National Institute for Health and Care Excellence

1

Draft

Thyroid Cancer

[A] Evidence reviews for ultrasound accuracy and threshold of nodule size and classification

NICE guideline <number>

Evidence reviews underpinning recommendations 1.2.6, and 1.2.8 to 1.2.10 in the NICE guideline

June 2022

Draft for Consultation

These evidence reviews were developed by the National Guideline Centre

1

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/or their carer or guardian.

Local commissioners and/or 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 All NICE guidance is subject to regular review and may be updated or withdrawn.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:

Contents

1

1. Ultrasound

1.1. Review question: What is the diagnostic accuracy of ultrasound for identifying i) thyroid nodule malignancies or ii) nodules with malignant potential?

1.1.1. Introduction

Ultrasound is the initial imaging used to assess thyroid enlargement. It is non-invasive, easy to use, and helps determine whether a thyroid cancer is likely to be present. A positive result means further investigation is needed. Ultrasound includes a variety of types such as conventional grey scale ultrasound, Doppler ultrasound, contrast enhanced ultrasound and elastography.

This review seeks to evaluate the diagnostic accuracy of different ultrasound (US) characteristics, or combinations of US characteristics, at different thresholds defining a positive US test, to detect thyroid cancer or nodules with malignant potential.

1.1.2. Summary of the protocol

For full details see the review protocol in Appendix A.1.

Table 1: PIRO characteristics of review question

Population	Inclusion: People aged 16 or over who are suspected of thyroid cancer. Exclusion: Children and young people under 16 years.
Target conditions	i) nodules with thyroid cancer malignancy ii) nodules with potential for malignant transformation
Index test	Ultrasound: this umbrella term includes conventional grey scale ultrasound, Doppler ultrasound, contrast enhanced ultrasound, elastography, and combinations of these.
Reference standard	Surgical histopathological findings
Statistical measures	Sensitivity and specificity
Study design	Retrospective or prospective designs. Retrospective designs may have an inherent bias in that the only people with histopathological findings may be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.

1.1.3. Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual. Methods specific to this review question are described in the review protocol in appendix A and the methods document.

Declarations of interest were recorded according to NICE's conflicts of interest policy.

1.11.4. Diagnostic evidence

1.1.12.1. Included studies

For the first part of the question, where the aim was to evaluate the accuracy of ultrasound in diagnosing *malignant* thyroid nodules, 133 studies were found and included in the review.[3, 7, 10, 13, 18, 21, 27, 33, 35, 36, 43-45, 55, 57, 60, 77, 81, 87, 104-106, 108, 111, 112, 117, 121, 139, 140, 142, 144, 147, 149, 154-156, 169, 171-173, 175, 179, 180, 188, 189, 195, 198, 204, 217-220, 223, 225, 232, 238, 240, 242, 243, 253, 255, 256, 260, 262, 263, 267, 268, 271, 275, 284, 286, 295, 296, 300, 303, 305-307, 311, 314, 315, 317, 318, 320-323, 333, 334, 336, 337, 343-346, 349, 350, 352-357, 361, 362, 364, 367, 370, 373, 375, 376, 380, 381, 383, 386, 388, 391, 392, 398, 400, 404, 405, 410, 412, 417, 423-425, 431, 433, 438] These studies are summarised in Table 2, and details of the scales used are provided in Table 3. Evidence from the included studies is summarised in the clinical evidence summaries below in Table 4 to Table 10.

For the second part of the question, where the aim was to evaluate the accuracy of ultrasound in diagnosing nodules with *malignant potential*, no relevant studies were found. Such studies would need to have utilised the gold standard of surgical histopathology to definitively differentiate nodules with malignant potential from all other nodules, and no such studies appear to currently exist in the literature.

Diagnostic accuracy was focused on sensitivity and specificity, which are independent of prevalence. Positive and negative predictive values, though important, were not directly calculated for each test because these values are dependent on the study prevalence of thyroid malignancy. Because the study prevalence often differs from the population prevalence such values may be misleading. The committee set clinical decision thresholds for sensitivity of 0.9, above which a test would be recommended, and 0.8, below which a test would be deemed of no clinical use. They also set clinical decision thresholds for specificity of 0.5, above which a test would be recommended, and 0.1, below which a test would be deemed of no clinical use.

The index test of ultrasound is a complex entity, and it could not be evaluated as a single all-encompassing test. It was agreed post-hoc that it needed to be looked at in all its many forms, such as grey scale, Doppler, contrast enhanced, or elastography and at appropriate thresholds of the many ultrasound characteristics or combinations of characteristics that could be evaluated within those forms. In every case, diagnostic accuracy could only be ascertained if the index test outcome were dichotomised – that is, if all possible results for that test were split into two mutually exclusive and exhaustive positive and negative categories (so that these two index test categories can be cross-tabulated with the similarly binary gold standard positive and negative categories to assess agreement). For index tests looking at single characteristics such as 'microcalcifications' this was a simple undertaking as the positive category is simply 'the existence of microcalcifications' and the negative category must therefore be 'the absence of microcalcifications'. For ordinal scales such as the various TIRADS systems, it is more complex, as the binary negative/positive threshold can be placed in between any of the possible scores; for example, while 'negative' can be all scores below 2 and 'positive' can be all scores of 2 and above, 'negative' can also be defined as all scores below 3 and 'positive' as all scores of 3 and above, and so on. Therefore, for scales-based tests there are a number of possible accuracy results, each relating to a possible position of the threshold demarcating a negative and positive test. It is important to note that for such ordinal scales it is not appropriate to select one score as positive and the others as negative, as would be appropriate for a nominal categorisation. On an ordinal scale, where a chosen score may fall along a continuum of scores that code for an increasing level of malignancy suspicion, it would not be sensible to use scores that are both below and above the chosen positive score to denote a negative test because this would be unlikely to have any diagnostic potential (you would want the negative category to be the lower risk group, but this would be less likely if participants are partially derived from those with higher risk scores). Instead, the strategy is to demarcate positive and negative as the threshold between the 'chosen score and above' and the 'scores below the chosen score'. This strategy will allow

the committee to decide not only the optimum form of test, but the optimum threshold of the test as well.

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

Index test data were divided into different types: i) individual grey-scale characteristics, ii) informal and largely simple combinations of grey-scale characteristics, iii) more formal combinations of predominantly grey scale characteristics, organised into scales such as EU TIRADS, iv) individual doppler characteristics; v) individual contrast enhanced ultrasound (CEUS) characteristics, vi) elastography characteristics and vii) combinations of all types of US characteristics and methods. Please see Table 3 for further information on the scales used in type iii and vi. The index test results have been presented separately in these seven groups for clarity.

The gold standard for all studies was surgical histopathological findings, on the basis that there are no contending alternatives with equivalent face validity, and this was adhered to strictly. Some of the excluded studies did not require surgical histopathology from those in the lowest FNAC category (benign), assuming that there would be no false negatives in this category. Evidence (i.e., https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4159900/) suggests that this assumption is not always true which is why all studies using this approach were excluded.

Both prospective and retrospective studies have been included in this review. It has been assumed that in prospective studies all participants received surgery in order to gain definitive histopathological gold standard data, and that this does not reflect any population bias. However, for retrospective studies this assumption could not be made, because the data were collected before any intention of evaluating diagnostic accuracy (and any contemplation of the nature of any gold standards) had been made. It was therefore possible that retrospective studies with surgical histopathological data would preferentially contain participants who were particularly indicated for surgery, making them different from the intended population for this review, which were a more general cross-section of people who are simply suspected of thyroid nodule malignancy. Thus, retrospective studies have been downgraded for population indirectness.

Data were meta-analysed with Bayesian methods using WinBugs software (see methods chapter) provided that at least 3 data cohorts with appropriately similar PIRO were available. If only two data cohorts were available the data were not meta-analysed, and the data from the two papers were simply presented side by side to allow transparent interpretation.

Medical training of the sonographer had been chosen during protocol development as the variable that could potentially influence accuracy. Therefore, if heterogeneity was noted in meta-analyses, the existence of medical training in the ultrasound tester was used to sub-group studies. Many meta-analyses demonstrated some degree of heterogeneity but in no cases did the medical training status of the sonographer resolve the heterogeneity within sub-groups. This indicated that medical training was not a significant factor influencing the variability in accuracy between studies. Therefore, all meta-analyses with heterogeneity were downgraded for inconsistency.

See also the study selection flow chart in Appendix C, sensitivity and specificity forest plots and sensitivity/1-specificity plots in Appendix F, and study evidence tables in Appendix D.

1.145.2. Excluded studies

See the excluded studies list in Appendix J.

1

1.12.5. Summary of studies aiming to detect nodule malignancy

3 **Table 2: Summary of studies included in the evidence review**

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
Abd_Alrahman, 2017[3]	Unclear	30	Patients with a solitary thyroid nodule for whom final diagnosis was available through surgical histopathology; US and elastography performed	Complex nodules with >30% cystic component; large nodules occupying >75% of the thyroid lobe with little or no discernible normal thyroid tissue; nodules with egg shell or completely calcified nodule, patients with abnormal US texture of the thyroid tissue and patients with previous thyroid surgery or previous radiation to the head and neck region	Grey scale US Doppler US elastography
Aggarwal, 1989[7]	India	36	Patients with ultrasonographically solitary and scintigraphically cold thyroid nodules	None reported	Grey scale US
Ahmadi, 2019[10]	USA	213	Adults with thyroid nodules >5mm undergoing thyroid surgery at a tertiary care hospital	Patients in whom pre-operative ultrasound imaging was not available, their ultrasound imaging was of poor quality, or their surgical pathology report was not available	Grey scale US
Akhaven, 2016[13]	Iran	90	Patients with a thyroid nodule referred to the research hospital who were candidates for surgical nodule resection	Patients with known thyroid autoimmune disease, age < 15, contraindications to anaesthesia, surgical resection	Grey scale US Doppler US
Appetecchia, 2006[18]	Italy	203	Patients submitted to surgery for solitary thyroid disease	None reported	Grey scale US Doppler US
Aslan, 2018[21]	Turkey	86	All consecutive patients scheduled for subtotal or total thyroidectomies due to multinodular goitre, or malignant or suspicious FNA results; nodular thyroid disease confirmed by grey scale US; co-operated with Duplex Doppler	Patients who could not co-operate with operator during DDUS exam; nodules intertwined or difficult to depict alone; nodules without vascularity on colour doppler US (CDUS); nodule without histopathological diagnosis; malignant tumour other than PTC.	Doppler US

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
			US (DDUS) evaluation; histopathological diagnosis of PTC or benign nodule		
Bakari, 2018[27]	China	54	1) Patients who are 18 years of age or older and of different genders; (2) Patients with single or multiple solid TNs bigger than 0.5 cm in size and scheduled for surgical treatment; (3) Patients with thyroid nodules of ≥0.5 cm with at least two of the following characteristic features hypoechoic, microcalcifications, irregular outline, taller than wider, increased central vascularity, isoechogenicity, and heterogeneity patterns; and (4) No invasive thyroid surgery or FNAB were performed before.	(1) Patients with fluid filled thyroid lesions; (2) Patients with thyroid nodules lesions of 0.5 cm regardless of their characteristic features.	Grey scale US Doppler US elastography
Berni, 2002[33]	Italy	108	None reported	None reported	Grey scale US Doppler US
Bora Makal, 2021[35]	Turkey	141	Patients undergoing thyroidectomy	patients without nodules; undergoing surgery because of hyperthyroidism, autoimmune thyroiditis or metastatic thyroid cancer; having previous neck radiation or surgery; not also being evaluated with shear wave elastography SWE	Grey scale US elastography
Borlea, 2020[36]	Romania	133	Patients presenting with solid thyroid nodules suitable for US assessment; surgical pathology report and US evaluation performed	Patients with previous thyroid surgery or radiation therapy or who presented with completely cystic nodules; autonomously functioning nodules	Grey scale US Doppler US elastography
Cakir, 2011[43]	Turkey	292	Patients admitted to outpatient thyroid clinic	Patients <15 years; previous history of thyroid surgery or percutaneous invasive procedures for thyroid nodules or radiotherapy to the head and neck region; cardiac or pulmonary disease; pure cystic nodules; haemorrhagic nodules, multinodular coalescent nodules with	Grey scale US elastography

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
				undetermined sizes; anaplastic carcinoma; Riedel thyroiditis; extensive cervical metastasis; patients refusing surgery.	
Cantisani, 2015 [44]	Italy	50	Presence of any thyroid nodule; FNAC and surgery performed during study period	Cystic nodules; coarse calcification; spongiform nodules; pregnancy; heart failure; severe pulmonary hypertension	elastography
Cao, 2021 [45]	China	355	Thyroid nodule resection performed for first time; pre-op conventional US and contrast enhanced ultrasonagraphy (CEUS) performed	Diffuse thyroid disease; prior history of thyroid surgery; no histopathology results (FNAC only); quality of dynamic contrast images was poor; special nodule cases such as cystic and spongiform nodules, uniform hyperechoic nodules of Hashimoto's thyroiditis, multiple predominantly solid nodules and/or predominantly cystic nodules with similar US appearance, uncertainty of internal characteristics because of calcifications, and nodules with snowstorm pattern of microcalcifications.	Grey scale US CEUS
Chen, 2016[55]	unclear	253	None reported	None reported	Grey scale US CEUS elastography
Chen, 2019[57]	China	120	Thyroid lesions confirmed by surgery and pathology	Nodules of >2cm diameter and <0.3cm diameter; Hashimoto's thyroiditis; cystic dominated thyroiditis; cystic dominated nodules; calcified nodules	Grey scale US
Chng, 2018[60]	Singapore	150	People with US prior to thyroid surgery	None reported	Grey scale US Doppler US
Deng, 2018[77]	China	92	1) Patients with thyroid diseases underwent both high frequency colour Doppler ultrasound (HFCDU) and pathological diagnosis; 2) Patients were willing to cooperate with the	1) Patients had other severe thyroid diseases in addition to thyroid nodules, or the acoustic halo couldn't be detected in thyroid nodules; 2) Patients experienced recurrence of malignant	Grey scale US Doppler US

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
			treatment; 3) Patients' medical records were complete	thyroid nodules after surgery, or had hyperplasia of one thyroid lobe caused by the hypoplasia of thyroid and parathyroid in the opposite lobe; 3) Patients had benign thyroid nodules generated by the scar and proliferation of the residual thyroid tissue or other factors after treatment.	
Dobruch-Sobczak, 2019[81]	Poland	428	Patients who had been admitted to the tertiary referral centre for thyroidectomy; population from a previously iodine deficient region	Symptomatic purely cystic lesions	Grey scale US
El-Hariri, 2014 [87]	Egypt	72	A solid nodule in one thyroid lobe	(1) Cystic component >15% of the nodule volume, (2) Large nodules occupying >75% of thyroid lobe volume because insufficient surrounding normal thyroid tissue to be used as reference and (3) nodules with peripheral calcifications.	elastography
Gao, 2019[104]	China	1758	Patients with total or nearly total thyroidectomy or lobectomy; complete preoperative US of nodules; surgical pathology	Non-mass-forming lesions and nodules that failed to meet the criteria for any pattern of ATA guidelines	Grey scale US
Garcia-Monco Fernandez, 2018[105]	Spain	263	Patients with total or partial thyroidectomy due to nodular thyroid disease	None reported	Grey scale US
Garg, 2018 [106]	India	97	Patients with thyroid nodules having elastography, and surgery	Patients detected to have pure cystic thyroid nodules (anechoic nodules without solid components), or nodules with eggshell calcification; patients with prior diagnosis of thyroid neoplasms, thyroid surgery, radioiodine therapy, or any severe comorbid states	elastography
Giammanco, 2002[108]	China	125	Patients with thyroid pathology receiving US B mode, doppler and thyroid surgery with histopathological confirmation of diagnosis	None reported	Doppler US
Goldfarb, 2011[267]	USA	624	Consecutive patients undergoing thyroidectomy; solitary or dominant thyroid	None reported	Grey scale US

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
			nodules; underwent surgeon-operated US before thyroidectomy		
Gorgulu, 2019[111]	Turkey	83	Patients admitted for thyroidectomy, according to thyroid US findings, FNA findings and clinical evaluations; single or multiple nodules of 40mm or less	Surgical history of the neck affecting the sternocleidomastoid muscle, pure cystic lesions, insufficient normal tissue around the measured nodule, isthmic nodules, rough calcification and autoimmune thyroiditis	Grey scale US
Gorgulu, 2021 [112]	Turkey	123	Patients who were admitted to the otorhinolaryngology and general surgery clinics in Adana Numune Research and Training Hospital (Adana, Turkey) for thyroidectomy according to preoperative clinical evaluation were included; single or multiple nodules ≤40 mm	The existence of pure cystic lesions, insufficient normal tissue surrounding the measured nodule, isthmic nodules, nodules larger than 40 mm, rough calcification and autoimmune thyroid disease were all exclusion criteria.	Elastography
Gray, 2014[117]	UK	78	Patients who underwent partial or total thyroidectomy	Ultrasonography report or suitable hard copy of ultrasound images unavailable, or an incomplete report; ultrasonography not originally performed by our head and neck radiologist; no definite thyroid nodule on the ultrasound image; repeat ultrasound for the same patient; follow-up ultrasound for histologically proven thyroid cancer; and pathology report unavailable	Grey scale US
Gu, 2011[121]	China	72	Patients undergoing thyroidectomy	Anatomic abnormalities of the neck, cystic lesions of a completely liquid nature, maximum nodule diameter of <6mm	Grey scale US Elastography
Hang, 2018[139]	China	262	(a) Age > 18 years; (b) no therapy or biopsy before US examination; (c) diameter of the thyroid nodule between 5 and 30mm; (d) patients with solid or mostly solid thyroid nodules as assessed by US examination (cystic part < 50%); (e) distance from the skin	None reported	Grey scale US elastography

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
			surface to nodular center was <25mm, as this could be fully included in the maximum range of the SWE colour overlay; (f) underwent conventional US and SWE examination before surgery		
Haskjold, 2021 [140]	Norway	101	Symptomatic thyroid nodules or incidentally discovered thyroid nodules referred to clinic that had US and surgery	No histopathology	Grey scale US
He, 2018[142]	China	83	Solid or cystic nodules with >50% solid; nodule diameter >6mm; normal tissue around the nodule; no surgery, drug or chemotherapy administered before the operation; thyroid surgery with histopathological results	None reported	Grey scale US elastography
Hekimsoy, 2021 [144]	Turkey	165	Patients who had undergone 7660 detailed US examinations of the thyroid gland during a 5-year period was obtained to compose a study population with histopathologically evaluated thyroid nodules.	Not reported	Grey scale US
Hong, 2009[147]	China	90	Patients with thyroid nodules referred for surgical treatment	None reported	Grey scale US Doppler US Elastography
Horvath, 2017[149]	Chile	210	Consecutive patients undergoing thyroid US and thyroidectomy	Patients with incomplete surgical or pathological information; undergoing surgery at other institutions; nodules whose anatomopathological characterisation was not possible due to tissue manipulation	Grey scale US
Huang, 2015 [156]	China	136	(1) Nodules underwent US, p-SWE and ARFI-induced SE. (2) Nodules were confirmed by histopathology after surgery.	(1) Maximum diameter of nodule was less than 7 mm. (2) Image data of nodules were not complete: US, p-SWE, or ARFI-induced SE image quality was poor. (3) Mixed cystic (< 50% solid) or almost cystic nodules. (4) There	elastography

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
				was not enough thyroid tissue surrounding the nodule.	
Huang, 2019[154]	China	109	Patients undergoing surgical resection of single thyroid nodules at or under 1cm	Pregnancy; nodules with predominantly cystic or coarse calcification	Colour Doppler US CEUS
Huang, 2020[155]	China	346	Patients with thyroid nodules treated surgically at the research hospital; TIRADS category 4 nodules; conventional US, FNAC and elastography performed before surgery; final diagnosis based on surgery	Surgery for hyperthyroidism; previous history of neck radiation or surgery	Grey scale US elastography
Jeong, 2016 [169]	South Korea	178	Patients with nodular hyperplasia, follicular adenoma or follicular carcinoma with a diagnosis made from surgical specimens between January 2002 and May 2013	Other follicular pattern lesions, such as the follicular variant of papillary cell carcinoma and Hurthle cell neoplasm, cases with multiple nodules in a lobe or multinodular goiter and cases without preoperative ultrasonography	Grey scale US
Jiang, 2014[172]	China	122	Patients undergoing thyroidectomy for previously diagnosed thyroid nodules, who had previously undergone contrast enhanced US.	None reported	Grey scale US
Jiang, 2015[171]	China	122	Patients who underwent surgery for thyroid nodules with calcification	None reported	Grey scale US CEUS
Jin, 2018[173]	China	94	Confirmed cases (from January 2011 to January 2015) of thyroid nodules by ultrasound	None reported	Grey scale US elastography
Kalantari, 2018[175]	Iran	63	Presence of thyroid nodules, with indication for surgery because of suspicious and/or indeterminate FNA and/or clinical risk factors for thyroid nodules	None reported	Grey scale US Doppler US
Kim, 2008[179]	South Korea	92	Patients with eggshell calcifications; patients undergoing thyroidectomy	Nodules with combinations of eggshell calcifications and other types of calcifications	Grey scale US

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
				such as microcalcifications and macrocalcifications	
Kim, 2008[188]	South Korea	174	Patients with macrocalcifications; patients undergoing thyroidectomy for thyroid nodules	Patients with microcalcifications (multiple punctate bright echoes <2mm with or without acoustic shadows)	Grey scale US
Kim, 2012[180]	South Korea	505	None reported	None reported	Grey scale US
Kim, 2016[189]	South Korea	85	Consecutive patients referred for total thyroidectomy	None reported	Grey scale US
Kobayashi, 2005[195]	Japan	910	Patients with follicular adenomas, adenomatous thyroid nodules and follicular carcinomas	Patients with papillary carcinomas, undifferentiated carcinomas, medullary carcinomas and malignant lymphomas	Grey scale US
Kong, 2017[198]	China	92	Presence of a signal or multiple nodular thyroid lesions; thyroidectomy surgery with full histopathological results	FNA performed before admission, which might affect vascularity	Grey scale US Doppler US
Kuru, 2021 [204]	Turkey	1122	Patients undergoing US, FNA and thyroidectomy	Not reported	Grey scale US
Li, 2015[218]	China	73	All subjects had small thyroid nodules, difficult to define on conventional ultrasound	None reported	CEUS elastography
Li, 2016[223]	China	762	Patients with thyroid nodules who underwent surgery and had a pathological diagnosis	None reported	Grey scale US Doppler US
Li, 2017[217]	China	89	Patients undergoing US and CEUS, but overall unclear	None reported	Grey scale US CEUS
Li, 2018[220]	China	68	1) Patients were diagnosed as having thyroid nodules by palpation and general neck ultrasound; 2) patients received surgery and pathologic diagnosis; 3) patients signed the informed consent	1) The nodules were completely cystic; 2) the nodules were too big or too close to the edge where there was no adequate thyroid tissue surrounded for comparison; 3) patients received head and neck radiotherapy in the past; 3) patients were obese with excessive fat	Grey scale US Doppler US elastography

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
				in the larynx that could affect the results of BUS	
Li, 2021 [219]	China	356	(1) The patients underwent thyroid surgery and had pathological results; (2) the patients had not been previously treated for thyroid nodules; and (3) the patients had no history of radiotherapy of the head and neck regions.	(1) More than 25% of the nodule consisted of the cystic component (because shear waves cannot propagate in liquid); (2) the nodule contained coarse or rim calcifications, which cause information loss in SWE images; (3) the nodule was located in the isthmus or adjacent to the cartilage of the trachea and common carotid artery (because it was difficult to distinguish between actual stiffness and artifacts); and (4) benign and malignant nodules appeared in the same thyroid lobe (because it is difficult to determine the pathological nature of the target nodule).	Grey scale US Elastography
Li, 2021 [225]	China	78	Patients who received a diagnosis of Follicular Thyroid Carcinoma by thorough histopathologic analysis of resected hemi-thyroidectomy or total thyroidectomy specimens from a pathology report database OR patients with a diagnosis of Follicular Thyroid Adenoma	Unavailability of data	Grey scale US
Lin, 2018 [232]	China	80	None reported	None reported	elastography
Liu, 2014[240]	China	49	Patients with thyroid nodules; surgery performed with histopathology performed within the study period; thyroid function and autoimmune bodies tested before surgery	Only diffuse thyroid disease; cystic nodule of completely liquid components; insufficient thyroid parenchyma surrounding the nodule; previous history of radiation therapy of head and neck	Grey scale US Doppler US Elastography
Liu, 2017[242]	China	100	Patients with thyroid nodules receiving an ultrasonic examination and operation	None reported	Grey scale US elastography

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
Liu, 2019[243]	China	174	Patients with thyroid nodules scheduled to undergo surgery imaged by greyscale US and SWE	Previous invasive procedures for thyroid nodules; no histopathological data	Grey scale US elastography
Liu,2020[238]	China	90	18-80 years; patient received US and SWE before surgery; proven by surgery and pathological result was FTC or FA	Nodule mostly cystic; image data incomplete; pathological result of the target nodule was unclear or uncertain when patient had multifarious pathological results	CEUS
Lyshchik, 2007[253]	Japan	56	Patients with solid thyroid nodules with the preoperative suspicion of thyroid cancer based on clinical, imaging and cytologic findings	Patients who refused to give informed consent or who did not receive surgical treatment	Doppler US
Ma, 2014[256]	China	144	>18 years; solid or mainly solid nodules on GSUS and CDUS; nodule size>0.5cm	Dominant cystic nodules; pregnancy; grade III-IV NYHA cardiac function; severe pulmonary hypertension; no surgical pathology of the thyroid gland	Grey scale US CEUS Doppler US
Ma, 2017[255]	China	125	Maximum nodule diameter <1cm; age 18 or over; surgery performed; histopathological diagnosis available	Pregnancy or lactation; not enough thyroid parenchyma surrounding the nodule; dominant cystic (>75%) nodules; NYHA grade III-IV; severe pulmonary hypertension; severe allergies	Grey scale US Doppler US CEUS elastography
Magri, 2020[260]	Italy	255	Age >18 years; detailed pre-op thyroid US exam; availability of data concerning thyroid autoimmunity; availability of US-guided FNA; total or partial thyroidectomy performed; availability of histological surgical diagnosis;	Any of the aforementioned procedures done in another clinical setting;	Grey scale US
Maia, 2011[262]	Brazil	143	Patients submitted to total or partial thyroid surgery; US examination	None reported	Grey scale US Doppler US
Maimati, 2016[263]	China	600	Thyroid nodules ≤ 10 mm in maximum diameter in 600 patients who underwent conventional ultrasonography examinations of the thyroid gland before surgery	No surgical evaluation; nodules >10mm	Grey scale US Doppler US

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
McClean, 2021 [268]	UK	296	All US reports for patients who underwent thyroid surgery were reviewed. Prior to 2014, patients were selected for surgery based on clinical assessment and FNA result. From 2014, patients were selected for surgery according to BTA guidelines	Patients who underwent FNA and surgery without US assessment were not included in the study	Grey scale US
Mohamed, 2013[271]	Egypt	45	Patients with solitary solid thyroid nodules, referred for surgical treatment	Patients with multiple nodules, and nodules with extensive macrocalcifications	Grey scale US Doppler US elastography
Mohey, 2013[275]	Egypt	46	Patients with solid thyroid nodules	Cystic thyroid nodules; calcified shell thyroid nodules; coalescent nodules in multi-nodular goitre	Grey scale US Doppler US elastography
Nemec, 2012[284]	Austria	46	Age 18-80; sonographically identified solitary, cold thyroid nodule on Tc-99m pertechnetate thyroid scintigraphy; scheduled surgery for the nodule	Pregnancy; multinodular thyroid goitre or hot nodules; history of thyroid malignancy; previous thyroid surgery or biopsy; previous adverse reactions to i.v. contrast agents	CEUS
Nilakantan, 2007[286]	India	106	All patients undergoing surgery for benign or malignant nodular thyroid disease	None reported	Grey scale US
Pagano, 2021 [295]	Italy	146	Consecutive patients who were referred to a Center for FNA cytology for suspected thyroid nodules and then underwent thyroid surgery	Cases undergoing surgery with a non-diagnostic cytology at FNA	Grey scale US
Paker, 2021 [296]	Israel	216	One inclusion criterion was the availability of sufficient data, which could be either a preoperative, detailed ultrasound report, containing all the sonographic features included in the aforementioned risk-stratification system or recorded preoperative pictures of the nodules on the ultrasound hard disc with sufficient	Thyroidectomies performed due to non-thyroid disease and non-differentiated thyroid lesions (lymphoma, anaplastic, amyloidosis).	Grey scale US

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
			details to determine the exact classification in each of the risk-stratification systems. Other inclusion criteria were postoperative surgical pathology of a differentiated thyroid carcinoma and the presence of 1–3 nodules in the excised lobe.		
Parikh, 2013[300]	USA	84	People undergoing surgeon-performed US before thyroidectomy, with provisional diagnosis of Hurtle cell neoplasms on FNA.	None reported	Grey scale US
Park, 2012[303]	South Korea	400	Patients undergoing US-FNA and surgery during the study period	None reported	Grey scale US
Pathirana, 2016[305]	Sri Lanka	132	Patients scheduled for thyroidectomy	Patients with goitres extending retrosternally or with diffuse goitres	Grey scale US
Peccin, 2002[306]	Brazil	80	Patients with palpable solitary thyroid nodules or multinodular goitres with a dominant nodule operated on for clinical and/or cytological suspicion of malignancy or for symptoms of compression	Patients with multinodular goitre without a dominant nodule, patients with toxic nodules, and patients whose thyroid nodules did not meet clinical or cytological criteria for surgery	Grey scale US
Pei, 2019[307]	China	170	(a) Those meeting the criteria for TI-RADS category 4; (b) all patients with complete data, including US indicators and pathological findings; (c) all nodules, in which RTE and SMI were successfully implemented; and (d) all thyroid nodules that were not subjected to minimally invasive surgery prior to US examination (such as puncture and ablation).	None reported	Grey scale US Doppler US elastography
Phuttharak, 2009[311]	Thailand	31	Patients with thyroid nodules presenting to ENT department; after FNA and clinical evaluation those thought to be at risk of thyroid malignancy; received US and thyroidectomy	None reported	Grey scale US Doppler US

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
Qi, 2021 [314]	China	884	Consecutive patients with thyroid nodules given US and followed up with thyroidectomy surgery	Pregnant and breastfeeding women	Grey scale US
Ragazzoni, 2012[315]	Italy	115	Presence of single or multiple thyroid nodules clearly distinguishable from surrounding parenchyma.	Cystic nodules and mixed nodules with liquid component >30% of the nodule total volume; nodules with eggshell calcifications (but internal calcifications NOT excluded)	Grey scale US Doppler US elastography
Rago, 1998[318]	Italy	104	Consecutive patients with a single thyroid nodule either in a normal thyroid or in a goiter, who underwent surgery for compressive symptoms or clinical suspicion of malignancy; nodules cold on scintiscans	None reported	Grey scale US Doppler US
Rago, 2007[317]	Italy	92	Patients who underwent thyroid surgery for compressive symptoms or suspicion of malignancy on FNA cytology	None reported	Grey scale US Doppler US elastography
Refaat, 2014 [320]	Egypt	30	Patients with solitary thyroid nodules referred for surgical treatment	Patients with multiple thyroid nodules (more than two nodules), previous surgery or radioiodine therapy and patients with thyroid nodules who refused or had any contraindication for thyroid surgery; patients with purely cystic (anechoic nodules without solid components) and egg shell-calcified nodules; patients with nodules of greatest diameter larger than 40 mm	elastography
Ren, 2015[321]	China	124	Patients with a preoperative diagnosis of thyroid nodules of any size by sonographers; patients undergoing total or partial thyroidectomy surgery with a histopathological diagnosis of malignancy or benignity; patients with predominantly solid nodules (>50%)	Malignancy that was not papillary TC	Grey scale US

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
Reverter, 2019[322]	Spain	300	Age >18 at time of diagnosis; total or nearly total thyroidectomy or lobectomy; preoperative US evaluation of thyroid nodules; available pathologic findings	None reported	Grey scale US
Rivo-Vazquez, 2013[323]	Spain	156	Patients on a surgical waiting list for thyroidectomy due to nodular thyroid disease	Diffuse goitre	elastography
Sancak, 2010[333]	Turkey and Germany	unclear	Patients with benign non-functioning nodules, papillary carcinomas and surrounding normal tissue	None reported	Grey scale US Colour Doppler
Schenke, 2018[336]	Germany	194	Patients who underwent thyroidectomy and sonoelastography and conventional US	Hyperfunctioning thyroid nodules in scintigraphy, incidental papillary microcarcinomas, sonoelastographies stored without colours and sonoelastographies with too high or too low levels at the quality indicator scale	Grey scale US elastography
Schenke, 2020[334]	Germany	140	Presence of thyroid nodules < 10 mm measured in B-mode ultrasound (independent of the histopathological size of the nodule), Kwak-TIRADS classification of the thyroid nodule during ultrasound investigation, and available histopathological results after surgery.	Incidental thyroid cancers detected at final histology and TNs > 10 mm measured with ultrasound.	Grey scale US
Schleder, 2015[337]	Germany	101	All patients with suspect thyroid nodules; final histopathology examination undertaken; CEUS and B mode US performed	None reported	Grey scale US Doppler US CEUS
Shweel, 2013[354]	Egypt	47	Surgery planned because of compressive symptoms or clinical and cytological suspicion of malignancy	Declining surgery; cystic nodules; history of external radiation; coarse marginal calcifications	Grey scale US Elastography

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
Shah, 2020 [343]	India	50	All patients above 12 years of age who were admitted with thyroid swelling in the Department of Surgery were included	Patients who did not give consent and those who were not willing for investigative procedures were excluded	Grey scale US
Shao, 2015[344]	China	297	People referred for thyroidectomy with US performed before surgery	None reported	Grey scale US Elastography
Sharma, 2019 [345]	India	48	Euthyroid cases of solitary thyroid nodule attending the Department of ENT. For the purpose of inclusion in this study, a solitary thyroid nodule (STN) was defined as a single clinically palpable discrete lesion involving either the lobe or the isthmus of the thyroid gland	Not reported	Grey scale US
Shen, 2019[346]	China	1568	Aged 18-80; nodules >5mm; underwent surgery for thyroid nodules	History of invasive procedures such as ablation or FNA; those without US data; those with any mismatch between US images and pathological results [meaning of this statement is unclear]	Grey scale US
Shi, 2020[349]	China	338	Nodules with both solid and cystic components; images complete and information on gender, age and histopathology available; no treatment history	Accepted thyroid hormone therapy; incomplete US images, clinical information or histopathology	Grey scale US
Shimura, 2005[350]	Japan	53	Patients having surgery for a thyroid nodule with prior grey-scale ultrasonography	Ultrasonograms showing cross-sections of additional nodules or did not show the nodule margin	Grey scale US
Shreyamsa, 2021[352]	India	139	All nodules measuring 4cm or less; availability of surgical histopathological findings	Patients with diffuse thyroid enlargement, autoimmune and inflammatory disorders; those unwilling to participate	Grey scale US
Shuzen, 2011[353]	China	244	None reported	None reported	Grey scale US
Skowronska, 2018[355]	Poland	52	1) preoperative US reassessment of the neck performed by two certified ultrasonographers;	None reported	Grey scale US

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
			2) preoperative FNAB of dominant or suspicious lesion; 3) surgical thyroid resection with histological evaluation.		
Sodagari, 2018[356]	Iran	114	All patients consecutively referred to the endocrine surgery clinic for thyroidectomy after being diagnosed with a cold thyroid nodule; positive family history of thyroid neoplasm	None reported	Grey scale US
Sohail, 2020 [357]	Pakistan	157	20-60 years; either gender; solid subcentimetre thyroid nodules suspected to be malignant on conventional US, with later surgical excision of the nodule and histology report	Diffuse background thyroid lesions including Grave's disease and Hashimoto's thyroiditis. Thyroid lesions on US occupying >75% of thyroid lobe, markedly calcified nodules, and complex nodules with both solid and cystic components	elastography
Stoian, 2015 #934	Romania	174	Patients with thyroid nodules who received surgery and histopathology.	None reported	Grey scale US
Stoian, 2020 [361]	Romania	261	Patients with a solid nodular goitre examined in an Ultrasound evaluation Unit between January 2016 and June 2018.	Absence of a pathology report	Grey scale US
Sui, 2016[364]	China	97	Surgery indication for palpable or impalpable thyroid nodules; ii) patients scheduled for surgical removal of thyroid nodules; iii) the final diagnosis was confirmed by histopathologic examination of resected thyroid gland tissue; and iv) the patients did not suffer from any serious allergies	None reported	Grey scale US CEUS elastography
Swan, 2019[367]	Denmark	329	>18 years; one or more thyroid nodules >10mm; indication for thyroid surgery providing histological specimens	Predominantly cystic nodules; insufficient solid area for SWE registration	Grey scale US Doppler US elastoplasty

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
Szczepanek-Parulska, 2013[370]	Poland	122	Patients with diagnosed TND admitted for thyroidectomy	None reported	Grey scale US Doppler US elastography
Taj, 2020 [373]	Pakistan	153	Patients with thyroid nodules. All nodules were papillary thyroid carcinoma (PTC). Although not stated explicitly, the study appears to have focussed only on PTC.	Patients who failed to give informed consent, have thyroid abscess and pregnant females	Grey scale US
Tan, 2010[375]	Singapore	112	All patients undergoing thyroid surgery and diagnostic evaluation in the research hospital	Diagnostic evaluation in other hospitals	Grey scale US
Tang, 2017[376]	China	323	Patients receiving thyroid surgery, US examination and elastography	None reported	Grey scale US elastography
Trimboli, 2019[380]	Switzerland	495	Adult patients on pathology database who had undergone thyroidectomy for all causes and who had had US images on the clinical information management system	Lesions <5mm; no US images	Grey scale US
Tuan,2020[381]	Vietnam	84	Patients undergoing thyroidectomy, US and SWE; nodules 10mm or more in widest dimension	Predominantly cystic nodules; inability to obtain SWE registration	Grey scale US Doppler US Elastography
Veyrieres, 2012 [383]	France	148	Patients referred for surgical care for dysthyroidism, local compression, suspicious nodule under US, prevalent adenopathy, positive or suspicious and unknown FNA	None reported	Grey scale US Doppler US Elastography
Vorlander, 2010[386]	Germany	309	Patients with solid dominant nodules on one or both thyroid lobes	Patients with an autonomous adenoma, hyperthyreosis or recurrent goitre	Grey scale US Doppler US Elastography

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
Wang, 2012 [392]	China	120	Solid lesions in one thyroid lobe	None reported	elastography
Wang, 2014 [391]	China	431	Patients with thyroid nodules of 10mm or less, located on both lobes of the thyroid, subsequently undergoing surgery	Abnormal neck anatomy; mass with eggshell calcifications	elastography
Wang, 2017 [398]	China	1011	Patients with US and surgical confirmation	(a) Patients with incomplete US information (103 nodules); (b) nodules with undetermined pathological results (26 nodules).	Grey scale US
Wang, 2018[388]	China	445	Patients with thyroid nodules undergoing US and FNA; receiving nodules confirmed by histopathologic analysis after surgery; no treatment previously performed on nodules	Nodules >40mm, nodules with a cystic component of >25%, nodules in the isthmus, nodules with calcification covering >25% of the nodule	Grey scale US
Watkins, 2021 [400]	UK	212	Patients undergoing preoperative thyroid ultrasound with eutopic thyroid histology results available	Ultrasound demonstrating diffuse thyroid disease such as thyroiditis or diffuse multinodular goitre rather than a discrete nodule or if it was not considered possible to reliably correlate imaging and histopathology, due to, for example, suboptimal image quality	Grey scale US
Wu, 2016[404]	China	970	Patients undergoing partial or total thyroidectomy because of suspicious US findings or suspicious FNA findings	None reported	Grey scale US
Wu, 2020 [405]	China	445	Patients who underwent US for nodular thyroid lesions followed by a neck CT scan within a close interval	None reported	Grey scale US
Xing, 2011[410]	China	86	Presence of single or multiple thyroid nodules whose size did not exceed 40mm	Cystic nodules, complex and partially cystic lesions, and nodules with a calcified shell; histologic findings of chronic inflammation	Grey scale US Doppler US Elastography
Xu, 2014 [412]	China	375	Thyroid nodules were detected in a prior conventional US examination and were at least 5 mm in maximal diameter. The indications for thyroid US were as follows: (a) thyroid nodule detected at US in a secondary	Nodules were excluded if they had been subjected to prior invasive procedures, signal loss or had no pathological confirmation.	Grey scale US Doppler US elastography

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
			or junior clinic, (b) discomfort or pressure symptoms in the cervical region, and (c) palpable thyroid nodules at physical examination.		
Yang, 2019 [417]	China	34	Patients undergoing thyroid surgery	Completely cystic nodes; no thyroid tissue around the node as a control; previous head and neck radiotherapy	elastography
Zhang, 2013[424]	China	155	None reported	Cystic nodules with a liquid nature; nodules near the carotid; nodules located in the margins of the thyroid gland; patients with poor breath holding capacity	elastography
Zhang, 2014[425]	China	59	Patients with nodules <10 mm, receiving US, ARFI and thyroid surgery for thyroid nodules;	Cystic lesion of a completely liquid nature	Grey scale US Doppler US Elastography
Zhang, 2016[433]	Unclear	111	At least one of: (1) diagnosed as follicular neoplasm follicular neoplasm, suspicious for malignancy or malignancy by fine needle aspiration cytology (FNAC), (2) the presence of BRAF V600E mutation, (3) Compressive symptoms or cosmetic complaints, (4) a significant increase in volume or a change in its ultrasound features during follow up, (5) diagnosed as non-diagnostic or indeterminate lesions by FNAC but showing two or more suspicious ultrasound criteria.	(1) The presence of a typical nodular goitre or scintigraphically functional (hot) thyroid nodules, (2) cystic nodules or nodules with egg shell calcifications, (3) incomplete elastography or time intensity curve data acquisition, (4) any condition of hyperthyroidism, heart failure, or severe pulmonary hypertension, or (5) previous adverse reaction to intravenous contrast agents.	Grey scale US CEUS elastography
Zhang, 2017[423]	China	128	Solid or almost solid nodules (<20% cystic); nodules at least 0.5cm, with sufficient peripheral gland at the same depth appearing normal; neither clinical treatment nor FNA/core biopsy performed before US	Unstable dynamic images on strain elastography; nodules with invalid SWV values	Grey scale US elastography
Zhang, 2021 [431]	China	241	(a) Patients aged >18 years; (b) patients with no treatment or biopsy examination before	(a) Patients with Hashimoto thyroiditis and patients with nodules more than 3 cm in size	Grey scale US

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Ultrasound tests evaluated
			ultrasound examinations; and (c) patients with solid or primarily solid (<25% cystic) nodules by conventional ultrasound examinations.	(overstepping the maximum coverage of the SWE color); (b) calcification within or around the nodules, which could result in the loss of elastic image information; (c) patients with nodules located in the isthmus or adjacent to the tracheal cartilage and the common carotid artery, as the lateral displacement of the pulsatile flow could cause the nodule hardness to produce artifacts.	Elastography
Zhuo, 2014 [438]	China	182	Patients with thyroid nodules referred for thyroidectomy	None reported	elastography

1 See Appendix D for full evidence tables

Table 3: Summary of the types of US scales used.

Scale name	Type of US measure	Description and scoring
0-5 colour grade system (Tang, 2017)	Real Time Elastography	0=normal - green in lesion and peripheral tissue; 1= green in lesion and adjacent tissue; 2= green or blue green in lesion; 3= mosaic blue green in lesion area with prominent blue colour; 4=mostly blue in lesion area; 5= all blue
0-IV colour grade system (Shuzen)	Real Time Elastography	0=normal – blue, red, green (or blue and red); I=green; II=blue; III=blue and green; IV=blue
1-6 ES scale (Mohammed, 2013; Hong, 2009)	Real Time Elastography	1=normal – low stiffness overall (green); 2=low stiffness mostly (green with blue spots); low stiffness at periphery (green) and stiff at centre (blue); 4=high stiffness over most but some green spots; 5=high stiffness over whole nodule; 6= high stiffness over nodule and surrounding tissue
1-6 scoring method (Huang 2015)	Real Time Elastography	1= normal – white; 2= white with some small amount of black; 3= white and black; 4=black with a small amount of white; 5=mostly black; 6=completely black
4 Pattern (Liu, 2019)	Real Time Elastography	1= normal – no high stiffness colour signal; 2=some high stiffness areas coded in colour, extending to interior or superficial cervical fascia; 3=some localised colour signal at nodule margin adjacent to thyroid parenchyma; 4=heterogeneous interior colours
AACE/ACE/AME	Grey scale and Doppler characteristics	1= Low risk. Cysts (fluid component >80%) mostly cystic nodules with reverberating artifacts and not associated with suspicious US signs OR isoechoic spongiform nodules, either confluent or with regular halo; 2= Intermediate risk. Slightly hypoechoic (vs. thyroid tissue) or isoechoic nodules, with ovoid-to-round shape, smooth or ill-defined margins May be present: Intranodular vascularization OR Elevated stiffness at elastography OR Macro or continuous rim calcifications OR Indeterminate hyperechoic Spots; 3=High risk. Nodules with at least 1 of the following features: Marked hypoechogenicity (vs. prethyroid muscles) OR Spiculated or lobulated margins OR Microcalcifications OR Taller-than-wide shape OR Extrathyroidal growth OR Pathologic adenopathy
ACR TIRADS	Grey scale characteristics	For the ACR TI-RADS, a cumulative score from five categories of ultrasound findings is determined (composition, echogenicity, shape, margins and presence of echogenic foci): 0 cumulative points=1 point on ACR scale; 2 points=2; 3 points=3; 4-6 points=4; ≥7 points=5
AI (artificial intelligence) TIRADS	Grey scale characteristics	Points are given for composition (3 for solid or almost completely solid but 0 for other types; if cystic and spongiform composition is observed the total score is set to zero regardless of findings in other categories) echogenicity (2 for hypoechoic but 0 for other types) shape (1 for taller than wide) margin (2 for lobulated or irregular and/or 3 for extra thyroidal extensions, but 0 for other types) and echogenic foci (2 for peripheral rim calcifications and 3 for punctate echogenic foci. Points are summed to determine the AI TIRADS level. AI TR1 is 1 point or less (benign). 2 points is AI

Scale name	Type of US measure	Description and scoring
		TR 2 (not suspicious). 3 points is AI TR 3 (mildly suspicious). 4-6 points is AI TR 4 (moderately suspicious). 7 points or more is AI TR 5 (highly suspicious).
Asteria 1-4 colour scale	Real Time Elastography	1= soft - Red (or green in some papers); 2= intermediate soft/hard - mostly red with some blue (or green with some red); 3=intermediate hard/soft – mostly blue, with some red (or red with some green); 4= hard – all blue (or all red)
ATA TIRADS	Grey scale characteristics	Cyst='Benign'; Spongiform, partially cystic no suspicious features='very low suspicion'; hyperechoic solid regular margin, OR isoechoic solid regular margin OR partially cystic with eccentric solid area ='low suspicion'; hypoechoic solid regular margin='intermediate suspicion'; microcalcifications, hypoechoic nodules and irregular margins OR hypoechoic irregular margins or hypoechoic and taller than wide OR hypoechoic, irregular margins and ETE OR hypoechoic, interrupted rim calcification with soft tissue extrusions OR nodule with irregular margins and suspicious lymph nodes='high suspicion'
BTA	Grey scale and Doppler characteristics	U1=normal; U2= benign A. Halo, isoechoic, mildly Hyperechoic B. Cystic change ± ring-down sign (colloid) C. Microcystic/spongiform D-E. Peripheral eggshell calcification F. Peripheral vascularity; U3= Indeterminate/equivocal A. Homogeneous, markedly hyperechoic, solid, halo (follicular lesion) B. Hypoechoic (?), equivocal echogenic foci, cystic change (irregular) C. Mixed/central vascularity; U4= suspicious A. Solid, hypoechoic (cf. thyroid) B. Solid, very hypoechoic (cf. strap muscle) C. Disrupted peripheral calcification, hypoechoic D. Lobulated outline; U5= Malignant A. Solid, hypoechoic, lobulated/irregular outline, microcalcification (papillary carcinoma?) B. Solid, hypoechoic, lobulated/ irregular outline, globular calcification (medullary carcinoma?) C. Intranodular vascularity D. Shape tall>wide (AP>TR) E. Characteristic associated lymphadenopathy
C TIRADS	Grey scale characteristics	Positive features are vertical orientation (+1), solid composition (+1), markedly hypoechoic (+1), microcalcifications (+1), ill-defined / irregular margin of extrathyroidal extension (+1). The negative feature is a comet tail artefact (-1). The scores are summed to yield points. No nodules is C-TR1, -1 point is C-TR2 (estimated malignancy risk of 0%), 0 points is C-TR3 (estimated malignancy risk of <2%), 1 point is C-TR4A (estimated malignancy risk of 2-10%), 2 points is C-TR4B (estimated malignancy risk of 10-50%), 3-4 points is C-TR4C (estimated malignancy risk of 50-90%, 5 points is C-TR5 (estimated malignancy risk of >90%). C-TR6 is reserved for nodules proved to be malignant on histopathology.
EU TIRADS	Grey scale characteristics	No nodules=1; cyst spongiform=2; iso-/hyperechoic, no suspicious feature =3; Oval, smooth margins, mild hypoechoic, no suspicious feature = 4; Suspicious features (min 1): irregular shape, irregular margins, microcalcifications, marked hypoechoic =5
French TIRADS	Grey scale characteristics and elastography	1= no nodules; 2=cyst/isolated macrocalcification/spongiform=2; oval, regular margins, iso/hyperechoic =3; oval, regular margins, mild hypoechoic=4A; High suspicion features (1-2) of: taller than wide, irregular margins, marked hypoechoic, microcalcifications, stiff on elastography=4B; high suspicion features (3-5) of taller than wide, irregular margins, marked hypoechoic, microcalcifications, stiff on elastography and/or lymph node metastasis=5

Scale name	Type of US measure	Description and scoring
Horvath TIRADS	Grey scale and Doppler characteristics	No nodules=1; Colloid/spongiform/mixed isoechoic=2; Hashimoto pseudo-nodule=3; hyper, iso, or hypoechoic nodule + thin capsule OR Hypoechoic, ill-defined borders, no calcifications OR hypervascularized, nodule with thick capsule and calcifications =4A; Hypoechoic, irregular shape and margins, penetrating vessels ±calcifications=4B; Iso/hypoechoic, nonencapsulated multiple peripheral microcalcifications hypervascular OR Nonencapsulated, isoechoic mixed, hypervascular ±calcifications=5
ITOH 1-5 colour scale	Real Time Elastography	1= normal – green overall; 2=mosaic pattern of green and red; 3=centre of lesion red, rest green; 4=lesion red overall; lesion and surrounding tissues were red
Korean TIRADS	Grey scale characteristics	1= no nodule; 2=spongiform or partially cystic nodule with comet tail artefact or pure cyst; 3= partially cystic or isohyperechoic nodule without any of 3 suspicious US features (microcalcification, taller than wide or spiculated/microlobulated margin); 4=solid hypoechoic nodule without any of 3 suspicious US features OR partially cystic or isohyperechoic nodule with any of 3 suspicious US features; 5=solid hypoechoic nodule with any of the 3 suspicious US features
Kwak TIRADS	Grey scale characteristics	Normal exam=1; predominantly cystic peripheral halo=2; no suspicious US features=3; one suspicious US feature=4a; 2 suspicious US features=4b; 3 or 4 suspicious US features=4c; Five suspicious US features including solid, hypoechogenicity, microlobulated or irregular margins, microcalcifications, taller than-wide shape=5
Park TIRADS	Grey scale characteristics	Normal exam=0; cystic predominant, peripheral halo=1; Circumscribed margin, solid predominant, heterogeneous echotexture, iso- to hyperechogenicity, eggshell or macrocalcification=2; Homogeneous echotexture, hypoechogenicity, circumscribed margin, solid, taller, without other US findings suggestive of malignancy=3; One or two US findings suggestive of malignancy, such as markedly hypoechoic, microcalcification, not-circumscribed margin, and lymph node abnormality=4; More than three US findings suggestive of malignancy, such as markedly hypoechoic, microcalcification, not-circumscribed margin, and lymph node abnormality=5.
Pathirana scale	Grey scale characteristics	Scoring system from 0-14, scored as follows: AP> width=2 points, AP=width=1 point, AP<width=0 points Solid=2 points, mixed=1 point, cystic thick wall=0 points Hypoechoic=0points, isoechoic=1point, hyperechoic=0 points Ill-defined margins= 2 points, well defined margins= 0 points Microcalcification 2 points, large coarse calcification 1 point, no calcification 0 points Thick halo= 2 points, thin halo= 1 point, no halo= 0 points Internal vascular flow= 2 points, peripheral flow= 1 point, no flow= 0 points

Scale name	Type of US measure	Description and scoring
R status	Grey scale and Doppler characteristics	R1= Definitely benign: simple cyst, spongiform nodule, or predominantly cystic nodule (>75% cystic) with no suspicious intranodular solid focus (e.g. hypervascular nodule or focus of intranodular microcalcification) & with comet-tail artefacts; R2= Probably benign: part cystic & part solid (25–75% cystic), or solid & hyperechoic (with no suspicious solid component); R3= Indeterminate (<25% cystic): solid & isoechoic, solid with heterogeneous echo-texture, or solid with slight hypoechogenicity (less than adjacent strap muscle); R4 Suspicious: any solid, markedly hypoechoic nodule (hypoechoicity equal to or more than that of strap muscle); or any solid markedly hypoechoic, slightly hypoechoic or isoechoic nodule with internal calcification, hypervascularity, irregular margins or an anteroposterior diameter larger than transverse diameter; R5 Malignant: any solid, markedly hypoechoic nodule displaying 2 of the above features
Rago 1-3 scale	Real Time Elastography	I=soft (blue); II=intermediate stiffness (blue and green); III completely or mostly stiff nodule (yellow/red)
Rago 1-5 scale	Real Time Elastography	1= entirely elastic (normal); 2=mostly soft, with areas of some stiffness; 3=elastic on edges and rigid in centre; 4= increased stiffness in whole nodule; 5=entirely stiff with stiff surroundings
RGB (red, green and blue) 0-4 colour scale	Real Time Elastography	0= soft - red, green and blue in lesion area; 1=uniform green in lesion area; green in >50% of lesion area; 3=blue in 50-90% of the total area; 4=hard tissue - blue in most of the lesion area and in >90% of the total area. In this scale, as in all the others, the hardness of tissue increases with the grade.
Russ TIRADS	Grey scale characteristics	1=normal exam; 2=simple cyst spongiform nodule ‘white knight’ isolated macrocalcification nodular hyperplasia=2; No sign of high suspicion: Regular shape and borders, no microcalcifications and Isoechoic or Hyperechoic=3; No sign of high suspicion -Mildly hypoechoic=4a; One or two signs -No metastatic- lymph node=4b; Three to five signs including marked hypoechogenicity, microlobulated or irregular margins, microcalcifications, taller than-wide shape and/or -metastatic -lymph node=5
SN-US	Grey scale characteristics	1= solid thyroid nodules with ≥3 US features of benignancy and no malignant or borderline US features; 2= probably benign (SN-US class II): solid thyroid nodules with 1 or 2 US features of benignancy and no malignant or borderline US features; 3= borderline (SN-US class III): solid thyroid nodules with ≥1 borderline US feature and no US features of malignancy, regardless of benign US features; 4= possibly malignant (SN-US class IV): solid thyroid nodules with 1 US feature of malignancy, regardless of borderline or benign US features; 5= malignant (SN-US class V): solid thyroid nodules with >2 US features of malignancy, regardless of borderline or benign US features
Thyroid Multimodal Imaging Comprehensive Risk Stratification System (TMC-RSS) categories	Grey scale and Doppler characteristics	Malignant characteristics: Plus (+) 3 points for: ES score 3/4, malignant nodes. (+) 1 point for: Taller than wider, microcalcification, hypoechogenicity, solid composition, ill-defined margins, central +/- peripheral vascularity. (+) 0.5 point for: irregular halo, size > 1 cm. Benign characteristics: Minus (-) 3 points for: Purely cystic, ES score-1; (-) 1 point for: spongiform, comet tail artefact, complete halo. (-) 0.5 point for: peripheral vascularity. Final TMC-RSS Score calculated by addition of all the points.

Scale name	Type of US measure	Description and scoring
Virtual Touch Imaging I-VI grade	Shear wave Elastography	No detailed description of the scales is given in the literature as far as is known, but as the grade increases the level of tissue hardness increases. Therefor Grade I denotes normal sift tissue and Grade VI denotes very hard (probably malignant) tissue

1

1.11.6. Summary of the evidence concerning detection of nodule malignancies with ultrasound

In the tables that follow, the index test will be defined by the definition of the positive test derived from that index test (the index test finding that would be intended to 'detect' thyroid cancer).

Table 4: Summary of evidence relating to individual grey-scale US characteristics

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Taller than wide	40	11,321	Pooled sensitivity (95% credible intervals): 0.402 (0.3241-0.4832)	Pooled specificity (95% credible intervals): 0.9367 (0.9107 – 0.9567)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	None[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	None[d]	VERY LOW
Solitary nodule	10	3,343	Pooled sensitivity (95% credible intervals): 0.4734 (0.2928-0.6566)	Pooled specificity (95% credible intervals): 0.6385 (0.4035 – 0.8216)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	None[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	Serious[d]	VERY LOW
Solidity	26	9,931	Pooled sensitivity (95% credible	Pooled specificity (95%	**Sensitivity**				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
			intervals): 0.8988 (0.8331-0.9443)	credible intervals): 0.4098 (0.3144 – 0.5062)	Very serious[a]	none[b]	serious[c]	Serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	Serious[d]	VERY LOW
microcalcifications	53	12,445	Pooled sensitivity (95% credible intervals): 0.540 (0.4796-0.5984)	Pooled specificity (95% credible intervals): 0.8864 (0.8476 – 0.918)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
hypoechoicity	39	7,396	Pooled sensitivity (95% credible intervals): 0.8046 (0.744-0.8557)	Pooled specificity (95% credible intervals): 0.615 (0.5535 – 0.6747)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
marked hypoechoicity	15	5,343	Pooled sensitivity (95% credible	Pooled specificity (95%	**Sensitivity**				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
			intervals): 0.3626 (0.2393-0.5129)	credible intervals): 0.9215 (0.8478 – 0.9622)	Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
Hypoechoic or markedly hypoechoic	6	1985	Pooled sensitivity (95% credible intervals): 0.8862 (0.6952-0.9663)	Pooled specificity (95% credible intervals): 0.6321 (0.4517 – 0.7863)	Sensitivity				
					Very serious[a]	none[b]	serious[c]	Very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
Poorly defined borders	23	4,465	Pooled sensitivity (95% credible intervals): 0.6925 (0.5517-0.8077)	Pooled specificity (95% credible intervals): 0.805 (0.701 – 0.885)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	Serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
Absent halo sign	15	3,085	Pooled sensitivity (95% credible	Pooled specificity (95%	**Sensitivity**				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
			intervals): 0.7584 (0.6066-0.8661)	credible intervals): 0.6192 (0.4329 – 0.781)	Very serious[a]	none[b]	serious[c]	Serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
Irregular borders	26	6,842	Pooled sensitivity (95% credible intervals): 0.5435 (0.44345-0.6496)	Pooled specificity (95% credible intervals): 0.8698 (0.8123 – 0.9134)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
Heterogeneous texture	14	2,564	Pooled sensitivity (95% credible intervals): 0.6518(0.4784-0.797)	Pooled specificity (95% credible intervals): 0.4629 (0.3065 – 0.6304)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
Macrocalcifications	4	1,695	Pooled sensitivity (95% credible	Pooled specificity (95%	**Sensitivity**				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
			intervals): 0.2325 (0.112-0.4202)	credible intervals): 0.8696 (0.7181 – 0.9448)	Serious[a]	none[b]	none[c]	none[d]	MODERATE
					Specificity				
					Serious[a]	none[b]	none[c]	none[d]	MODERATE
nodule diameter of 10mm or less	2	591	0.24 [0.14, 0.35]; 0.40 [0.35, 0.46]	0.96 [0.89, 0.99]; 0.82 [0.73, 0.89]	**Sensitivity**				
					Serious[a]	none[b]	serious[c]	none[d]	LOW
					Specificity				
					Serious[a]	none[b]	serious[c]	none[d]	LOW
nodule diameter of 20mm or less	3	1,029	Pooled sensitivity (95% credible intervals): 0.7467 (0.3628-0.9409)	Pooled specificity (95% credible intervals): 0.5355 (0.2304 – 0.819)	**Sensitivity**				
					Serious[a]	none[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Serious[a]	none[b]	serious[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
nodule diameter of 36mm or less	1	114	0.82 [0.68, 0.92]	0.46 [0.34, 0.59]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
nodule diameter of 40mm or less	2	1,004	0.60 [0.50, 0.69]; 0.79 [0.60, 0.92]	0.31 [0.28, 0.34]; 0.09 [0.03, 0.20]	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW

(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) Indirectness was assessed using the QUADAS-2 checklist items referring to applicability. The evidence was downgraded by 1 increment if the majority of studies were seriously indirect.
(c) Inconsistency was assessed by visual inspection of the sensitivity/specificity plots, or data (if 2 studies). The evidence was downgraded by 1 increment if there was no overlap of 95% confidence intervals. For single studies no evaluation was made and 'NA' was recorded.
(d) Imprecision was assessed based on inspection of the confidence region 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. The evidence was downgraded by 1 increment when the confidence interval around the point estimate crossed one of the clinical thresholds (0.90 or 0.80 for sensitivity and 0.5 or 0.1 for specificity), and downgraded by 2 increments when the confidence interval around the point estimate crossed both of the clinical thresholds. The upper clinical threshold marked the point above which recommendations would be possible, and the lower clinical threshold marked the point below which the tool would be regarded as of little clinical use.

Table 5: Summary of evidence relating to combined grey-scale US characteristics

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
microcalcifications AND absent halo	5	525	Pooled sensitivity (95% credible intervals):0.524 (0.2772-0.7601)	Pooled specificity (95% credible intervals): 0.9223 (0.7907 – 0.9783)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very erious[a]	none[b]	serious[c]	none[d]	VERY LOW
hypoechoicity AND absent halo	5	525	Pooled sensitivity (95% credible intervals):0.5062(0.3117-0.6981)	Pooled specificity (95% credible intervals): 0.8662 (0.7115 – 0.9499)	**Sensitivity**				
					Very serious[a]	none[b]	none[c]	none[d]	LOW
					Specificity				
					Serious[a]	none[b]	serious[c]	none[d]	LOW
hypoechoicity AND microcalcifications	5	525	Pooled sensitivity (95% credible intervals):0.4115 (0.2372-0.6082)	Pooled specificity (95% credible intervals): 0.8965 (0.7854 – 0.9592)	**Sensitivity**				
					Serious[a]	none[b]	none[c]	none[d]	MODERATE
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Serious[a]	none[b]	none[c]	none[d]	MODERATE
hypoechoicity AND microcalcifications AND absent halo	1	80	0.25 [0.07, 0.52]	0.97 [0.89, 1.00]	**Sensitivity**				
					Serious[a]	none[b]	NA[c]	none[d]	MODERATE
					Specificity				
					Serious[a]	none[b]	NA[c]	none[d]	MODERATE
hypoechoicity OR microcalcifications	1	80	0.69 [0.41, 0.89]	0.80 [0.68, 0.89]	**Sensitivity**				
					Serious[a]	none[b]	NA[c]	serious[d]	LOW
					Specificity				
					Serious[a]	none[b]	NA[c]	none[d]	MODERATE
hypoechoicity OR microcalcifications OR absent halo	1	80	0.81 [0.54, 0.96]	0.70 [0.58, 0.81]	**Sensitivity**				
					Serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Serious[a]	none[b]	NA[c]	none[d]	MODERATE
Microcalcifications OR absent halo	1	80	0.75 [0.48, 0.93]	0.77 [0.64, 0.86]	**Sensitivity**				
					Serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Serious[a]	none[b]	NA[c]	none[d]	MODERATE
At least one US sign (any allowed from a variety of selections that varied by study)	14	2,134	Pooled sensitivity (95% credible intervals):0.8182 (0.6666-0.912)	Pooled specificity (95% credible intervals): 0.7275 (0.5767 – 0.8433)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
At least two US signs (any allowed from a variety of selections that varied by study)	3	776	Pooled sensitivity (95% credible intervals):0.6393 (0.1065-0.9565)	Pooled specificity (95% credible intervals): 0.9046 (0.4579 – 0.994)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
At least three US signs (any allowed from a variety of selections that varied by study)	4	928	Pooled sensitivity (95% credible intervals):0.5124 (0.1092-0.9148)	Pooled specificity (95% credible intervals): 0.9708 (0.8426 – 0.9975)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
At least four US signs	1	512	0.57 [0.50, 0.64]	0.97 [0.95, 0.99]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
At least five US signs	1	512	0.37 [0.30, 0.44]	0.99 [0.97, 1.00]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
ALL of: Taller than wide, Sub capsular, Intense hypo echoic, Calcification, Suspect lymph nodes	1	261	0.54 [0.41, 0.68]	0.95 [0.91, 0.97]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
ALL of: Taller than wide, Sub capsular, Intense hypo echoic, Calcification, Suspect lymph nodes OR ALL of hypoechoic, sub-capsular position, inhomogeneity	1	261	0.89 [0.78, 0.96]	0.49 [0.42, 0.56]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	Very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Blurred margins plus at least one of the following: hypoechoicity, microcalcifications or taller than wide	3	207	Pooled sensitivity (95% credible intervals):0.9834 (0.9053-0.9984)	Pooled specificity (95% credible intervals): 0.6254 (0.2083 – 0.9125)	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
hypoechoicity plus at least one of the following: blurred margins, microcalcifications or taller than wide	3	207	Pooled sensitivity (95% credible intervals):0.9326 (0.7606-0.9855)	Pooled specificity (95% credible intervals): 0.7148 (0.3838 – 0.91)	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	serious[d]	VERY LOW
microcalcifications plus at least one of the following: blurred margins, hypoechoicity or taller than wide	3	207	Pooled sensitivity (95% credible intervals):0.4416 (0.1347-0.7927)	Pooled specificity (95% credible intervals): 0.938 (0.7642 – 0.9914)	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
microlobulated or irregular margins	3	1,795	Pooled sensitivity (95% credible intervals):0.3049 (0.1193-0.5898)	Pooled specificity (95% credible intervals): 0.9686 (0.8791 – 0.9947)	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
infiltrative/ETE or lobulated or irregular	2	1,309	0.80 [0.74, 0.86]; 0.71 [0.67, 0.75]	0.79 [0.70, 0.85]; 0.86 [0.83, 0.89]	**Sensitivity**				
					Very serious[a]	none[b]	none[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	none[c]	none[d]	LOW
spiculated or blurred/ill-defined margins	2	211	0.92 [0.80, 0.98]; 0.50 [0.25, 0.75]	0.84 [0.76, 0.91]; 0.96 [0.86, 1.00]	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	none[c]	none[d]	LOW
spiculated or microlobulated margins	1	338	0.32 [0.20, 0.47]	0.97 [0.94, 0.99]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
hypoechoic or microlobulated margins	1	93	0.78 [0.65, 0.88]	0.65 [0.46, 0.80]	**Sensitivity**				
					Serious[a]	none[b]	NA[c]	serious[d]	LOW
					Specificity				
					Serious[a]	none[b]	NA[c]	serious[d]	LOW

(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) Indirectness was assessed using the QUADAS-2 checklist items referring to applicability. The evidence was downgraded by 1 increment if the majority of studies were seriously indirect.
(c) Inconsistency was assessed by visual inspection of the sensitivity/specificity plots, or data (if 2 studies). The evidence was downgraded by 1 increment if there was no overlap of 95% confidence intervals. For single studies no evaluation was made and 'NA' was recorded.
(d) Imprecision was assessed based on inspection of the confidence region 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. The evidence was downgraded by 1 increment when the confidence interval around the point estimate crossed one of the clinical thresholds (0.90 or 0.80 for sensitivity and 0.5 or 0.1 for specificity), and downgraded by 2 increments when the confidence interval around the point estimate crossed both of the clinical thresholds. The upper clinical threshold marked the point above which recommendations would be possible, and the lower clinical threshold marked the point below which the tool would be regarded as of little clinical use.

Table 6: Summary of evidence relating to scales (such as TIRADS) based primarily on grey-scale US characteristics (though some include Doppler and elastography characteristics). See Table 3 for details of the scales and scoring criteria

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
ACR TIRADS of 2 or more	11	6445	Pooled sensitivity (95% credible intervals): 0.9978 (0.9935-0.9994)	Pooled specificity (95% credible intervals): 0.04019 (0.0193 – 0.0777)	**Sensitivity**				
					Very serious[a]	Serious[b]	none[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	none[c]	none[d]	VERY LOW
ACR TIRADS of 3 or more	13	8155	Pooled sensitivity (95% credible intervals): 0.9886 (0.9633-0.9975)	Pooled specificity (95% credible intervals): 0.2035 (0.1083 – 0.3475)	**Sensitivity**				
					Very serious[a]	Serious[b]	none[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	serious[c]	serious[d]	VERY LOW
ACR TIRADS of 4 or more	16	8577	Pooled sensitivity (95% credible intervals): 0.9074(0.8254-0.9545)	Pooled specificity (95% credible intervals): 0.5949 (0.4671 – 0.7127)	**Sensitivity**				
					Very serious[a]	Serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	serious[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
ACR TIRADS of 5	15	8456	Pooled sensitivity (95% credible intervals): 0.5048 (0.3609-0.6403)	Pooled specificity (95% credible intervals): 0.9368 (0.8731 – 0.9723)	**Sensitivity**				
					Very serious[a]	Serious[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	serious[c]	none[d]	VERY LOW
EU TIRADS of 2 or more	1	305	1.00 [0.96, 1.00]	0.00 [0.00, 0.03]	**Sensitivity**				
					Very serious[a]	Serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	NA[c]	none[d]	VERY LOW
EU TIRADS of 3 or more	10	6730	Pooled sensitivity (95% credible intervals): 0.998 (0.9949-0.9994)	Pooled specificity (95% credible intervals): 0.03443 (0.01319 – 0.0777)	**Sensitivity**				
					Very serious[a]	Serious[b]	none[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	serious[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
EU TIRADS of 4 or more	11	6870	Pooled sensitivity (95% credible intervals): 0.9328 (0.8346-0.9747)	Pooled specificity (95% credible intervals): 0.572 (0.3886 – 0.7423)	**Sensitivity**				
					Very serious[a]	Serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	serious[c]	serious[d]	VERY LOW
EU TIRADS of 5	10	6730	Pooled sensitivity (95% credible intervals): 0.7803 (0.5922-0.901)	Pooled specificity (95% credible intervals): 0.848 (0.697 – 0.931)	**Sensitivity**				
					Very serious[a]	Serious[b]	serious[c]	Very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	serious[c]	none[d]	VERY LOW
Kwak TIRADS of 3 or more	5	6507	Pooled sensitivity (95% credible intervals):0.9994 (0.9872-0.09999)	Pooled specificity (95% credible intervals): 0.02852 (0.003976 – 0.1693)	**Sensitivity**				
					Very serious[a]	Serious[b]	none[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	serious[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Kwak TIRADS of 4a or more	8	6922	Pooled sensitivity (95% credible intervals):0.9924 (0.9826-0.9969)	Pooled specificity (95% credible intervals): 0.2698 (0.1426 – 0.4291)	**Sensitivity**				
					Very serious[a]	Serious[b]	none[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	serious[c]	none[d]	VERY LOW
Kwak TIRADS of 4b or more	10	7574	Pooled sensitivity (95% credible intervals):0.9422 (0.78422-0.9823)	Pooled specificity (95% credible intervals): 0.6102 (0.4932 – 0.717)	**Sensitivity**				
					Very serious[a]	Serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	serious[c]	serious[d]	VERY LOW
Kwak TIRADS of 4c or more	9	7313	Pooled sensitivity (95% credible intervals):0.7809 (0.5152-0.9251)	Pooled specificity (95% credible intervals): 0.8793(0.741 – 0.9529)	**Sensitivity**				
					Very serious[a]	Serious[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	serious[c]	none[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Kwak TIRADS of 5	7	6812	Pooled sensitivity (95% credible intervals):0.1065(0.05256-0.1952)	Pooled specificity (95% credible intervals): 0.9894 (0.9685 – 0.9972)	**Sensitivity**				
					Very serious[a]	Serious[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	none[c]	none[d]	VERY LOW
ATA 'very low suspicion' or higher	3	1511	Pooled sensitivity (95% credible intervals):0.9984 (0.9844-0.9999)	Pooled specificity (95% credible intervals): 0.02199 (0.0017 – 0.1538)	**Sensitivity**				
					Very serious[a]	Serious[b]	none[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	serious[c]	serious[d]	VERY LOW
ATA 'low suspicion' or higher	8	6241	Pooled sensitivity (95% credible intervals):0.9914 (0.972-0.998)	Pooled specificity (95% credible intervals): 0.1464 (0.05965–0.2883)	**Sensitivity**				
					Very serious[a]	Serious[b]	none[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	serious[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
ATA 'intermediate suspicion' or higher	9	6408	Pooled sensitivity (95% credible intervals):0.8708(0.7249-0.9453)	Pooled specificity (95% credible intervals): 0.7008 (0.5731 – 0.8042)	**Sensitivity**				
					Very serious[a]	Serious[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	serious[c]	none[d]	VERY LOW
ATA 'high suspicion'	9	6408	Pooled sensitivity (95% credible intervals):0.664 (0.4304-0.8393)	Pooled specificity (95% credible intervals): 0.9317 (0.871 – 0.966)	**Sensitivity**				
					Very serious[a]	Serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	Serious[b]	serious[c]	none[d]	VERY LOW
Horvath TIRADS 3 or more	3	1,646	Pooled sensitivity (95% credible intervals):0.999 (0.9783-1.0)	Pooled specificity (95% credible intervals): 0.24 (0.04586 – 0.6849)	**Sensitivity**				
					Very serious[a]	none[b]	none[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	very serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Horvath TIRADS 4a or more	4	1,909	Pooled sensitivity (95% credible intervals):0.9759 (0.7304-0.9987)	Pooled specificity (95% credible intervals): 0.6447(0.2024 – 0.9329)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
Horvath TIRADS 4b or more	3	1,646	Pooled sensitivity (95% credible intervals):0.962 (0.643-0.9976)	Pooled specificity (95% credible intervals): 0.7236 (0.4148 – 0.9107)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
Horvath TIRADS 4c or more	2	1,513	0.81 [0.76, 0.86]; 0.58 [0.54, 0.63]	0.94 [0.90, 0.97]; 0.94 [0.92, 0.96]	**Sensitivity**				
					serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					serious[a]	none[b]	serious[c]	none[d]	LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Horvath TIRADS 5	3	1,646	Pooled sensitivity (95% credible intervals):0.301 (0.09403-0.6651)	Pooled specificity (95% credible intervals): 0.9837 (0.7321 – 0.9993)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
Park TIRADS 2 or more	1	1,011	1.00 [0.98, 1.00]	0.36 [0.32, 0.40]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Park TIRADS 3 or more	1	1,011	0.97 [0.95, 0.98]	0.71 [0.67, 0.75]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Park TIRADS 4 or more	1	1,011	0.83 [0.80, 0.87]	0.86 [0.83, 0.89]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Park TIRADS 5	1	1,011	0.12 [0.09, 0.15]	1.00 [0.99, 1.00]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Russ TIRADS 3 or more	3	1,215	Pooled sensitivity (95% credible intervals):0.9966(0.9248-0.999)	Pooled specificity (95% credible intervals): 0.2374 (0.0221 – 0.8196)	**Sensitivity**				
					Very serious[a]	none[b]	none[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	very serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Russ TIRADS 4a or more	3	1,585	Pooled sensitivity (95% credible intervals):0.9766(0.788-0.9984)	Pooled specificity (95% credible intervals): 0.4411 (0.1068 – 0.8351)	**Sensitivity**				
					Very serious[a]	none[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
Russ TIRADS 4b or more	3	1,215	Pooled sensitivity (95% credible intervals):0.8995(0.6752-0.9778)	Pooled specificity (95% credible intervals): 0.9414 (0.6859 – 0.9936)	**Sensitivity**				
					very serious[a]	none[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					very serious[a]	none[b]	none[c]	none[d]	LOW
Russ TIRADS 4c or more	1	30	0.57 [0.18, 0.90]	1.00 [0.85, 1.00]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Russ TIRADS 5	2	1,185	0.55 [0.36, 0.74]; 0.26 [0.22, 0.30]	0.99 [0.96, 1.00]; 0.99 [0.98, 1.00]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
French TIRADS 3 and above	1	133	1.00 [0.90, 1.00]	0.06 [0.02, 0.13]	**Sensitivity**				
					very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
French TIRADS 4a and above	1	133	1.00 [0.90, 1.00]	0.26 [0.17, 0.35]	**Sensitivity**				
					very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					very serious[a]	none[b]	NA[c]	none[d]	LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
French TIRADS 4b and above	2	301	0.91 [0.77, 0.98]; 0.83 [0.66, 0.93]	0.83 [0.74, 0.90]; 0.95 [0.89, 0.98]	**Sensitivity**				
					very serious[a]	none[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					very serious[a]	none[b]	none[c]	none[d]	LOW
French TIRADS 4c and above	1	168	0.46 [0.29, 0.63]	0.98 [0.95, 1.00]	**Sensitivity**				
					very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					very serious[a]	none[b]	NA[c]	none[d]	LOW
French TIRADS 5	1	133	0.51 [0.34, 0.69]	0.96 [0.90, 0.99]	**Sensitivity**				
					very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					very serious[a]	none[b]	NA[c]	none[d]	LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Korean TIRADS 3 and above	2	2239	1.00 [0.99, 1.00] 1.00 [0.99, 1.00]	0.04 [0.02, 0.05] 0.11 [0.09, 0.13]	Sensitivity				
					very serious[a]	none[b]	none[c]	none[d]	LOW
					Specificity				
					very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
Korean TIRADS 4 and above	3	2407	0.8633 (0.3078-0.9885)	0.6949(0.2546-0.942)	**Sensitivity**				
					very serious[a]	none[b]	serious[c]	Very serious[d]	VERY LOW
					Specificity				
					very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
Korean TIRADS 5	3	2407	0.5225(0.1046-0.906)	0.9329(0.4501-0.9965)	**Sensitivity**				
					very serious[a]	none[b]	serious[c]	Very serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
C TIRADS 3 or more	1	1096	1.00 [0.99, 1.00]	0.06 [0.05, 0.09]	**Sensitivity**				
					serious[a]	serious[b]	NA[c]	none[d]	LOW
					Specificity				
					serious[a]	serious[b]	NA[c]	none[d]	LOW
C TIRADS 4a or more	2	1484	1.00 [0.98, 1.00] 1.00 [0.98, 1.00]	0.18 [0.12, 0.25] 0.34 [0.30, 0.38]	**Sensitivity**				
					serious[a]	serious[b]	none[c]	none[d]	LOW
					Specificity				
					serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
C TIRADS 4b or more	2	1484	0.97 [0.94, 0.99] 0.93 [0.90, 0.95]	0.56 [0.48, 0.64] 0.54 [0.50, 0.58]	**Sensitivity**				
					serious[a]	serious[b]	none[c]	none[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					serious[a]	serious[b]	none[c]	none[d]	LOW
C TIRADS 4c or more	2	1484	0.79 [0.73, 0.84] 0.66 [0.61, 0.70]	0.75 [0.68, 0.82] 0.82 [0.79, 0.85]	**Sensitivity**				
					serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					serious[a]	serious[b]	none[c]	none[d]	LOW
C TIRADS 5	2	1484	0.08 [0.05, 0.12] 0.02 [0.01, 0.04]	0.99 [0.95, 1.00] 1.00 [0.99, 1.00]	**Sensitivity**				
					serious[a]	serious[b]	none[c]	none[d]	LOW
					Specificity				
					serious[a]	serious[b]	none[c]	none[d]	LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
AI TIRADS 2 and above	1	218	0.96 [0.89, 0.99]	0.31 [0.24, 0.40]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	LOW
AI TIRADS 3 and above	1	218	0.95 [0.87, 0.99]	0.34 [0.26, 0.42]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	LOW
AI TIRADS 4 and above	1	218	0.74 [0.63, 0.83]	0.60 [0.51, 0.68]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
AI TIRADS 5	1	218	0.49 [0.38, 0.61]	0.82 [0.74, 0.88]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	LOW
					serious[a]	serious[b]	NA[c]	none[d]	LOW
					Specificity				
					serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
TIRADS (0-10 version) 3 or above	1	298	1.00 [0.98, 1.00]	0.09 [0.05, 0.16]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
TIRADS (0-10 version) 4 or above	1	298	0.99 [0.96, 1.00]	0.31 [0.23, 0.40]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
TIRADS (0-10 version) 5 or above	1	298	0.91 [0.86, 0.95]	0.66 [0.57, 0.74]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
TIRADS (0-10 version) 6 or above	1	298	0.90 [0.84, 0.94]	0.74 [0.65, 0.81]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
TIRADS (0-10 version) 7 or above	1	298	0.75 [0.67, 0.81]	0.86 [0.78, 0.92]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
TIRADS (0-10 version) 8 or above	1	298	0.64 [0.56, 0.71]	0.94 [0.88, 0.98]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
TIRADS (0-10 version) 9 or above	1	298	0.61 [0.53, 0.68]	0.95 [0.90, 0.98]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
TIRADS (0-10 version) 10	1	298	0.27 [0.21, 0.34]	1.00 [0.97, 1.00]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
AACE/ACE/AME 2016 'Intermediate' or higher	2	1289	1.00 [0.95, 1.00] 1.00 [0.99, 1.00]	0.08 [0.03, 0.16] 0.05 [0.04, 0.07]	**Sensitivity**				
					serious[a]	none[b]	none[c]	none[d]	MODERATE
					Specificity				
					serious[a]	none[b]	none[c]	serious[d]	LOW
AACE/ACE/AME 2016 'high'.	2	1289	0.54 [0.42, 0.67] 0.42 [0.37, 0.46]	0.92 [0.84, 0.97] 0.92 [0.84, 0.97]	**Sensitivity**				
					serious[a]	serious[b]	none[c]	none[d]	LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					seriousa	seriousb	nonec	noned	LOW
BTA intermediate suspicion and higher (3 and higher)	3	686	0.9368 (0.7049, 0.9929)	0.3974 (0.1481 - 0.7167]	**Sensitivity**				
					very seriousa	seriousb	seriousc	very seriousd	VERY LOW
					Specificity				
					very seriousa	seriousb	seriousc	seriousd	VERY LOW
BTA 'suspicious' and higher (4 and higher)	3	686	0.67 (0.3255- 0.9021)	0.758 [0.385 - 0.9367)	**Sensitivity**				
					very seriousa	seriousb	NAc	Very seriousd	VERY LOW
					Specificity				
					very seriousa	seriousb	NAc	seriousd	VERY LOW
BTA 'malignant' (5)	3	686	0.3789 (0.1395- 0.6906)	0.9265 (0.7493- 0.9828)	**Sensitivity**				
					very seriousa	seriousb	NAc	noned	VERY LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
SN-US class II and above	1	505	0.96 [0.93, 0.98]	0.52 [0.45, 0.60]	**Sensitivity**				
					very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
SN-US class III and above	1	505	0.86 [0.82, 0.90]	0.83 [0.77, 0.88]	**Sensitivity**				
					very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					very serious[a]	none[b]	NA[c]	none[d]	LOW
SN-US class IV and above	1	505	0.77 [0.72, 0.82]	0.91 [0.86, 0.94]	**Sensitivity**				
					very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					very serious[a]	none[b]	NA[c]	none[d]	LOW
SN-US class V and above	1	505	0.57 [0.51, 0.62]	0.99 [0.96, 1.00]	**Sensitivity**				
					very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					very serious[a]	none[b]	NA[c]	none[d]	LOW
R2 and above	1	78	0.97 [0.83, 1.0]	0.06 [0.01, 0.18]	**Sensitivity**				
					serious[a]	none[b]	NA[c]	serious[d]	LOW
					Specificity				
					serious[a]	none[b]	NA[c]	serious[d]	LOW
R3 and above	1	78	0.87 [0.70, 0.96]	0.30 [0.17, 0.45]	**Sensitivity**				
					serious[a]	none[b]	NA[c]	very serious[d]	LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					seriousa	noneb	NAc	noned	MODERATE
R4 and above	1	78	0.74 [0.55, 0.88]	0.81 [0.67, 0.91]	**Sensitivity**				
					seriousa	noneb	NAc	seriousd	LOW
					Specificity				
					seriousa	noneb	NAc	noned	MODERATE
R5 and above	1	78	0.29 [0.14, 0.48]	1.00 [0.92, 1.00]	**Sensitivity**				
					seriousa	noneb	NAc	noned	MODERATE
					Specificity				
					seriousa	noneb	NAc	noned	MODERATE
TMC-RSS category 2 and above	1	168	0.91 [0.77, 0.98]	0.97 [0.92, 0.99]	**Sensitivity**				
					very seriousa	noneb	NAc	very seriousd	VERY LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					very serious[a]	none[b]	NA[c]	none[d]	VERY LOW
TMC-RSS category 3 and above	1	168	0.77 [0.60, 0.90]	1.00 [0.97, 1.00]	**Sensitivity**				
					very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					very serious[a]	none[b]	NA[c]	none[d]	VERY LOW
Pathirana score of 5 and above	1	189	1.00 [0.88, 1.00]	0.45 [0.37, 0.53]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
Pathirana score of 6 and above	1	189	0.64 [0.44, 0.81]	0.72 [0.64, 0.79]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
low-level echo, 'vertical/horizontal >1', fuzzy boundary, microcalcification and grade IV blood flow (TIRADS grades 4 and 5)	1	89	0.88 [0.76, 0.95]	0.79 [0.61, 0.91]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW

(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) Indirectness was assessed using the QUADAS-2 checklist items referring to applicability. The evidence was downgraded by 1 increment if the majority of studies were seriously indirect.
(c) Inconsistency was assessed by visual inspection of the sensitivity/specificity plots, or data (if 2 studies). The evidence was downgraded by 1 increment if there was no overlap of 95% confidence intervals. For single studies no evaluation was made and 'NA' was recorded.
(d) Imprecision was assessed based on inspection of the confidence region 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. The evidence was downgraded by 1 increment when the confidence interval around the point estimate crossed one of the clinical thresholds (0.90 or 0.80 for sensitivity and 0.5 or 0.1 for specificity), and downgraded by 2 increments when the confidence interval around the point estimate crossed both of the clinical thresholds. The upper clinical threshold marked the point above which recommendations would be possible, and the lower clinical threshold marked the point below which the tool would be regarded as of little clinical use.

Table 7: Summary of evidence relating to Doppler US characteristics (for visualisation of nodular vascularity)

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Any blood flow	8	1,897	Pooled sensitivity (95% credible intervals):0.7167(0.52854-0.8676)	Pooled specificity (95% credible intervals): 0.3631 (0.2176 – 0.5376)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
Central blood flow	25	4,652	Pooled sensitivity (95% credible intervals):0.563(0.4299-0.6906)	Pooled specificity (95% credible intervals): 0.7153 (0.6191 – 0.7972)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
mean systolic blood velocity of	1	63	0.67 [0.30, 0.93]	0.63 [0.49, 0.76]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
33.5 m/s or more					**Specificity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
pulsatility index 0.92 or more	1	139	0.83 [0.64, 0.94]	0.55 [0.46, 0.65]	**Sensitivity**				
					serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					serious[a]	none[b]	NA[c]	serious[d]	LOW
pulsatility index 0.945 or more	1	63	1.00 [0.66, 1.00]	0.91 [0.80, 0.97]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
	1	43	0.72 [0.53, 0.87]	1.00 [0.77, 1.00]	**Sensitivity**				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
normalised VI of 0.14 or more					none	none[b]	NA[c]	serious[d]	MODERATE
					Specificity				
					none	none[b]	NA[c]	none[d]	HIGH
normalised VI of 0.278 or more	1	86	0.83 [0.69, 0.92]	0.55 [0.38, 0.71]	**Sensitivity**				
					none[a]	none[b]	NA[c]	very serious[d]	LOW
					Specificity				
					none[a]	none[b]	NA[c]	serious[d]	MODERATE
resistive index of 0.68 or more	1	140	0.57 [0.37, 0.75]	0.83 [0.74, 0.89]	**Sensitivity**				
					serious[a]	none[b]	NA[c]	none[d]	MODERATE
					Specificity				
					serious[a]	none[b]	NA[c]	none[d]	MODERATE

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
resistive index of 0.715 or more	1	63	0.89 [0.52, 1.00]	0.80 [0.66, 0.89]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
resistive index of 0.75 or more	1	144	0.57 [0.45, 0.68]	0.59 [0.46, 0.71]	**Sensitivity**				
					none	none[b]	NA[c]	none[d]	HIGH
					Specificity				
					none	none[b]	NA[c]	serious[d]	MODERATE
systolic /diastolic ratio of 3.11 or more	1	140	0.60 [0.41, 0.77]	0.83 [0.74, 0.89]	**Sensitivity**				
					serious[a]	none[b]	NA[c]	none[d]	MODERATE
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					seriousa	noneb	NAc	noned	MODERATE
colour doppler VTQ of 2.910 m/s or more	1	71	0.72 [0.53, 0.86]	1.00 [0.91, 1.00]	**Sensitivity**				
					Very seriousa	noneb	NAc	seriousd	VERY LOW
					Specificity				
					Very seriousa	noneb	NAc	noned	LOW

(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) Indirectness was assessed using the QUADAS-2 checklist items referring to applicability. The evidence was downgraded by 1 increment if the majority of studies were seriously indirect.
(c) Inconsistency was assessed by visual inspection of the sensitivity/specificity plots, or data (if 2 studies). The evidence was downgraded by 1 increment if there was no overlap of 95% confidence intervals. For single studies no evaluation was made and 'NA' was recorded.
(d) Imprecision was assessed based on inspection of the confidence region 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. The evidence was downgraded by 1 increment when the confidence interval around the point estimate crossed one of the clinical thresholds (0.90 or 0.80 for sensitivity and 0.5 or 0.1 for specificity), and downgraded by 2 increments when the confidence interval around the point estimate crossed both of the clinical thresholds. The upper clinical threshold marked the point above which recommendations would be possible, and the lower clinical threshold marked the point below which the tool would be regarded as of little clinical use.

Table 8: Summary of evidence relating to contrast-enhanced US characteristics [CEUS]

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
CEUS heterogeneous AND low enhancement pattern	4	412	Pooled sensitivity (95% credible intervals):0.9041 (0.7429-0.971)	Pooled specificity (95% credible intervals): 0.9116 (0.7778 – 0.97)	**Sensitivity**				
					very serious[a]	none[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					very serious[a]	none[b]	none[c]	none[d]	LOW
CEUS heterogeneous OR low enhancement pattern	1	109	0.71 [0.60, 0.81]	0.63 [0.44, 0.79]	**Sensitivity**				
					very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
CEUS heterogeneous enhancement pattern	4	538	Pooled sensitivity (95% credible intervals):0.8063 (0.5576-0.9332)	Pooled specificity (95% credible intervals): 0.8448 (0.6172 – 0.9514)	**Sensitivity**				
					very serious[a]	none[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					very serious[a]	none[b]	none[c]	none[d]	LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
CEUS hypo-enhancement pattern	1	795	0.78 [0.68, 0.87]	0.55 [0.41, 0.69]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
CEUS: low enhancement, weak enhancement, late enhancement and uneven enhancement	1	89	0.93 [0.83, 0.98]	0.88 [0.72, 0.97]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
CEUS: incomplete or no ring enhancement pattern	2	307	0.97 [0.91, 0.99]; 0.99 [0.93, 1.00]	0.82 [0.72, 0.90]; 0.13 [0.05, 0.24]	**Sensitivity**				
					serious[a]	none[b]	none[c]	none[d]	MODERATE
					Specificity				
					serious[a]	none[b]	serious[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
CEUS: irregular shape	3	376	Pooled sensitivity (95% credible intervals):0.7129 (0.4044-0.9062)	Pooled specificity (95% credible intervals): 0.8362 (0.5815 – 0.951)	**Sensitivity**				
					very serious[a]	none[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
CEUS: ill-defined enhancement border	3	376	Pooled sensitivity (95% credible intervals):0.6994 (0.3949-0.8988)	Pooled specificity (95% credible intervals): 0.8697(0.5958 – 0.9708)	**Sensitivity**				
					very serious[a]	none[b]	none[c]	serious[d]	VERY LOW
					Specificity				
					very serious[a]	none[b]	none[c]	none[d]	VERY LOW
CEUS: peak interior echogenicity on contrast enhanced US - hypoechoic	1	172	0.77 [0.67, 0.85]	0.78 [0.67, 0.87]	**Sensitivity**				
					none	none[b]	NA[c]	serious[d]	MODERATE

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					none	none[b]	NA[c]	none[d]	HIGH
CEUS: peak peripheral echogenicity on contrast enhanced US - hypoechoic	1	172	0.66 [0.55, 0.75]	0.82 [0.72, 0.90]	**Sensitivity**				
					none	none[b]	NA[c]	none[d]	HIGH
					Specificity				
					none	none[b]	NA[c]	none[d]	HIGH
CEUS: relative arrival time of nodule on contrast enhanced US - later	1	172	0.54 [0.44, 0.65]	0.92 [0.84, 0.97]	**Sensitivity**				
					Serious[a]	none[b]	NA[c]	none[d]	MODERATE
					Specificity				
					Serious[a]	none[b]	NA[c]	none[d]	MODERATE
CEUS fast wash-out	2	521	0.57 [0.44, 0.70] 0.25 [0.20, 0.31]	0.66 [0.55, 0.76] 0.92 [0.87, 0.96]	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
CEUS: complete wash out	1	101	0.92 [0.75, 0.99]	0.81 [0.71, 0.89]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
CEUS: hypo-perfusion	1	145	0.44 [0.32, 0.58]	0.76 [0.65, 0.84]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
CEUS: peak ratio ≤1.06	1	145	0.81 [0.69, 0.90]	0.40 [0.30, 0.52]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
CEUS: score of 1.6 or higher on purpose built risk score	1	145	0.86 [0.75, 0.93]	0.68 [0.57, 0.78]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
CEUS: sharpness ratio of >1.6	1	145	0.40 [0.28, 0.53]	0.83 [0.73, 0.90]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
CEUS: TTP ratio < 1.15	1	145	0.79 [0.67, 0.89]	0.50 [0.39, 0.61]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
CEUS: area>50%	1	109	0.76 [0.64, 0.85]	0.49 [0.33, 0.65]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
CEUS: based on access speed, peak time, subsidence speed, access manner, peak intensity, evenness, pattern of enhancement and clarity of boundary. Values for these parameters taken to represent a positive test were unclear	1	319	0.88 [0.81, 0.93]	0.86 [0.80, 0.91]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW

1
2
3
4

(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) Indirectness was assessed using the QUADAS-2 checklist items referring to applicability. The evidence was downgraded by 1 increment if the majority of studies were seriously indirect.

(c) Inconsistency was assessed by visual inspection of the sensitivity/specificity plots, or data (if 2 studies). The evidence was downgraded by 1 increment if there was no overlap of 95% confidence intervals. For single studies no evaluation was made and 'NA' was recorded.

(d) Imprecision was assessed based on inspection of the confidence region 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. The evidence was downgraded by 1 increment when the confidence interval around the point estimate crossed one of the clinical thresholds (0.90 or 0.80 for sensitivity and 0.5 or 0.1 for specificity), and downgraded by 2 increments when the confidence interval around the point estimate crossed both of the clinical thresholds. The upper clinical threshold marked the point above which recommendations would be possible, and the lower clinical threshold marked the point below which the tool would be regarded as of little clinical use.

Table 9: Summary of evidence relating to elastography [Real time elastography (RTE) and Shear wave elastography (SWE)]

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Real Time Elastography - Asteria 1-4 colour scale: 2 and above	1	30	1.00[0.59, 1.00]	0.17 [0.05, 0.39]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
Real Time Elastography - Asteria 1-4 colour scale: 3 and above	10	1,541	Pooled sensitivity (95% credible intervals):0.8183(0.7293-0.8925)	Pooled specificity (95% credible intervals): 0.8097 (0.7324 – 0.8731)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
Real Time Elastography - Asteria 1-4 colour scale: 4	1	30	0.43[0.10, 0.82]	1.00 [0.85, 1.00]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Real Time Elastography - Rago 1-5 colour scale: 2 and above	3	649	Pooled sensitivity (95% credible intervals):0.9744(0.8362-0.9982)	Pooled specificity (95% credible intervals): 0.3033 (0.003623 – 0.9798)	**Sensitivity**				
					Very serious[a]	none[b]	none[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	very serious[d]	VERY LOW
Real Time Elastography - Rago 1-5 colour scale: 3 and above	6	600	Pooled sensitivity (95% credible intervals):0.8773(0.6919-0.9741)	Pooled specificity (95% credible intervals): 0.7686 (0.4343 – 0.9449)	**Sensitivity**				
					serious[a]	none[b]	serious[c]	very serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
Real Time Elastography - Rago 1-5 colour scale: 4 and above	9	878	Pooled sensitivity (95% credible intervals):0.7511(0.5552-0.8953)	Pooled specificity (95% credible intervals): 0.9385 (0.8546 – 0.9825)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
Real Time Elastography - Rago 1-5 colour scale: 5	3	302	Pooled sensitivity (95% credible intervals):0.3202(0.01978-0.9146)	Pooled specificity (95% credible intervals): 0.9882 (0.9265 – 0.9989)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	none[c]	none[d]	LOW
Real Time Elastography - ITOH 1-5 colour scale: 2 or more	3	175	Pooled sensitivity (95% credible intervals):0.9611(0.8195-0.9937)	Pooled specificity (95% credible intervals): Pooled spec: 0.3766 (0.05041 – 0.875)	**Sensitivity**				
					Very serious[a]	none[b]	none[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					Very serious[a]	none[b]	serious[c]	very serious[d]	VERY LOW
Real Time Elastography - ITOH 1-5 colour scale: 3 or more	5	1,395	Pooled sensitivity (95% credible intervals):0.9241(0.6609-0.9902)	Pooled specificity (95% credible intervals): Pooled spec: 0.6111 (0.2242 – 0.8963)	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
Real Time Elastography - ITOH 1-5 colour scale: 4 or more	4	958	Pooled sensitivity (95% credible intervals):0.6096(0.4028-0.7814)	Pooled specificity (95% credible intervals): Pooled spec: 0.9408 (0.7215 – 0.9927)	**Sensitivity**				
					Very serious[a]	none[b]	none[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
Real Time Elastography - ITOH 1-5 colour scale: 5	3	849	Pooled sensitivity (95% credible intervals):0.1827(0.05411-0.4449)	Pooled specificity (95% credible intervals): Pooled spec: 0.9835 (0.8631 – 0.9992)	**Sensitivity**				
					serious[a]	none[b]	none[c]	none[d]	MODERATE

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					serious[a]	none[b]	none[c]	none[d]	MODERATE
Real Time Elastography - RGB (Liu, 2017 #854; Li, 2015 #836; Jin, 2018 #809; Pei, 2019 #898) 0-4 colour scale: 2 or more	2	321	0.98 [0.90, 1.00]; 1.00 [0.97, 1.00]	0.04 [0.01, 0.12]; 0.37 [0.26, 0.49]	**Sensitivity**				
					Serious[a]	Serious[b]	none[c]	serious[d]	VERY LOW
					Specificity				
					Serious[a]	Serious[b]	serious[c]	serious[d]	VERY LOW
Real Time Elastography - RGB 0-4 colour scale: 3 or more	4	495	Pooled sensitivity (95% credible intervals):0.8168(0.5634-0.9457)	Pooled specificity (95% credible intervals): Pooled spec: 0.8688 (0.7288 – 0.9488)	**Sensitivity**				
					serious[a]	none[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					serious[a]	none[b]	none[c]	none[d]	MODERATE
Real Time Elastography - RGB 0-4 colour scale: 4	2	321	0.07 [0.02, 0.18]; 0.05 [0.02, 0.11]	0.97 [0.90, 1.00]; 1.00 [0.95, 1.00]	**Sensitivity**				
					Very serious[a]	Serious[b]	none[c]	none[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					Very serious[a]	Serious[b]	none[c]	none[d]	VERY LOW
Real Time Elastography - 1-3 Rago scale: 2 or more	2	638	0.82 [0.66, 0.92]; 0.33 [0.23, 0.44]	0.43 [0.36, 0.51]; 0.75 [0.70, 0.80]	**Sensitivity**				
					Very serious[a]	none[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
Real Time Elastography - 1-3 Rago scale: 3	1	309	0.70 [0.55, 0.82]	0.82 [0.77, 0.87]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Real Time Elastography - 1-6 ES scale (Mohammed,	2	190	0.88 [0.75, 0.95]; 0.89 [0.72, 0.98]	0.90 [0.82, 0.95]; 0.88 [0.64, 0.99]	**Sensitivity**				
					Very serious[a]	none[b]	none[c]	very serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
2013 #672; Hong, 2009 #792): 4 or more					**Specificity**				
					Very serious[a]	none[b]	none[c]	none[d]	LOW
Real Time Elastography - '4 pattern' (Liu, 2019 #669): 3 or more	1	174	0.89 [0.79, 0.95]	0.75 [0.65, 0.82]	**Sensitivity**				
					Serious[a]	serious[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Serious[a]	Serious[b]	NA[c]	none[d]	LOW
Real Time Elastography - 0-IV colour grade system (Shuzen): III or more	2	686	0.97 [0.83, 1.00] 0.83 [0.77, 0.88]	0.93 [0.84, 0.98] 0.91 [0.87, 0.94]	**Sensitivity**				
					Very serious[a]	none[b]	none[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	none[c]	none[d]	LOW
Real Time Elastography - 0-5 colour grade system	1	100	0.92 [0.86, 0.96]	0.96 [0.92, 0.98]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
(Tang, 2017 #686): 3 or more					**Specificity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Real Time Elastography – 1-6 scoring method (Huang, 2015 #797) – 2 or more	1	155	1.00 [0.94, 1.00]	0.02 [0.00, 0.07]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Real Time Elastography – 1-6 scoring method (Huang, 2015 #797) – 3 or more	1	155	0.90 [0.80, 0.96]	0.39 [0.29, 0.50]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Real Time Elastography – 1-6 scoring method (Huang,	1	155	0.74 [0.61, 0.84]	0.90 [0.83, 0.96]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
2015 #797) – 4 or more									
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Real Time Elastography – 1-6 scoring method (Huang, 2015 #797) – 5 or more	1	155	0.16 [0.08, 0.28]	0.98 [0.93, 1.00]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Real Time Elastography – 1-6 scoring method (Huang, 2015 #797) – 6	1	155	0.00 [0.00, 0.06]	0.99 [0.94, 1.00]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Shear Wave Elastography – virtual touch tissue imaging I to	2	593	0.99 [0.95, 1.00]; 1.00 [0.94, 1.00]	0.22 [0.17, 0.26]; 0.19 [0.11, 0.28]	**Sensitivity**				
					serious[a]	none[b]	none[c]	none[d]	MODERATE

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
VI grade – grade II or more					**Specificity**				
					serious[a]	none[b]	none[c]	serious[d]	LOW
Shear Wave Elastography – virtual touch tissue imaging I to VI grade – grade III or more	2	593	0.91 [0.84, 0.95]; 0.85 [0.73, 0.94]	0.69 [0.64, 0.74]; 0.64 [0.54, 0.73]	**Sensitivity**				
					serious[a]	none[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					serious[a]	none[b]	none[c]	none[d]	MODERATE
Shear Wave Elastography – virtual touch tissue imaging I to VI grade – grade IV or more	2	593	0.79 [0.71, 0.86]; 0.73 [0.59, 0.84]	0.94 [0.91, 0.97]; 0.90 [0.82, 0.95]	**Sensitivity**				
					serious[a]	none[b]	none[c]	serious[d]	LOW
					Specificity				
					serious[a]	none[b]	none[c]	none[d]	MODERATE
Shear Wave Elastography – virtual touch tissue imaging I to	2	593	0.32 [0.24, 0.41]; 0.13 [0.05, 0.24]	1.00 [0.98, 1.00]; 1.00 [0.96, 1.00]	**Sensitivity**				
					serious[a]	none[b]	none[c]	none[d]	MODERATE

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
VI grade – grade V or more					**Specificity**				
					serious[a]	none[b]	none[c]	none[d]	MODERATE
Shear Wave Elastography – virtual touch imaging quantification velocity of 2.4 m/s or more	1	78	0.93 [0.82, 0.98]	0.83 [0.63, 0.95]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Shear Wave Elastography – virtual touch imaging quantification velocity of 2.545 m/s or more	1	182	0.963 (no CIs provided)	0.962 (no CIs provided)	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
Shear Wave Elastography – virtual touch imaging	1	88	0.76 [0.56, 0.90]	0.95 [0.86, 0.99]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
quantification velocity of 2.565 m/s or more					**Specificity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Shear Wave Elastography – virtual touch imaging quantification velocity of 2.64 m/s or more	1	154	0.82 [0.70, 0.91]	0.77 [0.68, 0.85]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Shear Wave Elastography – virtual touch imaging quantification velocity of 2.84 m/s or more	1	155	0.97 [0.89, 1.00]	0.96 [0.89, 0.99]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Shear Wave Elastography – virtual touch imaging	1	152	0.78 [0.65, 0.88]	0.84 [0.75, 0.90]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
quantification velocity of 2.87 m/s or more					**Specificity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Shear Wave Elastography – virtual touch imaging quantification velocity of 2.91 m/s or more	1	71	0.72 [0.53, 0.86]	1.00 [0.91, 1.00]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Shear Wave Elastography – virtual touch imaging quantification velocity of 2.91 m/s or more	1	141	0.33 [0.21, 0.47]	1.00 [0.96, 1.00]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
Shear Wave Elastography – Elastic Index of	1	319	0.85 [0.77, 0.90]	0.84 [0.78, 0.89]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
27.65kpa or more					**Specificity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Shear Wave Elastography – Elastic Index of 31.0 kpa or more	1	404	0.46 [0.34, 0.57]	0.61 [0.55, 0.66]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Shear Wave Elastography – Elastic Index of 36.2 kpa or more	1	356	0.76 [0.71, 0.81]	0.79 [0.68, 0.87]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
Shear Wave Elastography – Elastic Index of 38.3 kpa or more	1	51	0.73[0.39, 0.94]	0.85 [0.70, 0.94]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Shear Wave Elastography – Elastic Index of 39.3 kpa or more	1	64	0.68[0.43, 0.87]	0.87 [0.73, 0.95]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Shear Wave Elastography – Elastic Index of 45.0 kpa or more	1	298	0.85 [0.79, 0.90]	0.54 [0.44, 0.63]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
Shear Wave Elastography – Elastic Index of 52.1 kpa or more	1	174	0.73 [0.61, 0.84]	0.76 [0.67, 0.84]	**Sensitivity**				
					serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					serious[a]	serious[b]	NA[c]	none[d]	LOW
Shear Wave Elastography – Elastic Index of 66 kpa or more	2	454	0.81 [0.62, 0.94]; 0.80 [0.63, 0.92]	0.92 [0.86, 0.96]; 0.90 [0.86, 0.94]	**Sensitivity**				
					serious[a]	none[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					serious[a]	none[b]	none[c]	none[d]	MODERATE
Shear Wave Elastography – Elastic Index of 69 kpa or more	1	298	0.51 [0.43, 0.58]	0.85 [0.78, 0.91]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
Shear Wave Elastography – Elastic Index of 74.5 kpa or more	1	94	0.74 [0.63, 0.84]	0.90 [0.68, 0.99]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Shear Wave Elastography – Elastic Index of 120 kpa or more	1	298	0.11 [0.07, 0.16]	0.98 [0.93, 0.99]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
Real Time Elastography – Strain ratio of 1.32 and above	1	155	0.92 [0.82, 0.97]	0.82 [0.72, 0.89]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Real Time Elastography – Strain ratio of 2.2 and above	1	35	0.86 [0.57, 0.98]	0.90 [0.70, 0.99]	**Sensitivity**				
					none[a]	none[b]	NA[c]	very serious[d]	LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					none[a]	none[b]	NA[c]	none[d]	HIGH
Real Time Elastography – Strain ratio of 2.37 and above	1	152	0.91 [0.80, 0.97]	0.74 [0.64, 0.83]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Real Time Elastography – Strain ratio of 2.88 and above	1	30	0.86 [0.42, 1.00]	0.91 [0.72, 0.99]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Real Time Elastography – Strain ratio of 2.9 and above	1	131	0.88 [0.71, 0.96]	0.93 [0.86, 0.97]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Real Time Elastography – Strain ratio of 3.5 and above	2	219	0.88 [0.69, 0.97]; 0.72 [0.61, 0.82]	0.86 [0.75, 0.94]; 0.84 [0.72, 0.92]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Real Time Elastography – Strain ratio of 3.59 and above	1	144	1.00 [0.88, 1.00]	0.84 [0.77, 0.91]	**Sensitivity**				
					serious[a]	none[b]	NA[c]	serious[d]	LOW
					Specificity				
					serious[a]	none[b]	NA[c]	none[d]	MODERATE
Real Time Elastography – Strain ratio of 3.65 and above	1	812	0.46 [0.43, 0.50]	0.86 [0.78, 0.92]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Real Time Elastography – Strain ratio of 3.79 and above	1	98	0.98 [0.88, 1.00]	0.85 [0.72, 0.93]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Real Time Elastography – Strain ratio of 4 and above	1	133	0.80 [0.63, 0.92]	0.88 [0.80, 0.94]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW

(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) Indirectness was assessed using the QUADAS-2 checklist items referring to applicability. The evidence was downgraded by 1 increment if the majority of studies were seriously indirect.

(c) Inconsistency was assessed by visual inspection of the sensitivity/specificity plots, or data (if 2 studies). The evidence was downgraded by 1 increment if there was no overlap of 95% confidence intervals. For single studies no evaluation was made and 'NA' was recorded.

(d) Imprecision was assessed based on inspection of the confidence region 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. The evidence was downgraded by 1 increment when the confidence interval around the point estimate crossed one of the clinical thresholds (0.90 or 0.80 for sensitivity and 0.5 or 0.1 for specificity), and downgraded by 2 increments when the confidence interval around the point estimate crossed both of the clinical thresholds. The upper clinical threshold marked the point above which recommendations would be possible, and the lower clinical threshold marked the point below which the tool would be regarded as of little clinical use.

Table 10: Summary of evidence relating to combinations of grey scale, Doppler, CEUS and elastography scales/characteristics

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Microcalcifications AND absent halo AND type III vascularisation	3	399	Pooled sensitivity (95% credible intervals):0.3104 (0.0115-0.9369)	Pooled specificity (95% credible intervals): 0.9365 (0.4293 – 0.9979)	**Sensitivity**				
					Very erious[a]	none[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	serious[c]	serious[d]	VERY LOW
Microcalcifications AND	3	399	Pooled sensitivity (95% credible	Pooled specificity (95% credible	**Sensitivity**				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
hypoechoicity AND type III vascularisation			intervals):0.195 (0.02299-0.6325)	intervals): 0.9432 (0.7106 – 0.9956)	Very serious[a]	none[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	none[c]	none[d]	LOW
Hypoechoicity AND absent halo AND type III vascularisation	3	399	Pooled sensitivity (95% credible intervals):0.07959(0.01404-0.3089)	Pooled specificity (95% credible intervals): 0.9923 (0.9452 – 0.9993)	**Sensitivity**				
					Very serious[a]	none[b]	none[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	none[c]	none[d]	LOW
Microcalcification AND type III vascularisation	2	307	0.23 [0.10, 0.42]; 0.81 [0.64, 0.92]	0.86 [0.77, 0.93]; 0.75 [0.68, 0.82]	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Hypoechoicity AND type III vascularisation	2	307	0.47 [0.30, 0.65]; 0.47 [0.28, 0.66]	0.75 [0.68, 0.82]; 0.70 [0.59, 0.80]	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
Absent halo AND type III vascularisation	2	307	0.83 [0.67, 0.94]; 0.50 [0.31, 0.69]	0.44 [0.36, 0.52]; 0.89 [0.80, 0.95]	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
combined doppler and grey scale characteristics: calcification OR resistive index >0.715 OR pulsatility index >0.945	1	63	0.89 [0.52, 1.00]	0.93 [0.82, 0.98]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
French TI-RADS 3 or more AND capsule interruption and increased intranodular vascularization	1	133	1.00 [0.90, 1.00]	0.06 [0.02, 0.13]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
French TI-RADS 4a or more AND capsule interruption and increased intranodular vascularization	1	133	1.00 [0.90, 1.00]	0.26 [0.17, 0.35]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
French TI-RADS 4b or more AND capsule interruption and increased	1	133	0.94 [0.81, 0.99]	0.76 [0.66, 0.84]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
intranodular vascularization					**Specificity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
French TI-RADS 5 AND capsule interruption and increased intranodular vascularization	1	133	0.60 [0.42, 0.76]	0.96 [0.90, 0.99]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
TIRADS (0-10) and elastography Emax score of 5 or more	1	298	1.00 [0.98, 1.00]	0.17 [0.10, 0.24]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
TIRADS (0-10) and elastography	1	298	0.97 [0.93, 0.99]	0.50 [0.41, 0.60]	**Sensitivity**				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Emax score of 6 or more					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
TIRADS (0-10) and elastography Emax score of 7 or more	1	298	0.92 [0.87, 0.96]	0.68 [0.59, 0.76]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
TIRADS (0-10) and elastography Emax score of 8 or more	1	298	0.88 [0.82, 0.93]	0.79 [0.71, 0.86]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Kwak TIRADS and ITOH combined score of 5 and above	1	392	0.92 [0.88, 0.95]	0.75 [0.67, 0.81]	**Sensitivity**				
					serious[a]	none[b]	NA[c]	serious[d]	LOW
					Specificity				
					serious[a]	none[b]	NA[c]	none[d]	MODERATE
SWE at 66kpa and above OR microcalcification OR central vascularisation	1	297	0.97 [0.85, 1.00]	0.55 [0.49, 0.61]	**Sensitivity**				
					serious[a]	none[b]	NA[c]	serious[d]	LOW
					Specificity				
					serious[a]	none[b]	NA[c]	serious[d]	LOW
elastography ARFI SWV of 2.64 m/s or more OR elastography 1-6 scoring method score of 4 or more	1	155	0.97 [0.89, 1.00]	0.71 [0.61, 0.80]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	serious[d]	VERY LOW
					Specificity				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
virtual touch quantification at 2.91 m/s and above OR markedly hypoechoic	1	71	0.91 [0.75, 0.98]	0.74 [0.58, 0.87]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	VERY LOW
virtual touch quantification at 2.91 m/s and above OR poorly defined margins	1	71	0.88 [0.71, 0.96]	0.77 [0.61, 0.89]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
virtual touch quantification at 2.91 m/s and above OR taller than wide	1	71	0.91 [0.75, 0.98]	0.97 [0.87, 1.00]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	very serious[d]	VERY LOW

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
spot microcalcification AND presence of hypoecho AND type II vascularisation	1	46	0.07 [0.00, 0.32]	1.00 [0.89, 1.00]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
Absence of halo AND presence of hypoecho AND type II vascularisation	1	46	0.07 [0.00, 0.32]	1.00 [0.89, 1.00]	**Sensitivity**				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW
spot microcalcification	1	46	0.07 [0.00, 0.32]	1.00 [0.89, 1.00]	**Sensitivity**				

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
AND absence of halo AND type II vascularisation					Very serious[a]	none[b]	NA[c]	none[d]	LOW
					Specificity				
					Very serious[a]	none[b]	NA[c]	none[d]	LOW

(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) Indirectness was assessed using the QUADAS-2 checklist items referring to applicability. The evidence was downgraded by 1 increment if the majority of studies were seriously indirect.
(c) Inconsistency was assessed by visual inspection of the sensitivity/specificity plots, or data (if 2 studies). The evidence was downgraded by 1 increment if there was no overlap of 95% confidence intervals. For single studies no evaluation was made and 'NA' was recorded.
(d) Imprecision was assessed based on inspection of the confidence region 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. The evidence was downgraded by 1 increment when the confidence interval around the point estimate crossed one of the clinical thresholds (0.90 or 0.80 for sensitivity and 0.5 or 0.1 for specificity), and downgraded by 2 increments when the confidence interval around the point estimate crossed both of the clinical thresholds. The upper clinical threshold marked the point above which recommendations would be possible, and the lower clinical threshold marked the point below which the tool would be regarded as of little clinical use.

11.7. Summary of studies evaluating the accuracy of ultrasound for detection of nodules with malignant potential

No evidence was found

11.8. Summary of the evidence from studies evaluating ultrasound for the detection of nodules with malignant potential

No evidence was found

1.2. Review Question: In people with thyroid nodules on ultrasound at initial presentation, for what size and classification is it clinically and cost effective to use active surveillance or discharge rather than biopsy?

