2016[137]	Inappropriate population
Lee 2011[139]	Inappropriate tests
Li 2014[146]	Inappropriate population
Li 2015[141]	Inappropriate population
Li 2015[142]	No usable outcomes
Li 2015[147]	Inappropriate tests
Li 2015[148]	Inappropriate tests
Li 2016[144]	Inappropriate population

Reference	Exclusion reason
Li 2017[140]	Inappropriate tests
Li 2017[145]	Inappropriate study design
Li 2018[143]	Inappropriate tests, no combination
Liang 2018[149]	Inappropriate population
Lim 2008[151]	Inappropriate tests
Lin 2005[152]	Inappropriate tests
Lingam 2013[153]	Inappropriate tests
Lippolis 2011[154]	Inappropriate population
Liu 2011[159]	Inappropriate tests
Liu 2014[156]	Inappropriate population
Liu 2017[155]	USE no combination
Liu 2017[158]	Inappropriate population
Liu 2018[157]	Inappropriate population
Lu 1994[160]	Inappropriate population
Lu 2011[161]	US no criteria
Luo 2011[162]	Inappropriate tests
Luo 2012[163]	Inappropriate tests
Lyshchik 2005[164]	Inappropriate tests
Ma 2014[165]	US no criteria
Maia 2011[167]	Inappropriate population
Maia 2011[168]	Inappropriate population
Maimaiti 2016[169]	Inappropriate population
Majstorov 2015[171]	US no criteria
Mallikarjunappa 2014[172]	US no criteria
Mansor 2012[173]	USE no combination
Marqusee 2000[174]	US no criteria
Mehrotra 2013[176]	No usable outcomes
Memon 2017[177]	Inappropriate tests
Merino 2011[178]	US no criteria
Migda 2018[180]	Inappropriate population
Migda 2018[181]	SR, references checked
Mohamed 2013[182]	USE no combination
Mohammadi 2013[183]	US no criteria
Mohey 2013[184]	US no criteria
Moon 2007[185]	US no criteria
Moon 2008[189]	Inappropriate test
Moon 2011[186]	US no criteria
Nam 2016[192]	USG no combination
Nachiappan 2018[191]	USE no combination
Nemec 2012[194]	Inappropriate test
Nixon 2010[195]	No usable outcomes
Nixon 2013[196]	No usable outcomes
Nobrega 2007[197]	Inappropriate population
Noda 2015[198]	Inappropriate population
Nonchev 2017[199]	Not in English

Reference	Exclusion reason
Okamoto 1994[201]	US no criteria
Okamoto 1995[200]	Erratum, not relevant
Okasha 2018[202]	Inappropriate population
Oliveira 2018[203]	Inappropriate tests
Ozel 2012[204]	US no criteria
Palaniappan 2016[205]	US no criteria
Pandey 2017[206]	ARFI no combination
Pang 2017[208]	US no criteria
Papini 2002[209]	Inappropriate tests
Park 2009[212]	Derivation of criteria
Park 2009[213]	Inappropriate test
Park 2012[214]	No usable outcomes
Park 2015[210]	USE no established criteria
Park 2017[211]	Inappropriate population
Pathirana 2016[216]	Inappropriate population
Peccin 2002[217]	Inappropriate test
Petrone 2012[219]	Derivation of criteria
Phuttharak 2009[220]	US no criteria
Pompili 2018[221]	Inappropriate population
Popli 2012[222]	No combination with conventional US
Popowicz 2009[223]	Inappropriate population
Ragazzoni 2012[224]	Inappropriate population
Raggiunti 2011[225]	USE no combination
Raghavendra 2017[226]	Inappropriate tests
Rago 1998[230]	USE no combination
Rago 2007[227]	Inappropriate population
Rago 2007[228]	Inappropriate tests
Rago 2017[229]	USE no combination
Ram 2015[232]	US no criteria
Rao 2014[233]	USD no combination
Razavi 2013[234]	SR, not PICO
Razek 2008[235]	Inappropriate population
Refaat 2014[236]	Inappropriate population
Reginelli 2014[237]	No usable outcomes
Rios 2016[239]	Inappropriate tests
Rios 2018[238]	Not in English
Rivo-Vazquez 2013[240]	Inappropriate tests
Rosario 2015[242]	USD no combination
Rosario 2018[241]	Inappropriate population
Russ 2011[243]	Abstract only
Russ 2011[244]	Abstract only
Sagazio 2014[245]	Abstract only
Sahbaz 2017[246]	Abstract only
Saito 2015[247]	Abstract only
Sajjadieh 2005[248]	US no criteria