1.2.1. Introduction

The size of the nodule is sometimes used as a factor in determining who should get further investigations. It is unclear whether it is a good predictor of malignancy, whether there is any need to treat smaller nodules and ultimately whether small nodules impact on a person's quality of life. It may be that rather than receiving invasive tests and treatment they could be monitored

This review seeks to determine the threshold of nodule size/classification below which biopsy is not required (where harm, relative to the reference, is not manifested in the outcomes).

1.2.2. Summary of the protocol

For full details see the review protocol in Appendix A.2.

Table 11: PICO characteristics of review question

Population	People aged 16 or over who are suspected of having thyroid cancer with potentially malignant nodules on ultrasound at initial presentation and who have been assigned to active surveillance/discharge on the basis of the initial US result
Interventions	Different groups characterised by nodule size/characteristics For example: <1 cm with hypoechoic solid nodule + irregular borders, calcification, taller than wide, ETE >1 cm with hypoechoic solid nodule + irregular borders, calcification, taller than wide, ETE <1 cm with isoechoic/hyperechoic spongy/cystic nodules >1 cm with isoechoic/hyperechoic spongy/cystic nodules
Comparisons	Each other
Outcomes	• mortality • quality of life • local cancer progression • incidence of distant metastases • decision to treat • adverse events
Study design	• Observational studies (prospective/retrospective cohorts)

1.2.3. Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual. Methods specific to this review question are described in the review protocol in appendix A and the methods document.

Declarations of interest were recorded according to NICE's conflicts of interest policy.

1.2.4. Effectiveness evidence

1.2.4.1. Included studies

One clinical study[327] comparing nodule characteristics or sizes in people assigned to active surveillance or discharge was identified.

See also the study selection flow chart in Appendix C.2, study evidence tables in Appendix D.2, forest plots in Appendix F.2 and GRADE tables in Appendix G.

1.2.4.2. 1.1.4.2 Excluded studies

See the excluded studies list in Appendix J.2.

1.2.5. Summary of studies included in the effectiveness evidence

Table 12: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Rozenbaum 2021[327]	The participants, who were all on active surveillance until they reached the end point of the study by virtue of being treated or having FNA/surgery, were divided into two groups according to the outcome of volumetric progression (>50% increase in size) or no such progression. The US characteristics and size characteristics have been compared between these groups (volumetric progression [n=28] and no volumetric progression [n=52]	All patients had received US, had at least one suspicious nodule and agreed to go on active surveillance for at least one year. Investigations such as FNAC or diagnostic surgery would only be undertaken in response to suspected progression and would represent the end point to the study. Inclusion: Age of 18 or older; EU TIRADS 5 nodules < or equal to 10mm in the largest diameter; at least 2 sequential US examinations; patient willing to avoid surgery and having understood the principles and constraints of active surveillance. Exclusion: extra thyroidal extension; suspicious latero-cervical lymph nodes by neck ultrasound.	Local cancer progression	No adjustments made for potential confounding

See Appendix D.2 for full evidence tables.

1.2.6. Summary of the effectiveness evidence

The GRADE table below is unusual in that the sole outcome of the included study, nodule volumetric progression, was used as the grouping variable. This has meant that the factors relating to US characteristics and nodule size occupy the first column. This is in contrast to the normal situation, where the grouping variable is determined by the intervention type or characteristic type, rather than the outcome category, and the first column is the preserve of the separate outcomes. The comparison for each US characteristics or size factor is for volumetric progression versus no volumetric progression, and so relative risks and mean differences should be interpreted with that in mind. For example, the lower value for the baseline nodule diameter in the volumetric progression group indicates that a lower nodule diameter at baseline is associated with progression of nodule size. Similarly, the greater proportion of people with irregular margins in the volumetric progression group relative to the group showing no volumetric progression indicates that irregular margins are associated with volumetric progression.

Table 13: Clinical evidence summary: volumetric progression versus no volumetric progression

Baseline US and size factors	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with no volumetric progression	Risk difference with volumetric progression (95% CI)
baseline volume	80 (1)	VERY LOW1,2,3			The mean baseline volume in the intervention groups was 0.03 lower (from 0.06 lower to 0 higher)
baseline diameter	80 (1)	VERY LOW1,2,3			The mean baseline diameter in the intervention groups was 0.7 lower (from 1.64 lower to 0.24 higher)
microcalcifications	80 (1)	VERY LOW1,2,3	RR 1.06 (0.34 to 3.32)	135 per 1000	8 more per 1000 (from 89 fewer to 312 more)
hypoechogenicity	80 (1)	VERY LOW1,2,3	RR 0.84 (0.33 to 2.19)	212 per 1000	34 fewer per 1000 (from 142 fewer to 252 more)
irregular margins	80 (1)	VERY LOW1,2,3	RR 1.14 (0.92 to 1.42)	750 per 1000	105 more per 1000 (from 60 fewer to 315 more)

Baseline US and size factors	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with no volumetric progression	Risk difference with volumetric progression (95% CI)
irregular shape	80 (1)	VERY LOW1,2,3	RR 0.98 (0.73 to 1.3)	731 per 1000	15 fewer per 1000 (from 197 fewer to 219 more)
2 or more criteria on EU TIRADS 5	80 (1)	VERY LOW1,2,3	RR 1.06 (0.79 to 1.43)	673 per 1000	40 more per 1000 (from 141 fewer to 289 more)
no vascularity	80 (1)	VERY LOW1,2,3	RR 1.66 (1.04 to 2.65)	365 per 1000	241 more per 1000 (from 15 more to 603 more)
peripheral vascularity	80 (1)	VERY LOW1,2,3	RR 0.66 (0.27 to 1.65)	269 per 1000	92 fewer per 1000 (from 197 fewer to 175 more)
central vascularity	80 (1)	VERY LOW1,2,3	RR 0.37 (0.12 to 1.17)	288 per 1000	182 fewer per 1000 (from 254 fewer to 49 more)

1. Risk of bias was assessed with the ROBINS tool. This yielded on an overall judgement of critical risk of bias, based on failure to adjust for any confounding.
2. Indirectness was deemed serious due to the outcome of nodular volumetric progression not necessarily relating to cancer progression
3. Imprecision was rated very serious if the 95% CIs crossed both MIDs and serious if they crossed one MID. MIDs were taken as + 0.5 x the standard deviation of the control group for continuous variables* and a RR of 0.8 and 1.25 for binary variables.

**MID for nodule volume was 0.042, based on sd of 0.084 in control group, and MID for nodule diameter was 1.05, based on sd of 2.1 in control group*

CI: Confidence interval; RR: Risk ratio;

See Appendix G for full GRADE tables

To facilitate interpretation of results, the results are also presented in tabular form below, with an indication of the percentages.

US characteristic or size characteristic	volumetric progression [n=28]	no volumetric progression [n=52]
Baseline nodule volume(cm3)	0.045(0.047)	0.074(0.084)
Baseline nodule diameter (mm)	4.9(2.0)	5.6(2.1)
microcalcifications	4/28 (14.3%)	7/52 (13.5%)
hypoechogenicity	5/28 (17.9%)	11/52 (21.2%)
irregular margins	24/28 (85.7%)	39/52 (75%)
irregular shape	20/28 (71.4%)	38/52 (73.1%)
2 or more criteria on EU TIRADS 5	20/28 (71.4%)	35/52 (67.3%)
US vascularity – none	17/28 (60.7%)	19/52 (36.5%)
US vascularity – peripheral only	5/28 (17.9%)	14/52 (26.9%)
US vascularity – central component only	3/28 (10.7%)	15/52 (28.8%)

1
2
3

1.3. Economic evidence and model

1.3.1. Included studies

No health economic studies were included.

1.3.2. Excluded studies

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

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

1.3.3. Summary of included economic evidence

None.

1.3.4. Economic model

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

1.3.5. Unit costs

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

Resource	Unit costs	Source
Ultrasound Scan with duration of less than 20 minutes, without Contrast	£52	NHS Reference Costs 2018/2019[285]
Ultrasound Scan with duration of 20 minutes and over, without Contrast	£67	NHS Reference Costs 2018/2019[285]

1.3.6. Economic evidence statements

- No relevant economic evaluations were identified.

1.4. The committee's discussion and interpretation of the evidence

1.4.1. The committee's discussion and interpretation of the evidence

This discussion includes information from the 2 reviews above. These are combined as the recommendations were informed by both reviews. We have included subheadings in some sections to clarify which review we are referring to.

1.4.1.1. The outcomes that matter most

Diagnostic accuracy of ultrasound

Sensitivity and specificity were the outcomes used in the diagnostic review of ultrasound. Sensitivity was identified as the primary measure in guiding decision-making. The committee realised that it was unlikely that any method or threshold of ultrasound would have sufficiently high sensitivity and specificity to be able to be used as a definitive stand-alone diagnostic test. It was deemed more likely that ultrasound would have utility as a first line test, prior to a more expensive and invasive but more definitive test such as fine needle aspiration cytology (FNAC). If a test is to be used as a first line screening test then as high a sensitivity as possible is essential to avoid people with true malignancy being lost from the system at the first hurdle. This is because first line screening usually works by only feeding through people with a positive result to the next test – thus it is essential to be sure that all people with negative test results – those that will be eliminated from further testing – are truly without disease, and this can only be assured with high sensitivity. It is also important to have reasonable specificity alongside the high sensitivity, as poor specificity would imply little value from a first line test. For example, using a test with low specificity would mean that there would be very few negative index test results, and there would thus be few people that would be eliminated from further testing. Since the purpose of first line testing is to ration access to the later stages of testing, there would be little point in using such a test as a first line test. It would simply result in almost everyone tested with ultrasound being retained for further investigation with the next test – and thus it would make more sense to send everyone straight to the second test without wasting time and resources on the first test. The committee therefore set clinical decision thresholds for sensitivity of 0.9 and above for recommending a test, and 0.8, below which a test would be deemed of no clinical use. They also set clinical decision thresholds for specificity of 0.5 and above for recommending a test, and 0.1, below which a test would be deemed of no clinical use.

Nodule size and ultrasound characteristics review

For the review evaluating the nodule size and US characteristics associated with a good outcome from active surveillance, outcomes were mortality, quality of life, local cancer progression, incidence of distant metastases and cancer recurrence. All were regarded as of critical importance to decision-making.

1.4.31.2. The quality of the evidence

Diagnostic accuracy of ultrasound

The quality of the evidence for the diagnostic review of ultrasound was graded as very low to high, although the majority of outcomes were very low. The main reasons for this were the serious or very serious risk of bias (as determined by QADAS 2) in the majority of outcomes due to insufficient data on patient selection, blinding and poor reporting of the time between index and gold standard testing. GRADE ratings were also downgraded due to indirectness in outcomes where the majority of studies were retrospective: retrospective studies might have different populations to those specified in the protocol because only participants with more severe disease might be given surgery (and therefore qualify for inclusion due to having the gold standard of surgical histopathology). Heterogeneity was common in most meta-analyses undertaken and since these were not resolved by the pre-hoc sub-grouping strategy (medical status of US tester) many outcomes were downgraded for inconsistency. Finally, many outcomes were downgraded for imprecision, partly because of small study sizes, but also secondary to unresolved heterogeneity in meta-analyses.

Nodule size and ultrasound characteristics review

The quality of evidence for the review evaluating associations between nodule size and ultrasound characteristics and outcome in people placed on active surveillance was very low.

The evidence was observational by nature (as it was a study of the effects of characteristics that were already present in participants, rather than a study amendable to randomisation) and unadjusted for selection bias. The evidence was also downgraded for indirectness because the outcome of nodule volumetric progression was a proxy for cancer progression, and in some patients, it was possible that the nodules were not malignant. Further downgrades were made for imprecision.

1.4.18.3. Benefits and harms

Diagnostic accuracy of ultrasound

The committee set clinical decision thresholds for sensitivity of 0.9 and above for recommending a test, and 0.5 for specificity, above which a test would be recommended. The committee agreed that an index test of choice would have good sensitivity and specificity but ideally also did not involve complex procedures or special training to use, was not invasive and was applicable to most patients.

Elastography

Real time elastography using 'ITOH 1-5 colour scale with 3 or more' had a good sensitivity of 0.92 and specificity of 0.61 pooled from 5 studies graded as very low quality. There was other elastography index test findings that also met these criteria from single studies or from two studies where only one of the studies met these criteria. The committee agreed that elastography, whilst containing some index tests with excellent sensitivity and specificity, was not a simple modality to use, had cost implications, would involve much new training, varied widely in form and function between manufacturers, and was not applicable to many patients. For example, it is not useful for people with follicular carcinoma, nodules with a calcified shell, cystic lesions, and multinodular goitre with coalescent nodules, because the tumour margins need to be well demarcated for proper interpretation. They also agreed that results from elastography are heterogenous varying with different manufacturers ultrasound devices.

Contrast enhanced ultrasound

Contrast enhanced ultrasound results were ranged from high to very low quality. The best diagnostic accuracy came from pooled results from 4 studies with a sensitivity 0.90 (very low quality) and specificity of 0.91 (low quality). Despite the good sensitivity and specificity, the committee agreed this modality should not be recommended because it was invasive and would involve significant amounts of special training. Complex combinations of diagnostic approaches, such as those involving elastography and doppler, were also regarded as impractical and unfeasible in the clinical setting.

Grey scale ultrasound

The committee therefore focussed consideration upon evidence relating to grey scale findings, simple combinations of grey scale findings, doppler US, and the more formal ordinal scales, such as the different TIRADS systems (see Table 3: Summary of the types of US scales used.). Within these groups, the committee agreed to focus on tests that met the pre-hoc criteria of 0.9 sensitivity and 0.5 specificity. There were no simple grey-scale findings meeting the accuracy criteria, but the combination index test where the positive category was 'blurred margins plus at least one of the following: hypoechoicity, microcalcifications or taller than wide' had a high sensitivity of 0.98 and specificity of 0.63. However, concern was raised over the risk of bias, which was very serious, the wide spread of credible intervals for specificity, and the fact that the data were derived from one study, where low representativeness of the data can be a concern. One Doppler index test, where the positive category was 'pulsatility index of 0.945 or more' also had excellent sensitivity and specificity of 1.00 and 0.91 respectively. However, the small size of this single study meant there was considerable uncertainty in the sensitivity estimate and the committee lacked confidence in making a recommendation based on this study alone.

Scales for classifying ultrasound results

The evidence review considered a number of ordinal scales. The most important consideration in the discussion of choice of scale was the sensitivity, which needed to be as close to unity as possible in order to prevent people with malignancy being 'missed'. However, specificity also had to be reasonably high to ensure that sufficient filtering took place before the second line test – otherwise there would be little purpose in first-line testing. The Horvath scale with the positive category at 4a and above, the Horvath scale with the positive category at 4b and above and the Park scale with the positive category at 3 and above had the highest sensitivity whilst also having specificity about the minimum level of 0.5. However, the committee noted that the evidence for these three scales/thresholds were based on fewer studies and patients than some of the other scales/thresholds. Furthermore, the committee noted that the Horvath scale was highly complex to use, and that the Park scale tended to lead to the detection of less-clinically relevant findings (such as microcarcinomas). The committee therefore considered two other scales/thresholds – the EU TIRADS with the positive category at 4 or more and Kwak TIRADS with positive category at 4b or more. Although the accuracy of these two latter scales were slightly less than seen for the Horvath and Park scales/thresholds previously considered, the difference was small, and the great advantages of the latter two scales over the Horvath and Park scales/thresholds were the fact that they were based on much more studies and patients, and the certainty of the accuracy estimates were far greater. Although the Kwak TIRADS with positive category at 4b or more had slightly better sensitivity than EU TIRADS with the positive category at 4 or more, the committee favoured the EU TIRADS with the positive category at 4 or more. This was because most of the evidence behind the Kwak TIRADS scale was from outside Europe, whereas the majority of the evidence for the EU TIRADS was from Europe. The committee agreed that the population of many of the non-European studies in the evidence base for the Kwak scale may have been biased by the presence of many micropapillary nodules that would not necessarily be representative of the nodules that would be scanned in UK practice. Finally, the committee agreed that the EU TIRADS with the positive category at 4 or more scale/threshold had the great advantage of simplicity over most of the other scales – it was regarded as implementable and would not be difficult for clinicians to learn how to use.

There was some concern that the EU TIRADS with the positive category at 4 or more sensitivity of 0.95 would still lead to an unacceptably high number of false negatives. If the first line test is to be used as a filter, and the conventional 'positives-only-to-the next-test' strategy is used, then any negatives will be dismissed from further testing. Based on the sensitivity values of 0.95, this means that 5% of people with true malignancy would be incorrectly identified as negative by the test, and thus lost from further testing. This prompted the125ommitee to consider other scales/thresholds with higher sensitivities, but all of these contravened the specificity minimum, or involved the Horvath and Park scales/thresholds that had been previously dismissed as impractical. The committee considered a post-hoc revision of the pre-hoc decision to set specificity at a minimum level of 0.5, but review of all available scales/thresholds revealed that the only scales/thresholds with appreciably better sensitivity than the EU TIRADS with the positive category at 4 or more (at around 0.99) had specificities in the region of 0.25. These low specificities would not justify a first line testing approach, as they would not reduce the burden on FNA testing appreciably. The committee therefore agreed that the EU TIRADS with the positive category at 4 or more was probably the best scale/threshold available, but remained concerned about the risk of false negatives. The committee therefore made a recommendation for those people with other reasons for clinical suspicion, but without a TIRADS of 4 or more, could also be put forward for further investigation, at the discretion of the attending clinician, or person/team performing the ultrasound. This would reduce the number of people with true malignancy who might otherwise be lost from the system.

The committee agreed that all the data were of very low quality, the data for an EU-TIRADS score of 4 or more had a few thou

The committee were aware that the positive threshold of 4 or more for the EU TIRADS was different to the threshold recommended by the EU TIRADS developers of 3 or more. Our review data showed that using a threshold of 3 or more would have a sensitivity of 0.999 but a specificity of around 0.035, thus making first line testing completely irrelevant. One reason put forward for the discrepancy was that there is now more evidence available and included in this review than had been available to the developers at the time of the recommendation. EU TIRADS also include nodule size when making recommendations, but this evidence review did not support use of size in addition to EU TIRADS classification or as an alternative.

The committee were aware that many clinicians in the UK use the BTA U classification although some may be using EU-TIRADS. Should people adopt EU TIRADS then it would be a change in practice. However, the committee agreed that the published accuracy of the BTA U classification was inferior to the EU TIRADS and that it was important to highlight that evidence. Our evidence showed that using a BTA U threshold of 3 or more would have a sensitivity of 0.9368 but a specificity of 0.3974. Although, all the data were of very low quality the committee agreed that the evidence from the EU-TIRADS score of 4 or more was from a 11 studies and involved over 6000 patients. They agreed that this appeared to offer the best evidence for a threshold for identifying malignancy. In addition, the EU TIRADS was regarded as non-complex and it was not regarded as a tool that would present problems for clinicians.

Overall, the committee agreed that while the evidence showed that the EU-TIRADS with a score of 4 or more showed the best evidence for identifying people with a likely malignancy, they did not think it was enough to warrant recommending a change in practice. The committee agreed it was also important to note that using EU-TIRADS without including criteria for size could lead to many more people being referred for FNAC than is currently happening. However, they agreed that it was important to highlight what they considered as the best available evidence. Therefore, they recommended that if EU-TIRADS was already being used that an EU-TIRADS score of 4 or more should be considered to select people for further investigation with FNAC.

The committee also recommended that people who did not reach this threshold might also be considered for further evaluation if there were extenuating clinical reasons for this. This is to ensure that clinicians also use their judgement when deciding who needs further investigation. Grey scale ultrasound was recommended because the TIRADS scale used grey scale characteristics.

Nodule size and ultrasound characteristics review

Very limited evidence was found for the review question concerning the size and characteristics of nodules that would lead to the best outcomes when using active surveillance or discharge rather than biopsy. This question arises because of an understanding that very small nodules are both difficult to biopsy but also unlikely to lead to metastases. Specific recommendations for very small nodules might reduce patient anxiety and health resource use.

One article of a population with micropapillary thyroid carcinoma was included in the review, but it was flawed by its use of nodule growth (>50% increase in volume) as the measure of progression. No assessment was made of ultimate diagnosis, so it was unclear if the nodule growth represented cancer progression (although given the sample of people with TIRADS 5 US findings, malignancy was likely in a reasonable proportion of the sample). In addition, a 50% increase in volume would mean relatively small increases in diameter. The study showed very weak and uncertain associations between greater nodule progression over the course of follow up and baseline US findings of smaller nodule volume and diameter, less central vascularity and more irregular margins. The committee thought that the association with smaller volume may have been artefactual, relating to the greater scope for smaller

nodules to grow. The association with less central vascularity was similarly counter-intuitive, conflicting with the established idea that central vascularity and malignancy are associated, and it was thought that one reason for this result might be the subjectivity of Doppler findings. The only feature that fitted with current knowledge was the association of progression with more irregular margins. Overall, the committee did not think that the results from this small and unconvincing study were enough to allow useful recommendations. The committee were unsurprised that no other good evidence existed, giving the opinion that such evidence would need follow-ups of decades rather than years. They agreed that high quality research had not been carried out because it was very difficult to carry out, and it was therefore decided that a research recommendation would not be feasible.

The committee considered whether they could make a consensus recommendation as the evidence used to support the recommendation did not include size as a criterion. The committee noted that in the developmental literature for the EU TIRADS it was suggested that for nodules of 4 and above FNA should only be considered if the size was of 10mm or more. This seemed to support a recommendation stating that further investigations should be instituted if nodules were EU TIRADS 4 or above AND the nodules were at least 10mm; this seemed to imply that if nodules were smaller than this then even though the risk of malignancy would remain substantially the same, the ultimate clinical outcome would be sufficiently better to warrant no further investigation. However, the basis for this claim was unclear, and the committee were therefore not confident to use 10mm as a cut off in a recommendation. The committee were concerned however that people with very small nodules, might be put through to further testing unnecessarily. It was agreed that some very small nodules, particularly those of 5mm or less, are technically difficult to FNAC and were unlikely to metastasise. They also agreed that even if they were to be malignant, they might not progress and might not need treatment. Overall, given that the evidence used to make a recommendation to use a threshold of an EU TIRADS score of 4 or more did not include mention of size the committee were unable to include size as part of the recommendations in relation to selecting people for FNAC.

The committee also agreed with the recommendations on investigating thyroid enlargement in the NICE guideline on thyroid disease. They discussed the importance of using a classification system that takes into account echogenicity, microcalcifications, border, shape in transverse plane, internal vascularity and lymphadenopathy and noted that the EU TIRADS and BTA U classification do this. They also agreed that reports of ultrasound findings should: specify which grading system has been used for the assessment; include information on the characteristics of the nodule; provide an overall assessment of malignancy; confirm that both lobes have been assessed; and document assessment of cervical lymph nodes. This can help improve diagnosis by ensuring all the data are available to clinicians when assessing the patient.

1.4.10.4. Cost effectiveness and resource use

No health economics evidence was found for this question. The committee made a recommendation to offering grayscale ultrasound as the initial test for people with nodules. This represent current practice and, as such, is not expected to have any impact to the NHS.

Due to the low quality of the evidence, the committee decided not to recommend any classification system over the widely used in current practice BTA U scale, and therefore, the recommendations are unlikely to persuade clinicians to switch to another scale. The recommendation on EU TIRADS threshold reflects the clinical evidence and it is expected to improve the accuracy of clinicians already using this system.

There were some concerns on false negatives as a number of malignant nodules are expected to be missed during the ultrasound. Hence, the committee made a consider

recommendation to institute active surveillance, FNAC or diagnostic hemithyroidectomy for people who do not meet the threshold for FNAC if there are other reasons for clinical concern. This represents a change in practice and may require additional resource in terms of more outpatient visits and yearly US scan. However, this should reduce the number of people ending up with a delayed thyroid cancer diagnosis, which should ultimately reduce treatment and surgery cost occurring downstream and improve quality of life and survival of people with malignant nodules.

1.4.18.5. Other factors the committee took into account

The committee emphasised that the accuracy of any test, and particularly ultrasound, depends on the expertise of the operator. It was pointed out how the accuracies of the studies may be superior to those seen in the real world, because they may tend to utilise the most experienced operators in order to obtain the best levels of accuracy possible.

Our data did include some information on the expertise of the operators, though it was incomplete. The medical status of the tester was used as a sub-grouping strategy, and so data were collected on the status of the tester. Most studies did not specify who collected the data, and only 40% of studies specified that a medically trained tester was used. However, this does not demonstrate that the expertise was perhaps less than might be expected, because a failure to specify the tester does not imply the tester is inexpert. On the other hand, the medical status of the testers is of little value in deciding expertise, and so overall the information gained from studies was unhelpful in deciding if the expertise in the studies was representative of that in the real world.

The committee acknowledged the potential for health inequalities for people with a high BMI, who may have difficulty lying supine with neck extended. It was agreed that this could affect the quality of images obtained, and that special attention should therefore be given to such patients to ensure optimal positioning and comfort. However, they anticipated that it is standard practice for people doing imaging to consider this.

1.4.24.2. Recommendations supported by this evidence review

This evidence review supports recommendation 1.2.6, and recommendations 1.2.8 – 1.2.10.

References

1. Abbasian Ardakani A, Bitarafan-Rajabi A, Mohammadi A, Hekmat S, Tahmasebi A, Shiran MB et al. CAD system based on B-mode and color Doppler sonographic features may predict if a thyroid nodule is hot or cold. European Radiology. 2019; 29(8):4258-4265

2. Abbasian Ardakani A, Gharbali A, Mohammadi A. Application of texture analysis method for classification of benign and malignant thyroid nodules in ultrasound images. Iranian Journal of Cancer Prevention. 2015; 8(2):116-124

3. Abd_Alrahman ZA, Hummadi NA, Hussaini HA, S HA-H, Al-Saadi WI. Conventional sonography and strain sonoelastography in the evaluation of solid thyroid nodules. American Journal of BioMedicine. 2017; 5(10):538-557

4. Abdelrahman SF, Ali FH, El-Sayed Khalil M, El Masry MR. Ultrasound elastography in the diagnostic evaluation of indeterminate thyroid nodule s. Egyptian Journal of Radiology and Nuclear Medicine. 2015; 46(3):639-648

5. 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

6. 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

7. Aggarwal SK, Jayaram G, Kakar A, Goel GD, Prakash R, Pant CS. Fine needle aspiration cytologic diagnosis of the solitary cold thyroid nodule. Comparison with ultrasonography, radionuclide perfusion study and xeroradiography. Acta Cytologica. 1989; 33(1):41-47

8. Aghaghazvini L, Maheronnaghsh R, Soltani A, Rouzrokh P, Chavoshi M. Diagnostic value of shear wave sonoelastography in differentiation of benign from malignant thyroid nodules. European Journal of Radiology. 2020; 126:108926

9. Ahmadi S, Herbst R, Oyekunle T, Jiang X, Strickland K, Roman S et al. Using the Ata and Acr Ti-Rads sonographic classifications as adjunctive predictors of malignancy for indeterminate thyroid nodules. Endocrine Practice. 2019; 25(9):908-917

10. Ahmadi S, Oyekunle T, Jiang X, Scheri R, Perkins J, Stang M et al. A direct comparison of the ata and ti-rads ultrasound scoring systems. Endocrine Practice. 2019; 25(5):413-422

11. 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

12. Ajmal S, Rapoport S, Ramirez Batlle H, Mazzaglia PJ. The natural history of the benign thyroid nodule: what is the appropriate follow-up strategy? Journal of the American College of Surgeons. 2015; 220(6):987-992

13. 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):269-273

14. Al-Chalabi H, Karthik S, Vaidyanathan S. Radiological-pathological correlation of the British Thyroid Association ultrasound classification of thyroid nodules: a real-world validation study. Clinical Radiology. 2019; 74(9):702-711

15. Amin M, Zahra M, Siddiqui MA, Kamal R, Javed S, Maqsood A et al. Diagnostic accuracy of a new sonographic classification system in diagnosing malignant thyroid nodules. Pakistan Journal of Medical and Health Sciences. 2021; 15(6):1865-1867

16. An P, Ye Y, Ning Y, Feng W, Li Y, Zhou S et al. Analysis of the influence factors on the differential diagnosis between benign and malignant 4-class thyroid nodules by the ratio of the elastic strain rate. Journal of Medical Imaging and Health Informatics. 2020; 10(3):731-735

17. Angell TE, Frates MC, Medici M, Liu X, Kwong N, Cibas ES et al. Afirma benign thyroid nodules show similar growth to cytologically benign nodules during follow-up. Journal of Clinical Endocrinology and Metabolism. 2015; 100(11):E1477-1483

18. 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):249-256

19. Arambewela MH, Wijesinghe AM, Randhawa K, Bull M, Wadsley J, Balasubramanian SP. A pragmatic assessment of the British Thyroid Association "U classification" of thyroid nodules with a focus on their follow-up. Clinical Radiology. 2020; 75(6):466-473

20. Arpana, Panta OB, Gurung G, Pradhan S. Ultrasound findings in thyroid nodules: A radio-cytopathologic correlation. Journal of Medical Ultrasound. 2018; 26(2):90-93

21. Aslan A, Sancak S, Aslan M, Ayaz E, Inan I, Ozkanli SS et al. Diagnostic value of duplex doppler ultrasound parameters in papillary thyroid carcinoma. Acta Endocrinologica. 2018; 14(1):43-48

22. Aydin R, Elmali M, Polat AV, Danaci M, Akpolat I. Comparison of muscle-to-nodule and parenchyma-to-nodule strain ratios in the differentiation of benign and malignant thyroid nodules: which one should we use? European Journal of Radiology. 2014; 83(3):e131-136

23. 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):734-740

24. 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

25. Baek HJ, Ryu KH, An HJ, Kim JP, Jung EJ, Kim DW. Validation of diagnostic performance and interobserver agreement of DTD-TIRADS for diffuse thyroid disease on ultrasound: A single-center study. AJR American Journal of Roentgenology. 2021; 216(5):1329-1334

26. Bafaraj SM. Assessment of sensitivity, specificity, and accuracy of nuclear medicines, CT scan, and ultrasound in diagnosing thyroid disorders. Current Medical Imaging Reviews. 2020; 16(3):193-198

27. Bakari KH, Simai KA, Mghanga FP, Mayala HA, Mingxing X. Role of shear wave elastography in diagnosis of benign and malignant thyroid nodules. Journal of Clinical and Diagnostic Research. 2018; 12(9):TC05-TC08

28. Bas H, Ustuner E, Kula S, Konca C, Demirer S, Elhan AH. Elastography and doppler may bring a new perspective to TIRADS, altering conventional ultrasonography dominance. Academic Radiology. 2022; 29(3):e25-e38

29. Baz AAA, Mohamed AHI, El-Esawy YFG, El-kaffas KH. Conventional ultrasound, color Doppler, TI-RADS, and shear wave elastography for thyroid nodule differentiation: a study of efficacy compared with the histopathology results. Egyptian Journal of Radiology and Nuclear Medicine. 2021; 52(1):103

30. Becker D, Bair HJ, Becker W, Gunter E, Lohner W, Lerch S et al. Thyroid autonomy with color-coded image-directed Doppler sonography: internal hypervascularization for the recognition of autonomous adenomas. Journal of Clinical Ultrasound. 1997; 25(2):63-69

31. Bederina EL, Orlinskaya NY, Konovalov VA, Zubeev PS. Diagnostic value sonoelastography in differential diagnosis of thyroid nodules. Sovremennye Tehnologii v Medicine. 2014; 6(1):43-45

32. Berker D, Aydin Y, Ustun I, Gul K, Tutuncu Y, Isik S et al. The value of fine-needle aspiration biopsy in subcentimeter thyroid nodules. Thyroid. 2008; 18(6):603-608

33. Berni A, Tromba L, Falvo L, Marchesi M, Grilli P, Peparini N. Malignant thyroid nodules: comparison between color Doppler diagnosis and histological examination of surgical samples. Chirurgia Italiana. 2002; 54(5):643-647

34. Bonavita JA, Mayo J, Babb J, Bennett G, Oweity T, Macari M et al. Pattern recognition of benign nodules at ultrasound of the thyroid: which nodules can be left alone? AJR American Journal of Roentgenology. 2009; 193(1):207-213

35. Bora Makal G, Aslan A. The diagnostic value of the american college of radiology thyroid imaging reporting and data system classification and shear-wave elastography for the differentiation of thyroid nodules. Ultrasound in Medicine and Biology. 2021; 47(5):1227-1234

36. Borlea A, Borcan F, Sporea I, Dehelean CA, Negrea R, Cotoi L et al. Ti-rads diagnostic performance: Which algorithm is superior and how elastography and 4d vascularity improve the malignancy risk assessment. Diagnostics. 2020; 10(4):26

37. Bozbora A, Erbil Y, Ozarmagan S, Barbaros U, Sari S, Degirmenci B. Color Doppler sonography in cold thyroid nodules for malignancy prediction. Acta Chirurgica Belgica. 2002; 102(4):259-262

38. Brandenstein M, Wiesinger I, Jung F, Stroszczynski C, Jung EM. High-performance sonographical multimodal imaging of non cystic thyroid lesions: Chances of the preoperative diagnostics in relation to histopathology. Clinical Hemorheology and Microcirculation. 2021; 79(1):27-38

39. Brandler TC, Yee J, Zhou F, Cho M, Cangiarella J, Wei XJ et al. Does noninvasive follicular thyroid neoplasm with papillary-like nuclear features have distinctive features on sonography? Diagnostic Cytopathology. 2018; 46(2):139-147

40. Brito JP, Singh-Ospina N, Gionfriddo MR, Maraka S, Espinosa De Ycaza A, Rodriguez-Gutierrez R et al. Restricting ultrasound thyroid fine needle aspiration biopsy by nodule size: which tumors are we missing? A population-based study. Endocrine. 2016; 51(3):499-505

41. Brunese L, Romeo A, Iorio S, Napolitano G, Fucili S, Biondi B et al. A new marker for diagnosis of thyroid papillary cancer: B-flow twinkling sign. Journal of Ultrasound in Medicine. 2008; 27(8):1187-1194

42. 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):251-257

43. 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

44. 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):162-167

45. Cao H, Fan Q, Zhuo S, Qi T, Sun H, Rong X et al. The value of chinese thyroid imaging report and data system combined with contrast-enhanced ultrasound scoring in differential diagnosis of benign and malignant thyroid nodules. Journal of Ultrasound in Medicine. 2021; Epub

46. Cao J, Huang W, Huang P, Huang Y. ACR TI-RADS and ATA ultrasound classifications are helpful for the management of thyroid nodules located in the isthmus. Clinical Hemorheology and Microcirculation. 2021; Epub

47. Castellana M, Piccardo A, Virili C, Scappaticcio L, Grani G, Durante C et al. Can ultrasound systems for risk stratification of thyroid nodules identify follicular carcinoma? Cancer Cytopathology. 2020; 128(4):250-259

48. Celletti I, Fresilli D, De Vito C, Bononi M, Cardaccio S, Cozzolino A et al. TIRADS, SRE and SWE in INDETERMINATE thyroid nodule characterization: Which has better diagnostic performance? Radiologia Medica. 2021; 126(9):1189-1200

49. 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

50. Chanda R, Kandagaddala M, Moses V, Sigamani E, Keshava SN, Janakiraman R. Role of ultrasound acoustic radiation force impulse in differentiating benign from malignant superficial lymph nodes. Journal of Clinical Imaging Science. 2020; 10(1):18

51. Chang Y, Paul AK, Kim N, Baek JH, Choi YJ, Ha EJ et al. Computer-aided diagnosis for classifying benign versus malignant thyroid nodules based on ultrasound images: A comparison with radiologist-based assessments. Medical Physics. 2016; 43(1):554

52. Chen BD, Xu HX, Zhang YF, Liu BJ, Guo LH, Li DD et al. Calcification of thyroid nodules increases shear-wave speed (SWS) measurement: using multiple calcification-specific SWS cutoff values outperforms a single uniform cutoff value in diagnosing malignant thyroid nodules. Oncotarget. 2016; 7(40):66149-66159

53. Chen L, Zhan J, Diao XH, Liu YC, Shi YX, Chen Y et al. Additional value of superb microvascular imaging for thyroid nodule classification with the thyroid imaging reporting and data system. Ultrasound in Medicine and Biology. 2019; 45(8):2040-2048

54. 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

55. 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

56. Chen SJ, Chang CY, Chang KY, Tzeng JE, Chen YT, Lin CW et al. Classification of the thyroid nodules based on characteristic sonographic textural feature and correlated histopathology using hierarchical support vector machines. Ultrasound in Medicine and Biology. 2010; 36(12):2018-2026

57. Chen X, Gao M, Hu L, Zhu J, Zhang S, Wei X. The diagnostic value of the ultrasound gray scale ratio for different sizes of thyroid nodules. Cancer Medicine. 2019; 8(18):7644-7649

58. 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):4689-4692

59. 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):541-547

60. 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 prediction of malignancy in histologically proven thyroid nodules. Singapore Medical Journal. 2018; 59(11):578-583

61. Cho YJ, Ha EJ, Han M, Choi JW. Us elastography using carotid artery pulsation may increase the diagnostic accuracy for thyroid nodules with us-pathology discordance. Ultrasound in Medicine and Biology. 2017; 43(8):1587-1595

62. Choi JW, Lee JH, Baek JH, Choi BS, Jeong KS, Ryu JS et al. Diagnostic accuracy of ultrasound and 18-F-FDG PET or PET/CT for patients with suspected recurrent papillary thyroid carcinoma. Ultrasound in Medicine and Biology. 2010; 36(10):1608-1615

63. Choi WJ, Park JS, Kim KG, Kim SY, Koo HR, Lee YJ. Computerized analysis of calcification of thyroid nodules as visualized by ultrasonography. European Journal of Radiology. 2015; 84(10):1949-1953

64. 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

65. Chung SR, Ahn HS, Choi YJ, Lee JY, Yoo RE, Lee YJ et al. Diagnostic performance of the modified korean thyroid imaging reporting and data system for thyroid malignancy: A multicenter validation study. Korean journal of radiology. 2021; 22(9):1579-1586

66. Ciledag N, Arda K, Aribas BK, Aktas E, Kose SK. The utility of ultrasound elastography and MicroPure imaging in the differentiation of benign and malignant thyroid nodules. AJR American Journal of Roentgenology. 2012; 198(3):W244-249

67. Clark TJ, McKinney K, Jensen A, Patel NU. Risk threshold algorithm for thyroid nodule management demonstrates increased specificity and diagnostic accuracy as compared with american college of radiology thyroid imaging, reporting and data system; society of radiologists in ultrasound; and american thyroid association management guidelines. Ultrasound Quarterly. 2019; 35(3):224-227

68. Cohen O, Tzelnick S, Lahav Y, Schindel D, Halperin D, Yehuda M. Selection of atypia/follicular lesion of unknown significance patients for surgery versus active surveillance, without using genetic testing: A single institute experience, prospective analysis, and recommendations. Thyroid. 2017; 27(7):928-935

69. Cohen O, Zornitzki T, Yarkoni TR, Lahav Y, Schindel D, Halperin D et al. Follow-up of large thyroid nodules without surgery: Patient selection and long-term outcomes. Head and Neck. 2019; 41(6):1696-1702

70. Colakoglu B, Yildirim D, Alis D, Ucar G, Samanci C, Ustabasioglu FE et al. Elastography in distinguishing benign from malignant thyroid nodules. Journal of Clinical Imaging Science. 2016; 6:51

71. Cordes M, Nagel H, Horstrup K, Sasiadek M, Kuwert T. Ultrasound characteristics of thyroid nodules diagnosed as follicular neoplasms by fine-needle aspiration cytology. A prospective study with histological correlation. Nuclear-Medizin. 2016; 55(3):93-98

72. 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

73. da Silva PF, de Araujo Arcoverde LC, de Siqueira Barbosa Arcoverde L, Wanderley Fernandes Lima GT, de Medeiros Lima TP, do Amaral FJ et al. Agreement between ACR TI-RADS and EU TI-RADS scoring systems in the diagnosis of 473 thyroid nodules from a single-center in Brazil. Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists. 2021; 27(11):1108-1113

74. Daniels KE, Xu J, Liu JB, Chen X, Huang K, Patel J et al. Diagnostic value of TI-RADS classification system and next generation genetic sequencing in indeterminate thyroid nodules. Academic Radiology. 2021; 28(12):1685-1691

75. Dawoud MM, Dawoud RM. Added value of strain elastosonography in prediction of malignancy in solitary thyroid nodule. Egyptian Journal of Radiology and Nuclear Medicine. 2017; 48(4):905-912

76. 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 [Electronic Resource]. 2014; 9(3):e90674

77. 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

78. Deveci MS, Deveci G, LiVolsi VA, Gupta PK, Baloch ZW. Concordance between thyroid nodule sizes measured by ultrasound and gross pathology examination: effect on patient management. Diagnostic Cytopathology. 2007; 35(9):579-583

79. Dhayalan PM, Phansalkar DS, Kiran CM, Rao H, Philip PK, Palanisamy M et al. Ultrasound elastography in differentiating benign and malignant thyroid lesions and its correlation with fine needle aspiration cytology or biopsy results. Journal of Clinical and Diagnostic Research. 2018; 12(12):TC05-TC09

80. Dighe M, Bae U, Richardson ML, Dubinsky TJ, Minoshima S, Kim Y. Differential diagnosis of thyroid nodules with US elastography using carotid artery pulsation. Radiology. 2008; 248(2):662-669

81. Dobruch-Sobczak K, Adamczewski Z, Szczepanek-Parulska E, Migda B, Wolinski K, Krauze A et al. Histopathological verification of the diagnostic performance of the eu-tirads classification of thyroid nodules-results of a multicenter study performed in a previously iodine-deficient region. Journal of Clinical Medicine. 2019; 8(11):25

82. Dong Y, Mao M, Zhan W, Zhou J, Zhou W, Yao J et al. Size and ultrasound features affecting results of ultrasound-guided fine-needle aspiration of thyroid nodules. Journal of Ultrasound in Medicine. 2018; 37(6):1367-1377

83. 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

84. 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; 9:1311-1317

85. Dy JG, 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

86. 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

87. 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

88. 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):148-158

89. Esfahanian F, Aryan A, Ghajarzadeh M, Yazdi MH, Nobakht N, Burchi M. Application of sonoelastography in differential diagnosis of benign and malignant thyroid nodules. International Journal of Preventive Medicine. 2016; 7:55

90. Fang D, Ma W, Xu L, Liu Y, Ma X, Lu H. A predictive model to distinguish papillary thyroid carcinomas from benign thyroid nodules using ultrasonographic features: A single-center, retrospective analysis. Medical Science Monitor. 2019; 25:9409-9415

91. Farghadani M, Tabatabaei SA, Barikbin R, Shahsanai A, Riahinezhad M, Jafarpishe S. Comparing the sensitivity and specificity of two-dimensional shear wave elastography and fine needle aspiration in determining malignant thyroid nodules. Advanced Biomedical Research. 2019; 8:30

92. 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(5):263-271

93. Ferrari FS, Megliola A, Scorzelli A, Guarino E, Pacini F. Ultrasound examination using contrast agent and elastosonography in the evaluation of single thyroid nodules: Preliminary results. Journal of Ultrasound. 2008; 11(2):47-54

94. Freire da Silva P, Correa de Araujo Arcoverde L, de Siqueira Barbosa Arcoverde L, Tenorio Wanderley Fernandes Lima G, Paes de Medeiros Lima T, Jose do Amaral F et al. Agreement between american and european thyroid imaging, reporting, and data system (TIRADS) in the diagnosis of 473 thyroid nodules from a single center in brazil. Endocrine Practice. 2021; 27(11):1108-1113

95. Fresilli D, Grani G, De Pascali ML, Alagna G, Tassone E, Ramundo V et al. Computer-aided diagnostic system for thyroid nodule sonographic evaluation outperforms the specificity of less experienced examiners. Journal of Ultrasound. 2020; 23(2):169-174

96. Friedrich-Rust M, Romenski O, Meyer G, Dauth N, Holzer K, Grunwald F et al. Acoustic Radiation Force Impulse-Imaging for the evaluation of the thyroid gland: a limited patient feasibility study. Ultrasonics. 2012; 52(1):69-74

97. Friedrich-Rust M, Sperber A, Holzer K, Diener J, Grunwald F, Badenhoop K et al. Real-time elastography and contrast-enhanced ultrasound for the assessment of thyroid nodules. Experimental and Clinical Endocrinology and Diabetes. 2010; 118(9):602-609

98. Friedrich-Rust M, Vorlaender C, Dietrich CF, Kratzer W, Blank W, Schuler A et al. Evaluation of strain elastography for differentiation of thyroid nodules: Results of a prospective DEGUM multicenter study. Ultraschall in der Medizin. 2016; 37(3):262-270

99. Fukuhara T, Matsuda E, Donishi R, Koyama S, Miyake N, Fujiwara K et al. Clinical efficacy of novel elastography using acoustic radiation force impulse (ARFI) for diagnosis of malignant thyroid nodules. Laryngoscope investigative otolaryngology. 2018; 3(4):319-325

100. Gacayan RJ, Kasala R, Puno-Ramos P, Mojica DJ, Castro K. Comparison of the diagnostic performance of ultrasound-based thyroid imaging reporting and data system (TIRADS) classification with american thyroid association (ATA) guidelines in the prediction of thyroid malignancy in a single tertiary center in Manila, Philippines. Journal of the Asean Federation of Endocrine Societies. 2021; 36(1):69-75

101. Galimzianova A, Siebert SM, Kamaya A, Rubin DL, Desser TS. Quantitative framework for risk stratification of thyroid nodules with ultrasound: A step toward automated triage of thyroid cancer. AJR American Journal of Roentgenology. 2020; 214(4):885-892

102. 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

103. 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

104. Gao L, Xi X, Jiang Y, Yang X, Wang Y, Zhu S et al. Comparison among TIRADS (ACR TI-RADS and KWAK- TI-RADS) and 2015 ATA Guidelines in the diagnostic efficiency of thyroid nodules. Endocrine. 2019; 64(1):90-96

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

106. Garg M, Khandelwal D, Aggarwal V, Raja KB, Kalra S, Agarwal B et al. Ultrasound elastography is a useful adjunct to conventional ultrasonography and needle aspiration in preoperative prediction of malignancy in thyroid nodules: A northern india perspective. Indian Journal of Endocrinology and Metabolism. 2018; 22(5):589-596

107. Ghai S, O'Brien C, Goldstein DP, Sawka AM, Canadian Thyroid Cancer Active Surveillance Study G. Ultrasound in active surveillance for low-risk papillary thyroid

cancer: imaging considerations in case selection and disease surveillance. Insights Into Imaging. 2021; 12(1):130

108. Giammanco M, Di Gesu G, Massenti MF, Di Trapani B, Vetri G. Role of color flow Doppler sonography in pre-operative diagnostics of the thyroid pathology. Minerva Endocrinologica. 2002; 27(1):1-10

109. Gitto S, Grassi G, De Angelis C, Monaco CG, Sdao S, Sardanelli F et al. A computer-aided diagnosis system for the assessment and characterization of low-to-high suspicion thyroid nodules on ultrasound. Radiologia Medica. 2019; 124(2):118-125

110. Goel S, Malhotra A, Agarwal A, Chandak S, Kumar A, Khan A. Comparative efficacy of ultrasonography and acoustic radiation force impulse (ARFI) elastography in prediction of malignancy in thyroid nodules. Journal of Diagnostic Medical Sonography. 2020; 36(5):433-443

111. Gorgulu FF. Which is the best reference tissue for strain elastography in predicting malignancy in thyroid nodules, the sternocleidomastoid muscle or the thyroid parenchyma? Journal of Ultrasound in Medicine. 2019; 38(11):3053-3064

112. Gorgulu O, Gorgulu FF, Koc AS. Can the unnecessary operations for suspected thyroid nodules be avoided by the combined use of the strain ratio and elastography score? Revista Brasileira de Otorrinolaringologia. 2021; 87(3):338-345

113. 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

114. Goundan PN, Mamou J, Rohrbach D, Smith J, Patel H, Wallace KD et al. A preliminary study of quantitative ultrasound for cancer-risk assessment of thyroid nodules. Frontiers in Endocrinology. 2021; 12:627698

115. Grani G, Brenta G, Trimboli P, Falcone R, Ramundo V, Maranghi M et al. Sonographic risk stratification systems for thyroid nodules as rule-out tests in older adults. Cancers. 2020; 12(9):30

116. Grani G, D'Alessandri M, Carbotta G, Nesca A, Del Sordo M, Alessandrini S et al. Grey-scale analysis improves the ultrasonographic evaluation of thyroid nodules. Medicine. 2015; 94(27):e1129

117. Gray SL, O'Neill G, McGarry G. The predictive value of structured ultrasonographic staging for thyroid nodules. Journal of Laryngology and Otology. 2014; 128(10):914-921

118. Grazhdani H, Cantisani V, Lodise P, Di Rocco G, Proietto MC, Fioravanti E et al. Prospective evaluation of acoustic radiation force impulse technology in the differentiation of thyroid nodules: accuracy and interobserver variability assessment. Journal of Ultrasound. 2014; 17(1):13-20

119. Gregory A, Bayat M, Kumar V, Denis M, Kim BH, Webb J et al. Differentiation of benign and malignant thyroid nodules by using comb-push ultrasound shear elastography: A preliminary two-plane view study. Academic Radiology. 2018; 25(11):1388-1397

120. Griffin A, Brito JP, Bahl M, Hoang JK. Applying criteria of active surveillance to low-risk papillary thyroid cancer over a decade: How many surgeries and complications can be avoided? Thyroid. 2017; 27(4):518-523

121. 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):763-771

122. Guan Q, Wang Y, Du J, Qin Y, Lu H, Xiang J et al. Deep learning based classification of ultrasound images for thyroid nodules: a large scale of pilot study. Annals of Translational Medicine. 2019; 7(7):137

123. 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):464-472

124. Gulcelik NE, Gulcelik MA, Kuru B. Risk of malignancy in patients with follicular neoplasm: predictive value of clinical and ultrasonographic features. Archives of Otolaryngology -- Head & Neck Surgery. 2008; 134(12):1312-1315

125. Guo BL, Ouyang FS, Ouyang LZ, Liu ZW, Lin SJ, Meng W et al. Development and validation of an ultrasound-based nomogram to improve the diagnostic accuracy for malignant thyroid nodules. European Radiology. 2019; 29(3):1518-1526

126. Ha EJ, Baek JH, Lee JH, Lee HY, Song DE, Kim JK et al. A focal marked hypoechogenicity within an isoechoic thyroid nodule: is it a focal malignancy or not? Acta Radiologica. 2015; 56(7):814-819

127. 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

128. 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 college of radiology guidelines. Thyroid. 2018; 28(11):1532-1537

129. Ha EJ, Shin JH, Na DG, Jung SL, Lee YH, Paik W et al. Comparison of the diagnostic performance of the modified Korean Thyroid Imaging Reporting and Data System for thyroid malignancy with three international guidelines. Ultrasonography. 2021; 40(4):594-601

130. 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

131. Ha SM, Chung YJ, Ahn HS, Baek JH, Park SB. Echogenic foci in thyroid nodules: diagnostic performance with combination of TIRADS and echogenic foci. BMC Medical Imaging. 2019; 19(1):28

132. Hachim MH. The value of ultrasound examination in iraqi patients with thyroid nodules: A cross-sectional study. Journal of Pharmaceutical Sciences and Research. 2018; 10(9):2412-2414

133. Hahn SY, Shin JH, Ko EY, Bae JM, Choi JS, Park KW. Complementary role of elastography using carotid artery pulsation in the ultrasonographic assessment of thyroid nodules: A prospective study. Korean journal of radiology. 2018; 19(5):992-999

134. 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. Radiologia Medica. 2015; 120(6):579-583

135. Han DY, Sohn YM, Seo M, Yun SJ, Park WS, Jeon SH et al. Shear-wave elastography in thyroid ultrasound: Can be a predictor of extrathyroidal extension and

cervical lymph node metastasis in papillary thyroid carcinoma? Medicine. 2020; 99(52):e23654

136. Han M, Ha EJ, Park JH. Computer-aided diagnostic system for thyroid nodules on ultrasonography: Diagnostic performance based on the thyroid imaging reporting and data system classification and dichotomous outcomes. AJNR American journal of neuroradiology. 2021; 42(3):559-565

137. Han RJ, Du J, Li FH, Zong HR, Wang JD, Shen YL et al. Comparisons and combined application of two-dimensional and three-dimensional real-time shear wave elastography in diagnosis of thyroid nodules. Journal of Cancer. 2019; 10(9):1975-1984

138. Han Y, Wu JQ, Hou XJ, Sun JW, Piao ZY, Teng F et al. Strain imaging in the evaluation of thyroid nodules: The associated factors leading to misdiagnosis. Ultrasound in Medicine and Biology. 2021; 47(12):3372-3383

139. Hang J, Li F, Qiao XH, Ye XH, Li A, Du LF. Combination of maximum shear wave elasticity modulus and tirads improves the diagnostic specificity in characterizing thyroid nodules: A retrospective study. International Journal of Endocrinology Print. 2018; 2018:4923050

140. Haskjold OI, Foshaug HS, Iversen TB, Kjoren HC, Brun VH. Prediction of thyroid nodule histopathology by expert ultrasound evaluation. Endocrine Connections. 2021; 10(7):776-781

141. Hayes C, Shvarts Y, Sewgolam R, Nguyen T, Ussher S. Reducing unnecessary thyroid fine needle aspirations using American College of Radiology's thyroid imaging reporting and data system: A 5-year retrospective audit. Sonography. 2021; 9(1):29-32

142. He Y, Wang XY, Hu Q, Chen XX, Ling B, Wei HM. Value of contrast-enhanced ultrasound and acoustic radiation force impulse imaging for the differential diagnosis of benign and malignant thyroid nodules. Frontiers in Pharmacology. 2018; 9:1363

143. He YP, Xu HX, Wang D, Li XL, Ren WW, Zhao CK et al. First experience of comparisons between two different shear wave speed imaging systems in differentiating malignant from benign thyroid nodules. Clinical Hemorheology and Microcirculation. 2017; 65(4):349-361

144. Hekimsoy I, Ozturk E, Ertan Y, Orman MN, Kavukcu G, Ozgen AG et al. Diagnostic performance rates of the ACR-TIRADS and EU-TIRADS based on histopathological evidence. Diagnostic & Interventional Radiology. 2021; 27(4):511-518

145. Hess J, Schafernak K, Newbern D, Vern-Gross T, Foote J, Van Tassel D et al. Ultrasound is superior to palpation for thyroid cancer detection in high-risk childhood cancer and BMT survivors. Supportive Care in Cancer. 2020; 28(11):5117-5124

146. Hong MJ, Na DG, Baek JH, Sung JY, Kim JH. Impact of nodule size on malignancy risk differs according to the ultrasonography pattern of thyroid nodules. Korean journal of radiology. 2018; 19(3):534-541

147. 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):861-867

148. 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):707-716

149. Horvath E, Silva CF, Majlis S, Rodriguez I, Skoknic V, Castro A et al. Prospective validation of the ultrasound based TIRADS (Thyroid Imaging Reporting And Data System) classification: results in surgically resected thyroid nodules. European Radiology. 2017; 27(6):2619-2628

150. Hu L, Liu X, Pei C, Xie L, He N. Assessment of perinodular stiffness in differentiating malignant from benign thyroid nodules. Endocrine Connections. 2021; 10(5):492-501

151. Hu S, Zhang H, Sun Z, Ge Y, Li J, Yu C et al. Preoperative assessment of extrathyroidal extension of papillary thyroid carcinomas by ultrasound and magnetic resonance imaging: a comparative study. Radiologia Medica. 2020; 125(9):870-876

152. Hu Y, Li P, Jiang S, Li F. Quantitative analysis of suspicious thyroid nodules by contrast-enhanced ultrasonography. International Journal of Clinical and Experimental Medicine. 2015; 8(7):11786-11793

153. Hu YL, Cao XY, Zhou YR, Ye XH, Wang JX, Li X et al. Management of sonographically suspicious thyroid nodules 1 cm or smaller and candidacy for active surveillance: Experience of a tertiary center in China. Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists. 2021; 27(9):903-911

154. Huang K, Bai Z, Bian D, Yang P, Li X, Liu Y. Diagnostic accuracy of contrast-enhanced ultrasonography in papillary thyroid microcarcinoma stratified by size. Ultrasound in Medicine and Biology. 2020; 46(2):269-274

155. Huang S, Meng N, Pan M, Yu B, Liu J, Deng K et al. Diagnostic performances of the KWAK-TIRADS classification, elasticity score, and Bethesda System for Reporting Thyroid Cytopathology of TI-RADS category 4 thyroid nodules. International Journal of Clinical and Experimental Pathology. 2020; 13(5):1159-1168

156. 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):10956-10963

157. Huang Y, Hong Y, Xu W, Song K, Huang P. Contrast-Enhanced Ultrasound Improves the Accuracy of the ACR TI-RADS in the Diagnosis of Thyroid Nodules Located in the Isthmus. Ultraschall in der Medizin. 2021; 03:03

158. Huang Y, Zhou H, Zhang C, Hong Y, Ye Q, Huang P. Diagnostic performance of ultrasound strain elastography in transverse and longitudinal views in predicting malignant thyroid nodules. Ultrasound in Medicine and Biology. 2019; 45(9):2289-2297

159. Huaqun Z, Xueling L, Bei L, Ping C, Jian L, Yedong W et al. Advantages of the korean thyroid imaging reporting and data system combined with contrast-enhanced ultrasound in diagnosing papillary thyroid microcarcinoma. Iranian Journal of Radiology. 2019; 16(4):e83460

160. Huh S, Yoon JH, Lee HS, Moon HJ, Park VY, Kwak JY. Comparison of diagnostic performance of the ACR and Kwak TIRADS applying the ACR TIRADS' size thresholds for FNA. European Radiology. 2021; 31(7):5243-5250

161. Hussain M, Hameed N, Abbas N, Ali S, Akram S, Maqsood S. Positive predictive value of ultrasound in determination of malignant thyroid nodule. Pakistan Journal of Medical and Health Sciences. 2020; 14(2):408-410

162. Hwang HS, Orloff LA. Efficacy of preoperative neck ultrasound in the detection of cervical lymph node metastasis from thyroid cancer. Laryngoscope. 2011; 121(3):487-491

163. Ibrahim Y, Mohamed SE, Deniwar A, Al-Qurayshi ZH, Khan AN, Moroz K et al. The impact of thyroid nodule size on the risk of malignancy in follicular neoplasms. Anticancer Research. 2015; 35(3):1635-1639

164. 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):1269-1276

165. Ivanac G, Brkljacic B, Ivanac K, Huzjan R, Skreb F, Cikara I. Vascularisation of benign and malignant thyroid nodules: CD US evaluation. Ultraschall in der Medizin. 2007; 28(5):502-506

166. Jainulabdeen T, Ramaswamy B, Devaraja K, Paruthikunnan SM, Bhandarkar AM. Preoperative staging of differentiated thyroid carcinomas: Comparison of usg and ct with intraoperative findings and histopathology. Indian Journal of Otolaryngology & Head & Neck Surgery. 2019; 71(3):327-333

167. Jalan S, Sengupta S, Ray R, Mondal R, Phukan J, Bardhan J et al. A comparative evaluation of USG-guided FNAC with conventional FNAC in the preoperative assessment of thyroid lesions: A particular reference to cyto-histologically discordant cases. Bangladesh Journal of Medical Science. 2017; 16(2):274-281

168. Jeong EY, Kim HL, Ha EJ, Park SY, Cho YJ, Han M. Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators. European Radiology. 2019; 29(4):1978-1985

169. Jeong SH, Hong HS, Lee EH. Can nodular hyperplasia of the thyroid gland be differentiated from follicular adenoma and follicular carcinoma by ultrasonography? Ultrasound Quarterly. 2016; 32(4):349-355

170. Jiang J, Huang L, Zhang H, Ma W, Shang X, Zhou Q et al. Contrast-enhanced sonography of thyroid nodules. Journal of Clinical Ultrasound. 2015; 43(3):153-156

171. 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):138-144

172. Jiang J, Shang X, Zhang H, Ma W, Xu Y, Zhou Q et al. Correlation between maximum intensity and microvessel density for differentiation of malignant from benign thyroid nodules on contrast-enhanced sonography. Journal of Ultrasound in Medicine. 2014; 33(7):1257-1263

173. 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

174. 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):230-234

175. Kalantari S. The diagnostic value of color doppler ultrasonography in predicting thyroid nodules malignancy. International Tinnitus Journal. 2018; 22(1):35-39

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

177. Kaya O, Gulek B, Arslan M, Sozutok S, Soker G, Inan I. Does ultrasound-guided fine-needle aspiration biopsy indicated for subcen-timeter thyroid nodules? - Retrospective evaluation of 6-years single-center results. Acta Medica Mediterranea. 2019; 35(1):33-36

178. Kikovic S, Marinkovic DM, Ristic P, Karajovic J, Kuzmic-Jankovic S, Djuran Z et al. Role and importance of elastography in the diagnosis of differentiated thyroid carcinomas regarding the clinical, echosonographic, biohumoral and cytological examination and correlation of these results with definitive histopathological findings - A retrospective study. Vojnosanitetski Pregled. 2021; 78(6):589-598

179. Kim BM, Kim MJ, Kim EK, Kwak JY, Hong SW, Son EJ et al. Sonographic differentiation of thyroid nodules with eggshell calcifications. Journal of Ultrasound in Medicine. 2008; 27(10):1425-1430

180. Kim DW, Park JS, In HS, Choo HJ, Ryu JH, Jung SJ. Ultrasound-based diagnostic classification for solid and partially cystic thyroid nodules. AJNR: American Journal of Neuroradiology. 2012; 33(6):1144-1149

181. Kim GR, Lee E, Kim HR, Yoon JH, Park VY, Kwak JY. Convolutional neural network to stratify the malignancy risk of thyroid nodules: Diagnostic performance compared with the american college of radiology thyroid imaging reporting and data system implemented by experienced radiologists. AJNR: American Journal of Neuroradiology. 2021; 42(8):1513-1519

182. 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):2532-2537

183. Kim H, Kim JA, Son EJ, Youk JH, Chung TS, Park CS et al. Preoperative prediction of the extrathyroidal extension of papillary thyroid carcinoma with ultrasonography versus MRI: a retrospective cohort study. International Journal of Surgery. 2014; 12(5):544-548

184. 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):1583-1589

185. 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. 2019; 34(4):850-857

186. 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

187. Kim MH, Luo S, Ko SH, Jung SL, Lim DJ, Kim Y. Elastography can effectively decrease the number of fine-needle aspiration biopsies in patients with calcified thyroid nodules. Ultrasound in Medicine and Biology. 2014; 40(10):2329-2335

188. 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):1179-1184

189. 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):3353-3360

190. Kim SJ, Moon WK, Cho N. Sonographic criteria for fine-needle aspiration cytology in a Korean female population undergoing thyroid ultrasound screening. Acta Radiologica. 2010; 51(5):475-481

191. 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):W343-351

192. Kim SY, Lee HS, Kim EK, Moon HJ, Yoon JH, Hong JH et al. Follow-up ultrasound may be enough for thyroid nodules from 5 mm to 1 cm in size. Endocrine. 2016; 52(1):130-138

193. Ko SY, Kim EK, Moon HJ, Yoon JH, Kim HY, Kwak JY. Application of thyroid imaging reporting and data system in the ultrasound assessment of thyroid nodules according to physician experience. Ultrasound Quarterly. 2016; 32(2):126-131

194. Ko SY, Kim EK, Sung JM, Moon HJ, Kwak JY. Diagnostic performance of ultrasound and ultrasound elastography with respect to physician experience. Ultrasound in Medicine and Biology. 2014; 40(5):854-863

195. Kobayashi K, Fukata S, Miyauchi A. Diagnosis of follicular carcinoma of the thyroid: Role of sonography in preoperative diagnosis of follicular nodules. Journal of Medical Ultrasonics. 2005; 32(4):153-158

196. 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

197. 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):334-337

198. Kong J, Li JC, Wang HY, Wang YH, Zhao RN, Zhang Y et al. Role of superb micro-vascular imaging in the preoperative evaluation of thyroid nodules: Comparison with power doppler flow imaging. Journal of Ultrasound in Medicine. 2017; 36(7):1329-1337

199. Kong W, Yue X, Ren J, Tao X. A comparative analysis of diffusion-weighted imaging and ultrasound in thyroid nodules. BMC Medical Imaging. 2019; 19(1):92

200. 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

201. Kuma K, Matsuzuka F, Kobayashi A, Hirai K, Morita S, Miyauchi A et al. Outcome of long standing solitary thyroid nodules. World Journal of Surgery. 1992; 16(4):583-587; discussion 587-588

202. Kuma K, Matsuzuka F, Yokozawa T, Miyauchi A, Sugawara M. Fate of untreated benign thyroid nodules: results of long-term follow-up. World Journal of Surgery. 1994; 18(4):495-498; discussion 499

203. Kuo TC, Wu MH, Chen KY, Hsieh MS, Chen A, Chen CN. Ultrasonographic features for differentiating follicular thyroid carcinoma and follicular adenoma. Asian Journal of Surgery. 2020; 43(1):339-346

204. Kuru B, Kefeli M, Danaci M. Comparison of 5 thyroid ultrasound stratification systems for differentiation of benign and malignant nodules and to avoid biopsy using histology as reference standard. Endocrine Practice. 2021; 27(11):1093-1099

205. 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 [Electronic Resource]. 2016; 11(2):e0148567

206. Lampung N, Khalid S, Khalid M, Rashid BA, Arif SH. Evaluation of the nature of thyroid nodules by duplex ultrasound and assessment of the diagnostic accuracy of TIRADS classification. Journal of Clinical and Diagnostic Research. 2018; 12(7):TC01-TC04

207. Lang BHH, Shek TWH, Chan AOK, Lo CY, Wan KY. Significance of size of persistent/recurrent central nodal disease on surgical morbidity and response to therapy in reoperative neck dissection for papillary thyroid carcinoma. Thyroid. 2017; 27(1):67-73

208. Latif MA, El Rakhawy MM, Saleh MF. Diagnostic accuracy of B-mode ultrasound, ultrasound elastography and diffusion weighted MRI in differentiation of thyroid nodules (prospective study). Egyptian Journal of Radiology and Nuclear Medicine. 2021; 52(1):256

209. Leboulleux S, Girard E, Rose M, Travagli JP, Sabbah N, Caillou B et al. Ultrasound criteria of malignancy for cervical lymph nodes in patients followed up for differentiated thyroid cancer. Journal of Clinical Endocrinology and Metabolism. 2007; 92(9):3590-3594

210. Lee HK, Hur MH, Ahn SM. Diagnosis of occult thyroid carcinoma by ultrasonography. Yonsei Medical Journal. 2003; 44(6):1040-1044

211. Lee HS, Park HS, Kim SW, Choi G, Park HS, Hong JC et al. Clinical characteristics of papillary thyroid microcarcinoma less than or equal to 5 mm on ultrasonography. European Archives of Oto-Rhino-Laryngology. 2013; 270(11):2969-2974

212. Lee JH, Han K, Kim EK, Moon HJ, Yoon JH, Park VY et al. Risk stratification of thyroid nodules with atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS) cytology using ultrasonography patterns defined by the 2015 ata guidelines. Annals of Otology, Rhinology and Laryngology. 2017; 126(9):625-633

213. Lee JH, Han K, Kim EK, Moon HJ, Yoon JH, Park VY et al. Validation of the modified 4-tiered categorization system through comparison with the 5-tiered categorization system of the 2015 American Thyroid Association guidelines for classifying small thyroid nodules on ultrasound. Head and Neck. 2017; 39(11):2208-2215

214. Lee MJ, Kim EK, Kwak JY, Kim MJ. Partially cystic thyroid nodules on ultrasound: Probability of malignancy and sonographic differentiation. Thyroid. 2009; 19(4):341-346

215. Lee SK, Rho BH, Woo SK. Hurthle cell neoplasm: correlation of gray-scale and power Doppler sonographic findings with gross pathology. Journal of Clinical Ultrasound. 2010; 38(4):169-176

216. 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):559-567

217. 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

218. 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

219. Li H, Kang C, Xue J, Jing L, Miao J. Influence of lesion size on shear wave elastography in the diagnosis of benign and malignant thyroid nodules. Scientific Reports. 2021; 11(1):21616

220. 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

221. 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):2183-2190

222. Li M, Xi X, Zhang W. The diagnostic value of ultrasound and CT combined with miR-200c in thyroid cancer. Acta Medica Mediterranea. 2020; 36(3):1849-1856

223. Li RQ, Yuan GH, Chen M, Shao YM, Zhu SN, Zhang JQ et al. Evaluation of diagnostic efficiency of ultrasound features on malignant thyroid nodules in chinese patients. Chinese Medical Journal. 2016; 129(15):1784-1788

224. Li T, Jiang Z, Lu M, Zou S, Wu M, Wei T et al. Computer-aided diagnosis system of thyroid nodules ultrasonography: Diagnostic performance difference between computer-aided diagnosis and 111 radiologists. Medicine. 2020; 99(23):e20634

225. Li W, Song Q, Lan Y, Li J, Yan L, Li Y et al. The value of sonography in distinguishing follicular thyroid carcinoma from adenoma. Cancer Management and Research. 2021; 13:3991-4002

226. Li X, Gao F, Li F, Han XX, Shao SH, Yao MH et al. Qualitative analysis of contrast-enhanced ultrasound in the diagnosis of small, TR3-5 benign and malignant thyroid nodules measuring <=1 cm. British Journal of Radiology. 2020; 93(1111):20190923

227. Liang J, Huang X, Hu H, Liu Y, Zhou Q, Cao Q et al. Predicting malignancy in thyroid nodules: Radiomics score versus 2017 american college of radiology thyroid imaging, reporting and data system. Thyroid. 2018; 28(8):1024-1033

228. Liao LJ, Chen HW, Hsu WL, Chen YS. Comparison of strain elastography, shear wave elastography, and conventional ultrasound in diagnosing thyroid nodules. Journal of Medical Ultrasound. 2019; 27(1):26-32

229. 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

230. Lim-Dunham JE, Toslak IE, Reiter MP, Martin B. Assessment of the american college of radiology thyroid imaging reporting and data system for thyroid nodule malignancy risk stratification in a pediatric population. AJR American Journal of Roentgenology. 2019; 212(1):188-194

231. 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):853-858

232. Lin J. Clinical value of ultrasonic elastography in diagnosing thyroid carcinoma. Biomedical Research (India). 2018; 29(2):293-295

233. Lin JD, Huang BY, Weng HF, Jeng LB, Hsueh C. Thyroid ultrasonography with fine-needle aspiration cytology for the diagnosis of thyroid cancer. Journal of Clinical Ultrasound. 1997; 25(3):111-118

234. Lin WC, Kan NN, Chen HL, Luo SD, Tung YC, Chen WC et al. Efficacy and safety of single-session radiofrequency ablation for benign thyroid nodules of different sizes: a retrospective study. International Journal of Hyperthermia. 2020; 37(1):1082-1089

235. 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):617-624

236. 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):E1826-1830

237. Liu B, Liang J, Zheng Y, Xie X, Huang G, Zhou L et al. Two-dimensional shear wave elastography as promising diagnostic tool for predicting malignant thyroid nodules: a prospective single-centre experience. European Radiology. 2015; 25(3):624-634

238. Liu BJ, Zhang YF, Zhao CK, Wang HX, Li MX, Xu HX. Conventional ultrasound characteristics, TI-RADS category and shear wave speed measurement between follicular adenoma and follicular thyroid carcinoma. Clinical Hemorheology and Microcirculation. 2020; 75(3):291-301

239. 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

240. 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

241. Liu J, Guo Y, Xiao J, Chen L, Liang Z. Comparison of the efficacy and safety of the american thyroid association guidelines and american college of radiology TI-RADS. Endocrine Practice. 2021; 27(7):661-667

242. 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

243. Liu QQ, Yang ZW, Tian J, Xing P, Tang HL, Qiu QY et al. The application of qualitative shear wave elastography in differential diagnosis of thyroid nodules. International Journal of Clinical and Experimental Medicine. 2019; 12(4):3569-3579

244. Liu X, Medici M, Kwong N, Angell TE, Marqusee E, Kim MI et al. Bethesda categorization of thyroid nodule cytology and prediction of thyroid cancer type and prognosis. Thyroid. 2016; 26(2):256-261

245. Liu Y, Wu H, Zhou Q, Gou J, Xu J, Liu Y et al. Diagnostic value of conventional ultrasonography combined with contrast-enhanced ultrasonography in thyroid

imaging reporting and data system (ti-rads) 3 and 4 thyroid micronodules. Medical Science Monitor. 2016; 22:3086-3094

246. Liu Z, Jing H, Han X, Shao H, Sun YX, Wang QC et al. Shear wave elastography combined with the thyroid imaging reporting and data system for malignancy risk stratification in thyroid nodules. Oncotarget. 2017; 8(26):43406-43416

247. 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):933-938

248. Lu R, Meng Y, Zhang Y, Zhao W, Wang X, Jin M et al. Superb microvascular imaging (SMI) compared with conventional ultrasound for evaluating thyroid nodules. BMC Medical Imaging. 2017; 17(1):65

249. Lu Y, Shi XQ, Zhao X, Song D, Li J. Value of computer software for assisting sonographers in the diagnosis of thyroid imaging reporting and data system grade 3 and 4 thyroid space-occupying lesions. Journal of Ultrasound in Medicine. 2019; 38(12):3291-3300

250. Luo S, Lim DJ, Kim Y. Objective ultrasound elastography scoring of thyroid nodules using spatiotemporal strain information. Medical Physics. 2012; 39(3):1182-1189

251. 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):604-613

252. Lyshchik A, Higashi T, Asato R, Tanaka S, Ito J, Mai JJ et al. Thyroid gland tumor diagnosis at US elastography. Radiology. 2005; 237(1):202-211

253. Lyshchik A, Moses R, Barnes SL, Higashi T, Asato R, Miga MI et al. Quantitative analysis of tumor vascularity in benign and malignant solid thyroid nodules. Journal of Ultrasound in Medicine. 2007; 26(6):837-846

254. Lyu YJ, Shen F, Yan Y, Situ MZ, Wu WZ, Jiang GQ et al. Ultrasound-guided fine-needle aspiration biopsy of thyroid nodules <10 mm in the maximum diameter: Does size matter? Cancer Management and Research. 2019; 11:1231-1236

255. Ma HJ, Yang JC, Leng ZP, Chang Y, Kang H, Teng LH. Preoperative prediction of papillary thyroid microcarcinoma via multiparameter ultrasound. Acta Radiologica. 2017; 58(11):1303-1311

256. 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):355-363

257. 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

258. Maddaloni E, Briganti SI, Crescenzi A, Beretta Anguissola G, Perrella E, Taffon C et al. Usefulness of color doppler ultrasonography in the risk stratification of thyroid nodules. European Thyroid Journal. 2021; 10(4):339-344

259. Magri F, Chytiris S, Capelli V, Gaiti M, Zerbini F, Carrara R et al. Comparison of elastographic dtrain index and thyroid fine-Needle aspiration cytology in 631 thyroid nodules. Journal of Clinical Endocrinology and Metabolism. 2013; 98(12):4790-4797

260. Magri F, Chytiris S, Croce L, Molteni M, Bendotti G, Gruosso G et al. Performance of the ACR TI-RADS and EU TI-RADS scoring systems in the diagnostic work-up of

thyroid nodules in a real-life series using histology as reference standard. European Journal of Endocrinology. 2020; 183(5):521-528

261. Magri F, Chytiris S, Zerbini F, Capelli V, Gaiti M, Carbone A et al. Maximal stiffness evaluation by real-time ultrasound elastography, an improved tool for the differential diagnosis of thyroid nodules. Endocrine Practice. 2015; 21(5):474-481

262. 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(1):17

263. 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

264. Maino F, Bufano A, Dalmazio G, Campanile M, Pilli T, Forleo R et al. Validation of american thyroid association ultrasound risk-adapted approach for repeating cytology in benign thyroid nodules. Thyroid. 2021; 31(3):446-451

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

266. 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

267. Marturano I, Russo M, Malandrino P, Buscema M, La Rosa GL, Spadaro A et al. Combined use of sonographic and elastosonographic parameters can improve the diagnostic accuracy in thyroid nodules at risk of malignancy at cytological examination. Minerva Endocrinologica. 2020; 45(1):3-11

268. McClean S, Omakobia E, England RJA. Comparing ultrasound assessment of thyroid nodules using BTA U classification and ACR TIRADS measured against histopathological diagnosis. Clinical Otolaryngology. 2021; 46(6):1286-1289

269. Merhav G, Zolotov S, Mahagneh A, Malchin L, Mekel M, Beck-Razi N. Validation of TIRADS ACR risk assessment of thyroid nodules in comparison to the ATA guidelines. Journal of Clinical Imaging Science. 2021; 11:37

270. Miao S, Jing M, Sheng R, Cui D, Lu S, Zhang X et al. The analysis of differential diagnosis of benign and malignant thyroid nodules based on ultrasound reports. Gland Surgery. 2020; 9(3):653-660

271. 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

272. 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):641-648

273. Mohanapriya G, Chandrasekaran M. Is a surgeon-performed ultrasound good enough in diagnosing thyroid malignancy? Indian Journal of Endocrinology and Metabolism. 2018; 22(2):181-184

274. Mohebbi M, Dehaki MG, Mozaffari M, Hosseinpanah F. Comparison between ultrasonographic findings and fine needle aspiration cytology in differentiating

malignant and benign thyroid nodules. European Journal of Translational Myology. 2019; 29(3)

275. 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

276. 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):1410-1416

277. Moon HJ, Kwak JY, Kim MJ, Son EJ, Kim EK. Can vascularity at power Doppler US help predict thyroid malignancy? Radiology. 2010; 255(1):260-269

278. 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):1002-1013

279. 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):762-770

280. 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):562-572

281. Nabahati M, Moazezi Z, Fartookzadeh S, Mehraeen R, Ghaemian N, Sharbatdaran M. The comparison of accuracy of ultrasonographic features versus ultrasound-guided fine-needle aspiration cytology in diagnosis of malignant thyroid nodules. Journal of Ultrasound. 2019; 22(3):315-321

282. 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):775-782

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

284. 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):1357-1365

285. NHS England and NHS Improvement. {Confirm access date with HE} National Cost Collection 2019/20 Report. 2021. Available from: https://www.england.nhs.uk/publication/2019-20-national-cost-collection-data-publication/ Last accessed:

286. Nilakantan A, Venkatesh MD, Raghavan D, Datta R, Sharma V. Ultrasonography: Its role in nodular thyroid disease. Indian Journal of Otolaryngology & Head & Neck Surgery. 2007; 59(4):332-335

287. Ning CP, Jiang SQ, Zhang T, Sun LT, Liu YJ, Tian JW. The value of strain ratio in differential diagnosis of thyroid solid nodules. European Journal of Radiology. 2012; 81(2):286-291

288. Oh HS, Kwon H, Song E, Jeon MJ, Kim TY, Lee JH et al. Tumor volume doubling time in active surveillance of papillary thyroid carcinoma. Thyroid. 2019; 29(5):642-649

289. 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):243-247

290. 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

291. Okasha HH, Mansor M, Sheriba N, Assem M, Abdelfattah Y, Ashoush OA et al. Role of elastography strain ratio and TIRADS score in predicting malignant thyroid nodule. Archives of Endocrinology & Metabolism. 2021; 64(6):735-742

292. 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

293. Orhan Soylemez UP, Gunduz N. Diagnostic accuracy of five different classification systems for thyroid nodules: A prospective, comparative study. Journal of Ultrasound in Medicine. 2021; Epub

294. 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):24-28

295. Pagano L, Falco EC, Bisceglia A, Gambella A, Rossetto R, Garberoglio S et al. Retrospective analysis of the ultrasound features of resected thyroid nodules. Endocrine. 2021; 72(2):486-494

296. Paker M, Goldman T, Masalha M, Shlizerman L, Mazzawi S, Ashkenazi D et al. A comparison of two widely used risk stratification systems for thyroid nodule sonographic evaluation. Israel Medical Association Journal: IMAJ. 2021; 23(11):714-719

297. 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

298. 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 [Electronic Resource]. 2017; 12(12):e0188987

299. Paredes-Manjarrez C, Arreola-Chaidez D, Magdalena-Buitrago A, Ferreira-Hermosillo A, Avelar-Garnica JF, Arreola-Rosales R. Shear-wave elastography as a tool in the assessment of thyroid nodules. Gaceta Medica de México. 2021; 157(1):18-23

300. Parikh PP, Allan BJ, Lew JI. Surgeon-performed ultrasound predictors of malignancy in patients with Hurthle cell neoplasms of the thyroid. Journal of Surgical Research. 2013; 184(1):247-252

301. 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

302. Park JS, Son KR, Na DG, Kim E, Kim S. Performance of preoperative sonographic staging of papillary thyroid carcinoma based on the sixth edition of the AJCC/UICC TNM classification system. AJR American Journal of Roentgenology. 2009; 192(1):66-72

303. 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-E190

304. Park YJ, Kim JA, Son EJ, Youk JH, Kim EK, Kwak JY et al. Thyroid nodules with macrocalcification: sonographic findings predictive of malignancy. Yonsei Medical Journal. 2014; 55(2):339-344

305. 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):32-34

306. 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

307. Pei S, Cong S, Zhang B, Liang C, Zhang L, Liu J et al. Diagnostic value of multimodal ultrasound imaging in differentiating benign and malignant TI-RADS category 4 nodules. International Journal of Clinical Oncology. 2019; 24(6):632-639

308. Pei S, Zhang B, Cong S, Liu J, Wu S, Dong Y et al. Ultrasound real-time tissue elastography improves the diagnostic performance of the ACR thyroid imaging reporting and data system in differentiating malignant from benign thyroid nodules: A summary of 1525 thyroid nodules. International Journal of Endocrinology Print. 2020; 2020:1749351

309. 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

310. 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):720-724

311. 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):759-764

312. Polat YD, Ozturk VS, Ersoz N, Anik A, Karaman CZ. Is thyroid imaging reporting and data system useful as an adult ultrasonographic malignancy risk stratification method in pediatric thyroid nodules? Journal of Medical Ultrasound. 2019; 27(3):141-145

313. Pompili G, Tresoldi S, Primolevo A, De Pasquale L, Di Leo G, Cornalba G. Management of thyroid follicular proliferation: an ultrasound-based malignancy score to opt for surgical or conservative treatment. Ultrasound in Medicine and Biology. 2013; 39(8):1350-1355

314. Qi Q, Zhou A, Guo S, Huang X, Chen S, Li Y et al. Explore the diagnostic efficiency of chinese thyroid imaging reporting and data systems by comparing with the other four systems (ACR TI-RADS, Kwak-TIRADS, KSThR-TIRADS, and EU-TIRADS): a single-center study. Frontiers in Endocrinology. 2021; 12:763897

315. 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):1154-1162

316. 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):136-141

317. 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):2917-2922

318. 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):41-46

319. Ramundo V, Di Gioia CRT, Falcone R, Lamartina L, Biffoni M, Giacomelli L et al. Diagnostic performance of neck ultrasonography in the preoperative evaluation for extrathyroidal extension of suspicious thyroid nodules. World Journal of Surgery. 2020; 44(8):2669-2674

320. 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

321. Ren J, Liu B, Zhang LL, Li HY, Zhang F, Li S et al. A taller-than-wide shape is a good predictor of papillary thyroid carcinoma in small solid nodules. Journal of Ultrasound in Medicine. 2015; 34(1):19-26

322. Reverter JL, Vazquez F, Puig-Domingo M. Diagnostic performance evaluation of a computer-assisted imaging analysis system for ultrasound risk stratification of thyroid nodules. AJR American Journal of Roentgenology. 2019; 213(1):169-174

323. 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):887-891

324. Rosario PW, de Faria S, Bicalho L, Alves MF, Borges MA, Purisch S et al. Ultrasonographic differentiation between metastatic and benign lymph nodes in patients with papillary thyroid carcinoma. Journal of Ultrasound in Medicine. 2005; 24(10):1385-1389

325. 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

326. Rossing M, Nygaard B, Nielsen FC, Bennedbaek FN. High prevalence of papillary thyroid microcarcinoma in danish patients: a prospective study of 854 consecutive patients with a cold thyroid nodule undergoing fine-needle aspiration. European Thyroid Journal. 2012; 1(2):110-117

327. Rozenbaum A, Buffet C, Bigorgne C, Royer B, Rouxel A, Bienvenu M et al. Outcomes of active surveillance of EU-TIRADS 5 thyroid nodules. European Journal of Endocrinology. 2021; 184(5):677-686

328. Rubaltelli L, Stramare R, Tregnaghi A, Scagliori E, Cecchelero E, Mannucci M et al. The role of sonoelastography in the differential diagnosis of neck nodules. Journal of Ultrasound. 2009; 12(3):93-100

329. Sahin M, Oguz A, Tuzun D, Akkus G, Torun GI, Bahar AY et al. Effectiveness of TI-RADS and ATA classifications for predicting malignancy of thyroid nodules. Advances in Clinical & Experimental Medicine. 2021; 30(11):1133-1139

330. Sahli ZT, Karipineni F, Hang JF, Canner JK, Mathur A, Prescott JD et al. The association between the ultrasonography TIRADS classification system and surgical pathology among indeterminate thyroid nodules. Surgery. 2019; 165(1):69-74

331. Samir AE, Dhyani M, Anvari A, Prescott J, Halpern EF, Faquin WC et al. Shear-wave elastography for the preoperative risk stratification of follicular-patterned lesions of the thyroid: Diagnostic accuracy and optimal measurement plane. Radiology. 2015; 277(2):565-573

332. Sanabria A. Experience with active surveillance of thyroid low-risk carcinoma in a developing country. Thyroid. 2020; 30(7):985-991

333. Sancak S, Hardt A, Gartner R, Eszlinger M, Aslan A, Eren FT et al. Comparison of Color Flow Doppler Sonography (CFDS) and immunohistologic detection of microvessels for the assessment of the malignancy of thyroid nodules. Hormone and Metabolic Research. 2010; 42(9):670-676

334. Schenke S, Klett R, Seifert P, Kreissl MC, Gorges R, Zimny M. Diagnostic Performance of Different Thyroid Imaging Reporting and Data Systems (Kwak-TIRADS, EU-TIRADS and ACR TI-RADS) for risk stratification of small thyroid nodules (<=10 mm). Journal of Clinical Medicine. 2020; 9(1):16

335. 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

336. Schenke S, Zimny M. Combination of Sonoelastography and TIRADS for the Diagnostic Assessment of Thyroid Nodules. Ultrasound in Medicine and Biology. 2018; 44(3):575-583

337. Schleder S, Janke M, Agha A, Schacherer D, Hornung M, Schlitt HJ et al. Preoperative differentiation of thyroid adenomas and thyroid carcinomas using high resolution contrast-enhanced ultrasound (CEUS). Clinical Hemorheology and Microcirculation. 2015; 61(1):13-22

338. 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):5281-5288

339. Seifert P, Schenke S, Zimny M, Stahl A, Grunert M, Klemenz B et al. Diagnostic Performance of Kwak, EU, ACR, and Korean TIRADS as Well as ATA guidelines for the ultrasound risk stratification of non-autonomously functioning thyroid nodules in a region with long history of iodine deficiency: A german multicenter trial. Cancers. 2021; 13(17):04

340. Sengul D, Sengul I. Association between Tsukuba elasticity scores 4 and 5 on elastography and Bethesda undetermined cytology on US-guided FNA with 27-G needle, verified by histopathology: a cut-off point of 20 mm of diameter designated for thyroid nodules. Journal of BUON. 2019; 24(1):382-390

341. Sengul D, Sengul I, Egrioglu E, Ozturk T, Aydin I, Kesicioglu T et al. Can cut-off points of 10 and 15 mm of thyroid nodule predict malignancy on the basis of three diagnostic tools: i) strain elastography, ii) the Bethesda System for Reporting Thyroid Cytology with 27-gauge fine-needle, and iii) histopathology? Journal of BUON. 2020; 25(2):1122-1129

342. Seo YL, Yoon DY, Baek S, Ku YJ, Rho YS, Chung EJ et al. Detection of neck recurrence in patients with differentiated thyroid cancer: Comparison of ultrasound, contrast-enhanced CT and 18F-FDG PET/CT using surgical pathology as a reference standard: Ultrasound vs. CT vs. 18F-FDG PET/CT in recurrent thyroid cancer. European Radiology. 2012; 22(10):2246-2254

343. Shah B, Mahesan R, Vijayasekar I, Shah BB. Diagnostic performance of the american college of radiology thyroid imaging reporting and data system. World Journal of Endocrine Surgery. 2020; 12(3):113-116

344. 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):254-260

345. Sharma VK, Paulose AA, Singh P, Sonkhya N. Diagnostic efficacy of ultrasonography and fine-needle aspiration cytology in correlation with histopathology in euthyroid patients having solitary thyroid nodule. Clinical Medicine and Research. 2019; 8(1):1-5

346. Shen Y, Liu M, He J, Wu S, Chen M, Wan Y et al. Comparison of different risk-stratification systems for the diagnosis of benign and malignant thyroid nodules. Frontiers in Oncology. 2019; 9:378

347. Shi X, Liu R, Gao L, Xia Y, Jiang Y. Diagnostic value of sonographic features in distinguishing malignant partially cystic thyroid nodules: A systematic review and meta-analysis. Frontiers in Endocrinology. 2021; 12:624409

348. Shi YX, Chen L, Liu YC, Zhan J, Diao XH, Fang L et al. Differences among the thyroid imaging reporting and data system proposed by korean, the american college of radiology and the european thyroid association in the diagnostic performance of thyroid nodules. Translational Cancer Research. 2020; 9(8):4958-4967

349. Shi YZ, Jin Y, Zheng L. Partially cystic thyroid nodules on ultrasound: The associated factors for malignancy. Clinical Hemorheology and Microcirculation. 2020; 74(4):373-381

350. 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):251-258

351. Shin I, Kim YJ, Han K, Lee E, Kim HJ, Shin JH et al. Application of machine learning to ultrasound images to differentiate follicular neoplasms of the thyroid gland. Ultrasonography. 2020; 39(3):257-265

352. Shreyamsa M, Mishra A, Ramakant P, Parihar A, Singh KR, Rana C et al. Comparison of multimodal ultrasound imaging with conventional ultrasound risk stratification systems in presurgical risk stratification of thyroid nodules. Indian Journal of Endocrinology and Metabolism. 2020; 24(6):537-542

353. Shuzhen C. Comparison analysis between conventional ultrasonography and ultrasound elastography of thyroid nodules. European Journal of Radiology. 2012; 81(8):1806-1811

354. Shweel M, Mansour E. Diagnostic performance of combined elastosonography scoring and high-resolution ultrasonography for the differentiation of benign and malignant thyroid nodules. European Journal of Radiology. 2013; 82(6):995-1001

355. Skowronska A, Milczarek-Banach J, Wiechno W, Chudzinski W, Zach M, Mazurkiewicz M et al. Accuracy of the European Thyroid Imaging Reporting and Data System (EU-TIRADS) in the valuation of thyroid nodule malignancy in reference to the post-surgery histological results. Polish Journal of Radiology. 2018; 83:e579-e586

356. Sodagari N, Jafari M, Nasiri S, Molavi B, Notash AY, Mir A et al. A practical score model for detecting malignant thyroid nodules by using ultrasonographic findings. Acta Medica Iranica. 2018; 56(7):465-473

357. Sohail S, Kaliq Uz Zaman S. Diagnostic value of mean elasticity index as a quantitative shear wave elastography parameter for prediction of malignancy in small suspicious solid thyroid nodules. Jcpsp, Journal of the College of Physicians & Surgeons - Pakistan. 2020; 30(7):683-687

358. Song G, Xue F, Zhang C. A model using texture features to differentiate the nature of thyroid nodules on sonography. Journal of Ultrasound in Medicine. 2015; 34(10):1753-1760

359. Song Q, Tian X, Jiao Z, Yan L, Lan Y, Zhu Y et al. Value of conventional ultrasonography with contrast-enhanced ultrasonography in the differential diagnosis of partial cystic thyroid nodules. Ultrasound in Medicine and Biology. 2021; 47(9):2494-2501

360. Songsaeng D, Prechaverakul P, Soodchuen S, Suwanbundit A. Accuracy of elastography alone and in combination with siriraj TIRADS in predicting thyroid nodule malignancy. Journal of the Medical Association of Thailand. 2019; 102(7):730-737

361. Stoian D, Ivan V, Sporea I, Florian V, Mozos I, Navolan D et al. Advanced ultrasound application - impact on presurgical risk stratification of the thyroid nodules. Therapeutics and Clinical Risk Management. 2020; 16:21-30

362. 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):327-332

363. Studeny T, Kratzer W, Schmidberger J, Graeter T, Barth TFE, Hillenbrand A. Analysis of vascularization in thyroid gland nodes with superb microvascular imaging (SMI) and CD34 expression histology: a pilot study. BMC Medical Imaging. 2021; 21(1):159

364. 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

365. Sultan LR, Xiong H, Zafar HM, Schultz SM, Langer JE, Sehgal CM. Vascularity assessment of thyroid nodules by quantitative color Doppler ultrasound. Ultrasound in Medicine and Biology. 2015; 41(5):1287-1293

366. Sun C, Zhang Y, Chang Q, Liu T, Zhang S, Wang X et al. Evaluation of a deep learning-based computer-aided diagnosis system for distinguishing benign from malignant thyroid nodules in ultrasound images. Medical Physics. 2020; 47(9):3952-3960

367. Swan KZ, Bonnema SJ, Jespersen ML, Nielsen VE. Reappraisal of shear wave elastography as a diagnostic tool for identifying thyroid carcinoma. Endocrine Connections. 2019; 8(8):1195-1205

368. Sych YP, Fadeev VV, Fisenko EP, Kalashnikova M. Reproducibility and interobserver agreement of different thyroid imaging and reporting data systems (TIRADS). European Thyroid Journal. 2021; 10(2):161-167

369. Szczepanek-Parulska E, Wolinski K, Dobruch-Sobczak K, Antosik P, Ostalowska A, Krauze A et al. S-Detect Software vs. EU-TIRADS Classification: A Dual-Center Validation of Diagnostic Performance in Differentiation of Thyroid Nodules. Journal of Clinical Medicine. 2020; 9(8):03

370. 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 [Electronic Resource]. 2013; 8(11):e81532

371. 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):461-466

372. 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

373. Taj B, Ayaz Z, Shakoor S, Saad M, Tahir M, Laique T. Role of ultrasonography in papillary thyroid carcinoma diagnosis. Pakistan Journal of Medical and Health Sciences. 2020; 14(3):1520-1521

374. Tan S, Sun PF, Xue H, Fu S, Zhang ZP, Mei F et al. Evaluation of thyroid micro-carcinoma using shear wave elastography: Initial experience with qualitative and quantitative analysis. European Journal of Radiology. 2021; 137:109571

375. Tan WJ, Sanghvi K, Liau KH, Low CH. An audit study of the sensitivity and specificity of ultrasound, fine needle aspiration cytology and frozen section in the evaluation of thyroid malignancies in a tertiary institution. Annals of the Academy of Medicine, Singapore. 2010; 39(5):359-362

376. Tang M, Wang J, Lu L. Comparative analysis between ultrasound elastography and pathological examination on papillary thyroid carcinoma. International Journal of Clinical and Experimental Medicine. 2017; 10(3):5075-5081

377. 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):348-353

378. Tian Q, Zhu H, Li H. Significance of contrast-enhanced ultrasonography in differential diagnosis of thyroid nodules. Medicine. 2018; 97(40):e12688

379. Tian W, Hao S, Gao B, Jiang Y, Zhang S, Guo L et al. Comparison of diagnostic accuracy of real-time elastography and shear wave elastography in differentiation malignant from benign thyroid nodules. Medicine. 2015; 94(52):e2312

380. Trimboli P, Ngu R, Royer B, Giovanella L, Bigorgne C, Simo R et al. A multicentre validation study for the EU-TIRADS using histological diagnosis as a gold standard. Clinical Endocrinology. 2019; 91(2):340-347

381. Tuan PA, Duc NM, An M, Vien MV, Giang BV. The role of shear wave elastography in the discrimination between malignant and benign thyroid nodules. Acta Informatica Medica. 2020; 28(4):248-253

382. 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):1031-1038

383. 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):3965-3972

384. 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 and Diabetes. 2012; 120(10):635-638

385. Vinayak S, Sande JA. Avoiding unnecessary fine-needle aspiration cytology by accuractely predicting the benign nature of thyroid nodules using ultrasound. Journal of Clinical Imaging Science. 2012; 2:23

386. Vorlander C, Wolff J, Saalabian S, Lienenluke RH, Wahl RA. Real-time ultrasound elastography--a noninvasive diagnostic procedure for evaluating dominant thyroid nodules. Langenbecks Archives of Surgery. 2010; 395(7):865-871

387. Walsh J, An L, Mills B, Hah Z, Moalem J, Miller M et al. Quantitative crawling wave sonoelastography of benign and malignant thyroid nodules. Otolaryngology - Head & Neck Surgery. 2012; 147(2):233-238

388. 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

389. Wang F, Chang C, Gao Y, Chen YL, Chen M, Feng LQ. Does shear wave elastography provide additional value in the evaluation of thyroid nodules that are suspicious for malignancy? Journal of Ultrasound in Medicine. 2016; 35(11):2397-2404

390. 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):50-55

391. 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):1945-1949

392. 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):124-127

393. Wang J, He X, Ma L, Li M, Sun L, Jiang J et al. Multimode ultrasonic technique is recommended for the differential diagnosis of thyroid cancer. PeerJ. 2020; 8:e9112

394. Wang J, Li P, Sun L, Sun Y, Fang S, Liu X. Diagnostic value of strain ratio measurement in differential diagnosis of thyroid nodules coexisted with Hashimoto thyroiditis. International Journal of Clinical and Experimental Medicine. 2015; 8(4):6420-6426

395. Wang J, Wei W, Guo R. Ultrasonic elastography and conventional ultrasound in the diagnosis of thyroid micro-nodules. Pakistan Journal of Medical Sciences. 2019; 35(6):1526-1531

396. Wang M, Sun P, Zhao X, Sun Y. Ultrasound parameters of thyroid nodules and the risk of malignancy: A retrospective analysis. Cancer Control. 2020; 27(1):1073274820945976

397. 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):8625-8628

398. 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(1):11560

399. Wang Y, Li L, Wang YX, Feng XL, Zhao F, Zou SM et al. Ultrasound findings of papillary thyroid microcarcinoma: a review of 113 consecutive cases with histopathologic correlation. Ultrasound in Medicine and Biology. 2012; 38(10):1681-1688

400. Watkins L, O'Neill G, Young D, McArthur C. Comparison of British Thyroid Association, American College of Radiology TIRADS and Artificial Intelligence TIRADS with histological correlation: diagnostic performance for predicting thyroid malignancy and unnecessary fine needle aspiration rate. British Journal of Radiology. 2021; 94(1123):20201444

401. Weller A, Sharif B, Qarib MH, St Leger D, De Silva HS, Lingam RK. British Thyroid Association 2014 classification ultrasound scoring of thyroid nodules in predicting malignancy: Diagnostic performance and inter-observer agreement. Ultrasound. 2020; 28(1):4-13

402. Wettasinghe MC, Rosairo S, Ratnatunga N, Wickramasinghe ND. Diagnostic accuracy of ultrasound characteristics in the identification of malignant thyroid nodules. BMC Research Notes. 2019; 12(1):193

403. Woon TK, Sanamandra SK, Salkade PR, Venkatanarasimha N, Ganguly R, Ho CL et al. Subcentimetre thyroid nodules: Sonographic features associated with malignancy. Ultrasound. 2020; 28(3):155-163

404. Wu H, Deng Z, Zhang B, Liu Q, Chen J. Classifer model based on machine learning algorithms: Application to differential diagnosis of suspicious thyroid nodules via sonography. American Journal of Roentgenology. 2016; 207(4):859-864

405. Wu JH, Zeng W, Wu RG, Wang M, Ye F, Fu MY. Comparison of Ultrasonography and CT for Determining the Preoperative Benign or Malignant Nature of Thyroid Nodules: Diagnostic Performance According to Calcification. Technology in Cancer Research & Treatment. 2020; 19:1533033820948183

406. 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

407. 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

408. Xia S, Dong Y, Kang H, Zhan W. Ultrasonography is valuable in evaluation of papillary thyroid microcarcinoma based on 5 mm tumor size. Journal of Cancer Research and Therapeutics. 2018; 14(Suppl 1):S319-S323

409. Xiao J, Xiao Q, Cong W, Li T, Ding S, Shao C et al. Discriminating malignancy in thyroid nodules: The nomogram versus the kwak and ACR TI-RADS. Otolaryngology - Head & Neck Surgery. 2020; 163(6):1156-1165

410. 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):663-669

411. Xu J, Wang P, Yue W, Luo Y, Li Z. Application of ultrasonic shear wave elastography and contrast-enhanced ultrasound in the differential diagnosis of patients with benign and malignant thyroid lesions. Gland Surgery. 2020; 9(6):2136-2143

412. 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):577-586

413. 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

414. 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. 2019; 64(2):299-307

415. Xu Y, Qi X, Zhao X, Ren W, Ding W. Clinical diagnostic value of contrast-enhanced ultrasound and TI-RADS classification for benign and malignant thyroid tumors: One comparative cohort study. Medicine. 2019; 98(4):e14051

416. 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):913-916

417. Yang Q, Zhou W, Li J, Wu G, Ding F, Tian X. Comparative analysis of diagnostic value for shear wave elastography and real-time elastographic imaging for thyroid nodules. Journal of Medical Imaging and Health Informatics. 2019; 9(2):334-338

418. Yang R, Zou X, Zeng H, Zhao Y, Ma X. Comparison of diagnostic performance of five different ultrasound TI-RADS classification guidelines for thyroid nodules. Frontiers in Oncology. 2020; 10:598225

419. Yoo MH, Kim HJ, Choi IH, Park S, Yun S, Park HK et al. Efficacy of differential diagnosis of thyroid nodules by shear wave elastography-the stiffness map. Journal of the Endocrine Society. 2021; 5(11):bvab154

420. 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

421. Young JK, Lumapas CG, Mirasol R. Sonographically guided fine-needle aspiration biopsy of thyroid nodules: Correlation between cytologic and histopathologic findings. Phillippine Journal of Internal Medicine. 2011; 49(1):8-14

422. Zhang F, Chen W. Sonographic features of follicular variant of papillary thyroid carcinoma (FV-PTC) and diagnostic performance of the 2017 ACR TI-RADS in FV-PTC. Endocrine. 2020; 67(2):379-386

423. 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

424. 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):e686-690

425. 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):257-264

426. 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):5911-5917

427. Zhang L, Ding Z, Dong F, Wu H, Liang W, Tian H et al. Diagnostic performance of multiple sound touch elastography for differentiating benign and malignant thyroid nodules. Frontiers in Pharmacology. 2018; 9:1359

428. Zhang Q, Liao L, Peng Q, Tang K, Xu Y, Zhang R et al. Value of contrast-enhanced ultrasound in differentiating clinically atypical subacute thyroiditis from papillary thyroid carcinomas. Ultrasound in Medicine and Biology. 2021; 47(12):3384-3392

429. Zhang WB, Xu HX, Zhang YF, Guo LH, Xu SH, Zhao CK et al. Comparisons of ACR TI-RADS, ATA guidelines, Kwak TI-RADS, and KTA/KSThR guidelines in malignancy risk stratification of thyroid nodules. Clinical Hemorheology and Microcirculation. 2020; 75(2):219-232

430. 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

431. Zhang YX, Xue JP, Li HZ, Miao JW, Kang CS. Clinical value of shear wave elastography color scores in classifying thyroid nodules. International Journal of General Medicine. 2021; 14:8007-8018

432. 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

433. 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

434. 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

435. 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. 2019; 72(1):95-106

436. Zhao W, Lu R, Yin L, Guo R. The value of superb microvascular imaging (SMI) scoring assignment method in differentiating benign and malignant thyroid nodules by conventional ultrasound. Clinical Hemorheology and Microcirculation. 2021; 78(4):355-363

437. Zhu H, Yang Y, Wu S, Chen K, Luo H, Huang J. Diagnostic performance of US-based FNAB criteria of the 2020 Chinese guideline for malignant thyroid nodules: comparison with the 2017 American College of Radiology guideline, the 2015 American Thyroid Association guideline, and the 2016 Korean Thyroid Association guideline. Quantitative Imaging in Medicine & Surgery. 2021; 11(8):3604-3618

438. Zhuo J, Ma Z, Fu WJ, Liu SP. Differentiation of benign from malignant thyroid nodules with acoustic radiation force impulse technique. British Journal of Radiology. 2014; 87(1035):20130263

Appendices

Appendix A Review protocols

A.1 Diagnostic accuracy of Ultrasound for detecting thyroid malignancy

Field	Content
PROSPERO registration number	CRD42021244436
Review title	Diagnostic accuracy of ultrasound for identifying 1) thyroid nodule malignancies or 2) nodules with malignant potential (potential for malignant transformation)
Review question	What is the diagnostic accuracy of ultrasound for identifying thyroid nodule malignancies or nodules with malignant potential?
Objective	To identify the most accurate methods of detecting thyroid cancer in this population

Searches	The following databases from inception will be searched: • Cochrane Central Register of Controlled Trials (CENTRAL) • Cochrane Database of Systematic Reviews (CDSR) • Embase • MEDLINE Searches will be restricted by: • English language • Human studies • Letters and comments excluded Other searches: • Inclusion lists of relevant systematic reviews will be checked by the reviewer. None The searches may be re-run 6 weeks before final committee meeting and further studies retrieved for inclusion if relevant. The full search strategies will be published in the final review.

	Medline search strategy to be quality assured using the PRESS evidence-based checklist (see methods chapter for full details).
Condition or domain being studied	Thyroid cancer
Population	Inclusion: People aged 16 or over who are suspected of thyroid cancer. Exclusion: Children and young people under 16 years.
Index Test	Ultrasound Different thresholds of a positive US finding for 1) nodule malignancies: Use any thresholds used in the literature 2) nodules with malignant potential (ie follicular neoplasms): Use any thresholds used in the literature. These will be analysed separately
Comparator/Reference standard/Confounding factors	Later histopathological findings of 1) malignancy or 2) follicular adenomas with potential for malignant transformation

Types of study to be included	Cross-sectional/prospective/retrospective diagnostic studies, or any study containing a diagnostic accuracy analysis. Retrospective studies will be downgraded for indirectness (see comments on right)
Other exclusion criteria	• Studies that do not report sensitivity and specificity, or insufficient data to derive these values. • Studies where all participants do not receive the gold standard measure of histopathology Non-English language studies.
Context	Ultrasound is commonly used clinically as the first line test, but there are few up to date recommendations about the optimum thresholds to use that yield the most appropriate sensitivity and specificity for a first line test
Primary outcomes (critical outcomes)	• Sensitivity • Specificity • Raw data to calculate 2x2 tables to calculate sensitivity and specificity (number of true positives, true negatives, false positives and false negatives).
Data extraction (selection and coding)	EndNote will be used for reference management, sifting, citations and bibliographies. All references identified by the searches and from other sources will be screened for inclusion. 10% of the abstracts will be reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer.

	The full text of these potentially eligible studies will be retrieved and assessed in line with the criteria outlined above. A standardised form will be used to extract data from the included studies (see Developing NICE guidelines: the manual section 6.4). 10% of all evidence reviews are quality assured by a senior research fellow. This includes checking: • papers were included /excluded appropriately • a sample of the data extractions • correct methods are used to synthesise data • a sample of the risk of bias assessments Disagreements between the review authors over the risk of bias in particular studies will be resolved by discussion, with involvement of a third review author where necessary.
Risk of bias (quality) assessment	Risk of bias quality assessment will be assessed using QUADAS-2.
Strategy for data synthesis	Where possible data will be meta-analysed where appropriate (if at least 3 studies reporting data at the same diagnostic threshold) in WinBUGS. Summary diagnostic outcomes will be reported from the meta-analyses with their 95% confidence intervals in adapted GRADE tables. Heterogeneity will be assessed by visual inspection of the sensitivity and specificity plots and summary area under the curve (AUC) plots. Particular attention will be placed on sensitivity, determined by the committee to be the primary outcome for decision making. GRADE will be carried out for both sensitivity and specificity.

	If meta-analysis is not possible, data will be presented as individual values in adapted GRADE profile tables and plots of un-pooled sensitivity and specificity from RevMan software.
Analysis of sub-groups	Stratify: none If heterogeneity is identified, where data is available, subgroup analysis will be carried out for the following subgroups: Subgroups to investigate if heterogeneity is present 1.Expertise of individual undertaking the US (not known/medic/non medic) Decision making thresholds Sensitivity: 0.90 (threshold for possible recommendation), 0.80 (threshold for clinical usefulness) Specificity: 0.50 (threshold for possible recommendation), 0.10 (threshold for clinical usefulness) These apply to both 1) malignancy and 2) nodules with malignant potential
Type and method of review	☐ Intervention ☒ Diagnostic ☐ Prognostic ☐ Qualitative ☐ Epidemiologic ☐ Service Delivery

	☐ Other (please specify)
Language	English
Country	England
Named contact	**Named contact** National Guideline Centre **Organisational affiliation of the review** National Institute for Health and Care Excellence (NICE) and the National Guideline Centre
Review team members	From the National Guideline Centre: Carlos Sharpin, Guideline lead Mark Perry, Senior systematic reviewer Alfredo Mariani, Health economist Lina Gulhane, Head of Information specialists
Funding sources/sponsor	This systematic review is being completed by the National Guideline Centre which receives funding from NICE.
Conflicts of interest	All guideline committee members and anyone who has direct input into NICE guidelines (including the evidence review team and expert witnesses) must declare any potential conflicts of interest in line with NICE's code of practice for declaring and dealing with conflicts of interest. Any relevant interests, or changes to interests, will also be declared publicly at the start of each guideline committee meeting. Before each meeting, any potential conflicts of interest will be considered by the guideline committee Chair and a senior member of the development team. Any decisions to exclude a person from all or part of a meeting will be documented. Any changes to a member's declaration of interests will be recorded in the minutes of the meeting. Declarations of interests will be published with the final guideline.

Collaborators	Development of this systematic review will be overseen by an advisory committee who will use the review to inform the development of evidence-based recommendations in line with section 3 of Developing NICE guidelines: the manual. Members of the guideline committee are available on the NICE website: https://www.nice.org.uk/guidance/indevelopment/gid-ng10150/documents
Other registration details	N/A
Reference/URL for published protocol	https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=244436
Dissemination plans	NICE may use a range of different methods to raise awareness of the guideline. These include standard approaches such as: • notifying registered stakeholders of publication • publicising the guideline through NICE's newsletter and alerts • issuing a press release or briefing as appropriate, posting news articles on the NICE website, using social media channels, and publicising the guideline within NICE. [Add in any additional agree dissemination plans.]
Keywords	Diagnosis, Thyroid cancer
Details of existing review of same topic by same authors	N/A
Additional information	N/A

Details of final publication	www.nice.org.uk

1

A.2 Threshold of size and classification of thyroid nodule below which it is clinically and cost effective to use active surveillance

3

Field	Content
PROSPERO registration number	Not registered
Review title	The threshold of size and classification of thyroid nodule below which it is clinically and cost effective to use active surveillance or discharge rather than biopsy, in people with thyroid nodules on ultrasound at initial presentation
Review question	In people with thyroid nodules on ultrasound at initial presentation, for what size and classification is it clinically and cost effective to use active surveillance or discharge rather than biopsy?
Objective	To determine the threshold of nodule size/classification below which biopsy is not required (where harm, relative to the reference, is not manifested in the outcomes).
Searches	The following databases (from inception) will be searched: • Cochrane Central Register of Controlled Trials (CENTRAL) • Cochrane Database of Systematic Reviews (CDSR) • Embase • MEDLINE

	Searches will be restricted by: • English language • Human studies • Letters and comments are excluded. Other searches: • Inclusion lists of relevant systematic reviews will be checked by the reviewer. The searches may be re-run 6 weeks before final committee meeting and further studies retrieved for inclusion if relevant. The full search strategies will be published in the final review. Medline search strategy to be quality assured using the PRESS evidence-based checklist (see methods chapter for full details).
Condition or domain being studied	Thyroid cancer
Population	Inclusion: People aged 16 or over who are suspected of having thyroid cancer with thyroid nodules on ultrasound at initial presentation. and who have been assigned to active surveillance/discharge on the basis of the US result Exclusion: • Children and young people under 16 years.

	• People given biopsy/FNA
Intervention/Exposure/Test	Different US nodule size/characteristics. May use any established classification system, such as BTA 2014 guidance, US classifications
Comparator/Reference standard/Confounding factors	Each other
Types of study to be included	• Observational studies (prospective/retrospective cohorts)
Other exclusion criteria	Non-English language studies. Abstracts will be excluded as it is expected there will be sufficient full text published studies available.
Context	It is believed that many people with lower stage malignancies may not need active treatment. This review is aimed at identifying the threshold of US findings below which active surveillance/discharge may be a safe and effective option
Primary outcomes (critical outcomes)	All outcomes are considered equally important for decision making and therefore have all been rated as critical: • mortality • quality of life • local cancer progression • incidence of distant metastases • decision to treat • adverse events Minimum time of follow up: 2 years, except for quality of life, which is set to no minimum. This is on the basis that quality of life changes may be expected to occur rapidly.

Data extraction (selection and coding)	EndNote will be used for reference management, sifting, citations and bibliographies. Titles and/or abstracts of studies retrieved using the search strategy and those from additional sources will be screened for inclusion. The full text of potentially eligible studies will be retrieved and will be assessed for eligibility in line with the criteria outlined above. 10% of all evidence reviews are quality assured by a senior research fellow. This includes checking: • papers were included /excluded appropriately • a sample of the data extractions • correct methods are used to synthesise data • a sample of the risk of bias assessments Disagreements between the review authors over the risk of bias in particular studies will be resolved by discussion, with involvement of a third review author where necessary.
Risk of bias (quality) assessment	Risk of bias will be assessed using the appropriate checklist as described in Developing NICE guidelines: the manual. For Intervention reviews the following checklist will be used according to study design being assessed: • Non-randomised checklist: ROBINS-I
Strategy for data synthesis	• Where possible, data will be meta-analysed. Pairwise meta-analyses will be performed using Cochrane Review Manager (RevMan5). Fixed-effects (Mantel-Haenszel) techniques will be used to calculate risk ratios for the binary outcomes where possible. Continuous outcomes will be analysed using an inverse variance method for pooling weighted mean differences.

	Heterogeneity between the studies in effect measures will be assessed using the I^2 statistic and visually inspected. We will consider an I^2 value greater than 50% indicative of substantial heterogeneity. Sensitivity analyses will be conducted based on pre-specified subgroups using stratified meta-analysis to explore the heterogeneity in effect estimates. If this does not explain the heterogeneity, the results will be presented using random-effects. GRADE pro will be used to assess the quality of each outcome, taking into account individual study quality and the meta-analysis results. The 4 main quality elements (risk of bias, indirectness, inconsistency and imprecision) will be appraised for each outcome. Publication bias is tested for when there are more than 5 studies for an outcome. The risk of bias across all available evidence was evaluated for each outcome 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/. Where meta-analysis is not possible, data will be presented and quality assessed individually per outcome. If sufficient data is available to make a network of treatments, WinBUGS will be used for network meta-analysis.
Analysis of sub-groups	<u>Stratification</u> • None <u>Sub-groups that will be investigated if heterogeneity is present:</u> • Age (<u>≤</u>55, > 55) • Gender (male, female) • Prior radiation exposure (Y/N)
	☒ Intervention

Type and method of review	☐ Diagnostic ☐ Prognostic ☐ Qualitative ☐ Epidemiologic ☐ Service Delivery ☐ Other (please specify)
Language	English
Country	England
Named contact	**Named contact** National Guideline Centre **Organisational affiliation of the review** National Institute for Health and Care Excellence (NICE) and the National Guideline Centre
Review team members	From the National Guideline Centre: Calos Sharpin Mark Perry Vimal Bedia Alexandra Bonnon Lina Gulhane
Funding sources/sponsor	This systematic review is being completed by the National Guideline Centre which receives funding from NICE.
Conflicts of interest	All guideline committee members and anyone who has direct input into NICE guidelines (including the evidence review team and expert witnesses) must declare any potential conflicts of interest in line with NICE's code of practice for declaring and dealing with conflicts of interest. Any relevant interests, or changes to interests, will also be declared publicly at the start of each guideline committee meeting. Before each meeting, any potential conflicts of interest will be considered by the guideline committee Chair and a senior member of the development team. Any decisions to exclude a person from all or part of a meeting will be documented.

	Any changes to a member's declaration of interests will be recorded in the minutes of the meeting. Declarations of interests will be published with the final guideline.
Collaborators	Development of this systematic review will be overseen by an advisory committee who will use the review to inform the development of evidence-based recommendations in line with section 3 of Developing NICE guidelines: the manual. Members of the guideline committee are available on the NICE website: https://www.nice.org.uk/guidance/indevelopment/gid-ng10150/documents
Other registration details	N/A
Reference/URL for published protocol	N/A
Dissemination plans	NICE may use a range of different methods to raise awareness of the guideline. These include standard approaches such as: • notifying registered stakeholders of publication • publicising the guideline through NICE's newsletter and alerts • issuing a press release or briefing as appropriate, posting news articles on the NICE website, using social media channels, and publicising the guideline within NICE.
Keywords	Thyroid cancer
Details of existing review of same topic by same authors	N/A
Additional information	N/A
Details of final publication	www.nice.org.uk

1

2

A.3 Review protocol health economic evidence

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 2005, 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).[283] **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 2005 or later but that depend on unit costs and resource data entirely or predominantly from before 2005 will be rated as ‘Not applicable’.
- Studies published before 2005 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 these reviews are detailed below and complied with the methodology outlined in Developing NICE guidelines: the manual, 2014 (updated 2020) https://www.nice.org.uk/process/pmg20/chapter/identifying-the-evidence-literature-searching-and-evidence-submission.

For more information, please see the Methodology review published as part of the accompanying documents for this guideline.

Clinical literature search strategies

1.1 Ultrasound

This literature search strategy was used for the following review:

- What is the diagnostic accuracy of ultrasound for identifying thyroid nodule malignancies or nodules with malignant potential?

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 14: Database parameters, filters and limits applied

Database	Dates searched	Search filters and limits applied
Medline (OVID)	1946 – 10 December 2021	Diagnostic studies Exclusions (animal studies, letters, comments, editorials, case studies/reports, children) English language
Embase (OVID)	1974 – 10 December 2021	Diagnostic studies Exclusions (animal studies, letters, comments, editorials, case studies/reports, conference abstracts, children) English language
The Cochrane Library (Wiley)	Cochrane Database of Systematic Reviews to Issue 12 of 12, December 2021	Exclusions (clinical trials, conference abstracts) Review type: Diagnostic

Medline (Ovid) search terms

1.	exp Thyroid Neoplasms/
2.	(thyroid adj3 (cancer* or carcinom* or microcarcinoma* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or node* or nodul* or nodal or lump* or papillar* or swollen or swell* or aplastic or sarcoma* or cyst* or malignan*)).ti,ab.

3.	DTC.ti,ab.
4.	((papillar* or anaplastic) adj2 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nodul* or node* or lump* or lymphoma*)).ti,ab.
5.	or/1-4
6.	letter/
7.	editorial/
8.	news/
9.	exp historical article/
10.	Anecdotes as Topic/
11.	comment/
12.	case report/
13.	(letter or comment*).ti.
14.	or/6-13
15.	randomized controlled trial/ or random*.ti,ab.
16.	14 not 15
17.	animals/ not humans/
18.	exp Animals, Laboratory/
19.	exp Animal Experimentation/
20.	exp Models, Animal/
21.	exp Rodentia/
22.	(rat or rats or mouse or mice or rodent*).ti.
23.	or/16-22
24.	5 not 23
25.	limit 24 to english language
26.	(exp child/ or exp pediatrics/ or exp infant/) not (exp adolescent/ or exp adult/ or exp middle age/ or exp aged/)
27.	25 not 26
28.	Ultrasonography/
29.	Elasticity Imaging Techniques/
30.	Endosonography/
31.	Microscopy, Acoustic/
32.	exp Ultrasonography, Doppler/
33.	exp Ultrasonography, Interventional/
34.	(ultrasonograph* or ultrasound* or ultra sound* or sonograph* or sonogram* or echograph* or echotomograph* or elastography* or elastosonograph* or sonoelastograph* or doppler or endosonograph* or acoustic microscop* or elasticity imag*).ti,ab.
35.	or/28-34
36.	27 and 35
37.	exp "sensitivity and specificity"/
38.	(sensitivity or specificity).ti,ab.
39.	((pre test or pretest or post test) adj probability).ti,ab.
40.	(predictive value* or PPV or NPV).ti,ab.
41.	likelihood ratio*.ti,ab.
42.	likelihood function/
43.	((area under adj4 curve) or AUC).ti,ab.
44.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.

45.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness or precision or validat* or validity)).ti,ab.
46.	gold standard.ab.
47.	exp Diagnostic errors/
48.	(false positiv* or false negativ*).ti,ab.
49.	or/37-48
50.	36 and 49

1 **Embase (Ovid) search terms**

1.	exp Thyroid Cancer/
2.	(thyroid and (cancer* or carcinom* or microcarcinoma* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or node* or nodul* or nodal or lump* or papillar* or swollen or swell* or anaplastic or sarcoma* or cyst* or malignan*)).ti,ab.
3.	DTC.ti,ab.
4.	((papillar* or anaplastic) adj2 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nodul* or node* or lump*)).ti,ab.
5.	or/1-4
6.	letter.pt. or letter/
7.	note.pt.
8.	editorial.pt.
9.	case report/ or case study/
10.	(letter or comment*).ti.
11.	(conference abstract or conference paper).pt.
12.	or/6-11
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 or rodent*).ti.
22.	or/14-21
23.	5 not 22
24.	limit 23 to english language
25.	(exp child/ or exp pediatrics/) not (exp adult/ or exp adolescent/)
26.	24 not 25
27.	*Echograph/
28.	*Elastograph/
29.	*Echography/
30.	*Elastography/
31.	*Endoscopic ultrasonography/
32.	*Microscopy, Acoustic/
33.	exp *Doppler Ultrasonography/
34.	*Interventional Ultrasonography/

35.	*Contrast-enhanced Ultrasound/ or exp *High Frequency Ultrasound/ or *Radiofrequency Echographic Multi Spectrometry/ or *Real Time Echography/ or exp *Three Dimensional Echography/
36.	(ultrasonograph* or ultrasound* or ultra sound* or sonograph* or sonogram* or echograph* or echotomograph* or elastography* or elastosonograph* or sonoelastograph* or doppler or endosonograph* or acoustic microscop* or elasticity imag*).ti,ab.
37.	or/27-36
38.	26 and 37
39.	exp "sensitivity and specificity"/
40.	(sensitivity or specificity).ti,ab.
41.	((pre test or pretest or post test) adj probability).ti,ab.
42.	(predictive value* or PPV or NPV).ti,ab.
43.	likelihood ratio*.ti,ab.
44.	((area under adj4 curve) or AUC).ti,ab.
45.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
46.	diagnostic accuracy/
47.	diagnostic test accuracy study/
48.	gold standard.ab.
49.	exp diagnostic error/
50.	(false positiv* or false negativ*).ti,ab.
51.	differential diagnosis/
52.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness or precision or validat* or validity or differential or error*)).ti,ab.
53.	or/39-52
54.	38 and 53

Cochrane Library (Wiley) search terms

#1.	MeSH descriptor: [Thyroid Neoplasms] explode all trees
#2.	(thyroid near/3 (cancer* or carcinom* or microcarcinoma* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or node* or nodul* or nodal or lump* or papillar* or swollen or swell* or anaplastic or sarcoma* or cyst* or malignan*)):ti,ab
#3.	DTC:ti,ab
#4.	((papillar* or anaplastic) near/2 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nodul* or node* or lump*)):ti,ab
#5.	#1 or #2 or #3 or #4

1.4 Ultrasound Accuracy

This literature search strategy was used for the following review:

- In people with thyroid nodules on ultrasound at initial presentation, for what size and classification is it clinically and cost effective to use active surveillance or discharge rather than biopsy?

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.

1

2 **Table 15: Database parameters, filters and limits applied**

Database	Dates searched	Search filters and limits applied
Medline (OVID)	1946 – 13 January 2022	Observational studies Exclusions (animal studies, letters, comments, editorials, case studies/reports, children) English language
Embase (OVID)	1974 – 13 January 2022	Observational studies Exclusions (animal studies, letters, comments, editorials, case studies/reports, conference abstracts, children) English language

3 **Medline (Ovid) search terms**

1.	(tumo?r* or node* or nodul* or nodal or lump* or swollen or swell* or sarcoma* or cyst* or classif* or size or grad* or circumference* or diameter* or shape* or U1 or U2 or U3 or U4 or U5 or EU Tirads).ti,ab.
2.	Thyroid Gland/
3.	thyroid.ti,ab.
4.	1 and (2 or 3)
5.	exp Thyroid Neoplasms/
6.	(thyroid adj3 (cancer* or carcinom* or microcarcinoma* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or node* or nodul* or nodal or lump* or papillar* or swollen or swell* or aplastic or sarcoma* or cyst* or malignan*)).ti,ab.
7.	DTC.ti,ab.
8.	((papillar* or anaplastic) adj2 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nodul* or node* or lump* or lymphoma*)).ti,ab.
9.	or/4-8
10.	letter/
11.	editorial/
12.	news/
13.	exp historical article/
14.	Anecdotes as Topic/
15.	comment/
16.	case report/
17.	(letter or comment*).ti.
18.	or/10-17
19.	randomized controlled trial/ or random*.ti,ab.
20.	18 not 19
21.	animals/ not humans/
22.	exp Animals, Laboratory/

23.	exp Animal Experimentation/
24.	exp Models, Animal/
25.	exp Rodentia/
26.	(rat or rats or mouse or mice or rodent*).ti.
27.	or/20-26
28.	9 not 27
29.	limit 28 to english language
30.	(exp child/ or exp pediatrics/ or exp infant/) not (exp adolescent/ or exp adult/ or exp middle age/ or exp aged/)
31.	29 not 30
32.	Ultrasonography/
33.	Elasticity Imaging Techniques/
34.	Endosonography/
35.	Microscopy, Acoustic/
36.	exp Ultrasonography, Doppler/
37.	exp Ultrasonography, Interventional/
38.	(ultrasonograph* or ultrasound* or ultra sound* or sonograph* or sonogram* or echograph* or echotomograph* or elastography* or elastosonograph* or sonoelastograph* or doppler or endosonograph* or acoustic microscop* or elasticity imag*).ti,ab.
39.	or/32-38
40.	31 and 39
41.	Epidemiologic studies/
42.	Observational study/
43.	exp Cohort studies/
44.	(cohort adj (study or studies or analys* or data)).ti,ab.
45.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
46.	((longitudinal or retrospective or prospective) and (study or studies or review or analys* or cohort* or data)).ti,ab.
47.	Controlled Before-After Studies/
48.	Historically Controlled Study/
49.	Interrupted Time Series Analysis/
50.	(before adj2 after adj2 (study or studies or data)).ti,ab.
51.	exp case control study/
52.	case control*.ti,ab.
53.	Cross-sectional studies/
54.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
55.	or/41-54
56.	40 and 55

Embase (Ovid) search terms

1.	(tumo?r* or node* or nodul* or nodal or lump* or swollen or swell* or sarcoma* or cyst* or classif* or size or grad* or circumference* or diameter* or shape* or U1 or U2 or U3 or U4 or U5 or EU Tirads).ti,ab.

2.	Thyroid Gland/
3.	thyroid.ti,ab.
4.	1 and (2 or 3)
5.	exp Thyroid Cancer/
6.	(thyroid adj3 (cancer* or carcinom* or microcarcinoma* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or node* or nodul* or nodal or lump* or papillar* or swollen or swell* or anaplastic or sarcoma* or cyst* or malignan*)).ti,ab.
7.	DTC.ti,ab.
8.	((papillar* or anaplastic) adj2 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nodul* or node* or lump*)).ti,ab.
9.	or/4-8
10.	letter.pt. or letter/
11.	note.pt.
12.	editorial.pt.
13.	case report/ or case study/
14.	(letter or comment*).ti.
15.	(conference abstract or conference paper).pt.
16.	or/10-15
17.	randomized controlled trial/ or random*.ti,ab.
18.	16 not 17
19.	animal/ not human/
20.	nonhuman/
21.	exp Animal Experiment/
22.	exp Experimental Animal/
23.	animal model/
24.	exp Rodent/
25.	(rat or rats or mouse or mice or rodent*).ti.
26.	or/18-25
27.	9 not 26
28.	limit 27 to english language
29.	(exp child/ or exp pediatrics/) not (exp adult/ or exp adolescent/)
30.	28 not 29
31.	*Echograph/
32.	*Elastograph/
33.	*Echography/
34.	*Elastography/
35.	*Endoscopic ultrasonography/
36.	*Microscopy, Acoustic/
37.	exp *Doppler Ultrasonography/
38.	*Interventional Ultrasonography/
39.	*Contrast-enhanced Ultrasound/ or exp *High Frequency Ultrasound/ or *Radiofrequency Echographic Multi Spectrometry/ or *Real Time Echography/ or exp *Three Dimensional Echography/
40.	(ultrasonograph* or ultrasound* or ultra sound* or sonograph* or sonogram* or echograph* or echotomograph* or elastography* or elastosonograph* or sonoelastograph* or doppler or endosonograph* or acoustic microscop* or elasticity imag*).ti,ab.
41.	or/31-40

42.	30 and 41
43.	Clinical study/
44.	Observational study/
45.	family study/
46.	longitudinal study/
47.	retrospective study/
48.	prospective study/
49.	cohort analysis/
50.	follow-up/
51.	cohort*.ti,ab.
52.	50 and 51
53.	(cohort adj (study or studies or analys* or data)).ti,ab.
54.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
55.	((longitudinal or retrospective or prospective) and (study or studies or review or analys* or cohort* or data)).ti,ab.
56.	(before adj2 after adj2 (study or studies or data)).ti,ab.
57.	exp case control study/
58.	case control*.ti,ab.
59.	cross-sectional study/
60.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
61.	or/43-49,52-60
62.	42 and 61

Health Economics literature search strategy

Health economic evidence was identified by conducting searches using terms for a broad Thyroid Cancer population. The following databases were searched: NHS Economic Evaluation Database (NHS EED - this ceased to be updated after 31st March 2015), Health Technology Assessment database (HTA - this ceased to be updated from 31st March 2018) and The International Network of Agencies for Health Technology Assessment (INAHTA). Searches for recent evidence were run on Medline and Embase from 2014 onwards for health economics, and all years for quality-of-life studies.

Table 2: Database parameters, filters and limits applied

Database	Dates searched	Search filters and limits applied
Medline (OVID)	Health Economics 1 January 2014 – 16 December 2021	Health economics studies Quality of life studies
	Quality of Life 1946 – 16 December 2021	Exclusions (animal studies, letters, comments, editorials, case studies/reports, conference abstracts) English language
Embase (OVID)	Health Economics 1 January 2014 – 16 December 2021	Health economics studies Quality of life studies

Database	Dates searched	Search filters and limits applied
	Quality of Life 1974 – 16 December 2021	Exclusions (animal studies, letters, comments, editorials, case studies/reports, conference abstracts) English language
NHS Economic Evaluation Database (NHS EED) (Centre for Research and Dissemination - CRD)	Inception –31st March 2015	
Health Technology Assessment Database (HTA) (Centre for Research and Dissemination – CRD)	Inception – 31st March 2018	
The International Network of Agencies for Health Technology Assessment (INAHTA)	Inception - 16 December 2021	English language

1 **Medline (Ovid) search terms**

1.	exp Thyroid Neoplasms/
2.	(thyroid adj4 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or papillar* or follicul* or lymphoma* or anaplastic)).ti,ab.
3.	((papillar* or follicul* or medullary or anaplastic) adj4 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or lymphoma*)).ti,ab.
4.	or/1-3
5.	letter/
6.	editorial/
7.	news/
8.	exp historical article/
9.	Anecdotes as Topic/
10.	comment/
11.	case report/
12.	(letter or comment*).ti.
13.	or/5-12
14.	randomized controlled trial/ or random*.ti,ab.
15.	13 not 14
16.	animals/ not humans/
17.	exp Animals, Laboratory/
18.	exp Animal Experimentation/
19.	exp Models, Animal/
20.	exp Rodentia/
21.	(rat or rats or mouse or mice).ti.
22.	or/15-21
23.	4 not 22

24.	limit 23 to english language
25.	economics/
26.	value of life/
27.	exp "costs and cost analysis"/
28.	exp Economics, Hospital/
29.	exp Economics, medical/
30.	Economics, nursing/
31.	economics, pharmaceutical/
32.	exp "Fees and Charges"/
33.	exp budgets/
34.	budget*.ti,ab.
35.	cost*.ti.
36.	(economic* or pharmaco?economic*).ti.
37.	(price* or pricing*).ti,ab.
38.	(cost* adj2 (effectiv* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
39.	(financ* or fee or fees).ti,ab.
40.	(value adj2 (money or monetary)).ti,ab.
41.	or/25-40
42.	24 and 41
43.	quality-adjusted life years/
44.	sickness impact profile/
45.	(quality adj2 (wellbeing or well being)).ti,ab.
46.	sickness impact profile.ti,ab.
47.	disability adjusted life.ti,ab.
48.	(qal* or qtime* or qwb* or daly*).ti,ab.
49.	(euroqol* or eq5d* or eq 5*).ti,ab.
50.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
51.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
52.	(hui or hui1 or hui2 or hui3).ti,ab.
53.	(health* year* equivalent* or hye or hyes).ti,ab.
54.	discrete choice*.ti,ab.
55.	rosser.ti,ab.
56.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.
57.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.
58.	(sf20 or sf 20 or short form 20 or shortform 20 or shortform20).ti,ab.
59.	(sf12* or sf 12* or short form 12* or shortform 12* or shortform12*).ti,ab.
60.	(sf8* or sf 8* or short form 8* or shortform 8* or shortform8*).ti,ab.
61.	(sf6* or sf 6* or short form 6* or shortform 6* or shortform6*).ti,ab.
62.	or/52-70
63.	24 and 62

1 **Embase (Ovid) search terms**

1.	exp Thyroid Cancer/
2.	(thyroid adj4 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or papillar* or follicul* or lymphoma* or anaplastic)).ti,ab.

3.	((papillar* or follicul* or medullary or anaplastic) adj4 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or lymphoma*)).ti,ab.
4.	or/1-3
5.	letter.pt. or letter/
6.	note.pt.
7.	editorial.pt.
8.	case report/ or case study/
9.	(letter or comment*).ti.
10.	or/5-9
11.	randomized controlled trial/ or random*.ti,ab.
12.	10 not 11
13.	animal/ not human/
14.	nonhuman/
15.	exp Animal Experiment/
16.	exp Experimental Animal/
17.	animal model/
18.	exp Rodent/
19.	(rat or rats or mouse or mice).ti.
20.	or/12-19
21.	4 not 20
22.	limit 21 to english language
23.	health economics/
24.	exp economic evaluation/
25.	exp health care cost/
26.	exp fee/
27.	budget/
28.	funding/
29.	budget*.ti,ab.
30.	cost*.ti.
31.	(economic* or pharmaco?economic*).ti.
32.	(price* or pricing*).ti,ab.
33.	(cost* adj2 (effectiv* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
34.	(financ* or fee or fees).ti,ab.
35.	(value adj2 (money or monetary)).ti,ab.
36.	or/23-35
37.	22 and 36
38.	quality-adjusted life years/
39.	"quality of life index"/
40.	short form 12/ or short form 20/ or short form 36/ or short form 8/
41.	sickness impact profile/
42.	(quality adj2 (wellbeing or well being)).ti,ab.
43.	sickness impact profile.ti,ab.
44.	disability adjusted life.ti,ab.
45.	(qal* or qtime* or qwb* or daly*).ti,ab.
46.	(euroqol* or eq5d* or eq 5*).ti,ab.

47.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
48.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
49.	(hui or hui1 or hui2 or hui3).ti,ab.
50.	(health* year* equivalent* or hye or hyes).ti,ab.
51.	discrete choice*.ti,ab.
52.	rosser.ti,ab.
53.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.
54.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.
55.	(sf20 or sf 20 or short form 20 or shortform 20 or shortform20).ti,ab.
56.	(sf12* or sf 12* or short form 12* or shortform 12* or shortform12*).ti,ab.
57.	(sf8* or sf 8* or short form 8* or shortform 8* or shortform8*).ti,ab.
58.	(sf6* or sf 6* or short form 6* or shortform 6* or shortform6*).ti,ab.
59.	or/37-58
60.	22 and 59

1 **NHS EED and HTA (CRD) search terms**

#1.	MeSH DESCRIPTOR Thyroid Neoplasms EXPLODE ALL TREES
#2.	((thyroid NEAR4 (cancer* or carcinom* or tumour* or tumor* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or papillar* or follicul* or lymphoma* or anaplastic)))
#3.	(((papillar* or follicul* or medullary or anaplastic) NEAR4 (cancer* or carcinom* or tumour* or tumor* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or lymphoma*)))
#4.	#1 OR #2 OR #3

2 **INHATA search terms**

1.	(Thyroid Neoplasms)[mh] OR (thyroid neoplasms) AND (thyroid cancers)

3

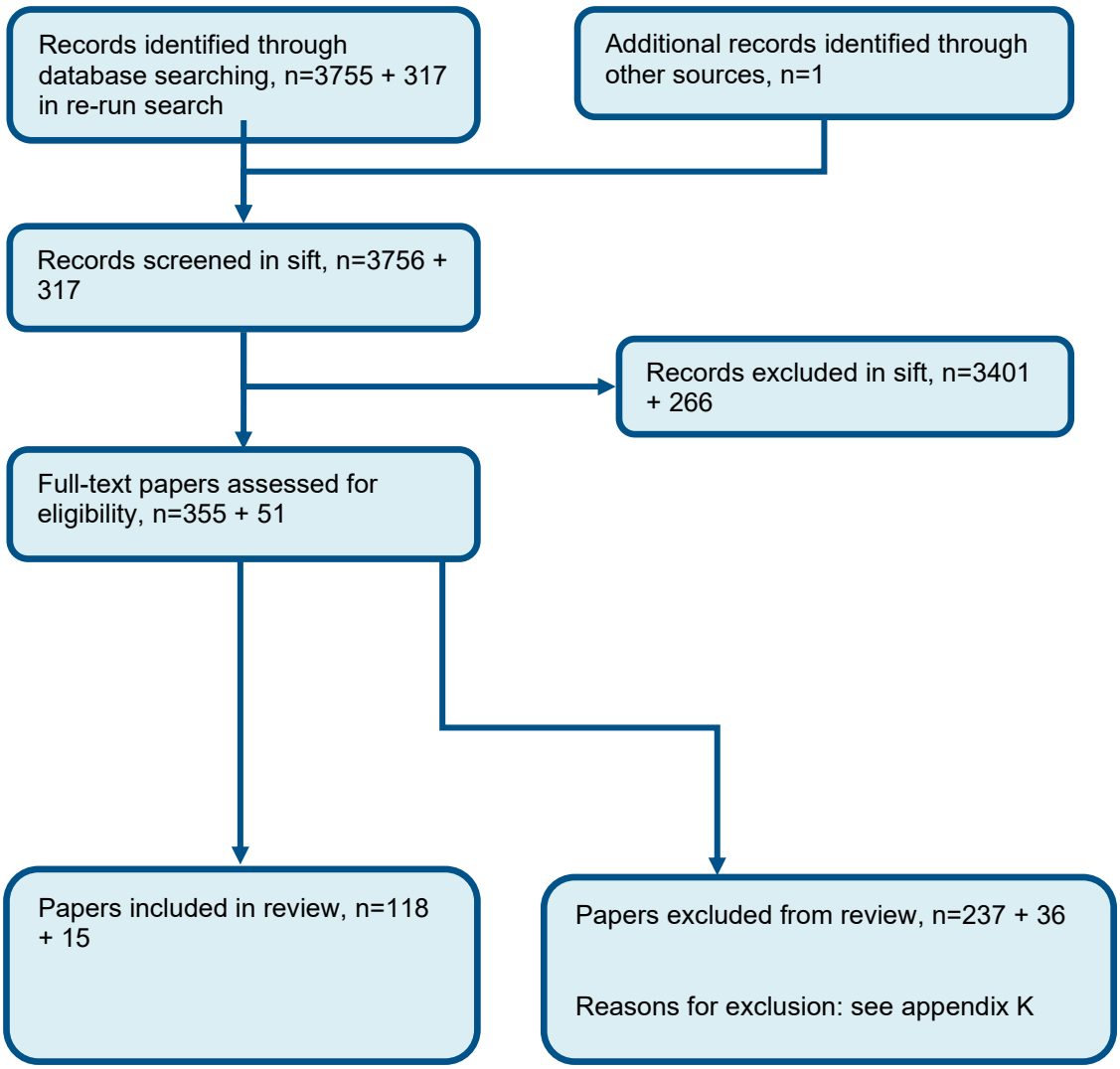
4

5

Appendix C Evidence study selection

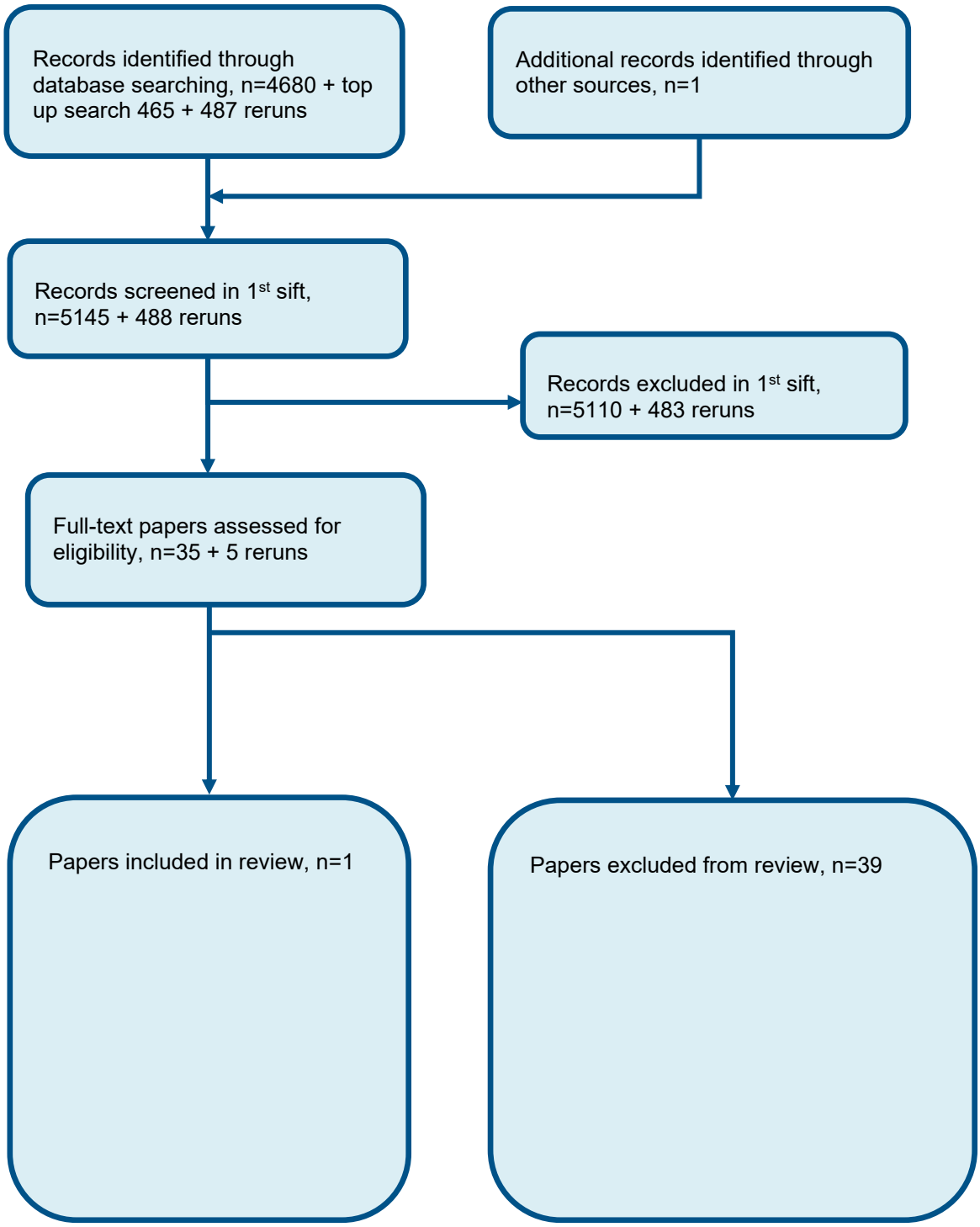
C.1 Diagnostic accuracy of Ultrasound

Figure 1: Flow chart of clinical study selection for the review of diagnostic accuracy of ultrasound



C.2 Threshold of size and classification of thyroid nodules

Figure 2: Flow chart of clinical study selection for the review of size and US characteristic thresholds for efficacy of active surveillance



Appendix D Effectiveness evidence

D.1 Diagnostic accuracy of evidence

Reference	**Trimboli, 2019[380]**
Study type	Retrospective
Number of patients	n = 495
Patient characteristics	*Age, mean (SD):* Switzerland 51.4(1.3), France 51.5(15.3), UK 53.6(18.8) *Gender (female to male ratio):* 381:114 *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Thyroid centre, oncology, Institute of Southern Switzerland; institute of endocrinology, Pitie Salpetriere hospital, Paris; Guys and St Thomas's head and neck thyroid imaging *Country*: Switzerland, France, UK *Inclusion criteria*: Adult patients on pathology database who had undergone thyroidectomy for all causes and who had US images on the clinical information management system *Exclusion criteria*: lesions <5mm; no US images
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using EU-TIRADS. 3 different machines were used at the different centres , with 18 MHz transducer in Switzerland, 10-18 MHz transducer in Paris and 12-15MHz transducer in UK. *Reference (gold) standard:*

Reference	**Trimboli, 2019[380]**
	Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	**257/1058 malignant on histopathology** **All nodules (1058 nodules from 495 patients; therefore we have some risk of unit of analysis errors)** *Index test 1: +ve= EU TIRADS >2* TP: 256 FN: 1 FP: 731 TN: 70; *sensitivity*: 0.996(0.979-0.999), *specificity*: 0.087(0.069-0.109) *Index test 2*: *+ve= EU TIRADS >3* TP: 239 FN: 18 FP: 257 TN: 544; *sensitivity*: 0.93(0.891-0.958), *specificity*: 0.679(0.645-0.711) *Index test 3*: *+ve= EU TIRADS >4* TP: 192 FN: 65 FP: 27 TN: 774; *sensitivity*: 0.747(0.689-0.799), *specificity*: 0.966(0.951-0.977)
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Aggarwal, 1989[7]**
Study type	prospective
Number of patients	n = 36
Patient characteristics	*Age, mean (SD):* not reported *Gender (female to male ratio):* not reported *Ethnicity*: not reported

Reference	**Aggarwal, 1989[7]**
	Expertise of US tester (medic/non medic/unknown): unknown *Setting*: Teaching Hospital departments of Surgery and Pathology, New Delhi. *Country*: India *Inclusion criteria*: patients with ultrasonographically solitary and scintigraphically cold thyroid nodules *Exclusion criteria*: None reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound (7.5MHz), using 1) echo texture and 2) degeneration *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*16/36 malignant* *Index test 1: US using echo texture (solid /mixed) with +ve test = solid texture* TP: unclear FN: unclear FP: unclear TN: unclear ; *sensitivity*: 0.545, *specificity*: 0.643 The text carried some details of the raw data, but these did not tally with the sensitivities and specificities given above. From the data in the text, the raw data would be: TP: 7, FN: 9, FP 9, TN 11. *Index test 2*: US using signs of degeneration in a 'mixed' echo nodule as the definition of +ve test TP: unclear FN: unclear FP: unclear TN: unclear ; *sensitivity*: 0.60, *specificity*: 1.00 The text carried some details of the raw data, but these did not tally with the sensitivities and specificities given above. From the data in the text, the raw data would be: TP: 5, FN: 11, FP 1, TN 19.
Source of funding	Not reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: None

Reference	**Aggarwal, 1989[7]**
Comments	

1

Reference	**Shao, 2015[344]**
Study type	prospective
Number of patients	n = 297 patients with 512 thyroid nodules
Patient characteristics	*Age, mean (SD):* 42.15(11.35) *Gender (female to male ratio):* 231:66 *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown for conventional US, but for elastography was medic (endocrinologist) *Setting*: Secondary care *Country*: China *Inclusion criteria*: People referred for thyroidectomy with US performed before surgery *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* 1. Grey scale ultrasound (6-15MHz transducer), using a 7-point scoring system based on the number of features observed: low echogenicity, solid rather than cystic, irregular margins, aspect ratio taller than wide, vascularity (on colour doppler) and calcifications 2. Ultrasound elastography. Elasticity scores were scored as follows: 0=lesions with a cystic component showing red and blue or blue-green and red; 1=lesions and surrounding tissue in an even green colour; 2=lesions in green and the surrounding area in blue; 3=lesions in blue and green; 4=lesions completely covered by blue *Reference (gold) standard:* Surgical histopathological findings

Reference	**Shao, 2015[344]**
	Time between measurement of index test and reference standard: Not clear
Results	Malignant nodules: 203/512 *CONVENTIONAL ULTRASOUND* *Index test 1: Conventional US: existence of solitary nodules (+ve = present)* TP: 174 FN: 29 FP: 111 TN: 198; *sensitivity*: 0.857 , *specificity*: 0.641 *Index test 2*: Conventional US: existence of irregular margins (+ve = present) TP: 90 FN: 113 FP: 20 TN: 289 ; *sensitivity*: 0.443 , *specificity*: 0.935 *Index test 3*: Conventional US: existence of hypo-echogenicity (+ve = present) TP: 135 FN: 68 FP: 78 TN: 231; *sensitivity*: 0.665 , *specificity*: 0.748 *Index test 4:* Conventional US: existence of microcalcifications (+ve = present) TP: 140 FN: 63 FP: 17 TN: 292; *sensitivity*: 0.690 , *specificity*: 0.945 *Index test 5:* Conventional US: existence of aspect ratio $\geq$1 (+ve = present) TP: 120 FN: 83 FP: 7 TN: 302; *sensitivity*: 0.591 , *specificity*: 0.977 *Index test 6:* Conventional US: intranodular blood flow (+ve = present) TP: 97 FN: 106 FP: 94 TN: 215; *sensitivity*: 0.478, *specificity*: 0.696 *Index test 7:* Conventional US: +ve = score of 1 or more (1 or more features occurring together); TP: 201 FN: 2 FP: 186 TN: 123; *sensitivity*: 0.990, *specificity*: 0.398 *Index test 8:* Conventional US: +ve = score of 2 or more (2 or more features occurring together); TP: 185 FN: 18 FP: 95 TN: 214; *sensitivity*: 0.911, *specificity*: 0.693 *Index test 9:* Conventional US: +ve = score of 3 or more (3 or more features occurring together); TP: 155 FN: 48 FP: 34 TN: 275; *sensitivity*: 0.764, *specificity*: 0.890 *Index test 10:* Conventional US: +ve = score of 4 or more (4 or more features occurring together); TP: 115 FN: 88 FP: 8 TN: 301; *sensitivity*: 0.567, *specificity*: 0.974 *Index test 11:* Conventional US: +ve = score of 5 or more (5 or more features occurring together); TP: 75 FN: 228 FP: 3 TN: 306; *sensitivity*: 0.369, *specificity*: 0.990 *ELASTOGRAPHY* *Index test 12:* Elastography: +ve = score of 1 or more on Shuzen method; TP: 203 FN: 0 FP: 263 TN: 46; *sensitivity*: 1.0 , *specificity*: 0.149 *Index test 13:* Elastography: +ve = score of 2 or more on Shuzen method; TP: 200 FN: 3 FP: 178 TN: 131; *sensitivity*: 0.985 , *specificity*: 0.424 *Index test 14:* Elastography: +ve = score of 3 or more on Shuzen method; TP: 169 FN: 34 FP: 28 TN: 281; *sensitivity* on Shuzen method: 0.832 , *specificity*: 0.909

Reference	**Shao, 2015[344]**
	Index test 15: Elastography: +ve = score of 4 on Shuzen method; TP: 32 FN: 271 FP: 0 TN: 309; sensitivity: 0.158 , specificity: 1.00 COMBINATION OF ULTRASOUND AND ELASTICITY SCORES (adding scores from conventional US and elastography) Index test 16: score of 1 or more TP: 203 FN: 0 FP: 275 TN: 34; sensitivity: 1.00 , specificity: 0.11 Index test 17: score of 2 or more TP: 202 FN: 1 FP: 230 TN: 79; sensitivity: 0.99, specificity: 0.256 Index test 18: score of 3 or more TP: 201 FN: 2 FP: 155 TN: 154; sensitivity: 0.990 , specificity: 0.498 Index test 19: score of 4 or more TP: 193 FN: 10 FP: 77 TN: 232; sensitivity: 0.951 , specificity: 0.751 Index test 20: score of 5 or more TP: 173 FN: 30 FP: 39 TN 270; sensitivity: 0.852 , specificity: 0.874 Index test 21: score of 6 or more TP: 147 FN: 56 FP: 14 TN: 295; sensitivity: 0.7241 , specificity: 0.9547 Index test 22: score of 7 or more TP: 113 FN: 90 FP: 3 TN: 306; sensitivity: 0.557 , specificity: 0.990
Source of funding	Academic and government (non-commercial)
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

Reference	**Shi, 2020[349]**
Study type	retrospective
Number of patients	n = 338
Patient characteristics	*Age, mean (SD): 54(15.5)[malignant]; 50(16.7)[benign]* *Gender (female to male ratio): 216:91* *Ethnicity*: not reported

Reference	**Shi, 2020[349]**
	Expertise of US tester (medic/non medic/unknown): tester unknown, but reviewed by medic *Setting*: Shanghai Ninth people's Hospital *Country*: China *Inclusion criteria*: Nodules with both solid and cystic components; images complete and information on gender, age and histopathology available; no treatment history *Exclusion criteria*: accepted thyroid hormone therapy; incomplete US images, clinical information or histopathology.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* • Grey scale ultrasound, using 5-12 MHz linear array transducer, looking for the following US findings of entire nodule: composition [solid >50% vs solid <50% vs spongy]; shape [ovoid to round vs taller than wide vs irregular]; margin [smooth vs spiculated or micro-lobulated vs ill-defined]; nodule vascularity [peripheral vs intramodular vs avascular]. For the US findings of the internal solid portion: configuration [eccentric vs no-eccentric]; rim [smooth vs non-smooth]; echogenicity [markedly hypoechoic vs hypoechoic vs isoechoic vs hyperechoic]; calcification [microcalcifications vs macrocalcifications vs mixed calcifications vs none]. • Colour doppler US using 700Hz pulse repetition frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Shi, 2020[349]**
Results	*Malignant nodules: 50, benign nodules: 288. Definitions of positive index test given below. A negative test would be anything that is mutually exclusive (i.e. anything else)* *Index test 1: nodules with taller than wide shape (+ve test)* TP: 12 FN: 38 FP: 0 TN: 288; *sensitivity*:0.24 , *specificity*: 1.00 *Index test 2*: nodules that are spiculated or microlobulated *(+ve test)* TP: 16 FN: 34 FP: 9 TN: 279; *sensitivity*: 0.32, *specificity*: 0.969 *Index test 3*: nodules with an eccentric configuration *(+ve test)* TP: 43 FN: 7 FP: 28 TN: 260; *sensitivity*: 0.86 , *specificity*: 0.903 *Index test 4*: nodules with a non-smooth rim *(+ve test)* TP: 37 FN: 13 FP: 12 TN: 276; *sensitivity*: 0.74 , *specificity*: 0.958 *Index test 5*: nodules with hypoechogenicity *(+ve test)* TP: 31 FN: 19 FP: 79 TN: 209; *sensitivity*: 0.62 , *specificity*: 0.726 *Index test 6*: nodules with microcalcification *(+ve test)* TP: 33 FN: 17 FP: 3 TN: 285; *sensitivity*: 0.66 , *specificity*: 0.99
Source of funding	Not stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus retrospective studies are downgraded for indirectness.
Comments	

1
2

Reference	**Magri, 2020[260]**
Study type	retrospective
Number of patients	n = 255 patients with 304 nodules
Patient characteristics	*Age, median (range): 56 (20-86)* *Gender (female to male ratio): 205:50*

Reference	Magri, 2020[260]
	Ethnicity: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: Outpatients, unit of endocrinology, secondary care *Country*: Italy *Inclusion criteria*: age >18 years; detailed pre-op thyroid US exam; availability of data concerning thyroid autoimmunity; availability of US-guided FNA; total or partial thyroidectomy performed; availability of histological surgical diagnosis; *Exclusion criteria*: Any of the aforementioned procedures done in another clinical setting;
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 7.5MHz linear transducer. The following features were examined: size, composition, echogenicity, shape, margins, echogenicity foci. US data classified according to the ACR and EU-TIRADS scoring systems *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Magri, 2020[260]**
Results	*Malignant 95 modules; benign 209 nodules* *Index test 1: ACR ≥2* TP: 95 FN: 0 FP: 208 TN: 1; *sensitivity*:1.0 , *specificity*: 0.005 *Index test 2*: ACR ≥3 TP: 92 FN: 3 FP: 133 TN: 76 ; *sensitivity*: 0.968, *specificity*: 0.364 *Index test 3*: ACR ≥4 TP: 74 FN: 21 FP: 46 TN: 163; *sensitivity*: 0.779, *specificity*: 0.780 *Index test 3*: ACR ≥5 TP: 29 FN: 66 FP: 3 TN: 206; *sensitivity*: 0.305, *specificity*: 0.986 *Index test 1: EU TIRADS ≥2* TP: 95 FN: 0 FP: 209 TN: 1; *sensitivity*:1.0 , *specificity*: 0.00 *Index test 2*: EU TIRADS ≥3 TP: 95 FN: 0 FP: 207 TN: 2; *sensitivity*: 1.000, *specificity*: 0.01 *Index test 3*: EU TIRADS ≥4 TP: 79 FN: 16 FP: 54 TN: 155; *sensitivity*: 0.832, *specificity*: 0.742 *Index test 3*: EU TIRADS ≥5 TP: 51 FN: 44 FP: 15 TN: 194; *sensitivity*: 0.537, *specificity*: 0.928
Source of funding	No specific funding
Limitations	*Risk of bias (QUADAS 2 – risk of bias): Very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus retrospective studies are downgraded for indirectness.*
Comments	

Reference	**Veyrieres, 2012 [383]**
Study type	prospective

Reference	**Veyrieres, 2012** [383]
Number of patients	n = 148 patients with 297 nodules
Patient characteristics	*Age, mean (SD): 52.5 (15.8)* *Gender (female to male ratio): 110:38* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Surgical endocrinology department *Country*: France *Inclusion criteria*: Patients referred for surgical care for dysthyroidism, local compression, suspicious nodule under US, prevalent adenopathy, positive or suspicious and unknown FNA *Exclusion criteria*: None stated
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale (B-scan) ultrasound with colour doppler, using 15-4 MHz broadband linear transducer. All nodules >5mm were studied. The aspects studied were hypo-echogenicity, central vascularisation, irregular margins, micro/macro calcifications and nodule size. Also SWE. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* 1 day

Reference	**Veyrieres, 2012** [383]
Results	*35 malignant nodules/297* *Index test 1: central vascularisation (+ve)* TP: 20 FN: 15 FP: 82 TN: 180; *sensitivity*: 0.57(0.39,0.73), *specificity*: 0.69 (0.62,0.73) *Index test 2*: microcalcifications (+ve) TP: 15 FN: 20 FP: 43 TN: 219 ; *sensitivity*: 0.42(0.26,0.60), *specificity*: 0.83(0.77,0.87) *Index test 3*: macrocalcifications (+ve) TP: 10 FN: 25 FP: 27 TN: 235; *sensitivity*: 0.28(0.15,0.46), *specificity*: 0.89(0.84,0.91) *Index test 4: halo (+ve)* TP: 3 FN: 32 FP: 41 TN: 221; *sensitivity*: 0.16(0.12,0.212), *specificity*: 0.914 (0.758,0.977) *Index test 5*: hypoechogenicity (+ve) TP: 25 FN: 10 FP: 120 TN: 142 ; *sensitivity*: 0.714(0.534,0.847), *specificity*: 0.54(0.479,0.602) *Index test 6*: irregular margins (+ve) TP: 8 FN: 27 FP: 32 TN: 230; *sensitivity*: 0.228(0.11,0.405), *specificity*: 0.872(0.824,0.908) *Index test 7*: SWE at threshold of 66kpa and above (+ve) TP: 28 FN: 7 FP: 25 TN: 237; *sensitivity*: 0.80, *specificity*: 0.905 *Index test 8*: SWE at threshold of 66kpa and above OR microcalcification OR central vascularisation (+ve) TP: 34 FN: 1 FP: 117 TN: 145 *sensitivity*: 0.553, *specificity*: 0.905
Source of funding	None stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

Reference	**Kobayashi, 2005**[195]
Study type	Retrospective
Number of patients	n = 910

Reference	**Kobayashi, 2005[195]**
Patient characteristics	*Age, mean (SD): 47.3(18.1) for malignant nodule patients; 49.7 (13.9) for benign nodule patients* *Gender (female to male ratio): 803: 117* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): Unknown ('well-trained') *Setting*: Unclear *Country*: Japan *Inclusion criteria*: Patients with follicular adenomas, adenomatous thyroid nodules and follicular carcinomas *Exclusion criteria*: Patients with papillary carcinomas, undifferentiated carcinomas, medullary carcinomas and malignant lymphomas
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using wither and 10MHz or 13 MHz mechanical sector probe. Size, border, echo pattern, solidity and echoicity were measured. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Kobayashi, 2005[195]**
Results	*Index test 1: size ≤40mm (+ve)* TP: 65 FN: 44 FP: 562 TN: 249 ; *sensitivity*: 0.596 , *specificity*: 0.307 *Index test 2*: Solitary nodule (+ve) TP: 70 FN: 39 FP: 519 TN:292 ; *sensitivity*: 0.642, *specificity*: 0.376 *Index test 3*: Solid pattern (+ve) TP: 87 FN: 22 FP: 525 TN:286 ; *sensitivity*: 0.798, *specificity*: 0.353 *Index test 4* level low (+ve) TP: 82 FN: 27 FP: 440 TN:371 ; *sensitivity*: 0.752, *specificity*: 0.457 *Index test 5*: jagged border (+ve) TP: 50 FN: 59 FP: 112 TN:699 ; *sensitivity*: 0.458, *specificity*: 0.862
Source of funding	No funding stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias): Very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus retrospective studies are downgraded for indirectness.*
Comments	

Reference	**Akhaven, 2016[13]**
Study type	prospective
Number of patients	n = 90
Patient characteristics	*Age, mean (SD): 45.95(12.3)* *Gender (female to male ratio): 73:17* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): U

Reference	**Akhaven, 2016[13]**
	Setting: Secondary care otorhinolaryngology clinic *Country*: Iran *Inclusion criteria*: patients with a thyroid nodule referred to the research hospital who were candidates for surgical nodule resection *Exclusion criteria*: patients with known thyroid autoimmune disease, age < 15, contraindications to anaesthesia, surgical resection
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using unclearly described equipment. Sonographic characteristics examined were length, width, area, tall shape, microcalcifications, solidity, irregular margins, echogenicity, vascularity and extracapsular extension *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Akhaven, 2016[13]**
Results	*Malignant n=16; benign n=74* *Index test 1: US overall. The threshold was not described clearly but appears to have been based on the existence of hypoechoicity, microcalcifications, irregular margins, incomplete halo, nodule taller than wide and enlargement of the nodule. However, it was unclear how many (one, some, all?) of these characteristics were necessary to meet the criterion of a positive test for malignancy.* TP: 9 FN: 7 FP: 3 TN: 71; *sensitivity*: 0.5625 , *specificity*: 0.959 *Index test 2: Single nodule (+ve)* TP: 8 FN: 8 FP: 33 TN: 41; *sensitivity*: 0.50 , *specificity*: 0.55 *Index test 3: microcalcification (+ve)* TP: 12 FN: 4 FP: 27 TN: 37; *sensitivity*: 0.75 , *specificity*: 0.50 *Index test 4: irregular border (+ve)* TP: 4 FN: 12 FP: 5 TN: 39; *sensitivity*: 0.25 , *specificity*: 0.527 *Index test 5: solid (+ve)* TP: 14 FN: 2 FP: 58 TN: 16; *sensitivity*: 0.875 , *specificity*: 0.216 *Index test 6: hypervascularity (+ve)* TP: 9 FN: 7 FP: 12 TN: 62; *sensitivity*: 0.5625 , *specificity*: 0.837 *Index test 7: tall shape (+ve)* TP: 1 FN: 15 FP: 2 TN: 72; *sensitivity*: 0.0625 , *specificity*: 0.972
Source of funding	None
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious *Indirectness (QUADAS 2 - applicability)*: None
Comments	

Reference	**Shuzen, 2011[353]**
Study type	prospective
Number of patients	n = 244 patients with 291 thyroid nodules

Reference	**Shuzen, 2011[353]**
Patient characteristics	*Age, mean (SD): 43.38 (0.83)* *Gender (female to male ratio): 183:61* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Unclear *Country*: China *Inclusion criteria*: Not reported *Exclusion criteria*: Not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 6-13 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*66 malignant and 225 benign nodules* *Index test 1: Conventional Ultrasonography. No description of the threshold used.* TP: 64 FN: 2 FP: 79 TN: 146 ; *sensitivity*: 0.97, *specificity*: 0.649

Reference	**Shuzen, 2011[353]**
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

2

Reference	**Liu, 2014[240]**
Study type	prospective
Number of patients	n = 49 patients with 64 focal thyroid nodules
Patient characteristics	*Age, mean (SD): 45.3 (13.1)* *Gender (female to male ratio): 36:13* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: University Hospital *Country*: China *Inclusion criteria*: Patients with thyroid nodules; surgery performed with histopathology performed within the study period; thyroid function and autoimmune bodies tested before surgery *Exclusion criteria*: only diffuse thyroid disease; cystic nodule of completely liquid components; insufficient thyroid parenchyma surrounding the nodule; previous history of radiation therapy of head and neck
Target condition(s)	Thyroid nodule malignancy

Reference	Liu, 2014[240]
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 4-15 MHz probe frequency; Real-time elastography; SWE *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 19 nodules* *benign n= 45 nodules* *Index test 1: hypoechogenicity (+ve)* TP: 14 FN: 5 FP: 19 TN: 26 *sensitivity*: 0.737, *specificity*: 0.578 *Index test 2: echotexture (unclear what echotexture characterised +ve)* TP: 12 FN: 7 FP: 30 TN: 15 *sensitivity*: 0.632, *specificity*: 0.333 *Index test 3: margins (unclear what margin characterised +ve)* TP: 11 FN: 8 FP: 9 TN: 36 *sensitivity*: 0.579, *specificity*: 0.800 *Index test 4: shape (unclear what shape characterised +ve)* TP: 10 FN: 9 FP: 8 TN: 37 *sensitivity*: 0.526, *specificity*: 0.822 *Index test 5: halo sign (+ve)* TP: 17 FN: 2 FP: 34 TN: 11 *sensitivity*: 0.895, *specificity*: 0.244 *Index test 6: micro-calcifications (+ve)* TP: 10 FN: 9 FP: 6 TN: 39 *sensitivity*: 0.526, *specificity*: 0.867 *Index test 7: doppler colour flow (+ve)* TP: 8 FN: 11 FP: 11 TN: 34 *sensitivity*: 0.421, *specificity*: 0.756 *Index test 8: RTE Rago score 2 or over (+ve)* TP: 19 FN: 0 FP: 40 TN: 5 *sensitivity*: 1.0, *specificity*: 0.111 *Index test 8: RTE Rago score 3 or over (+ve)*

Reference	**Liu, 2014[240]**
	TP: 19 FN: 0 FP: 23 TN: 22 *sensitivity*: 1.0, *specificity*: 0.488 *Index test 8: RTE Rago score 4 or over (+ve)* TP: 15 FN: 4 FP: 7 TN: 38 *sensitivity*: 0.789, *specificity*: 0.844 *Index test 8: RTE Rago score 5 (+ve)* TP: 5 FN: 14 FP: 1 TN: 44 *sensitivity*: 0.263, *specificity*: 0.977 *Index test 8: SWE elastic threshold of 39.3kpa or over (+ve)* TP: 13 FN: 6 FP: 6 TN: 39 *sensitivity*: 0.684 *specificity*: 0.867
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Berni, 2002[33]**
Study type	Appears to be retrospective but unclear
Number of patients	n = 108
Patient characteristics	*Age, mean: 45.5 (range 32-72)* *Gender (female to male ratio): 72:32 (noted that this does not add up to 108)* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown

Reference	**Berni, 2002[33]**
	Setting: Surgical Sciences Department in a University Hospital *Country*: Italy *Inclusion criteria*: Not reported *Exclusion criteria*: Not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 10 MHz probe frequency, with colour doppler, power doppler and spectral analysis flow velocity. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*Malignant nodules n= 54 people* *Benign nodules n= 54 people* *Index test 1: The authors used a 'personal classification' to differentiate between benign and malignant nodules as follows: A. vessel distribution: 1. Weak vascular spots, 2. Evident vascularisation rim, 3. Peripheral rim with intramodular bands, 4 spread vascularization; B. vessel morphology: 1. Straight, regular vessels, 2. Tortuosity and/or vessel interruption; C Flow velocity cm/s: 1. Slow flow, 2. Fast flow (systolic >50; diastolic >20); Staging: 1. Absence of signals in other structures, 2. Presence of signals in lymph nodes and/or adjacent structures. However, it is unclear how these criteria were used exactly, so the threshold is unclear.* TP: 48 FN: 6 FP: 10 TN: 44 *sensitivity*: 0.888 , *specificity*: 0.815
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Shweel, 2013[354]**
Study type	prospective
Number of patients	n = 47 patients with 66 nodules
Patient characteristics	*Age, mean (SD): 41 (11)* *Gender (female to male ratio): 35:12* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: unclear *Country*: Egypt *Inclusion criteria*: Surgery planned because of compressive symptoms or clinical and cytological suspicion of malignancy *Exclusion criteria*: declining surgery; cystic nodules; history of external radiation; coarse marginal calcifications
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* HIGH RESOLUTION ultrasound, using 7.5-13 MHz probe frequency; RTE; SWE *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* 2 weeks
Results	*malignant n= 16* *benign n= 50* *Index test 1: Internal content: predominantly solid (+ve for malignancy) [-ve for malignancy was 'completely solid']* TP: 10 FN: 6 FP: 10 TN: 40 *sensitivity*: 0.625 , *specificity*: 0.80

Reference	Shweel, 2013[354]
	Index test 2: Shape: taller than wide (+ve for malignancy) [-ve for malignancy was 'ovoid to round' or 'irregular'] TP: 11 FN: 5 FP: 1 TN: 49 *sensitivity*: 0.688 , *specificity*: 0.98 *Index test 3: Margin: speculated or ill-defined (+ve for malignancy) [-ve for malignancy was 'well defined']* TP: 8 FN: 8 FP: 2 TN: 48 *sensitivity*: 0.50 , *specificity*: 0.96 *Index test 4: Echogenicity: marked hypoechoic (+ve for malignancy) [-ve for malignancy was 'hypoechoic' or 'isoechoic-hyperechoic']* TP: 10 FN: 6 FP: 0 TN: 50 *sensitivity*: 0.625 , *specificity*: 1.0 *Index test 5: calcification: microcalcifications (+ve for malignancy) [-ve for malignancy was none or macro-calcifications]* TP: 8 FN: 8 FP: 0 TN: 50 *sensitivity*: 0.500 , *specificity*: 1.0 *Index test 6: Halo sign: absent (+ve for malignancy)* TP: 14 FN: 2 FP: 6 TN: 44 *sensitivity*: 0.875 , *specificity*: 0.88 The sensitivities and specificities above are calculated from the raw data in table 2 in the paper. These do not agree with the reported sensitivities and specificities in the text of the paper, but it is difficult to envisage how the authors of the paper arrived at the figures they quote. *Index test 7 Elastography ITOH scale 2 and more (+ve for malignancy)* TP: 16 FN: 0 FP: 20 TN: 30 *sensitivity*: 1.0 , *specificity*: 0.60 *Index test 8: Elastography ITOH scale 3 and more (+ve for malignancy)* TP: 14 FN: 2 FP: 8 TN: 42 *sensitivity*: 0.875 , *specificity*: 0.84 *Index test 9: Elastography ITOH scale 4 and more (+ve for malignancy)* TP: 9 FN: 7 FP: 1 TN: 49 *sensitivity*: 0.5625 , *specificity*: 0.98 *Index test 10: Elastography ITOH scale 5 (+ve for malignancy)* TP: 2 FN: 14 FP: 0 TN: 50 *sensitivity*: 0.5625 , *specificity*: 0.98
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Sancak, 2010[333]**
Study type	Prospective
Number of patients	n = unclear
Patient characteristics	*Age, mean (SD): Not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: single medical centre *Country*: Turkey and Germany *Inclusion criteria*: Patients with benign non-functioning nodules, papillary carcinomas and surrounding normal tissue. *Exclusion criteria*: Not reported
Target condition(s)	Thyroid nodule malignancy – papillary carcinoma
Index test(s) and reference standard	*Index test* Doppler ultrasound, using 7.5 MHz probe frequency probe for morphologic examination, and a 5.5 MHz transducer for colour flow doppler examination *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Sancak, 2010[333]**
Results	*malignant n=unclear* *benign n= unclear* *Index test 1: >70% of microvessels (threshold based on ROC analysis)* TP: FN: FP: TN: *sensitivity*: 0.92, *specificity*: 0.89
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Kalantari, 2018[175]**
Study type	prospective
Number of patients	n = 63
Patient characteristics	*Age range: 20-70* *Gender (female to male ratio): 55:8* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: Endocrinology clinic *Country*: Iran *Inclusion criteria*: Presence of thyroid nodules, with indication for surgery because of suspicious and/or indeterminate FNA and/or clinical risk factors for thyroid nodules *Exclusion criteria*: Not reported

Reference	**Kalantari, 2018[175]**
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound and colour doppler using 6-14 MHz probe frequency. Gray Scale US assessed lesion's features including number of nodules (solitary and multiple), echogenicity (hypo echoic, isoechoic and hyper echoic), texture (homogeneous and heterogeneous), margin (irregular and regular margin), calcification (microcalcification and coarse calcification), type (solid or cystic) and the existence of halo. Colour Doppler evaluated the vascular status of nodules (intranodular or perinodular vessels), Resistive Index (RI), Pulsatility Index (PI), and Mean Systolic Velocity (MSV). *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear but reported to follow US and FNA.
Results	*malignant n= 9* *benign n= 54* *Index test 1: multiplicity: solitary nodule (+ve for malignant)[multiple was taken as -ve]* TP: 5 FN: 4 FP: 11 TN: 43; *sensitivity*: 0.555, *specificity*: 0.796 *Index test 2: echogenicity: hypoechoic (+ve for malignant)[iso and hyper-echoic were taken as -ve]* TP: 6 FN: 3 FP: 11 TN: 43; sensitivity: 0.666, *specificity*: 0.796 *Index test 3: texture: heterogeneity (+ve for malignant)[homogeneity was taken as -ve]* TP: 7 FN: 2 FP: 41 TN: 13; *sensitivity*: 0.77, *specificity*: 0.24 *Index test 4: margin: irregular (+ve for malignant)[regular was taken as -ve]* TP: 3 FN: 6 FP: 4 TN: 50; *sensitivity*: 0.33, *specificity*: 0.92 *Index test 5: calcification: micro (+ve for malignant)[coarse was taken as -ve]* TP: 7 FN: 2 FP: 10 TN: 44; *sensitivity*: 0.77, *specificity*: 0.76 *Index test 6: types: solid (+ve for malignant)[cystic was taken as -ve]* TP: 5 FN: 4 FP: 8 TN: 46; *sensitivity*: 0.55, *specificity*: 0.79

Reference	**Kalantari, 2018[175]**
	Index test 7: halo: absent (+ve for malignant) TP: 7 FN: 2 FP: 35 TN: 19; *sensitivity*: 0.44, *specificity*: 0.35 *Index test 8: intra and peri nodular vascularity: absent (+ve for malignant)* TP: 7 FN: 2 FP: 28 TN: 26 *sensitivity*: 0.77, *specificity*: 0.48 *Index test 9: resistive index: ≥0.715 (+ve for malignant)* TP: 8 FN: 1 FP: 11 TN: 43 *sensitivity*: 0.889, *specificity*: 0.796 *Index test 10: pulsatility index: ≥0.945 (+ve for malignant)* TP: 9 FN: 0 FP: 5 TN: 49 *sensitivity*: 1.0, *specificity*: 0.907 *Index test 11: mean systolic velocity: ≥33.5 m/s (+ve for malignant)* TP: 6 FN: 3 FP: 20 TN: 34 *sensitivity*: 0.66, *specificity*: 0.629 *Index test 12: combined doppler and grey scale characteristics: calcification OR resistive index ≥0.715 OR pulsatility index ≥0.945 (+ve for malignant)* TP: 8 FN: 1 FP: 4 TN: 50 *sensitivity*: 0.889, *specificity*: 0.926
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Parikh, 2013[300]**
Study type	retrospective
Number of patients	n = 84
Patient characteristics	*Age, median (range): 53 (15-83)* *Gender (female to male ratio): 65:19*

Reference	**Parikh, 2013[300]**
	Ethnicity: Black 8/84; White 50/84; Other 15/84; unknown 11/84 *Expertise of US tester (medic/non medic/unknown*): medic (surgeon) *Setting*: Division of surgical endocrinology *Country*: USA *Inclusion criteria*: people undergoing surgeon-performed US before thyroidectomy, with provisional diagnosis of Hurtle cell neoplasms on FNA. *Exclusion criteria*: Not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 7.5-13 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Parikh, 2013[300]**
Results	*malignant n=29* *benign n= 55* *Index nodule only examined (1 per person)* *Index test 1: nodule size <4cm (+ve for malignancy)[-ve was ≥4cm]* TP: 23 FN: 6 FP: 50 TN: 5 *sensitivity*: 0.793, *specificity*: 0.10 *Index test 2: Type: solid (+ve for malignancy)[-ve was cystic or mixed or absent]* TP: 14 FN: 15 FP: 24 TN: 31 *sensitivity*: 0.483 , *specificity*: 0.563 *Index test 3: calcifications: micro (+ve for malignancy)[-ve was coarse or absent]* TP: 8 FN: 21 FP: 14 TN: 41 *sensitivity*: 0.276, *specificity*: 0.745 *Index test 4: nodule border: irregular (+ve for malignancy)[-ve was regular or absent]* TP: 7 FN: 22 FP: 11 TN: 44 *sensitivity*: 0.241 , *specificity*: 0.80 *Index test 5: echogenicity: hypoechoic (+ve for malignancy)[-ve was iso- or hyper-echoic or absent]* TP: 10 FN: 19 FP: 18 TN: 37 *sensitivity*: 0.345, *specificity*: 0.672
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias): Very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus retrospective studies are downgraded for indirectness.*
Comments	

1

Reference	**Gorgulu, 2019[111]**
Study type	prospective
Number of patients	n = 83 patients with 101 nodules
Patient characteristics	*Age, mean (SD): 45.98 (11.5)* *Gender (female to male ratio): 56:27*

Reference	Gorgulu, 2019[111]
	Ethnicity: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: otorhinolaryngology and general surgery clinics *Country*: Turkey *Inclusion criteria*: patients admitted for thyroidectomy, according to thyroid US findings, FNA findings and clinical evaluations; single or multiple nodules of 40mm or less *Exclusion criteria*: surgical history of the neck affecting the sternocleidomastoid muscle, pure cystic lesions, insufficient normal tissue around the measured nodule, isthmic nodules, rough calcification and autoimmune thyroiditis
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 4.8-11 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Gorgulu, 2019[111]**
Results	*malignant n=20* *benign n= 81* *Index test 1: nodule size≥15mm (+ve for malignancy) [<15mm was taken as -ve]* TP: 12 FN: 8 FP: 42 TN: 39 *sensitivity*: 0.60, *specificity*: 0.481 *Index test 2: microcalcification (+ve for malignancy) [absent microcalcification was taken as -ve]* TP: 12 FN: 8 FP: 10 TN: 71 *sensitivity*: 0.60 , *specificity*: 0.877 *Index test 3: hypoechogenic (+ve for malignancy) [iso-echogenic was taken as -ve]* TP: 17 FN: 3 FP: 25 TN: 56 *sensitivity*: 0.85 , *specificity*: 0.691 *Index test 4: heterogenous thyroid gland echo texture (+ve for malignancy) homogenous was taken as -ve]* TP: 19 FN: 1 FP: 68 TN: 13 *sensitivity*: 0.95, *specificity*: 0.160
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Gu, 2011[121]**
Study type	prospective
Number of patients	n = 72 with 98 thyroid nodules
Patient characteristics	*Age, mean (SD): 50.69(11.82)* *Gender (female to male ratio): 51:21* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Departments of US and pathology, at a university hospital *Country*: China

Reference	Gu, 2011[121]
	Inclusion criteria: Patients undergoing thyroidectomy *Exclusion criteria*: anatomic abnormalities of the neck, cystic lesions of a completely liquid nature, maximum nodule diameter of <6mm.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 9MHz probe frequency; VTI *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*Malignant nodules n=22* *Benign nodules n= 76* *Index test 1: hypoechoicity (+ve for malignancy) [-ve was taken as iso- or hyper-echoic]* TP: 20 FN: 2 FP: 33 TN: 43 *sensitivity*: 0.909, *specificity*: 0.566 *Index test 2: microcalcifications (+ve for malignancy) [-ve was taken as macro or other]* TP: 10 FN: 12 FP: 2 TN: 74 *sensitivity*: 0.455, *specificity*: 0.974 *Index test 3: Composite US test: morphologic characteristics [unclear which], boundary [not specified], hypoechoicity, absence of the halo sign, and presence of microcalcification (+ve for malignancy) [-ve was taken as macro or other]* TP: 15 FN: 7 FP: 10 TN: 66 *sensitivity*: 0.681, *specificity*: 0.868 *Index test 3: Elastography: VTI – stiffer or honeycomb (+ve for malignancy)* TP: 21 FN: 1 FP: 17 TN: 59 *sensitivity*: 0.954, *specificity*: 0.756 The paper did not report data sufficient to calculate accuracy indices for parameters such as the halo sign, etc. The results of index test 3 are not compatible with those of 1 and 2. If the composite score is based on the Boolean operator 'AND' (as it seems to be) then the

Reference	**Gu, 2011[121]**
	sensitivity cannot be better than the worst sensitivity in any of the composites, but it is better. If it is based on an 'OR' operator it cannot be worse than the best, but it is.
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Hong, 2009[147]**
Study type	prospective
Number of patients	n = 90 with 145 nodules
Patient characteristics	*Age, mean (SD): 46(13)* *Gender (female to male ratio): 74:16* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: University hospital *Country*: China *Inclusion criteria*: Patients with thyroid nodules referred for surgical treatment *Exclusion criteria*: none reported
Target condition(s)	Thyroid nodule malignancy

Reference	Hong, 2009[147]
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 6-13 MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=49* *benign n= 96* *Index test 1: hypoechogenicity (+ve for malignancy)* TP: 42 FN: 7 FP: 28 TN: 68 *sensitivity*: 0.86, *specificity*: 0.71 *Index test 2: spot microcalcifications (+ve for malignancy)* TP: 27 FN: 22 FP: 6 TN: 90 *sensitivity*: 0.55, *specificity*: 0.94 *Index test 3: blurred or spiculated margins (+ve for malignancy)* TP: 45 FN: 4 FP: 15 TN: 81 *sensitivity*: 0.92, *specificity*: 0.84 *Index test 4: AT ≥ 1cm (+ve for malignancy)* TP: 12 FN: 37 FP: 4 TN: 92 *sensitivity*: 0.24, *specificity*: 0.96 *Index test 5: intranodular blood flow (+ve for malignancy)* TP: 17 FN: 32 FP: 40 TN: 56 *sensitivity*: 0.35, *specificity*: 0.58 *Index test 6: Elastography 1-6 ES scale: 4 or higher (+ve for malignancy)* TP: 43 FN: 6 FP: 10 TN: 86 *sensitivity*: 0.88, *specificity*: 0.90
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Kim, 2008[179]**
Study type	prospective
Number of patients	n = 92 patients with 93 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: University hospital *Country*: South Korea *Inclusion criteria*: Patients with eggshell calcifications; patients undergoing thyroidectomy *Exclusion criteria*: nodules with combinations of eggshell calcifications and other types of calcifications such as microcalcifications and macrocalcifications
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 7-12MHz probe or 5-13MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* 27-63 days

Reference	**Kim, 2008[179]**
Results	*malignant n=59* *benign n= 34* *Index test 1: hypoechoic halo (+ve for malignancy)* TP: 37 FN: 59 FP: 2 TN: 32 *sensitivity*: 0.627 , *specificity*: 0.941 *Index test 2: disruption of calcifications (+ve for malignancy)* TP: 45 FN: 14 FP: 12 TN: 22 *sensitivity*: 0.763 , *specificity*: 0.647 *Index test 3: marked hypoechogenicity (+ve for malignancy)* TP: 24 FN: 35 FP: 7 TN: 27 *sensitivity*: 0.407 , *specificity*: 0.794 *Index test 4: irregular or microlobulated margin (+ve for malignancy)* TP: 21 FN: 38 FP: 1 TN: 33 *sensitivity*: 0.356 , *specificity*: 0.971 *Index test 5: taller than wide shape (+ve for malignancy)* TP: 33 FN: 26 FP: 5 TN: 29 *sensitivity*: 0.559 , *specificity*: 0.853 *Index test 6: halo OR calcification rim (+ve for malignancy)* TP: 55 FN: 4 FP: 12 TN: 22 *sensitivity*: 0.932 , *specificity*: 0.647 *Index test 6: hypoechogenicity OR microlobulated margins OR taller than wide (+ve for malignancy)* TP: 46 FN: 13 FP: 12 TN: 22 *sensitivity*: 0.78 , *specificity*: 0.647
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Kim, 2008[188]**
Study type	retrospective
Number of patients	n = 174 nodules

Reference	**Kim, 2008[188]**
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: University hospital *Country*: South Korea *Inclusion criteria*: Patients with macrocalcifications; patients undergoing thyroidectomy for thyroid nodules *Exclusion criteria*: patients with microcalcifications (multiple punctate bright echoes <2mm with or without acoustic shadows)
Target condition(s)	Thyroid nodule malignancy (papillary carcinoma)
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using unreported probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* 27-63 days

Reference	**Kim, 2008[188]**
Results	*malignant n=116* *benign n= 58* *Index test 1: solitary macrocalcifications (+ve for malignancy) [note that ALL had MACROcalcifications, the -ve or benign category was non-solitary]* TP: 26 FN: 90 FP: 24 TN: 34 *sensitivity*: 0.224, *specificity*: 0.586 *Index test 2: eggshell macrocalcifications (+ve for malignancy) [note that ALL had MACROcalcifications, the -ve or benign category was non-egg-shell]* TP: 52 FN: 64 FP: 29 TN: 29 *sensitivity*: 0.448, *specificity*: 0.50 *Index test 3: coarse not otherwise specified (NOS) macrocalcifications (+ve for malignancy) [note that ALL had MACROcalcifications, the -ve or benign category was non-NOS]* TP: 28 FN: 88 FP: 5 TN: 53 *sensitivity*: 0.241 , *specificity*: 0.914 *Index test 4: suspicious sonographic features: 1 or more of solitary/eggshell or coarse NOS (+ve for malignancy)* TP: 96 FN: 20 FP: 20 TN: 38 *sensitivity*: 0.828 , *specificity*: 0.655 *Index test 5: suspicious sonographic features: 2 or more of solitary/eggshell or coarse NOS (+ve for malignancy)* TP: 58 FN: 58 FP: 6 TN: 52 *sensitivity*: 0.50 , *specificity*: 0.897 *Index test 6: suspicious sonographic features: all 3 of solitary/eggshell or coarse NOS (+ve for malignancy)* TP: 28 FN: 88 FP: 0 TN: 58 *sensitivity*: 0.241 , *specificity*: 1.0
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Kong, 2017[198]**
Study type	retrospective

Reference	**Kong, 2017[198]**
Number of patients	n = 92 patients with 113 nodules
Patient characteristics	*Age, median (range): 42 (20-75)* *Gender (female to male ratio): 11:12* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: University hospital *Country*: China *Inclusion criteria*: presence of a signal or multiple nodular thyroid lesions; thyroidectomy surgery with full histopathological results *Exclusion criteria*: FNA performed before admission, which might affect vascularity
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale, doppler and 'superb microvascular imaging' (an advanced form of doppler imaging) ultrasound, using unreported MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Kong, 2017[198]**
Results	*malignant n=79* *benign n= 34* *Index test 1: solid component (+ve for malignancy)* TP: 66 FN: 13 FP: 17 TN: 17 *sensitivity*: 0.835, *specificity*: 0.50 *Index test 2: hypoechogenicity (+ve for malignancy)* TP: 69 FN: 10 FP: 19 TN: 15 *sensitivity*: 0.873, *specificity*: 0.441 *Index test 3: taller than wide (+ve for malignancy)* TP: 40 FN: 39 FP: 6 TN: 28 *sensitivity*: 0.506, *specificity*: 0.824 *Index test 4: irregular margin (+ve for malignancy)* TP: 42 FN: 36 FP: 7 TN: 30 *sensitivity*: 0.532, *specificity*: 0.794 *Index test 5: microcalcifications (+ve for malignancy)* TP: 51 FN: 28 FP: 8 TN: 26 *sensitivity*: 0.646, *specificity*: 0.765 *Index test 6: intranodular vascularity superb microvascular imaging grade III or higher (+ve for malignancy)* TP: 60 FN: 19 FP: 3 TN: 31 *sensitivity*: 0.759, *specificity*: 0.912
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Lyshchik, 2007[253]**
Study type	prospective
Number of patients	n = 56 patients with 86 nodules
Patient characteristics	*Age, mean (SD): 53.1(11.6)*

Reference	**Lyshchik, 2007[253]**
	Gender (female to male ratio): 44:12 *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: University hospital *Country*: Japan *Inclusion criteria*: patients with solid thyroid nodules with the preoperative suspicion of thyroid cancer based on clinical, imaging and cytologic findings *Exclusion criteria*: Patients who refused to give informed consent or who did not receive surgical treatment
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Power doppler ultrasound, using 5-9MHz probe frequency, with Doppler frequency range of 5-7MHz. Quantitative analysis (rather than visual analysis) used. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Lyshchik, 2007[253]**
Results	*malignant n= 46* *benign n= 40* *Index test 1: qualitative analysis: power doppler type 2* or higher (+ve for malignancy)[PD type 1=_ve]* TP: 34 FN: 12 FP: 30 TN: 10 *sensitivity*: 0.739, *specificity*: 0.75 *Index test 2: qualitative analysis: power doppler type 3* (+ve for malignancy)[PD types 1 and 2=-ve]* TP: 30 FN: 16 FP: 19 TN: 21 *sensitivity*: 0.652, *specificity*: 0.525 *type1=absence of flow signals, type 2=increased perinodular vascularisation, type 3= increased perinodular and intranodular vascularisation *Index test 3: quantitative analysis: normalised VI of >0.278 (+ve for malignancy)* TP: 38 FN: 8 FP: 18 TN: 22 *sensitivity*: 0.825, *specificity*: 0.543 *Index test 4: quantitative analysis: normalised VI of >0.14 (+ve for malignancy) RESTRICTED TO NODULES <2cm* TP: 21 FN: 8 FP: 0 TN: 14 *sensitivity*: 0.724, *specificity*: 1.00
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: No risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Ren, 2015[321]**
Study type	Retrospective
Number of patients	n = 124 patients with 207 nodules
Patient characteristics	*Age, mean (SD): 45(10)* *Gender (female to male ratio): 100:24* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown (‘expert sonologists’)

Reference	**Ren, 2015[321]**
	Setting: University Hospital *Country*: China *Inclusion criteria*: patients with a preoperative diagnosis of thyroid nodules of any size by sonographers; patients undergoing total or partial thyroidectomy surgery with a histopathological diagnosis of malignancy or benignity; patients with predominantly solid nodules (>50%) *Exclusion criteria*: malignancy that was not papillary TC
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 5-12 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=110* *benign n= 97* *The results in this paper were stratified into nodule size: <0.5cm [malignant n=43, benign = 31]; 0.5-1cm [malignant n=42, benign = 36]; >1cm [malignant n=25, benign = 30]* **<0.5cm** *Index test 1: A/T >1 (+ve for malignancy)* TP: 35 FN: 8 FP: 1 TN: 30 *sensitivity*: 0.814, *specificity*: 0.968 *Index test 2: blurred margins (+ve for malignancy)* TP: 42 FN: 1 FP:7 22 TN: 9 *sensitivity*: 0.977, *specificity*: 0.290 *Index test 2: hypoechogenicity (+ve for malignancy)* TP: 23 FN: 20 FP: 25 TN: 6 *sensitivity*: 0.53 *specificity*: 0.194

Reference	Ren, 2015[321]
	Index test 2: microcalcifications (+ve for malignancy) TP: 12 FN: 31 FP: 3 TN: 28 *sensitivity*: 0.279 *specificity*: 0.903 **<u>0.5-1cm</u>** *Index test 1: A/T >1 (+ve for malignancy)* TP: 33 FN: 9 FP: 2 TN: 34 *sensitivity*: 0.786, *specificity*: 0.944 *Index test 2: blurred margins (+ve for malignancy)* TP: 41 FN: 1 FP: 16 TN: 20 *sensitivity*: 0.976, *specificity*: 0.556 *Index test 2: hypoechogenicity (+ve for malignancy)* TP: 40 FN: 2 FP: 22 TN: 14 *sensitivity*: 0.952 *specificity*: 0.389 *Index test 2: microcalcifications (+ve for malignancy)* TP: 20 FN: 22 FP: 3 TN: 33 *sensitivity*: 0.476 *specificity*: 0.917 **<u>>1cm</u>** *Index test 1: A/T >1 (+ve for malignancy)* TP: 6 FN: 19 FP: 2 TN: 28 *sensitivity*: 0.250, *specificity*: 0.935 *Index test 2: blurred margins (+ve for malignancy)22* TP: 25 FN: 0 FP: 7 TN: 23 *sensitivity*: 1.00, *specificity*: 0.767 *Index test 2: hypoechogenicity (+ve for malignancy)* TP: 23 FN: 2 FP: 10 TN: 20 *sensitivity*: 0.917 *specificity*: 0.677 *Index test 2: microcalcifications (+ve for malignancy)* TP: 15 FN: 10 FP: 1 TN: 29 *sensitivity*: 0.600 *specificity*: 0.967 In combination with at least one of the other 3 criteria the following sensitivities and specificities were found for each sonographic feature in each of the strata: **<u><0.5cm</u>**

Reference	Ren, 2015[321]
	Index test 1: A/T >1 + at least 1 of the other 3 criteria (+ve for malignancy) TP: 35 FN: 8 FP: 1 TN: 30 *sensitivity*: 0.814, *specificity*: 0.968 *Index test 2: blurred margins + at least 1 of the other 3 criteria (+ve for malignancy)* TP: 42 FN: 1 FP: 19 TN: 12 *sensitivity*: 0.977, *specificity*: 0.387 *Index test 2: hypoechogenicity + at least 1 of the other 3 criteria (+ve for malignancy)* TP: 40 FN: 3 FP: 7 TN: 24 *sensitivity*: 0.93 *specificity*: 0.786 *Index test 2: microcalcifications + at least 1 of the other 3 criteria (+ve for malignancy)* TP: 12 FN: 31 FP: 3 TN: 28 *sensitivity*: 0.279 *specificity*: 0.903 **0.5-1cm** *Index test 1: A/T >1 + at least 1 of the other 3 criteria (+ve for malignancy)* TP: 33 FN: 9 FP: 2 TN: 34 *sensitivity*: 0.786, *specificity*: 0.944 *Index test 2: blurred margins + at least 1 of the other 3 criteria (+ve for malignancy)* TP: 41 FN: 1 FP: 14 TN: 22 *sensitivity*: 0.976, *specificity*: 0.611 *Index test 2: hypoechogenicity + at least 1 of the other 3 criteria (+ve for malignancy)* TP: 40 FN: 2 FP: 15 TN: 21 *sensitivity*: 0.952 *specificity*: 0.583 *Index test 2: microcalcifications + at least 1 of the other 3 criteria (+ve for malignancy)* TP: 20 FN: 22 FP: 3 TN: 33 *sensitivity*: 0.476 *specificity*: 0.917 **>1cm** *Index test 1: A/T >1 + at least 1 of the other 3 criteria (+ve for malignancy)* TP: 7 FN: 18 FP: 1 TN: 29 *sensitivity*: 0.280, *specificity*: 0.967 *Index test 2: blurred margins + at least 1 of the other 3 criteria (+ve for malignancy)* TP: 25 FN: 0 FP: 7 TN: 23 *sensitivity*: 1.00, *specificity*: 0.830 *Index test 2: hypoechogenicity + at least 1 of the other 3 criteria (+ve for malignancy)* TP: 22 FN: 3 FP: 7 TN: 23 *sensitivity*: 0.88 *specificity*: 0.767

Reference	**Ren, 2015[321]**
	Index test 2: microcalcifications + at least 1 of the other 3 criteria (+ve for malignancy) TP: 15 FN: 10 FP: 1 TN: 29 *sensitivity*: 0.600 *specificity*: 0.967
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Wang, 2018[388]**
Study type	Prospective
Number of patients	n = 445 nodules from 445 patients
Patient characteristics	*Age, mean (range): 44.1 (16-82)* *Gender (female to male ratio): 330: 115* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: single cancer centre *Country*: China *Inclusion criteria*: patients with thyroid nodules undergoing US and FNA; receiving nodules confirmed by histopathologic analysis after surgery; no treatment previously performed on nodules *Exclusion criteria*: nodules >40mm, nodules with a cystic component of >25%, nodules in the isthmus, nodules with calcification covering >25% of the nodule

Reference	**Wang, 2018[388]**
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using unreported MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 347* *benign n= 98* *Index test 1: maximum diameter ≤10mm (+ve for malignancy)* TP: 140 FN: 207 FP: 18 TN: 80 *sensitivity*: 0.403 *specificity*: 0.816 *Index test 2: maximum diameter ≤20mm (+ve for malignancy)* TP: 259 FN: 88 FP: 53 TN: 45 *sensitivity*: 0.746 *specificity*: 0.459 *Index test 3: hypoechoic (+ve for malignancy)* TP: 298 FN: 49 FP: 48 TN: 50 *sensitivity*: 0.858 *specificity*: 0.51 *Index test 4: microcalcification (+ve for malignancy)* TP: 91 FN: 256 FP: 5 TN: 93 *sensitivity*: 0.262 *specificity*: 0.949
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Xing, 2011[410]**
Study type	Prospective
Number of patients	n = 86 with 98 thyroid nodules
Patient characteristics	*Age, mean (SD): 47(11)* *Gender (female to male ratio): 71:15* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: University Hospital *Country*: China *Inclusion criteria*: presence of single or multiple thyroid nodules whose size did not exceed 40mm *Exclusion criteria*: cystic nodules, complex and partially cystic lesions, and nodules with a calcified shell; histologic findings of chronic inflammation
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 10 MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Xing, 2011[410]**
Results	*malignant n=45 nodules* *benign n= 53 nodules* *Index test 1: hypoechogenicity (+ve for malignancy)* TP: 32 FN: 13 FP: 18 TN: 35 *sensitivity*: 0.711 *specificity*: 0.66 *Index test 2: spot microcalcification (+ve for malignancy)* TP: 23 FN: 22 FP: 4 TN: 49 *sensitivity*: 0.511 *specificity*: 0.924 *Index test 3: speculated margins (+ve for malignancy)* TP: 29 FN: 16 FP: 7 TN: 46 *sensitivity*: 0.644 *specificity*: 0.867 *Index test 4: A/T ratio >1 (+ve for malignancy)* TP: 28 FN: 17 FP: 13 TN: 40 *sensitivity*: 0.622 *specificity*: 0.754 *Index test 5: Type 3 pattern of single intranodular blood flow (+ve for malignancy)* TP: 26 FN: 19 FP: 37 TN: 16 *sensitivity*: 0.577 *specificity*: 0.301 *Index test 6: Elastography: Asteria 1-4 colour score – 3 or more (+ve for malignancy)* TP: 40 FN: 5 FP: 10 TN: 43 *sensitivity*: 0.888 *specificity*: 0.811 *Index test 6: Elastography: strain ratio at cut off 3.79 or more (+ve for malignancy)* TP: 44 FN: 1 FP: 8 TN: 45 *sensitivity*: 0.978 *specificity*: 0.857
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

Reference	**Zhang, 2017[423]**
Study type	Prospective

Reference	**Zhang, 2017[423]**
Number of patients	n = 128 with 152 nodules
Patient characteristics	*Age, mean (SD): 48(11)* *Gender (female to male ratio): 104:24* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: University Hospital *Country*: China *Inclusion criteria*: solid or almost solid nodules (<20% cystic); nodules at least 0.5cm, with sufficient peripheral gland at the same depth appearing normal; neither clinical treatment nor FNA/core biopsy performed before US *Exclusion criteria*: unstable dynamic images on strain elastography; nodules with invalid SWV values
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using unreported MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=55* *benign n= 97* *Index test 1: 3 or more of the following: solid nodule, microcalcifications, marked hypoechogenicity, irregular margins, absence of halo and taller than wide shape (+ve for malignancy)* TP: 48 FN: 7 FP: 5 TN: 92 *sensitivity*: 0.873 *specificity*: 0.949

Reference	Zhang, 2017[423]
	Index test 2: elastography – virtual touch tissue imaging on I to VI grade – grade II or more (+ve for malignancy) TP: 55 FN: 0 FP: 79 TN: 18 *sensitivity*: 1.0 *specificity*: 0.186 *Index test 2: elastography – virtual touch tissue imaging on I to VI grade – grade III or more (+ve for malignancy)* TP: 47 FN: 8 FP: 35 TN: 62 *sensitivity*: 0.854 *specificity*: 0.639 *Index test 2: elastography – virtual touch tissue imaging on I to VI grade – grade IV or more (+ve for malignancy)* TP: 40 FN: 15 FP: 10 TN: 87 *sensitivity*: 0.720 *specificity*: 0.897 *Index test 2: elastography – virtual touch tissue imaging on I to VI grade – grade V or more (+ve for malignancy)* TP: 7 FN: 48 FP: 0 TN: 97 *sensitivity*: 0.127 *specificity*: 1.0 *Index test 2: elastography – shear wave velocity at 2.87 m/s and above (+ve for malignancy)* TP: 43 FN: 12 FP: 16 TN: 81 *sensitivity*: 0.782 *specificity*: 0.835 *Index test 2: elastography – strain ratio at 2.37 and above (+ve for malignancy)* TP: 50 FN: 5 FP: 25 TN: 72 *sensitivity*: 0.909 *specificity*: 0.742
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

2

Reference	Zhang, 2014[425]
Study type	Prospective
Number of patients	n = 59 with 71 nodules
Patient characteristics	*Age, mean (SD): 50.5(9.1)* *Gender (female to male ratio): 43:16* *Ethnicity*: not reported

Reference	Zhang, 2014[425]
	Expertise of US tester (medic/non medic/unknown): medic *Setting*: University Hospital *Country*: China *Inclusion criteria*: Patients with nodules <10 mm, receiving US, ARFI and thyroid surgery for thyroid nodules; *Exclusion criteria*: Cystic lesion of a completely liquid nature
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 4-9 MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=32* *benign n= 39* *Index test 1: taller than wide (+ve for malignancy) [-ve designated by oval to round]* TP: 18 FN: 14 FP: 1 TN: 38 *sensitivity*: 0.563 *specificity*: 0.974 *Index test 2: poorly defined boundary (+ve for malignancy) [-ve designated by well-defined]* TP: 18 FN: 14 FP: 9 TN: 30 *sensitivity*: 0.563 *specificity*: 0.769 *Index test 3: markedly hypoechoic (+ve for malignancy) [-ve designated by hypo-, iso-, or hyper-echoic]* TP: 19 FN: 13 FP: 10 TN: 29 *sensitivity*: 0.594 *specificity*: 0.744 *Index test 4: homogenous echo texture (+ve for malignancy) [-ve designated by heterogeneous]* TP: 23 FN: 9 FP: 32 TN: 7 *sensitivity*: 0.719 *specificity*: 0.179

Reference	**Zhang, 2014[425]**
	Index test 5: microcalcification (+ve for malignancy) [-ve designated by macro or no calcification] TP: 11 FN: 21 FP: 7 TN: 32 *sensitivity*: 0.344 *specificity*: 0.821 *Index test 6: Colour doppler: artery (+ve for malignancy) [-ve designated by vein or no vessel]* TP: 23 FN: 9 FP: 24 TN: 15 *sensitivity*: 0.719 *specificity*: 0.385 *Index test 6: Colour doppler: VTQ >2.910m/s (+ve for malignancy) [based on ROC analysis]* TP: 23 FN: 9 FP: 0 TN: 39 *sensitivity*: 0.719 *specificity*: 1.0 *Index test 6: elastography: virtual touch quantification at 2.91 m/s and above (+ve for malignancy) [based on ROC analysis]* TP: 23 FN: 9 FP: 0 TN: 39 *sensitivity*: 0.719 *specificity*: 1.0 *Index test 6: elastography: virtual touch quantification at 2.91 m/s and above OR taller than wide (+ve for malignancy)* TP: 29 FN: 3 FP: 1 TN: 38 *sensitivity*: 0.906 *specificity*: 0.974 *Index test 6: elastography: virtual touch quantification at 2.91 m/s and above OR poorly defined margins (+ve for malignancy)* TP: 28 FN: 4 FP: 9 TN: 30 *sensitivity*: 0.875 *specificity*: 0.769 *Index test 6: elastography: virtual touch quantification at 2.91 m/s and above OR markedly hypoechoic (+ve for malignancy)* TP: 29 FN: 3 FP: 10 TN: 29 *sensitivity*: 0.906 *specificity*: 0.744
Source of funding	National government funding (non commercial)
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Vorlander, 2010[386]**
Study type	Prospective
Number of patients	n = 309 with 309 dominant nodules
Patient characteristics	*Age, mean (SD): 47.2(13)* *Gender (female to male ratio): 207:102*

Reference	Vorlander, 2010[386]
	Ethnicity: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: clinic for endocrine surgery *Country*: Germany *Inclusion criteria*: patients with solid dominant nodules on one or both thyroid lobes *Exclusion criteria*: patients with an autonomous adenoma, hyperthyreosis or recurrent goitre
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 6-13 MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 50* *benign n= 259* *Index test 1: hypoechoicity (+ve for malignancy)* TP: 43 FN: 7 FP: 91 TN: 168 *sensitivity*: 0.86 *specificity*: 0.648 *Index test 1: colour doppler – increased perfusion (+ve for malignancy)* TP: 40 FN: 10 FP: 196 TN: 63 *sensitivity*: 0.80 *specificity*: 0.243 *Index test 1: elastography: L 3 point Rago scale – III (hard) (+ve for malignancy)* TP: 35 FN: 15 FP: 46 TN: 213 *sensitivity*: 0.70 *specificity*: 0.822

Reference	**Vorlander, 2010[386]**
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Giammanco, 2002[108]**
Study type	Prospective
Number of patients	n = 125 patients with 125 nodules
Patient characteristics	*Age, mean (range): 57.2 (35-70)* *Gender (female to male ratio): 98:27* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: University Hospital *Country*: China *Inclusion criteria*: Patients with thyroid pathology receiving US B mode, doppler and thyroid surgery with histopathological confirmation of diagnosis *Exclusion criteria*: none reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 7.5 MHz probe frequency. Colour doppler used low frequency impulse repetition, colour gain adjusted at 60% with low filters in order to exclude low frequency blood flow *Reference (gold) standard:* Surgical histopathological findings

Reference	**Giammanco, 2002[108]**
	Time between measurement of index test and reference standard: Not clear
Results	*malignant n=7* *benign n= 118* *Index test 1: Flow rate of II* or more (+ve for malignancy) [Flow rate of I was designated as -ve for malignancy]* TP: 7 FN: 0 FP: 89 TN: 29 *sensitivity*: 1.0 *specificity*: 0.246 *Index test 1: Flow rate of III* or more (+ve for malignancy) [Flow rate of I and II were designated as -ve for malignancy]* TP: 6 FN: 1 FP: 28 TN: 90 *sensitivity*: 0.857 *specificity*: 0.763 *Index test 1: Flow rate of IV* (+ve for malignancy) [Flow rate of I, II and III were designated as -ve for malignancy]* TP: 0 FN: 7 FP: 6 TN: 112 *sensitivity*: 0.0 *specificity*: 0.949 *I=absent blood flow; II=perinodular blood flow; III=peri and intra-nodular blood flow; IV=diffused blood flow
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Goldfarb, 2011[267]**
Study type	retrospective
Number of patients	n = 624
Patient characteristics	*Age, mean (SD): 50.57* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown)*: medic ('surgeon')

<table>
<tr><td>Reference</td><td>Goldfarb, 2011[267]</td></tr>
<tr><td></td><td>Setting: University Hospital

Country: USA

Inclusion criteria: consecutive patients undergoing thyroidectomy; solitary or dominant thyroid nodules; underwent surgeon-operated US before thyroidectomy

Exclusion criteria: not reported</td></tr>
<tr><td>Target condition(s)</td><td>Thyroid nodule malignancy</td></tr>
<tr><td>Index test(s) and reference standard</td><td>Index test
Grey scale ultrasound, using 7.5-13 MHz probe frequency

Reference (gold) standard:
Surgical histopathological findings

Time between measurement of index test and reference standard:
Not clear</td></tr>
<tr><td>Results</td><td>malignant n=217
benign n= 407

This study collected data retrospectively to develop a diagnostic algorithm to predict benignity. A logistic regression was carried out, evaluating the independent predictors of benignity, yielding a final model that gave a point each to the following 5 variables: size<1 cm, cystic components, no microcalcifications, regular borders and isoechoic. The threshold for benignity was more than or equal to 4 points. On validation (apparently within the same dataset as used for the development!) this yielded a sensitivity for predicting benignity of 0.106 and a specificity of 0.976. The 2x2 table was as follows (raw data calculated from the sensitivities and specificities and the known numbers of truly benign and malignant nodules):
<table>
<tr><td></td><td>Truly benign</td><td>Truly malignant</td></tr>
<tr><td>4 points or more (=benign)</td><td>43</td><td>5</td></tr>
<tr><td><4 points (=malignant)</td><td>364</td><td>212</td></tr>
</table>
From this it was easy to switch the rows and columns to yield sensitivities and specificities for predicting malignancy, in line with the aims of this review:
<table>
<tr><td></td><td>Truly malignant</td><td>Truly benign</td></tr>
<tr><td><4 points (=malignant)</td><td>212</td><td>364</td></tr>
</table></td></tr>
</table>

Reference	**Goldfarb, 2011[267]**
	4 points or more (=benign) | *5* | *43* *Index test 1: < 4 of the following US characteristics: size<1 cm, cystic components, no microcalcifications, regular borders and isoechoic. (+ve for malignancy)* TP: 212 FN: 5 FP: 364 TN: 43 *sensitivity*: 0.976 *specificity*: 0.106
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Ma, 2014[256]**
Study type	Prospective
Number of patients	n = 144 patients with 172 thyroid nodules
Patient characteristics	*Age, mean: 48.65* *Gender (female to male ratio): 105:39* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): tester unclear but reviewed by a medic *Setting*: University Hospital *Country*: China *Inclusion criteria*: >18 years; solid or mainly solid nodules on GSUS and CDUS; nodule size>0.5cm

Reference	**Ma, 2014[256]**
	Exclusion criteria: dominant cystic nodules; pregnancy; grade III-IV NYHA cardiac function; severe pulmonary hypertension; no surgical pathology of the thyroid gland
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using unknown MHz probe frequency; colour doppler; contrast enhanced US *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 94* *benign n= 78* *Index test 1: Shape on grey scale US – round or irregular (+ve for malignancy) [-ve result was designated by oval]* TP: 79 FN: 15 FP: 26 TN: 52 *sensitivity*: 0.84 *specificity*: 0.667 *Index test 2: orientation on grey scale US – round or taller than wide (+ve for malignancy) [-ve result was designated by wider than tall]* TP: 41 FN: 53 FP: 11 TN: 67 *sensitivity*: 0.436 *specificity*: 0.859 *Index test 3: interior echogenicity on grey scale US – hypoechoic (+ve for malignancy) [-ve result was designated by hyper-or iso-echoic]* TP: 73 FN: 21 FP: 16 TN: 62 *sensitivity*: 0.777 *specificity*: 0.795 *Index test 4: halo on grey scale US – incomplete or none (+ve for malignancy) [-ve result was designated by complete]* TP: 90 FN: 4 FP: 38 TN: 40 *sensitivity*: 0.957 *specificity*: 0.513 *Index test 5: microcalcification on grey scale US – present (+ve for malignancy) [-ve result was designated by absent]* TP: 62 FN: 32 FP: 4 TN: 74 *sensitivity*: 0.66 *specificity*: 0.949 *Index test 6: relative arrival time of nodule on contrast enhanced US – later (+ve for malignancy) [-ve result was designated by earlier or concurrent]* TP: 51 FN: 43 FP: 6 TN: 72 *sensitivity*: 0.543 *specificity*: 0.923

Reference	Ma, 2014[256]
	Index test 7: peak peripheral echogenicity on contrast enhanced US – hypoechoic (+ve for malignancy) [-ve result was designated by iso- or hyper-echoic] TP: 62 FN: 32 FP: 14 TN: 64 *sensitivity*: 0.66 *specificity*: 0.821 *Index test 8: peak interior echogenicity on contrast enhanced US – hypoechoic (+ve for malignancy) [-ve result was designated by iso- or hyper-echoic]* TP: 72 FN: 22 FP: 17 TN: 61 *sensitivity*: 0.766 *specificity*: 0.782 *Index test 9: ring enhancement on contrast enhanced US – incomplete or none (+ve for malignancy) [-ve result was designated by complete]* TP: 91 FN: 3 FP: 14 TN: 64 *sensitivity*: 0.968 *specificity*: 0.821 *Index test 10: homogeneity of enhancement on contrast enhanced US – heterogenous (+ve for malignancy) [-ve result was designated by homogeneous]* TP: 85 FN: 9 FP: 7 TN: 71 *sensitivity*: 0.904 *specificity*: 0.910 *Index test 11: homogeneity of enhancement on contrast enhanced US – heterogenous (+ve for malignancy) [-ve result was designated by homogeneous]* TP: 85 FN: 9 FP: 7 TN: 71 *sensitivity*: 0.904 *specificity*: 0.910 *Index test 12: vascularisation on colour doppler US – grade II and III (+ve for malignancy) [-ve result was designated by grade I]* TP: 73 FN: 21 FP: 61 TN: 17 *sensitivity*: 0.777 *specificity*: 0.218 *Index test 13: vascularisation on colour doppler US – grade III (+ve for malignancy) [-ve result was designated by grade I and II]* TP: 35 FN: 59 FP: 32 TN: 46 *sensitivity*: 0.372 *specificity*: 0.590 *Index test 14: RI on colour doppler US – ≥0.75 (+ve for malignancy) [-ve result was designated by <0.75]* TP: 46 FN: 35 FP: 26 TN: 37 *sensitivity*: 0.568 *specificity*: 0.587 *Index test 15: perforating branches on colour doppler US – presence (+ve for malignancy) [-ve result was designated by absence]* TP: 53 FN: 41 FP: 39 TN: 39 *sensitivity*: 0.564 *specificity*: 0.500 *Index test 15: vascular distribution on colour doppler US – some (+ve for malignancy) [-ve result was designated by none]* TP: 79 FN: 15 FP: 53 TN: 15 *sensitivity*: 0.840 *specificity*: 0.192
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: No risk of bias

Reference	**Ma, 2014[256]**
	Indirectness (QUADAS 2 - applicability): none
Comments	

1

Reference	**Park, 2012[303]**
Study type	Retrospective
Number of patients	n = 400 nodules in 400 patients
Patient characteristics	*Age, mean (range): 46(15-77)* *Gender (female to male ratio): 349:51* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: University Hospital *Country*: South Korea *Inclusion criteria*: Patients undergoing US-FNA and surgery during the study period *Exclusion criteria*: Not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 5-12 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Park, 2012[303]**
Results	*malignant n= 200* *benign n= 200* *Overall accuracy of US was stratified for nodule size. It was also conducted by 3 separate observers, testing the same people. The data from observer 1 only is included in this review.* ***Nodule size <5mm [malignant n=47, benign n=43]*** *Index test 1: Nodules showing one or more suspicious features from marked hypoechogenicity, microlobulated or spiculated margins, taller than wide, and microcalcifications (+ve for malignancy) [no suspicious features was -ve for malignancy]* TP: 41 FN: 6 FP: 31 TN: 12 *sensitivity*: 0.872 *specificity*: 0.279 ***Nodule size 5mm to <10mm [malignant n=50, benign n=53]*** *Index test 1: Nodules showing one or more suspicious features from marked hypoechogenicity, microlobulated or spiculated margins, taller than wide, and microcalcifications (+ve for malignancy) [no suspicious features was -ve for malignancy]* TP: 42 FN: 8 FP: 22 TN: 31 *sensitivity*: 0.840 *specificity*: 0.585 ***Nodule size 10mm to <20mm [malignant n=52, benign n=49]*** *Index test 1: Nodules showing one or more suspicious features from marked hypoechogenicity, microlobulated or spiculated margins, taller than wide, and microcalcifications (+ve for malignancy) [no suspicious features was -ve for malignancy]* TP: 37 FN: 15 FP: 13 TN: 36 *sensitivity*: 0.712 *specificity*: 0.735 ***Nodule size 20mm and higher [malignant n=51, benign n=55]*** *Index test 1: Nodules showing one or more suspicious features from marked hypoechogenicity, microlobulated or spiculated margins, taller than wide, and microcalcifications (+ve for malignancy) [no suspicious features was -ve for malignancy]* TP: 37 FN: 14 FP: 3 TN: 52 *sensitivity*: 0.725 *specificity*: 0.945
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may

Reference	**Park, 2012[303]**
	be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Bora Makal, 2021[35]**
Study type	Retrospective
Number of patients	n = 141
Patient characteristics	*Age, mean (SD): 47(12.7)* *Gender (female to male ratio): 90:51* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: Secondary care *Country*: Turkey *Inclusion criteria*: Patients undergoing thyroidectomy *Exclusion criteria*: patients without nodules; undergoing surgery because of hyperthyroidism, autoimmune thyroiditis or metastatic thyroid cancer; having previous neck radiation or surgery; not also being evaluated with SWE
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using unreported MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Bora Makal, 2021[35]**
Results	*malignant n=54* *benign n= 87* *Index test 1: ACR TI-RADS 3 or more (+ve for malignancy) [2 or less denotes -ve for malignancy]* TP: 54 FN: 0 FP: 75 TN: 12 *sensitivity*: 1.0 *specificity*: 0.138 *Index test 2: ACR TI-RADS 4 or more (+ve for malignancy) [3 or less denotes -ve for malignancy]* TP: 45 FN: 9 FP: 54 TN: 33 *sensitivity*: 0.83 *specificity*: 0.379 *Index test 3: ACR TI-RADS 5 (+ve for malignancy) [4 or less denotes -ve for malignancy]* TP: 24 FN: 30 FP: 15 TN: 72 *sensitivity*: 0.444 *specificity*: 0.828 *Index test 4: elastography SWE at 5 m/s or higher (+ve for malignancy) [based on ROC curve analysis]* TP: 18 FN: 36 FP: 0 TN: 87 *sensitivity*: 0.333 *specificity*: 1.0
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Huang, 2019[154]**
Study type	Retrospective
Number of patients	n = 109
Patient characteristics	*Age, mean (SD): 35.4(2.3)* *Gender (female to male ratio): 80:29* *Ethnicity*: not reported

Reference	Huang, 2019[154]
	Expertise of US tester (medic/non medic/unknown): unknown *Setting*: University Hospital *Country*: China *Inclusion criteria*: Patients undergoing surgical resection of single thyroid nodules at or under 1cm *Exclusion criteria*: pregnancy; nodules with predominantly cystic or coarse calcification
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Colour doppler ultrasound, using 5-12 MHz probe frequency Contrast enhanced US (CEUS) *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=77 (PTMC)* *benign n= 32 (nodular goitre)* *Index test 1: Colour doppler US TIRADS 4a,4b or 4c (+ve for malignancy) [3 or lower was taken as -ve for malignancy]* TP: 77 FN: 0 FP: 31 TN: 1 *sensitivity*: 1.0 *specificity*: 0.031 *Index test 2: Colour doppler US TIRADS 4b or 4c (+ve for malignancy) [4a or lower was taken as -ve for malignancy]* TP: 54 FN: 23 FP: 10 TN: 22 *sensitivity*: 0.701 *specificity*: 0.688 *Index test 3: Colour doppler US TIRADS 4c (+ve for malignancy) [4b or lower was taken as -ve for malignancy]* TP: 14 FN: 63 FP: 0 TN: 32 *sensitivity*: 0.182 *specificity*: 1.0

Reference	**Huang, 2019[154]**
	Index test 4: CEUS heterogeneous enhancement or hypo-enhancement (+ve for malignancy) [uniform hyper-enhancement or uniform iso-enhancement was taken as -ve for malignancy] TP: 55 FN: 22 FP: 12 TN: 20 *sensitivity*: 0.714 *specificity*: 0.625
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Ragazzoni, 2012[315]**
Study type	Prospective
Number of patients	n = 115 patients with 132 nodules
Patient characteristics	*Age, mean (SD): 54(13.37)* *Gender (female to male ratio): 92:23* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown (at least 5 years' experience but unclear if a medic) *Setting*: Thyroid surgery unit *Country*: Italy *Inclusion criteria*: presence of single or multiple thyroid nodules clearly distinguishable from surrounding parenchyma. *Exclusion criteria*: cystic nodules and mixed nodules with liquid component >30% of the nodule total volume; nodules with eggshell calcifications (but internal calcifications NOT excluded)
Target condition(s)	Thyroid nodule malignancy

Reference	**Ragazzoni, 2012[315]**
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 7.5-13 MHz probe frequency and US power doppler; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=40* *benign n= 92* *Index test 1: hypoechogenicity (+ve for malignancy)* TP: 35 FN: 5 FP: 39 TN: 53 *sensitivity*: 0.875 *specificity*: 0.576 *Index test 1: microcalcifications (+ve for malignancy)* TP: 22 FN: 18 FP: 4 TN: 88 *sensitivity*: 0.55 *specificity*: 0.956 *Index test 1: halo sign (+ve for malignancy)* TP: 25 FN: 15 FP: 12 TN: 80 *sensitivity*: 0.869 *specificity*: 0.792 *Index test 1: irregular margins (+ve for malignancy)* TP: 20 FN: 20 FP: 9 TN: 83 *sensitivity*: 0.500 *specificity*: 0.902 *Index test 1: type III vascularisation (+ve for malignancy) [type II and below denotes -ve]* TP: 25 FN: 15 FP: 33 TN: 59 *sensitivity*: 0.625 *specificity*: 0.641 *Index test 1: A/T ratio >1 (+ve for malignancy)* TP: 12 FN: 28 FP: 22 TN: 70 *sensitivity*: 0.30 *specificity*: 0.761 *Index test 1: elastography – Asteria 1-4 colour score: 3 or more (+ve for malignancy)* TP: 34 FN: 6 FP: 15 TN: 77 *sensitivity*: 0.85 *specificity*: 0.837
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious risk of bias

Reference	**Ragazzoni, 2012[315]**
	Indirectness (QUADAS 2 - applicability): none
Comments	

1

Reference	**Schenke, 2018[336]**
Study type	Retrospective
Number of patients	n = 194 with 244 nodules
Patient characteristics	*Age, mean (range): 47(17-76)* *Gender (female to male ratio): 150:44* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown (sonographers with >5 yrs experience but unclear if medics) *Setting*: unclear *Country*: Germany *Inclusion criteria*: Patients who underwent thyroidectomy and sonoelastography and conventional US *Exclusion criteria*: hyperfunctioning thyroid nodules in scintigraphy, incidental papillary microcarcinomas, sonoelastographies stored without colours and sonoelastographies with too high or too low levels at the quality indicator scale
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 5-10 MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Schenke, 2018[336]**
Results	*malignant n=38* *benign n= 206* *Index test 1: TIRADS 3 or greater (+ve for malignancy) [2 or less denotes -ve result for malignancy]* TP: 38 FN: 0 FP: 201 TN: 5 *sensitivity*: 1.0 *specificity*: 0.024 *Index test 2: TIRADS 4A or greater (+ve for malignancy) [3 or less denotes -ve result for malignancy]* TP: 37 FN: 1 FP: 155 TN: 51 *sensitivity*: 0.974 *specificity*: 0.248 *Index test 3: TIRADS 4B or greater (+ve for malignancy) [4A or less denotes -ve result for malignancy]* TP: 35 FN: 3 FP: 86 TN: 120 *sensitivity*: 0.921 *specificity*: 0.583 *Index test 4: TIRADS 4C or greater (+ve for malignancy) [4B or less denotes -ve result for malignancy]* TP: 31 FN: 7 FP: 25 TN: 181 *sensitivity*: 0.816 *specificity*: 0.878 *Index test 5: TIRADS 5 (+ve for malignancy) [4C or less denotes -ve result for malignancy]* TP: 9 FN: 29 FP: 0 TN: 206 *sensitivity*: 0.237 *specificity*: 1.0 *Index test 6: Shear wave velocity of 0.225 or higher (+ve for malignancy)* TP: 20 FN: 18 FP: 50 TN: 146 *sensitivity*: 0.529 *specificity*: 0.743 *Index test 6: Elastography – Rago 1-3 scale: 2 or higher (+ve for malignancy)* TP: 31 FN: 7 FP: 111 TN: 85 *sensitivity*: 0.816 *specificity*: 0.434 *Index test 6: Elastography – Asteria1-4 colour scale: 3 or higher (+ve for malignancy)* TP: 31 FN: 7 FP: 85 TN: 111 *sensitivity*: 0.816 *specificity*: 0.566
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Li, 2017[217]**
Study type	Retrospective but unclear
Number of patients	n = 89
Patient characteristics	*Age, mean (SD): 43.2(1.8)* *Gender (female to male ratio): 68:21* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown (senior ultrasonologist with >20 years' experience but unclear if medic) *Setting*: University Hospital *Country*: China *Inclusion criteria*: patients undergoing US and CEUS, but overall unclear *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy (PTC)
Index test(s) and reference standard	*Index test* Grey scale ultrasound and CEUS, using unreported MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Li, 2017[217]**
Results	*malignant n=56* *benign n= 33* *Index test 1: Conventional US:* low-level echo, 'vertical/horizontal ≥1', fuzzy boundary, microcalcification and grade IV blood flow - TIRADS grades 4 and 5 *(+ve for malignancy)* TP: 49 FN: 7 FP: 7 TN: 26 *sensitivity*: 0.875 *specificity*: 0.788 *Index test 2: CEUS: low enhancement, weak enhancement, late enhancement and uneven enhancement (+ve for malignancy)* TP: 52 FN: 4 FP: 4 TN: 29 *sensitivity*: 0.929 *specificity*: 0.879 *Index test 2: Combination of US and CEUS (unclear how combined)* TP: 54 FN: 2 FP: 2 TN: 31 *sensitivity*: 0.967 *specificity*: 0.927
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Aslan, 2018[21]**
Study type	Prospective
Number of patients	n = 86 patients with 140 nodules
Patient characteristics	*Age, mean (SD): unclear* *Gender (female to male ratio): 69:19* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown (10 years' experience in US but unknown if medic) *Setting*: University Hospital

Reference	Aslan, 2018[21]
	Country: Turkey *Inclusion criteria*: All consecutive patients scheduled for subtotal or total thyroidectomies due to multinodular goitre, or malignant or suspicious FNA results; nodular thyroid disease confirmed by grey scale US; co-operated with Duplex Doppler US (DDUS) evaluation; histopathological diagnosis of PTC or benign nodule *Exclusion criteria*: patients who could not co-operate with operator during DDUS exam; nodules intertwined or difficult to depict alone; nodules without vascularity on colour doppler US (CDUS); nodule without histopathological diagnosis; malignant tumour other than PTC.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Colour doppler and Duplex Doppler ultrasound, using multi-frequency linear array probe *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Within 1 week
Results	*malignant n=30 (PTC)* *benign n= 110* *Index test 1: systolic /diastolic ratio >3.11 (+ve for malignancy)* TP: 18 FN: 12 FP: 19 TN: 91 *sensitivity*: 0.593 *specificity*: 0.827 *Index test 2: pulsatility index >0.92 (+ve for malignancy)* TP: 24 FN: 5 FP: 49 TN: 61 *sensitivity*: 0.815 *specificity*: 0.555 *Index test 3: resistive index >0.68 (+ve for malignancy)* TP: 17 FN: 13 FP: 19 TN: 91 *sensitivity*: 0.556 *specificity*: 0.827

Reference	**Aslan, 2018[21]**
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Tuan,2020[381]**
Study type	Prospective
Number of patients	n = 84 patients, with 94 nodules
Patient characteristics	*Age, mean (SD): 46.94(12.69)* *Gender (female to male ratio): 77:9* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic (radiologist) *Setting*: University Hospital *Country*: Vietnam *Inclusion criteria*: patients undergoing thyroidectomy, US and SWE; nodules 10mm or more in widest dimension *Exclusion criteria*: predominantly cystic nodules; inability to obtain SWE registration
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 4-15 MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:*

Reference	**Tuan,2020[381]**
	Not clear
Results	*malignant n=74* *benign n= 20* *Index test 1: taller than wide shape (+ve for malignancy)* TP: 34 FN: 40 FP: 3 TN: 17 *sensitivity*: 0.459 *specificity*: 0.85 *Index test 2: irregular margin (+ve for malignancy)* TP: 51 FN: 23 FP: 6 TN: 14 *sensitivity*: 0.689 *specificity*: 0.70 *Index test 3: marked hypoechogenicity (+ve for malignancy)* TP: 23 FN: 51 FP: 1 TN: 19 *sensitivity*: 0.311 *specificity*: 0.95 *Index test 4: microcalcification (+ve for malignancy)* TP: 49 FN: 25 FP: 2 TN: 18 *sensitivity*: 0.662 *specificity*: 0.90 *Index test 5: intranodular vascularity (+ve for malignancy)* TP: 41 FN: 33 FP: 12 TN: 8 *sensitivity*: 0.554 *specificity*: 0.40 *Index test 6: Combination test: at least ONE* of taller than wide, irregular, marked hypoechogenicity and microcalcification (+ve for malignancy)* TP: 71 FN: 3 FP: 10 TN: 10 *sensitivity*: 0.959 *specificity*: 0.50 *this is assumed as this is not made clear in the paper *Index test 7: SWE of 74.5 kpa or higher (+ve for malignancy)* TP: 55 FN: 19 FP: 2 TN: 18 *sensitivity*: 0.743 *specificity*: 0.90
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Sodagari, 2018[356]**
Study type	Prospective
Number of patients	n = 114 patients
Patient characteristics	*Age, mean (SD): 43.13(13)* *Gender (female to male ratio): 90:51* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic (head and neck radiologist) *Setting*: Endocrine surgery clinic *Country*: Tehran *Inclusion criteria*: All patients consecutively referred to the endocrine surgery clinic for thyroidectomy after being diagnosed with a cold thyroid nodule; positive family history of thyroid neoplasm *Exclusion criteria*: Not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 5-7.5 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Sodagari, 2018[356]**
Results	*malignant n=45* *benign n= 69* *Index test 1: hypoechogenicity (+ve for malignancy) [hyper-or iso-echoic were deemed -ve for malignancy]* TP: 39 FN: 6 FP: 26 TN: 43 *sensitivity*: 0.86 *specificity*: 0.623 *Index test 2: nodule shape: oval (+ve for malignancy) [round deemed -ve for malignancy]* TP: 33 FN: 12 FP: 37 TN: 31 *sensitivity*: 0.73 *specificity*: 0.45 *Index test 3: nodule size ≤36mm (+ve for malignancy)* TP: 37 FN: 8 FP: 37 TN: 32 *sensitivity*: 0.822 *specificity*: 0.471 *Index test 4: calcification: positive (+ve for malignancy) [negative deemed -ve for malignancy]* TP: 26 FN: 19 FP: 11 TN: 58 *sensitivity*: 0.578 *specificity*: 0.841 *Index test 5: margin: ill-defined (+ve for malignancy) [well-defined deemed -ve for malignancy]* TP: 23 FN: 22 FP: 14 TN: 55 *sensitivity*: 0.511 *specificity*: 0.797 *Index test 6: pattern: heterogenic (+ve for malignancy) [homogeneous deemed -ve for malignancy]* TP: 37 FN: 8 FP: 41 TN: 28 *sensitivity*: 0.822 *specificity*: 0.406 *Index test 7: halo: thick or incomplete (+ve for malignancy) [homogeneous deemed -ve for malignancy]* TP: 36 FN: 9 FP: 17 TN: 52 *sensitivity*: 0.795 *specificity*: 0.754
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