Reference	Exclusion reason
Salehi 2014[249]	US no criteria
Salmaslioglu 2008[250]	Inappropriate population
Samulski 2015[251]	US no criteria
Sands 2011[252]	Inappropriate population
Sarabia 2017[253]	Abstract only
Schenke 2015[254]	Inappropriate study design
Schenke 2019[255]	Inappropriate population
Schueller-Weidekamm 2009[256]	Inappropriate population
Sebag 2010[257]	USE no combination
Seo 2012[260]	US no criteria
Seo 2015[258]	Inappropriate tests
Seo 2017[259]	Inappropriate tests
Shankar 2015[261]	Abstract only
Shao 2015[262]	Inappropriate population
Shi 2013[263]	Inappropriate population
Shimura 2005[264]	Inappropriate study design
Shrestha 2012[265]	Inappropriate tests
Shuzhen 2012[266]	Inappropriate population
Siderova 2016[267]	Abstract only
Simon 2017[268]	Abstract only
Singaporewalla 2017[269]	Inappropriate tests
Stacul 2007[270]	Inappropriate tests
Stoian 2015[271]	Inappropriate population
Sui 2016[272]	Inappropriate population
Sun 2014[273]	SR, not matching PICO
Swan 2017[274]	Inappropriate tests
Szczepanek-Parulska 2013[275]	Inappropriate population
Taghipour Zahir 2013[277]	Inappropriate population
Taha Ali 2017[278]	Inappropriate tests
Tahmasebi 2016[279]	US no criteria
Tamsel 2007[280]	Inappropriate tests
Tatar 2013[282]	USE no criteria
Tatar 2014[283]	US no criteria
Tezelman 2007[284]	Inappropriate population
Trimboli 2012[285]	RTE not combined with validated
Tugendsam 2018[286]	Inappropriate population
Tunca 2007[287]	Inappropriate population
Tuzun 2016[288]	Inappropriate population
Unluturk 2012[289]	Inappropriate tests
Vargas-Uricoechea 2017[290]	USE no combination
Varverakis 2002[291]	USD no combination
Veyrieres 2012[292]	USE no combination
Vidal-Casariego 2012[293]	Inappropriate tests
Vorlander 2010[294]	Inappropriate tests
Wang 2006[301]	Inappropriate tests

Reference	Exclusion reason
Wang 2012[299]	USE no combination
Wang 2013[297]	USE no combination
Wang 2014[298]	Inappropriate tests
Wang 2014[304]	US no criteria
Wang 2015[300]	US no criteria
Wang 2017[295]	USE no combination
Wang 2017[302]	Only surgical
Wang 2018[296]	USE no combination
Wang 2018[303]	Inappropriate population
Watters 1992[305]	Inappropriate tests
Wei 2014[306]	SR, checked for references
Wei 2016[307]	SR, checked for references
Wei 2016[308]	Inappropriate tests
Wharry 2014[310]	Only surgical
Witczak 2016[311]	Inappropriate tests
Wu 2013[315]	Inappropriate tests
Wu 2016[312]	Inappropriate population
Wu 2016[314]	Inappropriate tests
Wu 2017[313]	Only indetermine on previous USE
Xia 2017[316]	Machine learning
Xing 2011[318]	Only surgical
Xing 2016[317]	ARFI no combination
Xu 2014[319]	Inappropriate study design
Xu 2014[321]	ARFI no combination
Xu 2015[322]	Only surgical
Xu 2016[320]	VTI no combination
Xue 2016[326]	Only surgical/core Bx
Xue 2017[325]	Only surgical
Yang 2017[327]	USE no combination
Yerli 2017[328]	USE no combination
Yoon 2014[333]	Inappropriate population
Yoon 2018[330]	Inappropriate tests
Yu 2017[334]	Inappropriate tests
Yuan 2012[335]	Only surgical/core Bx
Yuan 2015[336]	CEUS no combination
Yunus 2010[337]	US no criteria
Zayadeen 2016[338]	No usable outcomes
Zhan 2017[339]	Inappropriate population
Zhang 2010[340]	No usable outcomes
Zhang 2012[356]	Only surgical
Zhang 2013[342]	Inappropriate study design
Zhang 2014[343]	ARFI, no combination with US criteria
Zhang 2014[345]	ARFI no combination
Zhang 2014[347]	Non-systematic review

Reference	Exclusion reason
Zhang 2014[349]	Only surgical
Zhang 2014[355]	Only surgical
Zhang 2015[351]	Not in English
Zhang 2015[357]	ARFI, no combination with US criteria
Zhang 2015[358]	Only high risk based on US
Zhang 2016[352]	CEUS no combination
Zhang 2016[359]	No combination with conventional US
Zhang 2017[341]	Inappropriate population
Zhang 2017[344]	VTUS, no combination with US criteria
Zhang 2017[348]	Inappropriate population
Zhang 2018[350]	Not in English
Zhao 2018[360]	No combination with conventional US
Zhao 2018[361]	Inappropriate population
Zheng 2013[362]	No combination with conventional US
Zhou 2014[366]	USD, no combination with US criteria
Zhou 2016[364]	Inappropriate population
Zhou 2017[365]	No combination with conventional US
Zhu 2013[368]	Inappropriate tests

I.2 Excluded health economic studies

None