Reference	**Ma, 2017[255]**
Study type	Retrospective, consecutive patient enrolment
Number of patients	n = 125 patients with 135 nodules
Patient characteristics	*Age, mean (SD): 48.55(12.03)* *Gender (female to male ratio): 104:31* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): tester unclear, but reviewed by a medic *Setting*: Secondary care *Country*: China *Inclusion criteria*: maximum nodule diameter <1cm; age 18 or over; surgery performed; histopathological diagnosis available *Exclusion criteria*: pregnancy or lactation; not enough thyroid parenchyma surrounding the nodule; dominant cystic (>75%) nodules; NYHA grade III-IV; severe pulmonary hypertension; severe allergies
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale and contrast enhanced ultrasound, using 3 scanners with 5-12, 3-9 and 6-13MHz probe frequencies; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	Ma, 2017[255]
Results	*malignant n=79 (PTMC)* *benign n= 56* *Index test 1: US: single (+ve for malignancy) [multiple deemed -ve for malignancy]* TP: 47 FN: 32 FP: 31 TN: 25 *sensitivity*: 0.595 *specificity*: 0.446 *Index test 2: US: taller than wide (+ve for malignancy) [wider than tall deemed -ve for malignancy]* TP: 64 FN: 15 FP: 13 TN: 43 *sensitivity*: 0.810 *specificity*: 0.768 *Index test 3: US: marked hypoechogenicity (+ve for malignancy) [hyper-, hypo- and iso-echoic deemed -ve for malignancy]* TP: 23 FN: 56 FP: 6 TN: 50 *sensitivity*: 0.291 *specificity*: 0.893 *Index test 4: US: heterogeneous echo texture (+ve for malignancy) [homogeneous deemed -ve for malignancy]* TP: 62 FN: 17 FP: 44 TN: 12 *sensitivity*: 0.785 *specificity*: 0.214 *Index test 5: US: poorly defined margins (+ve for malignancy) [well-defined deemed -ve for malignancy]* TP: 60 FN: 19 FP: 22 TN: 34 *sensitivity*: 0.759 *specificity*: 0.607 *Index test 6: US: microcalcification (+ve for malignancy) [macro or none deemed -ve for malignancy]* TP: 47 FN: 32 FP: 15 TN: 41 *sensitivity*: 0.595 *specificity*: 0.732 *Index test 7: US: halo absent (+ve for malignancy) [present deemed -ve for malignancy]* TP: 78 FN: 1 FP: 43 TN: 13 *sensitivity*: 0.987 *specificity*: 0.232 *Index test 8: US: vascular distribution I or more (+ve for malignancy) [0 deemed -ve for malignancy]* TP: 64 FN: 15 FP: 43 TN: 13 *sensitivity*: 0.810 *specificity*: 0.232 *Index test 9: US: vascular distribution II or more (+ve for malignancy) [0, I deemed -ve for malignancy]* TP: 34 FN: 45 FP: 22 TN: 34 *sensitivity*: 0.430 *specificity*: 0.607 *Index test 10: US: vascular distribution III (+ve for malignancy) [0, I and II deemed -ve for malignancy]* TP: 13 FN: 66 FP: 12 TN: 44 *sensitivity*: 0.164 *specificity*: 0.786 *Index test 11: CEUS: hypoenhancement pattern (+ve for malignancy) [hyper- or iso-enhancement deemed -ve for malignancy]* TP: 62 FN: 17 FP: 25 TN: 31 *sensitivity*: 0.785 *specificity*: 0.556

Reference	**Ma, 2017[255]**
	Index test 11: CEUS: heterogeneous enhancement pattern (+ve for malignancy) [hyper- or iso-enhancement deemed -ve for malignancy] TP: 66 FN: 13 FP: 9 TN: 47 *sensitivity*: 0.835 *specificity*: 0.839 *Index test 11: CEUS: Incomplete or no ring enhancement pattern (+ve for malignancy) [complete ring enhancement pattern deemed -ve for malignancy]* TP: 78 FN: 1 FP: 49 TN: 7 *sensitivity*: 0.987 *specificity*: 0.125 *Index test 12: elastography- Rago 1-5 scale: 4 or more (+ve for malignancy* TP: 57 FN: 22 FP: 16 TN: 40 *sensitivity*: 0.725 *specificity*: 0.714 *Index test 12: elastography- strain ratio of 3.5 or more (+ve for malignancy* TP: 57 FN: 22 FP: 9 TN: 47 *sensitivity*: 0.725 *specificity*: 0.839
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Kim, 2012[180]**
Study type	Prospective
Number of patients	n = 505
Patient characteristics	*Age, mean (SD): unclear for the 505 with surgical histopathological GS* *Gender (female to male ratio): unclear* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: University hospital

Reference	**Kim, 2012[180]**
	Country: South Korea *Inclusion criteria*: Not clearly reported *Exclusion criteria*: Not clearly reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 5-12MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=314* *benign n= 191* *Index test 1:SN-US class II and above (+ve for malignancy) [SN-US class I deemed -ve for malignancy]* TP:300 FN:14 FP: 91 TN: 100 *sensitivity*: 0.955 *specificity*: 0.523 *Index test 2:SN-US class III and above (+ve for malignancy) [SN-US class II and below deemed -ve for malignancy]* TP:271 FN:43 FP: 33 TN: 158 *sensitivity*: 0.863 *specificity*: 0.827 *Index test 3:SN-US class IV and above (+ve for malignancy) [SN-US class III and below deemed -ve for malignancy]* TP:242 FN:72 FP: 18 TN: 173 *sensitivity*: 0.771 *specificity*: 0.906 *Index test 4:SN-US class V (+ve for malignancy) [SN-US class IV and below deemed -ve for malignancy]* TP:178 FN:136 FP: 2 TN: 189 *sensitivity*: 0.567 *specificity*: 0.989
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: None

Reference	**Kim, 2012[180]**
Comments	

1

Reference	**Reverter, 2019[322]**
Study type	Retrospective
Number of patients	n = 300 patients with 300 nodules
Patient characteristics	*Age, mean (SD): 55(11)* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): tester unclear but reviewed by a medic *Setting*: Secondary care *Country*: Spain *Inclusion criteria*: Age >18 at time of diagnosis; total or nearly total thyroidectomy or lobectomy; preoperative US evaluation of thyroid nodules; available pathologic findings *Exclusion criteria*: Not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 5-15 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Reverter, 2019[322]**
Results	*malignant n= 135* *benign n= 165* *Index test 1: Clinical reporting using ATA system – at least one suspicious US sign (+ve for malignancy)* TP: 117 FN: 18 FP: 15 TN: 150 *sensitivity*: 0.870 *specificity*: 0.912
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Abd_Alrahman, 2017[3]**
Study type	Prospective
Number of patients	n = 30 nodules
Patient characteristics	*Age, mean (SD): 34.5(10.01)[females] and 46.5(10.7) [males]* *Gender (female to male ratio): 24:6* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): Unknown *Setting*: Unclear *Country*: Not reported *Inclusion criteria*: Patients with a solitary thyroid nodule for whom final diagnosis was available through surgical histopathology; US and elastography performed

Reference	**Abd_Alrahman, 2017[3]**
	Exclusion criteria: complex nodules with >30% cystic component; large nodules occupying >75% of the thyroid lobe with little or no discernible normal thyroid tissue ; nodules with egg shell or completely calcified nodule, patients with abnormal US texture of the thyroid tissue and patients with previous thyroid surgery or previous radiation to the head and neck region
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using unreported (high) MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=7* *benign n= 23* *Index test 1: hypoechoicity (+ve for malignancy) [hyper- or iso-echoic was deemed -ve for malignancy]* TP: 6 FN: 1 FP: 4 TN: 19 *sensitivity*: 0.857 *specificity*: 0.826 *Index test 2: taller than wide (+ve for malignancy) [wider than tall was deemed -ve for malignancy]* TP: 5 FN: 2 FP: 1 TN: 22 *sensitivity*: 0.714 *specificity*: 0.957 *Index test 3: fine calcification (+ve for malignancy) [none or coarse was deemed -ve for malignancy]* TP: 4 FN: 3 FP: 0 TN: 23 *sensitivity*: 0.571 *specificity*: 1.0 *Index test 4: irregular margins (+ve for malignancy) [regular was deemed -ve for malignancy]* TP: 6 FN: 1 FP: 2 TN: 21 *sensitivity*: 0.857 *specificity*: 0.913 *Index test 5: uncontinuous halo (+ve for malignancy) [continuous and no halo was deemed -ve for malignancy]* TP: 5 FN: 2 FP: 1 TN: 22 *sensitivity*: 0.714 *specificity*: 0.957 *Index test 6: central vascularity (+ve for malignancy) [peripheral and no vascularity was deemed -ve for malignancy]* TP: 4 FN: 3 FP: 1 TN: 22 *sensitivity*: 0.571 *specificity*: 0.957

Reference	**Abd_Alrahman, 2017[3]**
	Index test 7: TIRADS 3 and above (+ve for malignancy) [2 and below was deemed -ve for malignancy] TP: 7 FN: 0 FP: 7 TN: 16 *sensitivity*: 1.0 *specificity*: 0.696 *Index test 8: TIRADS 4b and above (+ve for malignancy) [3 and below was deemed -ve for malignancy]* TP: 7 FN: 0 FP: 1 TN: 22 *sensitivity*: 1.0 *specificity*: 0.957 *Index test 9: TIRADS 4c (+ve for malignancy) [4b and below was deemed -ve for malignancy]* TP: 4 FN: 3 FP: 0 TN: 23 *sensitivity*: 0.571 *specificity*: 1.0 *Index test 10: elastography – Asteria 1-4 colour score: 2 or more (+ve for malignancy)* TP: 7 FN: 0 FP: 19 TN: 4 *sensitivity*: 1.0 *specificity*: 0.174 *Index test 11: elastography – Asteria 1-4 colour score: 3 or more (+ve for malignancy)* TP: 6 FN: 1 FP: 4 TN: 19 *sensitivity*: 0.857 *specificity*: 0.826 *Index test 12: elastography – Asteria 1-4 colour score: 4 (+ve for malignancy)* TP: 3 FN: 4 FP: 0 TN: 23 *sensitivity*: 0.429 *specificity*: 1.0 *Index test 13: elastography – strain ratio of 2.88 or more (+ve for malignancy)* TP: 6 FN: 1 FP: 2 TN: 21 *sensitivity*: 0.857 *specificity*: 0.932
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Wu, 2016[404]**
Study type	Retrospective
Number of patients	n = 970 dominant nodules in 970 patients
Patient characteristics	*Age, mean (SD): 46.71(12.14)* *Gender (female to male ratio): 756:214*

Reference	**Wu, 2016[404]**
	Ethnicity: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: secondary care *Country*: China *Inclusion criteria*: patients undergoing partial or total thyroidectomy because of suspicious US findings or suspicious FNA findings *Exclusion criteria*: Not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 5-12 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=507 [PTC n=487, FTC n=12, MTC n=4, WDC n=3, CCC n=1)* *benign n= 463* *Two radiologists were used in this study, but only the values from the more experienced one are reported here.* *Index test 1: ill-defined margin (+ve for malignancy)* TP: 465 FN: 42 FP: 265 TN: 198 *sensitivity*: 0.917 *specificity*: 0.428 *Index test 2: taller than wide (+ve for malignancy)* TP: 198 FN: 309 FP: 48 TN: 415 *sensitivity*: 0.391 *specificity*: 0.897 *Index test 3: solidity (+ve for malignancy)* TP: 489 FN: 18 FP: 314 TN: 149 *sensitivity*: 0.965 *specificity*: 0.322

Reference	**Wu, 2016[404]**
	Index test 4: microcalcifications (+ve for malignancy) TP: 225 FN: 282 FP: 40 TN: 423 *sensitivity*: 0.444 *specificity*: 0.913
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Tan, 2010[375]**
Study type	Retrospective, consecutive patient enrolment
Number of patients	n = 112
Patient characteristics	*Age, mean (SD): 49(13)* *Gender (female to male ratio): 97:15* *Ethnicity*: 69% Chinese, 14% Malay, 5% Indian, 12% other ethnic groups. *Expertise of US tester (medic/non medic/unknown*): Unknown *Setting*: Tertiary care *Country*: Singapore *Inclusion criteria*: all patients undergoing thyroid surgery and diagnostic evaluation in the research hospital *Exclusion criteria*: Diagnostic evaluation in other hospitals
Target condition(s)	Thyroid nodule malignancy

Reference	**Tan, 2010[375]**
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=31* *benign n= 81* *Index test 1: US signs including any of hypoechoic lesions, ill-defined edges and microcalcifications (+ve for malignancy)* TP: 13 FN: 28 FP: 13 TN: 68 *sensitivity*:0.414 *specificity*: 0.839
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Phuttharak, 2009[311]**
Study type	Prospective
Number of patients	n = 31
Patient characteristics	*Age, mean (SD): 41.8(10.8)* *Gender (female to male ratio): 30:1* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): tester unclear but reviewed by a medic (radiologist)

Reference	**Phuttharak, 2009[311]**
	Setting: Secondary care *Country*: Thailand *Inclusion criteria*: Patients with thyroid nodules presenting to ENT department; after FNA and clinical evaluation those thought to be at risk of thyroid malignancy; received US and thyroidectomy *Exclusion criteria*: Not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 10 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* 1 day
Results	*malignant n=5* *benign n= 26* *Index test 1: predominantly solid (+ve for malignancy) [cystic or spongiform deemed -ve for malignancy]* TP: 5 FN: 0 FP: 18 TN: 8 *sensitivity*: 1.0 *specificity*: 0.308 *Index test 2: taller than wide (+ve for malignancy) [wider than tall deemed -ve for malignancy]* TP: 0 FN: 5 FP: 1 TN: 25 *sensitivity*:0.0 *specificity*: 0.962 *Index test 3: indistinct margin (+ve for malignancy) [well-defined deemed -ve for malignancy]* TP: 3 FN: 2 FP: 3 TN: 23 *sensitivity*:0.6 *specificity*: 0.885 *Index test 4: marked hypoechogenicity (+ve for malignancy) [hypo-, iso-, or hyper-echoic deemed -ve for malignancy]* TP: 2 FN: 3 FP: 11 TN: 15 *sensitivity*: 0.4 *specificity*: 0.577

Reference	**Phuttharak, 2009[311]**
	Index test 5: microcalcification (+ve for malignancy) [macro- or absence deemed -ve for malignancy] TP: 3 FN: 2 FP: 0 TN: 26 *sensitivity*: 0.6 *specificity*: 1.0 *Index test 6: absence of halo sign (+ve for malignancy) [presence deemed -ve for malignancy]* TP: 3 FN: 2 FP: 0 TN: 26 *sensitivity*: 0.6 *specificity*: 1.0 *Index test 7: invading adjacent structure (+ve for malignancy) [not invading deemed -ve for malignancy]* TP: 1 FN: 4 FP: 0 TN: 26 *sensitivity*: 0.2 *specificity*: 1.0 *Index test 8: perinodular flow or marked intrinsic hypervascularity (+ve for malignancy) [not invading deemed -ve for malignancy]* TP: 5 FN: 5 FP: 26 TN: 0 *sensitivity*: 1.0 *specificity*: 0.0
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious risk of bias *Indirectness (QUADAS 2 - applicability)*: None
Comments	

1

Reference	**Chen, 2019[57]**
Study type	Retrospective
Number of patients	n = 131 nodules in 120 patients
Patient characteristics	*Age, mean: not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown (>8 years' experience but unclear if medic) *Setting*: University Hospital *Country*: China *Inclusion criteria*: thyroid lesions confirmed by surgery and pathology

Reference	**Chen, 2019[57]**
	Exclusion criteria: nodules of >2cm diameter and <0.3cm diameter; Hashimoto's thyroiditis; cystic dominated thyroiditis; cystic dominated nodules; calcified nodules
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 5-12 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=281 [PTCs]* *benign n= 272* *Index test 1: Ultrasound grey-scale (UGSR*) ratio ≥0.692 (+ve for malignancy) [<0.692 was deemed -ve for malignancy]* TP: 247 FN: 34 FP: 73 TN: 199 *sensitivity*: 0.879 *specificity*: 0.732 *UGSR=ratio of the grey scale of the thyroid nodules to the surrounding normal thyroid tissues under the same operating conditions
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Pathirana, 2016[305]**
Study type	Prospective

Reference	**Pathirana, 2016[305]**
Number of patients	n = 132 people with 189 nodules
Patient characteristics	*Age, mean (range): 43.8 (16-70)* *Gender (female to male ratio): 116:16* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic (radiologist) *Setting*: University Hospital *Country*: Sri Lanka *Inclusion criteria*: Patients scheduled for thyroidectomy *Exclusion criteria*: patients with goitres extending retrosternally or with diffuse goitres
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 6-8 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 28* *benign n= 161* *A score was designed based on US features as shown below. The variables and weighting were based, according to the authors of the paper, on current evidence.*

Reference	Pathirana, 2016[305]
	Proposed scoring system using ultrasonographic features of thyroid Size: Anteroposterior (AP) > width – score 2; AP = width – score 1; AP< width – score 0 Internal contents: solid – score 2; mixed – score 1; cystic thick wall – score 0 Echogenecity: hypoechoic – score 2; isoechoic – score 1, hyperechoic – score 0 Margins: Ill-defined – score 2; well defined – score 0 Calcification: microcalcification – score 2; large coarse calcification – score 1; no calcification – score 0 Halo: thick – score 2, thin – score 1; No – score 0 Vascularity: internal flow – score 2; peripheral flow – score 1; no flow – score 0 Total score 14 *Index test 1: score of 2 or more (+ve for malignancy) [-ve was a score of 1]* TP: 28 FN: 0 FP: 160 TN: 1 *sensitivity*: 1.0 *specificity*: 0.006 *Index test 2: score of 3 or more (+ve for malignancy) [-ve was a score of 2 or less]* TP: 28 FN: 0 FP: 156 TN: 5 *sensitivity*: 1.0 *specificity*: 0.031 *Index test 3: score of 4 or more (+ve for malignancy) [-ve was a score of 3 or less]* TP: 28 FN: 0 FP: 126 TN: 35 *sensitivity*: 1.0 *specificity*: 0.217 *Index test 4: score of 5 or more (+ve for malignancy) [-ve was a score of 4 or less]* TP: 28 FN: 0 FP: 89 TN: 72 *sensitivity*: 1.0 *specificity*: 0.447 *Index test 5: score of 6 or more (+ve for malignancy) [-ve was a score of 5 or less]* TP: 18 FN: 10 FP: 45 TN: 116 *sensitivity*: 0.643 *specificity*: 0.720 *Index test 6: score of 7 or more (+ve for malignancy) [-ve was a score of 6 or less]* TP: 15 FN: 13 FP: 27 TN: 134 *sensitivity*: 0.536 *specificity*: 0.832 *Index test 7: score of 8 or more (+ve for malignancy) [-ve was a score of 7 or less]* TP: 15 FN: 13 FP: 11 TN: 150 *sensitivity*: 0.536 *specificity*: 0.932 *Index test 8: score of 9 or more (+ve for malignancy) [-ve was a score of 8 or less]* TP: 12 FN: 16 FP: 4 TN: 157 *sensitivity*: 0.429 *specificity*: 0.975 *Index test 9: score of 10 or more (+ve for malignancy) [-ve was a score of 9 or less]*

Reference	**Pathirana, 2016[305]**
	TP: 12 FN: 16 FP: 0 TN: 161 *sensitivity*: 0.429 *specificity*: 1.0 *Index test 10: score of 11 or more (+ve for malignancy) [-ve was a score of 10 or less]* TP: 9 FN: 19 FP: 0 TN: 161 *sensitivity*: 0.321 *specificity*: 1.0 *Index test 11: score of 12 or more (+ve for malignancy) [-ve was a score of 11 or less]* TP: 2 FN: 26 FP: 0 TN: 161 *sensitivity*: 0.071 *specificity*: 1.0 *Index test 12: score of 13 or more (+ve for malignancy) [-ve was a score of 12 or less]* TP: 1 FN: 27 FP: 0 TN: 161 *sensitivity*: 0.036 *specificity*: 1.0 *Index test 13: score of 14 (+ve for malignancy) [-ve was a score of 13 or less]* TP: 0 FN: 28 FP: 0 TN: 161 *sensitivity*: 0.0 *specificity*: 1.0
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Li, 2016[223]**
Study type	Retrospective, consecutive patient enrolment
Number of patients	n = 762
Patient characteristics	*Age, mean: 47.9* *Gender (female to male ratio): 570:192* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown (experienced sonographist) *Setting*: Teaching Hospital *Country*: China

Reference	**Li, 2016[223]**
	Inclusion criteria: Patients with thyroid nodules who underwent surgery and had a pathological diagnosis *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale and colour doppler ultrasound, using MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 424* *benign n=338* *Index test 1: irregular margins (+ve for malignancy)* TP: 306 FN: 118 FP: 67 TN: 271 *sensitivity*: 0.721 *specificity*: 0.802 *Index test 2: hypoechogenicity (+ve for malignancy)* TP: 380 FN:44 FP: 147 TN: 191 *sensitivity*: 0.896 *specificity*: 0.565 *Index test 3: solidity (+ve for malignancy)* TP: 378 FN: 46 FP: 178 TN: 160 *sensitivity*:0.892 *specificity*: 0.473 *Index test 4: microcalcifications (+ve for malignancy)* TP: 175 FN: 249 FP: 30 TN: 308 *sensitivity*: 0.413 *specificity*: 0.911 *Index test 5: taller than wide (+ve for malignancy)* TP: 71 FN: 353 FP: 5 TN: 333 *sensitivity*: 0.167 *specificity*: 0.985 *Index test 6: intranodular vascularity (+ve for malignancy)* TP: 202 FN: 222 FP: 75 TN: 263 *sensitivity*: 0.476 *specificity*: 0.778

Reference	Li, 2016[223]
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	Borlea, 2020[36]
Study type	Prospective
Number of patients	n = 133 nodules in 133 patients
Patient characteristics	*Age, mean: 45.3* *Gender (female to male ratio): 96:37* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown (but at least 10 years' experience) *Setting*: Teaching Hospital *Country*: Romania *Inclusion criteria*: Patients presenting with solid thyroid nodules suitable for US assessment; surgical pathology report and US evaluation performed

Reference	**Borlea, 2020[36]**
	Exclusion criteria: Patients with previous thyroid surgery or radiation therapy or who presented with completely cystic nodules; autonomously functioning nodules
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale and colour doppler ultrasound, using 5-18 or 5-13MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=35* *benign n= 98* *Index test 1: blurred margins (+ve for malignancy)* TP: 13 FN: 22 FP: 28 TN: 70 *sensitivity*: 0.371 *specificity*: 0.714 *Index test 2: microcalcification (+ve for malignancy)* TP: 11 FN: 24 FP: 9 TN: 89 *sensitivity*: 0.314 *specificity*: 0.908 *Index test 3; marked hypoechoicity (+ve for malignancy)* TP: 9 FN: 26 FP: 3 TN: 95 *sensitivity*: 0.257 *specificity*: 0.969 *Index test 4: taller than wide (+ve for malignancy)* TP: 15 FN: 20 FP: 15 TN: 83 *sensitivity*: 0.429 *specificity*: 0.847 *Index test 5: 4D Doppler – intranodular vascularity / interrupted capsule (+ve for malignancy)* TP: 23 FN: 12 FP: 14 TN: 84 *sensitivity*: 0.657 *specificity*: 0.857 *Index test 6: EU TI-RADS 3 or more (+ve for malignancy)* TP: 35 FN: 0 FP: 92 TN: 6 *sensitivity*: 1.0 *specificity*: 0.061

Reference	Borlea, 2020[36]
	Index test 7: EU TI-RADS 4 or more (+ve for malignancy) TP: 34 FN: 1 FP: 75 TN: 23 *sensitivity*: 0.971 *specificity*: 0.235 *Index test 8: EU TI-RADS 5 (+ve for malignancy)* TP: 22 FN: 13 FP: 19 TN: 79 *sensitivity*: 0.629 *specificity*: 0.806 *Index test 9: ACR TI-RADS 2 or more (+ve for malignancy)* TP: 35 FN: 0 FP: 93 TN: 5 *sensitivity*: 1.0 *specificity*: 0.051 *Index test 10: ACR TI-RADS 4 or more (+ve for malignancy)* TP: 33 FN: 2 FP: 70 TN: 28 *sensitivity*: 0.943 *specificity*: 0.286 *Index test 11: ACR TI-RADS 5 (+ve for malignancy)* TP: 17 FN: 18 FP: 22 TN: 76 *sensitivity*: 0.486 *specificity*: 0.776 *Index test 12: Horvath TI-RADS 3 or more (+ve for malignancy)* TP: 34 FN: 1 FP: 80 TN: 18 *sensitivity*: 0.971 *specificity*: 0.184 *Index test 13: Horvath TI-RADS 4A or more (+ve for malignancy)* TP: 34 FN: 1 FP: 75 TN: 23 *sensitivity*: 0.971 *specificity*: 0.237 *Index test 14: Horvath TI-RADS 4B or more (+ve for malignancy)* TP: 28 FN: 7 FP: 37 TN: 61 *sensitivity*: 0.800 *specificity*: 0.622 *Index test 15: Horvath TI-RADS 5 (+ve for malignancy)* TP: 16 FN: 19 FP: 13 TN: 85 *sensitivity*: 0.457 *specificity*: 0.867 *Index test 16: French TI-RADS 3 or more (+ve for malignancy)* TP: 35 FN: 0 FP: 92 TN: 6 *sensitivity*: 1.0 *specificity*: 0.061 *Index test 17: French TI-RADS 4A or more (+ve for malignancy)* TP: 35 FN: 0 FP: 73 TN: 25 *sensitivity*: 1.0 *specificity*: 0.255 *Index test 18: French TI-RADS 4B or more (+ve for malignancy)* TP: 32 FN: 3 FP: 17 TN: 81 *sensitivity*: 0.914 *specificity*: 0.827 *Index test 19: French TI-RADS 5 (+ve for malignancy)*

Reference	**Borlea, 2020[36]**
	TP: 18 FN: 17 FP: 4 TN: 94 *sensitivity*: 0.514 *specificity*: 0.959 *Index test 20: French TI-RADS + 4D CD 3 or more (+ve for malignancy)* TP: 35 FN: 0 FP: 92 TN: 6 *sensitivity*: 1.0 *specificity*: 0.061 *Index test 21: French TI-RADS + 4D CD 4A or more (+ve for malignancy)* TP: 35 FN: 0 FP: 73 TN: 25 *sensitivity*: 1.0 *specificity*: 0.255 *Index test 22: French TI-RADS + 4D CD 4B or more (+ve for malignancy)* TP: 33 FN: 2 FP: 24 TN: 74 *sensitivity*: 0.943 *specificity*: 0.755 *Index test 23: French TI-RADS + 4D CD 5 (+ve for malignancy)* TP: 21 FN: 14 FP: 4 TN: 94 *sensitivity*: 0.600 *specificity*: 0.959 *Index test 23: elastography: strain ratio > 4 (+ve for malignancy)* TP: 28 FN: 7 FP: 12 TN: 86 *sensitivity*: 0.800 *specificity*: 0.878
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Mohamed, 2013[271]**
Study type	Prospective
Number of patients	n = 45
Patient characteristics	*Age, mean (SD): 46.9(11.2)* *Gender (female to male ratio): 33:12* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown

Reference	**Mohamed, 2013[271]**
	Setting: Radiodiagnosis department *Country*: Egypt *Inclusion criteria*: Patients with solitary solid thyroid nodules, referred for surgical treatment *Exclusion criteria*: Patients with multiple nodules, and nodules with extensive macrocalcifications
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 10-12 MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=28* *benign n= 17* *Index test 1: ill defined margins (+ve for malignancy)* TP: 24 FN: 4 FP: 3 TN: 14 *sensitivity*: 0.857 *specificity*: 0.824 *Index test 2: hypoechoicity (+ve for malignancy)* TP: 15 FN: 13 FP: 6 TN: 11 *sensitivity*: 0.536 *specificity*: 0.647 *Index test 3: Halo sign absent (+ve for malignancy)* TP: 24 FN: 4 FP: 6 TN: 11 *sensitivity*: 0.857 *specificity*: 0.647 *Index test 4: AP/T diameter >1cm (+ve for malignancy)* TP: 24 FN: 4 FP: 4 TN: 13 *sensitivity*: 0.857 *specificity*: 0.882 *Index test 5: spot microcalcifications (+ve for malignancy)*

Reference	**Mohamed, 2013[271]**
	TP: 19 FN: 9 FP: 3 TN: 14 *sensitivity*: 0.678 *specificity*: 0.824 *Index test 6: intranodular blood flow – type III and above (+ve for malignancy)* TP: 19 FN: 9 FP: 7 TN: 10 *sensitivity*: 0.678 *specificity*: 0.588 *Index test 7: elastography 1-6 score: 4 or more (+ve for malignancy)* TP: 25 FN: 3 FP: 2 TN: 15 *sensitivity*: 0.893 *specificity*: 0.882
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Mohey, 2013[275]**
Study type	Prospective
Number of patients	n = 46
Patient characteristics	*Age, mean: 41.7* *Gender (female to male ratio): 31:15* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: General Surgery department *Country*: Egypt *Inclusion criteria*: Patients with solid thyroid nodules *Exclusion criteria*: cystic thyroid nodules; calcified shell thyroid nodules; coalescent nodules in multi-nodular goitre

Reference	**Mohey, 2013[275]**
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 10 MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=15* *benign n= 31* *Index test 1: hypoechogenicity (+ve for malignancy)* TP: 12 FN: 3 FP:12 TN: 19 *sensitivity*: 0.80 *specificity*: 0.612 *Index test 2: absent halo sign (+ve for malignancy)* TP: 9 FN: 6 FP: 6 TN:25 *sensitivity*: 0.60 *specificity*: 0.806 *Index test 3: spot microcalcifications (+ve for malignancy)* TP: 10 FN: 5 FP: 9 TN: 22 *sensitivity*: 0.667 *specificity*: 0.709 *Index test 4: pattern of intranodular blood flow – type III (+ve for malignancy)* TP: 1 FN: 14 FP: 1 TN: 30 *sensitivity*: 0.067 *specificity*: 0.967 *Index test 5: absence of halo sign AND presence of spot microcalcifications (+ve for malignancy)* TP: 9 FN: 6 FP: 2 TN:29 *sensitivity*: 0.60 *specificity*: 0.935 *Index test 6: absence of halo sign AND hypoechogenicity (+ve for malignancy)* TP: 9 FN: 6 FP: 2 TN:29 *sensitivity*: 0.60 *specificity*: 0.935 *Index test 7: hypoechogenicity AND presence of spot microcalcifications (+ve for malignancy)* TP: 8 FN: 7 FP: 3 TN: 28 *sensitivity*: 0.533 *specificity*: 0.903 *Index test 8: absence of halo sign AND hypoechogenicity AND type II vascularisation(+ve for malignancy)* TP: 1 FN: 14 FP: 0 TN:31 *sensitivity*: 0.067 *specificity*: 1.0

Reference	**Mohey, 2013[275]**
	Index test 9: hypoechogenicity AND presence of spot microcalcifications AND type II vascularisation(+ve for malignancy) TP: 1 FN: 14 FP: 0 TN:31 *sensitivity*: 0.067 *specificity*: 1.0 *Index test 10: absence of halo sign AND presence of spot microcalcifications AND type II vascularisation(+ve for malignancy)* TP: 1 FN: 14 FP: 0 TN:31 *sensitivity*: 0.067 *specificity*: 1.0 *Index test 11: elastography – Rago 1-5 score: 3 and above (+ve for malignancy)* TP: 15 FN: 0 FP: 6 TN: 25 *sensitivity*: 1.0 *specificity*: 0.806 *Index test 12: elastography – Rago 1-5 score: 4 and above (+ve for malignancy)* TP: 14 FN: 1 FP: 0 TN: 31 *sensitivity*: 0.93 *specificity*: 1.0 *Index test 13: elastography – Rago 1-5 score; 5 (+ve for malignancy)* TP: 12 FN: 3 FP: 0 TN: 31 *sensitivity*: 0.8 *specificity*: 1.0
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Cakir, 2011[43]**
Study type	Prospective
Number of patients	n = 391 nodules in 292 patients
Patient characteristics	*Age, mean (SD): 46.08(11.89)* *Gender (female to male ratio): 242: 50* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic ('experienced specialist') *Setting*: out-patient thyroid clinic

Reference	**Cakir, 2011[43]**
	Country: Turkey *Inclusion criteria*: Patients admitted to outpatient thyroid clinic *Exclusion criteria*: Patients <15 years; previous history of thyroid surgery or percutaneous invasive procedures for thyroid nodules or radiotherapy to the head and neck region; cardiac or pulmonary disease; pure cystic nodules; haemorrhagic nodules, multinodular coalescent nodules with undetermined sizes; anaplastic carcinoma; Riedel thyroiditis; extensive cervical metastasis; patients refusing surgery.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 5.5 – 12.5 MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 125* *benign n= 266* *Index test 1: microcalcification (+ve for malignancy)* TP: 119 FN: 6 FP: 232 TN: 34 *sensitivity*:0.952 *specificity*: 0.128 *Index test 2: hypoechoicity (+ve for malignancy)* TP: 115 FN: 10 FP: 214 TN: 52 *sensitivity*:0.920 *specificity*: 0.196 *Index test 2: absent halo (+ve for malignancy)* TP: 57 FN: 68 FP: 70 TN: 195 *sensitivity*:0.456 *specificity*: 0.736 *Index test 3: elastography – Itoh 1-5 scale: 2 or more (+ve for malignancy)* TP: 122 FN: 3 FP: 239 TN: 27 *sensitivity*:0.976 *specificity*: 0.102

Reference	**Cakir, 2011[43]**
	Index test 4: elastography – Itoh 1-5 scale: 3 or more (+ve for malignancy) TP: 94 FN: 31 FP: 168 TN: 98 *sensitivity*: 0.752 *specificity*: 0.368 *Index test 5: elastography – Itoh 1-5 scale: 4 or more (+ve for malignancy)* TP: 73 FN: 52 FP: 77 TN: 189 *sensitivity*:0.585 *specificity*: 0.711 *Index test 6: elastography – Itoh 1-5 scale: 5 (+ve for malignancy)* TP: 23 FN: 102 FP: 16 TN: 250 *sensitivity*:0.184 *specificity*: 0.940
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Gao, 2019[104]**
Study type	Retrospective
Number of patients	n = 2544 nodules in 1758 patients
Patient characteristics	*Age, mean: 44.9* *Gender (female to male ratio): 1788:756* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): tester unclear but reviewed by a medic (radiologist) *Setting*: Teaching Hospital *Country*: China *Inclusion criteria*: patients with total or nearly total thyroidectomy or lobectomy; complete preoperative US of nodules; surgical pathology *Exclusion criteria*: non-mass-forming lesions and nodules that failed to meet the criteria for any pattern of ATA guidelines

Reference	**Gao, 2019[104]**
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using either 5-12 or 8-15 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 1681* *benign n= 863* *Index test 1: KWAK-TIRADS 3 or more (+ve for malignancy)* TP: 1681 FN: 0 FP: 855 TN: 8 *sensitivity*: 1.0 *specificity*: 0.009 *Index test 2: KWAK-TIRADS 4a or more (+ve for malignancy)* TP: 1673 FN: 8 FP: 452 TN: 411 *sensitivity*: 0.995 *specificity*: 0.476 *Index test 3: KWAK-TIRADS 4b or more (+ve for malignancy)* TP: 1657 FN: 24 FP: 329 TN: 534 *sensitivity*: 0.986 *specificity*: 0.619 *Index test 4: KWAK-TIRADS 4c or more (+ve for malignancy)* TP: 1502 FN: 179 FP: 203 TN: 660 *sensitivity*: 0.894 *specificity*: 0.765 *Index test 5: KWAK-TIRADS 5 or more (+ve for malignancy)* TP: 244 FN: 1437 FP: 36 TN: 827 *sensitivity*:0.145 *specificity*: 0.958 *Index test 6: ATA 'low suspicion' or higher (+ve for malignancy)* TP: 1681 FN: 0 FP: 679 TN: 184 *sensitivity*: 1.0 *specificity*: 0.213 *Index test 7: ATA ''intermediate suspicion' or higher (+ve for malignancy)* TP: 1661 FN: 20 FP: 340 TN: 523 *sensitivity*: 0.988 *specificity*: 0.606 ***Index test 8: ATA ''high suspicion' (+ve for malignancy)***

Reference	**Gao, 2019[104]**
	TP: 1606 FN: 75 FP: 223 TN: 640 *sensitivity*: 0.955 *specificity*: 0.742 *Index test 9: ACR TI-RADs TR2 or higher (+ve for malignancy)* TP: 1681 FN: 0 FP: 855 TN: 8 *sensitivity*: 1.0 *specificity*: 0.09 *Index test 10: ACR TI-RADs TR3 or higher (+ve for malignancy)* TP: 1679 FN: 2 FP: 711 TN: 152 *sensitivity*: 0.999 *specificity*: 0.176 *Index test 11: ACR TI-RADs TR4 or higher (+ve for malignancy)* TP: 1651 FN: 30 FP: 431 TN: 432 *sensitivity*: 0.982 *specificity*: 0.501 *Index test 9: ACR TI-RADs TR5 (+ve for malignancy)* TP: 1372 FN: 309 FP: 179 TN: 684 *sensitivity*: 0.816 *specificity*: 0.793
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Pagano, 2021[295]**
Study type	Retrospective
Number of patients	n = 146 patients
Patient characteristics	*Age, mean (SD): 50.5(14.8)* *Gender (female to male ratio): 111:35* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): tester unclear but reviewed by medics (radiologists and endocrinologists) *Setting*: Centre for FNA cytology

Reference	Pagano, 2021[295]
	Country: Italy *Inclusion criteria*: Patients referred for suspected thyroid nodules who underwent thyroid surgery and US *Exclusion criteria*: Patients undergoing surgery with a non-diagnostic cytology at FNA
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 7-14 MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=68* *benign n= 78* *Index test 1: solidity (+ve for malignancy) [cystic or spongiform deemed -ve for malignancy]* TP: 68 FN: 0 FP: 74 TN: 4 *sensitivity*: 1.0 *specificity*: 0.051 *Index test 2: markedly hypo-echoic (+ve for malignancy) [an-, iso-, hypo- and hyper- deemed -ve for malignancy]* TP: 26 FN: 42 FP: 5 TN: 73 *sensitivity*: 0.382 *specificity*: 0.936 *Index test 3: ill defined margins (+ve for malignancy) [well-defined deemed -ve for malignancy]* TP: 24 FN: 44 FP: 0 TN: 78 *sensitivity*: 0.353 *specificity*: 1.0 *Index test 4: microcalcifications (+ve for malignancy) [macro or absent deemed -ve for malignancy]* TP: 21 FN: 47 FP: 2 TN: 76 *sensitivity*: 0.309 *specificity*: 0.974 *Index test 5: taller than wide (+ve for malignancy) [parallel deemed -ve for malignancy]* TP: 2 FN: 66 FP: 0 TN: 78 *sensitivity*: 0.029 *specificity*: 1.0

Reference	Pagano, 2021[295]
	Index test 6: perinodular and peri-endonodular vascularisation (+ve for malignancy) [endonodular or unknown deemed -ve for malignancy] TP: 57 FN: 11 FP: 68 TN: 10 *sensitivity*: 1.0 *specificity*: 0.051 *Index test 7: size 1-10mm (+ve for malignancy) [11 to >20mm deemed -ve for malignancy]* TP: 16 FN: 52 FP: 3 TN: 75 *sensitivity*: 0.236 *specificity*: 0.961 *Index test 8: ATA 'low suspicion' or higher (+ve for malignancy)* TP: 68 FN: 0 FP: 74 TN: 4 *sensitivity*: 1.0 *specificity*: 0.051 *Index test 9: ATA 'intermediate suspicion' or higher (+ve for malignancy)* TP: 57 FN: 11 FP: 37 TN: 41 *sensitivity*: 0.838 *specificity*: 0.526 *Index test 10: ATA 'high suspicion' (+ve for malignancy)* TP: 37 FN: 31 FP: 2 TN: 76 *sensitivity*: 0.544 *specificity*: 0.974 *Index test 11: AACE/ACE/AME 2016 'Intermediate' or higher (+ve for malignancy)* TP: 68 FN: 0 FP: 72 TN: 6 *sensitivity*: 1.0 *specificity*: 0.077 *Index test 12: AACE/ACE/AME 2016 'high' (+ve for malignancy)* TP: 37 FN: 31 FP: 6 TN: 72 *sensitivity*: 0.544 *specificity*: 0.923 *Index test 13: EU TI-RADS 3 (low risk) or higher (+ve for malignancy)* TP: 68 FN: 0 FP: 74 TN: 4 *sensitivity*: 1.0 *specificity*: 0.051 *Index test 14: EU TI-RADS 4 (intermediate risk) or higher (+ve for malignancy)* TP: 57 FN: 11 FP: 37 TN: 41 *sensitivity*: 0.838 *specificity*: 0.526 *Index test 15: EU TI-RADS 5 (high risk) (+ve for malignancy)* TP: 40 FN: 28 FP: 6 TN: 72 *sensitivity*: 0.588 *specificity*: 0.923 *Index test 15: elastography – Rago 1-5 scale: 3 and above (+ve for malignancy)* TP: 38 FN: 30 FP: 21 TN: 57 *sensitivity*: 0.559 *specificity*: 0.731 *Index test 16: elastography – Rago 1-5 scale: 4 and above (+ve for malignancy)* TP: 24 FN: 44 FP: 8 TN: 70 *sensitivity*: 0.353 *specificity*: 0.897

Reference	**Pagano, 2021[295]**
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: No risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Swan, 2019[367]**
Study type	Prospective
Number of patients	n = 329 patients with 413 nodules
Patient characteristics	*Age, mean (SD): 55(13)* *Gender (female to male ratio): 251:78* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unclear – some by research assistant and some by medic *Setting*: Department of otorhinolaryngology *Country*: Denmark *Inclusion criteria*: >18 years; one or more thyroid nodules >10mm; indication for thyroid surgery providing histological specimens *Exclusion criteria*: Predominantly cystic nodules; insufficient solid area for SWE registration
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 4-15 MHz probe frequency; elastography *Reference (gold) standard:*

Reference	Swan, 2019[367]
	Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=79* *benign n= 325* *Index test 1: solidity (+ve for malignancy)* TP: 57 FN: 22 FP: 160 TN: 165 *sensitivity*:0.722 *specificity*: 0.508 *Index test 2: hypoechoic (+ve for malignancy)* TP: 68 FN: 11 FP: 185 TN: 140 *sensitivity*:0.861 *specificity*: 0.431 *Index test 2: heterogeneous echogenicity (+ve for malignancy)* TP: 64 FN: 15 FP: 223 TN: 102 *sensitivity*:0.810 *specificity*: 0.314 *Index test 3: microcalcifications (+ve for malignancy)* TP: 51 FN: 28 FP: 150 TN: 175 *sensitivity*:0.646 *specificity*: 0.538 *Index test 4: taller than wide (+ve for malignancy)* TP: 12 FN: 67 FP: 37 TN: 288 *sensitivity*:0.152 *specificity*: 0.886 *Index test 5: irregular margins (+ve for malignancy)* TP: 37 FN: 42 FP: 103 TN: 222 *sensitivity*:0.468 *specificity*: 0.683 *Index test 6: Doppler flow – perinodular or equal (+ve for malignancy)* TP: 63 FN: 16 FP: 261 TN: 64 *sensitivity*: 0.797 *specificity*: 0.197 *Index test 7: TIRADS 4 and above (+ve for malignancy)* TP: 71 FN: 8 FP: 250 TN: 71 *sensitivity*:0.899 *specificity*: 0.218 *Index test 8: elastography – Rago 1-3 score: 2 and above (+ve for malignancy)* TP: 26 FN: 53 FP: 81 TN: 244 *sensitivity*:0.33 *specificity*: 0.75 *Index test 9: elastography – elasticity index of 31kpa and above (+ve for malignancy)*

Reference	**Swan, 2019[367]**
	TP: 36 FN: 43 FP: 127 TN: 198 *sensitivity*:0.46 *specificity*: 0.61
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Ahmadi, 2019[10]**
Study type	Retrospective
Number of patients	n = 323 nodules from 213 adults
Patient characteristics	*Age, median: 55* *Gender (female to male ratio): 24.4:75.6* *Ethnicity*: Black 26.3%; White 63.4%; Other/not reported/declined 10.3% *Expertise of US tester (medic/non medic/unknown*): tester unclear but reviewed by a medic (high volume experienced) *Setting*: tertiary care *Country*: USA *Inclusion criteria*: adult patients with thyroid nodules >5mm undergoing thyroid surgery at a tertiary care hospital *Exclusion criteria*: Patients in whom pre-operative ultrasound imaging was not available, their ultrasound imaging was of poor quality, or their surgical pathology report was not available

Reference	**Ahmadi, 2019[10]**
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using unreported MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=88* *benign n= 235* *Index test 1: ACR TI-RADS 2 or higher (+ve for malignancy)* TP: 88 FN: 0 FP: 229 TN: 6 *sensitivity*: 1.0 *specificity*: 0.026 *Index test 2: ACR TI-RADS 3 or higher (+ve for malignancy)* TP: 85 FN: 3 FP: 206 TN: 29 *sensitivity*: 0.966 *specificity*: 0.123 *Index test 3: ACR TI-RADS 4 or higher (+ve for malignancy)* TP: 69 FN: 19 FP: 63 TN: 172 *sensitivity*: 0.784 *specificity*: 0.732 *Index test 4: ACR TI-RADS 5 (+ve for malignancy)* TP: 29 FN: 59 FP: 3 TN: 232 *sensitivity*: 0.329 *specificity*: 0.987 *Index test 5: ATA 'very low' or higher or higher (+ve for malignancy)* TP: 88 FN: 0 FP: 230 TN: 5 *sensitivity*: 1.0 *specificity*: 0.021 *Index test 5: ATA 'low' or higher or higher (+ve for malignancy)* TP: 88 FN: 1 FP: 215 TN: 20 *sensitivity*: 0.988 *specificity*: 0.085 *Index test 5: ATA 'intermediate' or higher or higher (+ve for malignancy)* TP: 68 FN: 20 FP: 55 TN: 180 *sensitivity*: 0.772 *specificity*: 0.766 *Index test 5: ATA 'high' (+ve for malignancy)*

Reference	**Ahmadi, 2019[10]**
	TP: 44 FN: 44 FP: 8 TN: 227 *sensitivity*: 0.500 *specificity*: 0.966
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Garcia-Monco Fernandez, 2018[105]**
Study type	Retrospective
Number of patients	n = 263 with 263 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): tester unclear but reviewed by a medic *Setting*: Secondary care *Country*: Spain *Inclusion criteria*: Patients with total or partial thyroidectomy due to nodular thyroid disease *Exclusion criteria*: Not reported

Reference	**Garcia-Monco Fernandez, 2018[105]**
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound, using 14 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n=75* *benign n= 188* *Index test 1: TI-RADS of 4 or more (+ve for malignancy)* TP: 49 FN: 26 FP: 14 TN: 174 *sensitivity*: 0.652 *specificity*: 0.926
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Rago, 1998[318]**
Study type	Prospective
Number of patients	n = 104
Patient characteristics	*Age, mean (SD): 42.3* *Gender (female to male ratio): 70:34*

Reference	**Rago, 1998[318]**
	Ethnicity: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Secondary care *Country*: Italy *Inclusion criteria*: Consecutive patients with a single thyroid nodule either in a normal thyroid or in a goiter, who underwent surgery for compressive symptoms or clinical suspicion of malignancy; nodules cold on scintiscans *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound and colour flow doppler US, using 7.5 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 30* *benign n= 74* *Index test 1: absent halo sign (+ve for malignancy)* TP: 20 FN: 10 FP: 17 TN: 57 *sensitivity*: 0.666 *specificity*: 0.77 *Index test 2: microcalcifications (+ve for malignancy)* TP: 13 FN: 17 FP: 18 TN: 56 *sensitivity*: 0.54 *specificity*: 0.756 *Index test 3: hypoechogenicity (+ve for malignancy)* TP: 20 FN: 10 FP: 38 TN: 36 *sensitivity*: 0.666 *specificity*: 0.486

Reference	Rago, 1998[318]
	Index test 4: absent halo sign + hypoechogenicity (+ve for malignancy) TP: 18 FN: 12 FP: 13 TN: 61 *sensitivity*: 0.600 *specificity*: 0.824 *Index test 5: absent halo sign + microcalcifications (+ve for malignancy)* TP: 8 FN: 22 FP: 5 TN: 69 *sensitivity*: 0.26 *specificity*: 0.932 *Index test 6: hypogenicity + microcalcifications (+ve for malignancy)* TP: 9 FN: 21 FP: 15 TN: 59 *sensitivity*: 0.30 *specificity*: 0.797 *Index test 7: type III intranodular blood flow (+ve for malignancy)* TP: 20 FN: 10 FP: 38 TN: 36 *sensitivity*: 0.666 *specificity*: 0.286 *Index test 8: type III intranodular blood flow + absent halo (+ve for malignancy)* TP: 15 FN: 15 FP: 8 TN: 66 *sensitivity*: 0.50 *specificity*: 0.890 *Index test 9: type III intranodular blood flow + microcalcifications (+ve for malignancy)* TP: 7 FN: 23 FP: 10 TN: 64 *sensitivity*: 0.233 *specificity*: 0.864 *Index test 10: type III intranodular blood flow + hypoechogenicity (+ve for malignancy)* TP: 14 FN: 16 FP: 22 TN: 52 *sensitivity*: 0.44 *specificity*: 0.702 *Index test 11: type III intranodular blood flow + absent halo + microcalcifications (+ve for malignancy)* TP: 5 FN: 25 FP: 2 TN: 72 *sensitivity*: 0.166 *specificity*: 0.972 *Index test 12: type III intranodular blood flow + hypoechogenicity + absent halo (+ve for malignancy)* TP: 13 FN: 17 FP: 6 TN: 68 *sensitivity*: 0.433 *specificity*: 0.918 *Index test 13: type III intranodular blood flow + hypoechogenicity + microcalcifications (+ve for malignancy)* TP: 6 FN: 14 FP: 8 TN: 66 *sensitivity*: 0.20 *specificity*: 0.891
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

2

Reference	**Horvath, 2017[149]**
Study type	Prospective
Number of patients	n = 210 patients with 502 nodules
Patient characteristics	*Age, median (IQR): 46 (18)* *Gender (female to male ratio): 164:46* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medics *Setting*: Secondary care *Country*: Chile *Inclusion criteria*: Consecutive patients undergoing thyroid US and thyroidectomy *Exclusion criteria*: patients with incomplete surgical or pathological information; undergoing surgery at other institutions; nodules whose anatomopathological characterisation was not possible due to tissue manipulation
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound and colour flow doppler US, using 5-12 and 5-17 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Horvath, 2017[149]**
Results	*malignant n= 272* *benign n= 230* *Index test 1: TIRADS of 3 or more (+ve for malignancy)* TP: 272 FN: 0 FP: 114 TN: 116 *sensitivity*: 1.0 *specificity*: 0.504 *Index test 1: TIRADS of 4A or more (+ve for malignancy)* TP: 271 FN: 1 FP: 59TN: 171 *sensitivity*: 0.996 *specificity*: 0.743 *Index test 1: TIRADS of 4B or more (+ve for malignancy)* TP: 270 FN: 2 FP: 43 TN: 187 *sensitivity*: 0.993 *specificity*: 0.813 *Index test 1: TIRADS of 4C or more (+ve for malignancy)* TP: 221 FN: 51 FP: 14 TN: 216 *sensitivity*: 0.813 *specificity*: 0.939 *Index test 1: TIRADS of 5 (+ve for malignancy)* TP: 86 FN: 186 FP: 1 TN: 229 *sensitivity*: 0.316 *specificity*: 0.996
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Kim, 2016[189]**
Study type	Prospective
Number of patients	n = 85 people with 91 nodules
Patient characteristics	*Age, mean (range): 51 (28-83)* *Gender (female to male ratio): 73:12* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic

Reference	**Kim, 2016[189]**
	Setting: Secondary care *Country*: South Korea *Inclusion criteria*: consecutive patients referred for total thyroidectomy *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* 2D and 3D Grey scale ultrasound US, using 6-12MHz probe frequency for each type *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 61* *benign n= 30* *This study utilised 3 readers, but only the results of the first reader are given:* *Index test 1: 2D US: At least one of the following: marked hypoechogenicity, a spiculated margin, a taller than wide shape, microcalcifications and macrocalcifications (+ve for malignancy)* TP: 36 FN: 25 FP: 8 TN: 22 *sensitivity*: 0.59 *specificity*: 0.730 *Index test 2: 3D US: At least one of the following: marked hypoechogenicity, a spiculated margin, a taller than wide shape, microcalcifications and macrocalcifications (+ve for malignancy)* TP: 46 FN: 15 FP: 8 TN: 22 *sensitivity*: 0.754 *specificity*: 0.730
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious risk of bias

Reference	**Kim, 2016[189]**
	Indirectness (QUADAS 2 - applicability): none
Comments	

1

2

Reference	**Nemec, 2012[284]**
Study type	Prospective
Number of patients	n = 46 patients with 46 nodules
Patient characteristics	*Age, mean (SD): 52.1(13.6)* *Gender (female to male ratio): 36:10* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic (radiologist) *Setting*: tertiary referral centre *Country*: Austria *Inclusion criteria*: age 18-80; sonographically identified solitary, cold thyroid nodule on Tc-99m pertechnetate thyroid scintigraphy; scheduled surgery for the nodule *Exclusion criteria*: pregnancy; multinodular thyroid goitre or hot nodules; history of thyroid malignancy; previous thyroid surgery or biopsy; previous adverse reactions to i.v. contrast agents
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound and Contrast enhanced US, using 4-7 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:*

Reference	**Nemec, 2012[284]**
	3 days
Results	*malignant n= 13* *benign n= 33* *Index test 1: Relative enhancement of 2.35 and higher [20s after peak enhancement] (+ve for malignancy)* TP: 10 FN: 13 FP: 5 TN: 28 *sensitivity*: 0.769 *specificity*: 0.848
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: no serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

2

Reference	**Sui, 2016[364]**
Study type	Prospective
Number of patients	n = 97 patients with 109 nodules
Patient characteristics	*Age, mean (SD): 48.6 (12.4)* *Gender (female to male ratio): 47:50* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): tester unclear but reviewed by a medic (radiologist) *Setting*: Department of Gland Surgery *Country*: China *Inclusion criteria*: Surgery indication for palpable or impalpable thyroid nodules; ii) patients scheduled for surgical removal of thyroid nodules; iii) the final diagnosis was confirmed by histopathologic examination of resected thyroid gland tissue; and iv) the patients did not suffer from any serious allergies

Reference	**Sui, 2016[364]**
	Exclusion criteria: Not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound and CEUS US, using 9-12MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 66 (PTCs)* *benign n= 43* *Index test 1: US: irregular shape (+ve for malignancy) [regular was deemed -ve for malignancy]* TP: 37 FN: 29 FP: 20 TN:23 *sensitivity*: 0.561*specificity*: 0.535 *Index test 2: US: aspect ratio ≥1 (+ve for malignancy) [<1 was deemed -ve for malignancy]* TP: 34 FN: 32 FP: 9 TN: 34 *sensitivity*: 0.515 *specificity*: 0.791 *Index test 3: US: poorly defined margin (+ve for malignancy) [well-defined was deemed -ve for malignancy]* TP: 57 FN: 9 FP: 0 TN:43 *sensitivity*: 0.864 *specificity*: 1.0 *Index test 4: US: hypoechogenicity (+ve for malignancy) [iso- and hyper- deemed were deemed -ve for malignancy]* TP: 25 FN: 41 FP: 5 TN: 38 *sensitivity*: 0.379 *specificity*: 0.884 *Index test 5: US: heterogenous echotexture (+ve for malignancy) [homogeneous was deemed -ve for malignancy]* TP: 39 FN: 27 FP: 23 TN:20 *sensitivity*: 0.591 *specificity*: 0.465 *Index test 6: US: microcalcification (+ve for malignancy) [regular was deemed -ve for malignancy]* TP: 49 FN: 17 FP: 13 TN:30 *sensitivity*: 0.742 *specificity*: 0.698

Reference	Sui, 2016[364]
	Index test 7: CEUS: poorly defined margin (+ve for malignancy) [well-defined was deemed -ve for malignancy] TP: 51 FN: 15 FP: 3 TN: 40 *sensitivity*: 0.772 *specificity*: 0.930 *Index test 8: CEUS: irregular shape (+ve for malignancy) [regular was deemed -ve for malignancy]* TP: 51 FN: 15 FP: 5 TN: 38 *sensitivity*: 0.772 *specificity*: 0.884 *Index test 9: CEUS: area ≥50% (+ve for malignancy) [<50% was deemed -ve for malignancy]* TP: 50 FN: 16 FP: 22 TN: 21 *sensitivity*: 0.75 *specificity*: 0.488 *Index test 10: CEUS: Intensity low (+ve for malignancy) [high and equal was deemed -ve for malignancy]* TP: 53 FN: 13 FP: 4 TN: 39 *sensitivity*: 0.803 *specificity*: 0.907 *Index test 11: CEUS: heterogenous type (+ve for malignancy) homogenous and ring was deemed -ve for malignancy]* TP: 53 FN: 13 FP: 4 TN: 39 *sensitivity*: 0.803 *specificity*: 0.907 *Index test 12: elastography – ITOH I to V scale: II and more (+ve for malignancy)* TP: 61 FN: 5 FP: 18 TN: 25 *sensitivity*: 0.924 *specificity*: 0.581 *Index test 13: elastography – ITOH I to V scale: III and more (+ve for malignancy)* TP: 55 FN: 11 FP: 5 TN: 38 *sensitivity*: 0.833 *specificity*: 0.884 *Index test 14: elastography – ITOH I to V scale: IV and more (+ve for malignancy)* TP: 38 FN: 28 FP: 1 TN: 42 *sensitivity*: 0.576 *specificity*: 0.977 *Index test 15: CEUS and RTE (+ve for malignancy)* TP: 63 FN: 3 FP: 2 TN: 41 *sensitivity*: 0.955 *specificity*: 0.954
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

Reference	**Shen, 2019[346]**
Study type	Retrospective
Number of patients	n = 1568 patients with 1612 nodules
Patient characteristics	*Age, mean: 52.0* *Gender (female to male ratio): 1192:376* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic (radiologist) *Setting*: Secondary care *Country*: China *Inclusion criteria*: aged 18-80; nodules >5mm; underwent surgery for thyroid nodules *Exclusion criteria*: History of invasive procedures such as ablation or FNA; those without US data; those with any mismatch between US images and pathological results [*meaning of this statement is unclear*]
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound and colour flow doppler US, using 5-12MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	Shen, 2019[346]
Results	*malignant n= 773* *benign n= 839* *Index test 1: solidity (+ve for malignancy) [mixed cystic or cystic deemed -ve for malignancy]* TP: 712 FN:61 FP:517 TN: 322 *sensitivity*: 0.921 *specificity*: 0.384 *Index test 2: very hypoechoic (+ve for malignancy) [hypoechoic, iso-, hyper- or anechoic deemed -ve for malignancy]* TP: 45 FN:728 FP:18 TN: 821 *sensitivity*: 0.058 *specificity*: 0.978 *Index test 3: taller than wide (+ve for malignancy) [wider than tall deemed -ve for malignancy]* TP: 332 FN:441 FP:36 TN: 803 *sensitivity*: 0.429 *specificity*: 0.957 *Index test 4: lobulated or irregular margins (+ve for malignancy) [smooth, ill defined, ETT deemed -ve for malignancy]* TP: 201 FN: 572 FP: 39 TN: 800 *sensitivity*: 0.921 *specificity*: 0.384 *Index test 5: micro-calcification (+ve for malignancy) [macro-, none or peripheral deemed -ve for malignancy]* TP: 502 FN:271 FP:96 TN: 743 *sensitivity*: 0.649 *specificity*: 0.886 *Index test 6: ACR – TI-RADS 3 or more (mildly suspicious) (+ve for malignancy)* TP: 773 FN:0 FP:833 TN: 6 *sensitivity*: 1.0 *specificity*: 0.007 *Index test 7: ACR – TI-RADS 4 or more (moderately suspicious) (+ve for malignancy)* TP: 760 FN:13 FP: 564 TN: 275 *sensitivity*: 0.983 *specificity*: 0.328 *Index test 8: ACR – TI-RADS 5 (highly suspicious) (+ve for malignancy)* TP: 682 FN:91 FP: 105 TN: 734 *sensitivity*: 0.882 *specificity*: 0.874 *Index test 9: EU - TI-RADS 3 or more (low risk) (+ve for malignancy)* TP: 773 FN:0 FP:831 TN: 8 *sensitivity*: 1.0 *specificity*: 0.009 *Index test 10: EU - TI-RADS 4 or more (intermediate risk) (+ve for malignancy)* TP: 762 FN:11 FP:562 TN: 277 *sensitivity*: 0.986 *specificity*: 0.330 *Index test 11: EU - TI-RADS 5 (high risk) (+ve for malignancy)* TP: 721 FN:52 FP:160 TN: 679 *sensitivity*: 0.933 *specificity*: 0.809

Reference	Shen, 2019[346]
	Index test 12: Kwak - TI-RADS 3 or more (no suspicious features) (+ve for malignancy) TP: 773 FN:0 FP:836 TN: 3 *sensitivity*: 1.0 *specificity*: 0.004 *Index test 13: Kwak - TI-RADS 4A or more (1 suspicious feature) (+ve for malignancy)* TP: 767 FN:6 FP:586 TN: 253 *sensitivity*: 0.992 *specificity*: 0.302 *Index test 14: Kwak - TI-RADS 4B or more (2 suspicious features) (+ve for malignancy)* TP: 760 FN:13 FP: 480 TN: 359 *sensitivity*: 0.983 *specificity*: 0.428 *Index test 15: Kwak - TI-RADS 4C or more (3 or 4 suspicious features) (+ve for malignancy)* TP: 721 FN:52 FP:116 TN: 723 *sensitivity*: 0.933 *specificity*: 0.862 *Index test 16: Kwak - TI-RADS 5 (5 suspicious features) (+ve for malignancy)* TP: 38 FN:735 FP:2 TN: 837*sensitivity*: 0.049 *specificity*: 0.998 *Index test 17: ATA 'very low suspicion' or more (+ve for malignancy)* TP: 773 FN:0 FP:834 TN: 5 *sensitivity*: 1.0 *specificity*: 0.006 *Index test 17: ATA 'low suspicion' or more (+ve for malignancy)* TP: 767 FN:6 FP:647 TN: 192 *sensitivity*: 0.992 *specificity*: 0.229 *Index test 17: ATA 'intermediate suspicion' or more (+ve for malignancy)* TP: 750 FN:23 FP:498 TN: 341 *sensitivity*: 0.970 *specificity*: 0.406 *Index test 17: ATA 'high suspicion' (+ve for malignancy)* TP: 708 FN:65 FP:150 TN: 689 *sensitivity*: 0.916 *specificity*: 0.821
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

1

Reference	**Maia, 2011[262]**
Study type	Retrospective
Number of patients	n = 143 patients with 143 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: General Hospital *Country*: Brazil *Inclusion criteria*: Patients submitted to total or partial thyroid surgery; US examination *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound and colour flow doppler US, using MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Maia, 2011[262]**
Results	*malignant n= 50* *benign n= 93* *Index test 1: microcalcifications (+ve for malignancy)* TP: 20 FN:30 FP: 12 TN: 81 *sensitivity*: 0.40 *specificity*: 0.871 *Index test 2: irregular border (+ve for malignancy)* TP: 34 FN:16 FP: 15 TN: 78 *sensitivity*: 0.68 *specificity*: 0.839 *Index test 3: hypoechogenicity (+ve for malignancy)* TP: 32 FN:18 FP: 36 TN: 57 *sensitivity*: 0.64 *specificity*: 0.613 *Index test 4: size ≥2cm (+ve for malignancy)* TP: 23 FN:27 FP: 64 TN: 29 *sensitivity*: 0.46 *specificity*: 0.312 *Index test 5: internal blood flow (+ve for malignancy)* TP: 17 FN:33 FP: 13 TN: 80 *sensitivity*: 0.34 *specificity*: 0.860
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

2

Reference	**Appetecchia, 2006[18]**
Study type	Retrospective
Number of patients	n = 203
Patient characteristics	*Age, mean (range): 44.6(18-75)* *Gender (female to male ratio): 181:49*

Reference	Appetecchia, 2006[18]
	Ethnicity: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Cancer Institute *Country*: Italy *Inclusion criteria*: Patients submitted to surgery for solitary thyroid disease *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound and colour flow doppler US, using 13MHz transducer for standard study and 4.7 MHz probe frequency for colour flow doppler study (CFDS) *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 36* *benign n= 167* *Index test 1: US: absent halo (+ve for malignancy)* TP: 32 FN: 4 FP: 130 TN: 37 *sensitivity*: 0.889 *specificity*: 0.222 *Index test 2: US: hypoechoicity (+ve for malignancy)* TP: 18 FN: 18 FP: 65 TN: 102 *sensitivity*: 0.500 *specificity*: 0.611 *Index test 3: US: microcalcifications (+ve for malignancy)*

Reference	**Appetecchia, 2006[18]**
	TP: 30 FN: 6 FP: 56 TN: 111 *sensitivity*: 0.833 *specificity*: 0.665 *Index test 4: US: absent halo + hypoechoicity (+ve for malignancy)* TP: 15 FN:21 FP: 53 TN:114 *sensitivity*: 0.417 *specificity*: 0.683 *Index test 5: US: absent halo + microcalcifications (+ve for malignancy)* TP: 27 FN: 9 FP:47 TN:120 *sensitivity*: 0.750 *specificity*: 0.719 *Index test 6: US: hypoechoicity + microcalcifications (+ve for malignancy)* TP: 15 FN: 21 FP:22 TN: 145 *sensitivity*: 0.417 *specificity*: 0.868 *Index test 7: US: CFDS pattern III (+ve for malignancy)* TP: 33 FN: 3 FP: 109 TN: 58 *sensitivity*: 0.917 *specificity*: 0.347 *Index test 8: US: absent halo + hypoechoicity (+ve for malignancy)* TP: 15 FN:21 FP: 53 TN: 114 *sensitivity*: 0.417 *specificity*: 0.683 *Index test 9: US: absent halo + CFDS III (+ve for malignancy)* TP: 30 FN: 6 FP: 94 TN: 73 *sensitivity*: 0.833 *specificity*: 0.437 *Index test 10: US: hypoechoic + CFDS III (+ve for malignancy)* TP: 17 FN: 19 FP: 41 TN: 126 *sensitivity*: 0.472 *specificity*: 0.754 *Index test 11: US: microcalcifications + CFDS III (+ve for malignancy)* TP: 29 FN: 7 FP: 41 TN: 126 *sensitivity*: 0.806 *specificity*: 0.754 *Index test 12: US: absent halo + hypoechogenicity + CFDS III (+ve for malignancy)* TP: 3 FN:33 FP: 0 TN: 167 *sensitivity*: 0.083 *specificity*: 1.00 *Index test 13: US: absent halo + microcalcifications + CFDS III (+ve for malignancy)* TP: 26 FN: 10 FP: 38 TN: 129 *sensitivity*: 0.722 *specificity*: 0.772 *Index test 14: US: microcalcifications + hypoechogenicity + CFDS III (+ve for malignancy)* TP: 15 FN: 21 FP: 16 TN: 151 *sensitivity*: 0.417 *specificity*: 0.904
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias

Reference	**Appetecchia, 2006[18]**
	Indirectness (QUADAS 2 - applicability): serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

2

Reference	**Shreyamsa, 2021[352]**
Study type	Prospective
Number of patients	n = 168 nodules from 139 patients
Patient characteristics	*Age, mean (SD): 35.3(13.2)* *Gender (female to male ratio): 115:24* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic (radiologist) *Setting*: Department of endocrine surgery and radiodiagnosis *Country*: India *Inclusion criteria*: all nodules measuring 4cm or less; availability of surgical histopathological findings *Exclusion criteria*: patients with diffuse thyroid enlargement, autoimmune and inflammatory disorders; those unwilling to participate
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound and colour flow doppler US, using 12 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings

Reference	Shreyamsa, 2021[352]
	Time between measurement of index test and reference standard: Not clear
Results	*malignant n= 35* *benign n= 133* *Index test 1: K-TIRADS 4 'intermediate risk' and above (+ve for malignancy)* TP: 26 FN: 9 FP: 18 TN: 115 *sensitivity*: 0.743 *specificity*: 0.864 *Index test 2: K-TIRADS 5 'high risk' (+ve for malignancy)* TP: 14 FN: 21 FP: 2 TN: 131 *sensitivity*: 0.40 *specificity*: 0.985 *Index test 3: ACR-TIRADS 4 'intermediate risk' and above (+ve for malignancy)* TP: 26 FN: 9 FP: 15 TN: 118 *sensitivity*: 0.743 *specificity*: 0.887 *Index test 4: ACR-TIRADS 5 'high risk' (+ve for malignancy)* TP: 13 FN: 22 FP: 1 TN: 132 *sensitivity*: 0.371 *specificity*: 0.992 *Index test 5: ATA risk 'intermediate suspicion' and above (+ve for malignancy)* TP: 25 FN: 10 FP: 17 TN: 116 *sensitivity*: 0.714 *specificity*: 0.872 *Index test 6: ATA risk 'high suspicion' (+ve for malignancy)* TP: 24 FN: 11 FP: 7 TN: 126 *sensitivity*: 0.686 *specificity*: 0.947 *Index test 7: F-TIRADS 4B and above (+ve for malignancy)* TP: 29 FN: 6 FP: 7 TN: 126 *sensitivity*: 0.829 *specificity*: 0.947 *Index test 8: F-TIRADS 4C and above (+ve for malignancy)* TP: 16 FN: 21 FP: 2 TN: 131 *sensitivity*: 0.457*specificity*: 0.985 *Index test 7: TMC-RSS Category 2 and above (+ve for malignancy)* TP: 32 FN: 3 FP: 4 TN: 129 *sensitivity*: 0.914 *specificity*: 0.970 *Index test 8: TMC-RSS Category 3 and above (+ve for malignancy)* TP: 27 FN: 8 FP: 0 TN: 133 *sensitivity*: 0.771 *specificity*: 1.0

Reference	**Shreyamsa, 2021[352]**
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

2

Reference	**Nilakantan, 2007[286]**
Study type	Retrospective
Number of patients	n = 106 nodules in 106 patients
Patient characteristics	*Age, mean (range): 36.6(14-68)* *Gender (female to male ratio): 86:20* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic (radiologist) *Setting*: Tertiary referral centre *Country*: India *Inclusion criteria*: All patients undergoing surgery for benign or malignant nodular thyroid disease *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound and colour flow doppler US, using 2.6-8.6 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings

Reference	**Nilakantan, 2007[286]**
	Time between measurement of index test and reference standard: Not clear
Results	*malignant n= 20* *benign n= 86* *Index test 1: Any of absent halo sign, solidity, hypoechogenicity, heterogeneous echo structure, irregular margins, fine calcifications and extra glandular extensions (+ve for malignancy)* TP: 4 FN: 16 FP: 2 TN: 84 *sensitivity*: 0.20 *specificity*: 0.977
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

2

Reference	**Huang, 2020[155]**
Study type	Prospective
Number of patients	n = 346 patients with 392 nodules
Patient characteristics	*Age, mean (SD): 51.2 (benign) and 41.6 (malignant)* *Gender (female to male ratio): 280: 66* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown (‘excellent ultrasonologists’) *Setting*: Teaching hospital

Reference	**Huang, 2020[155]**
	Country: China *Inclusion criteria*: Patients with thyroid nodules treated surgically at the research hospital; TIRADS category 4 nodules; conventional US, FNAC and elastography performed before surgery; final diagnosis based on surgery *Exclusion criteria*: surgery for hyperthyroidism; previous history of neck radiation or surgery
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale ultrasound and colour flow doppler US, using MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 233* *benign n= 159* *Index test 1: KWAK TIRADS 4b and above (+ve for malignancy)* TP: 136 FN: 97 FP: 37 TN: 122 *sensitivity*: 0.584 *specificity*: 0.767 *Index test 2: KWAK TIRADS 4c and above (+ve for malignancy)* TP: 79 FN: 154 FP: 6 TN: 153 *sensitivity*: 0.339 *specificity*: 0.962 *Index test 3: elastography – ITOH 1-5 score: 3 and above (+ve for malignancy)* TP: 233 FN: 0 FP: 134 TN: 25 *sensitivity*: 1.0 *specificity*: 0.157 *Index test 4: elastography – ITOH 1-5 score: 4 and above (+ve for malignancy)* TP: 158 FN: 75 FP: 7 TN: 152 *sensitivity*: 0.678 *specificity*: 0.956 *Index test 5: elastography – ITOH 1-5 score: 5 (+ve for malignancy)* TP: 52 FN: 181 FP: 0 TN: 159 *sensitivity*: 0.233 *specificity*: 1.00

Reference	**Huang, 2020[155]**
	Index test 6: Kwak TIRADS and ITOH score: 5 and above (+ve for malignancy) TP: 214 FN: 19 FP: 40 TN: 119 *sensitivity*: 0.918 *specificity*: 0.751
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

2

Reference	**Deng, 2018[77]**
Study type	Retrospective
Number of patients	n = 92 with 107 nodules
Patient characteristics	*Age, mean (SD): 62.5(5)* *Gender (female to male ratio): 59:33* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: Secondary care *Country*: China *Inclusion criteria*: 1) Patients with thyroid diseases underwent both HFCDU and pathological diagnosis; 2) Patients were willing to cooperate with the treatment; 3) Patients' medical records were complete. *Exclusion criteria*: 1) Patients had other severe thyroid diseases in addition to thyroid nodules, or the acoustic halo couldn't be detected in thyroid nodules; 2) Patients experienced recurrence of malignant thyroid nodules after surgery, or had hyperplasia of one thyroid lobe caused by the hypoplasia of thyroid and parathyroid in the opposite lobe; 3) Patients had benign thyroid nodules generated by the scar and proliferation of the residual thyroid tissue or other factors after treatment.

Reference	Deng, 2018[77]
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Colour flow doppler US, using 3.5 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 45* *benign n= 62* *Index test 1:unclear boundary (+ve for malignancy)* TP: 39 FN:6 FP:16 TN:46 *sensitivity*: 0.86 *specificity*: 0.742 *Index test 2: irregular shape (+ve for malignancy)* TP: 38 FN:7 FP:29 TN:33 *sensitivity*: 0.84 *specificity*: 0.532 *Index test 3: aspect ratio >1 (+ve for malignancy)* TP: 30 FN:15 FP:13 TN:49 *sensitivity*: 0.667 *specificity*: 0.790 *Index test 4: internal echo uneven (+ve for malignancy)* TP: 40 FN:5 FP:27 TN: 35 *sensitivity*: 0.6889 *specificity*: 0.565 *Index test 5: no attenuation of posterior echo (+ve for malignancy)* TP: 26 FN:19 FP:49 TN:13 *sensitivity*: 0.57 *specificity*: 0.209 *Index test 6: calcification (+ve for malignancy)* TP: 22 FN:23 FP:11 TN:51 *sensitivity*: 0.489 *specificity*: 0.826 *Index test 7: Grade 1 blood flow and above (+ve for malignancy)* TP: 42 FN:3 FP:37 TN:25 *sensitivity*: 0.933 *specificity*: 0.403

Reference	**Deng, 2018[77]**
	Index test 7: Grade 2 blood flow and above (+ve for malignancy) TP: 34 FN:11 FP:18 TN:44 *sensitivity*: 0.755 *specificity*: 0.597 *Index test 7: Grade 3 blood flow and above (+ve for malignancy)* TP: 19 FN: 26 FP:8 TN:54 *sensitivity*: 0.422 *specificity*: 0.871
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

2

Reference	**Li, 2018[220]**
Study type	Prospective
Number of patients	n = 68 patients with 102 nodules
Patient characteristics	*Age, mean (SD): 45.8(9.2)* *Gender (female to male ratio): 28:40* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Secondary care *Country*: China *Inclusion criteria*: 1) patients were diagnosed as having thyroid nodules by palpation and general neck ultrasound; 2) patients received surgery and pathologic diagnosis; 3) patients signed the informed consent.

Reference	Li, 2018[220]
	Exclusion criteria: 1) the nodules were completely cystic; 2) the nodules were too big or too close to the edge where there was no adequate thyroid tissue surrounded for comparison; 3) patients received head and neck radiotherapy in the past; 3) patients were obese with excessive fat in the larynx that could affect the results of BUS
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Colour doppler US, using 5-14 MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 41* *benign n= 61* *Index test 1:* Any of the following US characteristics: *unclear boundary, irregular shape, internal hypoecho, posterior echo attenuation, micro-calcification, rich blood flow, anteroposterior/transverse diameter ratio (A/T) ≥1, and vascular resistance index (RI) ≥0.7 (+ve for malignancy)* TP: 19 FN:22 FP: 20 TN: 41 *sensitivity*:0.463 *specificity*: 0.672 *Index test 2:* elastography – Asteria score 1-4: 3 or higher *(+ve for malignancy)* TP: 31 FN:10 FP: 9 TN: 52 *sensitivity*:0.7561 *specificity*: 0.8525
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Liu, 2019[243]**
Study type	Retrospective, consecutive patient enrolment
Number of patients	n = 174 patients with 174 nodules (only one nodule per patient included – most suspicious included)
Patient characteristics	*Age, mean (SD): 47.4 (10.5)* *Gender (female to male ratio): 140:34* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic (radiologist) *Setting*: Teaching hospital *Country*: China *Inclusion criteria*: Consecutive patients with thyroid nodules scheduled to undergo surgery imaged by greyscale US and SWE *Exclusion criteria*: previous invasive procedures for thyroid nodules; no histopathological data
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Colour flow doppler US, using 4-15 MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Liu, 2019[243]**
Results	*malignant n= 64* *benign n= 110* *Index test 1: marked hypoechogenicity (+ve for malignancy)* TP: 31 FN: 33 FP: 8 TN: 102 *sensitivity*: 0.484 *specificity*: 0.927 *Index test 2: poorly defined margin (+ve for malignancy)* TP: 36 FN: 28 FP: 23 TN: 87 *sensitivity*: 0.563 *specificity*: 0.790 *Index test 3: microcalcification (+ve for malignancy)* TP: 41 FN: 23 FP: 17 TN: 93 *sensitivity*: 0.641 *specificity*: 0.845 *Index test 4: taller than wide shape (+ve for malignancy)* TP: 20 FN: 44 FP: 5 TN: 105 *sensitivity*: 0.313 *specificity*: 0.954 *Index test 5: one or more of marked hypoechogenicity, poorly defined margin, microcalcifications, or taller than wide (+ve for malignancy)* TP: 61 FN: 3 FP: 36 TN: 74 *sensitivity*: 0.953 *specificity*: 0.673 *Index test 6: elastography – EI of 52.1 kpa or higher (+ve for malignancy)* TP: 47 FN: 17 FP: 26 TN: 84 *sensitivity*: 0.734 *specificity*: 0.764 *Index test 7: elastography – '4 pattern': 3 or more (+ve for malignancy)* TP: 57 FN: 7 FP: 28 TN: 82 *sensitivity*: 0.891 *specificity*: 0.746
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Maimati, 2016[263]**
Study type	Prospective

Reference	**Maimati, 2016[263]**
Number of patients	n = 600 nodules in 600 patients
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): 489:111* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Secondary care *Country*: China *Inclusion criteria*: thyroid nodules ≤ 10 mm in maximum diameter in 600 patients who underwent conventional ultrasonography examinations of the thyroid gland before surgery *Exclusion criteria*: No surgical evaluation; nodules >10mm;
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Colour flow doppler US, using unknown MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 356* *benign n= 244* *Note: the sensitivities and specificities below are calculated from the raw data provided in the paper. These sensitivity and specificity values differ from the sensitivities and specificities provided in the paper. It is assumed that it is more likely that the raw data in the paper is correct than the final calculated accuracy data in the paper is correct, given that the latter carries additional scope for human error.*

Reference	Maimati, 2016[263]
	Given that I am highly confident that there have been no errors made in calculating accuracy values from the raw data on my part, the values below are the ones that will be used. *Index test 1: hypoechogenicity (+ve for malignancy) [iso- and hyper- echoic deemed -ve for malignancy]* TP: 339 FN: 17 FP: 172 TN: 72 *sensitivity*: 0.952 *specificity*: 0.295 *Index test 2: Solid (+ve for malignancy) [cystic >50% and predominantly solid deemed -ve for malignancy]* TP: 301 FN: 55 FP: 121 TN: 123 *sensitivity*: 0.846 *specificity*: 0.504 *Index test 3: taller than wide (+ve for malignancy) [oval to round deemed -ve for malignancy]* TP: 155 FN: 201 FP: 24 TN: 220 *sensitivity*: 0.435 *specificity*: 0.902 *Index test 4: poorly defined boundary (+ve for malignancy) [well defined deemed -ve for malignancy]* TP: 213 FN: 143 FP: 51 TN: 193 *sensitivity*: 0.598 *specificity*: 0.791 *Index test 5: irregular shape (+ve for malignancy) [regular deemed -ve for malignancy]* TP: 187 FN: 169 FP: 26 TN: 218 *sensitivity*: 0.525 *specificity*: 0.893 *Index test 6: complete capsule (+ve for malignancy) [incomplete deemed -ve for malignancy]* TP: 302 FN: 54 FP: 237 TN: 7 *sensitivity*: 0.848 *specificity*: 0.029 *Index test 7: central vascularity (+ve for malignancy) [negative, peripheral or central/peripheral combined deemed -ve for malignancy]* TP: 190 FN: 166` FP: 114 TN: 130 *sensitivity*: 0.534 *specificity*: 0.533 *Index test 8: microcalcification (+ve for malignancy) [macro- or none deemed -ve for malignancy]* TP: 222 FN: 134 FP: 79 TN: 165 *sensitivity*: 0.624 *specificity*: 0.676 *Index test 9: solitary nodule (+ve for malignancy) [multifocality deemed -ve for malignancy]* TP: 232 FN: 124 FP: 218 TN: 26 *sensitivity*: 0.652 *specificity*: 0.107 *Index test 10: Solid (+ve for malignancy) [cystic >50% and predominantly solid deemed -ve for malignancy]* TP: 301 FN: 55 FP: 121 TN: 123 *sensitivity*: 0.846 *specificity*: 0.504
Source of funding	None reported

Reference	**Maimati, 2016[263]**
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Tang, 2017[376]**
Study type	Prospective
Number of patients	n = 323 nodules in 323 patients
Patient characteristics	*Age, median (range): 43.85 (17-72)* *Gender (female to male ratio): 176:147* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Secondary care *Country*: China *Inclusion criteria*: Patients receiving thyroid surgery, US examination and elastography *Exclusion criteria*: none reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* US, using 6-13MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Tang, 2017[376]**
Results	*malignant n= 116* *benign n= 207* *Index test 1: US – based on any of the following [details not provided]: number, size, morphology, boundary, length/width ratio, surrounding halo, internal echo, calcification or expansion of neck lymph nodes (+ve for malignancy)* TP: 102 FN: 14 FP: 13 TN: 194 *sensitivity*: 0.879 *specificity*: 0.937 *Index test 2: elastography 0-5 colour method (different to all others): 3 or more (+ve for malignancy)* TP: 107 FN: 9 FP: 9 TN: 198 *sensitivity*: 0.922 *specificity*: 0.947
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Pei, 2019[307]**
Study type	Retrospective
Number of patients	n = 170 patients with 196 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): 112:58* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown (‘ultrasonographers’) *Setting*: Teaching hospital *Country*: China *Inclusion criteria*: (a) those meeting the criteria for TI-RADS category 4; (b) all patients with complete data, including US indicators and pathological findings; (c) all nodules, in which RTE and SMI were successfully implemented; and (d) all thyroid nodules that were not subjected to minimally invasive surgery prior to US examination (such as puncture and ablation).

Reference	Pei, 2019[307]
	Exclusion criteria: Not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Colour flow doppler US, using high frequency probe (undefined); elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 118* *benign n= 78* *Index test 1: solid (+ve for malignancy) [mixed, cystic or sponge-like deemed -ve for malignancy]* TP: 112 FN: 6 FP:63 TN:15 *sensitivity*: 0.949 *specificity*: 0.192 *Index test 2: marked hypoechogenicity (+ve for malignancy) [hypo-, iso-, hyper- or an-echogenicity deemed -ve for malignancy]* TP: 85 FN: 33 FP:31 TN:47 *sensitivity*: 0.720 *specificity*: 0.603 *Index test 3: taller than wide (+ve for malignancy) wider than tall deemed -ve for malignancy]* TP: 58 FN: 60 FP:24 TN:54 *sensitivity*: 0.492 *specificity*: 0.692 *Index test 4: poorly defined margin (+ve for malignancy) [well-defined, irregular/lobulating or extracapsular spread deemed -ve for malignancy]* TP: 91 FN: 27 FP:57 TN:21 *sensitivity*: 0.771 *specificity*: 0.269 *Index test 5: microcalcification (+ve for malignancy) [macro-, none, peripheral, or mixed micro/macro deemed -ve for malignancy]* TP: 53 FN: 65 FP:4 TN:74 *sensitivity*: 0.449 *specificity*: 0.949 *Index test 6: Superb Microvascular Imaging III or more (+ve for malignancy) [mixed, cystic or sponge-like deemed -ve for malignancy]* TP: 108 FN: 10 FP: 12 TN: 66 *sensitivity*: 0.915 *specificity*: 0.846

Reference	**Pei, 2019[307]**
	Index test 7: elastography RGB 0-4 scale: 2 or more (+ve for malignancy) TP: 118 FN: 0 FP: 49 TN: 29 *sensitivity*: 1.0 *specificity*: 0.372 *Index test 8: elastography RGB 0-4 scale: 3 or more (+ve for malignancy)* TP: 95 FN: 23 FP: 12 TN: 66 *sensitivity*: 0.805 *specificity*: 0.846 *Index test 9: elastography RGB 0-4 scale: 4 (+ve for malignancy)* TP: 6 FN: 112 FP: 0 TN: 78 *sensitivity*: 0.051 *specificity*: 1.0
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Hang, 2018[139]**
Study type	Retrospective
Number of patients	n = 262 patients with 298 thyroid nodules
Patient characteristics	*Age, mean (SD): 45.57(12.1)* *Gender (female to male ratio): 247:51* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): tester unclear but reviewed by a medic *Setting*: Secondary care *Country*: China *Inclusion criteria*: (a) age > 18 years; (b) no therapy or biopsy before US examination; (c) diameter of the thyroid nodule between 5 and 30mm; (d) patients with solid or mostly solid thyroid nodules as assessed by US examination (cystic part < 50%); (e) distance from the

Reference	Hang, 2018[139]
	skin surface to nodular center was <25mm, as this could be fully included in the maximum range of the SWE color overlay; (f) underwent conventional US and SWE examination before surgery *Exclusion criteria*: Not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* US, using 4-15MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 177* *benign n= 121* *Note: TIRADS was measured on a score system that differed from other papers: up to and over 10 points* *Index test 1: solid (+ve for malignancy) [mixed or cystic/spongiform deemed -ve for malignancy]* TP: 177 FN: 0 FP: 110 TN: 11 *sensitivity*:1.0 *specificity*: 0.091 *Index test 2: very hypoechoic (+ve for malignancy) [hypo-, iso-, hyper- or an-echoic deemed -ve for malignancy]* TP: 22 FN: 155 FP: 2 TN: 119 *sensitivity*:0.124 *specificity*: 0.983 *Index test 3: taller than wide (+ve for malignancy) [wider than tall deemed -ve for malignancy]* TP: 39 FN: 138 FP: 6 TN: 115 *sensitivity*: 0.220 *specificity*: 0.950 *Index test 4 Extrathyroidal extension or lobulated or irregular (+ve for malignancy) [smooth or ill defined deemed -ve for malignancy]* TP: 142 FN: 35 FP: 26 TN: 95 *sensitivity*: 0.802 *specificity*: 0.785 *Index test 5: Punctate echogenic (+ve for malignancy) [peripheral rim calcifications or macrocalcifications or comet tail artifacts or no echogenic foci deemed -ve for malignancy]* TP: 100 FN: 77 FP: 110 TN: 11 *sensitivity*: 0.565 *specificity*: 0.091

Reference	Hang, 2018[139]
	Index test 6: TIRADS score of 3 or more (+ve for malignancy) TP: 177 FN: 0 FP: 110 TN: 11 *sensitivity*: 1.00 *specificity*: 0.091 *Index test 7: TIRADS score of 4 or more (+ve for malignancy)* TP: 175 FN: 2 FP: 83 TN: 38 *sensitivity*: 0.989 *specificity*: 0.314 *Index test 8: TIRADS score of 5 or more (+ve for malignancy)* TP: 161 FN: 16 FP: 41 TN: 80 *sensitivity*: 0.910 *specificity*: 0.661 *Index test 9: TIRADS score of 6 or more (+ve for malignancy)* TP: 159 FN: 18 FP: 32 TN: 89 *sensitivity*: 0.898 *specificity*: 0.736 *Index test 10: TIRADS score of 7 or more (+ve for malignancy)* TP: 132 FN: 45 FP: 17 TN: 104 *sensitivity*: 0.746 *specificity*: 0.860 *Index test 11: TIRADS score of 8 or more (+ve for malignancy)* TP: 113 FN: 64 FP: 7 TN: 114 *sensitivity*: 0.638 *specificity*: 0.942 *Index test 11: TIRADS score of 9 or more (+ve for malignancy)* TP: 108 FN: 69 FP: 6 TN: 115 *sensitivity*: 0.610 *specificity*: 0.950 *Index test 11: TIRADS score of 10 or more (+ve for malignancy)* TP: 48 FN: 129 FP: 0 TN: 121 *sensitivity*: 0.271 *specificity*: 1.0 *Index test 11: elastography – Emax of 45 kpa (2pts) or more (+ve for malignancy)* TP: 151 FN: 26 FP: 56 TN: 65 *sensitivity*: 0.853 *specificity*: 0.537 *Index test 11: elastography – Emax of 69 kpa (3pts) or more (+ve for malignancy)* TP: 90 FN: 87 FP: 18 TN: 103 *sensitivity*: 0.508 *specificity*: 0.851 *Index test 11: elastography – Emax of 120 kpa (4pts) or more (+ve for malignancy)* TP: 19 FN: 158 FP: 3 TN: 118 *sensitivity*: 0.107 *specificity*: 0.975 *Index test 11: combined TIRADS (0-10) and Emax score: 5 or more (+ve for malignancy)* TP: 177 FN: 0 FP: 101 TN: 20 *sensitivity*: 1.0 *specificity*: 0.165

Reference	**Hang, 2018[139]**
	Index test 11: combined TIRADS (0-10) and Emax score: 6 or more (+ve for malignancy) TP: 171 FN: 6 FP: 60 TN: 61 *sensitivity*: 0.966 *specificity*: 0.504 *Index test 11: combined TIRADS (0-10) and Emax score: 7 or more (+ve for malignancy)* TP: 163 FN: 14 FP: 39 TN: 82 *sensitivity*: 0.921 *specificity*: 0.678 *Index test 11: combined TIRADS (0-10) and Emax score: 8 or more (+ve for malignancy)* TP: 156 FN: 21 FP: 25 TN: 96 *sensitivity*: 0.881 *specificity*: 0.793 *Index test 11: combined TIRADS (0-10) and Emax score: 9 or more (+ve for malignancy)* TP: 145 FN: 32 FP: 14 TN: 107 *sensitivity*: 0.819 *specificity*: 0.884
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Bakari, 2018[27]**
Study type	Prospective
Number of patients	n = 54 patients with 78 nodules
Patient characteristics	*Age, median (range): 45.9 (21-79)* *Gender (female to male ratio): 90:10* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Teaching hospital *Country*: China

Reference	Bakari, 2018[27]
	Inclusion criteria: (1) Patients who are 18 years of age or older and of different genders; (2) Patients with single or multiple solid TNs bigger than 0.5 cm in size and scheduled for surgical treatment; (3) Patients with thyroid nodules of ≥0.5 cm with at least two of the following characteristic features hypoechoic, microcalcifications, irregular outline, taller than wider, increased central vascularity, isoechogenicity, and heterogeneity patterns; and (4) No invasive thyroid surgery or FNAB were performed before. *Exclusion criteria*: (1) Patients with fluid filled thyroid lesions; (2) Patients with thyroid nodules lesions of 0.5 cm regardless of their characteristic features.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* B mode US, using unreported MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Bakari, 2018[27]**
Results	*malignant n= 54* *benign n= 24* *Index test 1: hypoechoicity (+ve for malignancy)* TP: 53 FN: 1 FP: 4 TN: 20 *sensitivity*: 0.981 *specificity*: 0.833 *Index test 2: spot microcalcification (+ve for malignancy)* TP: 51 FN: 3 FP: 5 TN: 19 *sensitivity*: 0.944 *specificity*: 0.792 *Index test 3: irregular margin (+ve for malignancy)* TP: 52 FN: 1 FP: 7 TN: 17 *sensitivity*: 0.981 *specificity*: 0.708 *Index test 4: heterogeneity(+ve for malignancy)* TP: 5 FN: 49 FP: 3 TN: 21 *sensitivity*: 0.093 *specificity*: 0.875 *Index test 5: central vascularity (+ve for malignancy)* TP: 53 FN: 1 FP: 15 TN: 9 *sensitivity*: 0.981 *specificity*: 0.375 *Index test 5: elastography – SWE velocity of 2.4 m/s or above (+ve for malignancy)* TP: 50 FN: 4 FP: 4 TN: 20 *sensitivity*: 0.925 *specificity*: 0.833
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Rago, 2007[317]**
Study type	Prospective
Number of patients	n = 92
Patient characteristics	*Age, mean (SD): 43(15)* *Gender (female to male ratio): 63:29*

Reference	Rago, 2007[317]
	Ethnicity: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Secondary care *Country*: Italy *Inclusion criteria*: Patients who underwent thyroid surgery for compressive symptoms or suspicion of malignancy on FNA cytology. *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Colour flow doppler US, using 10 MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 31* *benign n= 61* *Index test 1: hypoechogenicity (+ve for malignancy)* TP: 25 FN: 6 FP:23 TN: 38 *sensitivity*: 0.806 *specificity*: 0.623 *Index test 2: absent halo sign (+ve for malignancy)* TP: 19 FN: 12 FP:11 TN: 50 *sensitivity*: 0.613 *specificity*: 0.820 *Index test 3: spot microcalcifications (+ve for malignancy)* TP: 20 FN: 11 FP:17 TN: 44 *sensitivity*: 0.645 *specificity*: 0.721

Reference	**Rago, 2007[317]**
	Index test 4: type III vascularisation (+ve for malignancy) TP: 2 FN: 29 FP:2 TN: 59 *sensitivity*: 0.065 *specificity*: 0.967 *Index test 5: absent halo AND hypoechogenicity [BOTH PRESENT] (+ve for malignancy)* TP: 18 FN: 13 FP:4 TN: 57 *sensitivity*:0.581 *specificity*: 0.934 *Index test 6: absent halo AND spot microcalcifications [BOTH PRESENT] (+ve for malignancy)* TP: 19 FN: 12 FP:3 TN: 58 *sensitivity*0.613 *specificity*: 0.951 *Index test 7: hypoechogenicity AND spot microcalcifications [BOTH PRESENT] (+ve for malignancy)* TP: 16 FN: 15 FP:6 TN: 55 *sensitivity*: 0.516 *specificity*: 0.902 *Index test 8: absent halo AND hypoechogenicity AND type III vasc. [ALL PRESENT] (+ve for malignancy)* TP: 1 FN: 30 FP:0 TN: 61 *sensitivity*: 0.032 *specificity*: 1.00 *Index test 9: spot microcalcifications AND hypoechogenicity AND type III vasc. [ALL PRESENT] (+ve for malignancy)* TP: 2 FN: 29 FP:0 TN: 61 *sensitivity*: 0.065 *specificity*: 1.00 *Index test 10: spot microcalcifications AND absent halo sign AND type III vasc. [ALL PRESENT] (+ve for malignancy)* TP: 1 FN: 30 FP:0 TN: 61 *sensitivity*: 0.032 *specificity*: 1.00 *Index test 11: elastography- Rago I-V scale: 4 or more (+ve for malignancy)* TP: 30 FN: 1 FP:0 TN: 61 *sensitivity*: 0.97 *specificity*: 1.00
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Dobruch-Sobczak, 2019[81]**
Study type	Retrospective
Number of patients	n = 428 with 842 nodules

Reference	Dobruch-Sobczak, 2019[81]
Patient characteristics	*Age, mean (range): 62.7 (14-86)* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: Tertiary referral centres *Country*: Poland *Inclusion criteria*: patients who had been admitted to the tertiary referral centre for thyroidectomy; population from a previously iodine deficient region *Exclusion criteria*: symptomatic purely cystic lesions
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale US, using 5-12, 7-18 or 5-15 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 229* *benign n= 613* *Index test 1:markedly hypoechoic (+ve for malignancy)* TP: 135 FN: 94 FP: 173 TN: 440 *sensitivity*: 0.590 *specificity*: 0.718 *Index test 2: irregular margins (+ve for malignancy)*

Reference	**Dobruch-Sobczak, 2019[81]**
	TP: 173 FN: 56 FP:112 TN: 501 *sensitivity*: 0.755 *specificity*: 0.817 *Index test 3:microcalcifications (+ve for malignancy)* TP: 123 FN: 106 FP: 148 TN: 465 *sensitivity*: 0.537 *specificity*: 0.759 *Index test 4:macrocalcifications (+ve for malignancy)* TP: 51 FN: 178 FP: 93 TN: 520 *sensitivity*: 0.223 *specificity*: 0.848 *Index test 5:solid/almost solid (+ve for malignancy)* TP: 212 FN: 17 FP:343 TN:270 *sensitivity*: 0.926 *specificity*: 0.440 *Index test 6: taller than wide (+ve for malignancy)* TP: 105 FN: 124 FP: 91 TN: 522 *sensitivity*: 0.459 *specificity*: 0.852 *Index test 7: EU TIRADS of 3 or more (+ve for malignancy)* TP: 229 FN:0 FP: 459 TN: 154 *sensitivity*: 1.0 *specificity*: 0.251 *Index test 8: EU TIRADS of 4 or more (+ve for malignancy)* TP: 226 FN:3 FP: 369 TN: 244 *sensitivity*: 0.987 *specificity*: 0.398 *Index test 9: EU TIRADS of 5 (+ve for malignancy)* TP: 214 FN:15 FP: 278 TN: 335 *sensitivity*: 0.934 *specificity*: 0.546
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Schenke, 2020[334]**
Study type	Retrospective, consecutive patient enrolment
Number of patients	n = 140 with 145 nodules

Reference	Schenke, 2020[334]
Patient characteristics	*Age, mean: 48.6* *Gender (female to male ratio): 113: 27* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Secondary care *Country*: *Inclusion criteria*: presence of thyroid nodules ≤ 10 mm measured in B-mode ultrasound (independent of the histopathological size of the nodule), Kwak-TIRADS classification of the thyroid nodule during ultrasound investigation, and available histopathological results after surgery. *Exclusion criteria*: incidental thyroid cancers detected at final histology and TNs > 10 mm measured with ultrasound.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* B mode US, using 8-13 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 76* *benign n= 69* *Index test 1:Kwak TIRADS 4A and higher (+ve for malignancy)* TP: 76 FN: 0 FP: 60 TN: 9 *sensitivity*: 1.0 *specificity*: 0.130

Reference	Schenke, 2020[334]
	Index test 2:Kwak TIRADS 4B and higher (+ve for malignancy) TP: 76 FN: 0 FP: 41 TN: 28 *sensitivity*: 1.0 *specificity*: 0.406 *Index test 3:Kwak TIRADS 4C and higher (+ve for malignancy)* TP: 74 FN: 2 FP: 31 TN: 38 *sensitivity*: 0.974 *specificity*: 0.551 *Index test 4:Kwak TIRADS 5 (+ve for malignancy)* TP: 17 FN: 59 FP: 4 TN: 65 *sensitivity*: 0.224 *specificity*: 0.942 *Index test 5:ACR TIRADS TR2 and higher (+ve for malignancy)* TP: 76 FN: 0 FP: 65 TN: 4 *sensitivity*: 1.0 *specificity*: 0.058 *Index test 6:ACR TIRADS TR3 and higher (+ve for malignancy)* TP: 76 FN: 0 FP: 57 TN: 12 *sensitivity*:1.0 *specificity*: 0.174 *Index test 7:ACR TIRADS TR4 and higher (+ve for malignancy)* TP: 76 FN: 0 FP: 41 TN: 28 *sensitivity*: 1.0 *specificity*: 0.406 *Index test 8:ACR TIRADS TR5 (+ve for malignancy)* TP: 53 FN: 23 FP: 23 TN: 46 *sensitivity*: 0.697 *specificity*: 0.667 *Index test 9:EU TIRADS 3 and higher (+ve for malignancy)* TP: 76 FN: 0 FP: 69 TN: 0 *sensitivity*: 1.0 *specificity*: 0.0 *Index test 10:EU TIRADS 4 and higher (+ve for malignancy)* TP: 75 FN: 1 FP: 46 TN: 23 *sensitivity*: 0.987 *specificity*: 0.333 *Index test 11:EU TIRADS 5 (+ve for malignancy)* TP: 74 FN: 2 FP: 35 TN: 34 *sensitivity*: 0.974 *specificity*: 0.493
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.

Reference	**Schenke, 2020[334]**
Comments	

1

Reference	**Peccin, 2002[306]**
Study type	Prospective
Number of patients	n = 80
Patient characteristics	*Age, mean (SD): 45.3 (16.8)* *Gender (female to male ratio): 64:16* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: Secondary care *Country*: Brazil *Inclusion criteria*: patients with palpable solitary thyroid nodules or multinodular goitres with a dominant nodule operated on for clinical and/or cytological suspicion of malignancy or for symptoms of compression *Exclusion criteria*: Patients with multinodular goitre without a dominant nodule, patients with toxic nodules, and patients whose thyroid nodules did not meet clinical or cytological criteria for surgery
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey scale US, using MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	Peccin, 2002[306]
Results	*malignant n= 16* *benign n= 64* *Index test 1: microcalcifications (+ve for malignancy)* TP: 9 FN:7 FP:4 TN: 60 *sensitivity*:0.563 *specificity*: 0.938 *Index test 2: hypoechogenicity (+ve for malignancy)* TP: 7 FN:9 FP:11 TN: 53 *sensitivity*:0.438 *specificity*: 0.828 *Index test 3: absent halo (+ve for malignancy)* TP: 9 FN:7 FP:13 TN: 51 *sensitivity*:0.563 *specificity*: 0.797 *Index test 4: solid content (+ve for malignancy)* TP: 11 FN:5 FP:30 TN: 34 *sensitivity*:0.688 *specificity*: 0.531 *Index test 5: microcalcifications AND hypoechoicity (+ve for malignancy)* TP: 5 FN:11 FP:2 TN: 62 *sensitivity*:0.31 *specificity*: 0.97 *Index test 6: microcalcifications AND absent halo (+ve for malignancy)* TP: 6 FN:10 FP:2 TN: 62 *sensitivity*:0.38 *specificity*: 0.97 *Index test 7: absent halo AND hypoechoicity (+ve for malignancy)* TP: 5 FN:11 FP:7 TN: 57 *sensitivity*:0.31 *specificity*: 0.89 *Index test 8: microcalcifications AND hypoechoicity AND absent halo (+ve for malignancy)* TP: 4 FN:12 FP:2 TN: 62 *sensitivity*:0.25 *specificity*: 0.97 *Index test 9: microcalcifications OR hypoechoicity (+ve for malignancy)* TP: 11 FN:5 FP:13 TN: 51 *sensitivity*:0.69 *specificity*: 0.80 *Index test 10: microcalcifications OR absent halo (+ve for malignancy)* TP: 12 FN:4 FP:15 TN: 49 *sensitivity*:0.75 *specificity*: 0.77 *Index test 11: absent halo OR hypoechoicity (+ve for malignancy)* TP: 11 FN:5 FP:17 TN: 47 *sensitivity*:0.69 *specificity*: 0.73

Reference	**Peccin, 2002[306]**
	Index test 12: microcalcifications OR hypoechoicity OR absent halo (+ve for malignancy) TP: 13 FN:3 FP:19 TN: 45 *sensitivity*: 0.81 *specificity*: 0.70
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Gray, 2014[117]**
Study type	Prospective
Number of patients	n = 78 patients with 78 nodules
Patient characteristics	*Age, median (range): 51 (17-80)* *Gender (female to male ratio): 63:15* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: Secondary care *Country*: UK *Inclusion criteria*: patients who underwent partial or total thyroidectomy *Exclusion criteria*: ultrasonography report or suitable hard copy of ultrasound images unavailable, or an incomplete report; ultrasonography not originally performed by our head and neck radiologist; no definite thyroid nodule on the ultrasound image; repeat ultrasound for the same patient; follow-up ultrasound for histologically proven thyroid cancer; and pathology report unavailable
Target condition(s)	Thyroid nodule malignancy

Reference	**Gray, 2014[117]**
Index test(s) and reference standard	*Index test* Colour flow doppler US, using MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 31* *benign n= 47* *Index test 1: R staging: R2 and above (+ve for malignancy)* TP: 30 FN:1 FP:44 TN:3 *sensitivity*: 0.968 *specificity*: 0.064 *Index test 1: R staging: R3 and above (+ve for malignancy)* TP: 27 FN:4 FP:33 TN:14 *sensitivity*: 0.871 *specificity*: 0.298 *Index test 1: R staging: R4 and above (+ve for malignancy)* TP: 23 FN:8 FP:9 TN:38 *sensitivity*: 0.742 *specificity*: 0.809 *Index test 1: R staging: R5 (+ve for malignancy)* TP: 9 FN:22 FP:0 TN:47 *sensitivity*: 0.29 *specificity*: 1.0
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Schleder, 2015[337]**
Study type	Prospective
Number of patients	n = 101 patients with 101 nodules

Reference	Schleder, 2015[337]
Patient characteristics	*Age, median (range): 54 (25-82)* *Gender (female to male ratio): 55:46* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown (experienced but unclear if medic) *Setting*: Teaching Hospital *Country*: Germany *Inclusion criteria*: All patients with suspect thyroid nodules; final histopathology examination undertaken; CEUS and B mode US performed *Exclusion criteria*: None reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Colour flow doppler US, using MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 26* *benign n= 75* *Index test 1: heterogenous echo (+ve for malignancy)* TP: 10 FN:16 FP: 35 TN: 40 *sensitivity*: 0.385 *specificity*: 0.533 *Index test 2: Hypoechogenicity (+ve for malignancy)* TP: 10 FN:16 FP: 17 TN: 58 *sensitivity*: 0.385 *specificity*: 0.773

Reference	Schleder, 2015[337]
	Index test 3: complex cystic echo pattern (+ve for malignancy) TP: 4 FN: 22 FP: 16 TN: 59 *sensitivity*: 0.154 *specificity*: 0.786 *Index test 4: hyperechogenic pattern (+ve for malignancy)* TP: 2 FN: 24 FP: 7 TN: 68 *sensitivity*: 0.077 *specificity*: 0.906 *Index test 5: well defined contour (+ve for malignancy)* TP: 9 FN: 17 FP: 53 TN: 22 *sensitivity*: 0.346 *specificity*: 0.293 *Index test 6: marginal vessel in CCDS/PD (+ve for malignancy)* TP: 7 FN: 17 FP: 55 TN: 20 *sensitivity*: 0.269 *specificity*: 0.266 *Index test 7: extensive internal flow in CCDS/PD (+ve for malignancy)* TP: 14 FN: 12 FP: 14 TN: 61 *sensitivity*: 0.538 *specificity*: 0.813 *Index test 7: CEUS: complete wash out (+ve for malignancy)* *TP: 24 FN: 2 FP: 14 TN: 61 sensitivity: 0.92 specificity: 0.813*
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

2

Reference	Liu,2020[238]
Study type	Retrospective
Number of patients	n = 90 nodules from 90 patients
Patient characteristics	*Age, mean (SD): 48(14.2)* *Gender (female to male ratio): 63:27* *Ethnicity*: not reported

Reference	Liu,2020[238]
	Expertise of US tester (medic/non medic/unknown): unknown *Setting*: Teaching hospital *Country*: China *Inclusion criteria*: 18-80 years; patient received US and SWE before surgery; proven by surgery and pathological result was FTC or FA *Exclusion criteria*: nodule mostly cystic; image data incomplete; pathological result of the target nodule was unclear or uncertain when patient had multifarious pathological results
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Colour flow doppler US, using MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 28* *benign n= 62* *Index test 1: solidity (+ve for malignancy)* TP: 25 FN: 3 FP:47 TN: 15 *sensitivity*: 0.893 *specificity*: 0.242 *Index test 2: hypoechoic (+ve for malignancy)* TP: 18 FN: 10 FP: 13 TN: 49 *sensitivity*: 0.643 *specificity*: 0.790 *Index test 3: taller than wide (+ve for malignancy)* TP: 4 FN: 24 FP: 2 TN: 60 *sensitivity*: 0.143 *specificity*: 0.968

Reference	Liu,2020[238]
	Index test 4: lobulated or irregular (+ve for malignancy) TP: 9 FN: 19 FP: 1 TN: 61 *sensitivity*: 0.321 *specificity*: 0.984 *Index test 5: microcalcifications (+ve for malignancy)* TP: 1 FN: 27 FP: 0 TN: 62 *sensitivity*: 0.036 *specificity*: 1.0 *Index test 6: ATA TIRADS category 'low' or higher (+ve for malignancy)* TP: 28 FN: 0 FP: 62 TN: 0 *sensitivity*: 1.0 *specificity*: 0.0 *Index test 7: ATA TIRADS category 'intermediate' or higher (+ve for malignancy)* TP: 20 FN: 8 FP: 14 TN: 48 *sensitivity*: 0.714 *specificity*: 0.774 *Index test 8: ATA TIRADS category 'high' (+ve for malignancy)* TP: 11 FN: 17 FP: 2 TN: 60 *sensitivity*: 0.393 *specificity*: 0.968 *Index test 9: ACR TIRADS 3 and above (+ve for malignancy)* TP: 24 FN: 4 FP: 50 TN: 12 *sensitivity*: 0.857 *specificity*: 0.194 *Index test 10: ACR TIRADS 4 and above (+ve for malignancy)* TP: 20 FN: 8 FP: 14 TN: 48 *sensitivity*: 0.714 *specificity*: 0.774 *Index test 11: ACR TIRADS 5 (+ve for malignancy)* TP: 7 FN: 21 FP: 2 TN: 60 *sensitivity*: 0.250 *specificity*: 0.968 *Index test 12: number of significant (hypoechogenicity, lobulated or irr. margin and microcalcif.) US features – 1 or more (+ve for malignancy)* TP: 23 FN: 5 FP: 18 TN: 44 *sensitivity*: 0.821 *specificity*: 0.710 *Index test 13: number of significant US features (hypoechogenicity, lobulated or irr. margin and microcalcif.) – 2 or more (+ve for malignancy)* TP: 10 FN: 18 FP: 1 TN: 61 *sensitivity*: 0.357 *specificity*: 0.984 *Index test 14: number of significant US features (hypoechogenicity, lobulated or irr. margin and microcalcif.) – 3 or more (+ve for malignancy)* TP: 4 FN: 24 FP: 0 TN: 62 *sensitivity*: 0.143 *specificity*: 1.0

Reference	**Liu,2020[238]**
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Jiang, 2014[172]**
Study type	Retrospective
Number of patients	n = 122 patients with 122 nodules
Patient characteristics	*Age, mean (SD): 45(9.1)* *Gender (female to male ratio): 85:37* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): tester unclear but reviewed by medics *Setting*: Teaching hospital *Country*: China *Inclusion criteria*: Patients undergoing thyroidectomy for previously diagnosed thyroid nodules, who had previously undergone contrast enhanced US. *Exclusion criteria*: None reported
Target condition(s)	Thyroid nodule malignancy

Reference	**Jiang, 2014[172]**
Index test(s) and reference standard	*Index test* Contrast enhanced US, using 7 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 62* *benign n= 60* *Index test 1: Heterogeneous low enhancement pattern (+ve for malignancy)* TP: 60 FN: 2 FP:3 TN: 57 *sensitivity*: 0.968 *specificity*: 0.95
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Shimura, 2005[350]**
Study type	Retrospective
Number of patients	n = 53
Patient characteristics	*Age, mean (SD): unclear* *Gender (female to male ratio): unclear* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medics

Reference	**Shimura, 2005[350]**
	Setting: Secondary care *Country*: Japan *Inclusion criteria*: Patients having surgery for a thyroid nodule with prior grey-scale ultrasonography *Exclusion criteria*: ultrasonograms showing cross-sections of additional nodules or did not show the nodule margin
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Grey-scale US, using >7.5 MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= unclear* *benign n= unclear* *The results below were based on a multiple regression analysis. The raw data were not calculable.* *Index test 1: Jagged border AND hypoechoicity (+ve for malignancy)* *sensitivity*: 0.93 *specificity*: 0.92 *Index test 2: irregular shape AND hypoechoicity (+ve for malignancy)* *sensitivity*: 0.89 *specificity*: 0.92
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias

Reference	**Shimura, 2005[350]**
	Indirectness (QUADAS 2 - applicability): serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**He, 2018[142]**
Study type	Prospective
Number of patients	n = 88 nodules from 83 patients
Patient characteristics	*Age, mean (SD): 46(15.2)* *Gender (female to male ratio): unclear* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Secondary care *Country*: China *Inclusion criteria*: solid or cystic nodules with >50% solid; nodule diameter >6mm; normal tissue around the nodule; no surgery, drug or chemotherapy administered before the operation; thyroid surgery with histopathological results *Exclusion criteria*: none reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Contrast-enhanced US, using 15 MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings

Reference	**He, 2018[142]**
	Time between measurement of index test and reference standard: Not clear
Results	*malignant n= 29* *benign n= 59* *Index test 1: Inhomogeneous, low or equal enhancement (+ve for malignancy)* TP: 23 FN: 6 FP: 5 TN: 54 *sensitivity*: 0.793 *specificity*: 0.915 *Index test 2: elastography – SWV of 2.565 m/s or higher (+ve for malignancy)* TP: 22 FN: 7 FP: 3 TN: 56 *sensitivity*: 0.759 *specificity*: 0.949
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

Reference	**Jiang, 2015[171]**
Study type	Prospective
Number of patients	n = 122 with 122 nodules
Patient characteristics	*Age, mean (SD):* 46 +/- 12 years *Gender (female to male ratio):* 85:37 *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: The 2nd Affiliated Hospital, Xi'an JiaoTong University School of Medicine, Number 157 Xiwu Road, Xincheng District, Xi'an, Shanxi Province 710004, China.

Reference	**Jiang, 2015[171]**
	Country: China *Inclusion criteria*: Patients who underwent surgery for thyroid nodules with calcification *Exclusion criteria*: Not stated
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* 1. Conventional ultrasound 2. Contrast-enhanced ultrasound (CEUS) *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*Malignant nodules n= 49* *Benign nodules n= 73* *Index test 1* (conventional ultrasound): +ve = unclear ('dependent on nodular morphological and blood flow observation') TP: 24 FN: 25 FP: 17 TN: 56 ; *sensitivity*: 0.490 , *specificity*: 0.767 *Index test 2* (CEUS): +ve = inhomogeneous hypo-enhancement TP: 44 FN: 5 FP: 6 TN: 67 ; *sensitivity*: 0.898 , *specificity*: 0.918
Source of funding	Not stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Zhang, 2016[433]**
Study type	Prospective
Number of patients	n = 111 with 145 solid nodules
Patient characteristics	*Age, mean (SD):* Reported only by histopathology outcome: 53.56+/-10.97 (benign group); 42.19+/-13.44 (malignant) *Gender (female to male ratio): 91:20* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Not reported *Country*: Not reported *Inclusion criteria*: At least one of: (1) diagnosed as follicular neoplasm follicular neoplasm, suspicious for malignancy or malignancy by fine needle aspiration cytology (FNAC), (2) the presence of BRAF V600E mutation, (3) Compressive symptoms or cosmetic complaints, (4) a significant increase in volume or a change in its ultrasound features during follow up, (5) diagnosed as non-diagnostic or indeterminate lesions by FNAC but showing two or more suspicious ultrasound criteria. *Exclusion criteria*: (1) the presence of a typical nodular goitre or scintigraphically functional (hot) thyroid nodules, (2) cystic nodules or nodules with egg shell calcifications, (3) incomplete elastography or time intensity curve data acquisition, (4) any condition of hyperthyroidism, heart failure, or severe pulmonary hypertension, or (5) previous adverse reaction to intravenous contrast agents.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index tests* 1. High resolution US (HRUS) 2. Real-time elastography (RTE) 3. Contrast enhanced US (CEUS) 4. HRUS + CEUS 5. elastography *Reference (gold) standard:* Surgical histopathological findings

Reference	**Zhang, 2016[433]**
	Time between measurement of index test and reference standard: Not clear
Results	*Malignant nodules n=63* *Benign nodules n= 82* *Index test 1:* HRUS(+ve = risk score 3 or higher on purpose built risk score*) TP: 47 FN: 16 FP: 10 TN: 72 ; *sensitivity*: 0.746 , *specificity*: 0.878 *Index test 2*: CEUS (+ve = risk score 1.6 or higher on risk score*) TP: 54 FN: 9 FP: 26 TN: 56 ; *sensitivity*: 0.857 , *specificity*: 0.683 *Index test 3:* HRUS + CEUS (+ve = risk score 4.25 or higher on risk score*) TP: 47 FN: 16 FP: 10 TN: 72 ; *sensitivity*: 0.746 , *specificity*: 0.878 *Index test 4:* HRUS ill defined border (+ve) TP: 36 FN: 27 FP: 7 TN: 75 ; *sensitivity*: 0.571 , *specificity*: 0.915 *Index test 5:* HRUS microcalcification (+ve) TP: 32 FN: 31 FP: 9 TN: 73 ; *sensitivity*: 0.508 , *specificity*: 0.890 *Index test 6:* HRUS hypoechoic (+ve) TP: 61 FN: 2 FP: 55 TN: 27 ; *sensitivity*: 0.968 , *specificity*: 0.329 *Index test 7:* HRUS irregular shape (+ve) TP: 34 FN: 29 FP: 15 TN: 67 ; *sensitivity*: 0.539 , *specificity*: 0.817 *Index test 8:* HRUS taller than wide (+ve) TP: 18 FN: 45 FP: 4 TN: 78 ; *sensitivity*: 0.286 , *specificity*: 0.951 *Index test 9:* CEUS TTP ratio <1.15 (+ve) TP: 50 FN: 13 FP: 41 TN: 41 ; *sensitivity*: 0.794 , *specificity*: 0.500 *Index test 10:* CEUS sharpness ratio ≥1.6 (+ve) TP: 25 FN: 38 FP: 14 TN: 68 ; *sensitivity*: 0.397 , *specificity*: 0.829

Reference	**Zhang, 2016[433]**
	Index test 11: CEUS peak ratio ≤1.06 (+ve) TP: 51 FN: 12 FP: 49 TN: 33 ; *sensitivity*: 0.810 , *specificity*: 0.402 *Index test 12:* CEUS ill defined enhancement border (+ve) TP: 37 FN: 26 FP: 9 TN: 73 ; *sensitivity*: 0.587 , *specificity*: 0.890 *Index test 13:* CEUS irregular enhancement shape (+ve) TP: 37 FN: 26 FP: 13 TN: 69 ; *sensitivity*: 0.587 , *specificity*: 0.841 *Index test 14:* CEUS hypo-perfusion (+ve) TP: 28 FN: 35 FP: 20 TN: 62 ; *sensitivity*: 0.444 , *specificity*: 0.756 *Index test 15:* CEUS fast wash-out (+ve) TP: 36 FN: 27 FP: 28 TN: 54 ; *sensitivity*: 0.571 , *specificity*: 0.658 *Index test 15:* elastography ASTERIA 1-4 colour scale: 3 or more (+ve) TP: 46 FN: 17 FP: 10 TN: 72 ; *sensitivity*: 0.73 , *specificity*: 0.878 Derivation of risk score: χ^2 test for significance determined ultrasound features significantly associated with malignancy. These were entered into a logistic regression model. The beta coefficient of each ultrasound feature was summed to derive a risk score.
Source of funding	Supported by grants to X-HW from the National Natural Science Foundation of China (81261120566), Jiangsu Province key medical personnel project (RC2011068) and the Priority Academic Program Development of Jiangsu Higher Education Institutions
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 – applicability)*: none
Comments	

Reference	**Chen, 2016[55]**
Study type	Retrospective
Number of patients	n = 253 patients with 319 thyroid nodules
Patient characteristics	*Age, mean (SD)*: male 44 (13) years; female 43 (11) years *Gender (female to male ratio): 167:86*

Reference	Chen, 2016[55]
	Ethnicity: not reported *Expertise of US tester (medic/non medic/unknown*): not reported *Setting*: not reported *Country*: not reported *Inclusion criteria*: not reported *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index tests* 1. 2DUS (conventional 2 -dimensional ultrasound) 2. CEUS (contrast-enhanced ultrasound) 3. elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Chen, 2016[55]**
Results	*Malignant nodules n= 136* *Benign nodules n= 183* *Index test 1:* Conventional (2D) ultrasound: (+ve = nodule solidity, aspect ratio ≥1, calcification diameter <2 mm, irregular shape and unclear boundaries – unclear how many of these needed to be present) TP: 97 FN: 39 FP: 42 TN: 141 ; *sensitivity*: 0.713 , *specificity*: 0.770 *Index test 2*: CEUS (+ve = based on access speed, peak time, subsidence speed, access manner, peak intensity, evenness, pattern of enhancement and clarity of boundary. Values for these parameters taken to represent a positive test were unclear) TP: 119 FN: 17 FP: 25 TN: 158 ; *sensitivity*: 0.875 , *specificity*: 0.863 *Index test 2*: elastography – SWE at 27.65 kpa or more TP: 115 FN: 21 FP: 29 TN: 154 ; *sensitivity*: 0.8455 , *specificity*: 0.8415
Source of funding	Not reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

Reference	**Jin, 2018[173]**
Study type	Prospective
Number of patients	n = 94, with 94 nodules
Patient characteristics	*Age, mean (SD): 43.5 (4.5)* *Gender (female to male ratio): 48:46*

Reference	Jin, 2018[173]
	Ethnicity: not reported *Expertise of US tester (medic/non medic/unknown*): not reported *Setting*: Zhengzhou Central Hospital Affiliated to Zhengzhou University (Zhengzhou, China) *Country*: China *Inclusion criteria*: confirmed cases (from January 2011 to January 2015) of thyroid nodules by ultrasound *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test:* B-mode ultrasound; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*Malignant nodules n=22* *Benign nodules n= 72* *Index test 1:* B-mode ultrasound (+ve = irregular forms, unclear boundary, the aspect ratio was ≥1, inside low echo, micro-calcification, rear echo reduction, rich blood flow. The resistive index of blood flow was ≥0.7 Unclear how many of these features were required) TP: 17 FN: 5 FP: 10 TN: 62; *sensitivity*: 0.773, *specificity*: 0.861 *Index test 2:* elastography – RCG 0-4 colour scale: 3 and above

Reference	**Jin, 2018[173]**
	TP: 18 FN: 4 FP: 9 TN: 63; *sensitivity*: 0.8182, *specificity*: 0.875
Source of funding	
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

Reference	**Li, 2015[218]**
Study type	Retrospective
Number of patients	n = 73, with 80 nodules
Patient characteristics	*Age, mean (SD):* 39.5 (10.3) *Gender (female to male ratio):* 52:21 *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown (experienced ultrasonographers) *Setting*: Shaanxi Provincial Cancer Hospital Affiliated to Medical School (Xi'an, Shaanxi, China). *Country*: China *Inclusion criteria*: all subjects had small thyroid nodules, difficult to define on conventional ultrasound. *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Contrast enhanced ultrasound (CEUS); elastography *Reference (gold) standard:*

Reference	**Li, 2015[218]**
	Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*Malignant nodules = 50* *Benign nodules = 30* *Index test 1:* CEUS (+ve = inhomogeneous perfusion and whole course low enhancement) TP: 44 FN: 6 FP: 6 TN: 24 ; *sensitivity*: 0.88 , *specificity*: 0.80 *Index test 1:* elastography – RGB 0-4 pt colour scale: 3 and above TP: 47 FN: 3 FP: 3 TN: 27 ; *sensitivity*: 0.94 , *specificity*: 0.90
Source of funding	Not reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1
2
3

Reference	**Liu, 2017[242]**
Study type	prospective
Number of patients	n = 100, with 125 nodules for conventional ultrasound assessment and 122 nodules for contrast ultrasound assessment
Patient characteristics	*Age, mean (SD): 40.26 (8.03)* *Gender (female to male ratio): 67:33* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): physicians with years of experience in ultrasound diagnosis.

Reference	**Liu, 2017[242]**
	Setting: Affiliated Yantai Yuhuangding Hospital of Qingdao University *Country*: China *Inclusion criteria*: patients with thyroid nodules receiving an ultrasonic examination and operation *Exclusion criteria*: Not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* 1. Conventional US 2. Contrast US 3. elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	Liu, 2017[242]
Results	***Conventional ultrasound:*** *Malignant nodules = 55* *Benign nodules = 70* ***Contrast ultrasound**:* *Malignant nodules = 57* *Benign nodules = 65* **Index test 1: Conventional ultrasound** *Index test 1:* hypoechoicity (+ve) TP: 51 FN: 4 FP: 56 TN: 14 ; *sensitivity*: 0.927, *specificity*: 0.200 *Index test 1:* heterogenous echo condition (+ve) TP: 52 FN: 3 FP: 54 TN: 16 ; *sensitivity*: 0.945 , *specificity*: 0.229 *Index test 1:* irregular shape (+ve) TP: 25 FN: 30 FP: 13 TN: 57 ; *sensitivity*: 0.455, *specificity*: 0.814 *Index test 1:* unclear boundary (+ve) TP: 32 FN: 23 FP: 8 TN: 62 ; *sensitivity*: 0.582, *specificity*: 0.886 *Index test 1:* taller than wide (+ve) TP: 12 FN: 43 FP: 6 TN: 64 ; *sensitivity*: 0.218, *specificity*: 0.914 *Index test 1:* microcalcification (+ve) TP: 32 FN: 23 FP: 14 TN: 56 ; *sensitivity*: 0.582, *specificity*: 0.80 *Index test 1:* irregular echo halo (+ve) TP: 29 FN: 26 FP: 10 TN: 60 ; *sensitivity*: 0.527, *specificity*: 0.857

Reference	**Liu, 2017[242]**
	Index test 2: Contrast ultrasound *Index test 1:* unclear boundary (+ve) TP: 41 FN: 16 FP: 16 TN: 49 ; *sensitivity*: 0.719, *specificity*: 0.754 *Index test 1:* irregular shape (+ve) TP: 43 FN: 14 FP: 15 TN: 50 ; *sensitivity*: 0.754, *specificity*: 0.769 *Index test 1:* non enhanced perfusion intensity (+ve) TP: 35 FN: 22 FP: 17 TN: 48 ; *sensitivity*: 0.614, *specificity*: 0.738 *Index test 1:* non homogeneous enhancement (+ve) TP: 34 FN: 23 FP: 23 TN: 42 ; *sensitivity*: 0.596, *specificity*: 0.646 *Index test 1:* perfusion defects (+ve) TP: 38 FN: 19 FP: 14 TN: 51 ; *sensitivity*: 0.667, *specificity*: 0.785 *Index test 1:* elastography – RGB 0-4 colour scale 2 or more (+ve) TP: 54 FN: 1 FP: 67 TN: 3 ; *sensitivity*: 0.981, *specificity*: 0.043 *Index test 1:* elastography – RGB 0-4 colour scale 3 or more (+ve) TP: 34 FN: 21 FP: 11 TN: 59 ; *sensitivity*: 0.618, *specificity*: 0.843 *Index test 1:* elastography – RGB 0-4 colour scale 4 (+ve) TP: 4 FN: 51 FP: 2 TN: 68 ; *sensitivity*: 0.072, *specificity*: 0.971
Source of funding	
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

Reference	**Taj, 2020 [373]**
Study type	Prospective

Reference	**Taj, 2020** [373]
Number of patients	n = 153, with an unknown number of nodules. 113 patients had multiple nodules. It is unclear whether more than one nodule was examined from these patients.
Patient characteristics	*Age, mean (SD): 49 (2.13)* *Gender (female to male ratio): 127:26* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Department of Otorhinolarynogology Khyber Teaching Hospital, Peshawar-Pakistan *Country*: Pakistan *Inclusion criteria*: Patients with thyroid nodules. All nodules were papillary thyroid carcinoma (PTC). Although not stated explicitly, the study appears to have focussed only on PTC. *Exclusion criteria*: Patients who failed to give informed consent, have thyroid abscess and pregnant females
Target condition(s)	Thyroid nodule malignancy (PTC)
Index test(s) and reference standard	The unit of analysis may have been the individual patient rather than the individual nodule. *Index test* ultrasound (no further details reported) *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Taj, 2020** [373]
Results	*Malignant nodules =130* *Benign nodules = 23* *Index test 1:* (+ve = solid echo structure, hypoechogenicity, fine or micro clarification, and ill-defined margin. Unclear how many were required for a positive test) TP: 48 FN: 82 FP: 7 TN: 16; *sensitivity*: 0.369 , *specificity*: 0.696 (Incorrectly reported as Sn 75% and Sp 92%)
Source of funding	None
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Szczepanek-Parulska, 2013**[370]
Study type	Prospective
Number of patients	n = 122 with 393 nodules
Patient characteristics	*Age, mean (SD): 51 (13.6)* *Gender (female to male ratio): 19:103* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown 'experienced sonographers' *Setting*: not reported *Country*: Poland *Inclusion criteria*: patients with diagnosed TND admitted for thyroidectomy

Reference	**Szczepanek-Parulska, 2013[370]**
	Exclusion criteria: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* 1. Conventional ultrasound 2. Power doppler (PD) 3. elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*Malignant: 22 nodules in 22 patients* *Benign: 371 nodules in 100 patients* **Index test 1: Conventional US** Hypoechogenicity (+ve) TP: 21 FN: 1 FP: 244 TN: 127 ; *sensitivity*: 0.955 , *specificity*: 0.341 Microcalcifications (+ve) TP: 9 FN: 13 FP: 67 TN: 304 ; *sensitivity*: 0.429 , *specificity*: 0.819 Macrocalcifications (+ve) TP: 5 FN: 17 FP: 29 TN: 342 ; *sensitivity*: 0.227 , *specificity*: 0.923 Diffuse margins (+ve)

Reference	**Szczepanek-Parulska, 2013[370]**
	TP: 16 FN: 6 FP: 98 TN: 273 ; *sensitivity*: 0.727 , *specificity*: 0.736 Taller than wide (+ve) TP: 6 FN: 16 FP: 28 TN: 343 ; *sensitivity*: 0.273 , *specificity*: 0.925 Solid composition (+ve) TP: 19 FN: 3 FP: 220 TN: 151 ; *sensitivity*: 0.864 , *specificity*: 0.408 Solitary nodule (+ve) TP: 7 FN: 15 FP: 21 TN: 350 ; *sensitivity*: 0.318 , *specificity*: 0.943 **Index test 2: Power doppler** Pattern 4 (+ve) TP: 7 FN: 15 FP: 25 TN: 346 ; *sensitivity*: 0.308 , *specificity*: 0.933 **Elastography** Rago 5 pt scale: 2 or more TP: 21 FN: 1 FP: 170 TN: 201 ; *sensitivity*: 0.955 , *specificity*: 0.543
Source of funding	Not reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1
2

Reference	**Skowronska, 2018[355]**
Study type	prospective
Number of patients	n = 52, with 140 nodules
Patient characteristics	*Age, mean (SD): 55 (14)* *Gender (female to male ratio): 44:8* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown [certified ultrasonographers (certificate of Polish Ultrasound Society) with two and 15 years of experience in thyroid US]. *Setting*: Academic referral centre. *Country*: Poland *Inclusion criteria*: 1) preoperative US reassessment of the neck performed by two certified ultrasonographers; 2) preoperative FNAB of dominant or suspicious lesion; 3) surgical thyroid resection with histological evaluation. *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Hitachi Avius Medical ultrasound technique system equipped with a 7.5-12 MHz high-frequency linear array transducer. All images were examined on real-time two-dimensional B-mode grey-scale and Doppler imaging. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Skowronska, 2018[355]**
Results	*Malignant nodules = 8* *Benign nodules = 132* *Index test 1:* EU-TIRADS ≥4 (+ve) TP: 6 FN: 2 FP: 8 TN: 124 ; *sensitivity*: 0.75 , *specificity*: 0.941 *Index test 2:* solidity (+ve) TP: 7 FN: 1 FP: 34 TN: 98 ; *sensitivity*: 0.875 , *specificity*: 0.742 *Index test 3:* hypo/markedly hypoechoic (+ve) TP: 8 FN: 0 FP: 65 TN: 67 ; *sensitivity*: 1.00 , *specificity*: 0.508 *Index test 4:* non-circumscribed margins (+ve) TP: 6 FN: 2 FP: 0 TN: 132 ; *sensitivity*: 0.75 , *specificity*: 1.0 *Index test 5:* microcalcifications (+ve) TP: 3 FN: 5 FP: 2 TN: 130 ; *sensitivity*: 0.375 , *specificity*: 0.984 *Index test 6:* irregular shape (+ve) TP: 8 FN: 0 FP: 0 TN: 132 ; *sensitivity*: 1.0 , *specificity*: 1.0 *Index test 7:* taller than wide (+ve) TP: 5 FN: 3 FP: 1 TN: 131 ; *sensitivity*: 0.625 , *specificity*: 0.992
Source of funding	Not reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

Reference	**Xu, 2014** [412]
Study type	prospective
Number of patients	n = 375, with 441 nodules
Patient characteristics	*Age, mean (SD): 51(11)* *Gender (female to male ratio): 281:94* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown [9 years of experience in thyroid US] *Setting*: tertiary care *Country*: China *Inclusion criteria*: thyroid nodules were detected in a prior conventional US examination and were at least 5 mm in maximal diameter. The indications for thyroid US were as follows: (a) thyroid nodule detected at US in a secondary or junior clinic, (b) discomfort or pressure symptoms in the cervical region, and (c) palpable thyroid nodules at physical examination. *Exclusion criteria*: Nodules were excluded if they had been subjected to prior invasive procedures, signal loss or had no pathological confirmation.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Conventional ultrasound; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	 *Malignant nodules = 116*

Reference	Xu, 2014 [412]
	Benign nodules = 325 ***Index test: Conventional ultrasound*** **Internal nodule component** 1. Solid (+) versus [≤25% cystic component OR 26-50% OR 51-75%] (-) TP: 112 FN: 4 FP: 155 TN: 170 ; *sensitivity*: 0.965 , *specificity*: 0.523 2. [Solid or ≤25% cystic] (+) versus [26-75% cystic] (-) TP: 116 FN: 0 FP: 274 TN: 51 ; *sensitivity*: 1.0 , *specificity*: 0.157 3. [Solid or <50% cystic] (+) versus [51-75% cystic] (-) TP: 116 FN: 0 FP: 300 TN: 25 ; *sensitivity*: 1.0 , *specificity*: 0.77 **Mean nodule size** 1. ≤10mm (+) versus [11-20mm OR ≥21mm] (-) TP: 53 FN:63 FP: 66 TN: 259 ; *sensitivity*: 0.457 , *specificity*: 0.797 2. ≤20mm (+) versus ≥21mm (-) TP: 102 FN:14 FP: 177 TN: 148 ; *sensitivity*: 0.879 , *specificity*: 0.455 **Single/multiple (unit of analysis = patient)** Single nodule (+) versus multiple (-) TP: 18 FN: 88 FP: 57 TN: 212 ; *sensitivity*: 0.170, *specificity*: 0.788 **Echogenicity**

Reference	Xu, 2014 [412]
	1. Markedly hypoechoic (+) versus [moderately hypoechoic OR isoechoic OR mixed echoic OR hyperechoic] (-) TP: 63 FN:53 FP: 15 TN: 310 ; *sensitivity*: 0.543 , *specificity*: 0.954 2. [Markedly OR moderately hypoechoic] (+) versus [Isoechoic OR mixed echoic OR hyperechoic] (-) TP: 108 FN:8 FP: 98 TN: 227 ; *sensitivity*: 0.931 , *specificity*: 0.698 3. [Markedly OR moderately hypoechoic OR isoechoic] (+) versus [mixed echoic OR hyperechoic] (-) TP: 110 FN:6 FP: 151 TN: 174 ; *sensitivity*: 0.948 , *specificity*: 0.535 4. [Markedly OR moderately hypoechoic OR isoechoic OR mixed echoic] (+) versus hyperechoic (-) TP: 114 FN:2 FP: 321 TN: 4 ; *sensitivity*: 0.983 , *specificity*: 0.012 **Shape** Irregular (+) versus regular (-) TP: 58 FN:58 FP: 55 TN: 270 ; *sensitivity*: 0.500 , *specificity*: 0.831 **Margin** Poorly defined (+) versus well defined (-) TP: 61 FN:55 FP: 53 TN: 272 ; *sensitivity*: 0.526 , *specificity*: 0.837 **Calcification** 1. Microcalcification (+) versus [macrocalcification OR eggshell calcification OR no calcification] (-) TP: 61 FN:55 FP: 58 TN: 267 ; *sensitivity*: 0.526 , *specificity*: 0.821 2. [Microcalcification OR macrocalcification] (+) versus [eggshell OR no calcification] (-)

Reference	Xu, 2014 [412]
	TP: 76 FN:40 FP: 73 TN: 252 ; *sensitivity*: 0.655 , *specificity*: 0.775 3. [microcalcification OR macrocalcification OR eggshell calcification] (+) versus no calcification (-) TP: 77 FN:39 FP: 90 TN: 235 ; *sensitivity*: 0.664 , *specificity*: 0.723 **Vascularity** 1. Rich internal flow (+) versus [rare internal flow OR peripheral flow OR no visible flow] (-) TP: 26 FN:90 FP: 71 TN: 254 ; *sensitivity* 0.224 , *specificity*: 0.781 2. [Rich OR rare internal flow] (+) versus [peripheral flow OR no visible flow] (-) TP: 99 FN:17 FP: 274 TN: 51 ; *sensitivity* 0.853 , *specificity*: 0.157 3. [Rich OR rare internal flow OR peripheral flow] (+) versus no visible flow (-) TP: 114 FN:2 FP: 322 TN: 3 ; *sensitivity* 0.983 , *specificity*: 0.009 **Halo** Absent (+) versus present (-) TP: 89 FN:27 FP: 133 TN: 192 ; *sensitivity* 0.767 , *specificity*: 0.591 **Height and width** Taller than wide (+) versus wider than tall (-) TP: 60 FN:56 FP: 14 TN: 311 ; *sensitivity* 0.517 , *specificity*: 0.957 **Contact with capsule**

Reference	Xu, 2014 [412]
	1. >50% of perimeter (+) versus [26-50% OR <25% OR no contact] (-) TP: 23 FN:93 FP: 99 TN: 226 ; *sensitivity* 0.198 , *specificity*: 0.381 2. ≥26% of perimeter (+) versus <25% (-) TP: 71 FN:45 FP: 201 TN: 124 ; *sensitivity* 0.612 , *specificity*: 0.381 3. Any degree of contact (+) versus no contact (-) TP: 100 FN:16 FP: 282 TN: 43 ; *sensitivity* 0.862, *specificity*: 0.132 **Thyroid background at ultrasound (unit of analysis = patient)** Even background (+) versus coarse background (-) TP: 79 FN:27 FP: 234 TN: 35 ; *sensitivity* 0.745 , *specificity*: 0.130 **Elastography** Asteria 1-4 colour scale: 3 or more TP: 74 FN:42 FP: 92 TN: 233 ; *sensitivity* 0.638 , *specificity*: 0.717 VTI I-VI scale: II or more TP: 115 FN: 1 FP: 255 TN: 70 ; *sensitivity* 0.991 , *specificity*: 0.215 VTI I-VI scale: III or more TP: 105 FN: 11 FP: 101 TN: 224 ; *sensitivity* 0.905 , *specificity*: 0.689 VTI I-VI scale: IV or more TP: 92 FN: 24 FP: 18 TN: 307 ; *sensitivity* 0.793 , *specificity*: 0.945 VTI I-VI scale: V or more TP: 37 FN:79 FP: 1 TN: 324 ; *sensitivity* 0.319 , *specificity*: 0.997

Reference	**Xu, 2014** [412]
Source of funding	Supported in part by the Chinese Ministry of Education (grant NCET-06-0723) and Shanghai Talent Development Project from Shanghai Human Resource and Social Security Bureau (grant 2012045).
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1
2
3
4

Reference	**Wang, 2017** [398]
Study type	prospective
Number of patients	n = 1011, with 1011 nodules
Patient characteristics	*Age, mean (SD): 51 (13.7)* *Gender (female to male ratio):768:243* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): two radiologists with 6 and 13 years of experience respectively in thyroid US. *Setting*: not reported *Country*: China *Inclusion criteria*: Patients with US and surgical confirmation *Exclusion criteria*: (a) patients with incomplete US information (103 nodules); (b) nodules with undetermined pathological results (26 nodules).
Target condition(s)	Thyroid nodule malignancy

Reference	Wang, 2017 [398]
Index test(s) and reference standard	*Index test 1* Conventional ultrasound including doppler *Index test 2* TI-RADSs published by Horvath E et al. (TI-RADS H) *Index test 3* TI-RADSs published by Park et al. (TI-RADS P) *Index test 4* TI-RADSs published by Kwak et al. (TI-RADS K) *Index test 4* TI-RADSs published by Russ et al. (TI-RADS R). *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Wang, 2017** [398]
Results	*Malignant = 464* *Benign = 547* ***Index test 1:*** **Conventional ultrasound with doppler** **Composition** 1. Solid (+) versus [predominantly solid OR predominantly cystic OR spongiform] (-) TP: 452 FN: 12 FP: 288 TN: 259 ; *sensitivity*: 0.974 , *specificity*: 0.473 2. [Solid OR predominantly solid] (+) versus [predominantly cystic OR spongiform] (-) TP: 463 FN: 1 FP: 385 TN: 162 ; *sensitivity*: 0.998 , *specificity*: 0.296 3. [Solid OR predominantly solid OR predominantly cystic] (+) versus spongiform (-) TP: 464 FN: 0 FP: 530 TN: 17 ; *sensitivity*: 1.0 , *specificity* 0.031 **Echogenicity** 1. Marked hypoechogenicity (+) versus [hypoechogenicity OR iso-hypoechogenicity] (-) TP: 57 FN: 407 FP: 8 TN: 539 ; *sensitivity*: 0.123 , *specificity* 0.475 2. [Marked hypoechogenicity OR hypoechogenicity] (+) versus iso-hypoechogenicity] (-) TP: 447 FN: 17 FP: 287 TN: 260 ; *sensitivity*: 0.963 , *specificity* 0.475 **Echo structure** Heterogeneous (+) versus homogeneous (-) TP: 329 FN: 135 FP: 447 TN: 100 ; *sensitivity*: 0.709 , *specificity*: 0.183

Reference	Wang, 2017 [398]
	Margin 1. Infiltrative (+) versus ['microlobulated or irregular' OR well circumscribed] (-) TP: 4 FN: 460 FP: 1 TN: 546 ; *sensitivity*: 0.009 , *specificity*: 0.998 2. [Infiltrative OR 'microlobulated or irregular'] (+) versus well circumscribed (-) TP: 330 FN: 134 FP: 75 TN: 472 ; *sensitivity*: 0.711 , *specificity*: 0.863 **Calcification** 1. Hyperechoic spot (HES) (+) versus [mixed calcification OR microcalcification OR macrocalcification OR no calcification] (-) TP: 0 FN: 464 FP: 59 TN: 488 ; *sensitivity*: 0.00 , *specificity*: 0.892 2. [HES OR mixed calcification] (+) versus microcalcification OR macrocalcification OR no calcification] (-) TP: 43 FN: 421 FP: 66 TN: 481 ; *sensitivity*: 0.093 , *specificity*: 0.879 3. [HES OR mixed calcification OR microcalcification] (+) versus [macrocalcification or no calcification] (-) TP: 256 FN: 208 FP: 100 TN: 447 ; *sensitivity*: 0.552 , *specificity*: 0.817 4. [HES OR mixed calcification OR microcalcification OR macrocalcification] (+) versus no calcification (-) TP: 274 FN: 190 FP: 139 TN: 408 ; *sensitivity*: 0.590 , *specificity*: 0.746 **Shape** Taller than wide (+) versus wider than tall (-) TP: 147 FN: 317 FP: 25 TN: 522 ; *sensitivity*: 0.317 , *specificity*: 0.954 **Vascularisation**

Reference	Wang, 2017 [398]
	1. 'Hypervascular or penetrating vessel' (+) versus [hypovascular OR avascular] (-) TP: 64 FN: 400 FP: 109 TN: 438 ; *sensitivity*: 0.138 , *specificity*: 0.801 2. ['Hypervascular or penetrating vessel' OR hypovascular] (+) versus avascular (-) TP: 264 FN: 200 FP: 332 TN: 215 ; *sensitivity*: 0.569 , *specificity*: 0.393 **Halo** 1. Absent (+) versus [partly OR complete fine] (-) TP: 420 FN: 44 FP: 414 TN: 133 ; *sensitivity*: 0.905 , *specificity*: 0.243 2. [Absent OR partly] (+) versus complete fine (-) TP: 424 FN: 40 FP: 440 TN: 107 ; *sensitivity*: 0.914 , *specificity*: 0.196 **Capsule** Present (+) versus absent (-) TP: 19 FN: 445 FP: 87 TN: 460 ; *sensitivity*: 0.041 , *specificity*: 0.841 **Cervical lymph node** Lymphadenopathy (+) versus normal (-) TP: 54 FN: 410 FP: 10 TN: 537 ; *sensitivity*: 0.116 , *specificity*: 0.982 Index test 1: TIRADS H 3 or higher (+ve for malignancy) TP: 464 FN: 0 FP: 480 TN: 67 ; sensitivity: 1.0 , specificity: 0.122

Reference	Wang, 2017 [398]
	Index test 2: TIRADS H 4a or higher (+ve for malignancy) TP: 459 FN: 5 FP: 279 TN: 268 ; sensitivity: 0.989 , specificity: 0.490 Index test 3: TIRADS H 4b or higher (+ve for malignancy) TP: 448 FN: 16 FP: 158 TN: 389 ; sensitivity: 0.966 , specificity: 0.711 Index test 4: TIRADS H 4c or higher (+ve for malignancy) TP: 271 FN: 193 FP: 33 TN: 514 ; sensitivity: 0.584 , specificity: 0.940 Index test 5: TIRADS H 5 (+ve for malignancy) TP: 83 FN: 381 FP: 3 TN: 544 ; sensitivity: 0.179 , specificity: 0.995 Index test 6: TIRADS P 2 or higher (+ve for malignancy) TP: 462 FN: 2 FP: 349 TN: 198 ; sensitivity: 0.996 , specificity: 0.362 Index test 7: TIRADS P 3 or higher (+ve for malignancy) TP: 449 FN: 15 FP: 157 TN: 390 ; *sensitivity*: 0.968 , *specificity*: 0.713 Index test 8: TIRADS P 4 or higher (+ve for malignancy) TP: 387 FN: 77 FP: 76 TN: 471 ; *sensitivity*: 0.834 , *specificity*: 0.861 Index test 9: TIRADS P 5 (+ve for malignancy) TP: 55 FN: 409 FP: 0 TN: 547 ; *sensitivity*: 0.119 , *specificity*: 1.0 Index test 10: TIRADS K 3 or higher (+ve for malignancy) TP: 464 FN: 0 FP: 393 TN: 154 ; *sensitivity*: 1.0 , *specificity*: 0.282 Index test 11: TIRADS P 4a or higher (+ve for malignancy) TP: 460 FN: 4 FP: 260 TN: 287 ; *sensitivity*: 0.991 , *specificity*: 0.525 Index test 12: TIRADS P 4b or higher (+ve for malignancy) TP: 449 FN: 15 FP: 137 TN: 410 ; *sensitivity*: 0.968 , *specificity*: 0.750 Index test 13: TIRADS P 4c or higher (+ve for malignancy) TP: 393 FN: 71 FP: 45 TN: 502 ; *sensitivity*: 0.847 , *specificity*: 0.918

Reference	**Wang, 2017** [398]
	Index test 14: TIRADS P 5 (+ve for malignancy) TP: 48 FN: 416 FP: 3 TN: 544 ; *sensitivity*: 0.103 , *specificity*: 0.995 Index test 15: TIRADS R 3 or higher (+ve for malignancy) TP: 464 FN: 0 FP: 479 TN: 68 ; *sensitivity*: 1.0 , *specificity*: 0.124 Index test 16: TIRADS R 4a or higher (+ve for malignancy) TP: 461 FN: 3 FP: 300 TN: 247 ; *sensitivity*: 0.994 , *specificity*: 0.452 Index test 17: TIRADS R 4b or higher (+ve for malignancy) TP: 419 FN: 45 FP: 86 TN: 461 ; *sensitivity*: 0.903 , *specificity*: 0.843 Index test 18: TIRADS R 5 (+ve for malignancy) TP: 120 FN: 344 FP: 6 TN: 541 ; *sensitivity*: 0.259 , *specificity*: 0.989
Source of funding	Not reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Chng, 2018**[60]
Study type	Retrospective.
Number of patients	n = 150, with 167 nodules
Patient characteristics	*Age, mean (SD): benign: 53.6(13); malignant: 54.4(12.4)* *Gender (female to male ratio): benign: 88.1:11.9; malignant:83.7:16.3* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown [two independent observers with 2–8 years' experience in US-FNA examined historic US records.]

Reference	**Chng, 2018[60]**
	Setting: Secondary care *Country*: Singapore *Inclusion criteria*: people with US prior to thyroid surgery *Exclusion criteria*: Not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test 1* Conventional ultrasound including doppler *Index test 2* **ATA guideline** *Index test 3* **BTA guideline** *Index test 4* **TI-RADS** *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Chng, 2018[60]**
Results	*Malignant = 52* *Benign = 115* *Index test 1:* Solid nodule (+) versus not solid nodule (-) TP: 40 FN: 11 FP: 70 TN: 42 ; *sensitivity*: 0.784 , *specificity*: 0.375 *Index test 2:* Hypoechogenicity (+) versus no hypoechogenicity (-) TP: 36 FN: 15 FP: 32 TN: 80 ; *sensitivity*: 0.706 , *specificity*: 0.714 *Index test 3:* Irregular margin (+) versus not irregular margin (-) TP: 18 FN: 33 FP: 9 TN: 103 ; *sensitivity*: 0.353 , *specificity*: 0.920 *Index test 4:* Microcalcification (+) versus no microcalcification (-) TP: 17 FN: 34 FP: 9 TN: 103 ; *sensitivity*: 0.333 , *specificity*: 0.920 *Index test 5:* Macrocalcification (+) versus no macrocalcification TP: 11 FN: 40 FP: 26 TN: 86 ; *sensitivity*: 0.216 , *specificity*: 0.768 *Index test 6:* Intranodular vascularity (+) versus no macrocalcification TP: 12 FN: 39 FP: 14 TN: 98 ; *sensitivity*: 0.235 , *specificity*: 0.875 *Index test 7:* ATA 'very low suspicion' or higher (+ve for malignancy) TP: 50 FN: 0 FP: 107 TN: 3 ; *sensitivity*: 1.00 , *specificity*: 0.027 *Index test 8:* ATA 'low suspicion' or higher (+ve for malignancy) TP: 49 FN: 1 FP: 91 TN: 19 ; *sensitivity*: 0.980 , *specificity*: 0.173 *Index test 9:* ATA 'intermediate suspicion' or higher (+ve for malignancy) TP: 39 FN: 11 FP: 31 TN: 79 ; *sensitivity*: 0.780 , *specificity*: 0.718 *Index test 10:* ATA 'high suspicion' (+ve for malignancy) TP: 27 FN: 23 FP: 13 TN: 97 ; *sensitivity*: 0.540 , *specificity*: 0.882 *Index test 11:* BTA 'intermediate suspicion' and higher (+ve for malignancy) TP: 45 FN: 5 FP: 54 TN: 56 ; *sensitivity*: 0.900 , *specificity*: 0.509

Reference	**Chng, 2018[60]**
	Index test 12: BTA 'suspicious and higher (+ve for malignancy) TP: 38 FN: 12 FP: 30 TN: 80 ; *sensitivity*: 0.760 , *specificity*: 0.727 *Index test 13:* BTA 'malignant' (+ve for malignancy) TP: 25 FN: 25 FP: 14 TN: 96 ; *sensitivity*: 0.500 , *specificity*: 0.873 *Index test 14:* TIRADS 4A and higher (+ve for malignancy) TP: 48 FN: 3 FP: 79 TN: 31 ; *sensitivity*: 0.941 , *specificity*: 0.282 *Index test 15:* TIRADS 4B and higher (+ve for malignancy) TP: 40 FN: 10 FP: 40 TN: 70 ; *sensitivity*: 0.800 , *specificity*: 0.636 *Index test 16:* TIRADS 4C and higher (+ve for malignancy) TP: 23 FN: 27 FP: 10 TN: 100 ; *sensitivity*: 0.460 , *specificity*: 0.909 *Index test 14:* TIRADS 5 (+ve for malignancy) TP: 1 FN: 49 FP: 0 TN: 110 ; *sensitivity*: 0.02 , *specificity*: 1.0
Source of funding	Not reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

Reference	**Wu, 2020 [405]**
Study type	Retrospective
Number of patients	n = 445.
Patient characteristics	*Age, mean (SD): 48.3 (12.5)* *Gender (female to male ratio): 333:112*

Reference	Wu, 2020 [405]
	Ethnicity: not reported *Expertise of US tester (medic/non medic/unknown*): medics: two experienced head and neck radiologists reviewed all of the US images *Setting*: Not reported *Country*: China *Inclusion criteria*: patients who underwent US for nodular thyroid lesions followed by a neck CT scan within a close interval *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Ultrasound *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Wu, 2020** [405]
Results	*Malignant = 94 patients* *Benign = 351 patients* *Index test 1: Ultrasound (+ve test based on size, internal content, the presence of a spongiform appearance, shape, margin, echotexture, echogenicity of solid portions, and calcification but details not given.)* *Based on patient as unit of analysis, and from reported Sn and Sp:* TP: 60 FN: 34 FP: 28 TN: 323 ; *sensitivity*: 0.64 , *specificity*: 0.92
Source of funding	Not reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

Reference	**Stoian, 2020** [361]
Study type	prospective
Number of patients	n = 261, with 261 nodules.
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): one operator with more than 10 years' experience in conventional US *Setting*: Ultrasound evaluation Unit

Reference	**Stoian, 2020** [361]
	Country: Romania *Inclusion criteria*: patients with a solid nodular goitre examined in an Ultrasound evaluation Unit between January 2016 and June 2018. *Exclusion criteria*: absence of a pathology report.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* conventional ultrasound (2B). Data were also available for volumetric doppler, but these were only reported in aggregation with data for elastography and were not, therefore, extracted. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* less than 2 months before surgery.
Results	*Index test 1: ALL of:* Taller than wide, Sub capsular, Intense hypo echoic, Calcification, Suspect lymph nodes (+ve for malignancy) TP: 31 FN: 26 FP: 11 TN: 193 ; *sensitivity*: 0.544 , *specificity*: 0.946 *Index test 2: ALL of:* Taller than wide, Sub capsular, Intense hypo echoic, Calcification, Suspect lymph nodes OR ALL of hypoechoic, sub-capsular position, inhomogeneity (+ve for malignancy) TP: 51 FN: 6 FP: 104 TN: 100 ; *sensitivity*: 0.895 , *specificity*: 0.490
Source of funding	Not reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Jeong, 2016** [169]
Study type	Retrospective (case-control)
Number of patients	n = 178, with 178 nodules
Patient characteristics	*Age, mean (SD): 46.62 (14.01)* *Gender (female to male ratio): 160:18* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): tester unclear but reviewed by 2 faculty radiologists with 25 and 5 years of experience, respectively, who specialize in thyroid imaging reviewed original (historic) ultrasound findings. *Setting*: not reported *Country*: Korea *Inclusion criteria*: patients with nodular hyperplasia, follicular adenoma or follicular carcinoma with a diagnosis made from surgical specimens between January 2002 and May 2013. *Exclusion criteria*: other follicular pattern lesions, such as the follicular variant of papillary cell carcinoma and Hurthle cell neoplasm, cases with multiple nodules in a lobe or multinodular goiter and cases without preoperative ultrasonography.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Conventional ultrasound and doppler *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Jeong, 2016** [169]
Results	*Benign = 156 (100 nodular hyperplasia and 56 follicular adenoma)* *Malignant = 22 (follicular carcinoma)* *Index test 1:* Maximum tumour diameter, absence of cystic changes and spongiform appearance and presence of peripheral vascularity (+ve for malignancy) [unclear if all had to be present]. TP: 17 FN: 5 FP: 66 TN: 90 ; *sensitivity*: 0.755 , *specificity*: 0.580 *Index test 2:* irregular shape (+ve for malignancy) TP: 1 FN: 21 FP: 2 TN: 154 ; *sensitivity*: 0.045 , *specificity*: 0.987 *Index test 3:* ill margin (+ve for malignancy) TP: 4 FN: 18 FP: 19 TN: 137 ; *sensitivity*: 0.182 , *specificity*: 0.878
Source of funding	Not reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious *Indirectness (QUADAS 2 - applicability)*: serious (Retrospective observational studies may have an inherent bias in that the only people with histopathological findings will be those at the highest level of presumed risk in these studies. This will mean that the population may be altered from what would be expected from the population of people who would normally be tested. Thus, retrospective studies are downgraded for indirectness.
Comments	

1

Reference	**Rivo-Vazquez, 2013**[323]
Study type	Prospective
Number of patients	n = 156
Patient characteristics	*Age, mean (SD): 52 (14.42)* *Gender (female to male ratio): 134:22* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown

Reference	Rivo-Vazquez, 2013[323]
	Setting: Secondary care *Country*: Spain *Inclusion criteria*: Patients on a surgical waiting list for thyroidectomy due to nodular thyroid disease *Exclusion criteria*: Diffuse goitre
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 28* *benign n= 164* *Index test 1: Elastography- Rago 1-5 score 2 or more (+ve for malignancy)* TP: 27 FN: 1 FP: 156 TN: 8 *sensitivity*: 0.964 *specificity*: 0.049 *Index test 1: Elastography- Rago 1-5 score 3 or more (+ve for malignancy)* TP: 21 FN: 7 FP: 89 TN: 75 *sensitivity*: 0.750 *specificity*: 0.457 *Index test 1: Elastography- Rago 1-5 score 4 or more (+ve for malignancy)* TP: 11 FN: 17 FP: 21 TN: 143 *sensitivity*: 0.393 *specificity*: 0.872 *Index test 1: Elastography- Rago 1-5 score 5 (+ve for malignancy)* TP: 2 FN: 26 FP: 2 TN: 162 *sensitivity*: 0.071 *specificity*: 0.988

Reference	**Rivo-Vazquez, 2013[323]**
Source of funding	Mutua Madrilena Foundation
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Zhang, 2013 [424]**
Study type	Prospective
Number of patients	n = 155 patients with 155 nodules
Patient characteristics	*Age, mean (SD): 42.56 (10.23)* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Secondary care *Country*: *Inclusion criteria*: not reported *Exclusion criteria*: cystic nodules with a liquid nature; nodules near the carotid; nodules located in the margins of the thyroid gland; patients with poor breath holding capacity
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Elastography – SWV and SWR

Reference	**Zhang, 2013** [424]
	Reference (gold) standard: Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 62* *benign n= 93* *Index test 1: Shear wave velocity higher than 2.84 m/s(+ve for malignancy)* TP: 60 FN: 2 FP: 4 TN: 89 *sensitivity*: 0.968 *specificity*: 0.957 *Index test 2: Shear wave ratio of 1.32 and higher (+ve for malignancy)* TP: 57 FN: 5 FP: 17 TN: 76 *sensitivity*: 0.919 *specificity*: 0.817
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Yang, 2019** [417]
Study type	Prospective
Number of patients	n = 34 with 51 nodules
Patient characteristics	*Age, mean (SD): 44.5 (15.6)* *Gender (female to male ratio): 23:11* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Secondary care

Reference	Yang, 2019 [417]
	Country: China *Inclusion criteria*: Patients undergoing thyroid surgery *Exclusion criteria*: completely cystic nodes; no thyroid tissue around the node as a control; previous head and neck radiotherapy
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Colour flow doppler US, using MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 11* *benign n= 40* *Index test 1: SWE – 38.3 kPa or higher (+ve for malignancy)* TP: 8 FN: 3 FP: 6 TN: 34 *sensitivity*: 0.727 *specificity*: 0.85 *Index test 1: RTE – Rago 1-5 score: 4 or higher (+ve for malignancy)* TP: 9 FN: 2 FP: 5 TN: 35 *sensitivity*: 0.818 *specificity*: 0.875
Source of funding	Government grant
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Cantisani, 2015** [44]
Study type	Prospective
Number of patients	n = 50 with 54 nodules
Patient characteristics	*Age, mean (range): 58 (38-78)* *Gender (female to male ratio): 46:4* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: Secondary care *Country*: Italy *Inclusion criteria*: Presence of any thyroid nodule; FNAC and surgery performed during study period *Exclusion criteria*: cystic nodules; coarse calcification; spongiform nodules; pregnancy; heart failure; severe pulmonary hypertension
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Colour flow doppler US, using MHz probe frequency *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 18* *benign n= 36* *Data for operator 1 only:* *Index test 1: axial peri-intranodular elasticity contrast index (ECI) of 3 or more (+ve for malignancy)*

Reference	**Cantisani, 2015** [44]
	Raw data do not fit reported accuracy data *sensitivity*: 0.91; *specificity*: 0.90 *Index test 2: axial intranodular elasticity contrast index (ECI) of 3.07 or more (+ve for malignancy)* Raw data do not fit reported accuracy data *sensitivity*: 0.90; *specificity*: 0.932 *Index test 2: longitudinal intranodular elasticity contrast index (ECI) of 3.03 or more (+ve for malignancy)* Raw data do not fit reported accuracy data *sensitivity*: 0.80; *specificity*: 0.805
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Wang, 2014** [391]
Study type	Prospective
Number of patients	n = 431 patients with 487 microcarcinomas
Patient characteristics	*Age, mean (range): 47.34 (21-80)* *Gender (female to male ratio): 295:136* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Secondary care *Country*: China *Inclusion criteria*: Patients with thyroid nodules of 10mm or less, located on both lobes of the thyroid, subsequently undergoing surgery *Exclusion criteria*: abnormal neck anatomy; mass with eggshell calcifications

Reference	**Wang, 2014** [391]
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 375* *benign n= 112* *Index test 1: Elasticity score of 3 or more [ITOH] (+ve for malignancy)* TP: 300 FN:25 FP: 31 TN:81 *sensitivity*: 0.799 *specificity*:0.723 *Index test 2: Strain ratio of 3.65 or more (+ve for malignancy)* TP: 325 FN: 375 FP: 16 TN: 96 *sensitivity*: 0.866 *specificity*: 0.8527
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Lin, 2018** [232]
Study type	Prospective
Number of patients	n = 80 patients with 100 nodules
Patient characteristics	*Age, mean (SD): 47.79 (7.35)* *Gender (female to male ratio): 50:30*

Reference	**Lin, 2018** [232]
	Ethnicity: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Secondary care *Country*: China *Inclusion criteria*: not reported *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 30* *benign n= 70* *Index test 1: elastography - 0-IV colour grade system [Shuzhen method]: III and above (+ve for malignancy)* TP: 29 FN: 1 FP:5 TN:65 *sensitivity*:0.9667 *specificity*: 0.9286
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none.
Comments	

1

Reference	**Zhuo, 2014** [438]
Study type	Prospective
Number of patients	n = 182 patients (191 nodules)
Patient characteristics	*Age, mean (range): 53.67 (27-83)* *Gender (female to male ratio): 94-88* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: Secondary care *Country*: China *Inclusion criteria*: Patients with thyroid nodules referred for thyroidectomy *Exclusion criteria*: Not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Elastography – acoustic radiation force impulse technique *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 69* *benign n= 122* *Index test 1: ARFI with velocity of 2.545 m/s or higher (+ve for malignancy)*

Reference	**Zhuo, 2014** [438]
	Raw data does not fit accuracy data as the accuracy data is based on a regression *sensitivity*: 0.963 *specificity*: 0.962
Source of funding	This work was supported by grants from Shandong Province Natural Science Foundation, Independent Innovation Foundation of Shandong University and Shandong Scientific Technology, and Population and Family Planning Commission of Shandong Province Science and Technology Research Projects (2010No. 14).
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Wang, 2012** [392]
Study type	Prospective
Number of patients	n = 120 patients with 131 nodules
Patient characteristics	*Age, mean (range): 45.85 (18-70)* *Gender (female to male ratio): 78-41* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Secondary care *Country*: China *Inclusion criteria*: solid lesions in one thyroid lobe *Exclusion criteria*: none reported
Target condition(s)	Thyroid nodule malignancy

Reference	**Wang, 2012** [392]
Index test(s) and reference standard	*Index test* Elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 32* *benign n= 99* *Index test 1: Fukunari pattern 0-4 elastic colour scoring (similar to Asteria): 3 and over (+ve for malignancy)* TP: 25 FN: 7 FP: 17 TN: 82 *sensitivity*: 0.781 *specificity*: 0.828 *Index test 1: Strain ratio of 2.9 and over (+ve for malignancy)* TP: 28 FN: 4 FP: 7 TN: 92 *sensitivity*: 0.875 *specificity*: 0.929
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**El-Hariri, 2014** [87]
Study type	Prospective
Number of patients	n = 72 patients with 84 nodules
Patient characteristics	*Age, mean (SD): 45.84* *Gender (female to male ratio): 43:29* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown

Reference	**El-Hariri, 2014** [87]
	Setting: Secondary care *Country*: Egypt *Inclusion criteria*: A solid nodule in one thyroid lobe *Exclusion criteria*: (1) cystic component >15% of the nodule volume, (2) Large nodules occupying >75% of thyroid lobe volume because insufficient surrounding normal thyroid tissue to be used as reference and (3) nodules with peripheral calcifications.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 25* *benign n= 59* *Index test 1: Elastography 1-4 colour scale (similar to Asteria): 3 and above (+ve for malignancy)* TP: 21 FN: 4 FP: 9 TN: 50 *sensitivity*: 0.84 *specificity*: 0.847 *Index test 2: Strain ratio of 3.5 and above (+ve for malignancy)* TP: 22 FN: 3 FP: 8 TN: 51 *sensitivity*: 0.88 *specificity*: 0.864
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Refaat, 2014** [320]
Study type	Prospective
Number of patients	n = 30 patients with 35 thyroid nodules
Patient characteristics	*Age, mean (SD): 37.8 (people with malignant nodules) and 42.6 (people with benign nodules)* *Gender (female to male ratio): 21:9* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: Secondary care *Country*: Egypt *Inclusion criteria*: Patients with solitary thyroid nodules referred for surgical treatment *Exclusion criteria*: Patients with multiple thyroid nodules (more than two nodules), previous surgery or radioiodine therapy and patients with thyroid nodules who refused or had any contraindication for thyroid surgery; patients with purely cystic (anechoic nodules without solid components) and egg shell-calcified nodules; patients with nodules of greatest diameter larger than 40 mm
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* 7 days
Results	*malignant n= 14* *benign n= 21*

Reference	**Refaat, 2014** [320]
	Index test 1: elastography – Rago 1-5 colour scoring: 3 and above (+ve for malignancy) TP: 14 FN: 0 FP: 6 TN: 15 *sensitivity*: 1.0 *specificity*: 0.714 *Index test 2: elastography – Rago 1-5 colour scoring: 4 and above (+ve for malignancy)* TP: 11 FN: 3 FP: 0 TN: 21 *sensitivity*: 0.786 *specificity*: 1.0 *Index test 3: elastography – strain ratio of 2.2 and above (+ve for malignancy)* TP: 12 FN: 2 FP: 2 TN: 19 *sensitivity*: 0.857 *specificity*: 0.905
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: no serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Garg, 2018** [106]
Study type	Prospective
Number of patients	n = 97 patients with 117 thyroid nodules
Patient characteristics	*Age, mean: 43* *Gender (female to male ratio): 82:15* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Secondary care *Country*: India *Inclusion criteria*: Patients with thyroid nodules having elastography, and surgery

Reference	**Garg, 2018** [106]
	Exclusion criteria: Patients detected to have pure cystic thyroid nodules (anechoic nodules without solid components), or nodules with eggshell calcification; patients with prior diagnosis of thyroid neoplasms, thyroid surgery, radioiodine therapy, or any severe comorbid states
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 33* *benign n= 84* *Index test 1: Elastography using Rago 1-5 colour score: 3 or above (+ve for malignancy)* TP: 29 FN:4 FP: 0 TN: 84 *sensitivity*: 0.8788 *specificity*: 1.0 *Index test 2: Elastography using Rago 1-5 colour score: 4 or above (+ve for malignancy)* TP: 23 FN:10 FP: 0 TN: 84 *sensitivity*: 0.697 *specificity*: 1.0
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Huang, 2015** [156]
Study type	Retrospective

Reference	**Huang, 2015** [156]
Number of patients	n = 136 patients with 155 nodules
Patient characteristics	*Age, mean (SD): 51(12)* *Gender (female to male ratio): 103:33* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Secondary care *Country*: China *Inclusion criteria*: (1) Nodules underwent US, p-SWE and ARFI-induced SE. (2) Nodules were confirmed by histopathology after surgery. *Exclusion criteria*: (1) Maximum diameter of nodule was less than 7 mm. (2) Image data of nodules were not complete: US, p-SWE, or ARFI-induced SE image quality was poor. (3) Mixed cystic (< 50% solid) or almost cystic nodules. (4) There was not enough thyroid tissue surrounding the nodule.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 61* *benign n= 94* *Index test 1: elastography 1-6 scoring method (different to other authors): 2 or more (+ve for malignancy)*

Reference	**Huang, 2015** [156]
	TP: 61 FN: 0 FP:92 TN:2 *sensitivity*: 1.0 *specificity*: 0.021 *Index test 1: elastography 1-6 scoring method (different to other authors): 3 or more (+ve for malignancy)* TP: 55 FN: 6 FP:57 TN:37 *sensitivity*: 0.902 *specificity*: 0.394 *Index test 1: elastography 1-6 scoring method (different to other authors): 4 or more (+ve for malignancy)* TP: 45 FN: 16 FP: 9 TN: 85 *sensitivity*: 0.738 *specificity*: 0.904 *Index test 1: elastography 1-6 scoring method (different to other authors): 5 or more (+ve for malignancy)* TP: 10 FN: 51 FP: 2 TN: 92 *sensitivity*: 0.164 *specificity*: 0.979 *Index test 1: elastography 1-6 scoring method (different to other authors): 6 (+ve for malignancy)* TP: 0 FN: 61 FP: 1 TN: 93 *sensitivity*: 0.0 *specificity*: 0.989 *Index test 1: elastography ARFI SWV of 2.64 m/s or more (+ve for malignancy)* TP: 50 FN: 11 FP: 21 TN: 72 *sensitivity*: 0.82 *specificity*: 0.766 *Index test 1: elastography ARFI SWV of 2.64 m/s or more OR elastography 1-6 scoring method score of 4 or more (+ve for malignancy)* TP: 59 FN: 2 FP: 27 TN: 67 *sensitivity*: 0.967 *specificity*: 0.712
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none.
Comments	

1

Reference	**Sohail, 2020** [357]
Study type	Prospective
Number of patients	n = 157
Patient characteristics	*Age, mean (SD): 37.64(9.44)* *Gender (female to male ratio): 109:48* *Ethnicity*: not reported

Reference	Sohail, 2020 [357]
	Expertise of US tester (medic/non medic/unknown): medic *Setting*: Department of radiology *Country*: Pakistan *Inclusion criteria*: 20-60 years; either gender; solid subcentimetre thyroid nodules suspected to be malignant on conventional US, with later surgical excision of the nodule and histology report *Exclusion criteria*: Diffuse background thyroid lesions including Grave's disease and Hashimoto's thyroiditis. Thyroid lesions on US occupying >75% of thyroid lobe, markedly calcified nodules, and complex nodules with both solid and cystic components.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 27* *benign n= 130* *Index test 1: SWE Elasticity Index of 66 kPa or more (+ve for malignancy)* TP: 22 FN: 5 FP: 10 TN: 120 *sensitivity*: 0.815 *specificity*: 0.923
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none

Reference	**Sohail, 2020** [357]
Comments	

1

2

3

Reference	**Stoian, 2015 #934**
Study type	Prospective
Number of patients	n = 174 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unknown *Setting*: Secondary care *Country*: Romania *Inclusion criteria*: Patients with thyroid nodules who received surgery and histopathology

Reference	**Stoian, 2015 #934**
	Exclusion criteria: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Colour flow doppler US, using MHz probe frequency; elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*malignant n= 29* *benign n= 145* *Index test 1: Russ TIRADS 3 and higher (+ve for malignancy)* TP: 29 FN:0 FP: 130 TN: 15 *sensitivity*: 1.0 *specificity*: 0.103 *Index test 2: Russ TIRADS 4a and higher (+ve for malignancy)* TP: 28 FN:1 FP: 46 TN: 99 *sensitivity*:0.966 *specificity*: 0.683 *Index test 3: Russ TIRADS 4b and higher (+ve for malignancy)* TP: 25 FN:4 FP: 3 TN: 142 *sensitivity*: 0.862 *specificity*: 0.979 *Index test 4: Russ TIRADS 5 (+ve for malignancy)* TP: 16 FN:13 FP: 1 TN: 144 *sensitivity*:0.552 *specificity*: 0.993 Elastography was measured but results not given
Source of funding	None reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: Very serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none.
Comments	

1
2

Reference	**Watkins, 2021[400]**
Study type	Retrospective
Number of patients	n = 212 patients with 218 nodules
Patient characteristics	*Age, mean (SD): 58.5(29)* *Gender (female to male ratio): 161:51* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: Secondary care *Country*: UK *Inclusion criteria*: patients undergoing preoperative thyroid ultrasound with eutopic thyroid histology results available *Exclusion criteria*: ultrasound demonstrating diffuse thyroid disease such as thyroiditis or diffuse multinodular goitre rather than a discrete nodule or if it was not considered possible to reliably correlate imaging and histopathology, due to, for example, suboptimal image quality
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test:* B-mode ultrasound; blinded *Reference (gold) standard:* Surgical histopathological findings

Reference	Watkins, 2021[400]
	Time between measurement of index test and reference standard: Not clear
Results	*Malignant nodules n= 77* *Benign nodules n= 141* *Index test 1: BTA U3 and above [positive] / U2 taken as negative* TP: 76 FN: 1 FP: 101 TN: 40 ; *sensitivity*: 0.987, *specificity*: 0.284 *Index test 2: BTA U4 and above [positive] / U2 + U3 taken as negative* TP: 57 FN: 20 FP: 54 TN: 87 ; *sensitivity*: 0.740, *specificity*: 0.617 *Index test 3: BTA U5 [positive] / U2 -U4 taken as negative* TP: 28 FN: 49 FP: 11 TN: 130 ; *sensitivity*: 0.364, *specificity*: 0.922 *Index test 4: ACR TIRADS TR2 and above [positive] / TR1 taken as negative* TP: 77 FN: 0 FP: 127 TN: 14 ; *sensitivity*: 1.0, *specificity*: 0.099 *Index test 5: ACR TIRADS TR3 and above [positive] / TR1+TR2 taken as negative* TP: 74 FN: 3 FP: 98 TN: 43 ; *sensitivity*: 0.961, *specificity*: 0.305 *Index test 6: ACR TIRADS TR4 and above [positive] / TR1-TR3 taken as negative* TP: 60 FN: 17 FP: 63 TN: 78 ; *sensitivity*: 0.779, *specificity*: 0.553 *Index test 7: ACR TIRADS TR5 [positive] / TR1-TR4 taken as negative*

Reference	**Watkins, 2021[400]**
	TP: 35 FN: 42 FP: 20 TN: 121 ; *sensitivity*: 0.454, *specificity*: 0.858 *Index test 8: AI TIRADS TR2 and above [positive] / TR1 taken as negative* TP: 74 FN: 3 FP: 97 TN: 44 ; *sensitivity*: 0.961, *specificity*: 0.312 *Index test 9: AI TIRADS TR3 and above [positive] / TR1-TR2 taken as negative* TP: 73 FN: 4 FP: 93 TN: 48 ; *sensitivity*: 0.948, *specificity*: 0.340 *Index test 10: AI TIRADS TR4 and above [positive] / TR1-TR3 taken as negative* TP: 57 FN: 20 FP: 57 TN: 84 ; *sensitivity*: 0.740, *specificity*: 0.595 *Index test 10: AI TIRADS TR5 [positive] / TR1-TR4 taken as negative* TP: 38 FN: 39 FP: 26 TN: 115 ; *sensitivity*: 0.494, *specificity*: 0.816
Source of funding	Not reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: serious - retrospective
Comments	

Reference	**Li, 2021[225]**
Study type	Retrospective
Number of patients	n = 78, with 81 nodules
Patient characteristics	*Age, mean (SD): 45.1(17.3) for patients with malignant nodules and 48.6(12.2) for patients with benign nodules* *Gender (female to male ratio): 64:17* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unclear ('sonographers') *Setting*: secondary care *Country*: China *Inclusion criteria*: patients who received a diagnosis of Follicular Thyroid Carcinoma by thorough histopathologic analysis of resected hemi-thyroidectomy or total thyroidectomy specimens from a pathology report database OR patients with a diagnosis of Follicular Thyroid Adenoma *Exclusion criteria*: unavailability of data;
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test:* B-mode ultrasound; *Reference (gold) standard:* Surgical histopathological findings

Reference	Li, 2021[225]
	Time between measurement of index test and reference standard: Not clear
Results	*Malignant nodules n= 28* *Benign nodules n= 53* *Index test 1: Echoicity (hypoechoic or markedly hypoechoic = positive)[hyper or iso echoic deemed negative]* TP: 19 FN: 9 FP: 15 TN: 38 ; *sensitivity*: 0.679, *specificity*: 0.717 *Index test 2: Echo texture (heterogeneous = positive)[homogeneous or predominantly homogeneous deemed negative]* TP: 19 FN: 9 FP: 22 TN: 31 ; *sensitivity*: 0.679, *specificity*: 0.585 *Index test 3: Absent halo (= positive)[existence of halo deemed negative]* TP: 7 FN: 21 FP: 14 TN: 39 ; *sensitivity*: 0.250, *specificity*: 0.736 *Index test 4: cluster of grapes sign (= positive)[none deemed negative]* TP: 16 FN: 12 FP: 15 TN: 38 ; *sensitivity*: 0.571, *specificity*: 0.717 *Index test 5: irregular shape (= positive)[round to oval deemed negative]* TP: 16 FN: 12 FP: 12 TN: 41 ; *sensitivity*: 0.571, *specificity*: 0.774 *Index test 6: ill defined margin (= positive)[clear deemed negative]* TP: 3 FN: 25 FP: 0 TN: 53 ; *sensitivity*: 0.107, *specificity*: 1.00 *Index test 7: Solidity – mostly solid (= positive)[cystic deemed negative]*

Reference	**Li, 2021[225]**
	TP: 25 FN: 3 FP: 21 TN: 32 ; *sensitivity*: 0.893, *specificity*: 0.604 *Index test 8: Any calcifications (= positive)[none deemed negative]* TP: 16 FN: 12 FP: 10 TN: 43 ; *sensitivity*: 0.571, *specificity*: 0.811 *Index test 9: solitary (= positive)[not solitary deemed negative]* TP: 8 FN: 20 FP: 7 TN: 46 ; *sensitivity*: 0.286, *specificity*: 0.868
Source of funding	
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: serious
Comments	

Reference	**Sharma, 2019 [345]**
Study type	Prospective
Number of patients	n = 48, with 48 nodules

Reference	Sharma, 2019 [345]
Patient characteristics	*Age, mean (SD): 36.08(13.9)* *Gender (female to male ratio): 39:9* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): not reported *Setting*: Secondary care *Country*: India *Inclusion criteria*: euthyroid cases of solitary thyroid nodule attending the Department of ENT. For the purpose of inclusion in this study, a solitary thyroid nodule (STN) was defined as a single clinically palpable discrete lesion involving either the lobe or the isthmus of the thyroid gland *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test:* B-mode ultrasound; *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	Sharma, 2019 [345]
Results	*Malignant nodules n=8* *Benign nodules n= 40* *Index test 1: Solidity – solid = positive [cystic deemed negative]* TP: 8 FN: 0 FP: 30 TN: 10 ; *sensitivity*: 1.00, *specificity*: 0.25 *Index test 2:* Suspicious for malignancy based on unreported US characteristics TP: 7 FN: 1 FP: 3 TN:37 ; *sensitivity*: 0.875, *specificity*: 0.925
Source of funding	
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Haskjold, 2021[140]**
Study type	Prospective
Number of patients	n = 101 nodules
Patient characteristics	*Age, mean (SD):56(3) [patients with malignant nodules]; 50(2) [patients with benign nodules]* *Gender (female to male ratio): 82:19* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: secondary care *Country*: Norway *Inclusion criteria*: symptomatic thyroid nodules or incidentally discovered thyroid nodules referred to clinic that had US and surgery *Exclusion criteria*: no histopathology
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test:* B-mode ultrasound; blinded *Reference (gold) standard:* Surgical histopathological findings

Reference	**Haskjold, 2021[140]**
	Time between measurement of index test and reference standard: Not clear
Results	*Malignant nodules n= 37* *Benign nodules n= 64* *Index test 1: A predetermined custom scoring template was used with categories corresponding to the most common histological diagnoses: colloid nodule, adenomatoid colloid nodule, follicular* *adenoma, follicular carcinoma, follicular variant of papillary thyroid carcinoma, papillary thyroid carcinoma, or other thyroid cancer. The operator also indicated the confidence of the assignment to any category on a scale from 1 to 5 (1 = very uncertain, 2 = uncertain, 3 = neutral,* *4 = certain, 5 = very certain). 4 and 5 denoted a positive test [1-3 deemed -ve]* TP: 36 FN: 1 FP: 14 TN:50 ; *sensitivity*: 0.973, *specificity*: 0.781 *Index test 2:* ACR TIRADS 4-5 [2-3 deemed negative] TP: 36 FN: 1 FP: 47 TN: 17 ; *sensitivity*:0.973, *specificity*:0.266
Source of funding	Not reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: no serious risk of bias *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

Reference	**Qi, 2021[314]**
Study type	Retrospective
Number of patients	n = 884, with 1096 nodules
Patient characteristics	*Age, median (range): 43.91 (10-78)* *Gender (female to male ratio): 681:203* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: secondary care *Country*: China *Inclusion criteria*: consecutive patients with thyroid nodules given US and followed up with thyroidectomy surgery *Exclusion criteria*: pregnant and breastfeeding women
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test:* B-mode ultrasound; blinded *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:*

Reference	Qi, 2021[314]
	Not clear
Results	*Malignant nodules n= 414* *Benign nodules n= 682* *Index test 1: ACR TIRADS TR2 or more [TR1 deemed negative]* TP: 414 FN: 0 FP: 642 TN: 40 ; *sensitivity*: 1.0, *specificity*: 0.058 *Index test 2: ACR TIRADS TR3 or more [TR1-2 deemed negative]* TP: 413 FN: 1 FP: 497 TN: 185 ; *sensitivity*: 0.998, *specificity*: 0.271 *Index test 3: ACR TIRADS TR4 or more [TR1-3 deemed negative]* TP: 406 FN: 8 FP: 443 TN: 239 ; *sensitivity*: 0.981, *specificity*: 0.350 *Index test 4: ACR TIRADS TR5 [TR1-4 deemed negative]* TP: 306 FN: 108 FP: 201 TN: 481 ; *sensitivity*: 0.739, *specificity*: 0.705 *Index test 5: Kwak TIRADS 3 or more [2 deemed negative]* TP: 414 FN: 0 FP: 642 TN: 40 ; *sensitivity*: 1.0, *specificity*: 0.058 *Index test 6: Kwak TIRADS 4A or more [2-3 deemed negative]*

Reference	Qi, 2021[314]
	TP: 413 FN: 1 FP: 493 TN: 189 ; *sensitivity*: 0.996, *specificity*: 0.277 *Index test 7: Kwak TIRADS 4B or more [2-4A deemed negative]* TP: 405 FN: 9 FP: 395 TN: 287 ; *sensitivity*: 0.978, *specificity*: 0.421 *Index test 8: Kwak TIRADS 4C or more [2-4B deemed negative]* TP: 372 FN: 42 FP: 259 TN: 423 ; *sensitivity*: 0.899, *specificity*: 0.620 *Index test 9: Kwak TIRADS 5 or more [2-4B deemed negative]* TP: 46 FN: 368 FP: 13 TN: 669 ; *sensitivity*: 0.111, *specificity*: 0.981 *Index test 10: EU TIRADS 3 or more [2 deemed negative]* TP: 414 FN: 0 FP: 642 TN: 40 ; *sensitivity*: 1.0, *specificity*: 0.058 *Index test 11: EU TIRADS 4 or more [2-3 deemed negative]* TP: 405 FN: 9 FP: 439 TN: 243 ; *sensitivity*: 0.978, *specificity*: 0.356 *Index test 12: EU TIRADS 5 [2-4 deemed negative]* TP: 365 FN: 49 FP: 304 TN: 378 ; *sensitivity*: 0.882, *specificity*: 0.554

Reference	Qi, 2021[314]
	Index test 13: C TIRADS 3 or more [2 deemed negative] TP: 414 FN: 0 FP: 638 TN: 44 ; *sensitivity*: 1.0, *specificity*: 0.065 *Index test 14: C TIRADS 4A or more [2-3 deemed negative]* TP: 412 FN: 2 FP: 451 TN: 231 ; *sensitivity*: 0.995, *specificity*: 0.338 *Index test 15: C TIRADS 4B or more [2-4A deemed negative]* TP: 384 FN: 30 FP: 312 TN: 370 ; *sensitivity*: 0.928, *specificity*: 0.542 *Index test 16: C TIRADS 4C or more [2-4B deemed negative]* TP: 272 FN: 142 FP: 121 TN: 561 ; *sensitivity*: 0.657, *specificity*: 0.823 *Index test 17: C TIRADS 5 [2-4C deemed negative]* TP: 9 FN: 405 FP: 0 TN: 682 ; *sensitivity*: 0.022, *specificity*: 1.0 *Index test 18: KTA/KSThR-TIRADS 3 or more [2 deemed negative]* TP: 414 FN: 0 FP: 608 TN: 74 ; *sensitivity*: 1.0, *specificity*: 0.109

Reference	**Qi, 2021[314]**
	Index test 19: KTA/KSThR-TIRADS 4 or more [2-3 deemed negative] TP: 404 FN: 10 FP: 398 TN: 284 ; *sensitivity*: 0.976, *specificity*: 0.416 *Index test 19: KTA/KSThR-TIRADS 5 [2-4 deemed negative]* TP: 338 FN: 76 FP: 227 TN: 455 ; *sensitivity*: 0.816, *specificity*: 0.667
Source of funding	Not reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: serious *Indirectness (QUADAS 2 - applicability)*: serious
Comments	

Reference	**Zhang, 2021[431]**
Study type	Prospective

Reference	**Zhang, 2021[431]**
Number of patients	n = 241 patients with 261 nodules
Patient characteristics	*Age, mean (SD): 46.4(11.34) [people with malignant nodules]; 52.12(10.61[people with benign nodules]* *Gender (female to male ratio): 209:32* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): unclear *Setting*: Secondary care *Country*: China *Inclusion criteria*: (a) patients aged >18 years; (b) patients with no treatment or biopsy examination before ultrasound examinations; and (c) patients with solid or primarily solid (<25% cystic) nodules by conventional ultrasound examinations. *Exclusion criteria*: (a) patients with Hashimoto thyroiditis and patients with nodules more than 3 cm in size (overstepping the maximum coverage of the SWE color); (b) calcification within or around the nodules, which could result in the loss of elastic image information; (c) patients with nodules located in the isthmus or adjacent to the tracheal cartilage and the common carotid artery, as the lateral displacement of the pulsatile flow could cause the nodule hardness to produce artifacts.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test:* Elastography; *Reference (gold) standard:* Surgical histopathological findings

Reference	**Zhang, 2021[431]**
	Time between measurement of index test and reference standard: Not clear
Results	*Malignant nodules n=203* *Benign nodules n= 58* *Index test 1: elastography - red or orange hardest colour [green or blue hardest colour deemed negative]* TP: 160 FN: 43 FP: 12 TN:46 ; *sensitivity*: 0.788, *specificity*: 0.793 *Index test 2: elastography – red, green or orange main colour [blue main colour deemed negative]* TP: 161 FN: 42 FP: 10 TN:48 ; *sensitivity*: 0.793, *specificity*: 0.828 *Index test 3: elastography – stiff rim [no stiff rim deemed negative]* TP: 91 FN: 112 FP: 7 TN:51 ; *sensitivity*: 0.448, *specificity*: 0.879 *Index test 4: elastography – internal colour inhomogeneity [homogeneity deemed negative]* TP: 161 FN: 42 FP: 20 TN:38 ; *sensitivity*: 0.793, *specificity*: 0.655 *Index test 5: Kwak TIRADS 4b or above [1-4a deemed negative]*

Reference	**Zhang, 2021[431]**
	TP: 176 FN: 27 FP: 16 TN:42 ; *sensitivity*: 0.867, *specificity*: 0.724 *Index test 5: Kwak TIRADS 4b or above combined with SWE colour scores (unexplained) [1-4a deemed negative]* TP: 173 FN: 30 FP: 11 TN:47 ; *sensitivity*: 0.852, *specificity*: 0.810
Source of funding	
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

Reference	**Gorgulu, 2021[112]**
Study type	Prospective
Number of patients	n = 123 patients with 144 nodules

Reference	**Gorgulu, 2021[112]**
Patient characteristics	*Age, mean (SD): 45.33(12.47)* *Gender (female to male ratio): 87:36* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medics (experienced radiologists) *Setting*: secondary care *Country*: Turkey *Inclusion criteria*: Patients who were admitted to the otorhinolaryngology and general surgery clinics in Adana Numune Research and Training Hospital (Adana, Turkey) for thyroidectomy according to preoperative clinical evaluation were included; single or multiple nodules ≤40 mm *Exclusion criteria*: The existence of pure cystic lesions, insufficient normal tissue surrounding the measured nodule, isthmic nodules, nodules larger than 40 mm, rough calcification and autoimmune thyroid disease were all exclusion criteria.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test:* B-mode ultrasound; strain elastography *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Gorgulu, 2021[112]**
Results	*Malignant nodules n= 28* *Benign nodules n= 116* *Index test 1: strain ratio >3.59 [SR <=3.59 deemed negative]* TP: 28 FN: 0 FP: 18 TN: 98 ; *sensitivity*: 1.0, *specificity*: 0.844 *Index test 2:* elastography score 3-4 [1-2 deemed negative] TP: 28 FN: 0 FP: 16 TN: 100 ; *sensitivity*: 1.0, *specificity*: 0.862
Source of funding	
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

Reference	**Li, 2021[219]**
Study type	Prospective

Reference	**Li, 2021[219]**
Number of patients	n = XX, with XX nodules
Patient characteristics	*Age, median (range): 48(24-77)* *Gender (female to male ratio): 216:64* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: Secondary care *Country*: China *Inclusion criteria*: (1) the patients underwent thyroid surgery and had pathological results; (2) the patients had not been previously treated for thyroid nodules; and (3) the patients had no history of radiotherapy of the head and neck regions. *Exclusion criteria*: (1) more than 25% of the nodule consisted of the cystic component (because shear waves cannot propagate in liquid); (2) the nodule contained coarse or rim calcifications, which cause information loss in SWE images; (3) the nodule was located in the isthmus or adjacent to the cartilage of the trachea and common carotid artery (because it was difficult to distinguish between actual stiffness and artifacts); and (4) benign and malignant nodules appeared in the same thyroid lobe (because it is difficult to determine the pathological nature of the target nodule).
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test:* B-mode ultrasound; SWE *Reference (gold) standard:* Surgical histopathological findings

Reference	Li, 2021[219]
	Time between measurement of index test and reference standard: Not clear
Results	*Malignant nodules n=272* *Benign nodules n=84* *Index test 1: Hypoechoic [iso- or mixed deemed negative]* TP: 258 FN: 14 FP: 29 TN:55 ; *sensitivity*: 0.949, *specificity*: 0.655 *Index test 2: microcalcification [macro or none deemed negative]* TP: 120 FN: 152 FP: 10 TN:74 ; *sensitivity*: 0.441, *specificity*: 0.881 *Index test 3: minimal to marked vascularity [absent deemed negative]* TP: 135 FN: 137 FP: 52 TN:32 ; *sensitivity*: 0.496, *specificity*: 0.381 *Index test 4: SWE at 36.2 kPa or more [< 36.2 kpa deemed negative]* TP: 207 FN: 65 FP: 18 TN:66 ; *sensitivity*: 0.761, *specificity*: 0.784

Reference	**Li, 2021[219]**
Source of funding	
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

1

2

3

Reference	**Shah, 2020[343]**
Study type	Prospective
Number of patients	n = 50
Patient characteristics	*Age, mean (SD): 44.18(14.29)* *Gender (female to male ratio): 35:15* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: Secondary care *Country*: India

Reference	**Shah, 2020[343]**
	Inclusion criteria: All patients above 12 years of age who were admitted with thyroid swelling in the Department of Surgery were included *Exclusion criteria*: Patients who did not give consent and those who were not willing for investigative procedures were excluded.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test:* B-mode ultrasound; *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*Malignant nodules n=9* *Benign nodules n= 41* *Index test 1: ACR TIRADS of 2 or more [1 deemed negative]* TP: 9 FN: 0 FP: 32 TN:9 ; *sensitivity*: 1.0, *specificity*: 0.220 *Index test 2: ACR TIRADS of 3 or more [1-2 deemed negative]* TP: 8 FN: 1 FP: 15 TN:26 ; *sensitivity*: 0.889, *specificity*: 0.634

Reference	**Shah, 2020[343]**
	Index test 3: ACR TIRADS of 4 or more [1-3 deemed negative] TP: 7 FN: 2 FP: 6 TN: 35 ; *sensitivity*: 0.778, *specificity*: 0.854 *Index test 3: ACR TIRADS of 5 [1-3 deemed negative]* TP: 2 FN: 7 FP: 0 TN: 41 ; *sensitivity*: 0.222, *specificity*: 1.0
Source of funding	
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

Reference	**Cao, 2021[45]**
Study type	Retrospective

Reference	**Cao, 2021[45]**
Number of patients	n = 355, with 388 nodules
Patient characteristics	*Age, mean (SD): 49.7(12.4)* *Gender (female to male ratio): 256:99* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: Secondary care *Country*: China *Inclusion criteria*: Thyroid nodule resection performed for first time; pre-op conventional US and CEUS performed; *Exclusion criteria*: diffuse thyroid disease; prior history of thyroid surgery; no histopathology results (FNAC only); quality of dynamic contrast images was poor; special nodule cases such as cystic and spongiform nodules, uniform hyperechoic nodules of Hashimoto's thyroiditis, multiple predominantly solid nodules and/or predominantly cystic nodules with similar US appearance, uncertainty of internal characteristics because of calcifications, and nodules with snowstorm pattern of microcalcifications.
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test:* B-mode ultrasound; CEUS; blinded *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:*

Reference	Cao, 2021[45]
	Not clear
Results	*Malignant nodules n=233* *Benign nodules n= 155* *Index test 1: C TIRADS 4A or above [3 deemed negative]* TP: 233 FN: 0 FP: 127 TN: 28 ; *sensitivity*:1.0, *specificity*: 0.181 *Index test 2: C TIRADS 4B or above [3-4A deemed negative]* TP: 227 FN: 6 FP: 68 TN: 87 ; *sensitivity*:0.974, *specificity*: 0.561 *Index test 3: C TIRADS 4C or above [3-4B deemed negative]* TP: 184 FN: 49 FP: 38 TN: 117 ; *sensitivity*:0.790, *specificity*: 0.755 *Index test 4: C TIRADS 5 [3-4C deemed negative]* TP: 19 FN: 214 FP: 2 TN: 153 ; *sensitivity*:0.082, *specificity*: 0.987 *Index test 5: CEUS enhancement: Mild, moderate, high enhancement [no enhancement or scant punctate linear enhancement deemed negative]*

Reference	**Cao, 2021[45]**
	TP: 232 FN: 1 FP: 136 TN: 19 ; *sensitivity*:0.996, *specificity*: 0.123 *Index test 5: CEUS rapid wash-out [slow or isochronous deemed negative]* TP: 59 FN: 175 FP: 11 TN: 131 ; *sensitivity*:0.252, *specificity*: 0.923
Source of funding	Not reported
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: serious
Comments	

1

2

3

Reference	**McClean, 2021[268]**
Study type	Retrospective
Number of patients	n = 296, with 308 nodules

Reference	McClean, 2021[268]
Patient characteristics	*Age, mean (SD): 49 (not reported)* *Gender (female to male ratio): 77.3:22.7* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): not all were medics so unclear *Setting*: *Country*: UK *Inclusion criteria*: All US reports for patients who underwent thyroid surgery were reviewed. Prior to 2014, patients were selected for surgery based on clinical assessment and FNA result. From 2014, patients were selected for surgery according to BTA guidelines *Exclusion criteria*: Patients who underwent FNA and surgery without US assessment were not included in the study
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test:* B-mode ultrasound; *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**McClean, 2021[268]**
Results	*Malignant nodules n=135* *Benign nodules n= 173* *Index test 1: BTA U classification U3 or above [U2 deemed negative]* TP: 119 FN: 16 FP: 101 TN:72 ; *sensitivity*: 0.881, *specificity*: 0.416 *Index test 2: BTA U classification U4 or above [U2-3 deemed negative]* TP: 67 FN: 68 FP: 21 TN:152 ; *sensitivity*: 0.496, *specificity*: 0.878 *Index test 3: BTA U classification U5 [U2-4 deemed negative]* TP: 40 FN: 95 FP: 7 TN:166 ; *sensitivity*: 0.296, *specificity*: 0.960 TIRADS data also evaluated but unclear which TIRADS scale was used.
Source of funding	
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: none

Reference	**McClean, 2021[268]**
Comments	

Reference	**Kuru, 2021[204]**
Study type	Retrospective
Number of patients	n =1122 patients, with 1143 nodules
Patient characteristics	*Age, median (range): 49 (18-87)* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: Secondary care *Country*: Turkey *Inclusion criteria*: Patients undergoing US, FNA and thyroidectomy *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy

Reference	**Kuru, 2021[204]**
Index test(s) and reference standard	*Index test:* B-mode ultrasound; *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear
Results	*Malignant nodules n=516* *Benign nodules n= 627* *Index test 1: ACR TIRADS 2 and higher* TP: 516 FN: 0 FP: 604 TN: 23 ; *sensitivity*: 1.00, *specificity*: 0.037 *Index test 2: ACR TIRADS 3 and higher* TP: 508 FN: 8 FP: 520 TN: 107 ; *sensitivity*: 0.984, *specificity*: 0.171 *Index test 3: ACR TIRADS 4 and higher* TP: 381 FN: 135 FP: 207 TN: 420 ; *sensitivity*: 0.738, *specificity*: 0.669

Reference	Kuru, 2021[204]
	Index test 4: ACR TIRADS 5 TP: 148 FN: 368 FP: 18 TN: 609 ; *sensitivity*: 0.287, *specificity*: 0.971 *Index test 5: AACE/ACE/AME moderate or higher* TP: 515 FN: 1 FP: 594 TN: 33 ; *sensitivity*: 0.998, *specificity*: 0.053 *Index test 6: AACE/ACE/AME High* TP: 215 FN: 301 FP: 56 TN: 571 ; *sensitivity*: 0.417, *specificity*: 0.911 *Index test 7: EU TIRADS 3 or higher* TP: 516 FN: 0 FP: 608 TN: 19; *sensitivity*: 1.0, *specificity*: 0.030 *Index test 8: EU TIRADS 4 or higher* TP: 372 FN: 144 FP: 186 TN: 441; *sensitivity*: 0.721, *specificity*: 0.703 *Index test 9: EU TIRADS 5* TP: 215 FN: 301 FP: 56 TN: 571; *sensitivity*: 0.416, *specificity*: 0.911

Reference	Kuru, 2021[204]
	Index test 10: ATA TIRADS Very low suspicion or higher TP: 516 FN: 0 FP: 619 TN: 8; *sensitivity*: 1.0, *specificity*: 0.013 *Index test 11: ATA TIRADS low suspicion or higher* TP: 489 FN: 27 FP: 476 TN: 151; *sensitivity*: 0.947, *specificity*: 0.241 *Index test 12: ATA TIRADS intermediate suspicion or higher* TP: 357 FN: 159 FP: 164 TN: 463; *sensitivity*: 0.692, *specificity*: 0.738 *Index test 13: ATA TIRADS high suspicion* TP: 158 FN: 358 FP: 24 TN: 603; *sensitivity*: 0.306, *specificity*: 0.962 *Index test 13: K TIRADS 3 or higher* TP: 516 FN: 0 FP: 604 TN: 23; sensitivity: 1.0, specificity: 0.037 *Index test 14: K TIRADS 4 or higher* TP: 358 FN: 158 FP: 167 TN: 460; sensitivity: 0.694, specificity: 0.734 *Index test 15: K TIRADS 5* TP: 159 FN: 357 FP: 22 TN: 605; sensitivity: 0.308, specificity: 0.965

Reference	**Kuru, 2021[204]**
Source of funding	
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

Reference	**Paker, 2021 #1814[296]**
Study type	Retrospective
Number of patients	n = 216, with 238 nodules
Patient characteristics	*Age, mean (SD): 50.0(12.4)[benign nodule group]; 50.7(16.7)[malignant nodule group)* *Gender (female to male ratio): 171:45* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: secondary care *Country*: Israel

Reference	**Paker, 2021 #1814[296]**
	Inclusion criteria: One inclusion criterion was the availability of sufficient data, which could be either a preoperative, detailed ultrasound report, containing all the sonographic features included in the aforementioned risk-stratification system or recorded preoperative pictures of the nodules on the ultrasound hard disc with sufficient details to determine the exact classification in each of the risk-stratification systems. Other inclusion criteria were postoperative surgical pathology of a differentiated thyroid carcinoma and the presence of 1–3 nodules in the excised lobe. *Exclusion criteria*: thyroidectomies performed due to non-thyroid disease and non-differentiated thyroid lesions (lymphoma, anaplastic, amyloidosis).
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test:* B-mode ultrasound (blind); *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	**Paker, 2021 #1814[296]**
Results	*Malignant nodules n=115* *Benign nodules n= 123* *Index test 1: ACR TIRADS 2 or higher* TP: 115 FN: 0 FP: 111 TN: 12 ; *sensitivity*: 1.0, *specificity*: 0.097 *Index test 2: ACR TIRADS 3 or higher* TP: 111 FN: 4 FP: 62 TN: 61 ; *sensitivity*: 0.965, *specificity*: 0.496 *Index test 3: ACR TIRADS 4 or higher* TP: 107 FN: 8 FP: 31 TN: 92 ; *sensitivity*: 0.930, *specificity*: 0.748 *Index test 4: ACR TIRADS 5* TP: 73 FN: 42 FP: 8 TN: 115 ; *sensitivity*: 0.635, *specificity*: 0.935 *Index test 5: ATA TIRADS Very low or higher [only 222 used ATA]* TP: 106 FN: 0 FP: 106 TN: 10 ; *sensitivity*: 1.00, *specificity*: 0.086

Reference	**Paker, 2021 #1814[296]**
	Index test 6: ATA TIRADS Low or higher [only 222 used ATA] TP: 103 FN: 3 FP: 65 TN: 51 ; *sensitivity*: 0.972, *specificity*: 0.439 *Index test 6: ATA TIRADS intermediate or higher [only 222 used ATA]* TP: 95 FN: 11 FP: 26 TN: 90 ; *sensitivity*: 0.896, *specificity*: 0.776 *Index test 6: ATA TIRADS high [only 222 used ATA]* TP: 70 FN: 36 FP: 6 TN: 110 ; *sensitivity*: 0.660, *specificity*: 0.948
Source of funding	
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious *Indirectness (QUADAS 2 - applicability)*: none
Comments	

Reference	**Hekimsoy, 2021 #1826[144]**
Study type	Retrospective

Reference	**Hekimsoy, 2021 #1826[144]**
Number of patients	n = 165, with 251 nodules
Patient characteristics	*Age, mean (SD): 49.64 (13.50)* *Gender (female to male ratio): 131:34* *Ethnicity*: not reported *Expertise of US tester (medic/non medic/unknown*): medic *Setting*: Secondary care *Country*: Turkey *Inclusion criteria*: Patients who had undergone 7660 detailed US examinations of the thyroid gland during a 5-year period was obtained to compose a study population with histopathologically evaluated thyroid nodules. *Exclusion criteria*: not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test:* B-mode ultrasound; *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	Hekimsoy, 2021 #1826[144]
Results	*Malignant nodules n=62* *Benign nodules n= 189* *Index test 1: EU TIRADS 3 or more* TP: 62 FN: 0 FP: 186 TN: 3 ; *sensitivity*: 1.00, *specificity*: 0.016 *Index test 2: EU TIRADS 4 or more* TP: 45 FN: 17 FP: 38 TN: 151 ; *sensitivity*: 0.726, *specificity*: 0.799 *Index test 3: EU TIRADS 5* TP: 32 FN: 30 FP: 7 TN: 182 ; *sensitivity*: 0.516, *specificity*: 0.963 *Index test 4: ACR TIRADS 2 or more* TP: 62 FN: 0 FP: 186 TN: 3 ; sensitivity: 1.00, specificity: 0.016 *Index test 5: ACR TIRADS 3 or more* TP: 58 FN: 4 FP: 144 TN: 45 ; sensitivity: 0.935, specificity: 0.238

Reference	Hekimsoy, 2021 #1826[144]
	Index test 6: ACR TIRADS 4 or more TP: 44 FN: 18 FP: 47 TN: 142 ; sensitivity: 0.709, specificity: 0.751 *Index test 7: ACR TIRADS 5* TP: 25 FN: 37 FP: 4 TN: 185 ; sensitivity: 0.403, specificity: 0.979 *Index test 8: Solid or almost completely solid* TP: 57 FN: 5 FP: 150 TN: 39 ; sensitivity: 0.919, specificity: 0.206 *Index test 9: Hypo or very hypoechoic* TP: 40 FN: 22 FP: 31 TN: 158 ; sensitivity: 0.645, specificity: 0.836 *Index test 10: Taller than wide* TP: 17 FN: 45 FP: 2 TN: 187 ; sensitivity: 0.274, specificity: 0.989 *Index test 11: Lobulated or irregular margins* TP: 29 FN: 33 FP: 2 TN: 187 ; sensitivity: 0.468, specificity: 0.989 *Index test 12: Microcalcifications* TP: 11 FN: 14 FP: 3 TN: 48 ; sensitivity: 0.44, specificity: 0.941
Source of funding	
Limitations	*Risk of bias (QUADAS 2 – risk of bias)*: very serious

Reference	**Hekimsoy, 2021 #1826[144]**
	Indirectness (QUADAS 2 - applicability): none
Comments	

1

D.2 Threshold of size and classification of thyroid nodules

Study	**Rozenbaum, 2021[327]**
Study type	Non-randomised study
Number of studies (number of participants)	1 (n=80)
Countries and setting	Conducted in France; Setting: unclear
Line of therapy	Not applicable
Duration of study	At least one year of follow up, but mean follow up was 53 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	NA
Subgroup analysis within study	Not applicable
Inclusion criteria	Age of 18 or older; EU TIRADS 5 nodules < or equal to 10mm in the largest diameter; at least 2 sequential US examinations; patient willing to avoid surgery and having understood the principles and constraints of active surveillance.
Exclusion criteria	extra thyroidal extension; suspicious latero-cervical lymph nodes by neck ultrasound;
Recruitment/selection of patients	Retrospective collection of data
Age, gender and ethnicity	Age – 53.3 (13) at discovery of nodule. Gender (M:F): 11:69. Ethnicity: not reported.
Further population details	
Indirectness of population	No indirectness
Interventions	No interventions given. The comparators were different US characteristics and nodule sizes.

Funding	Funding not stated
RESULTS Protocol outcome 1: local cancer progression - Actual outcome: volumetric progression (>50% increase in size). The participants, who were all on active surveillance until they reached the end point of the study by virtue of being treated or having FNA/surgery, were divided into two groups according to the outcome of volumetric progression (>50% increase in size) or no such progression. The US characteristics and size characteristics have been compared between these groups (volumetric progression [n=28] and no volumetric progression [n=52], as follows: Baseline nodule volume(cm3): volumetric progression: 0.045(0.047); no volumetric progression: 0.074(0.084) Baseline nodule diameter (mm): volumetric progression: 4.9(2.0); no volumetric progression: 5.6(2.1) microcalcifications: volumetric progression: 4/28; no volumetric progression: 7/52 hypoechogenicity: volumetric progression: 5/28; no volumetric progression: 11/52 irregular margins: volumetric progression: 24/28; no volumetric progression: 39/52 irregular shape: volumetric progression: 20/28; no volumetric progression: 38/52 EU TIRADS criteria: 1: volumetric progression: 8/28; no volumetric progression: 17/52 EU TIRADS criteria: 2: volumetric progression: 15/28; no volumetric progression: 27/52 EU TIRADS criteria: 3: volumetric progression: 5/28; no volumetric progression: 8/52 EU TIRADS criteria: 4: volumetric progression: 0/28; no volumetric progression: 0/52 US vascularity – none: volumetric progression: 17/28; no volumetric progression: 19/52 US vascularity – peripheral only: volumetric progression: 5/28; no volumetric progression: 14/52 US vascularity – central component only: volumetric progression: 3/28; no volumetric progression: 15/52	
Protocol outcomes not reported by the study	Mortality; quality of life; incidence of distant metastases; decision to treat; adverse events

1

2

Appendix E QUADAS2 risk of bias assessment

E.1 Diagnostic accuracy of Ultrasound

Table 16: Summary of QUADAS2 risk of bias assessment

Study	Patient selection	Index test with blinding of gold standard test results	Gold standard test with blinding of index test results	Time interval between index and gold standard adequately short (within 1 month)	Overall risk of bias
Abd_Alrahman, 2017[3]	U	U	U	U	Very serious risk of bias
Aggarwal, 1989[7]	U	U	U	U	Very serious risk of bias
Ahmadi, 2019[10]	U	Y	U	U	Very serious risk of bias
Akhaven, 2016[13]	L	U	U	U	Very serious risk of bias
Appetecchia, 2006[18]	U	U	U	U	Very serious risk of bias
Aslan, 2018[21]	U	Y	U	Y	Serious risk of bias
Bakari, 2018[27]	U	U	U	U	Very serious risk of bias
Berni, 2002[33]	U	U	U	U	Very serious risk of bias
Bora Makal, 2021[35]	U	U	U	U	Very serious risk of bias
Borlea, 2020[36]	U	U	U	U	Very serious risk of bias
Cakir, 2011[43]	L	U	U	U	Very serious risk of bias
Cantisani, 2015 [44]	L	Y	U	U	Serious risk of bias
Cao, 2021[45]	U	Y	U	U	Very serious risk of bias
Chen, 2016[55]	U	U	U	U	Very serious risk of bias
Chen, 2019[57]	U	U	U	U	Very serious risk of bias
Chng, 2018[60]	U	U	U	U	Very serious risk of bias
Deng, 2018[77]	U	U	U	U	Very serious risk of bias
Dobruch-Sobczak, 2019[81]	U	Y	Y	U	Serious risk of bias
El-Hariri, 2014 [87]	U	U	U	U	Very serious risk of bias
Garcia-Monco Fernandez, 2018[105]	U	U	U	U	Very serious risk of bias

Study	Patient selection	Index test with blinding of gold standard test results	Gold standard test with blinding of index test results	Time interval between index and gold standard adequately short (within 1 month)	Overall risk of bias
Garg, 2018 [106]	L	U	U	U	Very serious risk of bias
Gao, 2019[104]	U	Y	U	U	Very serious risk of bias
Giammanco, 2002[108]	U	U	U	U	Very serious risk of bias
Goldfarb, 2011[267]	U	U	U	U	Very serious risk of bias
Gorgulu, 2019[111]	L	U	U	U	Very serious risk of bias
Gorgulu, 2021[112]	U	Y	U	U	Very serious risk of bias
Gray, 2014[117]	U	Y	U	U	Very serious risk of bias
Gu, 2012[121]	L	U	U	U	Very serious risk of bias
Hang, 2018[139]	U	Y	U	U	Very serious risk of bias
Haskjold, 2021[140]	L	Y	Y	U	No serious risk of bias
He, 2018[142]	U	U	U	U	Very serious risk of bias
Hekimsoy, 2021[144]	U	Y	U	U	Very serious risk of bias
Hong, 2009[147]	L	U	U	U	Very serious risk of bias
Horvath, 2017[149]	U	Y	U	U	Very serious risk of bias
Huang, 2015 [156]	U	U	U	U	Very serious risk of bias
Huang, 2019[154]	U	U	U	U	Very serious risk of bias
Huang, 2020[155]	L	Y	U	U	Serious risk of bias
Jeong, 2016[169]	U	U	U	U	Very serious risk of bias
Jiang, 2014[172]	U	U	U	U	Very serious risk of bias
Jiang, 2015[171]	U	U	U	U	Very serious risk of bias
Jin, 2018[173]	U	U	U	U	Very serious risk of bias
Kalantari, 2018[175]	U	U	U	U	Very serious risk of bias
Kim, 2008[179]	L	Y	U	N	Serious risk of bias
Kim, 2008[188]	U	Y	U	U	Very serious risk of bias
Kim, 2012[180]	U	U	U	Y	Very serious risk of bias

Study	Patient selection	Index test with blinding of gold standard test results	Gold standard test with blinding of index test results	Time interval between index and gold standard adequately short (within 1 month)	Overall risk of bias
Kim, 2016[189]	L	Y	U	U	Serious risk of bias
Kobayashi, 2005[195]	U	U	U	U	Very serious risk of bias
Kong, 2017[198]	U	U	U	U	Very serious risk of bias
Kuru, 2021[204]	U	U	U	U	Very serious risk of bias
Li, 2015[218]	U	Y	U	U	Very serious risk of bias
Li, 2016[223]	U	U	U	U	Very serious risk of bias
Li, 2017[217]	U	U	U	U	Very serious risk of bias
Li, 2018[220]	U	U	U	U	Very serious risk of bias
Li, 2021[219]	U	U	U	U	Very serious risk of bias
Li, 2021[225]	U	U	U	U	Very serious risk of bias
Lin, 2018 [232]	U	U	U	U	Very serious risk of bias
Liu, 2014[240]	L	U	U	U	Very serious risk of bias
Liu, 2017[242]	U	U	U	U	Very serious risk of bias
Liu, 2019[243]	L	Y	U	U	Serious risk of bias
Liu,2020[238]	U	U	U	U	Very serious risk of bias
Lyshchik, 2007[253]	L	Y	Y	U	No risk of bias
Ma, 2014[256]	L	Y	Y	U	No risk of bias
Ma, 2017[255]	L	U	U	U	Very serious risk of bias
Magri, 2020[260]	U	U	U	U	Very serious risk of bias
Maia, 2011[262]	U	U	U	U	Very serious risk of bias
Maimati, 2016[263]	U	U	U	U	Very serious risk of bias
McClean, 2021[268]	U	Y	U	U	Very serious risk of bias
Mohamed, 2013[271]	U	U	U	U	Very serious risk of bias
Mohey, 2013[275]	L	U	U	U	Very serious risk of bias
Nemec, 2012[284]	L	Y	U	Y	No serious risk of bias

Study	Patient selection	Index test with blinding of gold standard test results	Gold standard test with blinding of index test results	Time interval between index and gold standard adequately short (within 1 month)	Overall risk of bias
Nilakantan, 2007[286]	U	U	U	U	Very serious risk of bias
Pagano, 2021[295]	L	Y	Y	U	No serious risk of bias
Paker, 2021 #1814[296]	U	Y	U	U	Very serious risk of bias
Park, 2012[303]	Y	U	U	U	Very serious risk of bias
Parikh, 2013[300]	U	U	U	U	Very serious risk of bias
Pathirana, 2016[305]	L	U	U	U	Very serious risk of bias
Peccin, 2002[306]	U	Y	Y	U	Serious risk of bias
Pei, 2019[307]	H	Y	U	U	Very serious risk of bias
Phuttharak, 2009[311]	U	Y	U	Y	Serious risk of bias
Qi, 2021[314]	U	Y	Y	U	Serious risk of bias
Ragazzoni, 2012[315]	L	Y	U	U	Serious risk of bias
Rago, 1998[318]	L	Y	U	U	Serious risk of bias
Rago, 2007[317]	U	Y	U	U	Very serious risk of bias
Refaat, 2014 [320]	L	U	Y	Y	No serious risk of bias
Ren, 2015[321]	U	Y	U	U	Very serious risk of bias
Reverter, 2019[322]	U	Y	U	U	Very serious risk of bias
Rivo-Vazquez, 2013[323]	L	Y	U	U	Serious risk of bias
Sancak, 2010[333]	U	U	U	U	Very serious risk of bias
Schenke, 2018[336]	U	Y	U	U	Very serious risk of bias
Schenke, 2020[334]	U	U	U	U	Very serious risk of bias
Schleder, 2015[337]	U	U	U	U	Very serious risk of bias
Shweel, 2013[354]	L	U	U	Y	Serious risk of bias
Shah, 2020[343]	U	U	U	U	Very serious risk of bias
Sharma, 2019 [345]	U	U	U	U	Very serious risk of bias
Shao, 2015[344]	U	U	U	U	Very serious risk of bias

Study	Patient selection	Index test with blinding of gold standard test results	Gold standard test with blinding of index test results	Time interval between index and gold standard adequately short (within 1 month)	Overall risk of bias
Shen, 2019[346]	U	Y	U	U	Very serious risk of bias
Shi, 2020[349]	U	U	U	U	Very serious risk of bias
Shimura, 2005[350]	U	Y	U	U	Very serious risk of bias
Shreyamsa, 2021[352]	U	U	U	U	Very serious risk of bias
Shuzhen, 2012[353]	U	U	U	U	Very serious risk of bias
Skowronska, 2018[355]	L	U	U	U	Very serious risk of bias
Sodagari, 2018[356]	L	U	U	U	Very serious risk of bias
Sohail, 2020 [357]	L	Y	U	U	Serious risk of bias
Stoian, 2015{Stoian, 2015 #934}	U	U	U	U	Very serious risk of bias
Stoian, 2020[361]	U	U	U	U	Very serious risk of bias
Sui, 2016[364]	U	U	U	U	Very serious risk of bias
Swan, 2019[367]	H	U	U	U	Very serious risk of bias
Szczepanek-Parulska, 2013[370]	U	U	U	U	Very serious risk of bias
Taj, 2020[373]	U	U	U	U	Very serious risk of bias
Tan, 2010[375]	U	U	U	U	Very serious risk of bias
Tang, 2017[376]	U	Y	U	U	Very serious risk of bias
Trimboli, 2019[380]	U	Y	U	U	Very serious risk of bias
Tuan,2020[381]	U	U	U	U	Very serious risk of bias
Veyrieres, 2012 [383]	L	U	U	Y	Serious risk of bias
Vorlander, 2010[386]	L	U	U	U	Very serious risk of bias
Wang, 2012 [392]	U	U	U	U	Very serious risk of bias
Wang, 2014 [391]	L	Y	U	U	Serious risk of bias
Wang, 2017[398]	U	Y	U	U	Very serious risk of bias
Wang, 2018[388]	L	U	Y	U	Serious risk of bias
Watkins, 2021[400]	U	Y	U	U	Very serious risk of bias

Study	Patient selection	Index test with blinding of gold standard test results	Gold standard test with blinding of index test results	Time interval between index and gold standard adequately short (within 1 month)	Overall risk of bias
Wu, 2016[404]	U	U	U	U	Very serious risk of bias
Wu, 2020[405]	U	Y	Y	U	Serious risk of bias
Xing, 2011[410]	L	U	U	U	Very serious risk of bias
Xu, 2014[412]	L	Y	U	U	Serious risk of bias
Yang, 2019 [417]	L	U	U	U	Very serious risk of bias
Zhang, 2014[425]	U	U	U	U	Very serious risk of bias
Zhang, 2016[433]	L	U	U	U	Very serious risk of bias
Zhang, 2017[423]	L	U	U	U	Very serious risk of bias
Zhang, 2013 [424]	U	U	U	U	Very serious risk of bias
Zhang, 2021[431]	U	Y	U	U	Very serious risk of bias
Zhuo, 2014 [438]	U	U	U	U	Very serious risk of bias

(a) L=low risk, H=high risk, Y=Yes, N=No, U=unclear, which counts as 'No'.

Appendix F Forest plots Diagnostic accuracy of ultrasound: Coupled sensitivity and specificity forest plots

INDIVIDUAL GREY SCALE CHARACTERISTICS

Figure 3: Sensitivity and specificity of the presence of a 'taller than wide' shape for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abd_Alrahman, 2017 #633	5	1	2	22	0.71 [0.29, 0.96]	0.96 [0.78, 1.00]
Akhavan, 2016 #698	1	2	15	72	0.06 [0.00, 0.30]	0.97 [0.91, 1.00]
Borlea, 2020 #712	15	15	20	83	0.43 [0.26, 0.61]	0.85 [0.76, 0.91]
Deng, 2018 #663	30	13	15	49	0.67 [0.51, 0.80]	0.79 [0.67, 0.88]
Dobruch-Sobczak, 2019 #740	105	91	124	522	0.46 [0.39, 0.53]	0.85 [0.82, 0.88]
Hang, 2018 #788	39	6	138	115	0.22 [0.16, 0.29]	0.95 [0.90, 0.98]
Hekimsoy, 2021 #1826	17	2	45	187	0.27 [0.17, 0.40]	0.99 [0.96, 1.00]
Hong, 2009 #792	12	4	37	92	0.24 [0.13, 0.39]	0.96 [0.90, 0.99]
Kim, 2008 #812	33	5	26	29	0.56 [0.42, 0.69]	0.85 [0.69, 0.95]
Kong, 2017 #827	40	6	39	28	0.51 [0.39, 0.62]	0.82 [0.65, 0.93]
Li, 2016 #838	71	5	353	333	0.17 [0.13, 0.21]	0.99 [0.97, 1.00]
Liu, 2017 #854	12	6	43	64	0.22 [0.12, 0.35]	0.91 [0.82, 0.97]
Liu, 2019 #669	20	5	44	105	0.31 [0.20, 0.44]	0.95 [0.90, 0.99]
Liu, 2020 #851	4	2	24	60	0.14 [0.04, 0.33]	0.97 [0.89, 1.00]
Ma, 2014 #865	41	11	53	67	0.44 [0.33, 0.54]	0.86 [0.76, 0.93]
Ma, 2017 #864	64	13	15	43	0.81 [0.71, 0.89]	0.77 [0.64, 0.87]
Maimaiti, 2016 #655	155	24	201	220	0.44 [0.38, 0.49]	0.90 [0.86, 0.94]
Mohamed, 2013 #672	24	2	4	15	0.86 [0.67, 0.96]	0.88 [0.64, 0.99]
Pagano, 2020 #888	2	0	66	78	0.03 [0.00, 0.10]	1.00 [0.95, 1.00]
Pei, 2019 #898	58	24	60	54	0.49 [0.40, 0.59]	0.69 [0.58, 0.79]
Phuttharak, 2009 #901	0	1	5	25	0.00 [0.00, 0.52]	0.96 [0.80, 1.00]
Ragazzoni, 2012 #903	12	22	28	70	0.30 [0.17, 0.47]	0.76 [0.66, 0.84]
Ren, 2015 #907	35	1	8	30	0.81 [0.67, 0.92]	0.97 [0.83, 1.00]
Ren, 2015 #907b	6	2	19	28	0.24 [0.09, 0.45]	0.93 [0.78, 0.99]
Ren, 2015 #907c	33	2	9	34	0.79 [0.63, 0.90]	0.94 [0.81, 0.99]
Schweel, 2013 #929	11	1	5	49	0.69 [0.41, 0.89]	0.98 [0.89, 1.00]
Shao, 2015 #923	120	7	83	302	0.59 [0.52, 0.66]	0.98 [0.95, 0.99]
Shen, 2019 #924	332	36	441	803	0.43 [0.39, 0.47]	0.96 [0.94, 0.97]
Shi, 2020 #681	12	0	38	288	0.24 [0.13, 0.38]	1.00 [0.99, 1.00]
Skowronska, 2018 #930	5	1	3	131	0.63 [0.24, 0.91]	0.99 [0.96, 1.00]
Sui, 2016 #935	34	9	32	34	0.52 [0.39, 0.64]	0.79 [0.64, 0.90]
Swan, 2019 #938	12	37	67	288	0.15 [0.08, 0.25]	0.89 [0.85, 0.92]
Szczepanek-Parulska, 2013 #940	6	28	16	343	0.27 [0.11, 0.50]	0.92 [0.89, 0.95]
Tuan, 2020 #948	34	3	40	17	0.46 [0.34, 0.58]	0.85 [0.62, 0.97]
Wang, 2017 #964	147	25	317	522	0.32 [0.27, 0.36]	0.95 [0.93, 0.97]
Wu, 2016 #687	198	48	309	415	0.39 [0.35, 0.43]	0.90 [0.86, 0.92]
Xing, 2011 #973	28	13	17	40	0.62 [0.47, 0.76]	0.75 [0.62, 0.86]
Xu, 2014 #975	60	14	56	311	0.52 [0.42, 0.61]	0.96 [0.93, 0.98]
Zhang, 2014 #984	18	1	14	38	0.56 [0.38, 0.74]	0.97 [0.87, 1.00]
Zhang,2016 #991	18	4	45	78	0.29 [0.18, 0.41]	0.95 [0.88, 0.99]

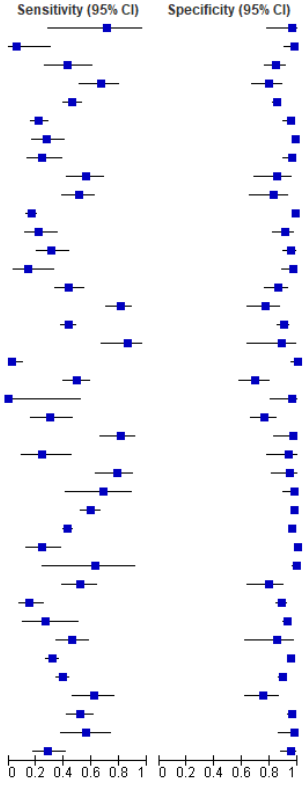
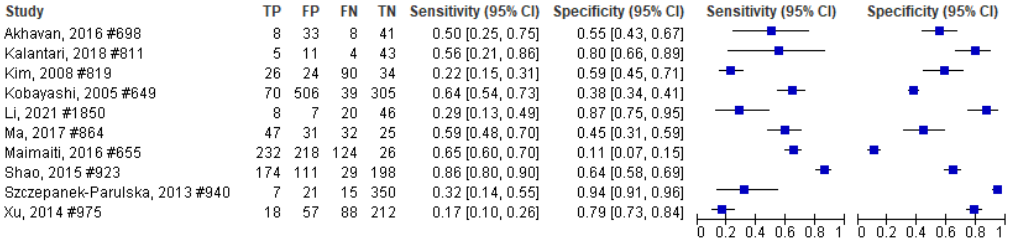
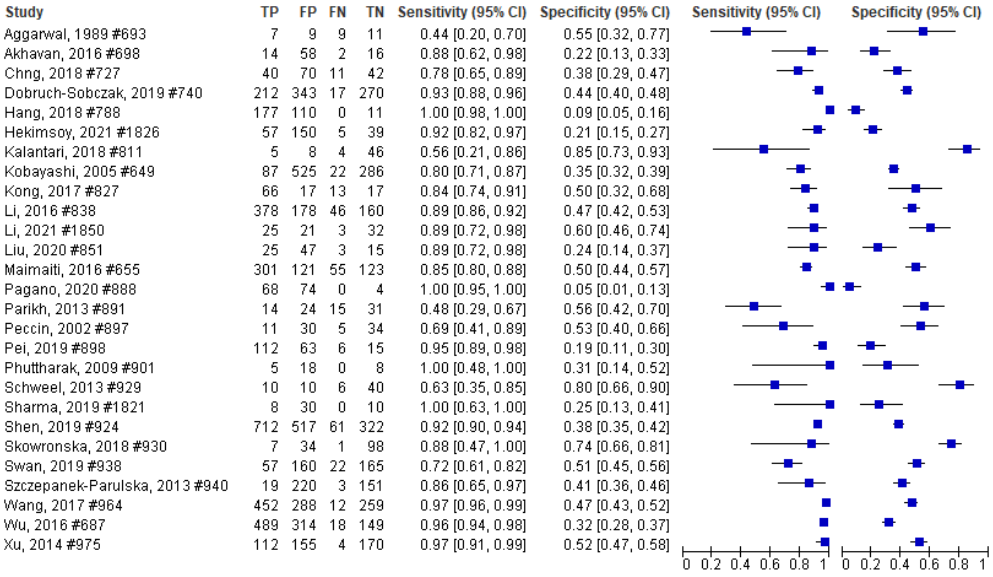


Figure 4: Sensitivity and specificity of the presence of a solitary nodule for diagnosis of malignancy in thyroid cancer



Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Akhavan, 2016 #698	8	33	8	41	0.50 [0.25, 0.75]	0.55 [0.43, 0.67]
Kalantari, 2018 #811	5	11	4	43	0.56 [0.21, 0.86]	0.80 [0.66, 0.89]
Kim, 2008 #819	26	24	90	34	0.22 [0.15, 0.31]	0.59 [0.45, 0.71]
Kobayashi, 2005 #649	70	506	39	305	0.64 [0.54, 0.73]	0.38 [0.34, 0.41]
Li, 2021 #1850	8	7	20	46	0.29 [0.13, 0.49]	0.87 [0.75, 0.95]
Ma, 2017 #864	47	31	32	25	0.59 [0.48, 0.70]	0.45 [0.31, 0.59]
Maimaiti, 2016 #655	232	218	124	26	0.65 [0.60, 0.70]	0.11 [0.07, 0.15]
Shao, 2015 #923	174	111	29	198	0.86 [0.80, 0.90]	0.64 [0.58, 0.69]
Szczepanek-Parulska, 2013 #940	7	21	15	350	0.32 [0.14, 0.55]	0.94 [0.91, 0.96]
Xu, 2014 #975	18	57	88	212	0.17 [0.10, 0.26]	0.79 [0.73, 0.84]

Figure 5: Sensitivity and specificity of the presence of solidity for diagnosis of malignancy in thyroid cancer



Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Aggarwal, 1989 #693	7	9	9	11	0.44 [0.20, 0.70]	0.55 [0.32, 0.77]
Akhavan, 2016 #698	14	58	2	16	0.88 [0.62, 0.98]	0.22 [0.13, 0.33]
Chng, 2018 #727	40	70	11	42	0.78 [0.65, 0.89]	0.38 [0.29, 0.47]
Dobruch-Sobczak, 2019 #740	212	343	17	270	0.93 [0.88, 0.96]	0.44 [0.40, 0.48]
Hang, 2018 #788	177	110	0	11	1.00 [0.98, 1.00]	0.09 [0.05, 0.16]
Hekimsoy, 2021 #1826	57	150	5	39	0.92 [0.82, 0.97]	0.21 [0.15, 0.27]
Kalantari, 2018 #811	5	8	4	46	0.56 [0.21, 0.86]	0.85 [0.73, 0.93]
Kobayashi, 2005 #649	87	525	22	286	0.80 [0.71, 0.87]	0.35 [0.32, 0.39]
Kong, 2017 #827	66	17	13	17	0.84 [0.74, 0.91]	0.50 [0.32, 0.68]
Li, 2016 #838	378	178	46	160	0.89 [0.86, 0.92]	0.47 [0.42, 0.53]
Li, 2021 #1850	25	21	3	32	0.89 [0.72, 0.98]	0.60 [0.46, 0.74]
Liu, 2020 #851	25	47	3	15	0.89 [0.72, 0.98]	0.24 [0.14, 0.37]
Maimaiti, 2016 #655	301	121	55	123	0.85 [0.80, 0.88]	0.50 [0.44, 0.57]
Pagano, 2020 #888	68	74	0	4	1.00 [0.95, 1.00]	0.05 [0.01, 0.13]
Parikh, 2013 #891	14	24	15	31	0.48 [0.29, 0.67]	0.56 [0.42, 0.70]
Peccin, 2002 #897	11	30	5	34	0.69 [0.41, 0.89]	0.53 [0.40, 0.66]
Pei, 2019 #898	112	63	6	15	0.95 [0.89, 0.98]	0.19 [0.11, 0.30]
Phuttharak, 2009 #901	5	18	0	8	1.00 [0.48, 1.00]	0.31 [0.14, 0.52]
Schweel, 2013 #929	10	10	6	40	0.63 [0.35, 0.85]	0.80 [0.66, 0.90]
Sharma, 2019 #1821	8	30	0	10	1.00 [0.63, 1.00]	0.25 [0.13, 0.41]
Shen, 2019 #924	712	517	61	322	0.92 [0.90, 0.94]	0.38 [0.35, 0.42]
Skowronska, 2018 #930	7	34	1	98	0.88 [0.47, 1.00]	0.74 [0.66, 0.81]
Swan, 2019 #938	57	160	22	165	0.72 [0.61, 0.82]	0.51 [0.45, 0.56]
Szczepanek-Parulska, 2013 #940	19	220	3	151	0.86 [0.65, 0.97]	0.41 [0.36, 0.46]
Wang, 2017 #964	452	288	12	259	0.97 [0.96, 0.99]	0.47 [0.43, 0.52]
Wu, 2016 #687	489	314	18	149	0.96 [0.94, 0.98]	0.32 [0.28, 0.37]
Xu, 2014 #975	112	155	4	170	0.97 [0.91, 0.99]	0.52 [0.47, 0.58]

Figure 6: Sensitivity and specificity of the presence of microcalcifications for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abd_Alrahman, 2017 #633	4	0	3	23	0.57 [0.18, 0.90]	1.00 [0.85, 1.00]
Akhavan, 2016 #698	12	27	4	47	0.75 [0.48, 0.93]	0.64 [0.52, 0.74]
Appetecchia, 2006 #699	30	56	6	111	0.83 [0.67, 0.94]	0.66 [0.59, 0.74]
Bakari, 2018 #659	51	5	3	19	0.94 [0.85, 0.99]	0.79 [0.58, 0.93]
Borlea, 2020 #712	11	9	24	89	0.31 [0.17, 0.49]	0.91 [0.83, 0.96]
Cakir, 2011 #716	119	232	6	34	0.95 [0.90, 0.98]	0.13 [0.09, 0.17]
Chng, 2018 #727	17	9	34	103	0.33 [0.21, 0.48]	0.92 [0.85, 0.96]
Dobruch-Sobczak, 2019 #740	123	148	106	465	0.54 [0.47, 0.60]	0.76 [0.72, 0.79]
Gorgulu, 2019 #766	12	10	8	71	0.60 [0.36, 0.81]	0.88 [0.78, 0.94]
Gu, 2012 #774	10	2	12	74	0.45 [0.24, 0.68]	0.97 [0.91, 1.00]
Hekimsoy, 2021 #1826	11	3	14	48	0.44 [0.24, 0.65]	0.94 [0.84, 0.99]
Hong, 2009 #792	27	6	22	90	0.55 [0.40, 0.69]	0.94 [0.87, 0.98]
Kalantari, 2018 #811	7	10	2	44	0.78 [0.40, 0.97]	0.81 [0.69, 0.91]
Kong, 2017 #827	51	8	28	26	0.65 [0.53, 0.75]	0.76 [0.59, 0.89]
Li, 2016 #838	175	30	249	308	0.41 [0.37, 0.46]	0.91 [0.88, 0.94]
Li, 2021 #1833	120	10	152	74	0.44 [0.38, 0.50]	0.88 [0.79, 0.94]
Liu, 2014 #853	10	6	9	39	0.53 [0.29, 0.76]	0.87 [0.73, 0.95]
Liu, 2017 #854	32	14	23	56	0.58 [0.44, 0.71]	0.80 [0.69, 0.89]
Liu, 2019 #669	41	17	23	93	0.64 [0.51, 0.76]	0.85 [0.76, 0.91]
Liu, 2020 #851	1	0	27	62	0.04 [0.00, 0.18]	1.00 [0.94, 1.00]
Ma, 2014 #865	62	4	32	74	0.66 [0.55, 0.75]	0.95 [0.87, 0.99]
Ma, 2017 #864	47	15	32	41	0.59 [0.48, 0.70]	0.73 [0.60, 0.84]
Maia, 2011 #654	20	12	30	81	0.40 [0.26, 0.55]	0.87 [0.79, 0.93]
Maimaiti, 2016 #655	222	79	134	165	0.62 [0.57, 0.67]	0.68 [0.61, 0.73]
Mohamed, 2013 #672	19	3	9	14	0.68 [0.48, 0.84]	0.82 [0.57, 0.96]
Mohey, 2013 #674	10	9	5	22	0.67 [0.38, 0.88]	0.71 [0.52, 0.86]
Pagano, 2020 #888	21	2	47	76	0.31 [0.20, 0.43]	0.97 [0.91, 1.00]
Parikh, 2013 #891	8	14	21	41	0.28 [0.13, 0.47]	0.75 [0.61, 0.85]
Peccin, 2002 #897	9	4	7	60	0.56 [0.30, 0.80]	0.94 [0.85, 0.98]
Pei, 2019 #898	53	4	65	74	0.45 [0.36, 0.54]	0.95 [0.87, 0.99]
Phuttharak, 2009 #901	3	0	2	26	0.60 [0.15, 0.95]	1.00 [0.87, 1.00]
Ragazzoni, 2012 #903	22	4	18	88	0.55 [0.38, 0.71]	0.96 [0.89, 0.99]
Rago, 1998 #905	13	18	17	56	0.43 [0.25, 0.63]	0.76 [0.64, 0.85]
Rago, 2007 #904	20	17	11	44	0.65 [0.45, 0.81]	0.72 [0.59, 0.83]
Ren, 2015 #907	12	3	31	28	0.28 [0.15, 0.44]	0.90 [0.74, 0.98]
Ren, 2015 #907b	15	1	10	29	0.60 [0.39, 0.79]	0.97 [0.83, 1.00]
Ren, 2015 #907c	20	3	22	33	0.48 [0.32, 0.64]	0.92 [0.78, 0.98]
Schweel, 2013 #929	8	0	8	50	0.50 [0.25, 0.75]	1.00 [0.93, 1.00]
Shao, 2015 #923	140	17	63	292	0.69 [0.62, 0.75]	0.94 [0.91, 0.97]
Shen, 2019 #924	502	96	271	743	0.65 [0.61, 0.68]	0.89 [0.86, 0.91]
Shi, 2020 #681	33	3	17	285	0.66 [0.51, 0.79]	0.99 [0.97, 1.00]
Skowronska, 2018 #930	3	2	5	130	0.38 [0.09, 0.76]	0.98 [0.95, 1.00]
Sui, 2016 #935	49	13	17	30	0.74 [0.62, 0.84]	0.70 [0.54, 0.83]
Swan, 2019 #938	51	150	28	175	0.65 [0.53, 0.75]	0.54 [0.48, 0.59]
Szczepanek-Parulska, 2013 #940	9	67	13	304	0.41 [0.21, 0.64]	0.82 [0.78, 0.86]
Tuan, 2020 #948	49	2	25	18	0.66 [0.54, 0.77]	0.90 [0.68, 0.99]
Veyrieres, 2012 #950	15	43	20	219	0.43 [0.26, 0.61]	0.84 [0.79, 0.88]
Wang, 2018 #955	91	5	256	93	0.26 [0.22, 0.31]	0.95 [0.88, 0.98]
Wu, 2016 #687	225	40	282	423	0.44 [0.40, 0.49]	0.91 [0.88, 0.94]
Xing, 2011 #973	23	4	22	49	0.51 [0.36, 0.66]	0.92 [0.82, 0.98]
Xu, 2014 #975	61	58	55	267	0.53 [0.43, 0.62]	0.82 [0.78, 0.86]
Zhang, 2014 #984	11	7	21	32	0.34 [0.19, 0.53]	0.82 [0.66, 0.92]
Zhang,2016 #991	32	9	31	73	0.51 [0.38, 0.64]	0.89 [0.80, 0.95]

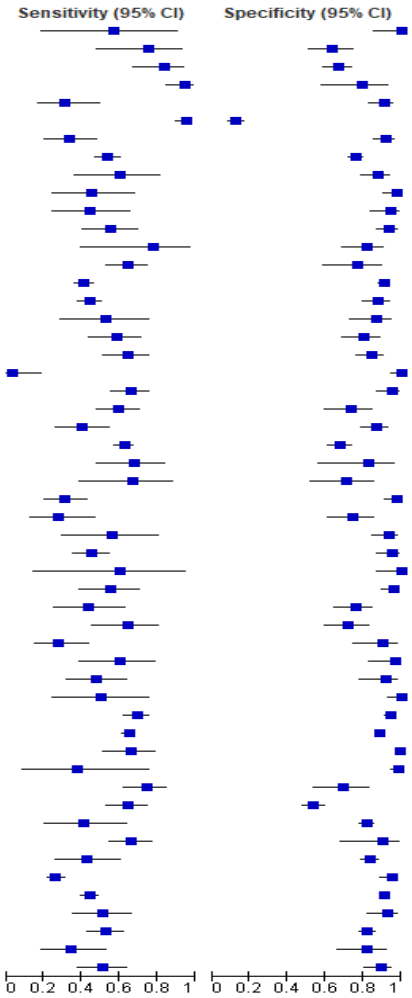


Figure 7: Sensitivity and specificity of the presence of hypoechoicity for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abd_Alrahman, 2017 #633	6	4	1	19	0.86 [0.42, 1.00]	0.83 [0.61, 0.95]
Appetecchia, 2006 #699	18	65	18	102	0.50 [0.33, 0.67]	0.61 [0.53, 0.69]
Bakari, 2018 #659	53	4	1	20	0.98 [0.90, 1.00]	0.83 [0.63, 0.95]
Cakir, 2011 #716	115	214	10	52	0.92 [0.86, 0.96]	0.20 [0.15, 0.25]
Chng, 2018 #727	36	32	15	80	0.71 [0.56, 0.83]	0.71 [0.62, 0.80]
Gorgulu, 2019 #766	17	25	3	56	0.85 [0.62, 0.97]	0.69 [0.58, 0.79]
Gu, 2012 #774	20	33	2	43	0.91 [0.71, 0.99]	0.57 [0.45, 0.68]
Hong, 2009 #792	42	28	7	68	0.86 [0.73, 0.94]	0.71 [0.61, 0.80]
Kalantari, 2018 #811	6	11	3	43	0.67 [0.30, 0.93]	0.80 [0.66, 0.89]
Kong, 2017 #827	69	19	10	15	0.87 [0.78, 0.94]	0.44 [0.27, 0.62]
Li, 2016 #838	380	147	44	191	0.90 [0.86, 0.92]	0.57 [0.51, 0.62]
Liu, 2014 #853	14	19	5	26	0.74 [0.49, 0.91]	0.58 [0.42, 0.72]
Liu, 2017 #854	51	56	4	14	0.93 [0.82, 0.98]	0.20 [0.11, 0.31]
Liu, 2020 #851	18	13	10	49	0.64 [0.44, 0.81]	0.79 [0.67, 0.88]
Ma, 2014 #865	73	16	21	62	0.78 [0.68, 0.86]	0.79 [0.69, 0.88]
Maia, 2011 #654	32	36	18	57	0.64 [0.49, 0.77]	0.61 [0.51, 0.71]
Maimaiti, 2016 #655	339	172	17	72	0.95 [0.92, 0.97]	0.30 [0.24, 0.36]
Mohamed, 2013 #672	15	6	13	11	0.54 [0.34, 0.72]	0.65 [0.38, 0.86]
Mohey, 2013 #674	12	12	3	19	0.80 [0.52, 0.96]	0.61 [0.42, 0.78]
Parikh, 2013 #891	10	18	19	37	0.34 [0.18, 0.54]	0.67 [0.53, 0.79]
Peccin, 2002 #897	7	11	9	53	0.44 [0.20, 0.70]	0.83 [0.71, 0.91]
Ragazzoni, 2012 #903	35	39	5	53	0.88 [0.73, 0.96]	0.58 [0.47, 0.68]
Rago, 1998 #905	20	38	10	36	0.67 [0.47, 0.83]	0.49 [0.37, 0.61]
Rago, 2007 #904	25	23	6	38	0.81 [0.63, 0.93]	0.62 [0.49, 0.74]
Ren, 2015 #907	40	7	3	24	0.93 [0.81, 0.99]	0.77 [0.59, 0.90]
Ren, 2015 #907b	23	10	2	20	0.92 [0.74, 0.99]	0.67 [0.47, 0.83]
Ren, 2015 #907c	40	15	2	21	0.95 [0.84, 0.99]	0.58 [0.41, 0.74]
Schleder, 2015 #920	10	17	16	58	0.38 [0.20, 0.59]	0.77 [0.66, 0.86]
Shao, 2015 #923	135	78	68	231	0.67 [0.60, 0.73]	0.75 [0.70, 0.80]
Shi, 2020 #681	31	79	19	209	0.62 [0.47, 0.75]	0.73 [0.67, 0.78]
Sodagari, 2018 #682	39	26	6	43	0.87 [0.73, 0.95]	0.62 [0.50, 0.74]
Sui, 2016 #935	25	5	41	38	0.38 [0.26, 0.51]	0.88 [0.75, 0.96]
Swan, 2019 #938	68	185	11	140	0.86 [0.76, 0.93]	0.43 [0.38, 0.49]
Szczepanek-Parulska, 2013 #940	21	244	1	127	0.95 [0.77, 1.00]	0.34 [0.29, 0.39]
Veyrieres, 2012 #950	25	120	10	142	0.71 [0.54, 0.85]	0.54 [0.48, 0.60]
Vorlander, 2010 #953	43	91	7	168	0.86 [0.73, 0.94]	0.65 [0.59, 0.71]
Wang, 2012 #959	298	48	49	50	0.86 [0.82, 0.89]	0.51 [0.41, 0.61]
Xing, 2011 #973	32	18	13	35	0.71 [0.56, 0.84]	0.66 [0.52, 0.78]
Zhang,2016 #991	61	55	2	27	0.97 [0.89, 1.00]	0.33 [0.23, 0.44]

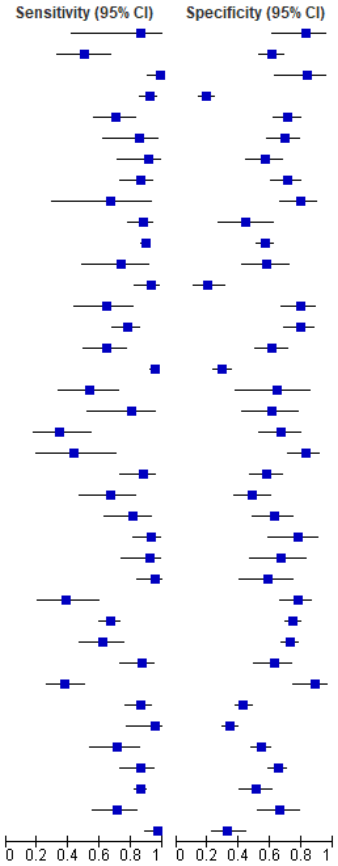


Figure 8: Sensitivity and specificity of the presence of marked hypoechoicity for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Borlea, 2020 #712	9	3	26	95	0.26 [0.12, 0.43]	0.97 [0.91, 0.99]
Dobruch-Sobczak, 2019 #740	135	173	94	440	0.59 [0.52, 0.65]	0.72 [0.68, 0.75]
Hang, 2018 #788	22	2	155	119	0.12 [0.08, 0.18]	0.98 [0.94, 1.00]
Kim, 2008 #812	24	7	35	27	0.41 [0.28, 0.54]	0.79 [0.62, 0.91]
Liu, 2019 #669	31	8	33	102	0.48 [0.36, 0.61]	0.93 [0.86, 0.97]
Ma, 2017 #864	23	6	56	50	0.29 [0.19, 0.40]	0.89 [0.78, 0.96]
Pagano, 2020 #888	26	5	42	73	0.38 [0.27, 0.51]	0.94 [0.86, 0.98]
Pei, 2019 #898	85	31	33	47	0.72 [0.63, 0.80]	0.60 [0.49, 0.71]
Phuttharak, 2009 #901	2	11	3	15	0.40 [0.05, 0.85]	0.58 [0.37, 0.77]
Schweel, 2013 #929	10	0	6	50	0.63 [0.35, 0.85]	1.00 [0.93, 1.00]
Shen, 2019 #924	45	18	728	821	0.06 [0.04, 0.08]	0.98 [0.97, 0.99]
Tuan, 2020 #948	23	1	51	19	0.31 [0.21, 0.43]	0.95 [0.75, 1.00]
Wang, 2017 #964	57	8	407	539	0.12 [0.09, 0.16]	0.99 [0.97, 0.99]
Xu, 2014 #975	63	15	53	310	0.54 [0.45, 0.64]	0.95 [0.93, 0.97]
Zhang, 2014 #984	19	10	13	29	0.59 [0.41, 0.76]	0.74 [0.58, 0.87]

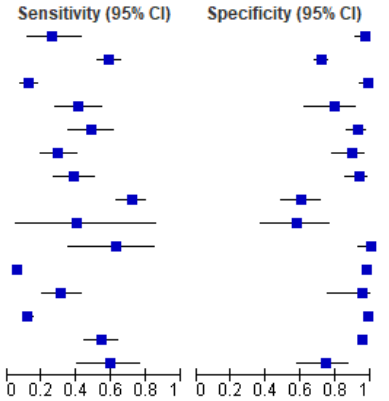


Figure 9: Sensitivity and specificity of the presence of hypoechoicity OR marked hypoechoicity for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Hekimsoy, 2021 #1826	40	31	22	158	0.65 [0.51, 0.76]	0.84 [0.78, 0.89]
Li, 2021 #1833	258	29	14	55	0.95 [0.92, 0.97]	0.65 [0.54, 0.76]
Li, 2021 #1850	19	15	9	38	0.68 [0.48, 0.84]	0.72 [0.58, 0.83]
Pagano, 2021 #1838	56	37	12	41	0.82 [0.71, 0.91]	0.53 [0.41, 0.64]
Skowronska, 2018 #930	8	65	0	67	1.00 [0.63, 1.00]	0.51 [0.42, 0.60]
Wang, 2017 #964	447	287	17	260	0.96 [0.94, 0.98]	0.48 [0.43, 0.52]

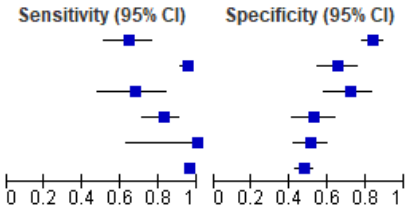


Figure 10: Sensitivity and specificity of the presence of poorly/ill defined margins or borders for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Borlea, 2020 #712	13	28	22	70	0.37 [0.21, 0.55]	0.71 [0.61, 0.80]
Deng, 2018 #663	39	16	6	46	0.87 [0.73, 0.95]	0.74 [0.62, 0.84]
Jeong, 2016 #805	4	19	18	137	0.18 [0.05, 0.40]	0.88 [0.82, 0.93]
Li, 2021 #1850	3	0	25	53	0.11 [0.02, 0.28]	1.00 [0.93, 1.00]
Liu, 2014 #853	11	9	8	36	0.58 [0.33, 0.80]	0.80 [0.65, 0.90]
Liu, 2017 #854	32	8	23	62	0.58 [0.44, 0.71]	0.89 [0.79, 0.95]
Liu, 2019 #669	36	23	28	87	0.56 [0.43, 0.69]	0.79 [0.70, 0.86]
Ma, 2017 #864	60	22	19	34	0.76 [0.65, 0.85]	0.61 [0.47, 0.74]
Maimaiti, 2016 #655	213	51	143	193	0.60 [0.55, 0.65]	0.79 [0.73, 0.84]
Mohamed, 2013 #672	24	3	4	14	0.86 [0.67, 0.96]	0.82 [0.57, 0.96]
Pagano, 2020 #888	24	0	44	78	0.35 [0.24, 0.48]	1.00 [0.95, 1.00]
Pei, 2019 #898	91	57	27	21	0.77 [0.68, 0.84]	0.27 [0.18, 0.38]
Phuttharak, 2009 #901	3	3	2	23	0.60 [0.15, 0.95]	0.88 [0.70, 0.98]
Ren, 2015 #907	42	22	1	9	0.98 [0.88, 1.00]	0.29 [0.14, 0.48]
Ren, 2015 #907b	25	7	0	23	1.00 [0.86, 1.00]	0.77 [0.58, 0.90]
Ren, 2015 #907c	41	14	1	22	0.98 [0.87, 1.00]	0.61 [0.43, 0.77]
Sodagari, 2018 #682	23	14	22	55	0.51 [0.36, 0.66]	0.80 [0.68, 0.88]
Sui, 2016 #935	57	0	9	43	0.86 [0.76, 0.94]	1.00 [0.92, 1.00]
Szczepanek-Parulska, 2013 #940	16	98	6	273	0.73 [0.50, 0.89]	0.74 [0.69, 0.78]
Wu, 2016 #687	465	265	42	198	0.92 [0.89, 0.94]	0.43 [0.38, 0.47]
Xu, 2014 #975	61	53	55	272	0.53 [0.43, 0.62]	0.84 [0.79, 0.88]
Zhang, 2014 #984	18	9	14	30	0.56 [0.38, 0.74]	0.77 [0.61, 0.89]
Zhang,2016 #991	36	7	27	75	0.57 [0.44, 0.70]	0.91 [0.83, 0.96]

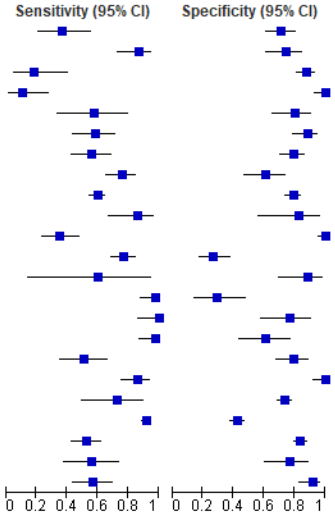


Figure 11: Sensitivity and specificity of the absence of the halo sign for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Appetecchia, 2006 #699	32	130	4	37	0.89 [0.74, 0.97]	0.22 [0.16, 0.29]
Cakir, 2011 #716	57	70	68	195	0.46 [0.37, 0.55]	0.74 [0.68, 0.79]
Kalantari, 2018 #811	7	35	2	19	0.78 [0.40, 0.97]	0.35 [0.23, 0.49]
Li, 2021 #1850	7	14	21	39	0.25 [0.11, 0.45]	0.74 [0.60, 0.85]
Ma, 2017 #864	78	43	1	13	0.99 [0.93, 1.00]	0.23 [0.13, 0.36]
Mohamed, 2013 #672	24	6	4	11	0.86 [0.67, 0.96]	0.65 [0.38, 0.86]
Mohey, 2013 #674	9	6	6	25	0.60 [0.32, 0.84]	0.81 [0.63, 0.93]
Peccin, 2002 #897	9	13	7	51	0.56 [0.30, 0.80]	0.80 [0.68, 0.89]
Phuttharak, 2009 #901	3	0	2	26	0.60 [0.15, 0.95]	1.00 [0.87, 1.00]
Rago, 1998 #905	20	17	10	57	0.67 [0.47, 0.83]	0.77 [0.66, 0.86]
Rago, 2007 #904	19	11	12	50	0.61 [0.42, 0.78]	0.82 [0.70, 0.91]
Schweel, 2013 #929	14	6	2	44	0.88 [0.62, 0.98]	0.88 [0.76, 0.95]
Veyrieres, 2012 #950	32	221	3	41	0.91 [0.77, 0.98]	0.16 [0.11, 0.21]
Wang, 2017 #964	420	414	44	133	0.91 [0.87, 0.93]	0.24 [0.21, 0.28]
Xu, 2014 #975	89	133	27	192	0.77 [0.68, 0.84]	0.59 [0.54, 0.64]

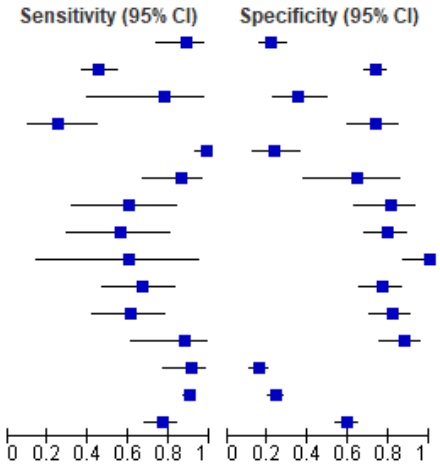


Figure 12: Sensitivity and specificity of the presence of irregular border for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abd_Alrahman, 2017 #633	6	2	1	21	0.86 [0.42, 1.00]	0.91 [0.72, 0.99]
Akhavan, 2016 #698	4	5	12	69	0.25 [0.07, 0.52]	0.93 [0.85, 0.98]
Bakari, 2018 #659	52	7	1	17	0.98 [0.90, 1.00]	0.71 [0.49, 0.87]
Chng, 2018 #727	18	9	33	103	0.35 [0.22, 0.50]	0.92 [0.85, 0.96]
Deng, 2018 #663	38	29	7	33	0.84 [0.71, 0.94]	0.53 [0.40, 0.66]
Dobruch-Sobczak, 2019 #740	173	112	56	501	0.76 [0.69, 0.81]	0.82 [0.78, 0.85]
Hekimsoy, 2021 #1826	29	2	33	187	0.47 [0.34, 0.60]	0.99 [0.96, 1.00]
Jeong, 2016 #805	1	2	21	154	0.05 [0.00, 0.23]	0.99 [0.95, 1.00]
Kalantari, 2018 #811	3	4	6	50	0.33 [0.07, 0.70]	0.93 [0.82, 0.98]
Kobayashi, 2005 #649	50	112	59	699	0.46 [0.36, 0.56]	0.86 [0.84, 0.88]
Kong, 2017 #827	42	7	37	27	0.53 [0.42, 0.64]	0.79 [0.62, 0.91]
Li, 2016 #838	306	67	118	271	0.72 [0.68, 0.76]	0.80 [0.76, 0.84]
Li, 2021 #1850	16	12	12	41	0.57 [0.37, 0.76]	0.77 [0.64, 0.88]
Liu, 2014 #853	10	8	9	37	0.53 [0.29, 0.76]	0.82 [0.68, 0.92]
Maia, 2011 #654	34	15	16	78	0.68 [0.53, 0.80]	0.84 [0.75, 0.91]
Maimaiti, 2016 #655	187	26	169	218	0.53 [0.47, 0.58]	0.89 [0.85, 0.93]
Parikh, 2013 #891	7	11	22	44	0.24 [0.10, 0.44]	0.80 [0.67, 0.90]
Ragazzoni, 2012 #903	20	9	20	83	0.50 [0.34, 0.66]	0.90 [0.82, 0.95]
Shao, 2015 #923	90	20	113	289	0.44 [0.37, 0.51]	0.94 [0.90, 0.96]
Skowronska, 2018 #930	8	0	0	132	1.00 [0.63, 1.00]	1.00 [0.97, 1.00]
Sui, 2016 #935	37	20	29	23	0.56 [0.43, 0.68]	0.53 [0.38, 0.69]
Swan, 2019 #938	37	103	42	222	0.47 [0.36, 0.58]	0.68 [0.63, 0.73]
Tuan, 2020 #948	51	6	23	14	0.69 [0.57, 0.79]	0.70 [0.46, 0.88]
Veyrieres, 2012 #950	8	32	27	230	0.23 [0.10, 0.40]	0.88 [0.83, 0.91]
Xu, 2014 #975	58	55	58	270	0.50 [0.41, 0.59]	0.83 [0.79, 0.87]
Zhang,2016 #991	34	15	29	67	0.54 [0.41, 0.67]	0.82 [0.72, 0.89]

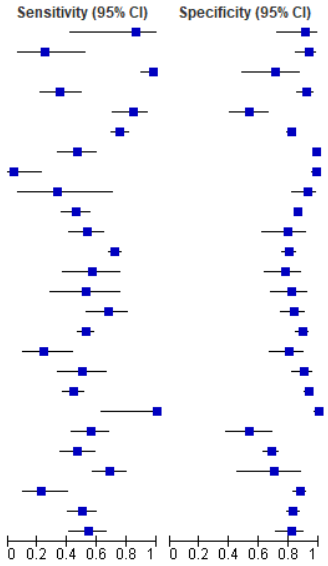


Figure 13: Sensitivity and specificity of the presence of heterogeneous texture for diagnosis of malignancy in thyroid cancer

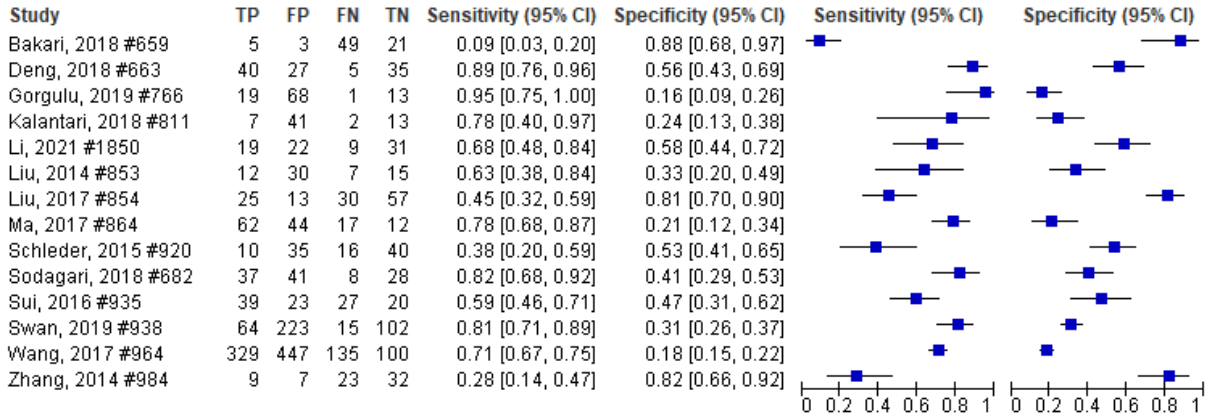


Figure 14: Sensitivity and specificity of the presence of macrocalcifications for diagnosis of malignancy in thyroid cancer

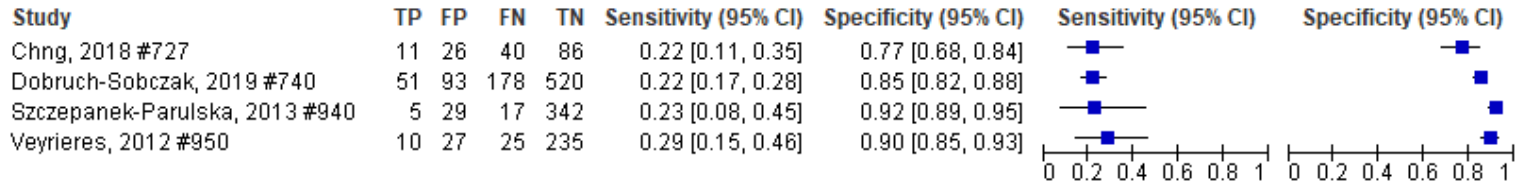


Figure 15: Sensitivity and specificity of the presence of nodules of 10mm diameter or less for diagnosis of malignancy in thyroid cancer

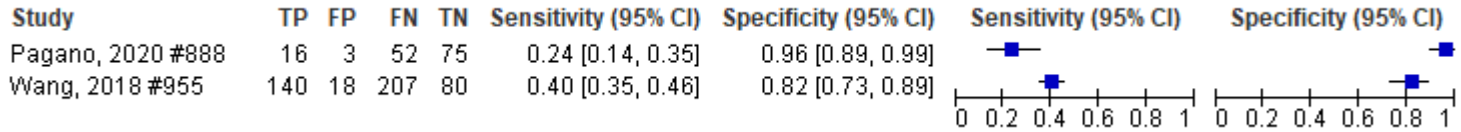


Figure 16: Sensitivity and specificity of the presence of nodules of 20mm diameter or less for diagnosis of malignancy in thyroid cancer

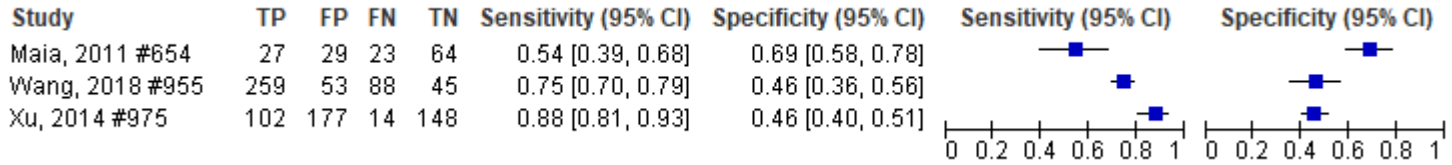


Figure 17: Sensitivity and specificity of the presence of nodules of 36mm diameter or less for diagnosis of malignancy in thyroid cancer

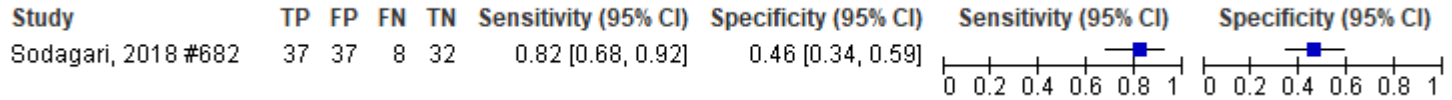
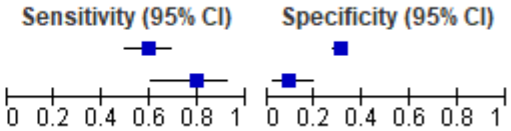


Figure 18: Sensitivity and specificity of the presence of nodules of 40mm diameter or less for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Kobayashi, 2005 #649	65	562	44	249	0.60 [0.50, 0.69]	0.31 [0.28, 0.34]
Parikh, 2013 #891	23	50	6	5	0.79 [0.60, 0.92]	0.09 [0.03, 0.20]



INFORMALLY COMBINED GREY SCALE CHARACTERISTICS

Figure 19: Sensitivity and specificity of microcalcifications AND absent halo for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Appetecchia, 2006 #699	27	47	9	120	0.75 [0.58, 0.88]	0.72 [0.64, 0.79]
Mohey, 2013 #674	9	2	6	29	0.60 [0.32, 0.84]	0.94 [0.79, 0.99]
Peccin, 2002 #897	6	2	10	62	0.38 [0.15, 0.65]	0.97 [0.89, 1.00]
Rago, 1998 #905	8	5	22	69	0.27 [0.12, 0.46]	0.93 [0.85, 0.98]
Rago, 2007 #904	19	3	12	58	0.61 [0.42, 0.78]	0.95 [0.86, 0.99]

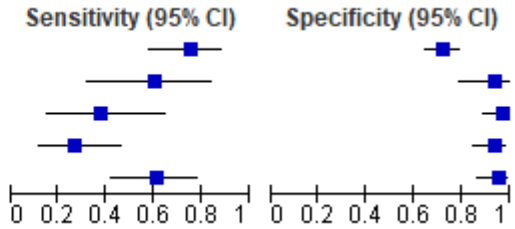


Figure 20: Sensitivity and specificity of hypoechoicity AND absent halo for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Appetecchia, 2006 #699	15	53	21	114	0.42 [0.26, 0.59]	0.68 [0.61, 0.75]
Mohey, 2013 #674	9	2	6	29	0.60 [0.32, 0.84]	0.94 [0.79, 0.99]
Peccin, 2002 #897	5	7	11	57	0.31 [0.11, 0.59]	0.89 [0.79, 0.95]
Rago, 1998 #905	18	13	12	61	0.60 [0.41, 0.77]	0.82 [0.72, 0.90]
Rago, 2007 #904	18	4	13	57	0.58 [0.39, 0.75]	0.93 [0.84, 0.98]

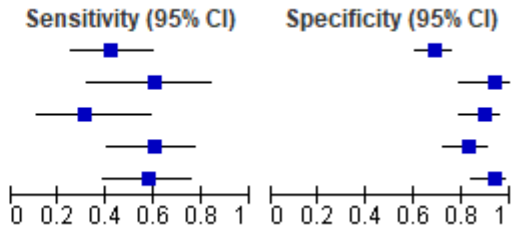


Figure 21: Sensitivity and specificity of hypoechoicity AND microcalcifications for diagnosis of malignancy in thyroid cancer

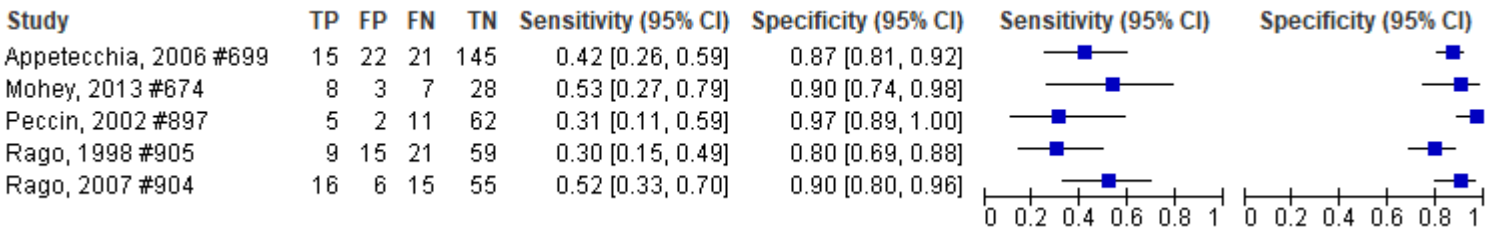


Figure 22: Sensitivity and specificity of hypoechoicity AND microcalcifications AND absent halo for diagnosis of malignancy in thyroid cancer

Figure 23: Sensitivity and specificity of hypoechoicity OR microcalcifications for diagnosis of malignancy in thyroid cancer

Figure 24: Sensitivity and specificity of hypoechoicity OR absent halo OR microcalcifications for diagnosis of malignancy in thyroid cancer

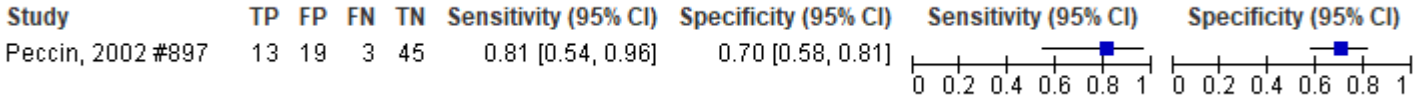


Figure 25: Sensitivity and specificity of microcalcifications OR absent halo for diagnosis of malignancy in thyroid cancer

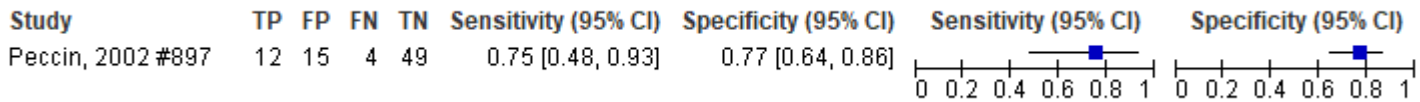


Figure 26: Sensitivity and specificity of hypoechoicity OR absent halo for diagnosis of malignancy in thyroid cancer

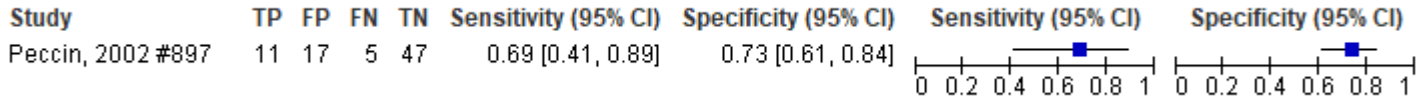


Figure 27: Sensitivity and specificity of at least one US sign detected (any allowed from a variety of selections) for diagnosis of malignancy in thyroid cancer

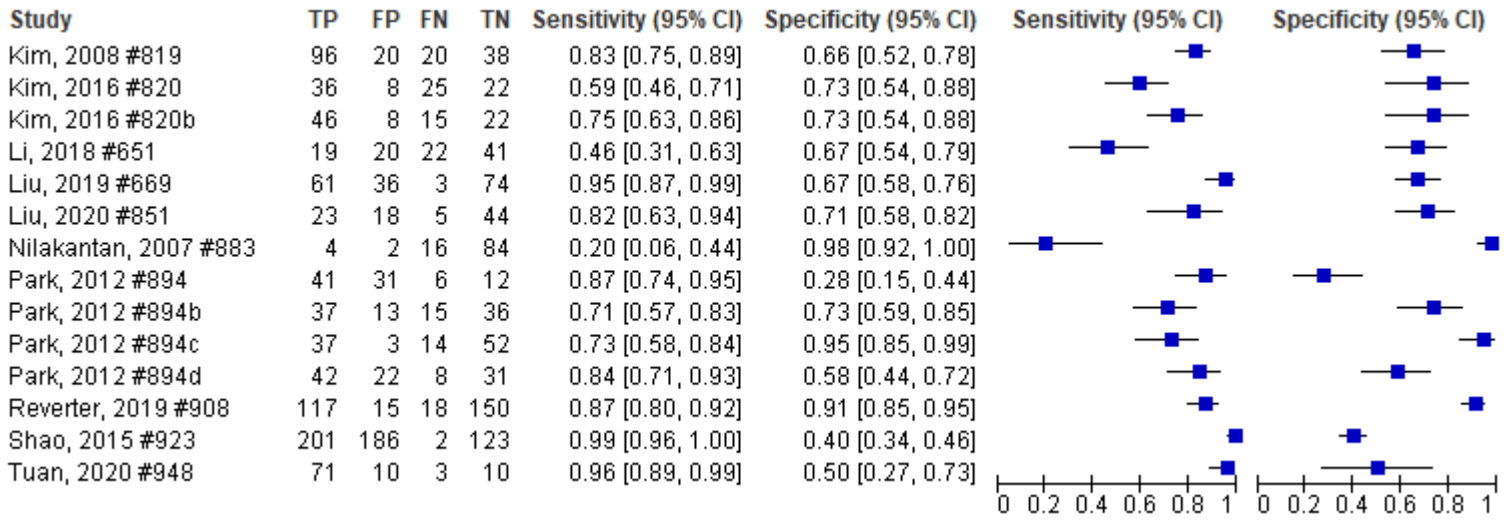


Figure 28: Sensitivity and specificity of at least 2 US signs detected for diagnosis of malignancy in thyroid cancer

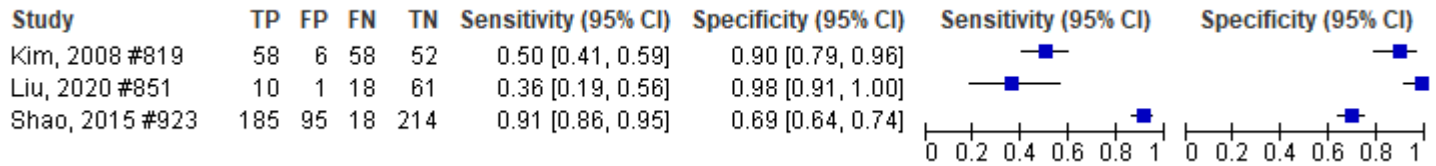


Figure 29: Sensitivity and specificity of at least 3 US signs detected for diagnosis of malignancy in thyroid cancer

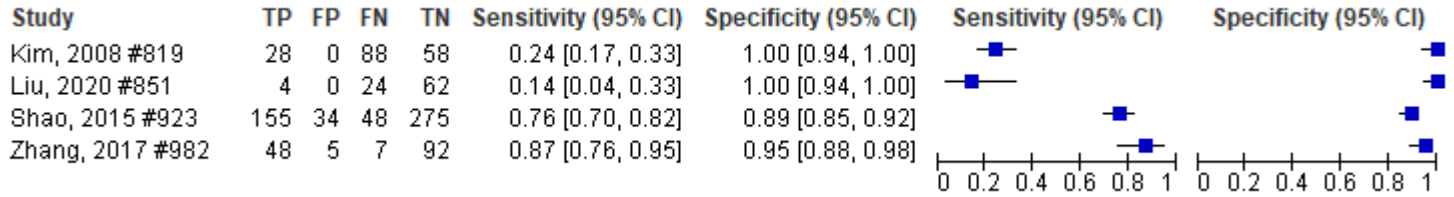


Figure 30: Sensitivity and specificity of at least 4 US signs detected for diagnosis of malignancy in thyroid cancer

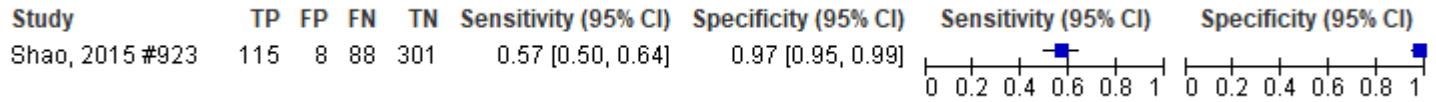


Figure 31: Sensitivity and specificity of at least 4 US signs detected for diagnosis of malignancy in thyroid cancer

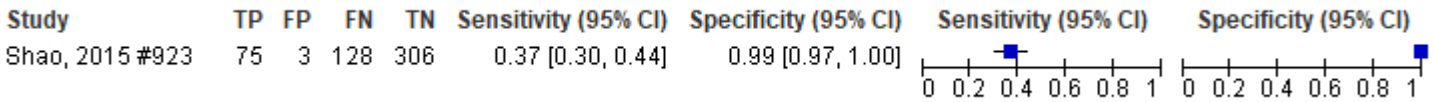


Figure 32: Sensitivity and specificity of ALL of: Taller than wide, Sub capsular, Intense hypo echoic, Calcification, Suspect lymph nodes for diagnosis of malignancy in thyroid cancer

Figure 33: Sensitivity and specificity of ALL of: Taller than wide, Sub capsular, Intense hypo echoic, Calcification, Suspect lymph nodes OR ALL of hypoechoic, sub-capsular position, inhomogeneity for diagnosis of malignancy in thyroid cancer

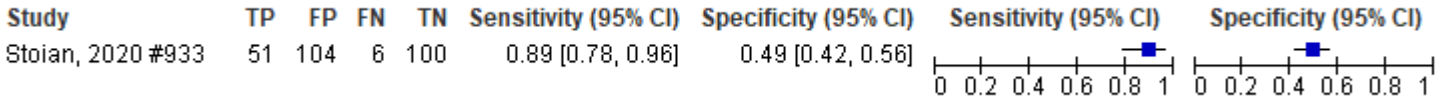


Figure 34: Sensitivity and specificity of blurred margins alongside any one of the following: hypoechoicity, microcalcification or taller than wide for diagnosis of malignancy in thyroid cancer

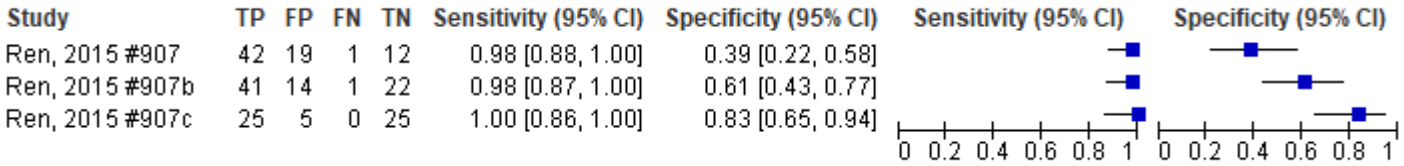


Figure 35: Sensitivity and specificity of hypoechoicity alongside any one of the following: blurred margins, microcalcification or taller than wide for diagnosis of malignancy in thyroid cancer

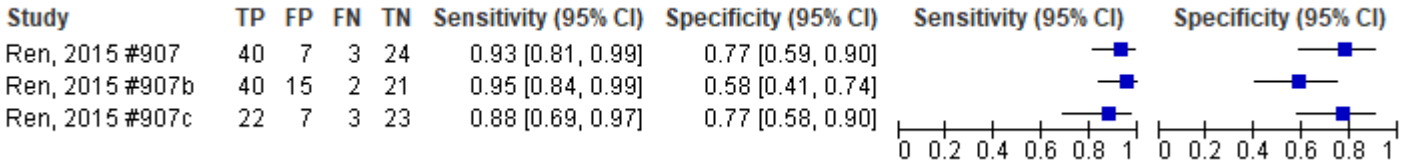


Figure 36: Sensitivity and specificity of microcalcifications alongside any one of the following: hypoechoicity, blurred margins or taller than wide for diagnosis of malignancy in thyroid cancer

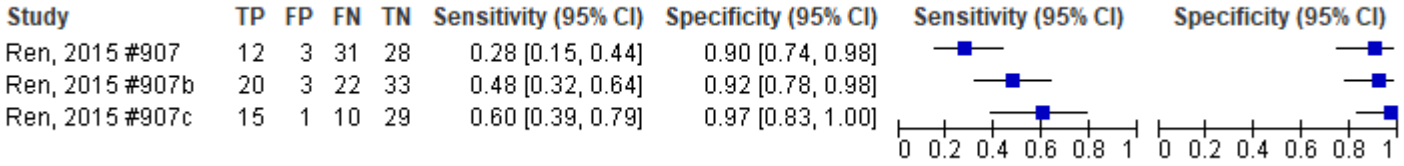


Figure 37: Sensitivity and specificity of taller than wide alongside any one of the following: hypoechoicity, microcalcification or blurred margins for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ren, 2015 #907	35	1	8	30	0.81 [0.67, 0.92]	0.97 [0.83, 1.00]
Ren, 2015 #907b	33	2	9	34	0.79 [0.63, 0.90]	0.94 [0.81, 0.99]
Ren, 2015 #907c	7	1	18	29	0.28 [0.12, 0.49]	0.97 [0.83, 1.00]

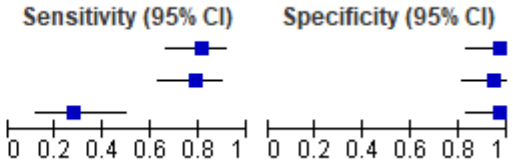


Figure 38: Sensitivity and specificity of microlobulated or irregular margins for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Kim, 2008 #812	21	1	38	33	0.36 [0.24, 0.49]	0.97 [0.85, 1.00]
Liu, 2020 #851	9	1	19	61	0.32 [0.16, 0.52]	0.98 [0.91, 1.00]
Shen, 2019 #924	201	39	572	800	0.26 [0.23, 0.29]	0.95 [0.94, 0.97]

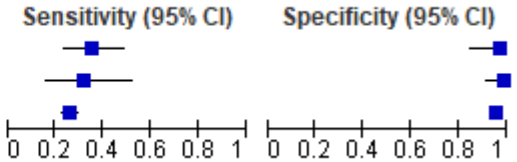


Figure 39: Sensitivity and specificity of infiltrative/ETE or lobulated or irregular for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Hang, 2018 #788	142	26	35	95	0.80 [0.74, 0.86]	0.79 [0.70, 0.85]
Wang, 2017 #964	330	75	134	472	0.71 [0.67, 0.75]	0.86 [0.83, 0.89]

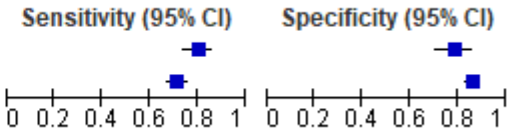


Figure 40: Sensitivity and specificity of spiculated or blurred/ ill-defined margins for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Hong, 2009 #792	45	15	4	81	0.92 [0.80, 0.98]	0.84 [0.76, 0.91]
Schweel, 2013 #929	8	2	8	48	0.50 [0.25, 0.75]	0.96 [0.86, 1.00]

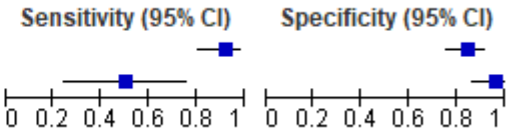


Figure 41: Sensitivity and specificity of spiculated or microlobulated nodules for diagnosis of malignancy in thyroid cancer

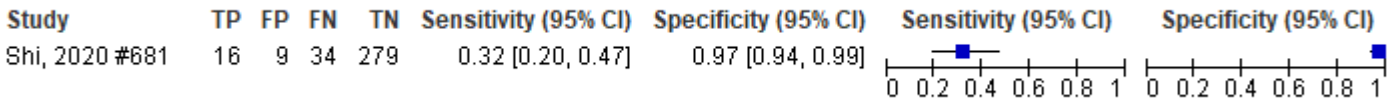
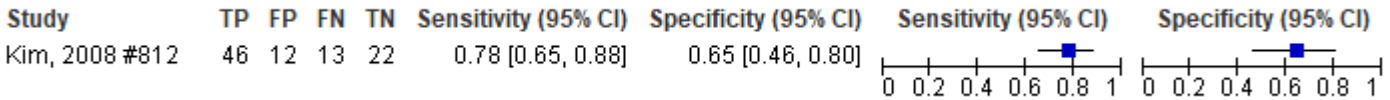


Figure 42: Sensitivity and specificity of hypoechoic or microlobulated margins for diagnosis of malignancy in thyroid cancer



FORMAL COMBINATIONS OF PREDOMINANTLY GREY SCALE CHARACTERISTICS (i.e. TIRADS, BTA U SCALE)

Figure 43: Sensitivity and specificity of ACR TIRADS score of 2 or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ahmadi, 2019 #696	88	229	0	6	1.00 [0.96, 1.00]	0.03 [0.01, 0.05]
Borlea, 2020 #712	35	93	0	5	1.00 [0.90, 1.00]	0.05 [0.02, 0.12]
Gao, 2019 #759	1681	855	0	8	1.00 [1.00, 1.00]	0.01 [0.00, 0.02]
Hekimsoy, 2021 #1826	62	186	0	3	1.00 [0.94, 1.00]	0.02 [0.00, 0.05]
Kuru, 2021 #1809	516	604	0	23	1.00 [0.99, 1.00]	0.04 [0.02, 0.05]
Magri, 2020 #867	95	208	0	1	1.00 [0.96, 1.00]	0.00 [0.00, 0.03]
Paker, 2021 #1814	115	111	0	12	1.00 [0.97, 1.00]	0.10 [0.05, 0.16]
Qi, 2021 #1831	414	642	0	40	1.00 [0.99, 1.00]	0.06 [0.04, 0.08]
Schenke, 2020 #917	76	65	0	4	1.00 [0.95, 1.00]	0.06 [0.02, 0.14]
Shah, 2020 #1823	9	32	0	9	1.00 [0.66, 1.00]	0.22 [0.11, 0.38]
Watkins, 2021 #1810	77	127	0	14	1.00 [0.95, 1.00]	0.10 [0.06, 0.16]

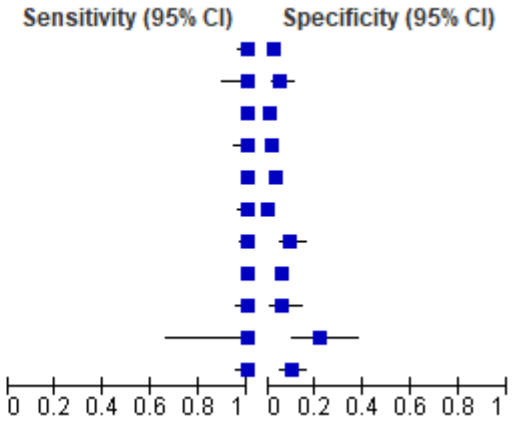


Figure 44: Sensitivity and specificity of ACR TIRADS score of 3 or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ahmadi, 2019 #696	85	206	3	29	0.97 [0.90, 0.99]	0.12 [0.08, 0.17]
Bora Makal, 2021 #711	54	75	0	12	1.00 [0.93, 1.00]	0.14 [0.07, 0.23]
Gao, 2019 #759	1679	711	2	152	1.00 [1.00, 1.00]	0.18 [0.15, 0.20]
Hekimsoy, 2021 #1826	58	144	4	45	0.94 [0.84, 0.98]	0.24 [0.18, 0.31]
Kuru, 2021 #1809	508	520	8	107	0.98 [0.97, 0.99]	0.17 [0.14, 0.20]
Liu, 2020 #851	24	50	4	12	0.86 [0.67, 0.96]	0.19 [0.10, 0.31]
Magri, 2020 #867	92	133	3	76	0.97 [0.91, 0.99]	0.36 [0.30, 0.43]
Paker, 2021 #1814	111	62	4	61	0.97 [0.91, 0.99]	0.50 [0.40, 0.59]
Qi, 2021 #1831	413	497	1	185	1.00 [0.99, 1.00]	0.27 [0.24, 0.31]
Schenke, 2020 #917	76	57	0	12	1.00 [0.95, 1.00]	0.17 [0.09, 0.28]
Shah, 2020 #1823	8	15	1	26	0.89 [0.52, 1.00]	0.63 [0.47, 0.78]
Shen, 2019 #924	773	833	0	6	1.00 [1.00, 1.00]	0.01 [0.00, 0.02]
Watkins, 2021 #1810	74	98	3	43	0.96 [0.89, 0.99]	0.30 [0.23, 0.39]

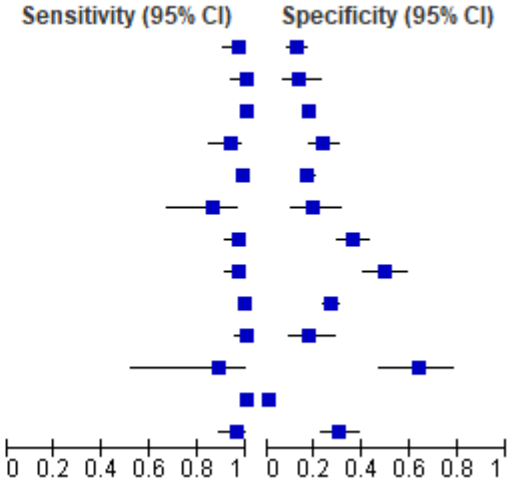


Figure 45: Sensitivity and specificity of ACR TIRADS score of 4 or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ahmadi, 2019 #696	69	63	19	172	0.78 [0.68, 0.86]	0.73 [0.67, 0.79]
Bora Makal, 2021 #711	45	54	9	33	0.83 [0.71, 0.92]	0.38 [0.28, 0.49]
Borlea, 2020 #712	33	70	2	28	0.94 [0.81, 0.99]	0.29 [0.20, 0.39]
Gao, 2019 #759	1651	431	30	432	0.98 [0.97, 0.99]	0.50 [0.47, 0.53]
Haskjold, 2021 #1834	26	47	1	17	0.96 [0.81, 1.00]	0.27 [0.16, 0.39]
Hekimsoy, 2021 #1826	44	47	18	142	0.71 [0.58, 0.82]	0.75 [0.68, 0.81]
Kuru, 2021 #1809	381	207	135	420	0.74 [0.70, 0.78]	0.67 [0.63, 0.71]
Liu, 2020 #851	20	14	8	48	0.71 [0.51, 0.87]	0.77 [0.65, 0.87]
Magri, 2020 #867	74	46	21	163	0.78 [0.68, 0.86]	0.78 [0.72, 0.83]
Paker, 2021 #1814	107	31	8	92	0.93 [0.87, 0.97]	0.75 [0.66, 0.82]
Qi, 2021 #1831	406	443	8	239	0.98 [0.96, 0.99]	0.35 [0.31, 0.39]
Schenke, 2020 #917	76	41	0	28	1.00 [0.95, 1.00]	0.41 [0.29, 0.53]
Shah, 2020 #1823	7	6	2	35	0.78 [0.40, 0.97]	0.85 [0.71, 0.94]
Shen, 2019 #924	760	564	13	275	0.98 [0.97, 0.99]	0.33 [0.30, 0.36]
Shreyamsa, 2020 #927	26	15	9	118	0.74 [0.57, 0.88]	0.89 [0.82, 0.94]
Watkins, 2021 #1810	60	63	17	78	0.78 [0.67, 0.87]	0.55 [0.47, 0.64]

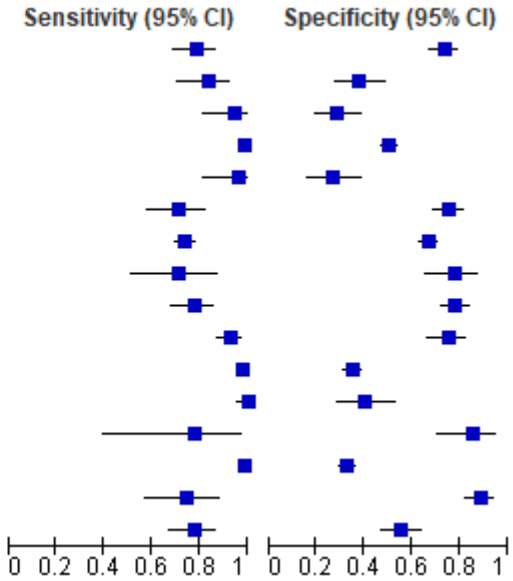


Figure 46: Sensitivity and specificity of ACR TIRADS score of 5 for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ahmadi, 2019 #696	29	3	59	232	0.33 [0.23, 0.44]	0.99 [0.96, 1.00]
Bora Makal, 2021 #711	24	15	30	72	0.44 [0.31, 0.59]	0.83 [0.73, 0.90]
Borlea, 2020 #712	17	22	18	76	0.49 [0.31, 0.66]	0.78 [0.68, 0.85]
Gao, 2019 #759	1372	179	309	684	0.82 [0.80, 0.83]	0.79 [0.76, 0.82]
Hekimsoy, 2021 #1826	25	4	37	185	0.40 [0.28, 0.54]	0.98 [0.95, 0.99]
Kuru, 2021 #1809	148	18	368	609	0.29 [0.25, 0.33]	0.97 [0.96, 0.98]
Liu, 2020 #851	7	2	21	60	0.25 [0.11, 0.45]	0.97 [0.89, 1.00]
Magri, 2020 #867	29	3	66	206	0.31 [0.21, 0.41]	0.99 [0.96, 1.00]
Paker, 2021 #1814	73	8	42	115	0.63 [0.54, 0.72]	0.93 [0.88, 0.97]
Qi, 2021 #1831	306	201	108	481	0.74 [0.69, 0.78]	0.71 [0.67, 0.74]
Schenke, 2020 #917	53	23	23	46	0.70 [0.58, 0.80]	0.67 [0.54, 0.78]
Shah, 2020 #1823	2	0	7	41	0.22 [0.03, 0.60]	1.00 [0.91, 1.00]
Shen, 2019 #924	682	105	91	734	0.88 [0.86, 0.90]	0.87 [0.85, 0.90]
Shreyamsa, 2020 #927	13	1	22	132	0.37 [0.21, 0.55]	0.99 [0.96, 1.00]
Watkins, 2021 #1810	35	20	42	121	0.45 [0.34, 0.57]	0.86 [0.79, 0.91]

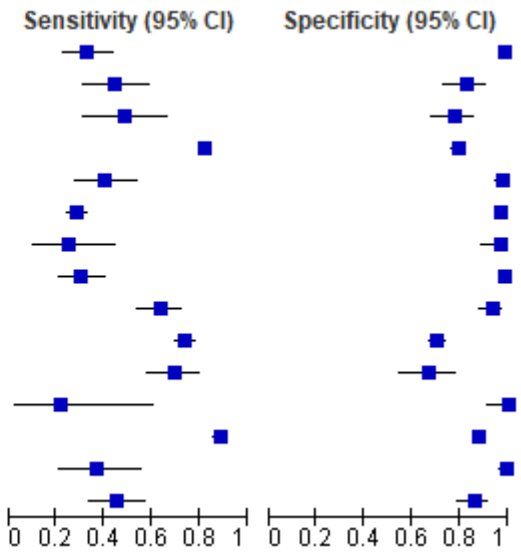


Figure 47: Sensitivity and specificity of EU TIRADS score of 2 or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Magri, 2020 #867	95	209	0	1	1.00 [0.96, 1.00]	0.00 [0.00, 0.03]

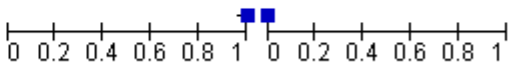


Figure 48: Sensitivity and specificity of EU TIRADS score of 3 or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Borlea, 2020 #712	35	92	0	6	1.00 [0.90, 1.00]	0.06 [0.02, 0.13]
Dobruch-Sobczak, 2019 #740	229	459	0	154	1.00 [0.98, 1.00]	0.25 [0.22, 0.29]
Hekimsoy, 2021 #1826	62	186	0	3	1.00 [0.94, 1.00]	0.02 [0.00, 0.05]
Kuru, 2021 #1809	516	608	0	19	1.00 [0.99, 1.00]	0.03 [0.02, 0.05]
Magri, 2020 #867	95	207	0	2	1.00 [0.96, 1.00]	0.01 [0.00, 0.03]
Pagano, 2020 #888	68	74	0	4	1.00 [0.95, 1.00]	0.05 [0.01, 0.13]
Qi, 2021 #1831	414	642	0	40	1.00 [0.99, 1.00]	0.06 [0.04, 0.08]
Schenke, 2020 #917	76	69	0	0	1.00 [0.95, 1.00]	0.00 [0.00, 0.05]
Shen, 2019 #924	773	831	0	8	1.00 [1.00, 1.00]	0.01 [0.00, 0.02]
Trimboli, 2019 #947	256	731	1	70	1.00 [0.98, 1.00]	0.09 [0.07, 0.11]

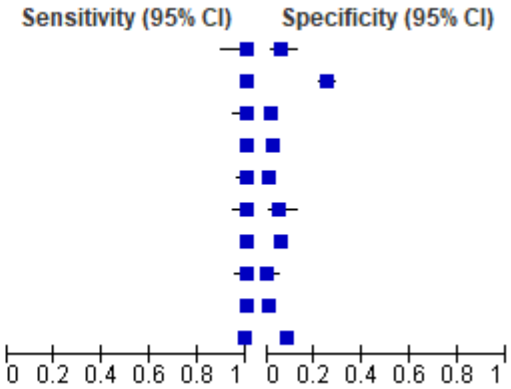


Figure 49: Sensitivity and specificity of EU TIRADS score of 4 or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Borlea, 2020 #712	34	75	1	23	0.97 [0.85, 1.00]	0.23 [0.15, 0.33]
Dobruch-Sobczak, 2019 #740	226	369	3	244	0.99 [0.96, 1.00]	0.40 [0.36, 0.44]
Hekimsoy, 2021 #1826	45	38	17	151	0.73 [0.60, 0.83]	0.80 [0.73, 0.85]
Kuru, 2021 #1809	372	186	144	441	0.72 [0.68, 0.76]	0.70 [0.67, 0.74]
Magri, 2020 #867	79	54	16	155	0.83 [0.74, 0.90]	0.74 [0.68, 0.80]
Pagano, 2020 #888	57	37	11	41	0.84 [0.73, 0.92]	0.53 [0.41, 0.64]
Qi, 2021 #1831	405	439	9	243	0.98 [0.96, 0.99]	0.36 [0.32, 0.39]
Schenke, 2020 #917	75	46	1	23	0.99 [0.93, 1.00]	0.33 [0.22, 0.46]
Shen, 2019 #924	762	562	11	277	0.99 [0.97, 0.99]	0.33 [0.30, 0.36]
Skowronska, 2018 #930	6	8	2	124	0.75 [0.35, 0.97]	0.94 [0.88, 0.97]
Trimboli, 2019 #947	239	257	18	544	0.93 [0.89, 0.96]	0.68 [0.65, 0.71]

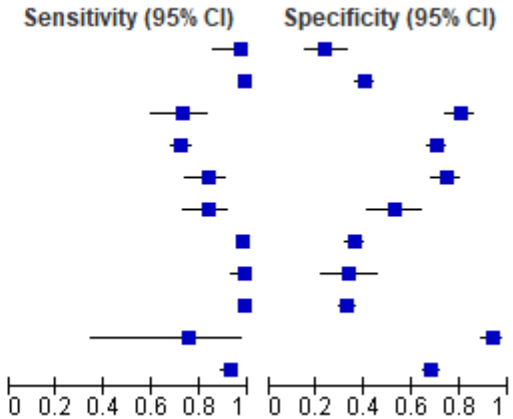


Figure 50: Sensitivity and specificity of EU TIRADS score of 5 or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Borlea, 2020 #712	22	19	13	79	0.63 [0.45, 0.79]	0.81 [0.71, 0.88]
Dobruch-Sobczak, 2019 #740	214	278	15	335	0.93 [0.89, 0.96]	0.55 [0.51, 0.59]
Hekimsoy, 2021 #1826	32	7	30	182	0.52 [0.39, 0.65]	0.96 [0.93, 0.98]
Kuru, 2021 #1809	215	56	301	571	0.42 [0.37, 0.46]	0.91 [0.89, 0.93]
Magri, 2020 #867	51	15	44	194	0.54 [0.43, 0.64]	0.93 [0.88, 0.96]
Pagano, 2020 #888	40	6	28	72	0.59 [0.46, 0.71]	0.92 [0.84, 0.97]
Qi, 2021 #1831	365	304	49	378	0.88 [0.85, 0.91]	0.55 [0.52, 0.59]
Schenke, 2020 #917	74	35	2	34	0.97 [0.91, 1.00]	0.49 [0.37, 0.62]
Shen, 2019 #924	721	160	52	679	0.93 [0.91, 0.95]	0.81 [0.78, 0.84]
Trimboli, 2019 #947	192	27	65	774	0.75 [0.69, 0.80]	0.97 [0.95, 0.98]

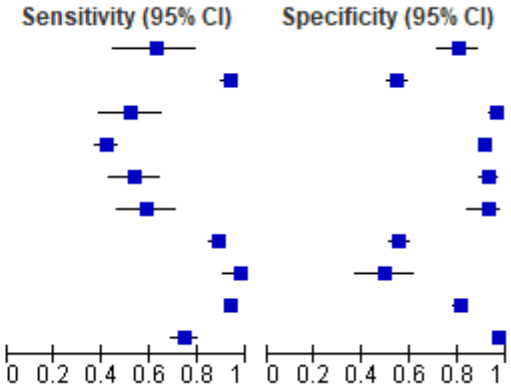


Figure 51: Sensitivity and specificity of Kwak TIRADS score of 3 or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Gao, 2019 #759	1681	855	0	8	1.00 [1.00, 1.00]	0.01 [0.00, 0.02]
Qi, 2021 #1831	414	642	0	40	1.00 [0.99, 1.00]	0.06 [0.04, 0.08]
Schenke, 2018 #919	38	201	0	5	1.00 [0.91, 1.00]	0.02 [0.01, 0.06]
Shen, 2019 #924	773	836	0	3	1.00 [1.00, 1.00]	0.00 [0.00, 0.01]
Wang, 2017 #964	464	393	0	154	1.00 [0.99, 1.00]	0.28 [0.24, 0.32]

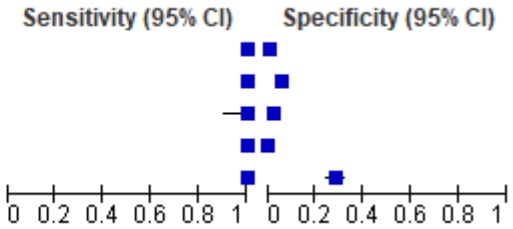


Figure 52: Sensitivity and specificity of Kwak TIRADS score of 4a or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chng, 2018 #727	48	79	3	31	0.94 [0.84, 0.99]	0.28 [0.20, 0.38]
Gao, 2019 #759	1673	452	8	411	1.00 [0.99, 1.00]	0.48 [0.44, 0.51]
Huang, 2020 #795	77	31	0	1	1.00 [0.95, 1.00]	0.03 [0.00, 0.16]
Qi, 2021 #1831	413	493	1	189	1.00 [0.99, 1.00]	0.28 [0.24, 0.31]
Schenke, 2018 #919	37	155	1	51	0.97 [0.86, 1.00]	0.25 [0.19, 0.31]
Schenke, 2020 #917	76	60	0	9	1.00 [0.95, 1.00]	0.13 [0.06, 0.23]
Shen, 2019 #924	767	586	6	253	0.99 [0.98, 1.00]	0.30 [0.27, 0.33]
Wang, 2017 #964	460	260	4	287	0.99 [0.98, 1.00]	0.52 [0.48, 0.57]

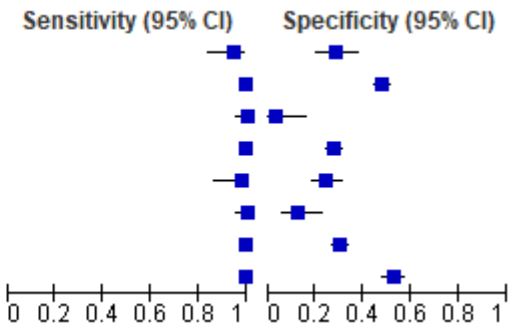


Figure 53: Sensitivity and specificity of Kwak TIRADS score of 4b or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chng, 2018 #727	40	40	10	70	0.80 [0.66, 0.90]	0.64 [0.54, 0.73]
Gao, 2019 #759	1657	329	24	534	0.99 [0.98, 0.99]	0.62 [0.59, 0.65]
Huang, 2020 #795	54	10	23	22	0.70 [0.59, 0.80]	0.69 [0.50, 0.84]
Huang, 2020 #796	136	37	97	122	0.58 [0.52, 0.65]	0.77 [0.69, 0.83]
Qi, 2021 #1831	405	395	9	287	0.98 [0.96, 0.99]	0.42 [0.38, 0.46]
Schenke, 2018 #919	35	86	3	120	0.92 [0.79, 0.98]	0.58 [0.51, 0.65]
Schenke, 2020 #917	76	41	0	28	1.00 [0.95, 1.00]	0.41 [0.29, 0.53]
Shen, 2019 #924	760	480	13	359	0.98 [0.97, 0.99]	0.43 [0.39, 0.46]
Wang, 2017 #964	449	137	15	410	0.97 [0.95, 0.98]	0.75 [0.71, 0.79]
Zhang, 2021 #1807	176	16	27	42	0.87 [0.81, 0.91]	0.72 [0.59, 0.83]

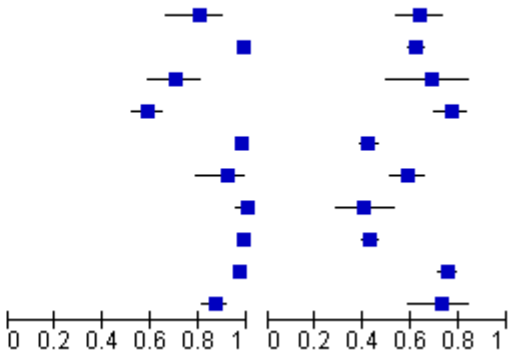


Figure 54: Sensitivity and specificity of Kwak TIRADS score of 4c or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chng, 2018 #727	23	10	27	100	0.46 [0.32, 0.61]	0.91 [0.84, 0.96]
Gao, 2019 #759	1502	203	179	660	0.89 [0.88, 0.91]	0.76 [0.74, 0.79]
Huang, 2020 #795	14	0	63	32	0.18 [0.10, 0.29]	1.00 [0.89, 1.00]
Huang, 2020 #796	79	6	154	153	0.34 [0.28, 0.40]	0.96 [0.92, 0.99]
Qi, 2021 #1831	372	259	42	423	0.90 [0.87, 0.93]	0.62 [0.58, 0.66]
Schenke, 2018 #919	31	25	7	181	0.82 [0.66, 0.92]	0.88 [0.83, 0.92]
Schenke, 2020 #917	74	31	2	38	0.97 [0.91, 1.00]	0.55 [0.43, 0.67]
Shen, 2019 #924	721	116	52	723	0.93 [0.91, 0.95]	0.86 [0.84, 0.88]
Wang, 2017 #964	393	45	71	502	0.85 [0.81, 0.88]	0.92 [0.89, 0.94]

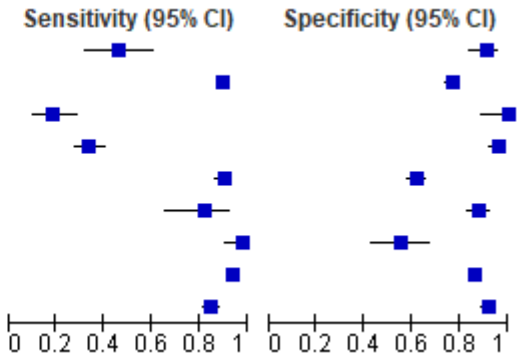


Figure 55: Sensitivity and specificity of Kwak TIRADS score of 5 for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chng, 2018 #727	1	0	49	110	0.02 [0.00, 0.11]	1.00 [0.97, 1.00]
Gao, 2019 #759	244	36	1437	827	0.15 [0.13, 0.16]	0.96 [0.94, 0.97]
Qi, 2021 #1831	46	13	368	669	0.11 [0.08, 0.15]	0.98 [0.97, 0.99]
Schenke, 2018 #919	9	0	29	206	0.24 [0.11, 0.40]	1.00 [0.98, 1.00]
Schenke, 2020 #917	17	4	59	65	0.22 [0.14, 0.33]	0.94 [0.86, 0.98]
Shen, 2019 #924	38	2	735	837	0.05 [0.04, 0.07]	1.00 [0.99, 1.00]
Wang, 2017 #964	48	3	416	544	0.10 [0.08, 0.13]	0.99 [0.98, 1.00]

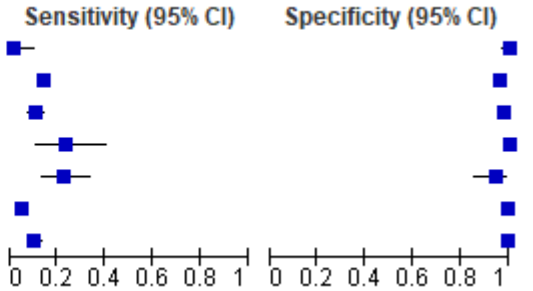


Figure 56: Sensitivity and specificity of ATA 'very low suspicion' or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Kuru, 2021 #1809	516	619	0	8	1.00 [0.99, 1.00]	0.01 [0.01, 0.02]
Pagano, 2020 #888	68	78	0	0	1.00 [0.95, 1.00]	0.00 [0.00, 0.05]
Paker, 2021 #1814	106	106	0	10	1.00 [0.97, 1.00]	0.09 [0.04, 0.15]

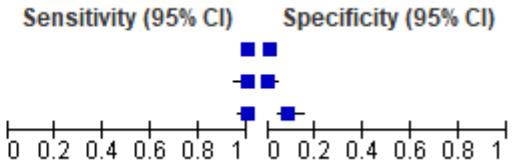


Figure 57: Sensitivity and specificity of ATA 'low suspicion' or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ahmadi, 2019 #696	88	215	1	20	0.99 [0.94, 1.00]	0.09 [0.05, 0.13]
Chng, 2018 #727	49	91	1	19	0.98 [0.89, 1.00]	0.17 [0.11, 0.26]
Gao, 2019 #759	1681	679	0	184	1.00 [1.00, 1.00]	0.21 [0.19, 0.24]
Kuru, 2021 #1809	489	476	27	151	0.95 [0.92, 0.97]	0.24 [0.21, 0.28]
Liu, 2020 #851	28	62	0	0	1.00 [0.88, 1.00]	0.00 [0.00, 0.06]
Pagano, 2020 #888	68	74	0	4	1.00 [0.95, 1.00]	0.05 [0.01, 0.13]
Paker, 2021 #1814	103	65	3	51	0.97 [0.92, 0.99]	0.44 [0.35, 0.53]
Shen, 2019 #924	767	647	6	192	0.99 [0.98, 1.00]	0.23 [0.20, 0.26]

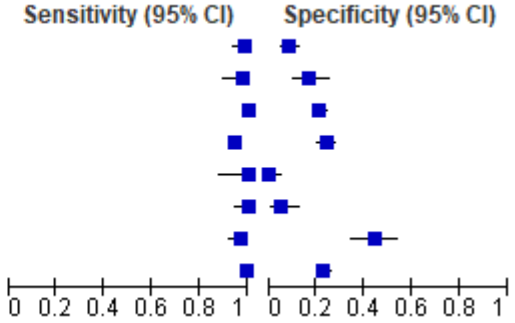


Figure 58: Sensitivity and specificity of ATA 'intermediate suspicion' or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ahmadi, 2019 #696	68	55	20	180	0.77 [0.67, 0.86]	0.77 [0.71, 0.82]
Chng, 2018 #727	39	31	11	79	0.78 [0.64, 0.88]	0.72 [0.62, 0.80]
Gao, 2019 #759	1661	340	20	523	0.99 [0.98, 0.99]	0.61 [0.57, 0.64]
Kuru, 2021 #1809	357	164	159	463	0.69 [0.65, 0.73]	0.74 [0.70, 0.77]
Liu, 2020 #851	20	14	8	48	0.71 [0.51, 0.87]	0.77 [0.65, 0.87]
Pagano, 2020 #888	57	37	11	41	0.84 [0.73, 0.92]	0.53 [0.41, 0.64]
Paker, 2021 #1814	95	26	11	90	0.90 [0.82, 0.95]	0.78 [0.69, 0.85]
Shen, 2019 #924	750	498	23	341	0.97 [0.96, 0.98]	0.41 [0.37, 0.44]
Shreyamsa, 2020 #927	25	17	10	116	0.71 [0.54, 0.85]	0.87 [0.80, 0.92]

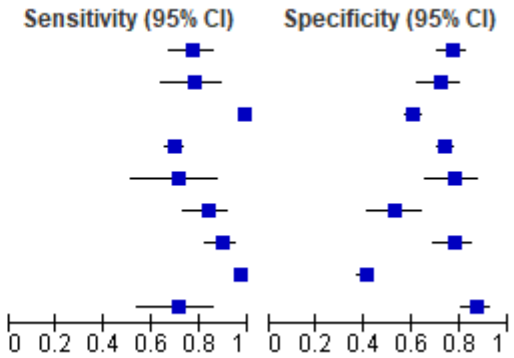


Figure 59: Sensitivity and specificity of ATA 'high suspicion' for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ahmadi, 2019 #696	44	8	44	227	0.50 [0.39, 0.61]	0.97 [0.93, 0.99]
Chng, 2018 #727	27	13	23	97	0.54 [0.39, 0.68]	0.88 [0.81, 0.94]
Gao, 2019 #759	1606	223	75	640	0.96 [0.94, 0.96]	0.74 [0.71, 0.77]
Kuru, 2021 #1809	158	24	358	603	0.31 [0.27, 0.35]	0.96 [0.94, 0.98]
Liu, 2020 #851	11	2	17	60	0.39 [0.22, 0.59]	0.97 [0.89, 1.00]
Pagano, 2020 #888	37	2	31	76	0.54 [0.42, 0.67]	0.97 [0.91, 1.00]
Paker, 2021 #1814	70	6	36	110	0.66 [0.56, 0.75]	0.95 [0.89, 0.98]
Shen, 2019 #924	708	150	65	689	0.92 [0.89, 0.93]	0.82 [0.79, 0.85]
Shreyamsa, 2020 #927	24	7	11	126	0.69 [0.51, 0.83]	0.95 [0.89, 0.98]

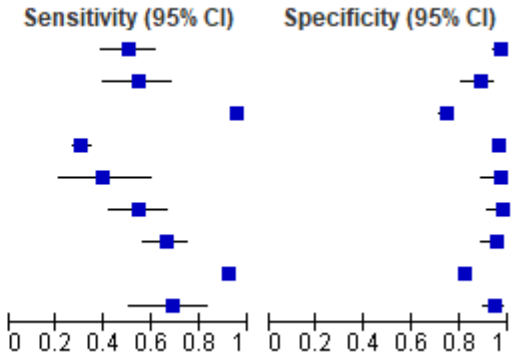


Figure 60: Sensitivity and specificity of Horvath TIRADS 3 or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Borlea, 2020 #712	34	80	1	18	0.97 [0.85, 1.00]	0.18 [0.11, 0.27]
Horvath, 2017 #793	272	114	0	116	1.00 [0.99, 1.00]	0.50 [0.44, 0.57]
Wang, 2017 #964	464	480	0	67	1.00 [0.99, 1.00]	0.12 [0.10, 0.15]

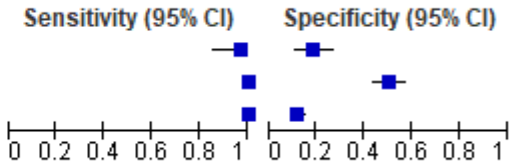


Figure 61: Sensitivity and specificity of Horvath TIRADS 4a for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Borlea, 2020 #712	34	75	1	23	0.97 [0.85, 1.00]	0.23 [0.15, 0.33]
Garcia-Monco Fernandez, 2018 #761	49	14	26	174	0.65 [0.53, 0.76]	0.93 [0.88, 0.96]
Horvath, 2017 #793	271	59	1	171	1.00 [0.98, 1.00]	0.74 [0.68, 0.80]
Wang, 2017 #964	459	279	5	268	0.99 [0.98, 1.00]	0.49 [0.45, 0.53]

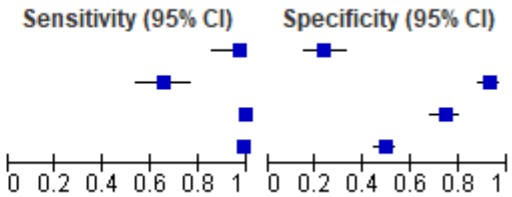


Figure 62: Sensitivity and specificity of Horvath TIRADS 4b and higher for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Borlea, 2020 #712	28	37	7	61	0.80 [0.63, 0.92]	0.62 [0.52, 0.72]
Horvath, 2017 #793	270	43	2	187	0.99 [0.97, 1.00]	0.81 [0.76, 0.86]
Wang, 2017 #964	448	158	16	389	0.97 [0.94, 0.98]	0.71 [0.67, 0.75]

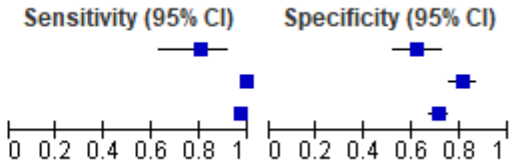


Figure 63: Sensitivity and specificity of Horvath TIRADS 4c and more for diagnosis of malignancy in thyroid cancer

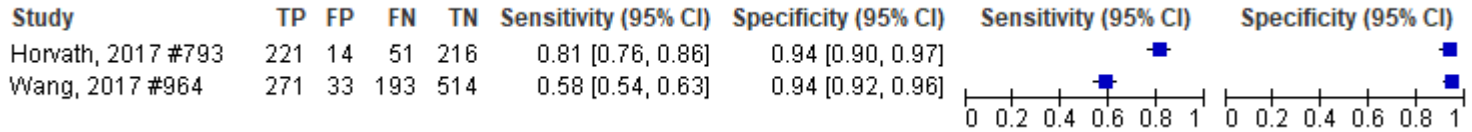


Figure 64: Sensitivity and specificity of Horvath TIRADS 5 for diagnosis of malignancy in thyroid cancer

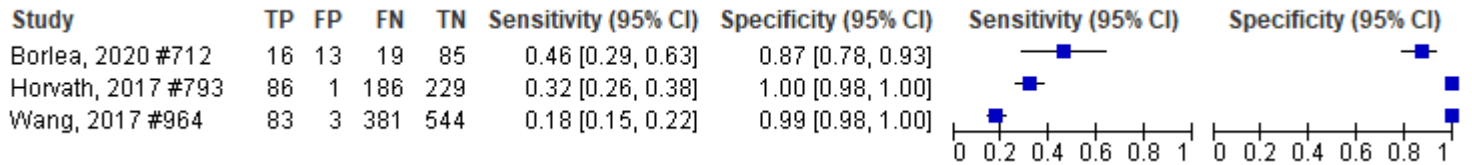


Figure 65: Sensitivity and specificity of Park TIRADS 2 or higher for diagnosis of malignancy in thyroid cancer

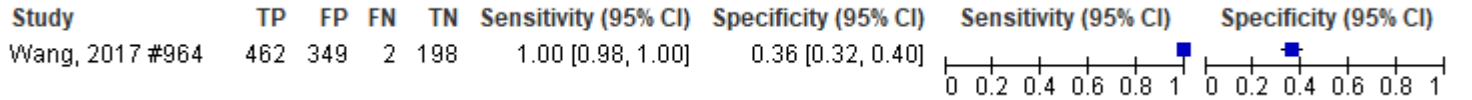


Figure 66: Sensitivity and specificity of Park TIRADS 3 or higher for diagnosis of malignancy in thyroid cancer

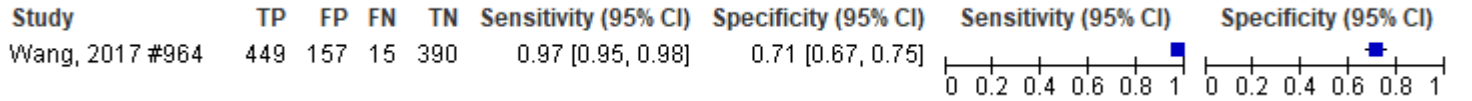


Figure 67: Sensitivity and specificity of Park TIRADS 4 or higher for diagnosis of malignancy in thyroid cancer

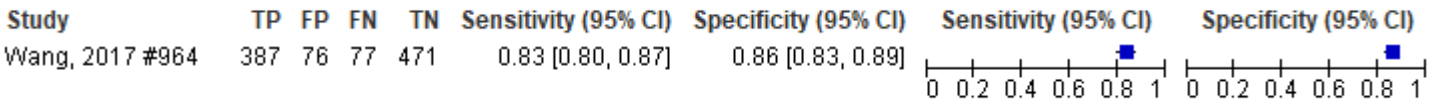


Figure 68: Sensitivity and specificity of Park TIRADS 5 for diagnosis of malignancy in thyroid cancer

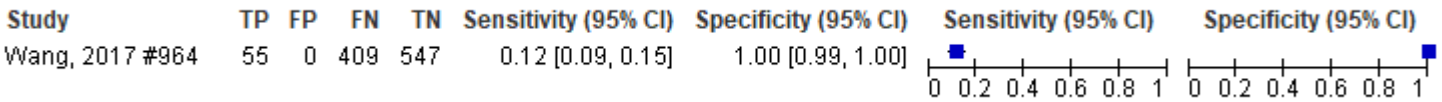


Figure 69: Sensitivity and specificity of Russ TIRADS 3 or higher for diagnosis of malignancy in thyroid cancer

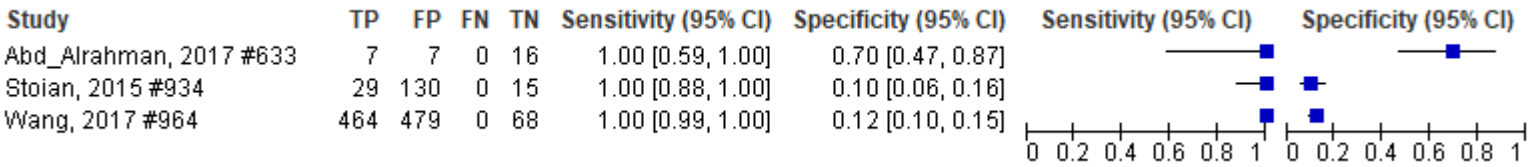


Figure 70: Sensitivity and specificity of Russ TIRADS 4a or higher for diagnosis of malignancy in thyroid cancer

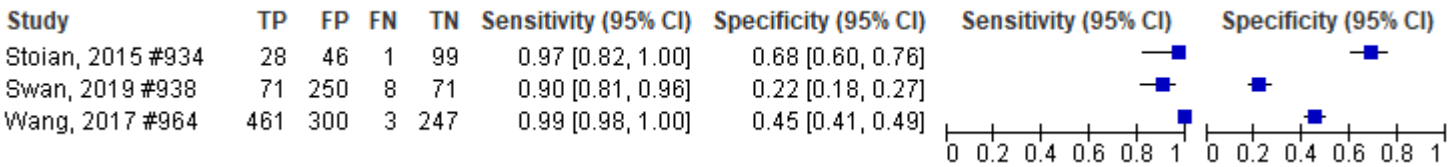


Figure 71: Sensitivity and specificity of Russ TIRADS 4b or higher for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abd_Alrahman, 2017 #633	7	1	0	22	1.00 [0.59, 1.00]	0.96 [0.78, 1.00]
Stoian, 2015 #934	25	3	4	142	0.86 [0.68, 0.96]	0.98 [0.94, 1.00]
Wang, 2017 #964	419	86	45	461	0.90 [0.87, 0.93]	0.84 [0.81, 0.87]

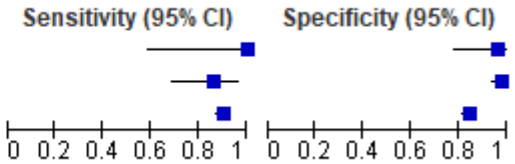


Figure 72: Sensitivity and specificity of Russ TIRADS 4c or higher for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abd_Alrahman, 2017 #633	4	0	3	23	0.57 [0.18, 0.90]	1.00 [0.85, 1.00]

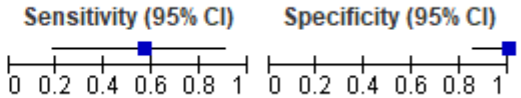


Figure 73: Sensitivity and specificity of Russ TIRADS 5 for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Stoian, 2015 #934	16	1	13	144	0.55 [0.36, 0.74]	0.99 [0.96, 1.00]
Wang, 2017 #964	120	6	344	541	0.26 [0.22, 0.30]	0.99 [0.98, 1.00]

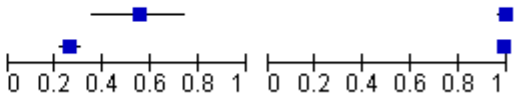


Figure 74: Sensitivity and specificity of French TIRADS 3 or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Borlea, 2020 #712	35	92	0	6	1.00 [0.90, 1.00]	0.06 [0.02, 0.13]

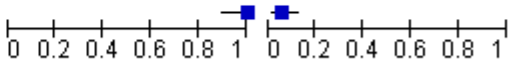


Figure 75: Sensitivity and specificity of French TIRADS 4a or more for diagnosis of malignancy in thyroid cancer

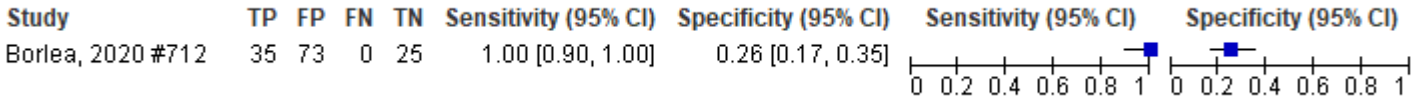


Figure 76: Sensitivity and specificity of French TIRADS 4b or more for diagnosis of malignancy in thyroid cancer

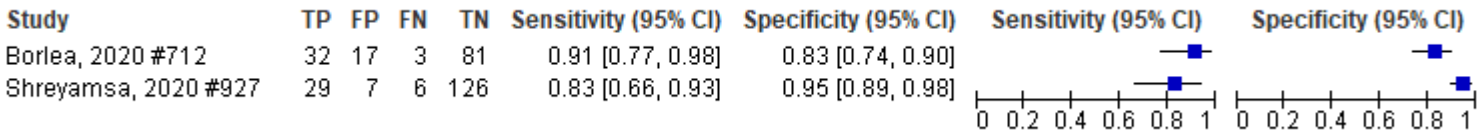


Figure 77: Sensitivity and specificity of French TIRADS 4c or more for diagnosis of malignancy in thyroid cancer

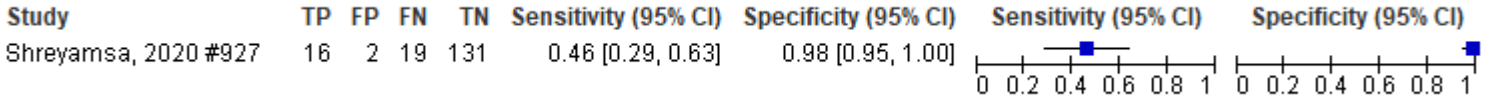


Figure 78: Sensitivity and specificity of French TIRADS 5 for diagnosis of malignancy in thyroid cancer

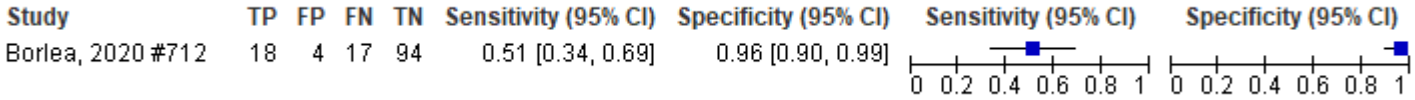


Figure 79: Sensitivity and specificity of Korean TIRADS 3 or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Kuru, 2021 #1809	516	604	0	23	1.00 [0.99, 1.00]	0.04 [0.02, 0.05]
Qi, 2021 #1831	414	608	0	74	1.00 [0.99, 1.00]	0.11 [0.09, 0.13]

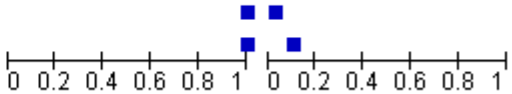


Figure 80: Sensitivity and specificity of Korean TIRADS 4 or more for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Kuru, 2021 #1809	358	167	158	460	0.69 [0.65, 0.73]	0.73 [0.70, 0.77]
Qi, 2021 #1831	404	398	10	284	0.98 [0.96, 0.99]	0.42 [0.38, 0.45]
Shreyamsa, 2020 #927	26	18	9	115	0.74 [0.57, 0.88]	0.86 [0.79, 0.92]

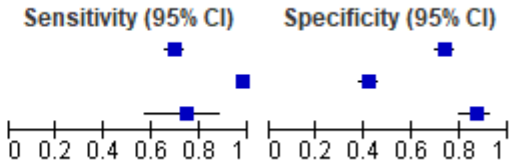


Figure 81: Sensitivity and specificity of Korean TIRADS 5 for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Kuru, 2021 #1809	159	22	357	605	0.31 [0.27, 0.35]	0.96 [0.95, 0.98]
Qi, 2021 #1831	338	227	76	455	0.82 [0.78, 0.85]	0.67 [0.63, 0.70]
Shreyamsa, 2020 #927	14	2	21	131	0.40 [0.24, 0.58]	0.98 [0.95, 1.00]

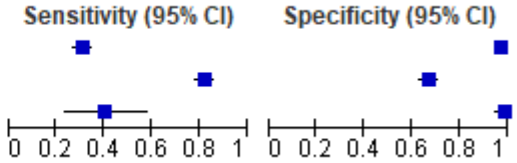


Figure 82: Sensitivity and specificity of C TIRADS 3 and above for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Qi, 2021 #1831	414	638	0	44	1.00 [0.99, 1.00]	0.06 [0.05, 0.09]

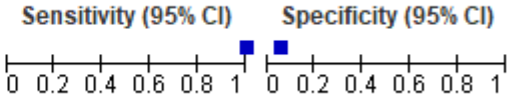


Figure 83: Sensitivity and specificity of C TIRADS 4a and above for diagnosis of malignancy in thyroid cancer

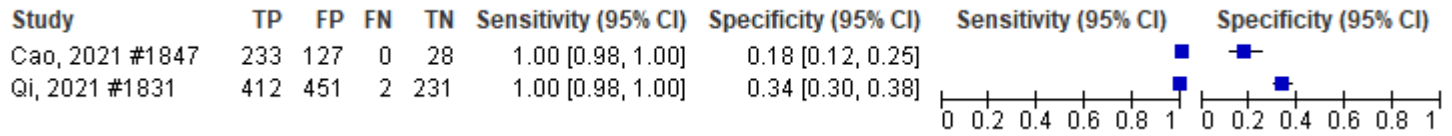


Figure 84: Sensitivity and specificity of C TIRADS 4b and above for diagnosis of malignancy in thyroid cancer

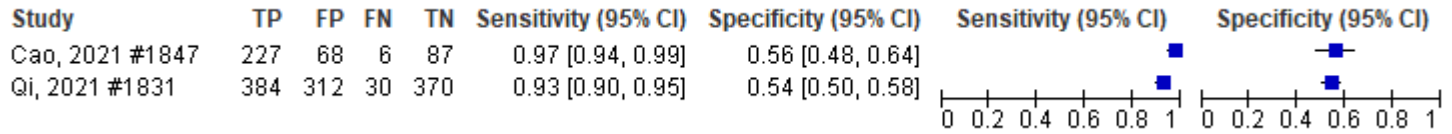


Figure 85: Sensitivity and specificity of C TIRADS 4c and above for diagnosis of malignancy in thyroid cancer

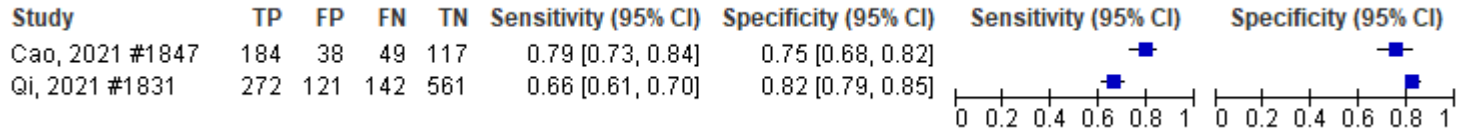


Figure 86: Sensitivity and specificity of C TIRADS 5 for diagnosis of malignancy in thyroid cancer

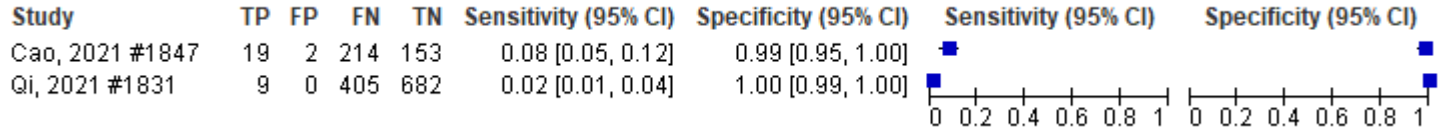


Figure 87: Sensitivity and specificity of AI TIRADS 2 and above for diagnosis of malignancy in thyroid cancer

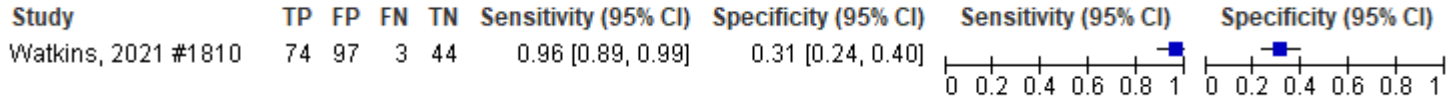


Figure 88: Sensitivity and specificity of AI TIRADS 3 and above for diagnosis of malignancy in thyroid cancer

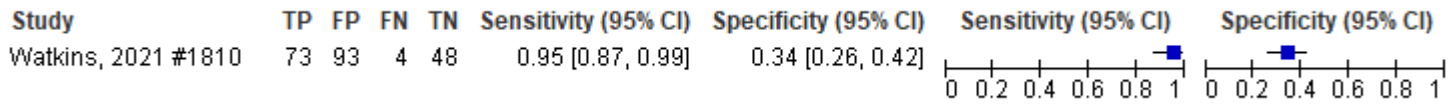


Figure 89: Sensitivity and specificity of AI TIRADS 4 and above for diagnosis of malignancy in thyroid cancer

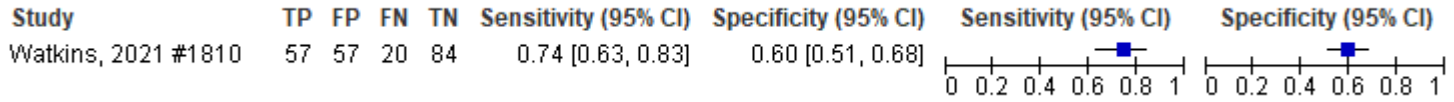


Figure 90: Sensitivity and specificity of AI TIRADS 5 for diagnosis of malignancy in thyroid cancer

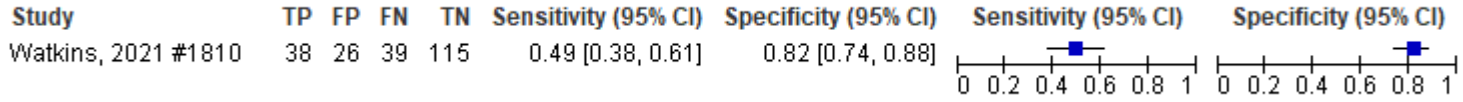


Figure 91: Sensitivity and specificity of KTA 3 and above for diagnosis of malignancy in thyroid cancer

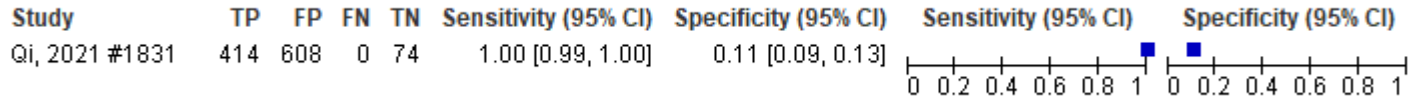


Figure 92: Sensitivity and specificity of KTA 4 and above for diagnosis of malignancy in thyroid cancer

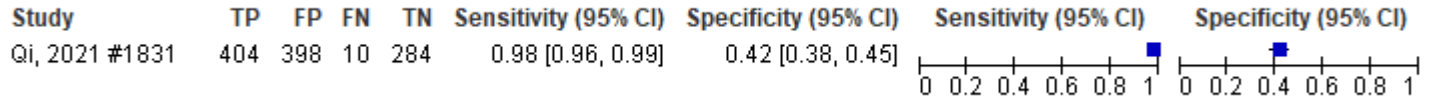


Figure 93: Sensitivity and specificity of KTA 5 for diagnosis of malignancy in thyroid cancer

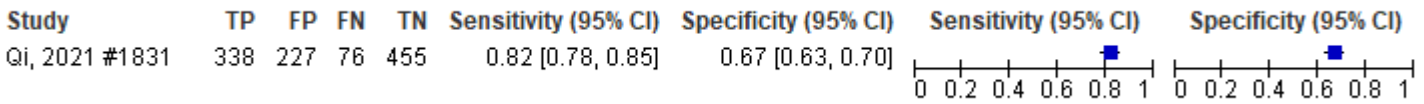


Figure 94: Sensitivity and specificity of TIRADS (0-10 version) 3 or more for diagnosis of malignancy in thyroid cancer

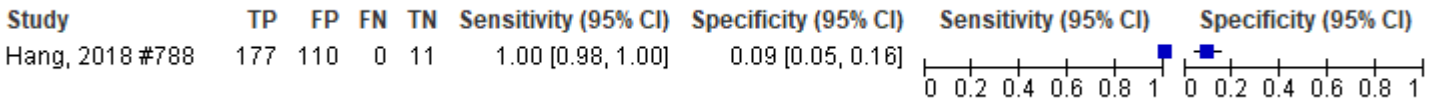


Figure 95: Sensitivity and specificity of TIRADS (0-10 version) 4 or more for diagnosis of malignancy in thyroid cancer

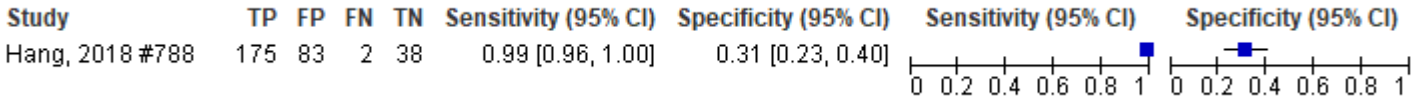


Figure 96: Sensitivity and specificity of TIRADS (0-10 version) 5 or more for diagnosis of malignancy in thyroid cancer

Figure 97: Sensitivity and specificity of TIRADS (0-10 version) 6 or more for diagnosis of malignancy in thyroid cancer

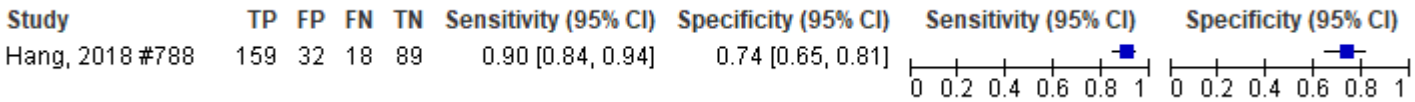


Figure 98: Sensitivity and specificity of TIRADS (0-10 version) 7 or more for diagnosis of malignancy in thyroid cancer

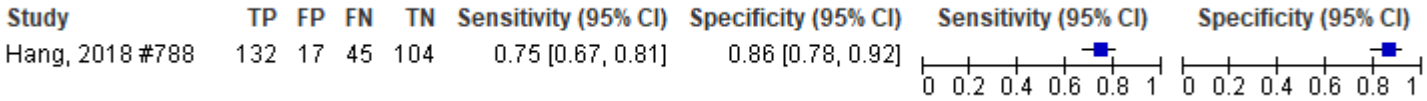


Figure 99: Sensitivity and specificity of TIRADS (0-10 version) 8 or more for diagnosis of malignancy in thyroid cancer

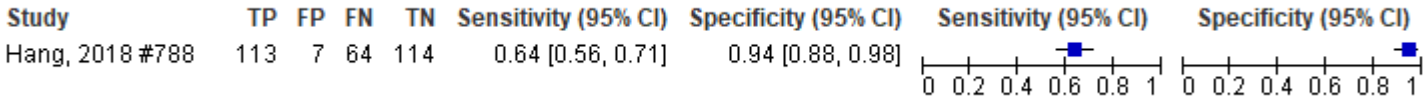


Figure 100: Sensitivity and specificity of TIRADS (0-10 version) 9 or more for diagnosis of malignancy in thyroid cancer

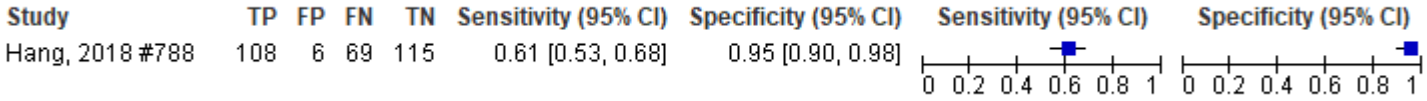


Figure 101: Sensitivity and specificity of TIRADS (0-10 version) 10 for diagnosis of malignancy in thyroid cancer

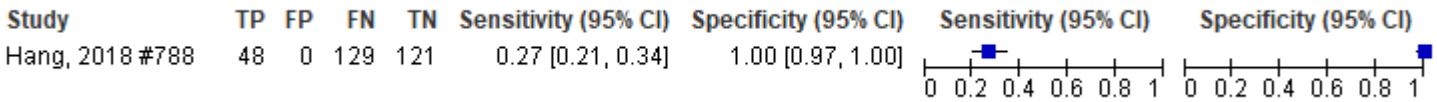


Figure 102: Sensitivity and specificity of AACE/ACE/AME 2016 'moderate or higher' for diagnosis of malignancy in thyroid cancer

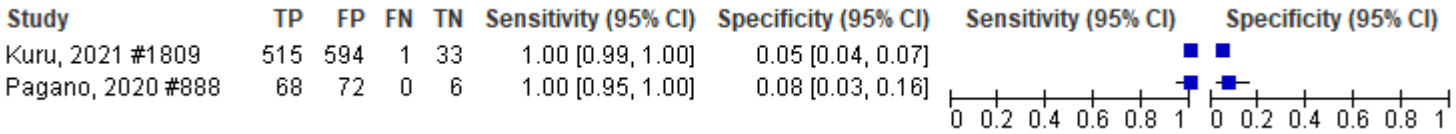


Figure 103: Sensitivity and specificity of AACE/ACE/AME 2016 'high' for diagnosis of malignancy in thyroid cancer

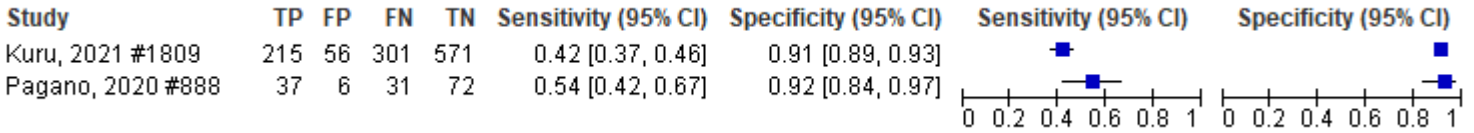


Figure 104: Sensitivity and specificity of BTA intermediate suspicion and higher for diagnosis of malignancy in thyroid cancer

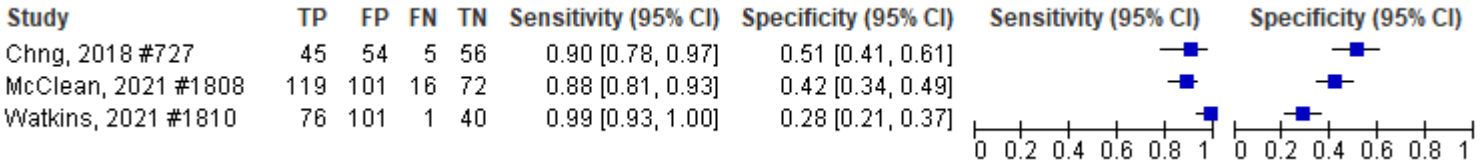


Figure 105: Sensitivity and specificity of BTA suspicious and higher for diagnosis of malignancy in thyroid cancer

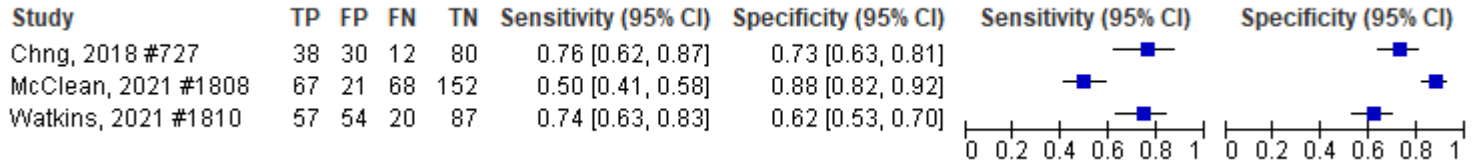


Figure 106: Sensitivity and specificity of BTA malignant for diagnosis of malignancy in thyroid cancer

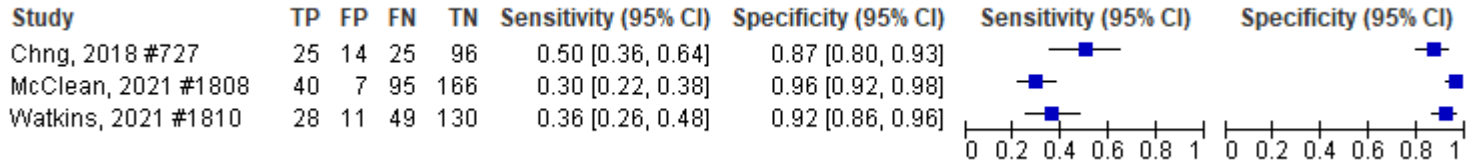


Figure 107: Sensitivity and specificity of SN-US class II and above for diagnosis of malignancy in thyroid cancer

Figure 108: Sensitivity and specificity of SN-US class III and above for diagnosis of malignancy in thyroid cancer

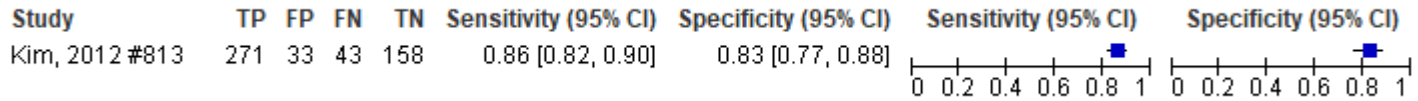


Figure 109: Sensitivity and specificity of SN-US class IV and above for diagnosis of malignancy in thyroid cancer

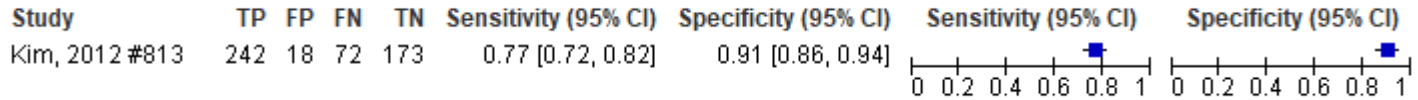


Figure 110: Sensitivity and specificity of SN-US class V and above for diagnosis of malignancy in thyroid cancer

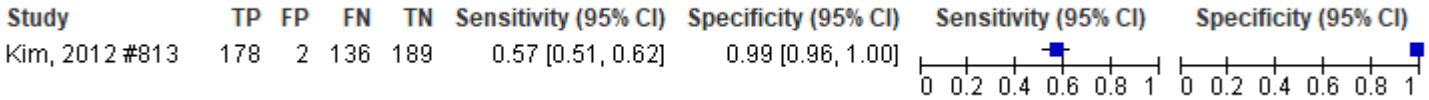


Figure 111: Sensitivity and specificity of R staging: R2 and above for diagnosis of malignancy in thyroid cancer

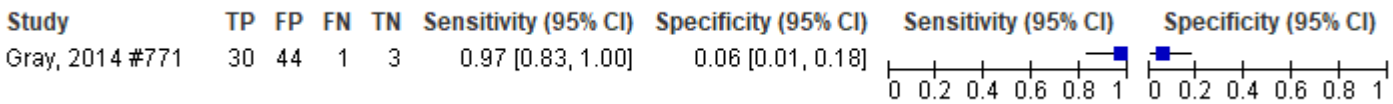


Figure 112: Sensitivity and specificity of R staging: R3 and above for diagnosis of malignancy in thyroid cancer

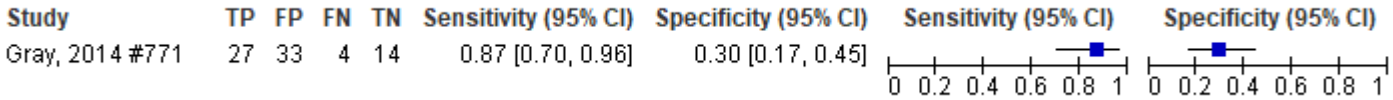


Figure 113: Sensitivity and specificity of R staging: R4 and above for diagnosis of malignancy in thyroid cancer

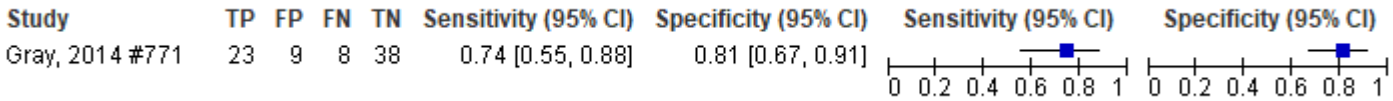


Figure 114: Sensitivity and specificity of R staging: R5 for diagnosis of malignancy in thyroid cancer

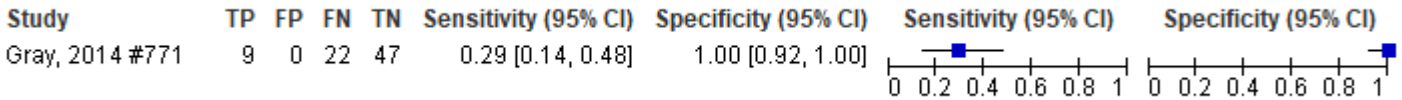


Figure 115: Sensitivity and specificity of TMC-RSS category 2 and above for diagnosis of malignancy in thyroid cancer

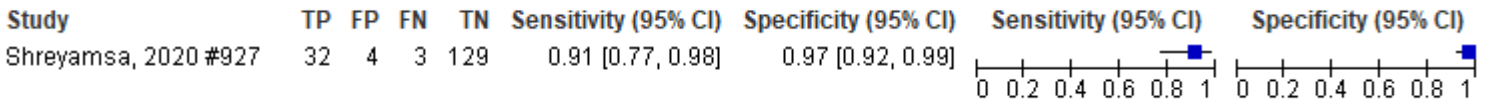


Figure 116: Sensitivity and specificity of TMC-RSS category 3 and above for diagnosis of malignancy in thyroid cancer

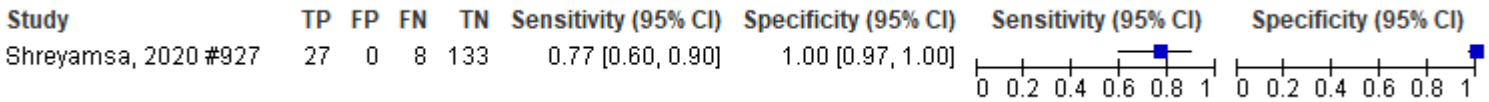


Figure 117: Sensitivity and specificity of Pathirana score of 5 and above for diagnosis of malignancy in thyroid cancer

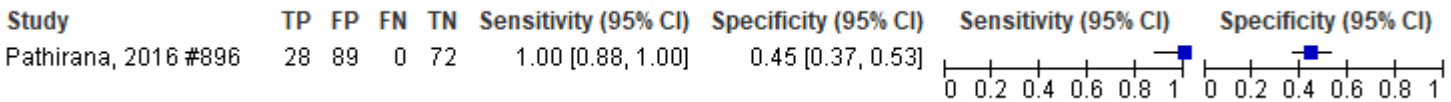


Figure 118: Sensitivity and specificity of Pathirana score of 6 and above for diagnosis of malignancy in thyroid cancer

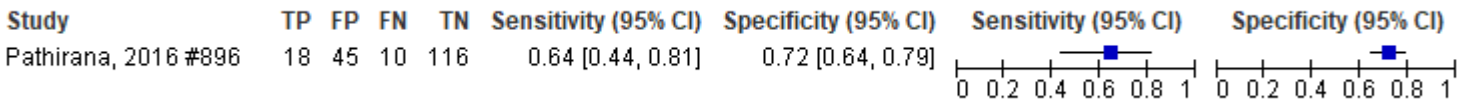
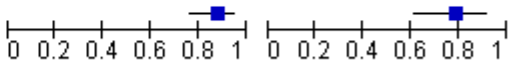


Figure 119: Sensitivity and specificity of low-level echo, ‘vertical/horizontal >1’, fuzzy boundary, microcalcification and grade IV blood flow (equivalent to TIRADS grades 4 and 5) for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Li, 2017 #835	49	7	7	26	0.88 [0.76, 0.95]	0.79 [0.61, 0.91]		



DOPPLER US

Figure 120: Sensitivity and specificity of any blood flow for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Deng, 2018 #663	42	37	3	25	0.93 [0.82, 0.99]	0.40 [0.28, 0.54]		
Giammanco, 2002 #763	7	89	0	29	1.00 [0.59, 1.00]	0.25 [0.17, 0.33]		
Li, 2021 #1833	135	52	137	32	0.50 [0.44, 0.56]	0.38 [0.28, 0.49]		
Lyshchik, 2007 #863	34	30	12	10	0.74 [0.59, 0.86]	0.25 [0.13, 0.41]		
Ma, 2014 #865	34	22	45	34	0.43 [0.32, 0.55]	0.61 [0.47, 0.74]		
Ma, 2017 #864	73	61	21	17	0.78 [0.68, 0.86]	0.22 [0.13, 0.33]		
Shao, 2015 #923	97	94	106	215	0.48 [0.41, 0.55]	0.70 [0.64, 0.75]		
Swan, 2019 #938	63	261	16	64	0.80 [0.69, 0.88]	0.20 [0.16, 0.24]		

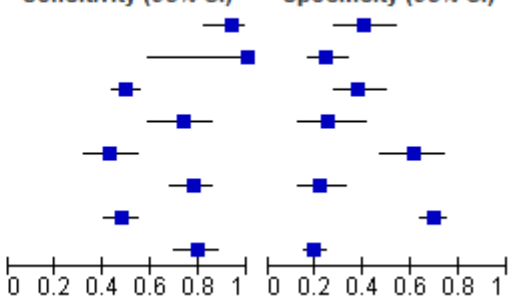


Figure 121: Sensitivity and specificity of central blood flow only for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abd_Alrahman, 2017 #633	4	1	3	22	0.57 [0.18, 0.90]	0.96 [0.78, 1.00]
Appetecchia, 2006 #699	33	109	3	58	0.92 [0.78, 0.98]	0.35 [0.28, 0.42]
Bakari, 2018 #659	53	15	1	9	0.98 [0.90, 1.00]	0.38 [0.19, 0.59]
Chng, 2018 #727	12	14	39	98	0.24 [0.13, 0.37]	0.88 [0.80, 0.93]
Giammanco, 2002 #763	6	28	1	90	0.86 [0.42, 1.00]	0.76 [0.68, 0.84]
Hong, 2009 #792	17	40	32	56	0.35 [0.22, 0.50]	0.58 [0.48, 0.68]
Kalantari, 2018 #811	7	28	2	26	0.78 [0.40, 0.97]	0.48 [0.34, 0.62]
Kong, 2017 #827	60	3	19	31	0.76 [0.65, 0.85]	0.91 [0.76, 0.98]
Li, 2016 #838	202	75	222	263	0.48 [0.43, 0.53]	0.78 [0.73, 0.82]
Liu, 2014 #853	8	11	11	34	0.42 [0.20, 0.67]	0.76 [0.60, 0.87]
Lyshchik, 2007 #863	30	19	16	21	0.65 [0.50, 0.79]	0.53 [0.36, 0.68]
Ma, 2017 #864	13	12	66	44	0.16 [0.09, 0.26]	0.79 [0.66, 0.88]
Maia, 2011 #654	17	13	33	80	0.34 [0.21, 0.49]	0.86 [0.77, 0.92]
Maimaiti, 2016 #655	190	114	166	130	0.53 [0.48, 0.59]	0.53 [0.47, 0.60]
Mohamed, 2013 #672	19	7	9	10	0.68 [0.48, 0.84]	0.59 [0.33, 0.82]
Mohey, 2013 #674	1	1	14	30	0.07 [0.00, 0.32]	0.97 [0.83, 1.00]
Pei, 2019 #898	108	12	10	66	0.92 [0.85, 0.96]	0.85 [0.75, 0.92]
Ragazzoni, 2012 #903	25	33	15	59	0.63 [0.46, 0.77]	0.64 [0.53, 0.74]
Rago, 1998 #905	20	38	10	36	0.67 [0.47, 0.83]	0.49 [0.37, 0.61]
Schleder, 2015 #920	14	14	12	61	0.54 [0.33, 0.73]	0.81 [0.71, 0.89]
Szczepanek-Parulska, 2013 #940	7	25	15	346	0.32 [0.14, 0.55]	0.93 [0.90, 0.96]
Tuan, 2020 #948	41	12	33	8	0.55 [0.43, 0.67]	0.40 [0.19, 0.64]
Veyrieres, 2012 #950	20	82	15	180	0.57 [0.39, 0.74]	0.69 [0.63, 0.74]
Xing, 2011 #973	26	37	19	16	0.58 [0.42, 0.72]	0.30 [0.18, 0.44]
Xu, 2014 #975	26	71	90	254	0.22 [0.15, 0.31]	0.78 [0.73, 0.83]

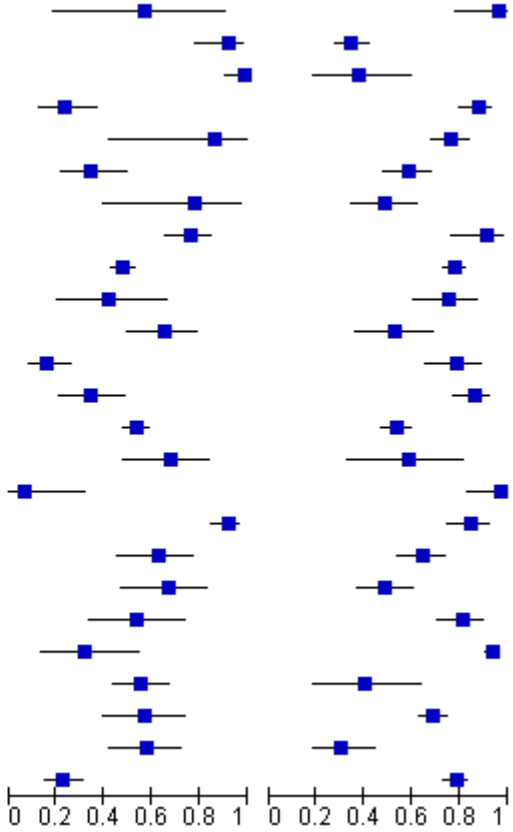


Figure 122 Sensitivity and specificity of mean systolic blood velocity of 33.5 m/s or more for diagnosis of malignancy in thyroid cancer

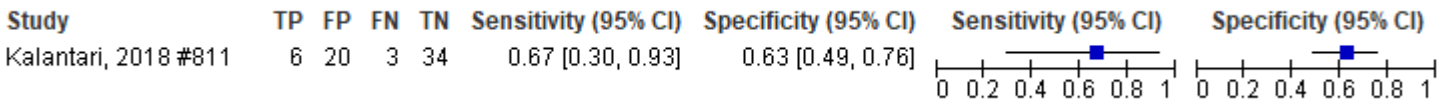


Figure 123: Sensitivity and specificity of pulsatility index 0.92 or more for diagnosis of malignancy in thyroid cancer

Figure 124: Sensitivity and specificity of pulsatility index 0.945 or more for diagnosis of malignancy in thyroid cancer

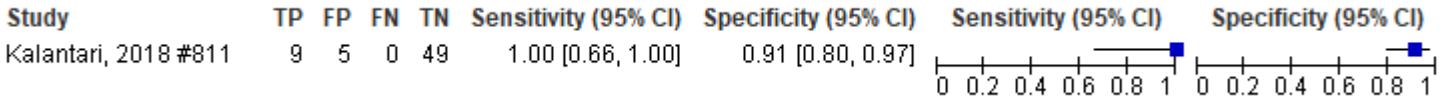


Figure 125: Sensitivity and specificity of normalised VI of 0.14 or more for diagnosis of malignancy in thyroid cancer

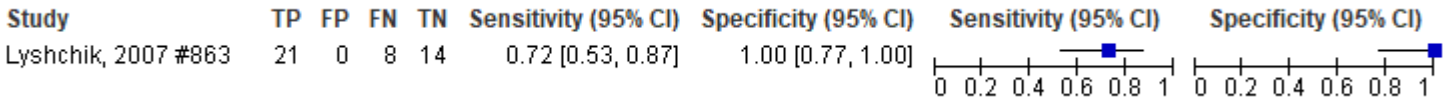


Figure 126: Sensitivity and specificity of normalised VI of 0.278 or more for diagnosis of malignancy in thyroid cancer

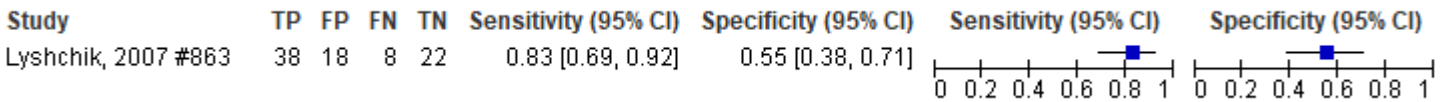


Figure 127: Sensitivity and specificity of resistive index of 0.68 or more for diagnosis of malignancy in thyroid cancer

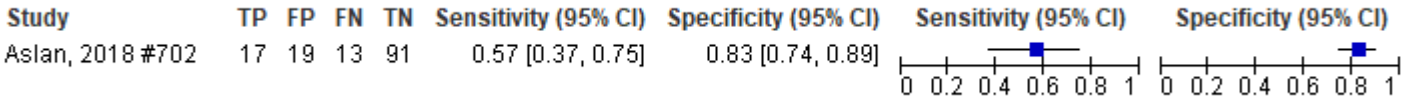


Figure 128: Sensitivity and specificity of resistive index of 0.715 or more for diagnosis of malignancy in thyroid cancer

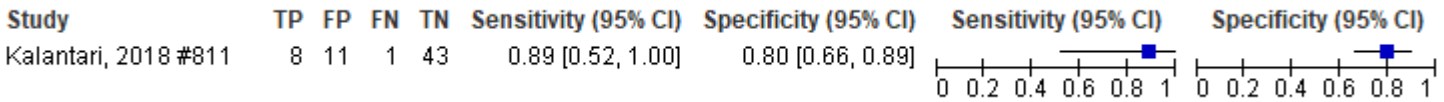


Figure 129: Sensitivity and specificity of resistive index of 0.75 or more for diagnosis of malignancy in thyroid cancer

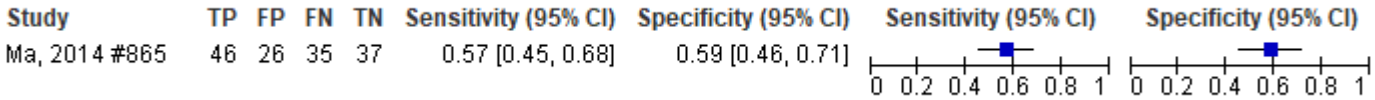


Figure 130: Sensitivity and specificity of systolic /diastolic ratio of 3.11 or more for diagnosis of malignancy in thyroid cancer

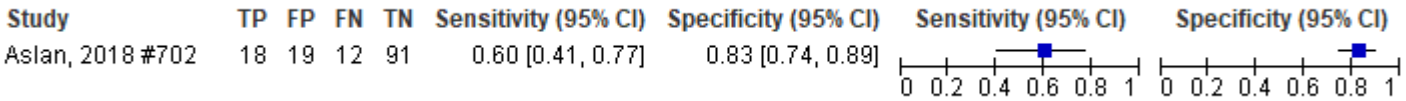
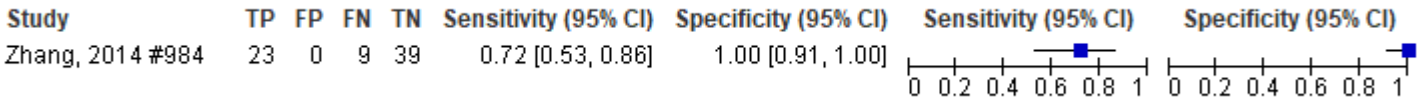


Figure 131: Sensitivity and specificity of colour doppler VTQ of 2.910 m/s or more for diagnosis of malignancy in thyroid cancer



CONTRAST ENHANCED ULTRASOUND

Figure 132: Sensitivity and specificity of CEUS heterogeneous AND low enhancement pattern for diagnosis of malignancy in thyroid cancer

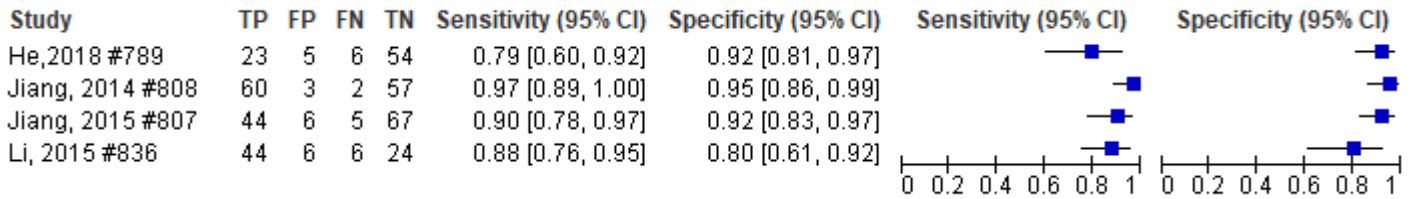


Figure 133: Sensitivity and specificity of CEUS heterogeneous enhancement OR low enhancement pattern for diagnosis of malignancy in thyroid cancer

Figure 134: Sensitivity and specificity of CEUS heterogeneous enhancement only for diagnosis of malignancy in thyroid cancer

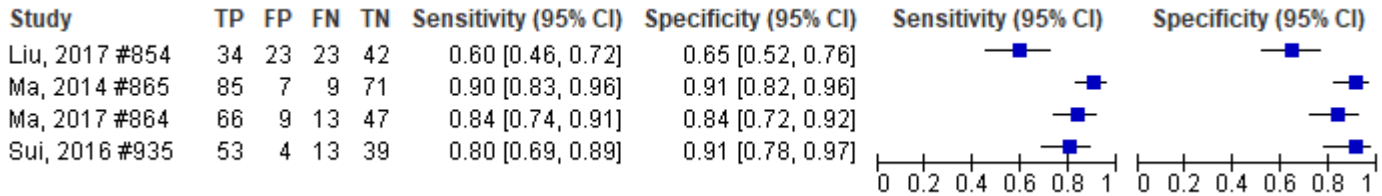


Figure 135: Sensitivity and specificity of CEUS hypo-enhancement for diagnosis of malignancy in thyroid cancer

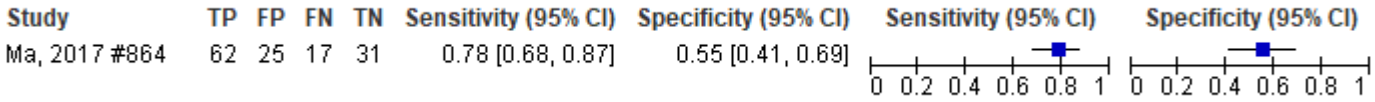


Figure 136: Sensitivity and specificity of CEUS low enhancement, weak enhancement, late enhancement and uneven enhancement for diagnosis of malignancy in thyroid cancer

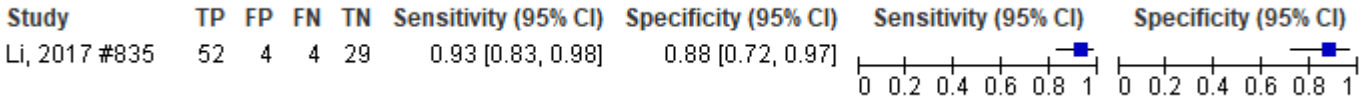


Figure 137: Sensitivity and specificity of incomplete or no ring-enhancement pattern for diagnosis of malignancy in thyroid cancer

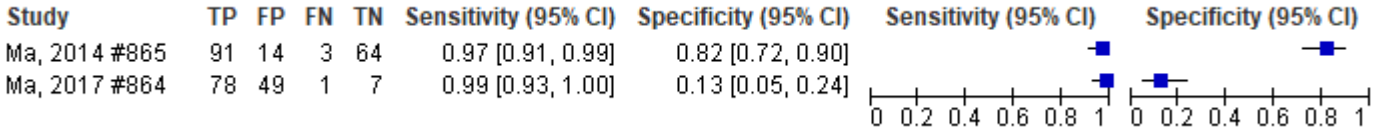


Figure 138: Sensitivity and specificity of CEUS irregular shape for diagnosis of malignancy in thyroid cancer

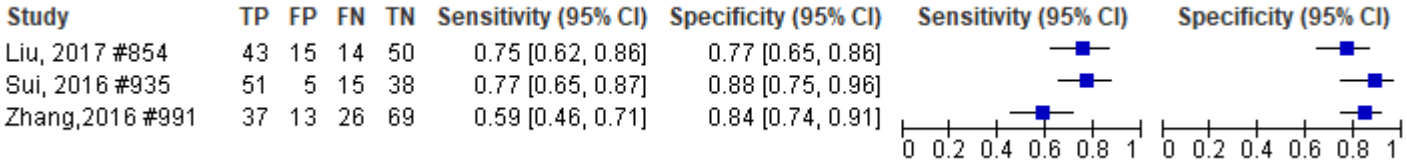


Figure 139: Sensitivity and specificity of CEUS ill-defined enhancement border for diagnosis of malignancy in thyroid cancer

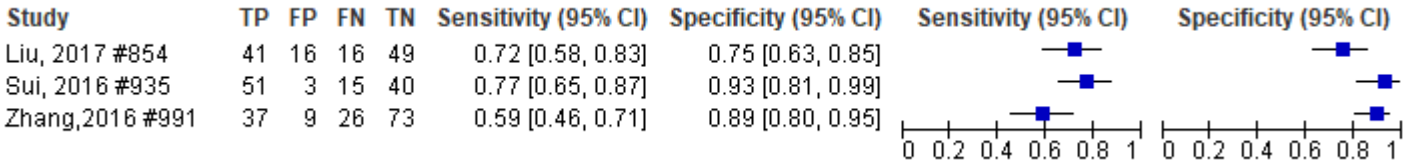


Figure 140: Sensitivity and specificity of CEUS peak interior echogenicity on contrast enhanced US – hypoechoic for diagnosis of malignancy in thyroid cancer

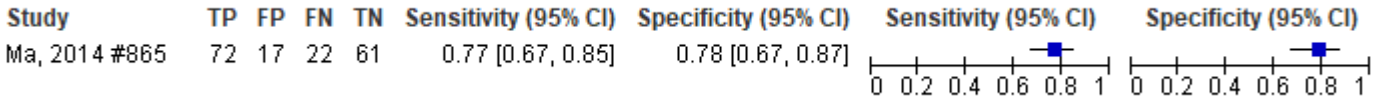


Figure 141: Sensitivity and specificity of CEUS peak peripheral echogenicity on contrast enhanced US – hypoechoic for diagnosis of malignancy in thyroid cancer

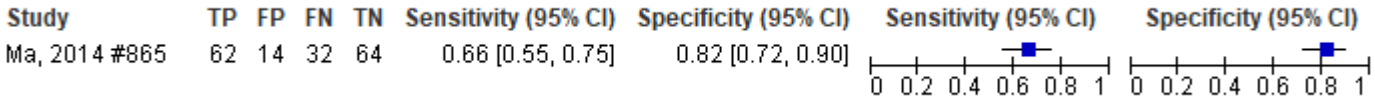


Figure 142: Sensitivity and specificity of later relative arrival time of nodule on contrast enhanced US for diagnosis of malignancy in thyroid cancer

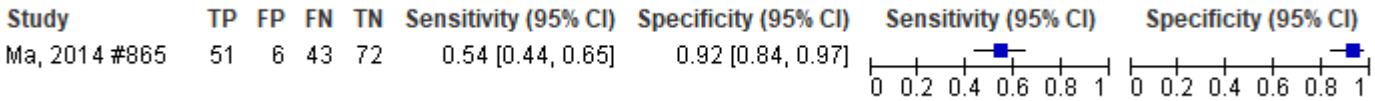


Figure 143: Sensitivity and specificity of CEUS fast wash-out for diagnosis of malignancy in thyroid cancer

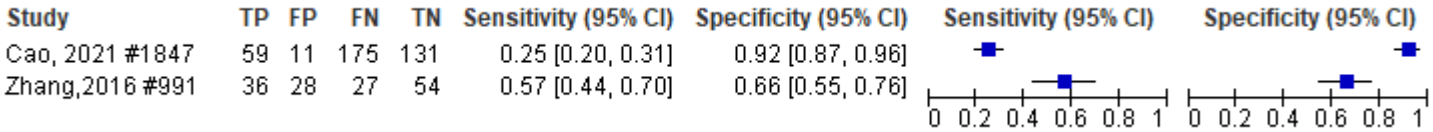


Figure 144: Sensitivity and specificity of CEUS: complete wash out for diagnosis of malignancy in thyroid cancer

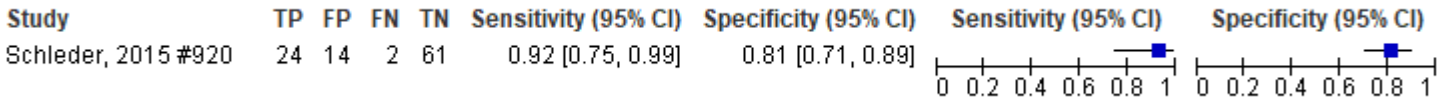


Figure 145: Sensitivity and specificity of CEUS hypo-perfusion for diagnosis of malignancy in thyroid cancer

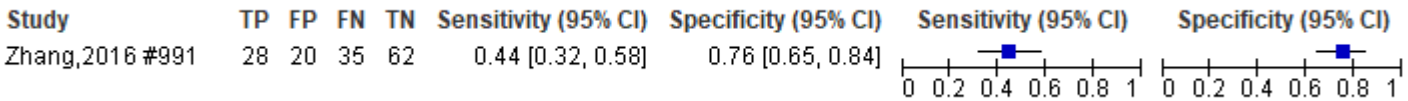


Figure 146: Sensitivity and specificity of CEUS peak ratio <1.06 for diagnosis of malignancy in thyroid cancer

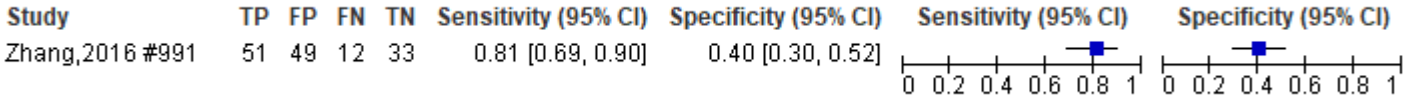


Figure 147: Sensitivity and specificity of CEUS score of 1.6 or higher on purpose built risk score for diagnosis of malignancy in thyroid cancer

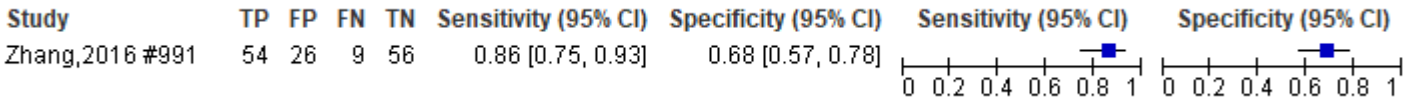


Figure 148: Sensitivity and specificity of CEUS sharpness ratio >1.6 for diagnosis of malignancy in thyroid cancer

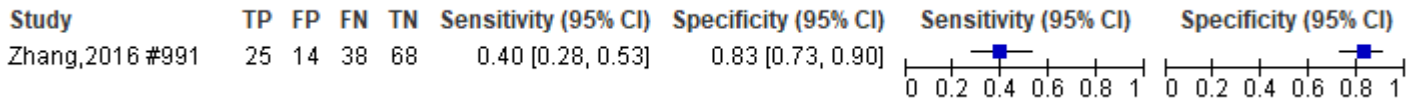


Figure 149: Sensitivity and specificity of CEUS TTP ratio <1.15 for diagnosis of malignancy in thyroid cancer

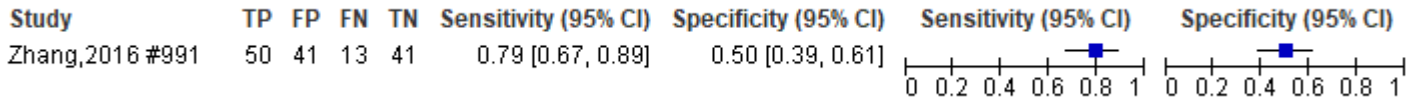


Figure 150: Sensitivity and specificity of CEUS: area >50% for diagnosis of malignancy in thyroid cancer

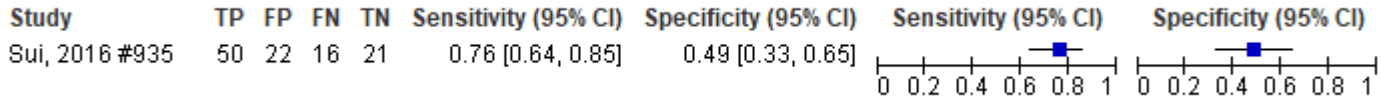
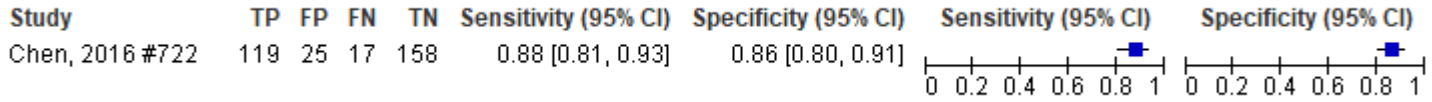


Figure 151: Sensitivity and specificity of CEUS: based on access speed, peak time, subsidence speed, access manner, peak intensity, evenness, pattern of enhancement and clarity of boundary (values for these parameters taken to represent a positive test were unclear) for diagnosis of malignancy in thyroid cancer



ELASTOGRAPHY

Figure 152: Sensitivity and specificity of elastography – Asteria 1-4 colour scale 2 and above - for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abd_Alrahman, 2017 #633	7	19	0	4	1.00 [0.59, 1.00]	0.17 [0.05, 0.39]

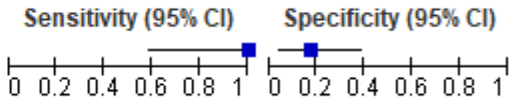


Figure 153: Sensitivity and specificity of elastography – Asteria 1-4 colour scale 3 and above - for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abd_Alrahman, 2017 #633	6	4	1	19	0.86 [0.42, 1.00]	0.83 [0.61, 0.95]
El-Hariri, 2014 #665	21	9	4	50	0.84 [0.64, 0.95]	0.85 [0.73, 0.93]
Gorgolu, 2020 #767	28	16	0	100	1.00 [0.88, 1.00]	0.86 [0.79, 0.92]
Li, 2018 #651	31	9	10	52	0.76 [0.60, 0.88]	0.85 [0.74, 0.93]
Ragazzoni, 2012 #903	34	15	6	77	0.85 [0.70, 0.94]	0.84 [0.75, 0.91]
Schenke, 2018 #919	31	85	7	111	0.82 [0.66, 0.92]	0.57 [0.49, 0.64]
Wang, 2012 #959	25	17	7	82	0.78 [0.60, 0.91]	0.83 [0.74, 0.90]
Xing, 2011 #973	40	10	5	43	0.89 [0.76, 0.96]	0.81 [0.68, 0.91]
Xu, 2014 #975	74	92	42	233	0.64 [0.54, 0.73]	0.72 [0.66, 0.77]
Zhang,2016 #991	46	10	17	72	0.73 [0.60, 0.83]	0.88 [0.79, 0.94]

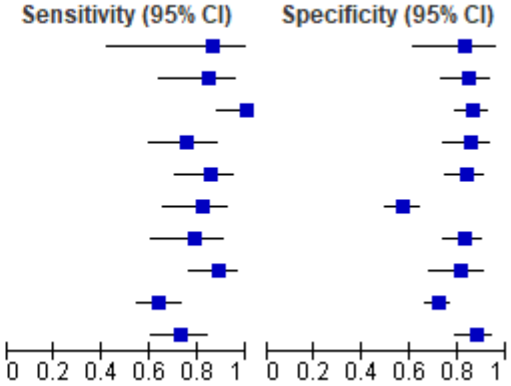


Figure 154: Sensitivity and specificity of elastography – Asteria 1-4 colour scale 4 - for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abd_Alrahman, 2017 #633	3	0	4	23	0.43 [0.10, 0.82]	1.00 [0.85, 1.00]

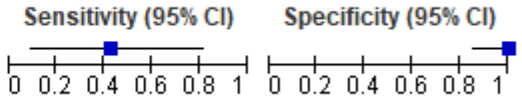


Figure 155: Sensitivity and specificity of elastography – Rago 1-5 colour scale: 2 or more - for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Liu, 2014 #853	19	40	0	5	1.00 [0.82, 1.00]	0.11 [0.04, 0.24]
Rivo-Vazquez, 2013 #909	27	156	1	8	0.96 [0.82, 1.00]	0.05 [0.02, 0.09]
Szczepanek-Parulska, 2013 #940	21	170	1	201	0.95 [0.77, 1.00]	0.54 [0.49, 0.59]

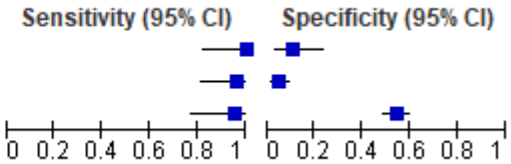


Figure 156: Sensitivity and specificity of elastography – Rago 1-5 colour scale: 3 or more - for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Garg, 2018 #762	29	0	4	84	0.88 [0.72, 0.97]	1.00 [0.96, 1.00]
Liu, 2014 #853	19	23	0	22	1.00 [0.82, 1.00]	0.49 [0.34, 0.64]
Mohey, 2013 #674	15	6	0	25	1.00 [0.78, 1.00]	0.81 [0.63, 0.93]
Pagano, 2020 #888	38	21	30	57	0.56 [0.43, 0.68]	0.73 [0.62, 0.82]
Refaat, 2014 #678	14	6	0	15	1.00 [0.77, 1.00]	0.71 [0.48, 0.89]
Rivo-Vazquez, 2013 #909	21	89	7	75	0.75 [0.55, 0.89]	0.46 [0.38, 0.54]

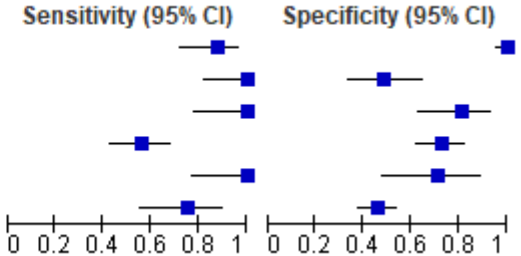


Figure 157: Sensitivity and specificity of elastography – Rago 1-5 colour scale: 4 or more - for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Garg, 2018 #762	23	0	10	84	0.70 [0.51, 0.84]	1.00 [0.96, 1.00]
Liu, 2014 #853	15	7	4	38	0.79 [0.54, 0.94]	0.84 [0.71, 0.94]
Ma, 2017 #864	57	16	22	40	0.72 [0.61, 0.82]	0.71 [0.58, 0.83]
Mohey, 2013 #674	14	0	1	31	0.93 [0.68, 1.00]	1.00 [0.89, 1.00]
Pagano, 2020 #888	24	8	44	70	0.35 [0.24, 0.48]	0.90 [0.81, 0.95]
Rago, 2007 #904	30	0	1	61	0.97 [0.83, 1.00]	1.00 [0.94, 1.00]
Refaat, 2014 #678	11	0	3	21	0.79 [0.49, 0.95]	1.00 [0.84, 1.00]
Rivo-Vazquez, 2013 #909	11	21	17	143	0.39 [0.22, 0.59]	0.87 [0.81, 0.92]
Yang, 2019 #689	9	5	2	35	0.82 [0.48, 0.98]	0.88 [0.73, 0.96]

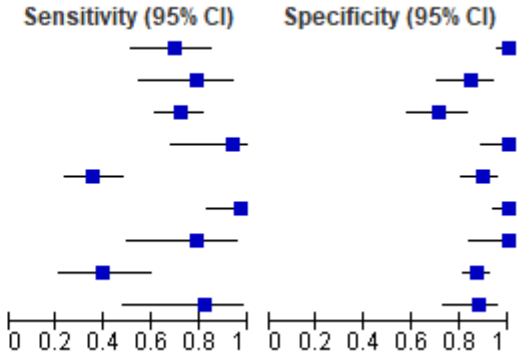


Figure 158: Sensitivity and specificity of elastography – Rago 1-5 colour scale: 50 - for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Liu, 2014 #853	5	1	14	44	0.26 [0.09, 0.51]	0.98 [0.88, 1.00]
Mohey, 2013 #674	12	0	3	31	0.80 [0.52, 0.96]	1.00 [0.89, 1.00]
Rivo-Vazquez, 2013 #909	2	2	26	162	0.07 [0.01, 0.24]	0.99 [0.96, 1.00]

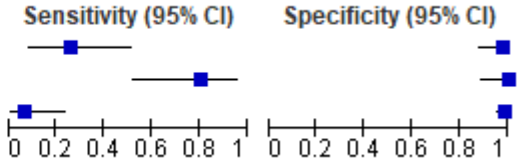


Figure 159: Sensitivity and specificity of elastography – ITOH 1-5 colour scale: 2 or more - for diagnosis of malignancy in thyroid cancer

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Cakir, 2011 #716	122	239	3	27	0.98 [0.93, 1.00]	0.10 [0.07, 0.14]
Schweel, 2013 #929	16	20	0	30	1.00 [0.79, 1.00]	0.60 [0.45, 0.74]
Sui, 2016 #935	61	18	5	25	0.92 [0.83, 0.97]	0.58 [0.42, 0.73]

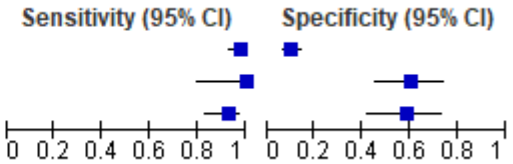


Figure 160: Sensitivity and specificity of elastography – ITOH 1-5 colour scale: 3 or more - for diagnosis of malignancy in thyroid cancer

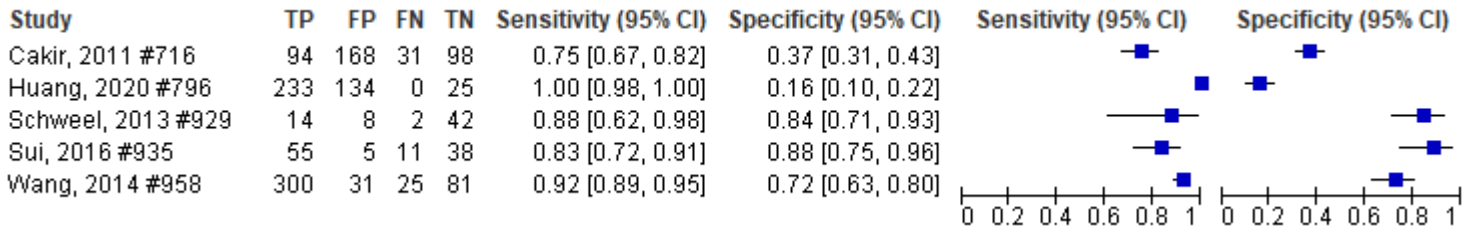


Figure 161: Sensitivity and specificity of elastography – ITOH 1-5 colour scale: 4 or more - for diagnosis of malignancy in thyroid cancer

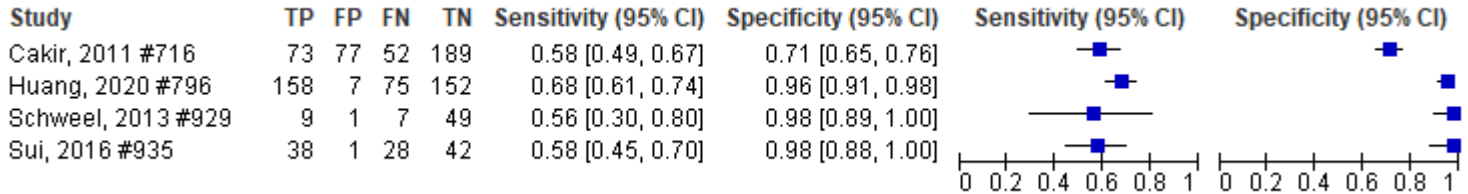


Figure 162: Sensitivity and specificity of elastography – ITOH 1-5 colour scale: 5 - for diagnosis of malignancy in thyroid cancer

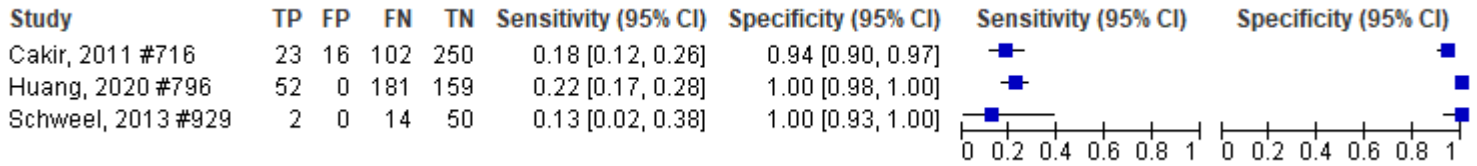


Figure 163: Sensitivity and specificity of elastography – RGB colour scale 0-4: 2 and more - for diagnosis of malignancy in thyroid cancer

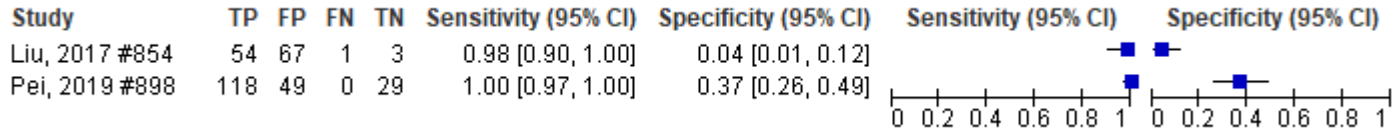


Figure 164: Sensitivity and specificity of elastography – RGB colour scale 0-4: 3 and more - for diagnosis of malignancy in thyroid cancer

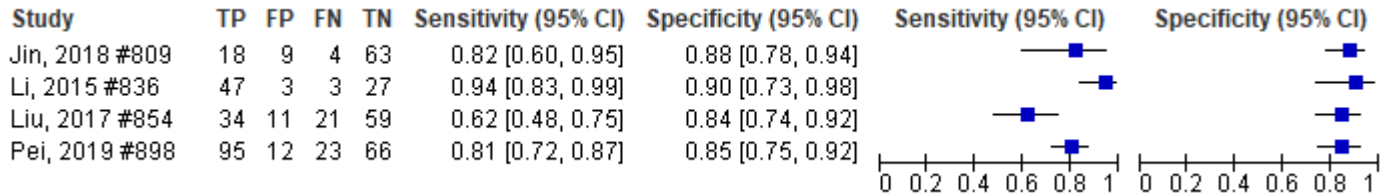


Figure 165: Sensitivity and specificity of elastography – RGB colour scale 0-4: 4 - for diagnosis of malignancy in thyroid cancer

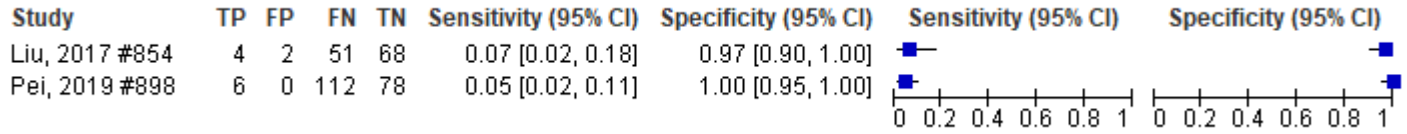


Figure 166: Sensitivity and specificity of elastography – 1-3 Rago scale: 2 or higher - for diagnosis of malignancy in thyroid cancer

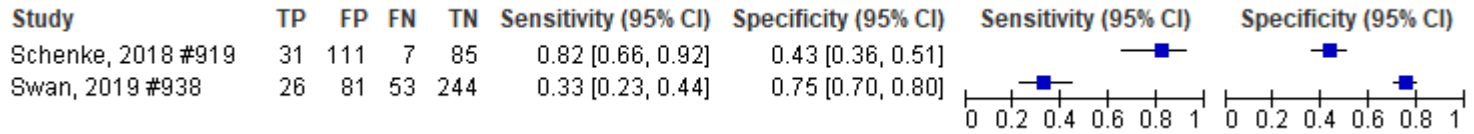


Figure 167: Sensitivity and specificity of elastography – 1-3 Rago scale: 3 - for diagnosis of malignancy in thyroid cancer

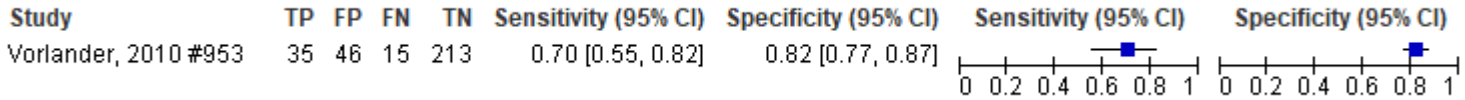


Figure 168: Sensitivity and specificity of elastography – 1-6 ES scale: 4 or more - for diagnosis of malignancy in thyroid cancer

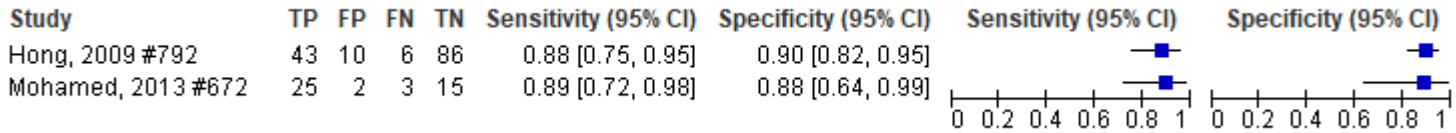


Figure 169: Sensitivity and specificity of elastography – '4 pattern': 3 or more - for diagnosis of malignancy in thyroid cancer

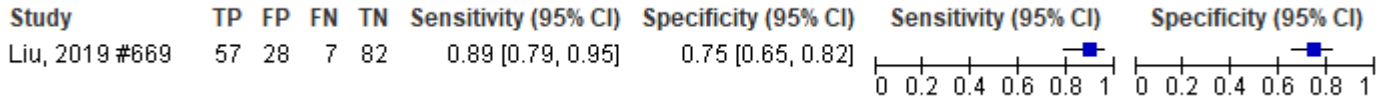


Figure 170: Sensitivity and specificity of elastography – 0-IV colour grade system (Shuzen): III and above - for diagnosis of malignancy in thyroid cancer

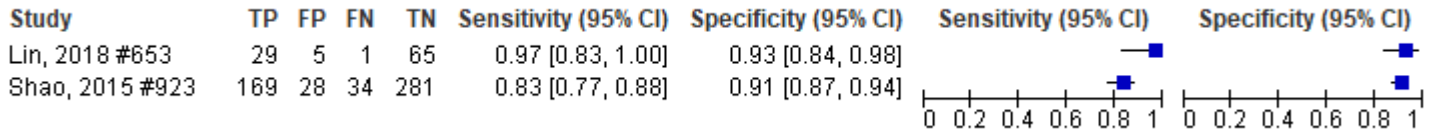


Figure 171: Sensitivity and specificity of elastography – 0-5 colour method (different to other studies): 3 or more - for diagnosis of malignancy in thyroid cancer

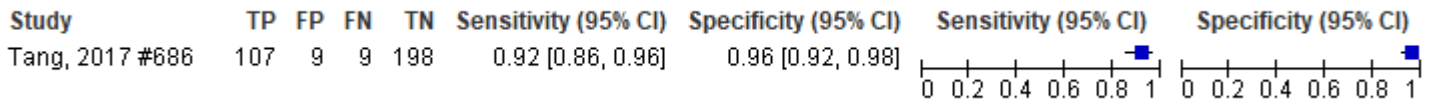


Figure 172: Sensitivity and specificity of elastography – VTI I – VI scale: II or more - for diagnosis of malignancy in thyroid cancer

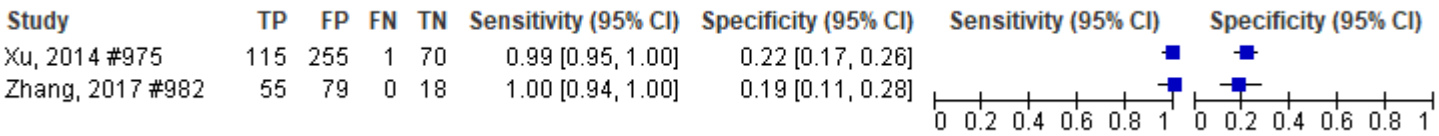


Figure 173: Sensitivity and specificity of elastography – VTI I – VI scale: III or more - for diagnosis of malignancy in thyroid cancer

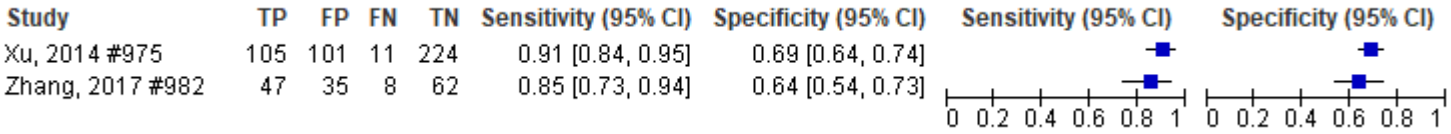


Figure 174: Sensitivity and specificity of elastography – VTI I – VI scale: IV or more - for diagnosis of malignancy in thyroid cancer

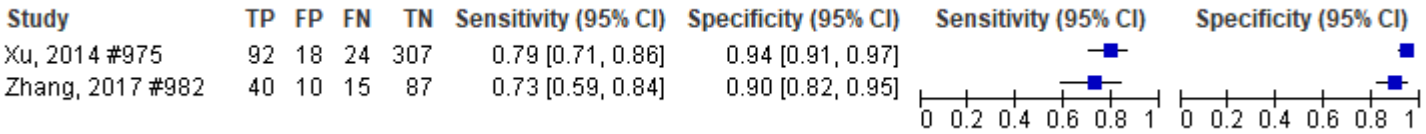


Figure 175: Sensitivity and specificity of elastography – VTI I – VI scale: V or more - for diagnosis of malignancy in thyroid cancer

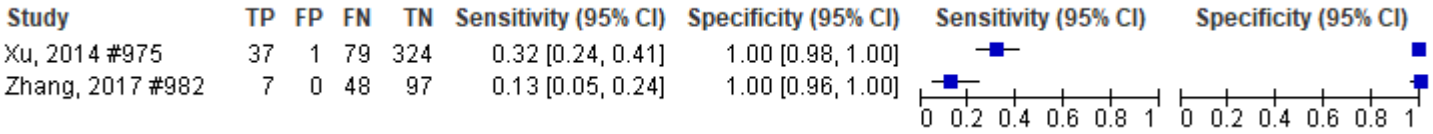


Figure 176: Sensitivity and specificity of elastography – 1-6 scoring method (Huang, 2015): 2 or more - for diagnosis of malignancy in thyroid cancer

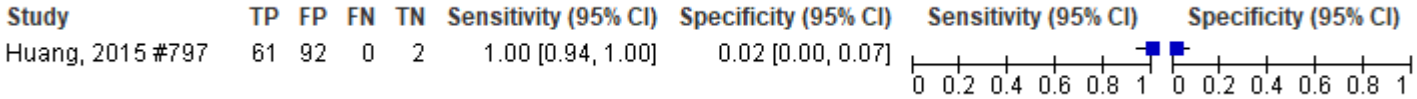


Figure 177: Sensitivity and specificity of elastography –1-6 scoring method (Huang, 2015): 3 or more - for diagnosis of malignancy in thyroid cancer

Figure 178: Sensitivity and specificity of elastography –1-6 scoring method (Huang, 2015): 4 or more - for diagnosis of malignancy in thyroid cancer

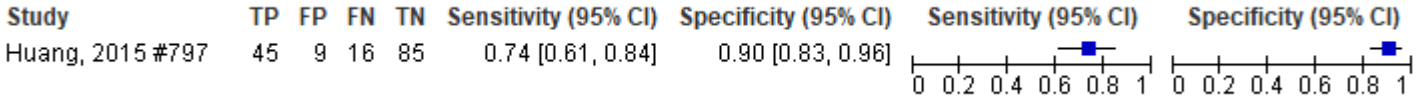


Figure 179: Sensitivity and specificity of elastography– 1-6 scoring method (Huang, 2015): 5 or more - for diagnosis of malignancy in thyroid cancer

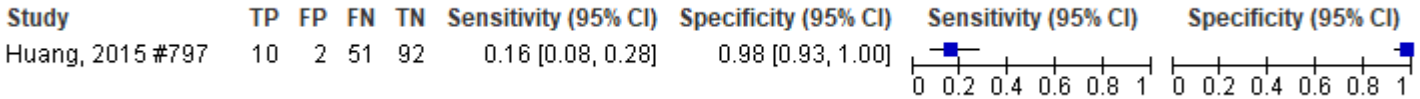


Figure 180: Sensitivity and specificity of elastography– 1-6 scoring method (Huang, 2015): 6 - for diagnosis of malignancy in thyroid cancer

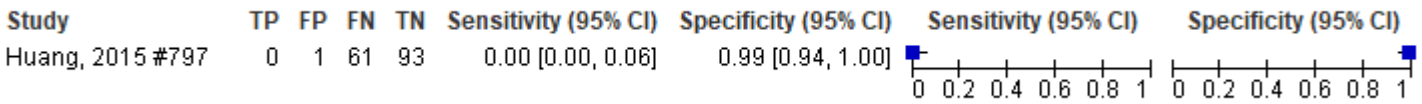


Figure 181: Sensitivity and specificity of elastography – VTIQ velocity of 2.4 m/s and above - for diagnosis of malignancy in thyroid cancer

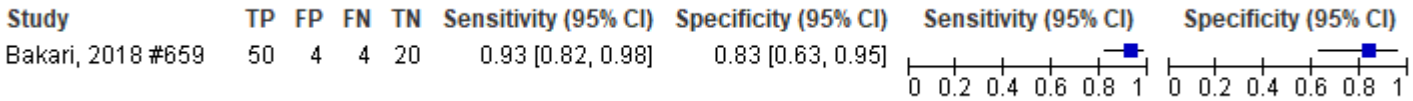


Figure 182: Sensitivity and specificity of elastography – VTIQ velocity of 2.565 m/s and above - for diagnosis of malignancy in thyroid cancer2.4

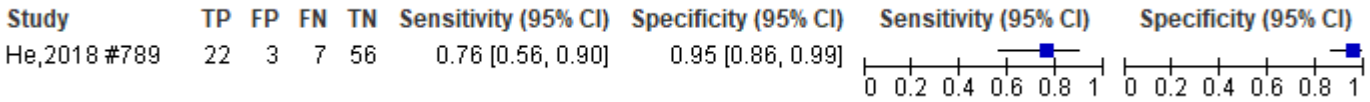


Figure 183: Sensitivity and specificity of elastography – VTIQ velocity of 2.64 m/s and above - for diagnosis of malignancy in thyroid cancer

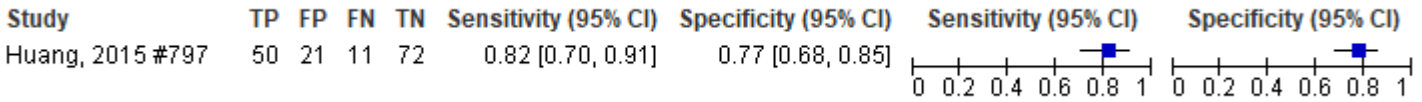


Figure 184: Sensitivity and specificity of elastography – VTIQ velocity of 2.84 m/s and above - for diagnosis of malignancy in thyroid cancer

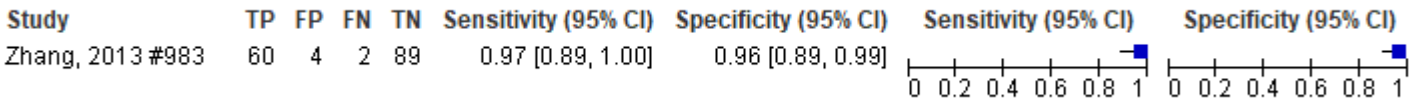


Figure 185: Sensitivity and specificity of elastography – VTIQ velocity of 2.87 m/s and above - for diagnosis of malignancy in thyroid cancer

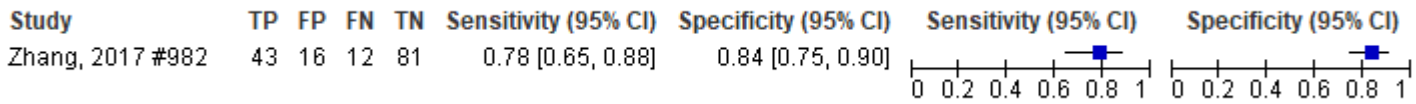


Figure 186: Sensitivity and specificity of elastography – VTIQ velocity of 2.91 m/s and above - for diagnosis of malignancy in thyroid cancer

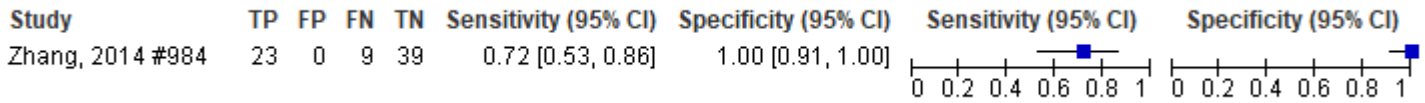


Figure 187: Sensitivity and specificity of elastography – VTIQ velocity of 5 m/s and above - for diagnosis of malignancy in thyroid cancer

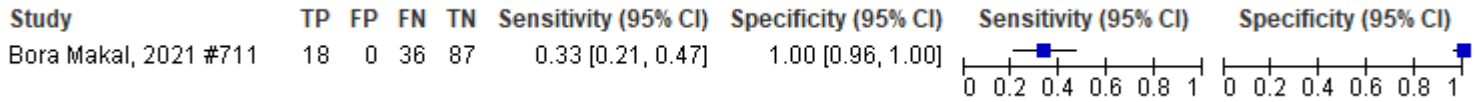


Figure 188: Sensitivity and specificity of elastography – EI of 27.65 kpa and above - for diagnosis of malignancy in thyroid cancer

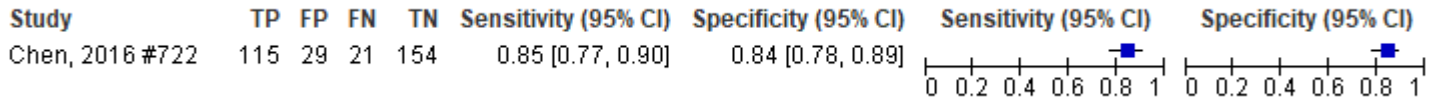


Figure 189: Sensitivity and specificity of elastography – EI of 31.0 kpa and above - for diagnosis of malignancy in thyroid cancer

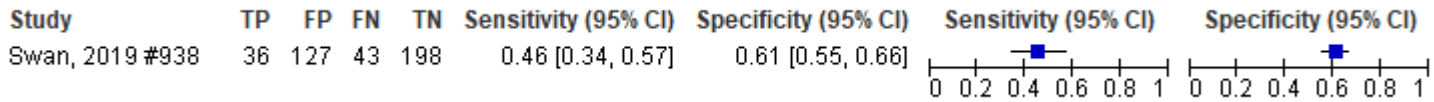


Figure 190: Sensitivity and specificity of elastography – EI of 36.2 kpa and above - for diagnosis of malignancy in thyroid cancer

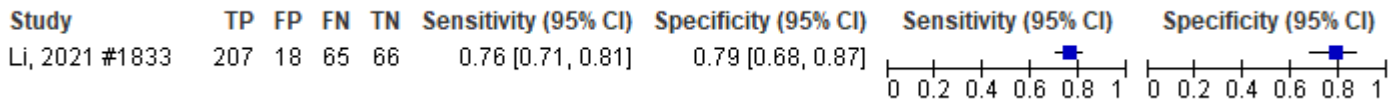


Figure 191: Sensitivity and specificity of elastography – EI of 38.3 kpa and above - for diagnosis of malignancy in thyroid cancer

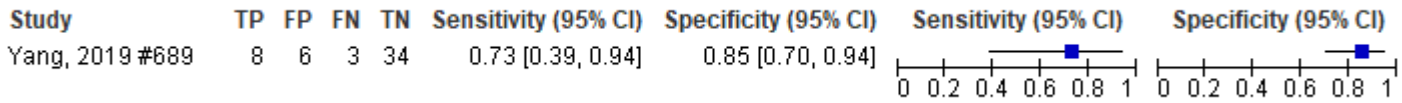


Figure 192: Sensitivity and specificity of elastography – EI of 39.3 kpa and above - for diagnosis of malignancy in thyroid cancer

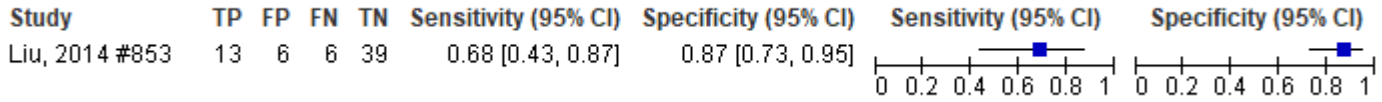


Figure 193: Sensitivity and specificity of elastography – EI of 45 kpa and above - for diagnosis of malignancy in thyroid cancer

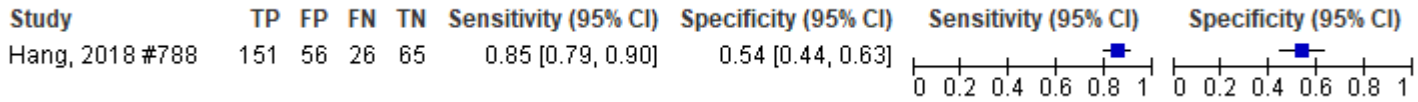


Figure 194: Sensitivity and specificity of elastography – EI of 52.1 kpa and above - for diagnosis of malignancy in thyroid cancer

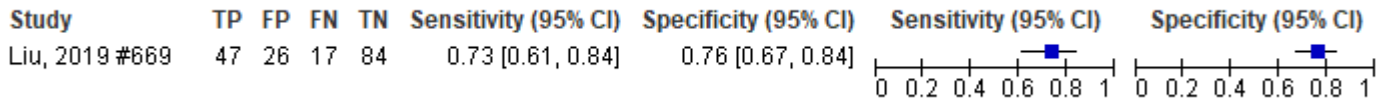


Figure 195: Sensitivity and specificity of elastography – EI of 66 kpa and above - for diagnosis of malignancy in thyroid cancer

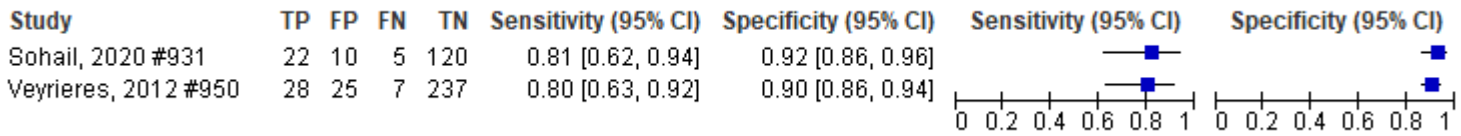


Figure 196: Sensitivity and specificity of elastography – EI of 69 kpa and above - for diagnosis of malignancy in thyroid cancer

Figure 197: Sensitivity and specificity of elastography – EI of 74.5 kpa and above - for diagnosis of malignancy in thyroid cancer

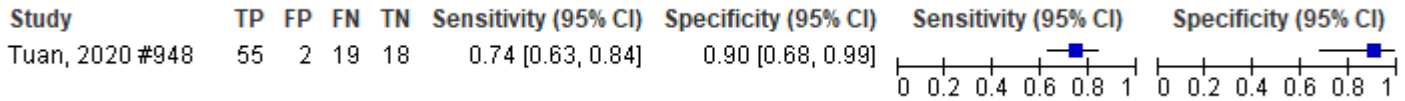


Figure 198: Sensitivity and specificity of elastography – EI of 120 kpa and above - for diagnosis of malignancy in thyroid cancer

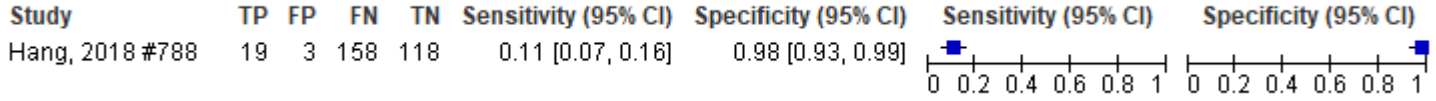


Figure 199: Sensitivity and specificity of elastography – SR of 1.32 and above - for diagnosis of malignancy in thyroid cancer

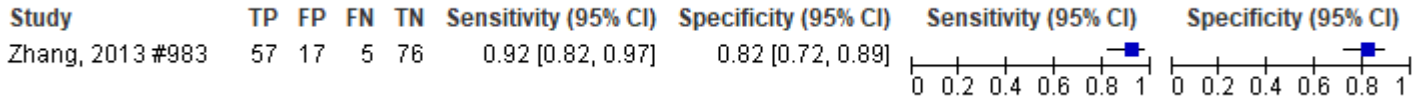


Figure 200: Sensitivity and specificity of elastography – SR of 2.2 and above - for diagnosis of malignancy in thyroid cancer

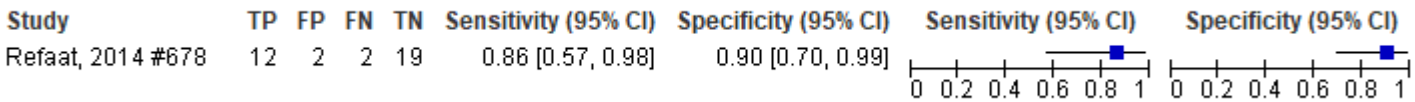


Figure 201: Sensitivity and specificity of elastography – SR of 2.37 and above - for diagnosis of malignancy in thyroid cancer

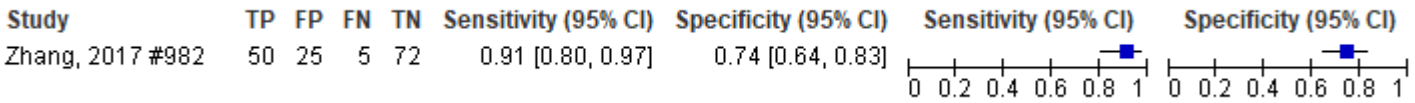


Figure 202: Sensitivity and specificity of elastography – SR of 2.88 and above - for diagnosis of malignancy in thyroid cancer

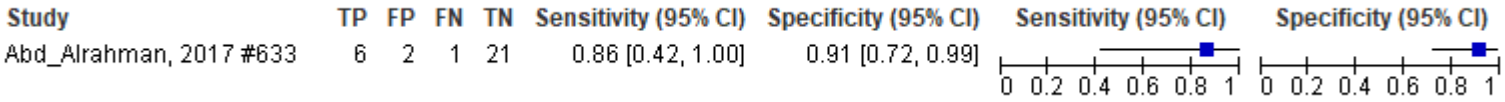


Figure 203: Sensitivity and specificity of elastography – SR of 2.9 and above - for diagnosis of malignancy in thyroid cancer

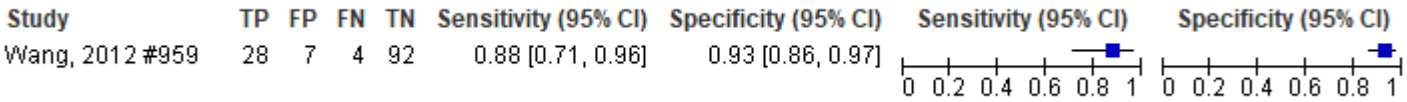


Figure 204: Sensitivity and specificity of elastography – SR of 3.5 and above - for diagnosis of malignancy in thyroid cancer

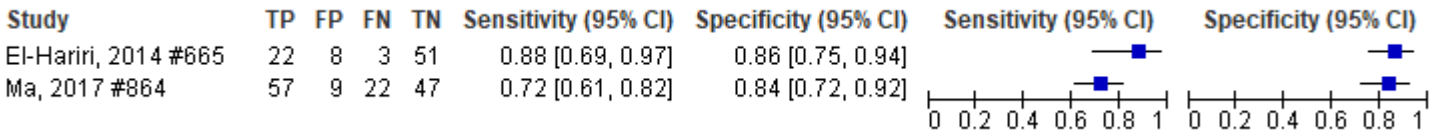


Figure 205: Sensitivity and specificity of elastography – SR of 3.59 and above - for diagnosis of malignancy in thyroid cancer

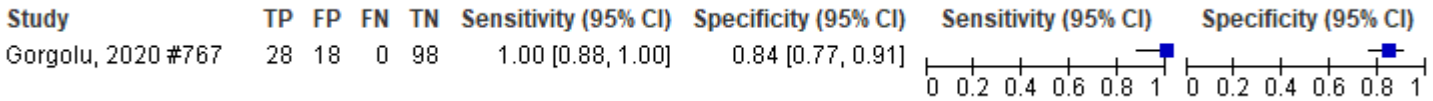


Figure 206: Sensitivity and specificity of elastography – SR of 3.65 and above - for diagnosis of malignancy in thyroid cancer

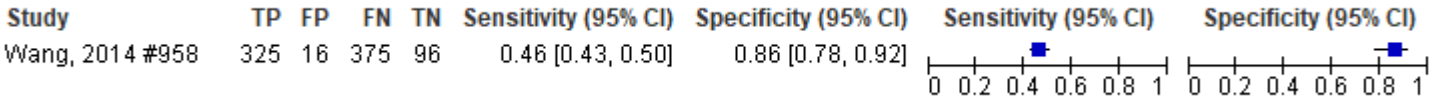


Figure 207: Sensitivity and specificity of elastography – SR of 3.79 and above - for diagnosis of malignancy in thyroid cancer

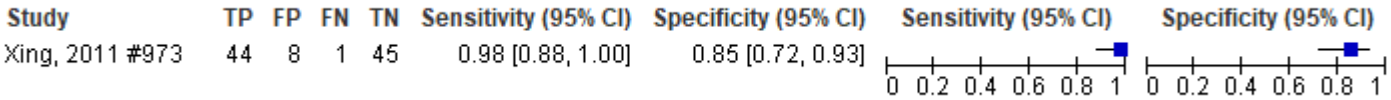


Figure 208: Sensitivity and specificity of elastography – SR of 4 and above - for diagnosis of malignancy in thyroid cancer

COMBINATIONS OF CHARACTERISTICS FROM DIFFERENT METHODS

Figure 209: Sensitivity and specificity of absent halo AND microcalcification AND type III vascularisation for diagnosis of malignancy in thyroid cancer

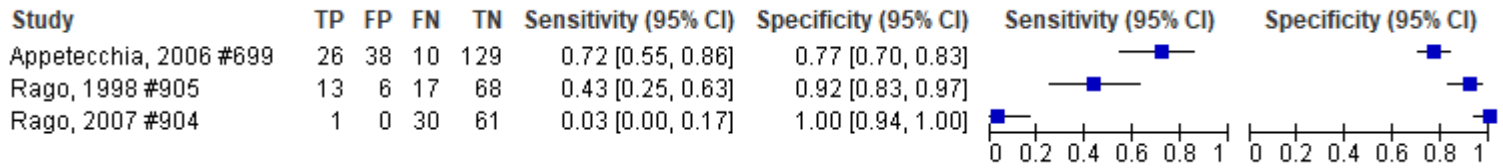


Figure 210: Sensitivity and specificity of hypoechoicity AND microcalcification AND type III vascularisation for diagnosis of malignancy in thyroid cancer

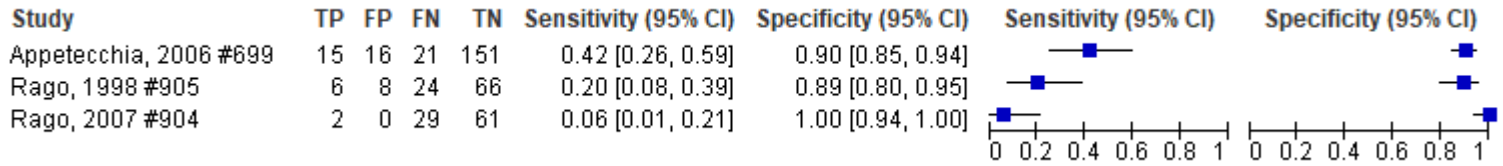


Figure 211: Sensitivity and specificity of absent halo AND hypoechoicity AND type III vascularisation for diagnosis of malignancy in thyroid cancer

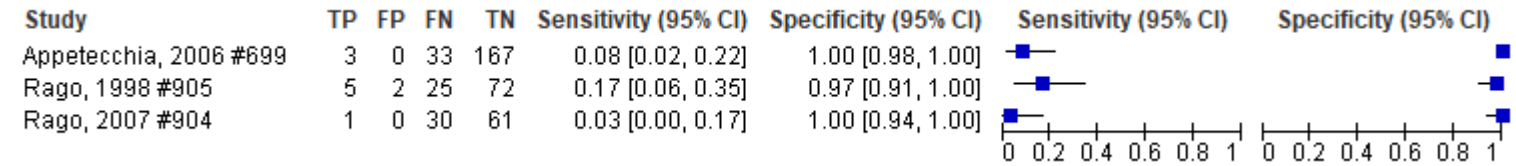


Figure 212: Sensitivity and specificity of microcalcifications AND type III vascularisation for diagnosis of malignancy in thyroid cancer

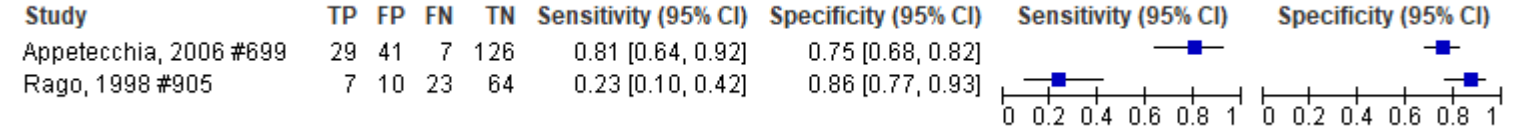


Figure 213: Sensitivity and specificity of hypoechoicity AND type III vascularisation for diagnosis of malignancy in thyroid cancer

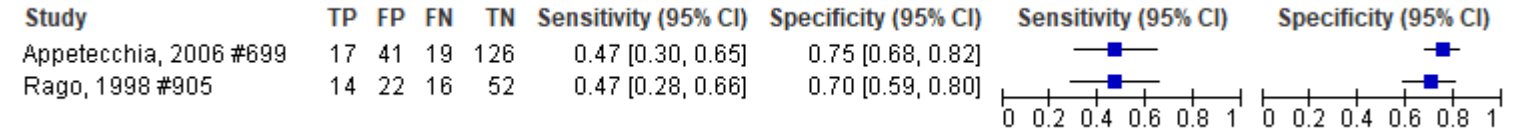


Figure 214: Sensitivity and specificity of absent halo AND type III vascularisation for diagnosis of malignancy in thyroid cancer

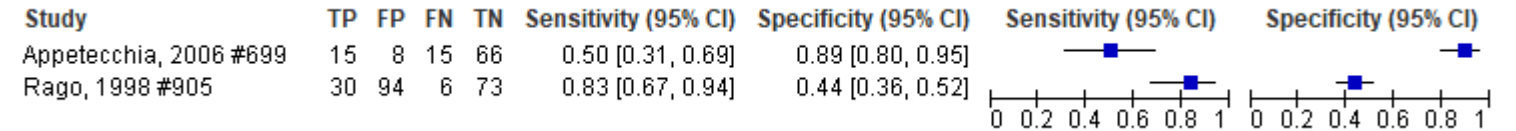


Figure 215: Sensitivity and specificity of combined doppler and grey scale characteristics: calcification OR resistive index >0.715 OR pulsatility index >0.945 for diagnosis of malignancy in thyroid cancer

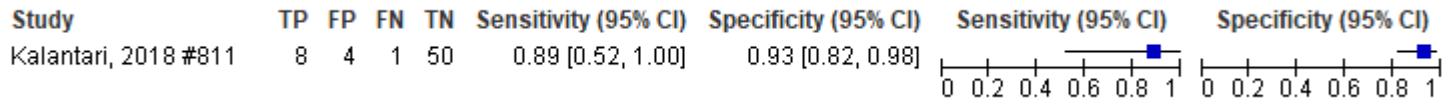


Figure 216: Sensitivity and specificity of French TI-RADS 3 or more AND capsule interruption and increased intranodular vascularization for diagnosis of malignancy in thyroid cancer

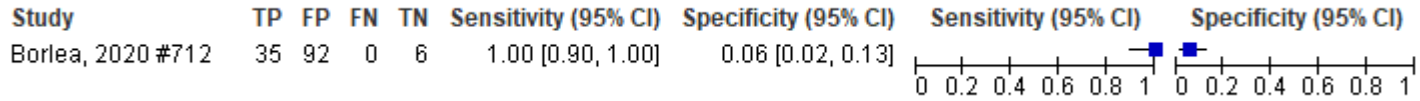


Figure 217: Sensitivity and specificity of French TI-RADS 4a or more AND capsule interruption and increased intranodular vascularization for diagnosis of malignancy in thyroid cancer

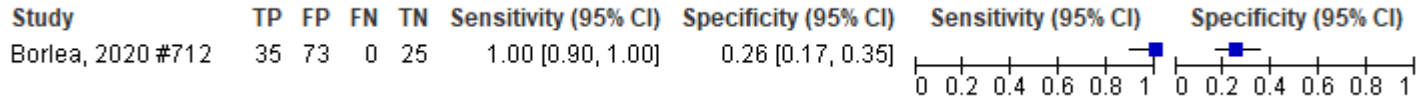


Figure 218: Sensitivity and specificity of French TI-RADS 4b or more AND capsule interruption and increased intranodular vascularization for diagnosis of malignancy in thyroid cancer

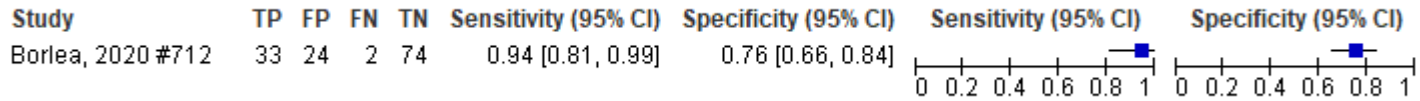


Figure 219: Sensitivity and specificity of French TI-RADS 5 AND capsule interruption and increased intranodular vascularization for diagnosis of malignancy in thyroid cancer

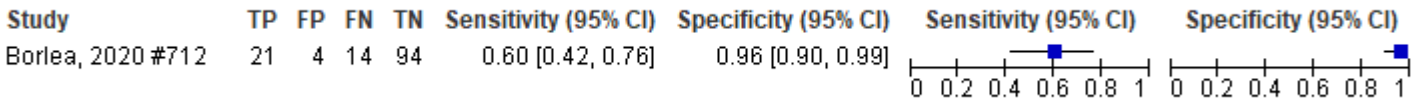


Figure 220: Sensitivity and specificity of TI-RADS (0-10) and elastography Emax score of 5 or more for diagnosis of malignancy in thyroid cancer

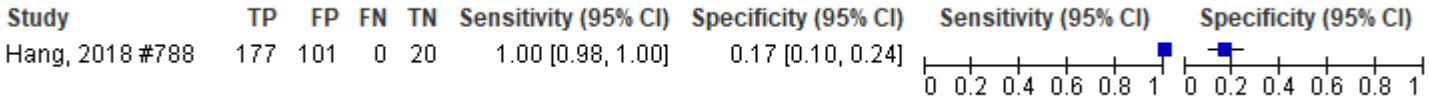


Figure 221: Sensitivity and specificity of TI-RADS (0-10) and elastography Emax score of 6 or more for diagnosis of malignancy in thyroid cancer

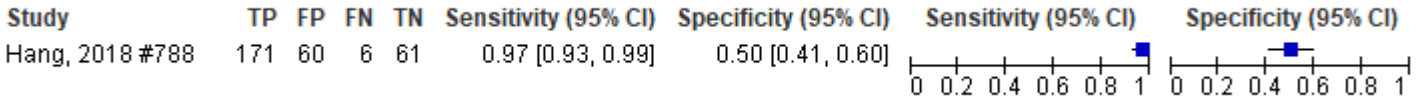


Figure 222: Sensitivity and specificity of TI-RADS (0-10) and elastography Emax score of 7 or more for diagnosis of malignancy in thyroid cancer

Figure 223: Sensitivity and specificity of TI-RADS (0-10) and elastography Emax score of 8 or more for diagnosis of malignancy in thyroid cancer

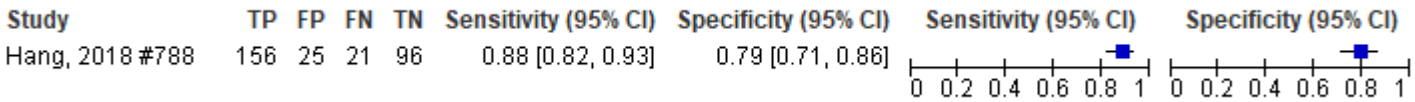


Figure 224: Sensitivity and specificity of Kwak TIRADS and ITOH combined score of 5 and above for diagnosis of malignancy in thyroid cancer

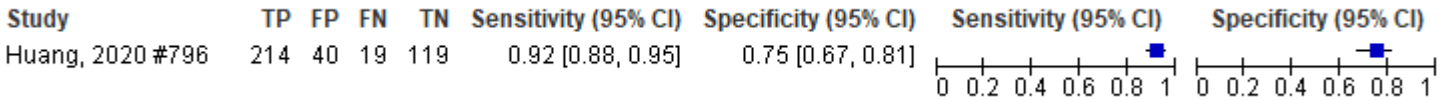


Figure 225: Sensitivity and specificity of SWE at 66kpa and above OR microcalcification OR central vascularisation for diagnosis of malignancy in thyroid cancer

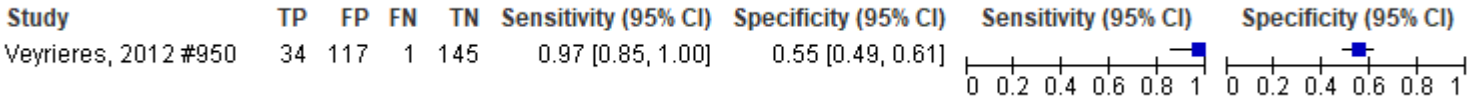


Figure 226: Sensitivity and specificity of elastography ARFI SWV of 2.64 m/s or more OR elastography 1-6 scoring method score of 4 or more for diagnosis of malignancy in thyroid cancer

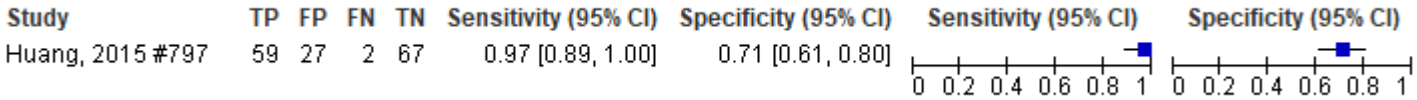


Figure 227: Sensitivity and specificity of virtual touch quantification at 2.91 m/s and above OR markedly hypoechoic for diagnosis of malignancy in thyroid cancer

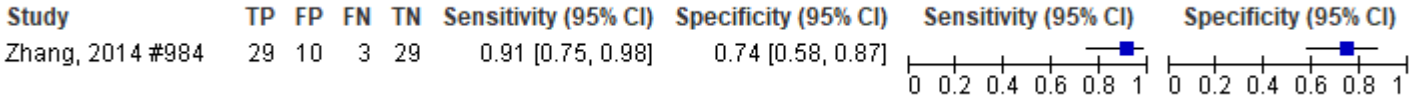


Figure 228: Sensitivity and specificity of virtual touch quantification at 2.91 m/s and above OR poorly defined margins for diagnosis of malignancy in thyroid cancer

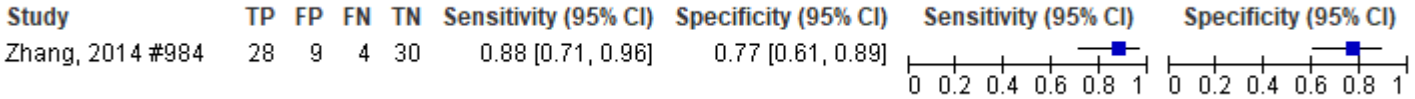


Figure 229: Sensitivity and specificity of virtual touch quantification at 2.91 m/s and above OR taller than wide for diagnosis of malignancy in thyroid cancer

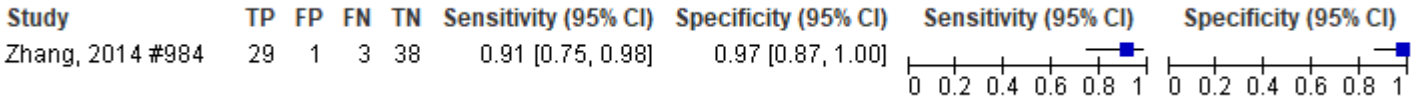


Figure 230: Sensitivity and specificity of spot microcalcification AND presence of hypoechoicity AND type II vascularisation for diagnosis of malignancy in thyroid cancer

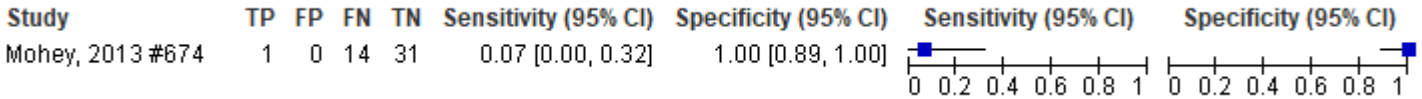


Figure 231: Sensitivity and specificity of absence of halo AND presence of hypoechoicity AND type II vascularisation for diagnosis of malignancy in thyroid cancer

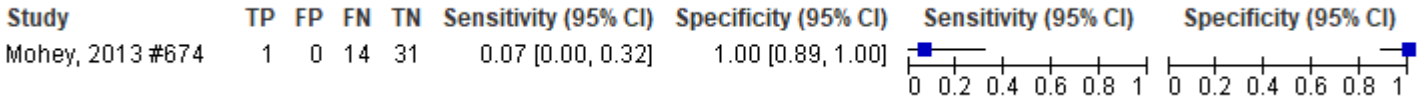
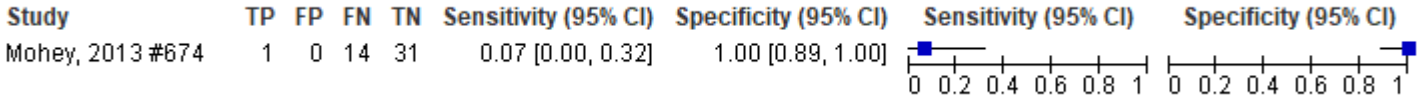
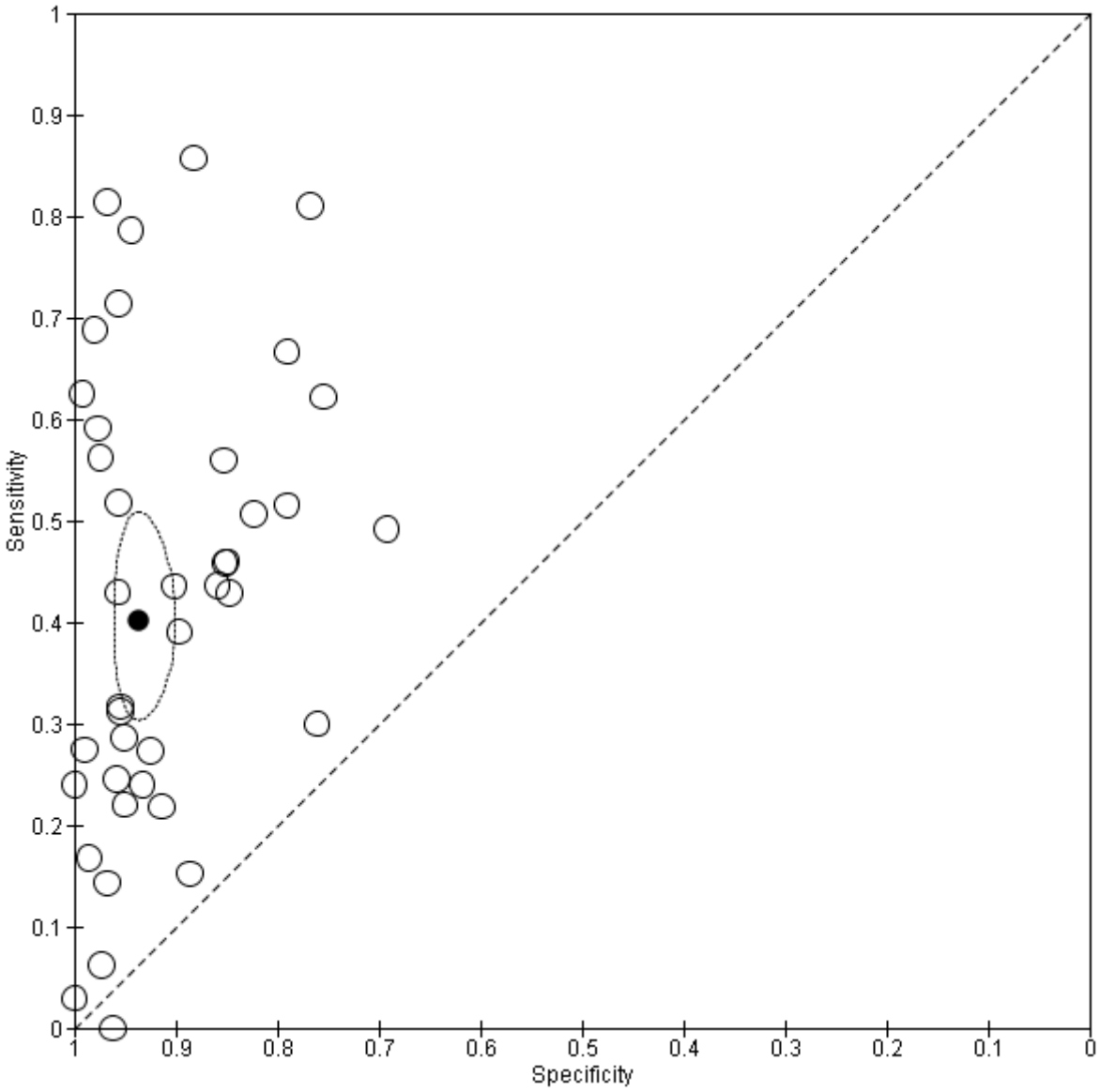


Figure 232: Sensitivity and specificity of spot microcalcification AND absence of halo AND type II vascularisation for diagnosis of malignancy in thyroid cancer

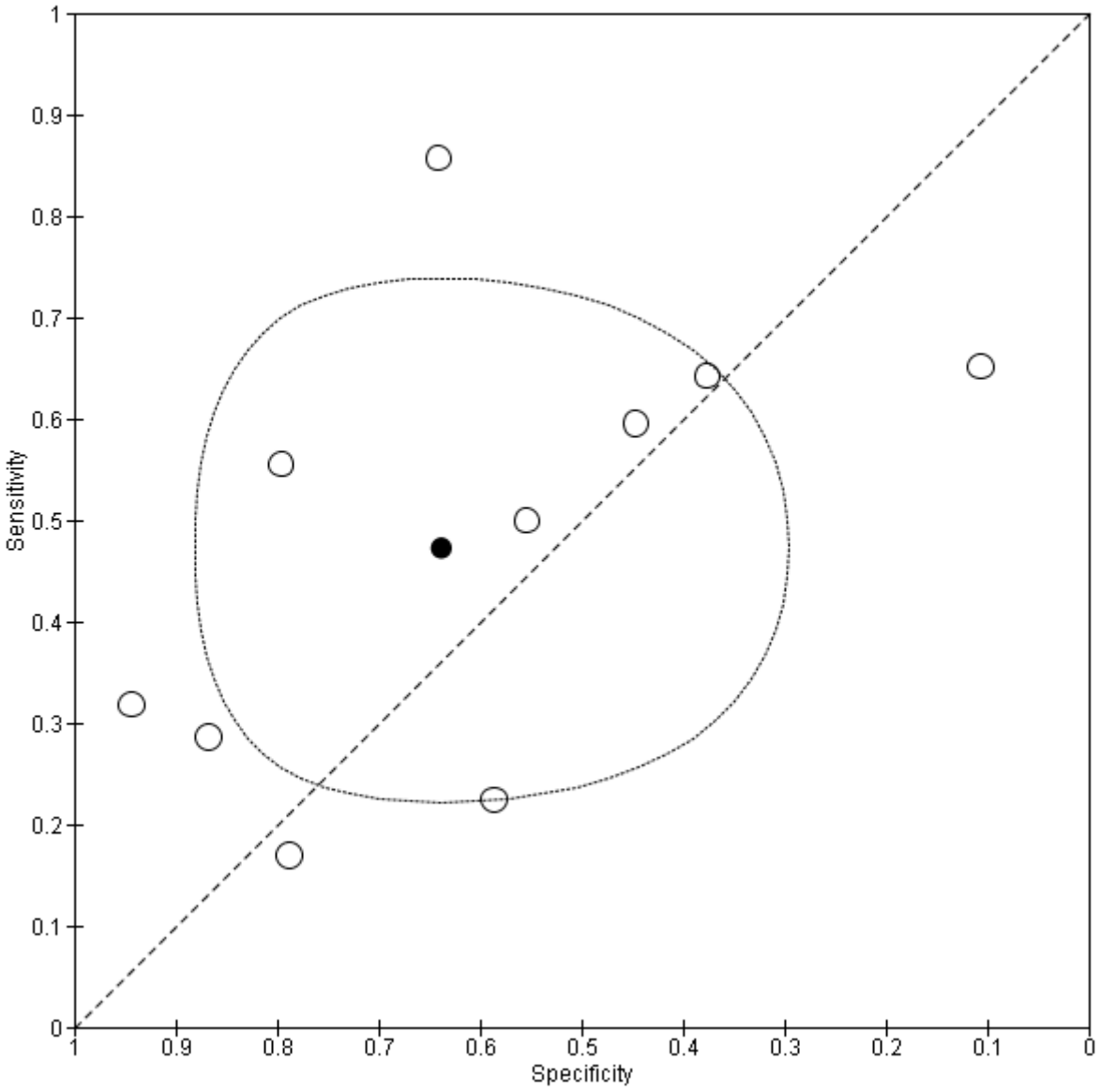


F.1 Sensitivity / 1-specificity plots

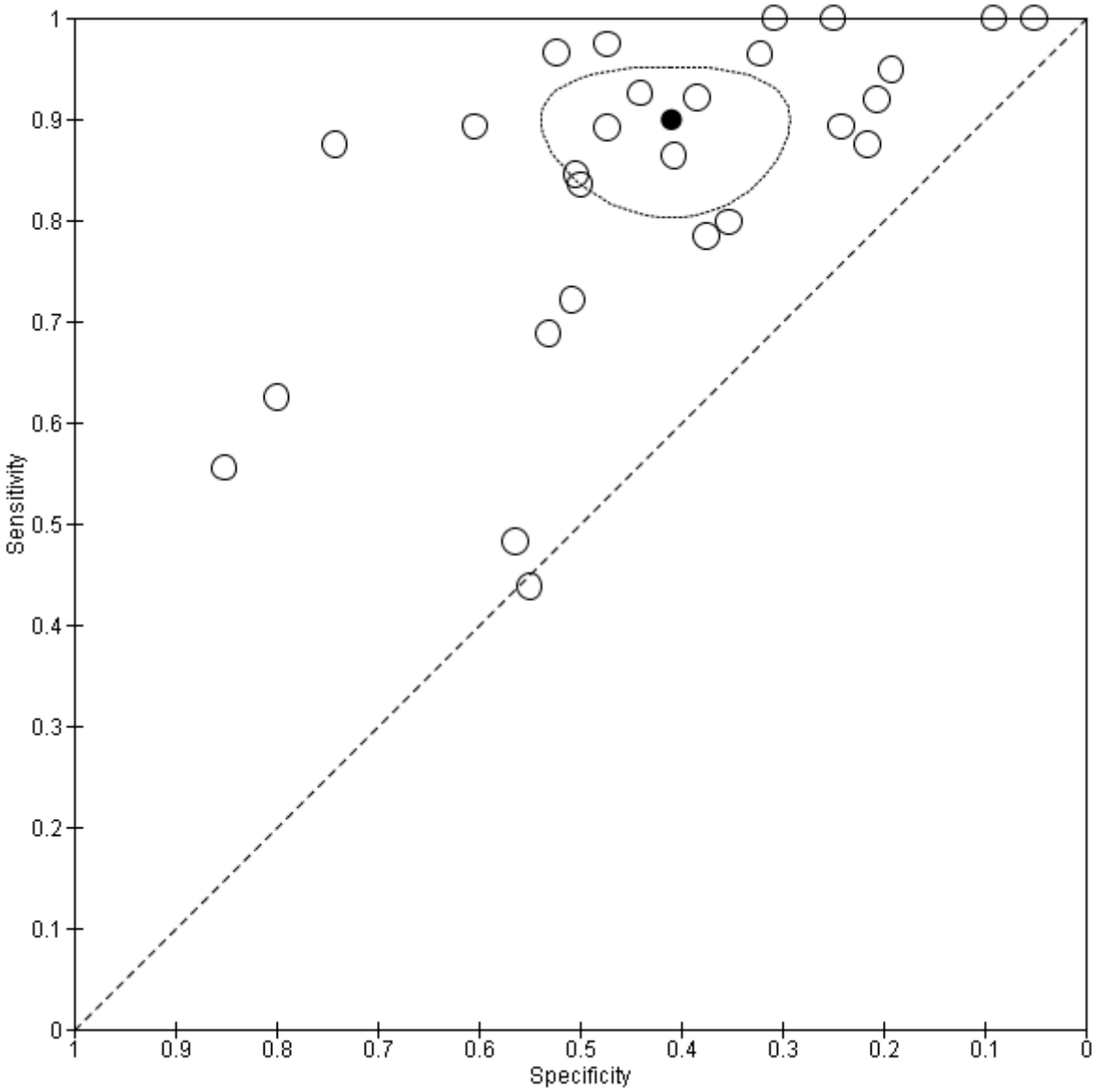
F.1.1 Taller than wide



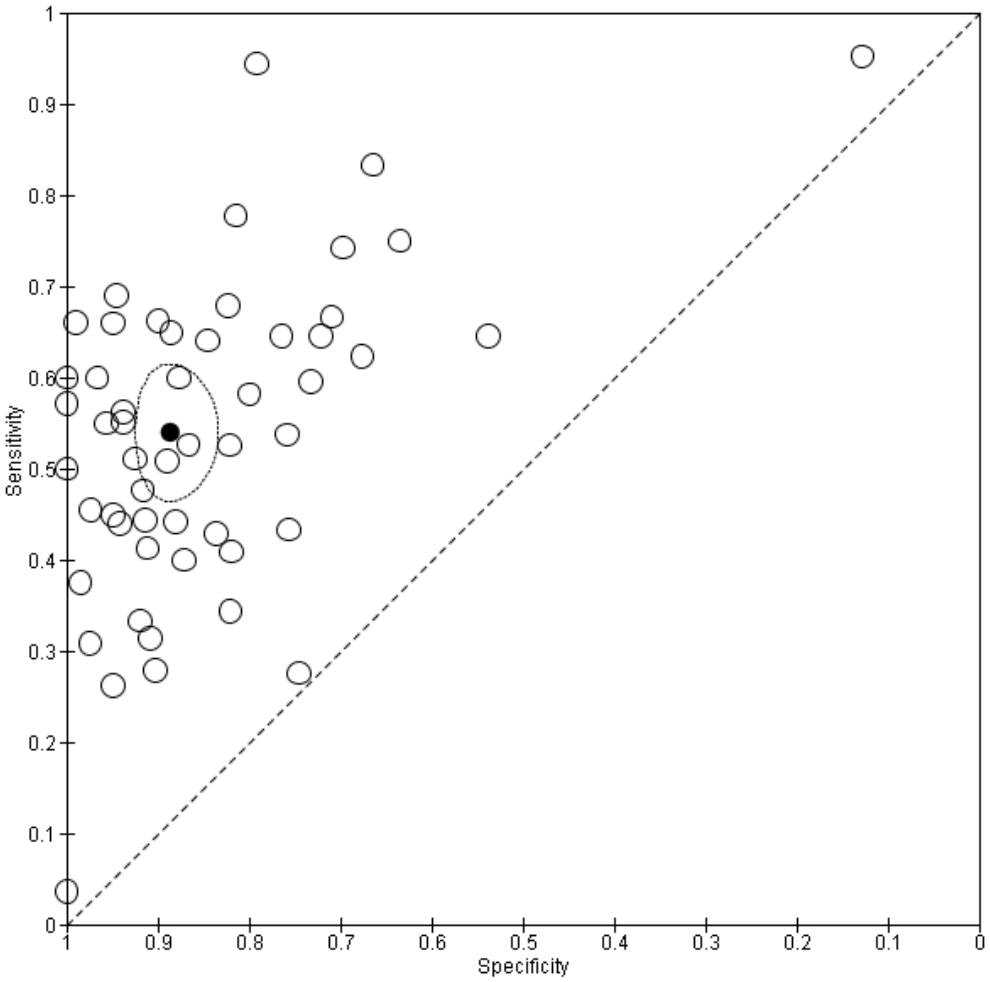
F1.2 Solitary nodule



F1.3 Solidity

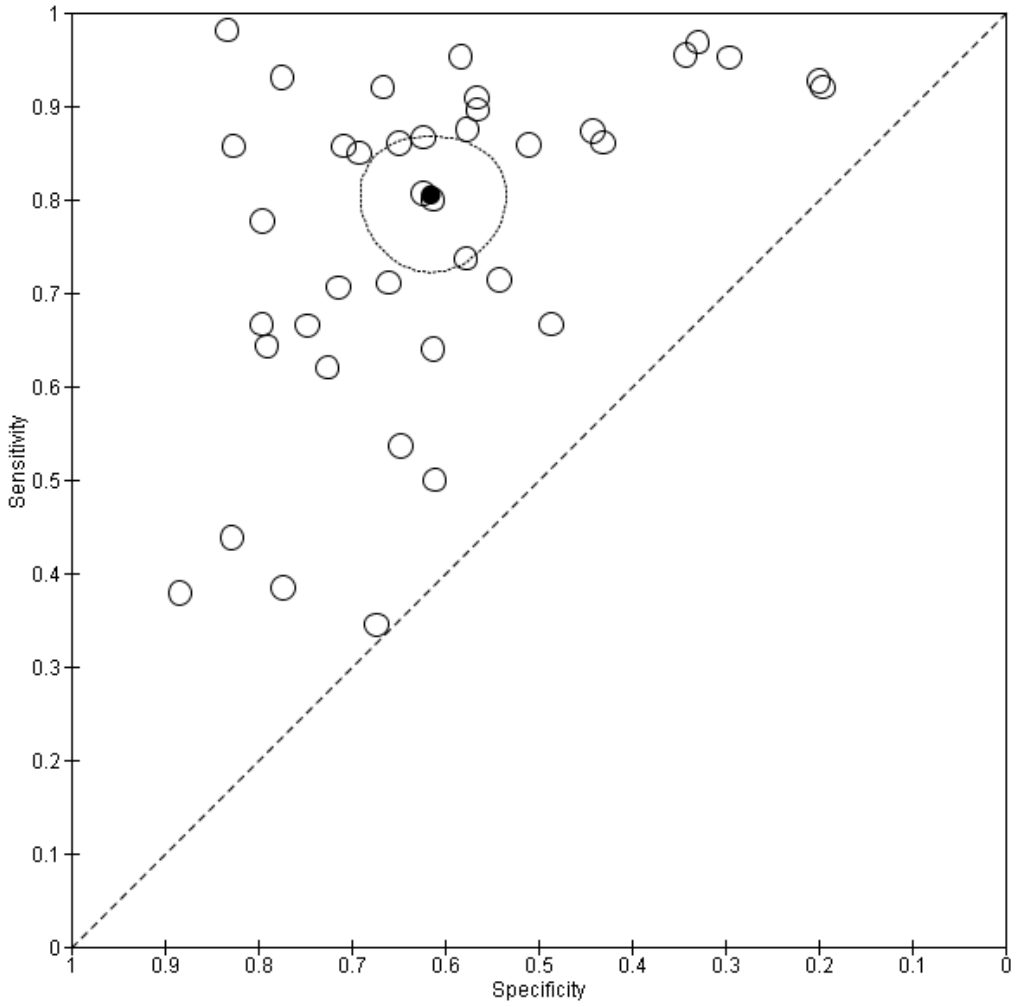


F1.4 Microcalcifications



1

F21.5 Hypoechoicity

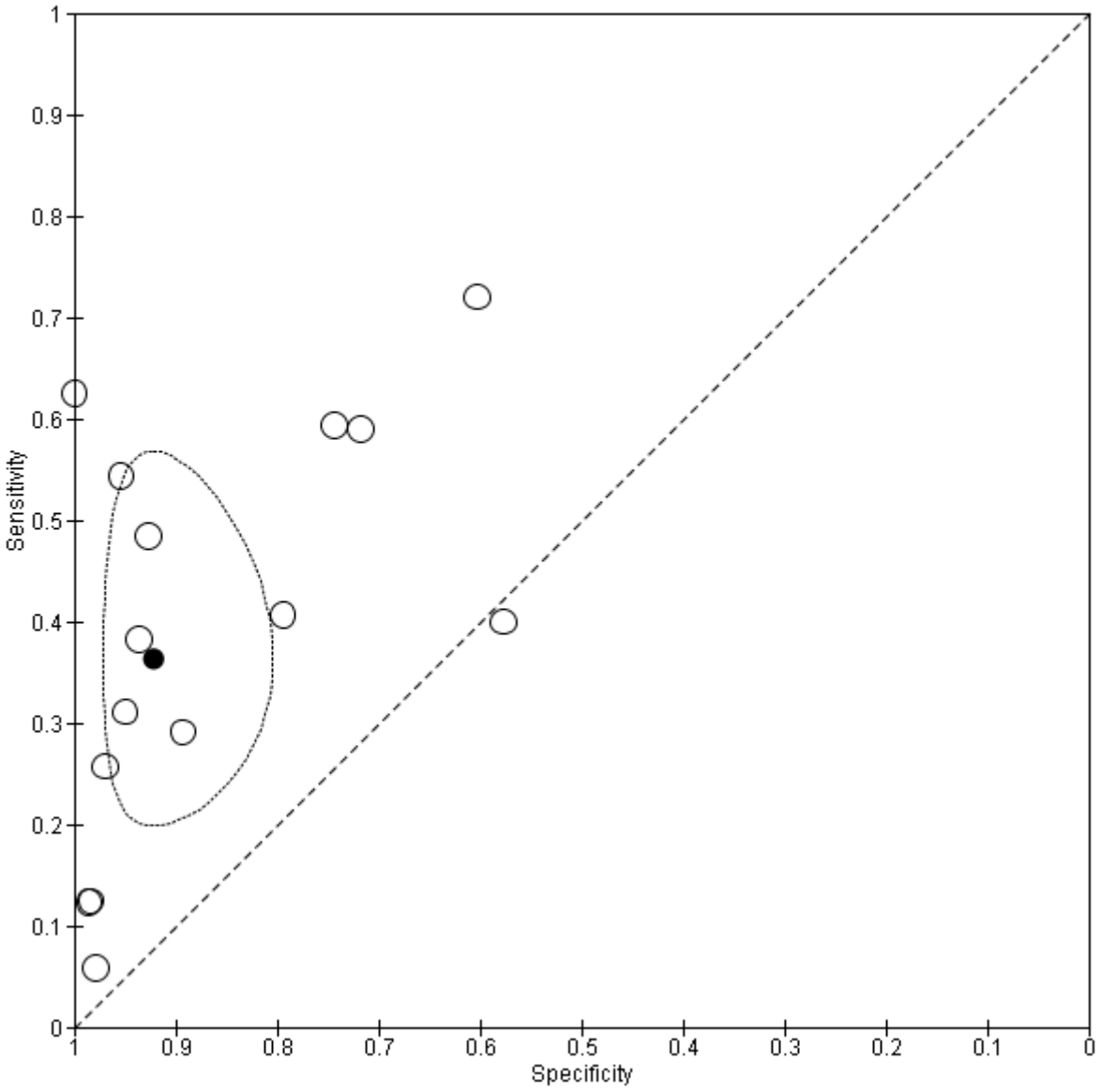


3

4

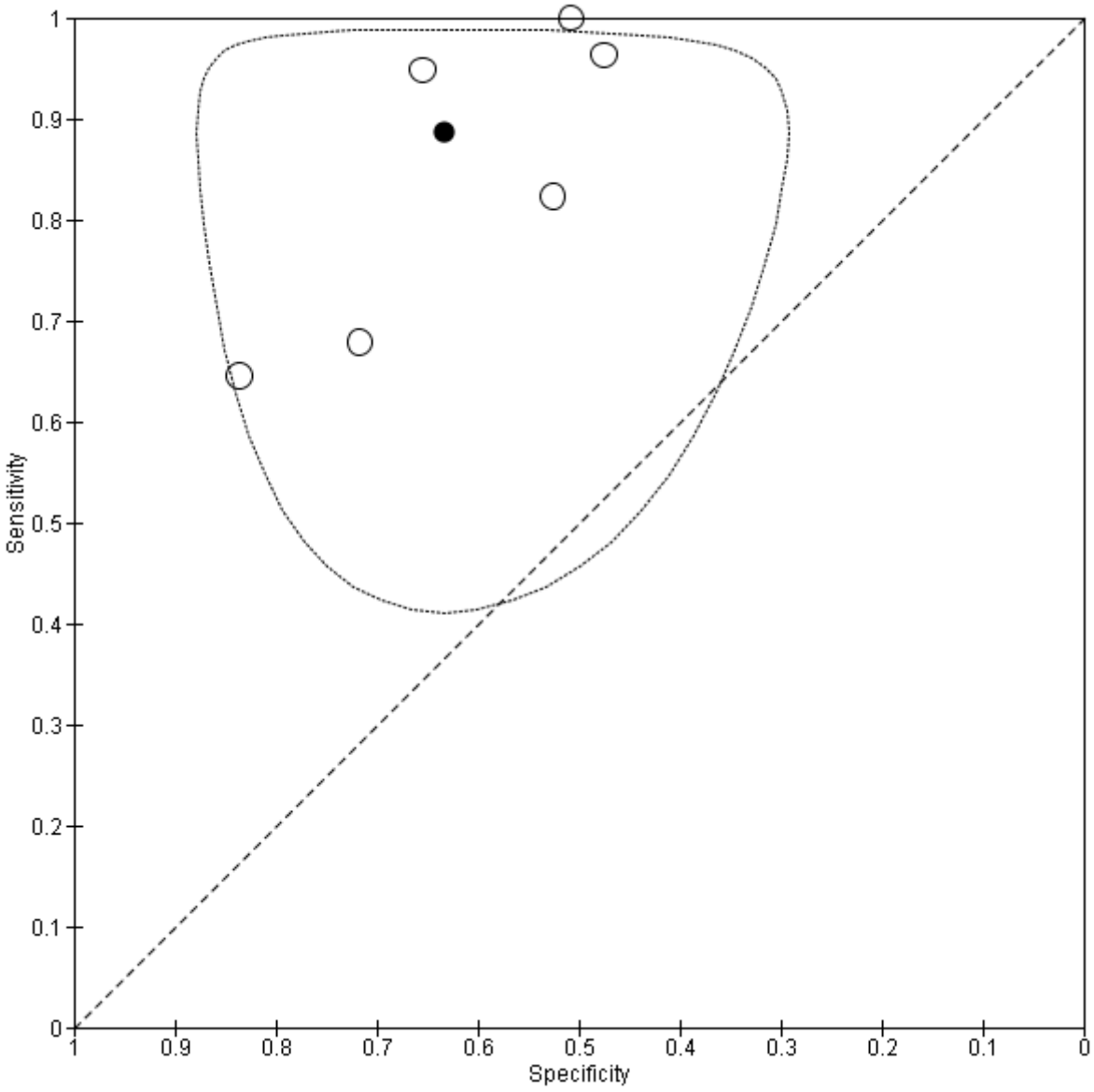
5

F1.6 Marked hypoechoicity



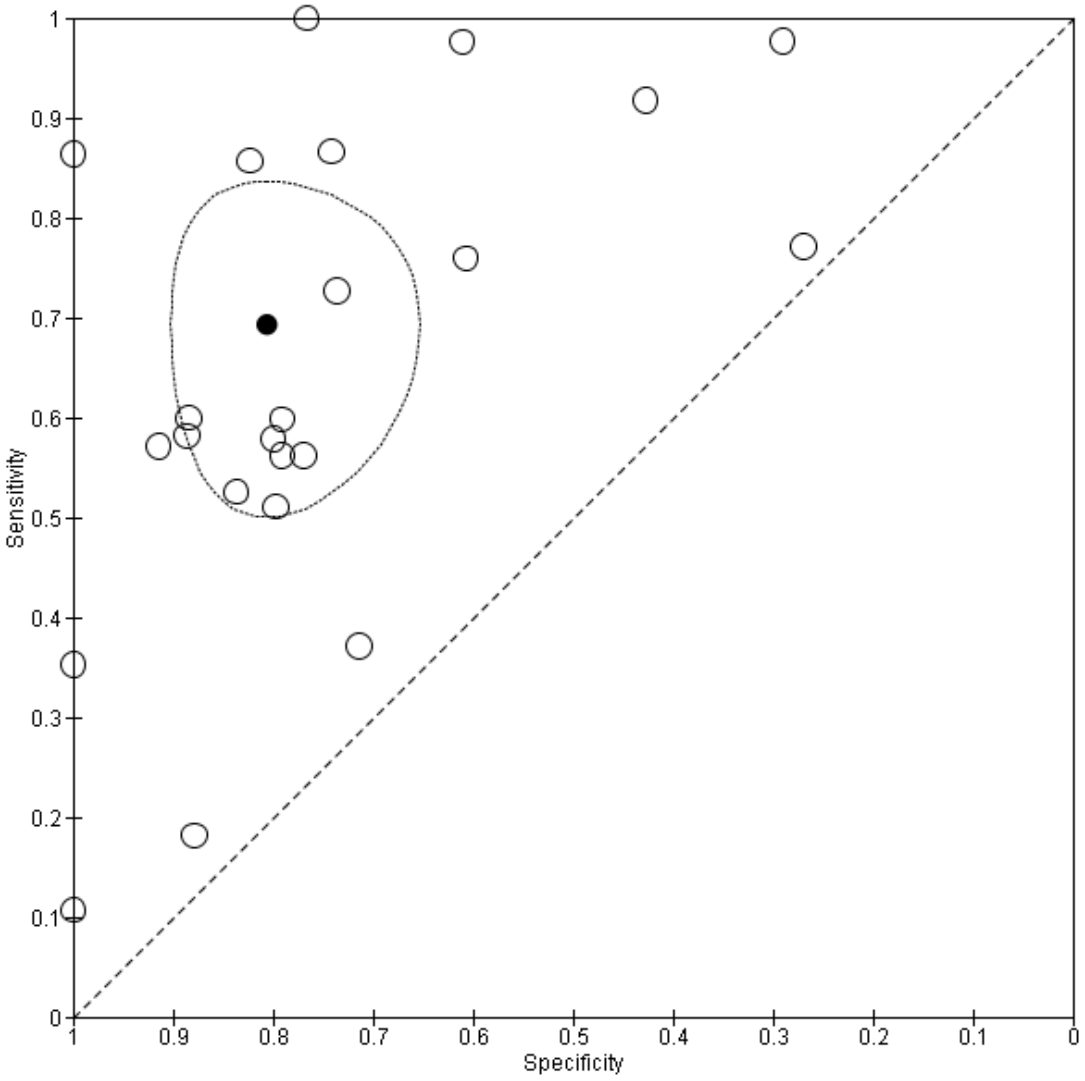
F1.7 Hypoechoicity OR marked hypoechoicity

2

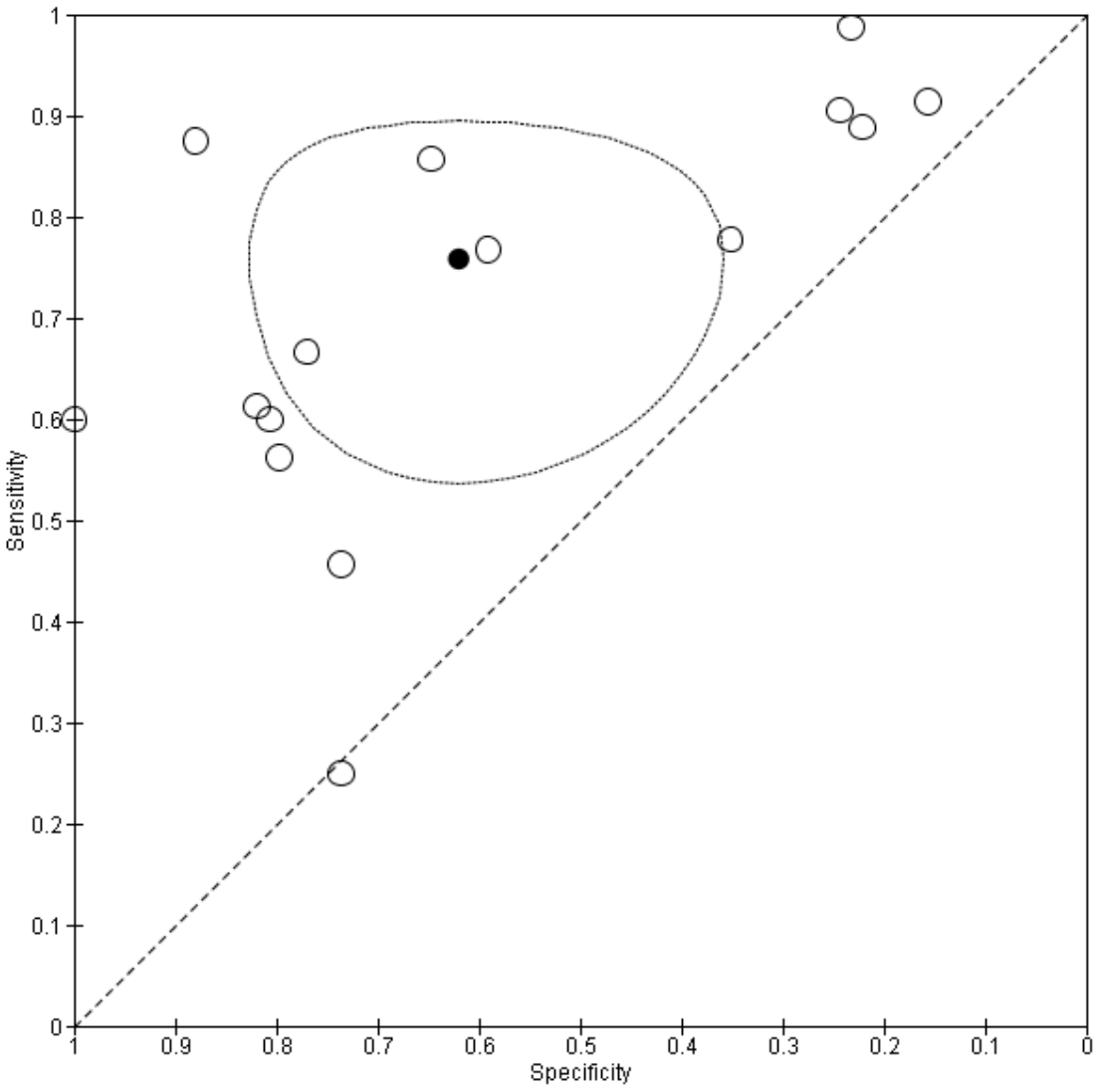


3

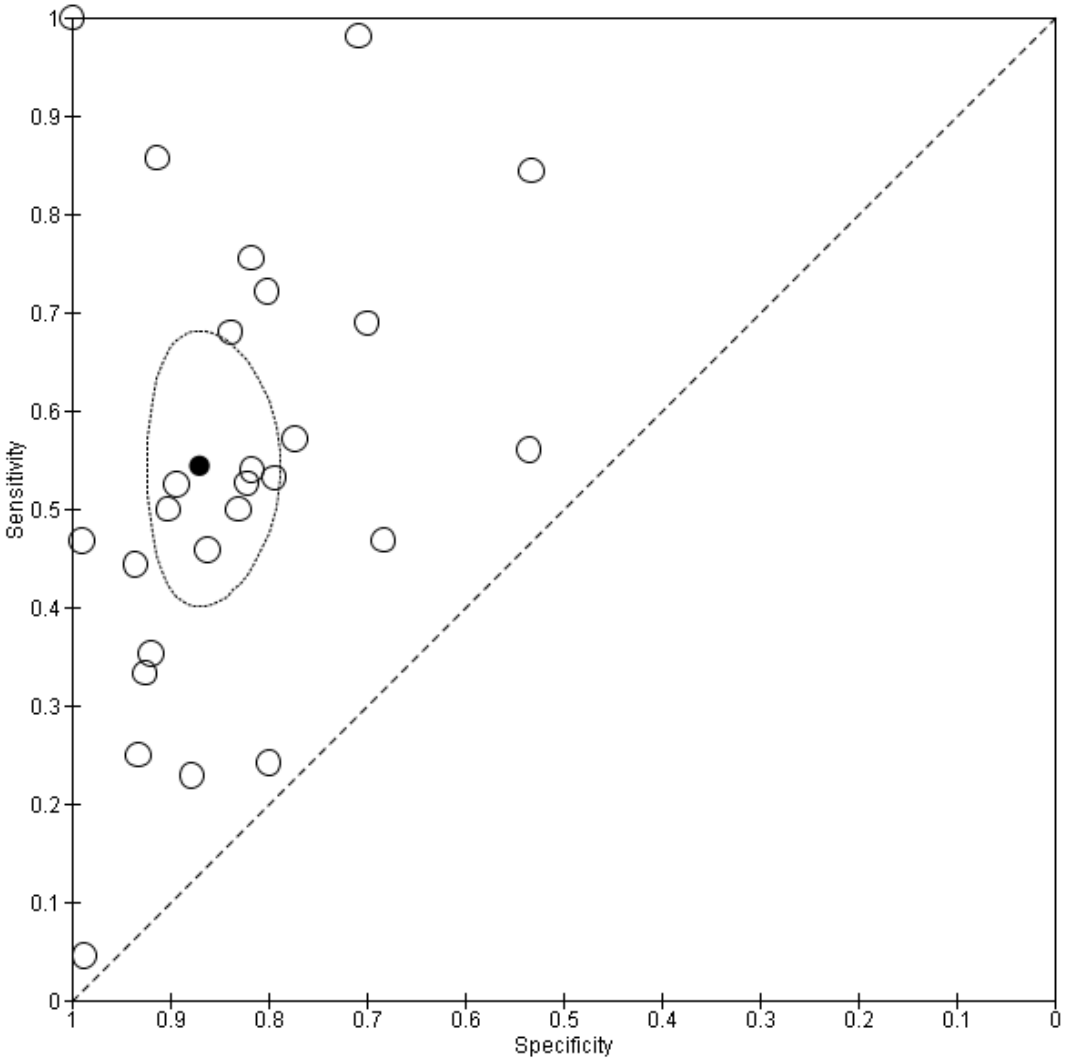
F1.8 Poorly defined margins/borders



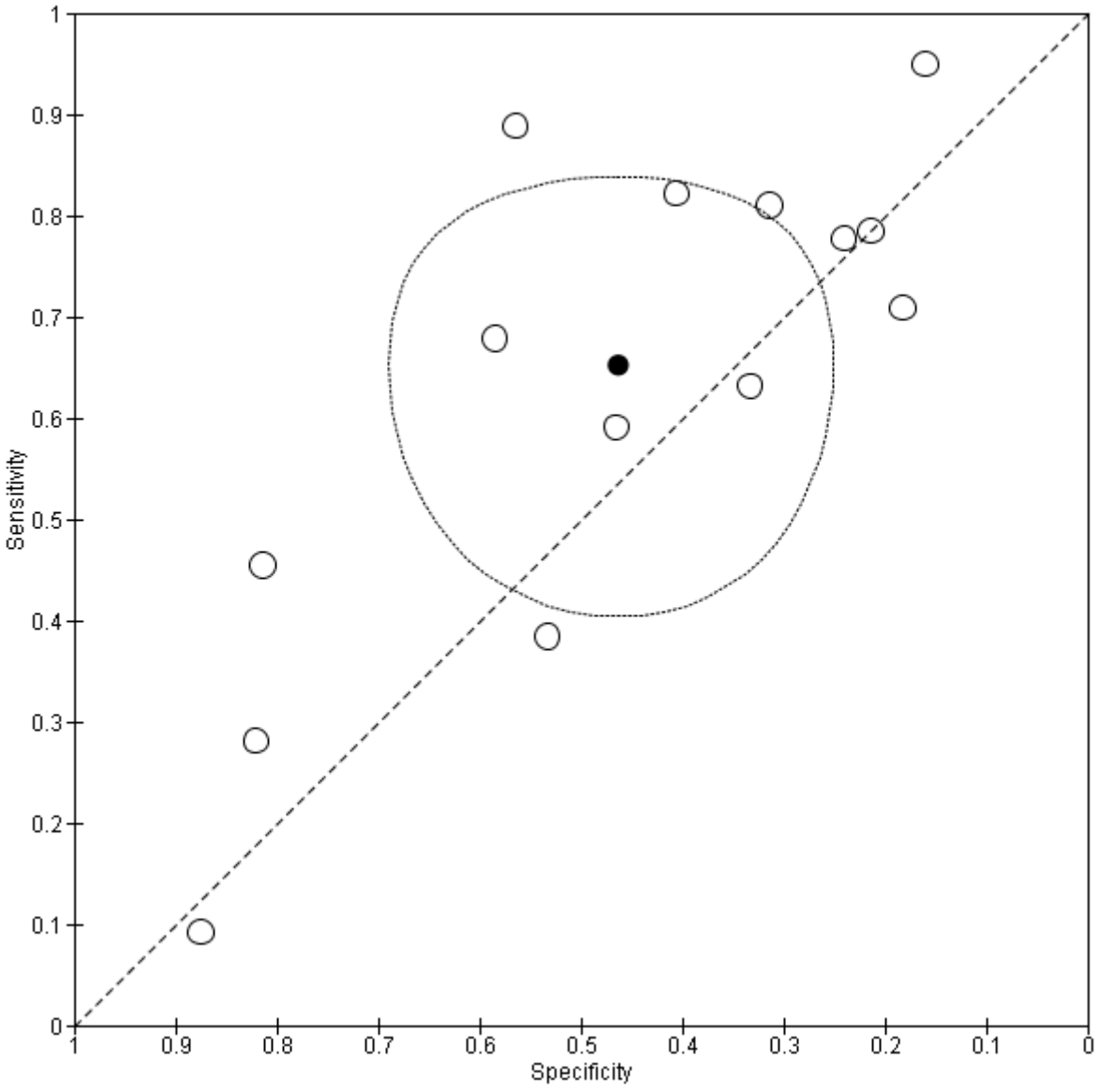
F1.9 Absent halo



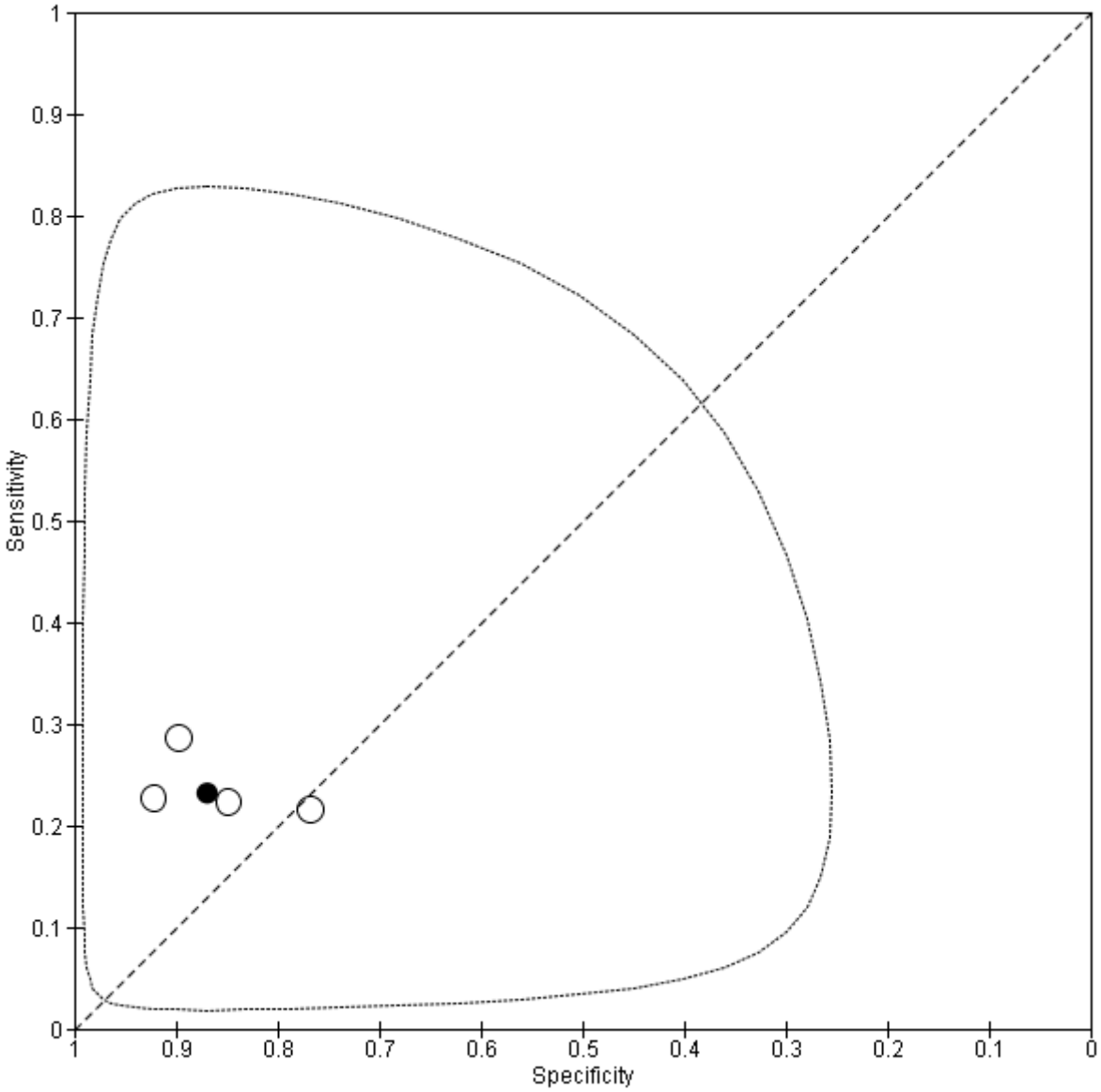
F.5.10 Irregular borders



F.9.11 Heterogeneous texture

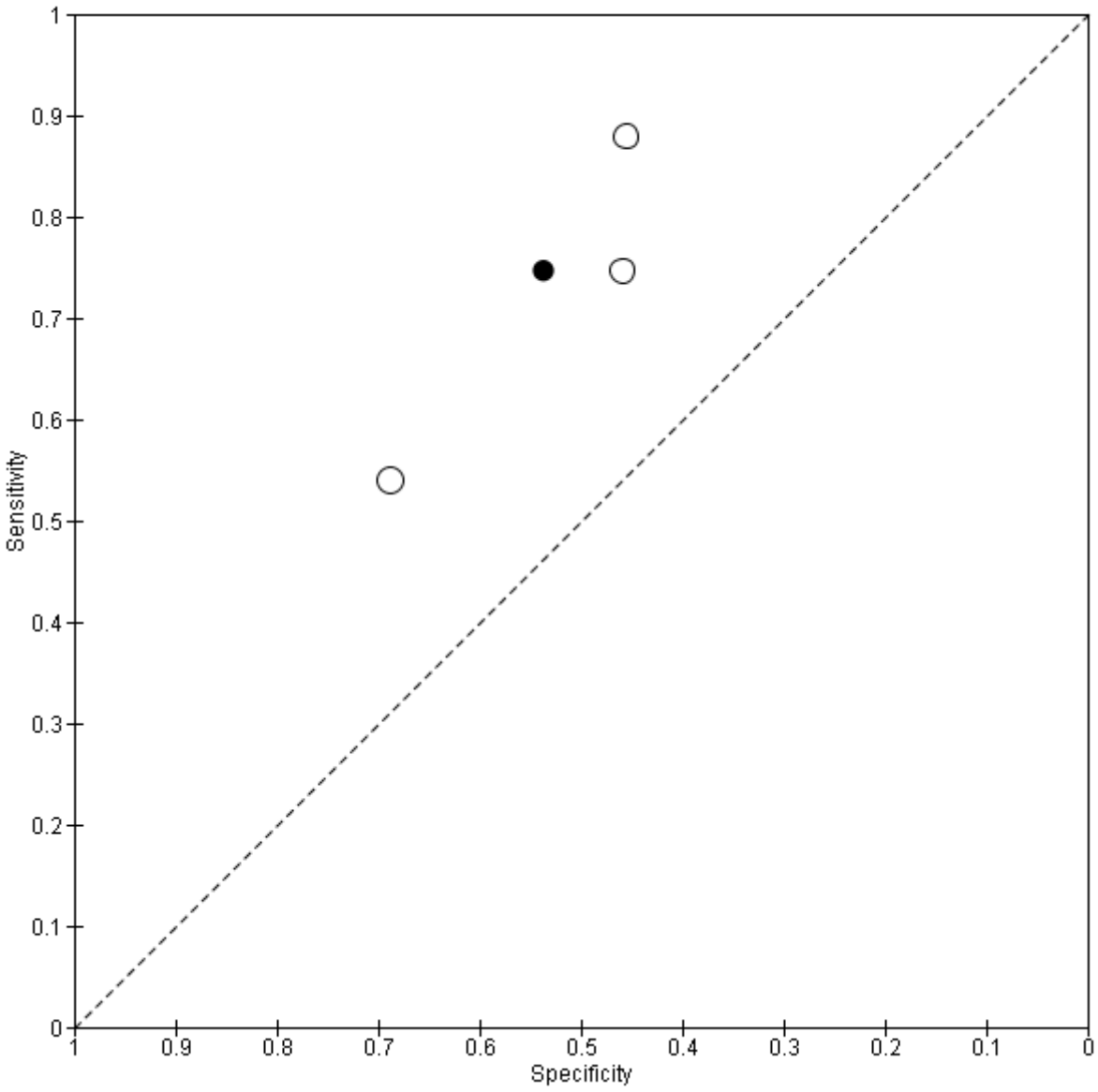


F.1.12 Macrocalcifications



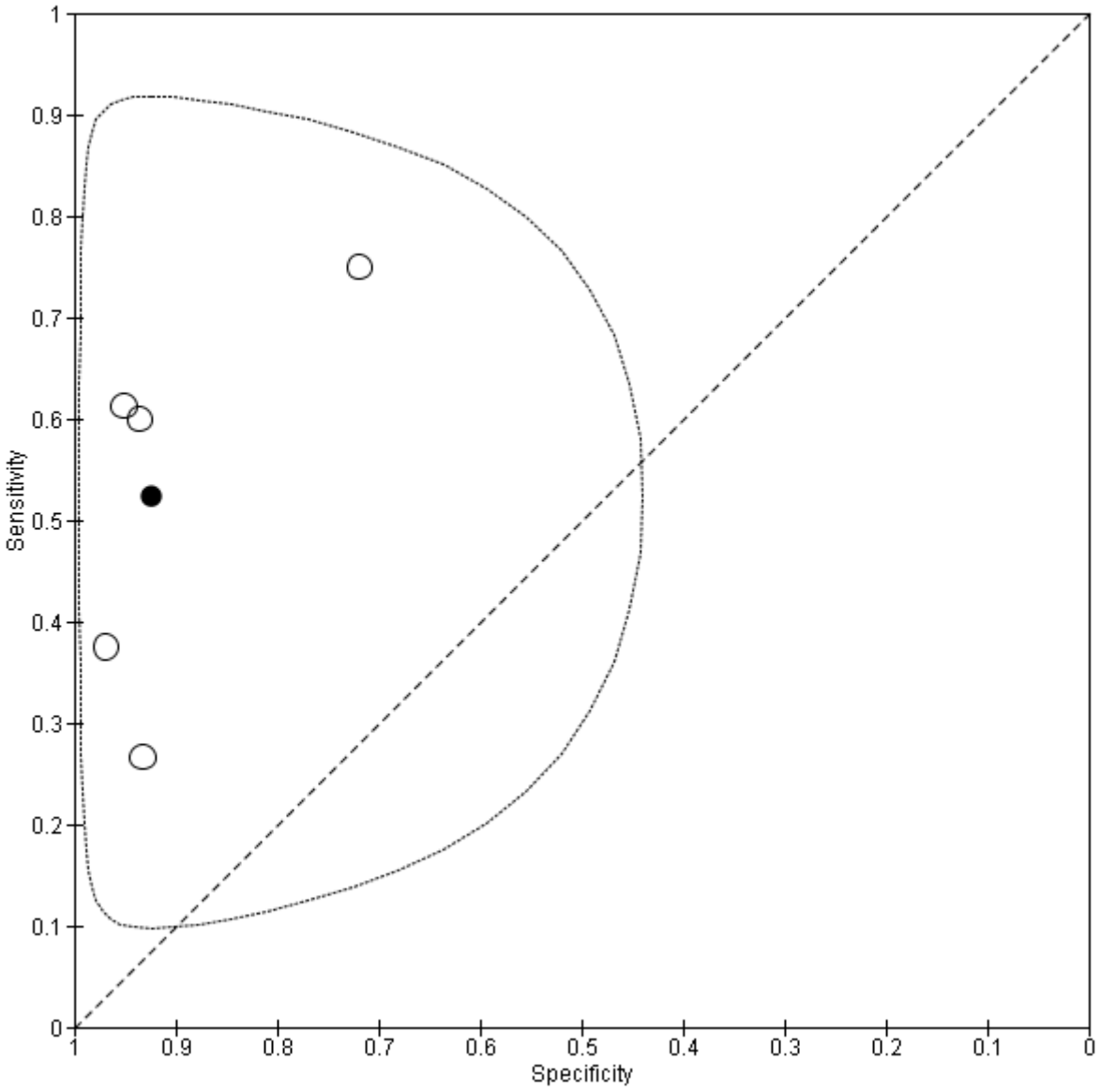
1

F.1.13 Nodule diameter or 20mm or less

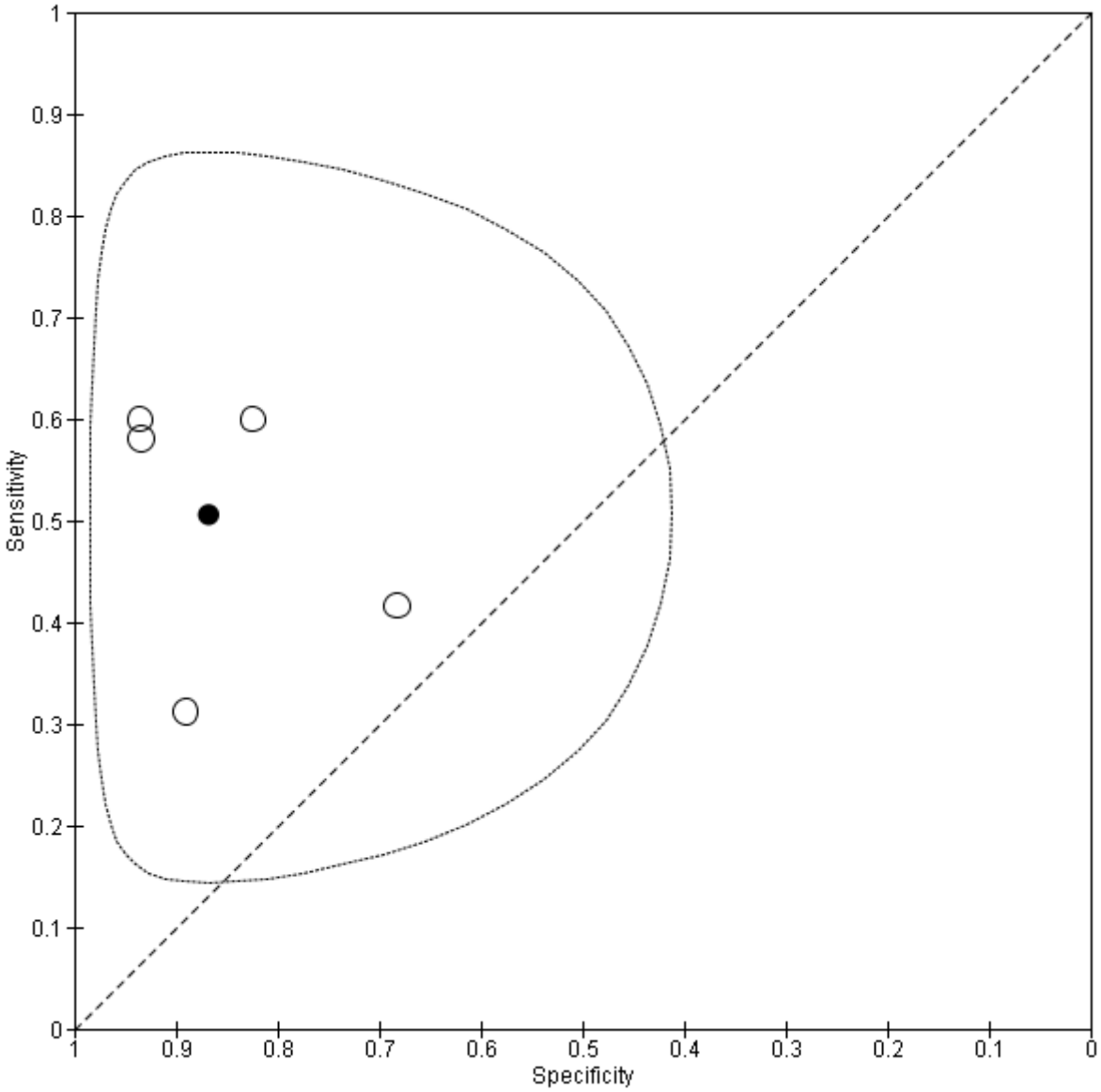


3

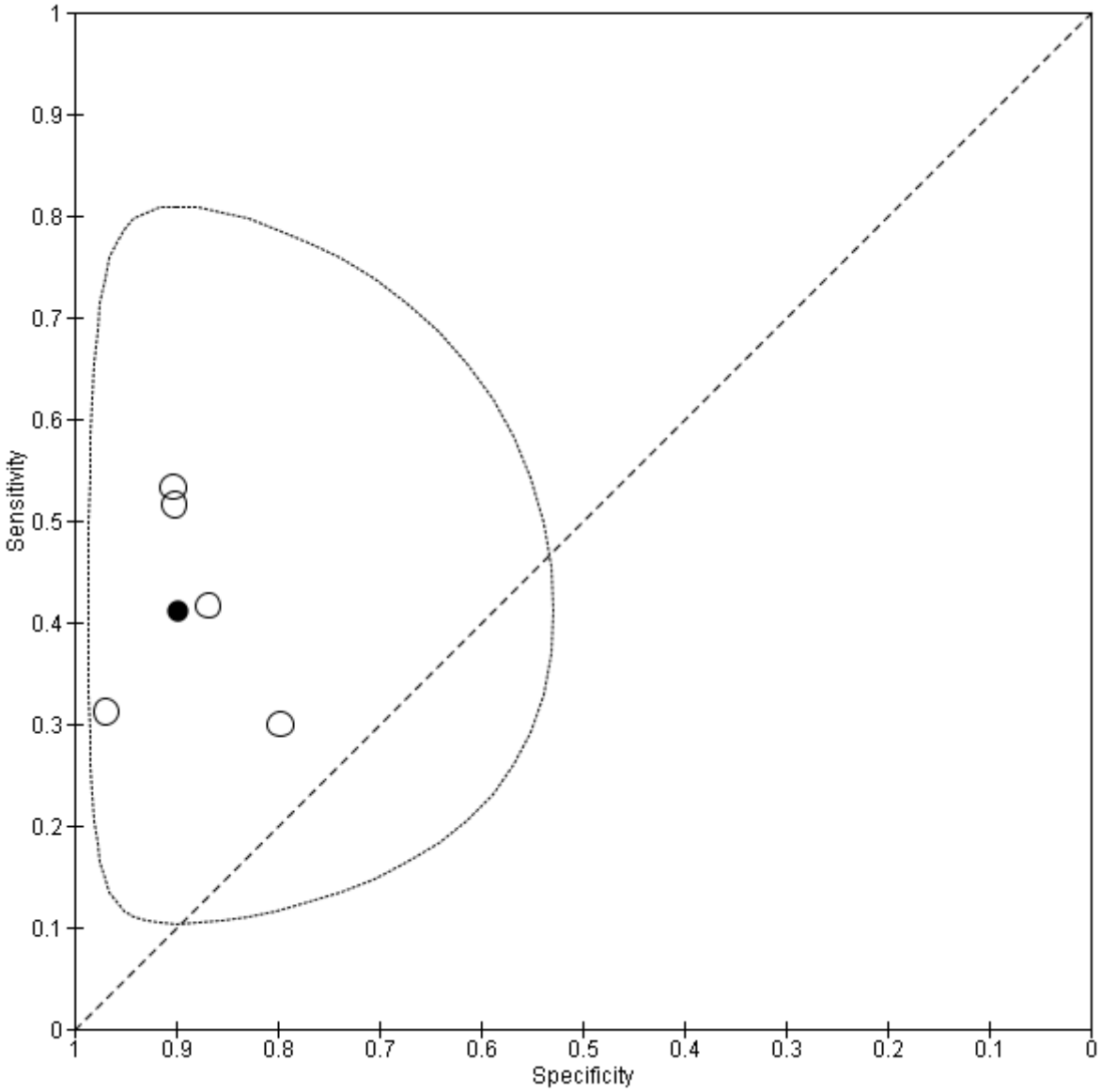
F.1.14 Microcalcifications AND absent halo



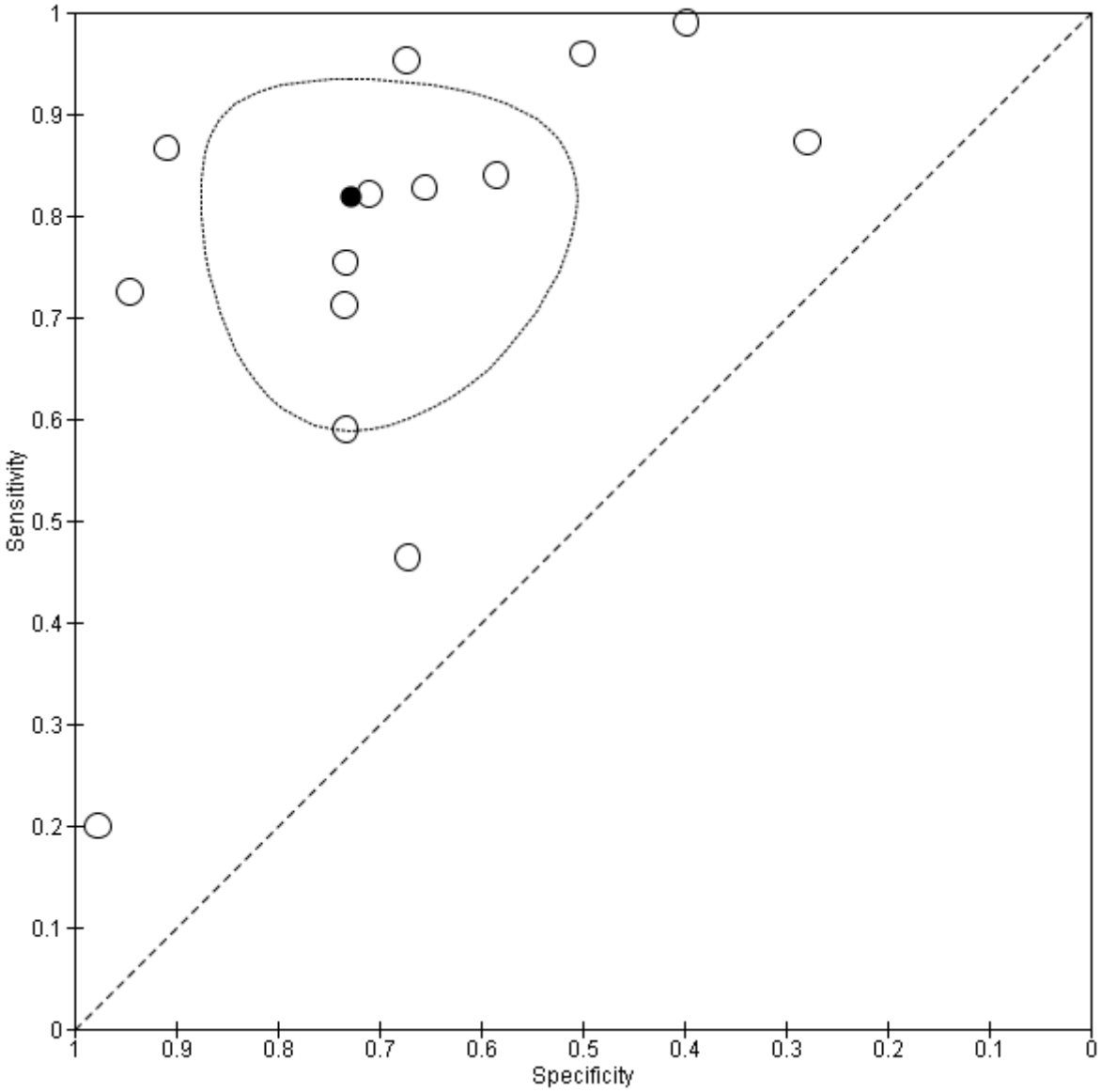
F.1.15 hypoechoicity AND absent halo



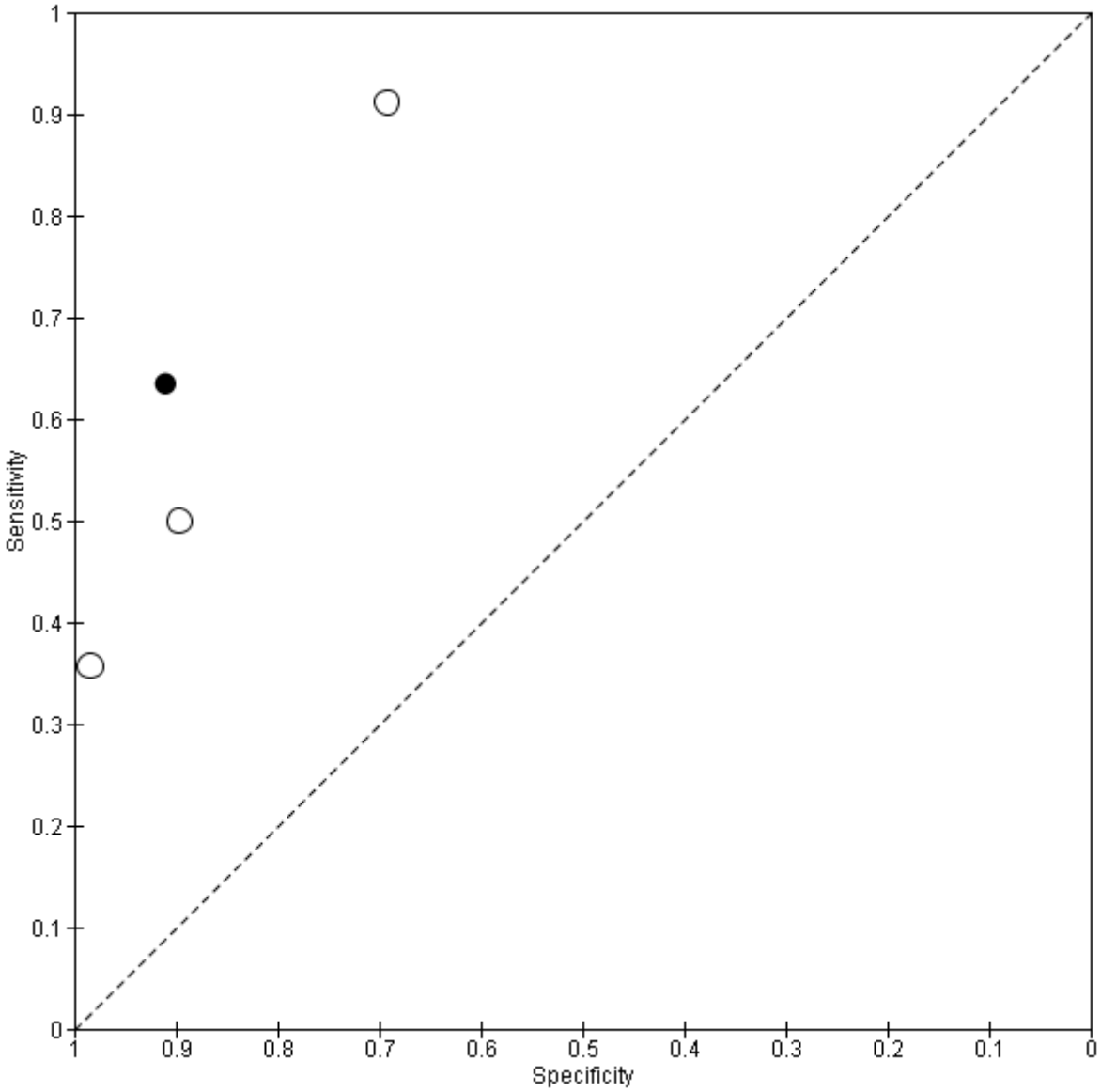
F.1.16 Microcalcifications AND hypoechoicity



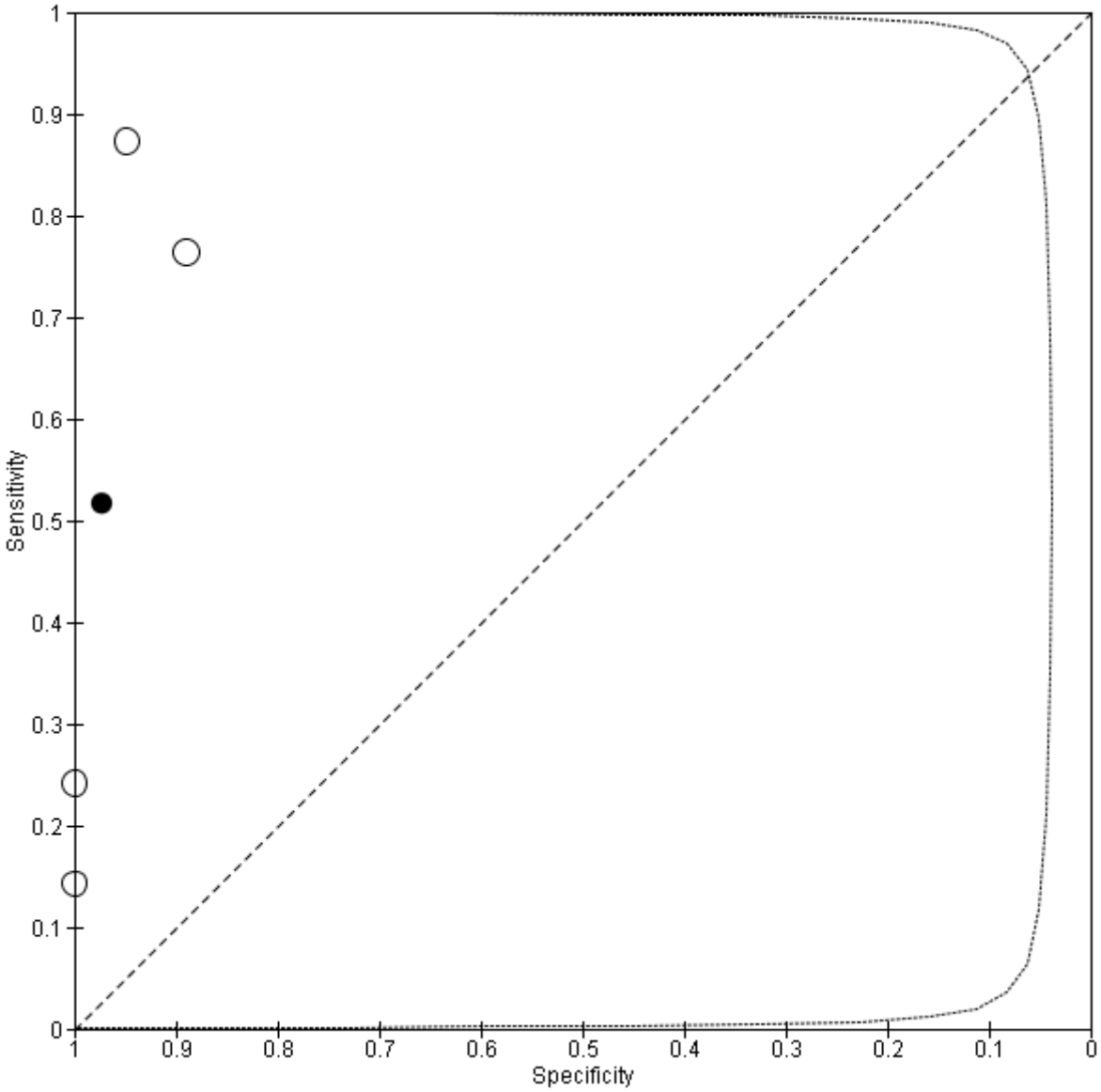
F.1.17 At least one US sign



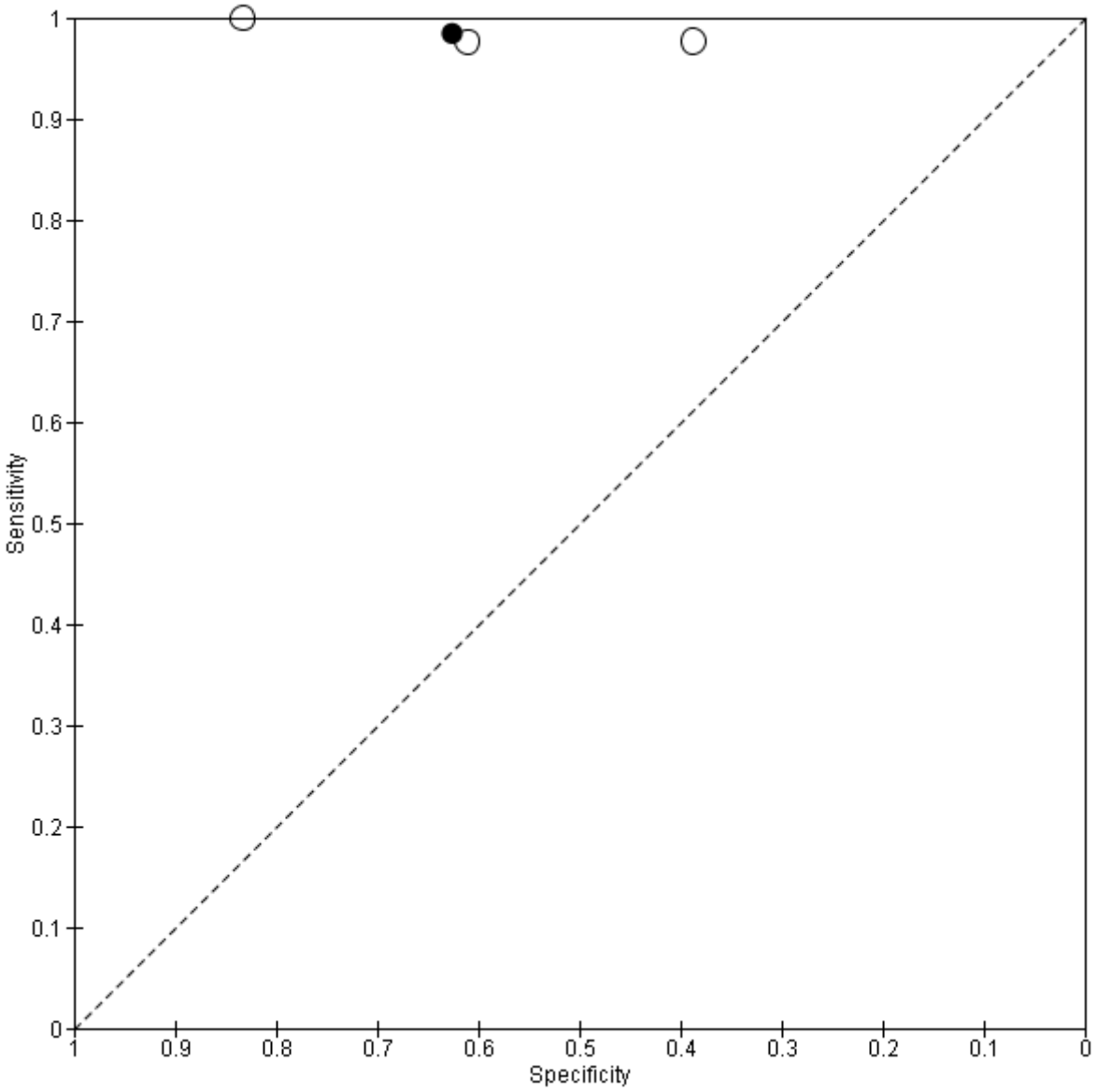
F.2.18 At least 2 US signs



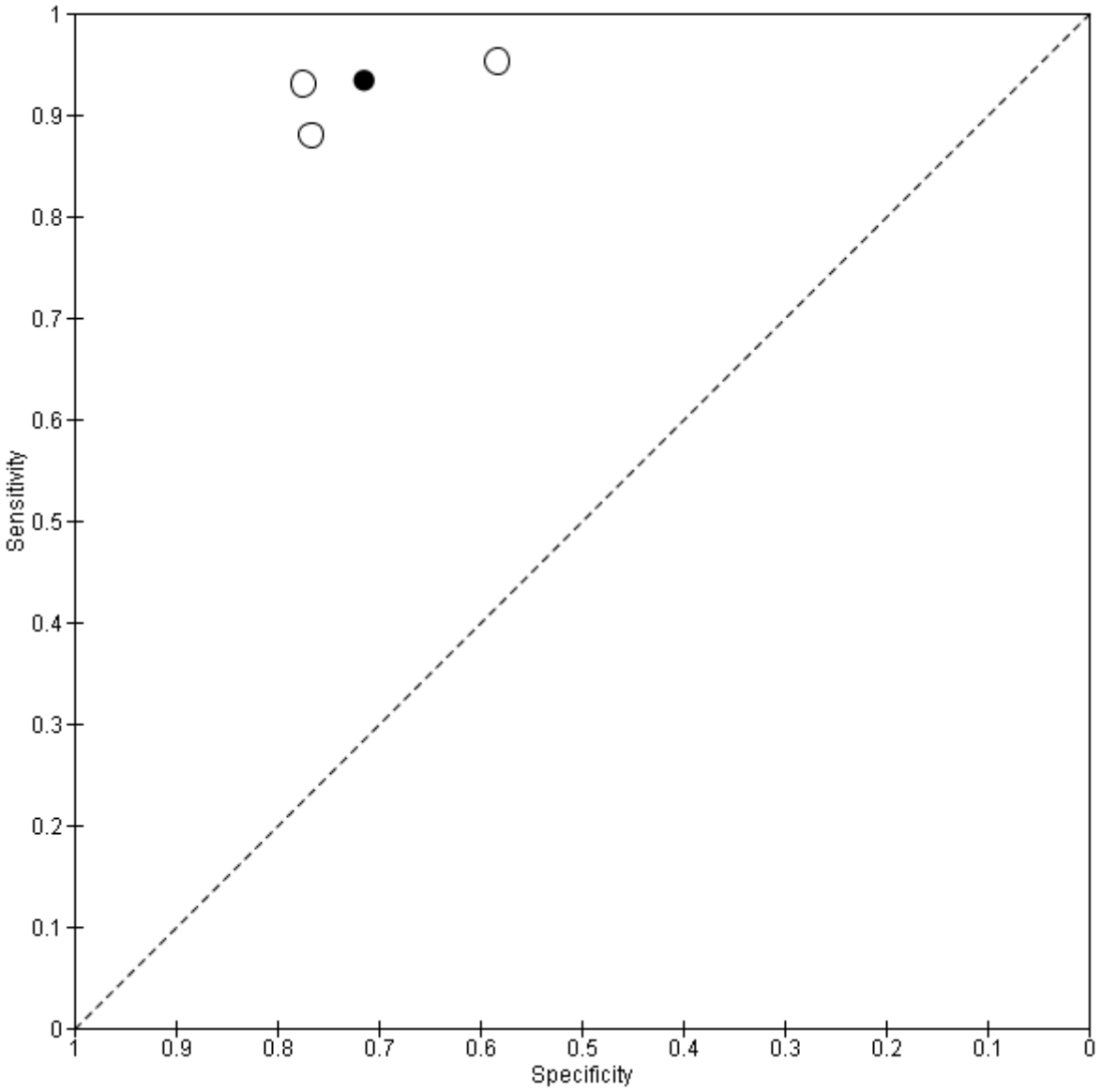
F.1.19 At least 3 US signs



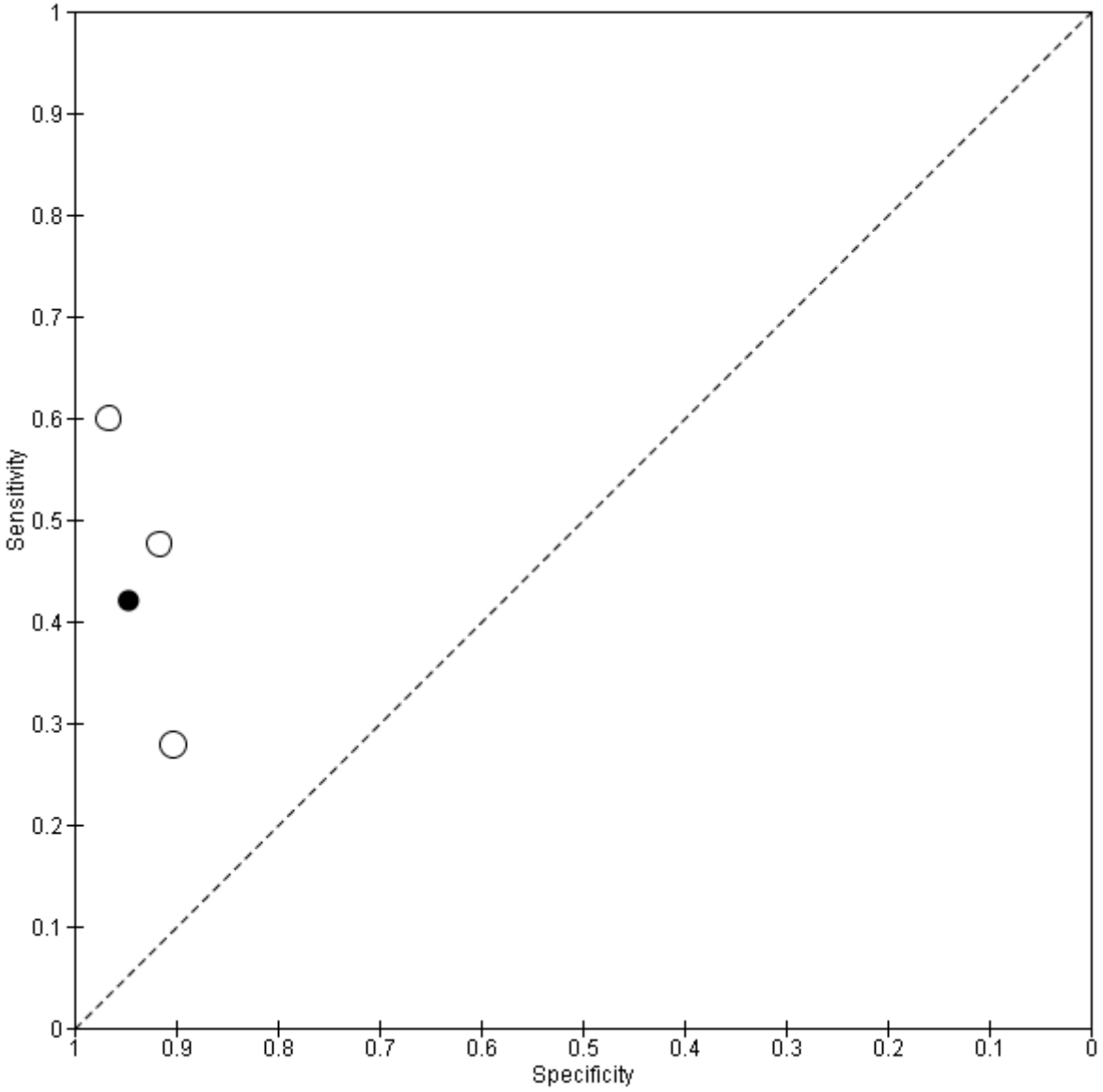
F.1.20 Blurred margins and any one of hypoechoicity, microcalcification or taller than wide



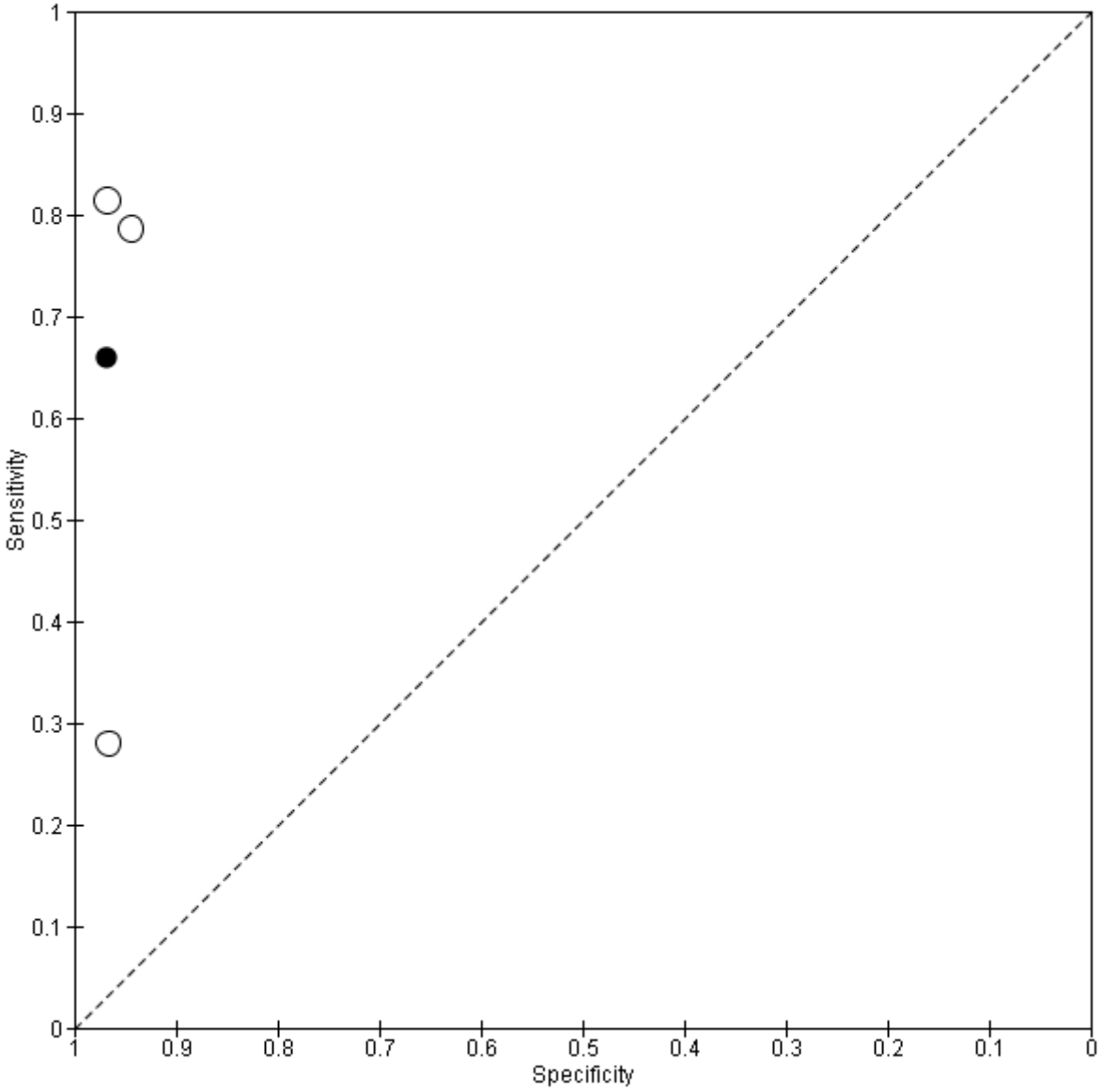
F.1.21 hypoechoicity and any one of blurred margins, microcalcification or taller than wide



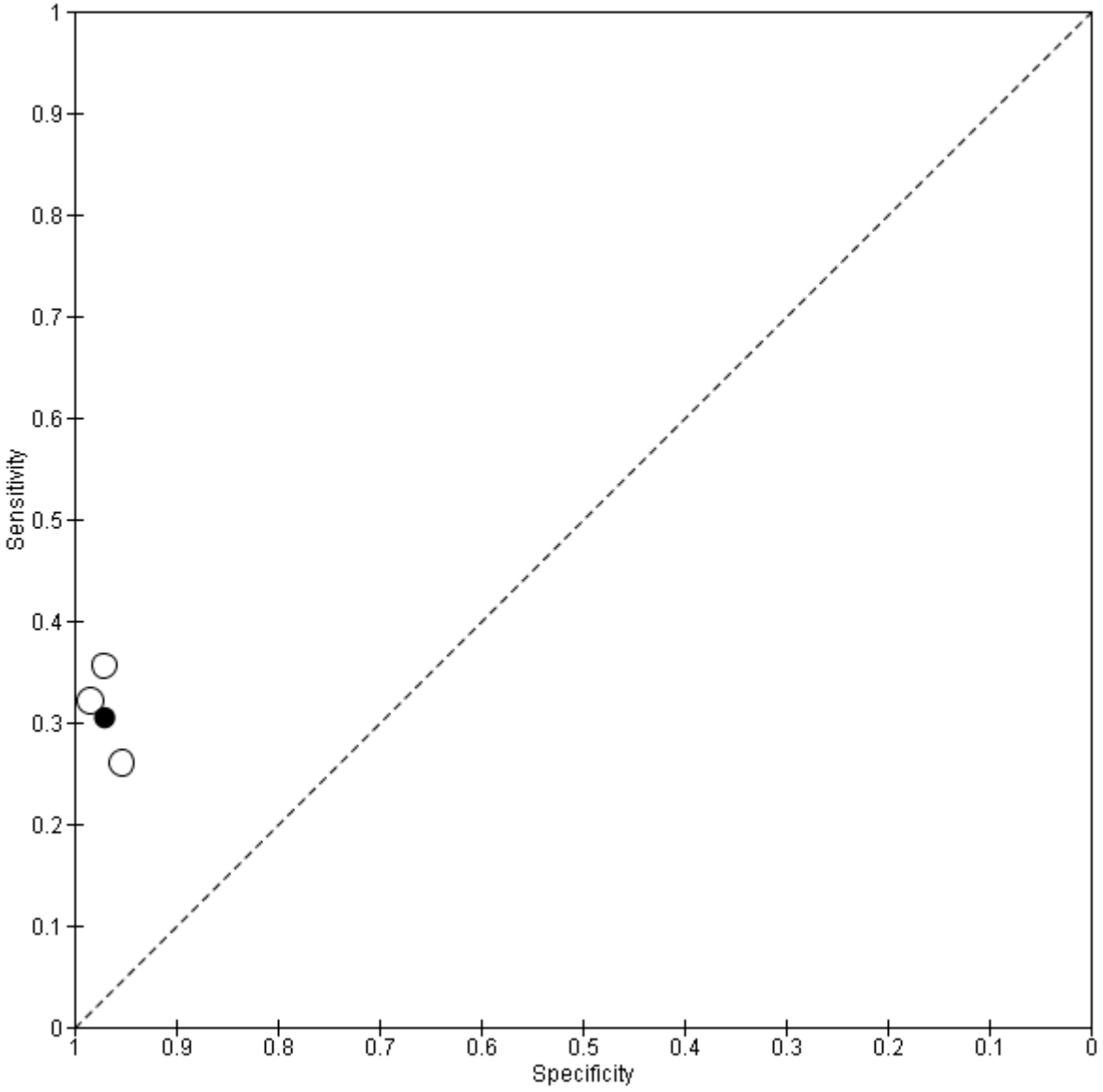
F.1.22 Microcalcifications and any one of blurred margins, hypoechoicity or taller than wide



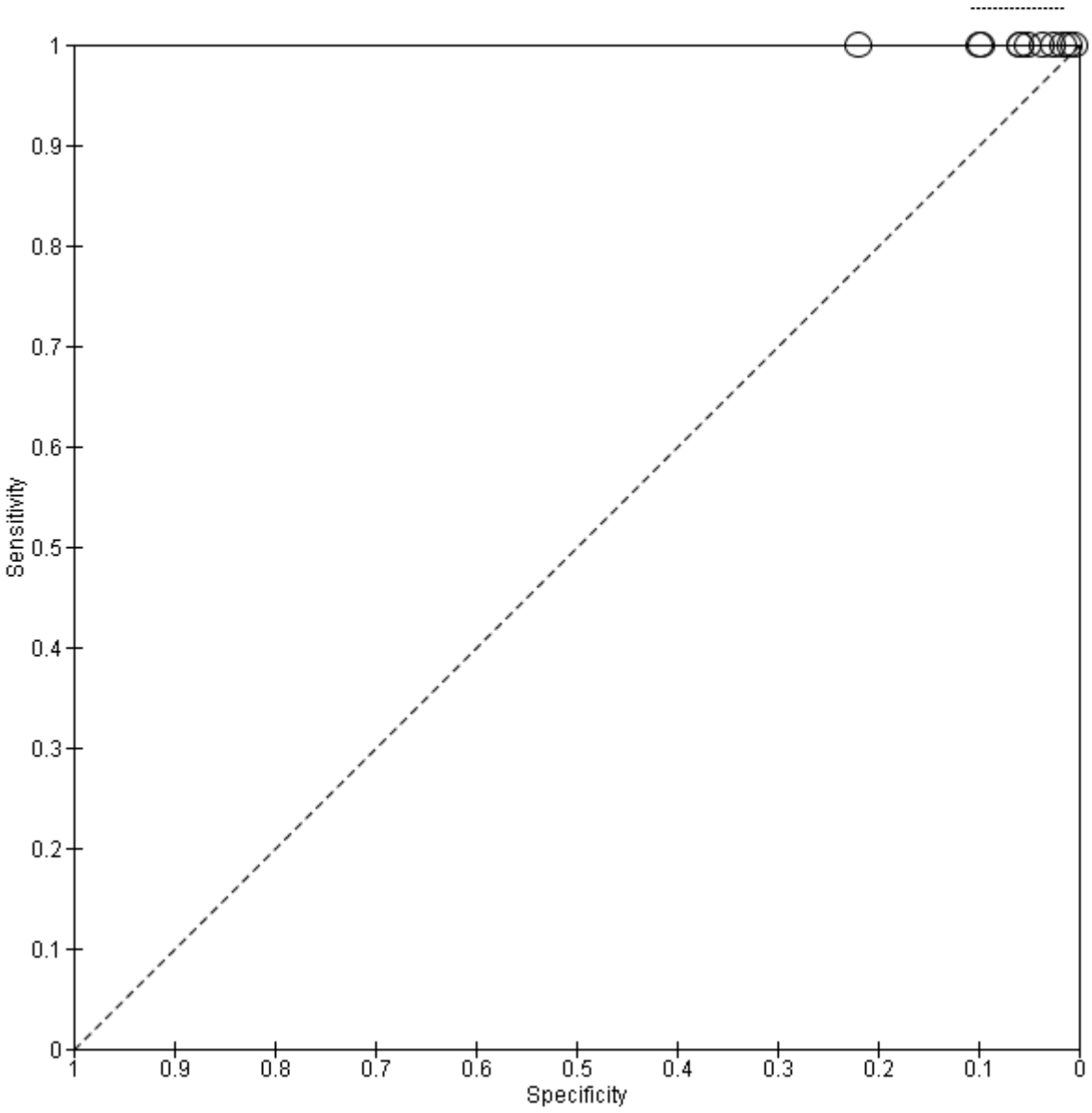
F.1.23 Taller than wide and any one of blurred margins, hypoechoicity or microcalcifications



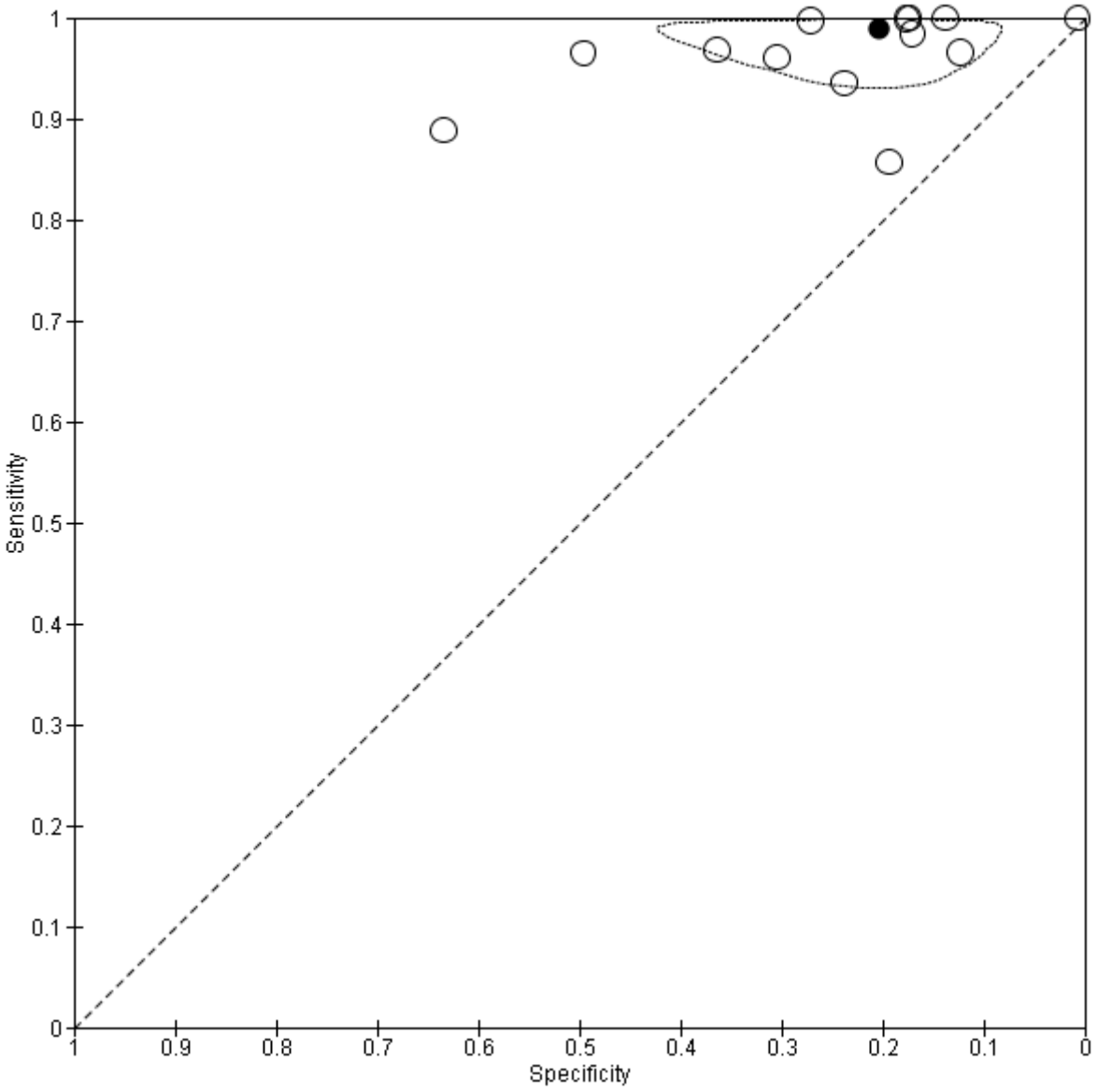
F.1.24 Microlobulated or irregular margins



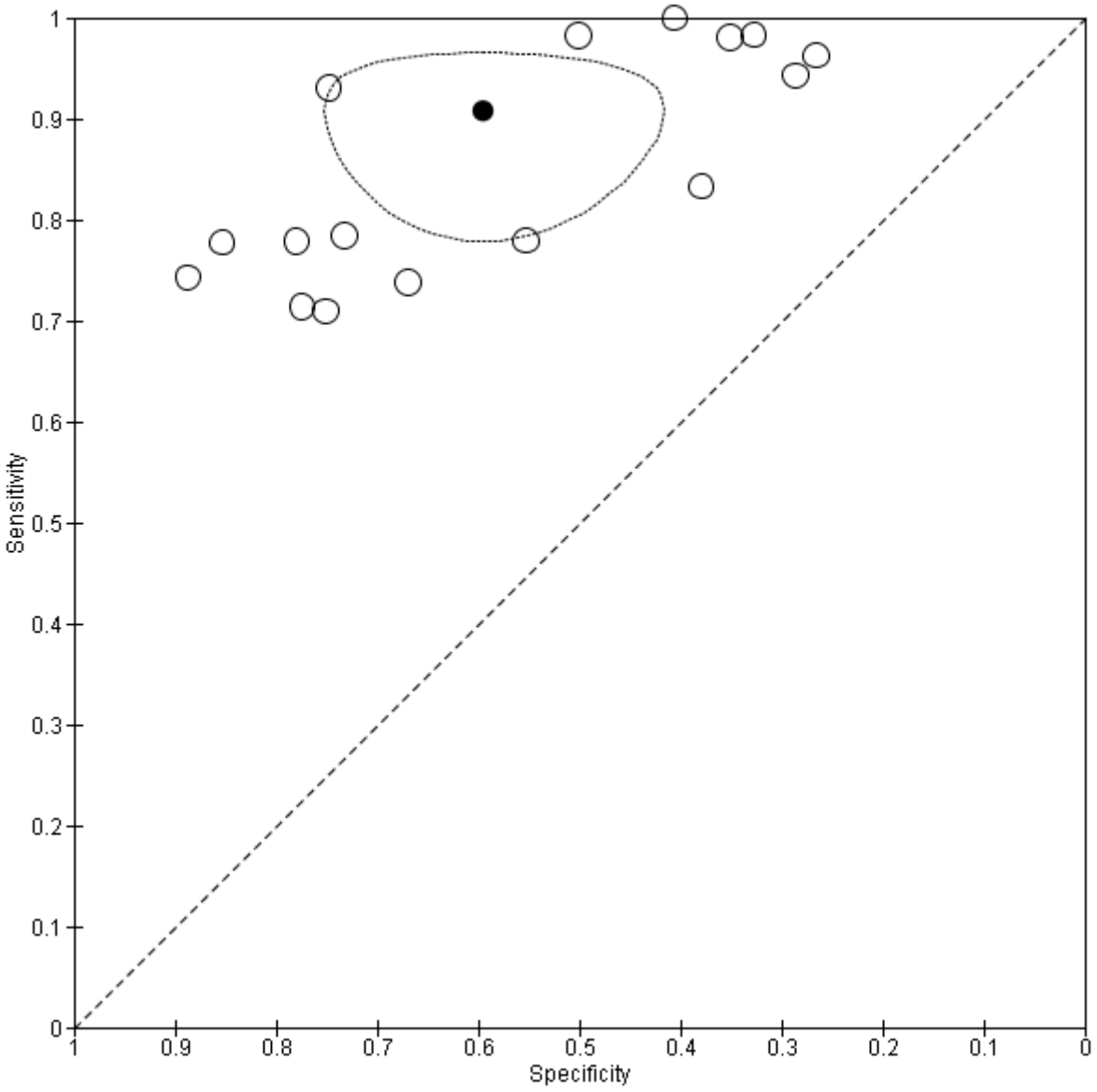
F.1.25 ACR TIRADS of 2 or more



F.1.26 ACR TIRADS 3 or more

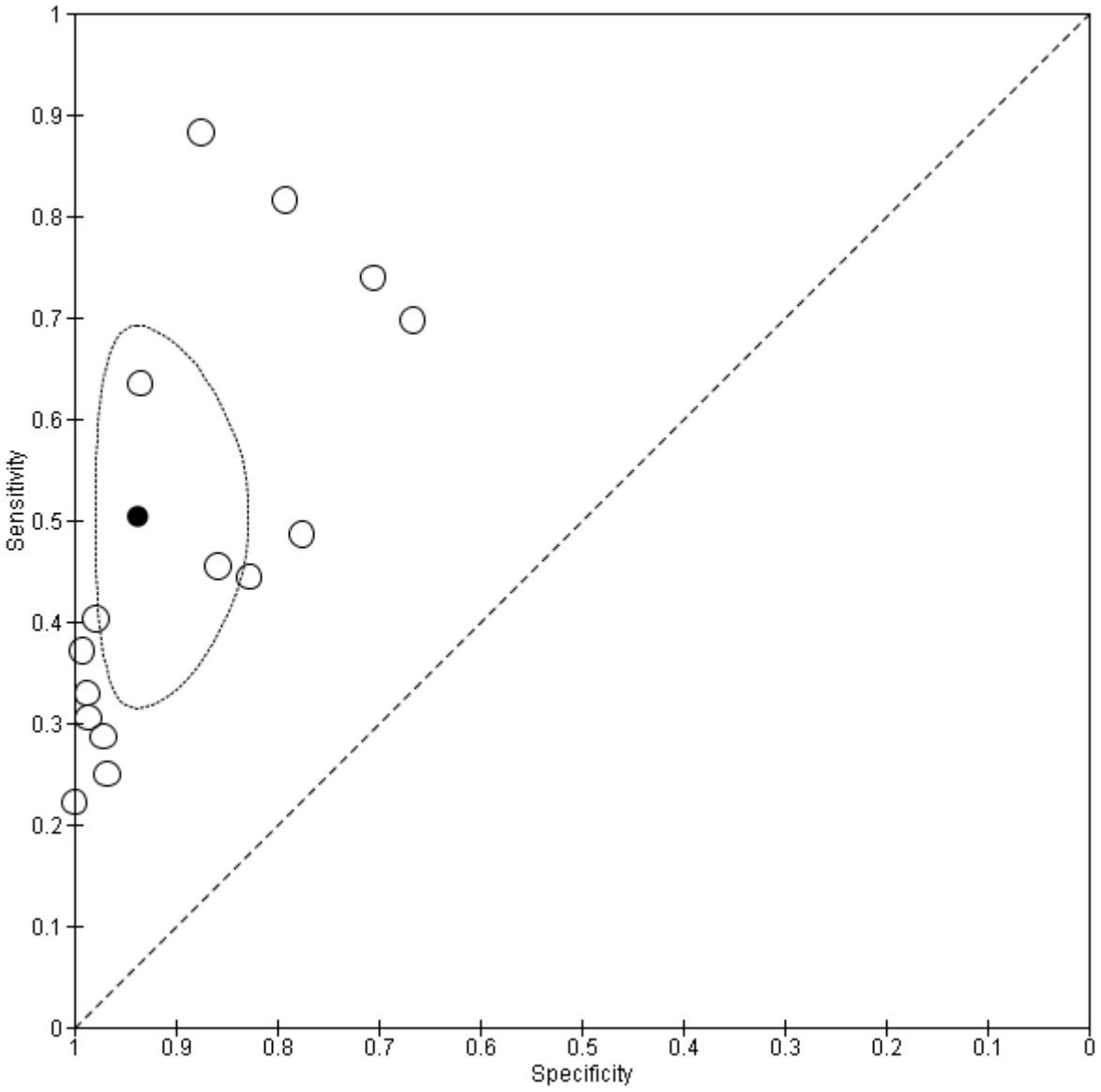


F.1.27 ACR TIRADS of 4 or more



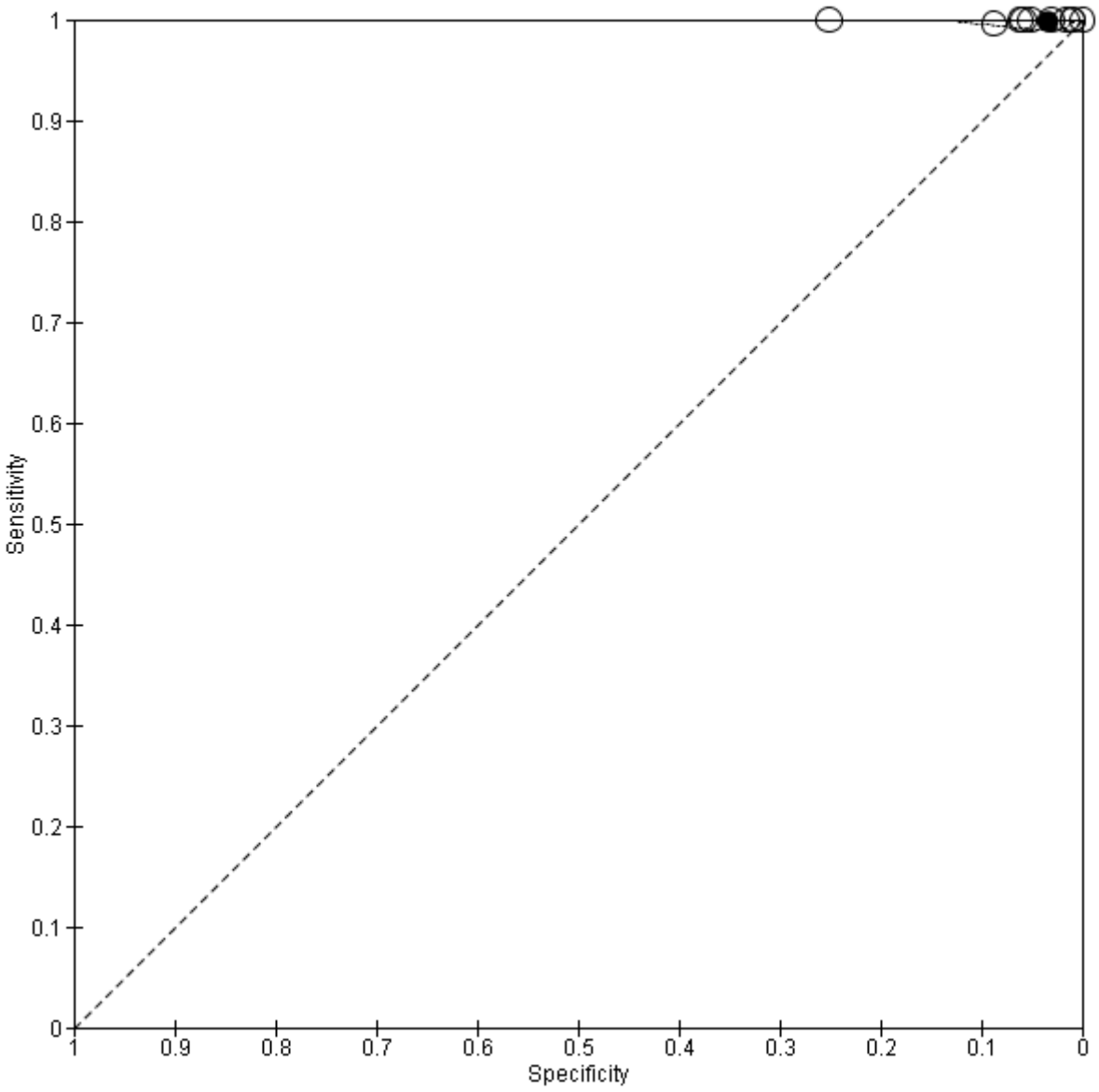
2

F.1.28 ACR TIRADS 5

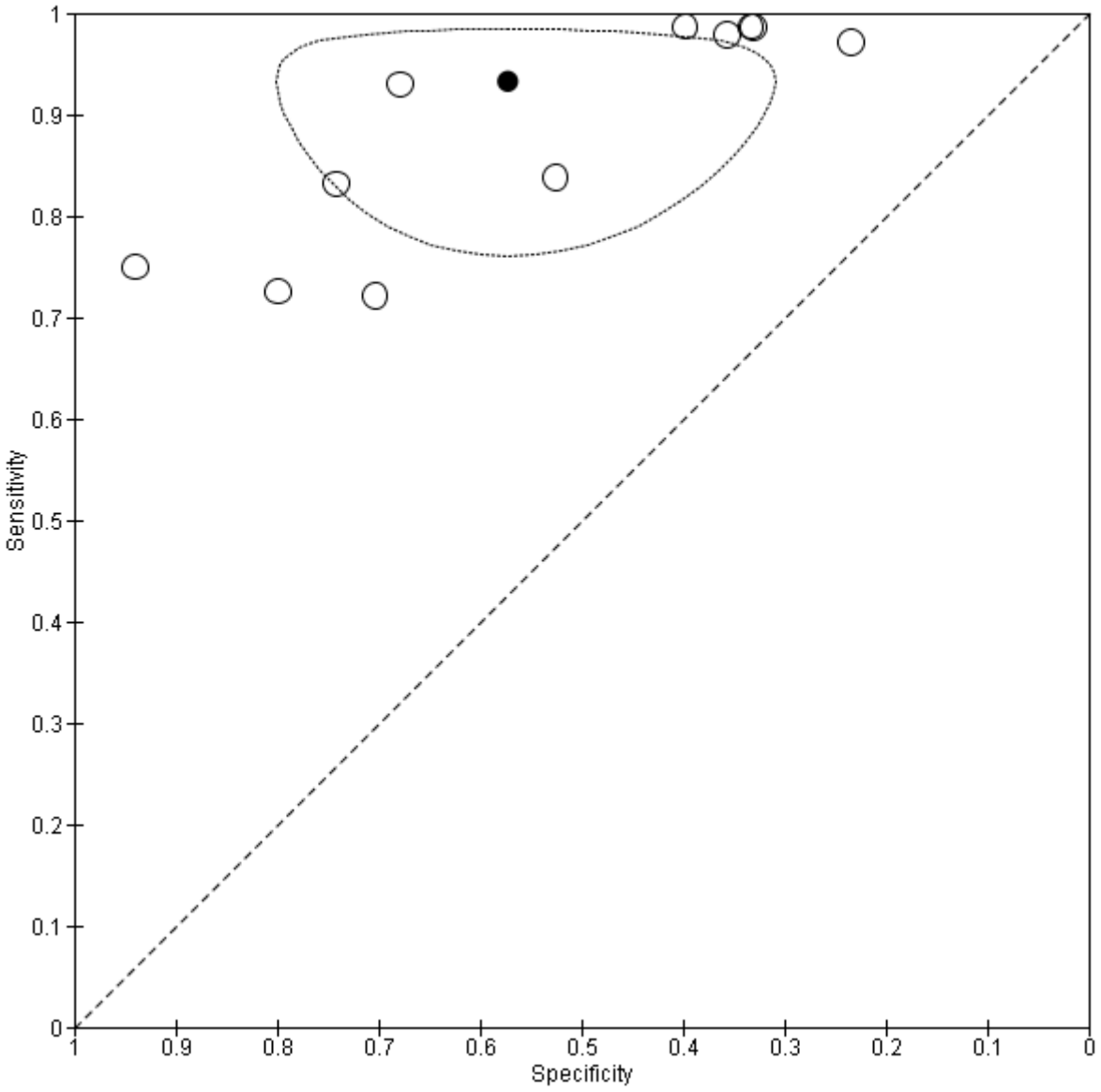


2

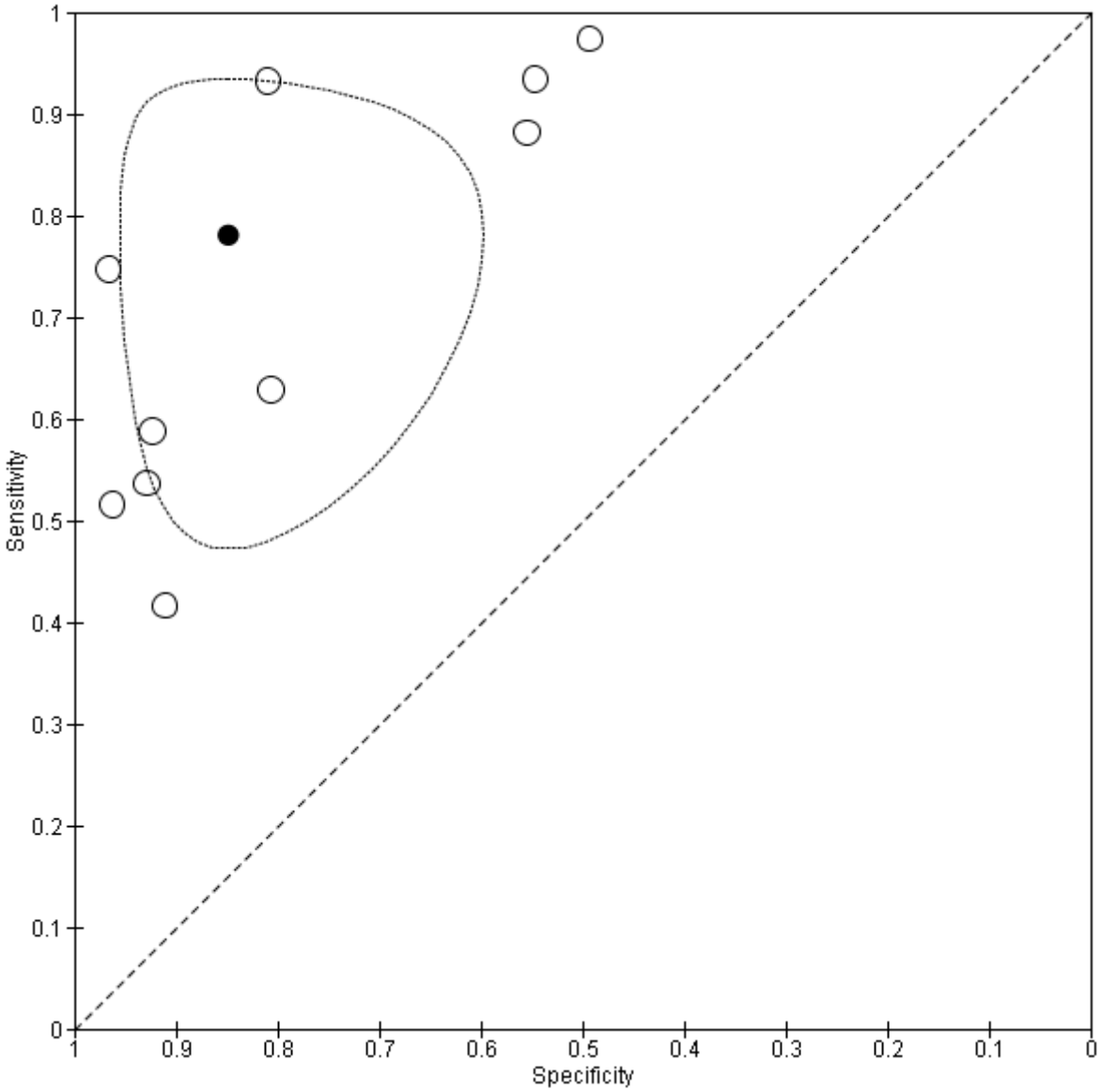
F.1.29 EU TIRADS 3 or more



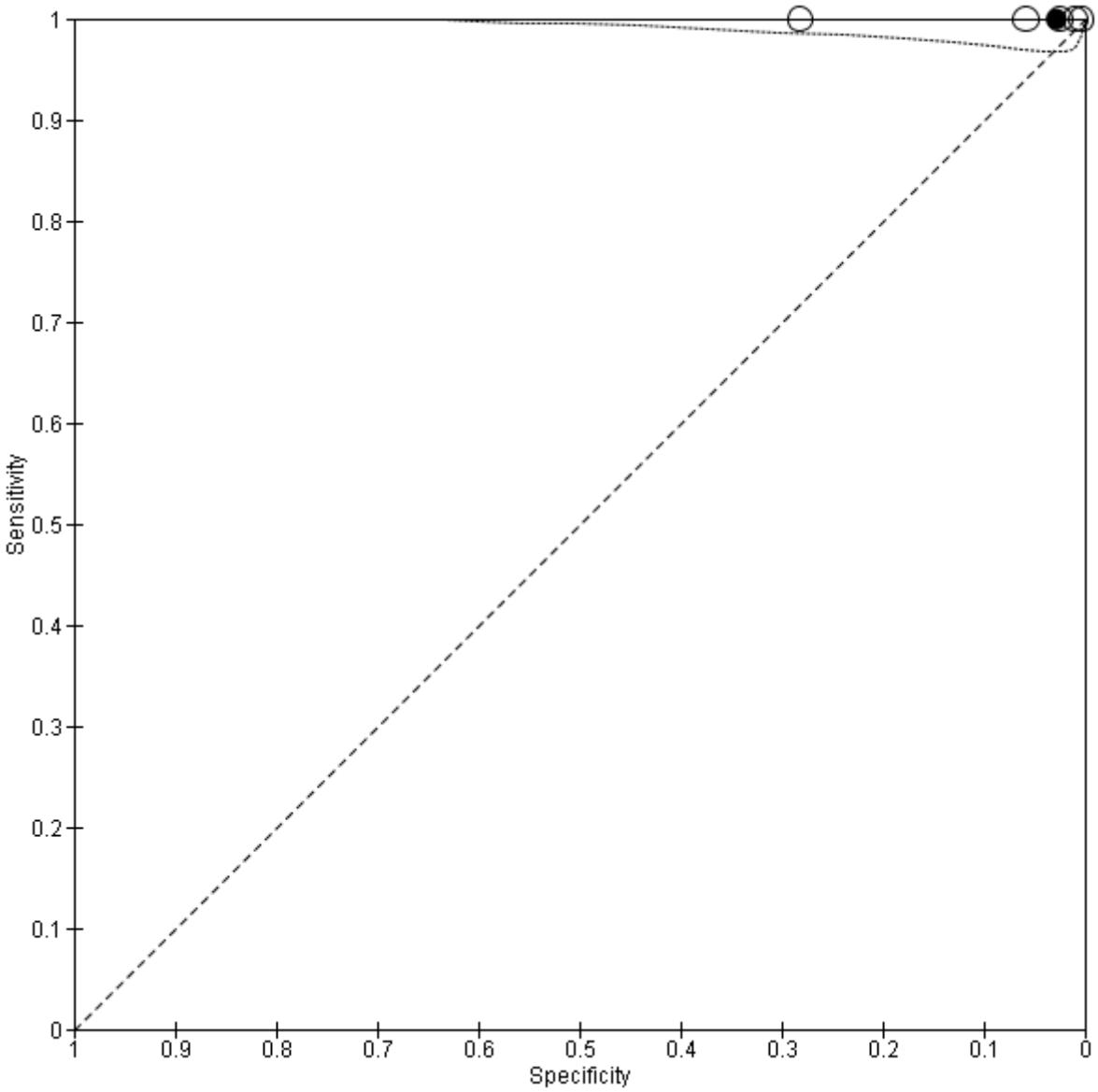
F.1.30 EU TIRADS 4 or more



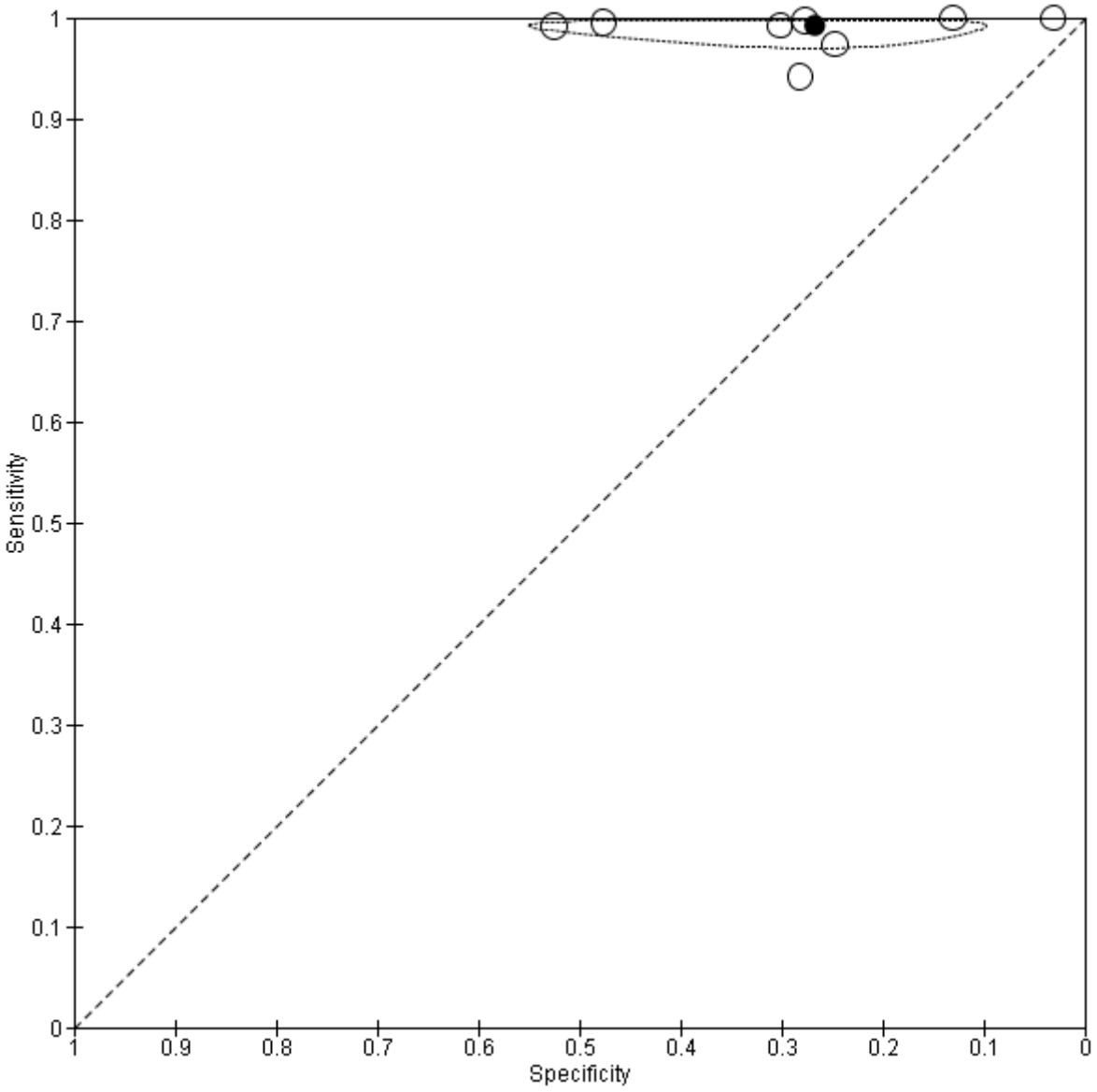
F.1.31 EU TIRADS 5



F.1.32 Kwak TIRADS 3 or more

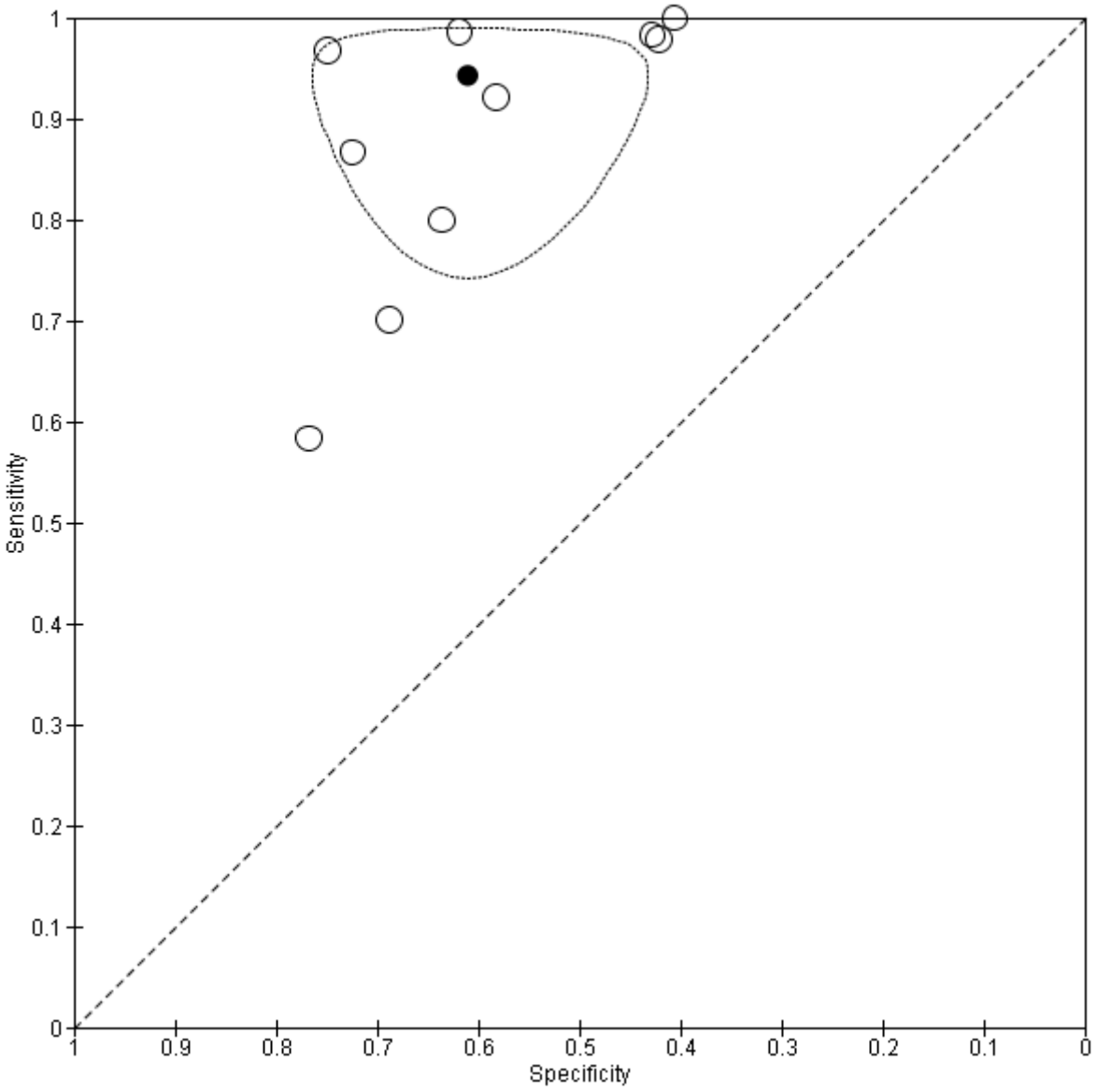


F.1.33 Kwak TIRADS 4a or more

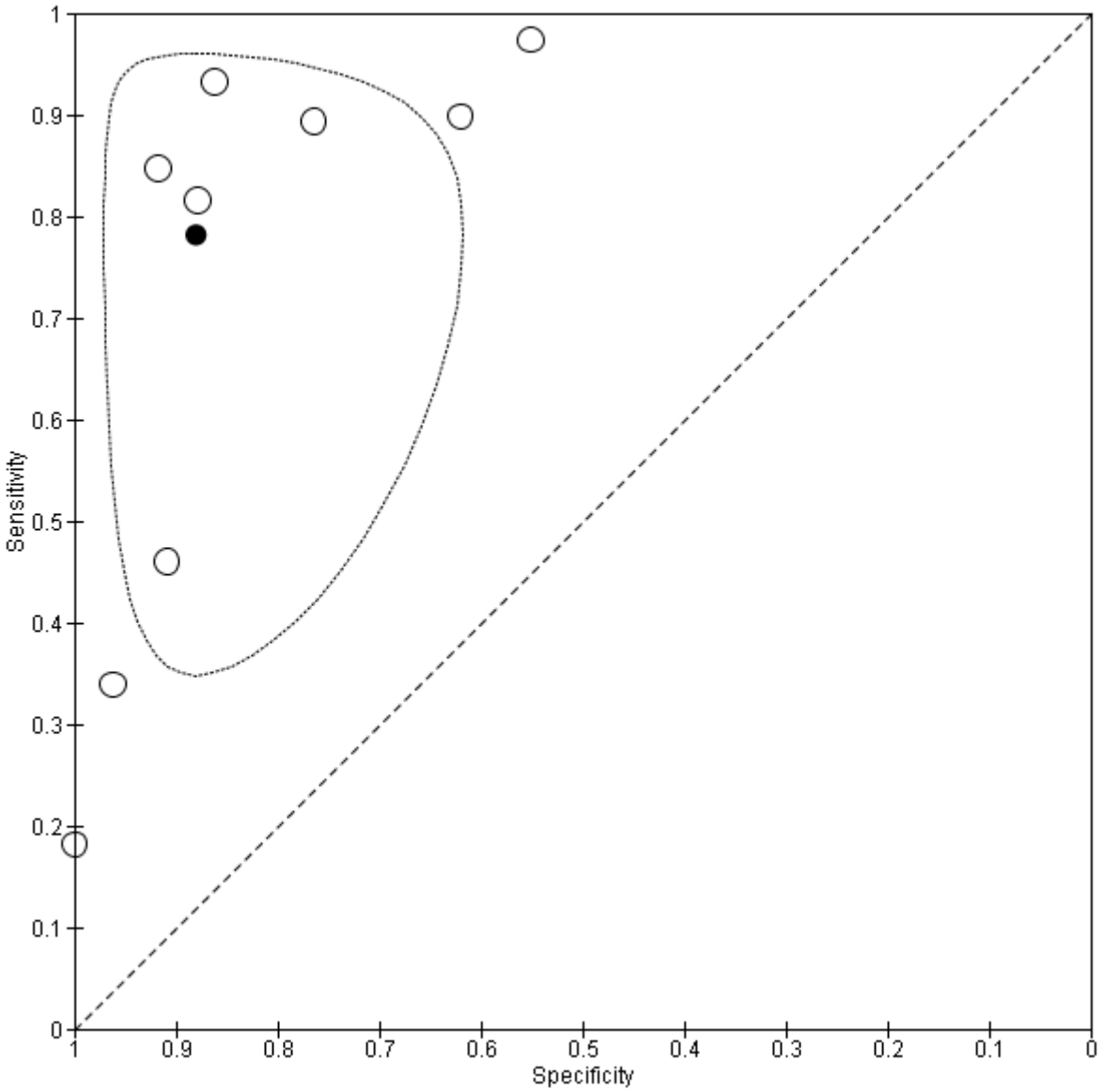


2

F.1.34 Kwak TIRADS 4b or more



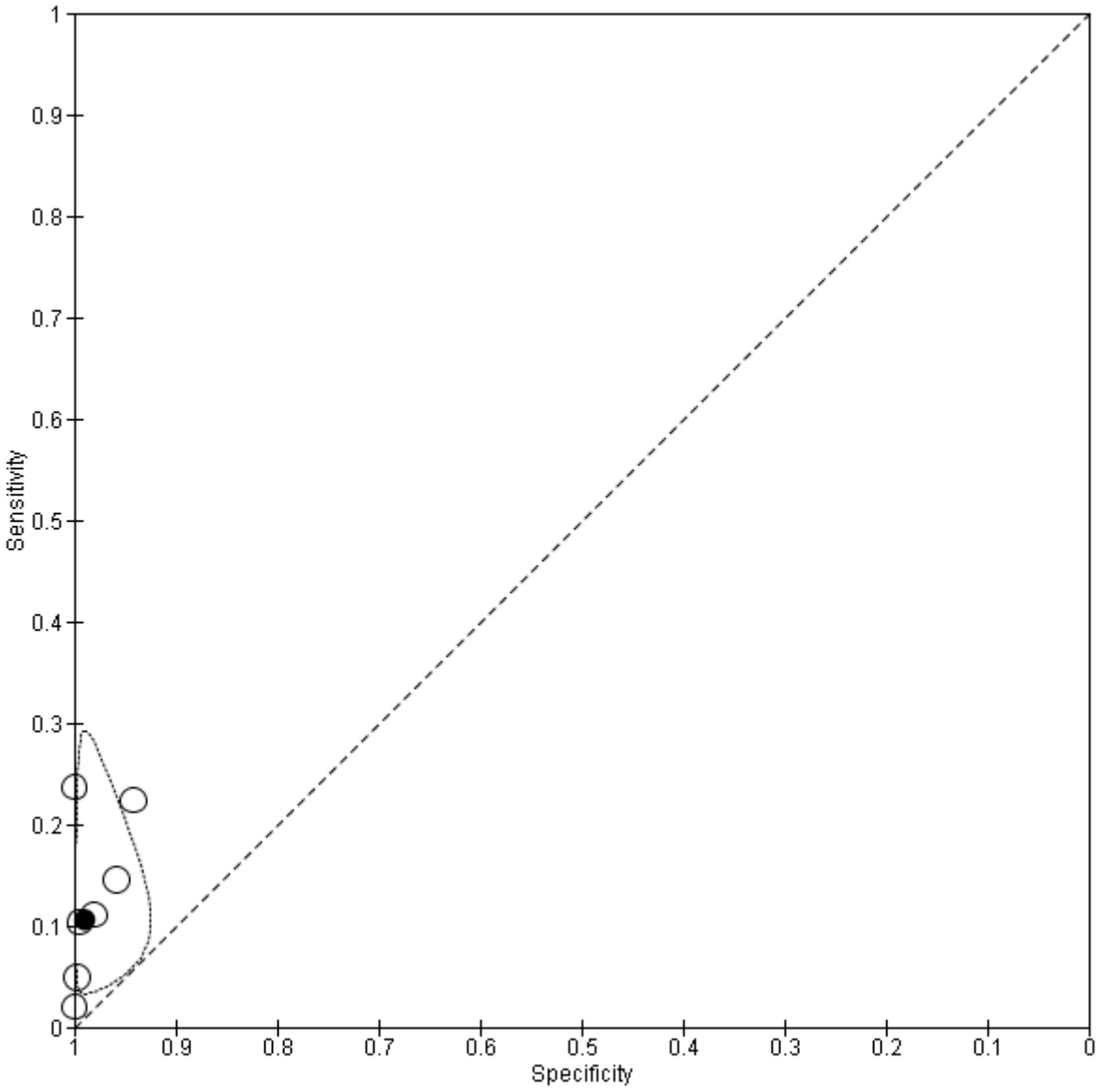
F.4.35 Kwak TIRADS 4c or more



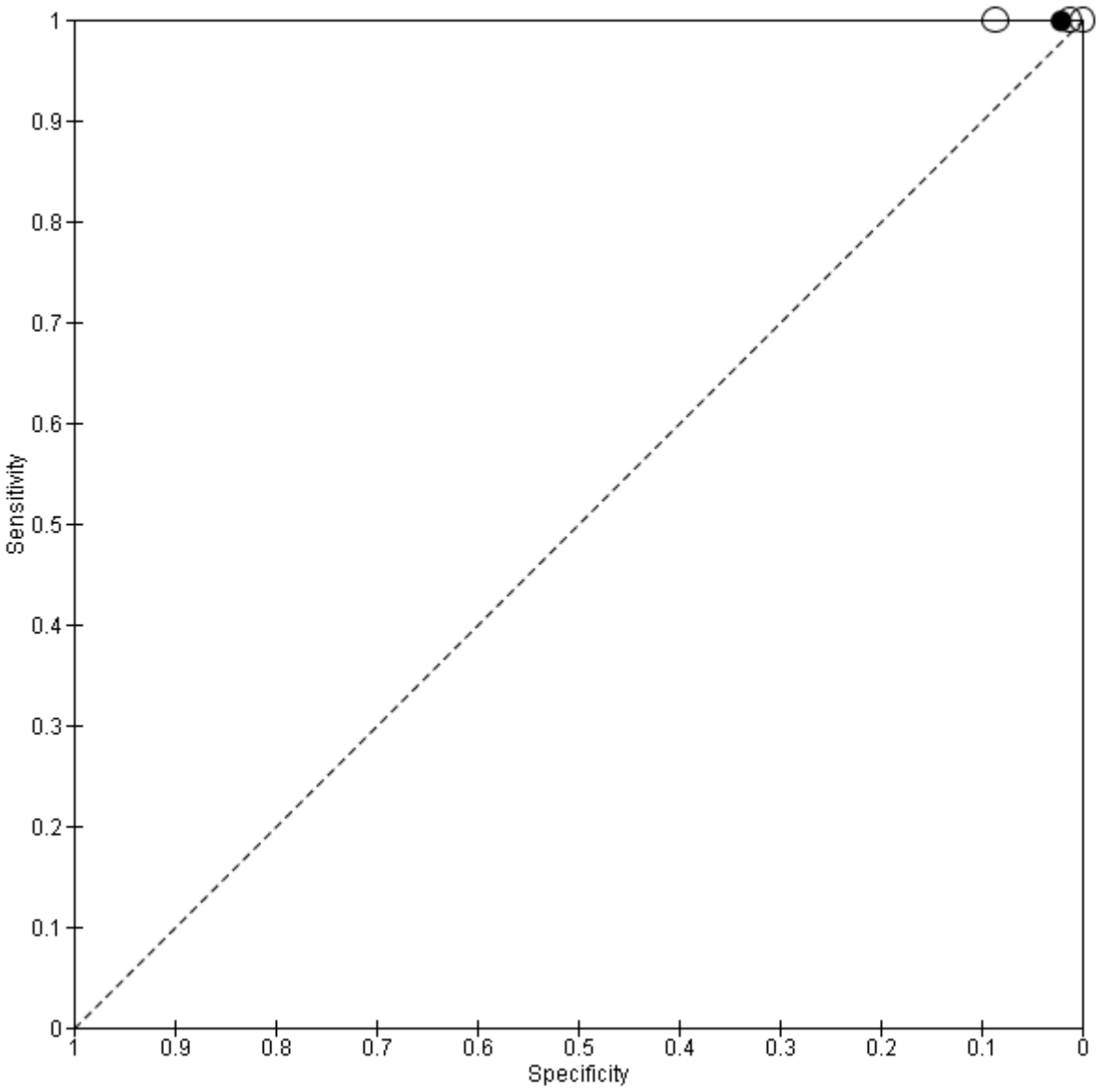
1

2

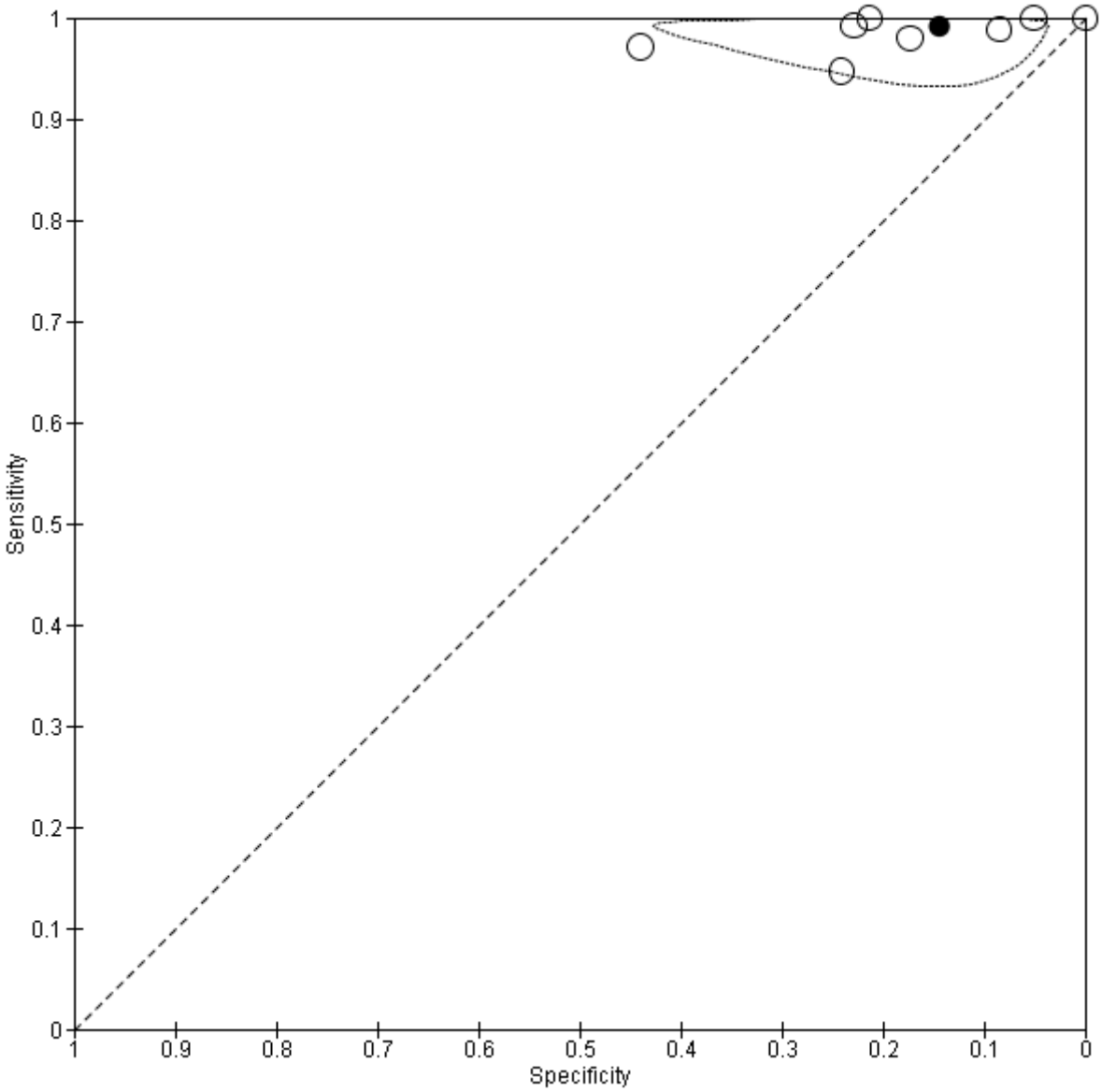
F.1.36 Kwak TIRADS 5



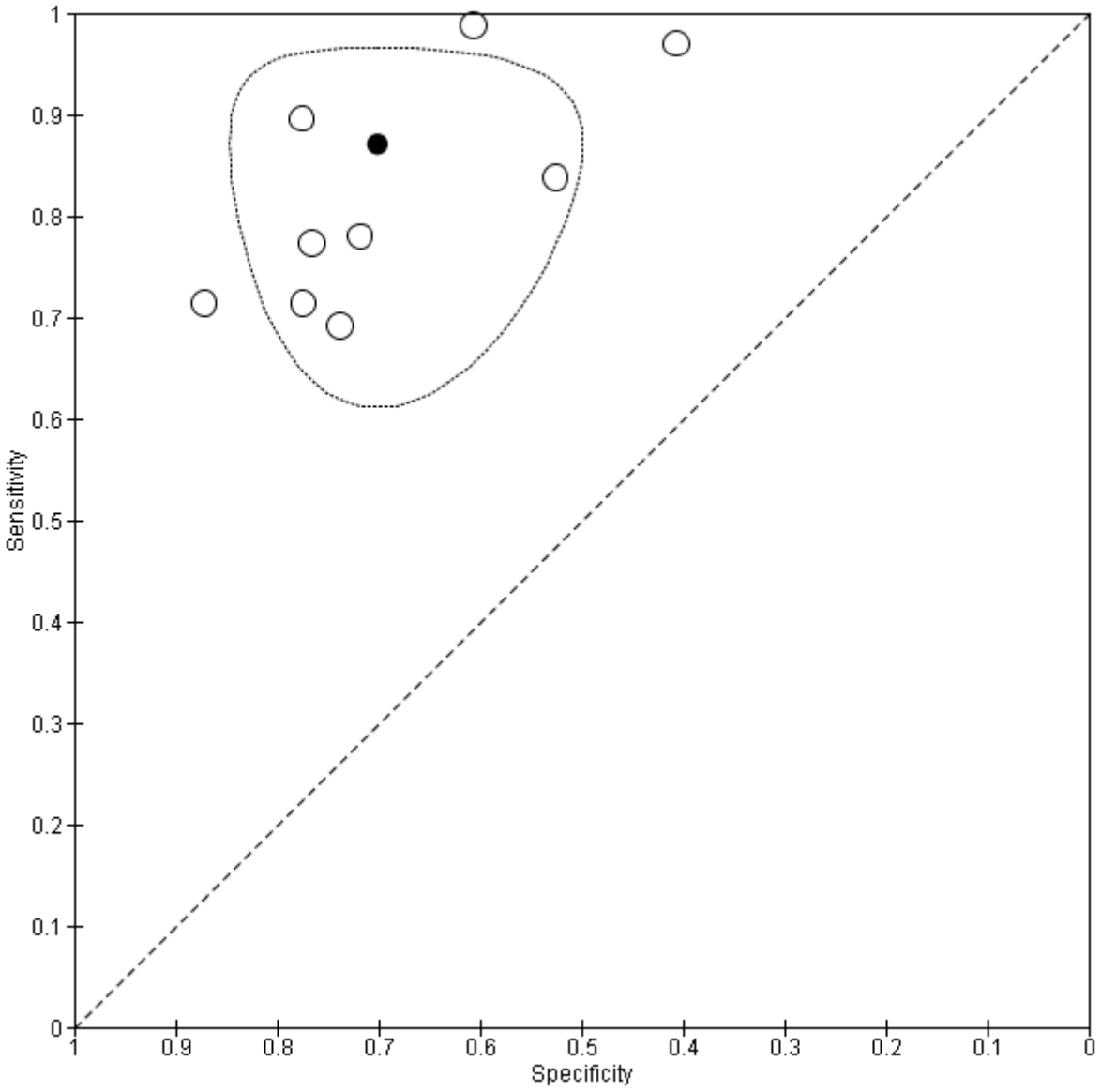
F.1.37 ATA ‘very low suspicion’ or more



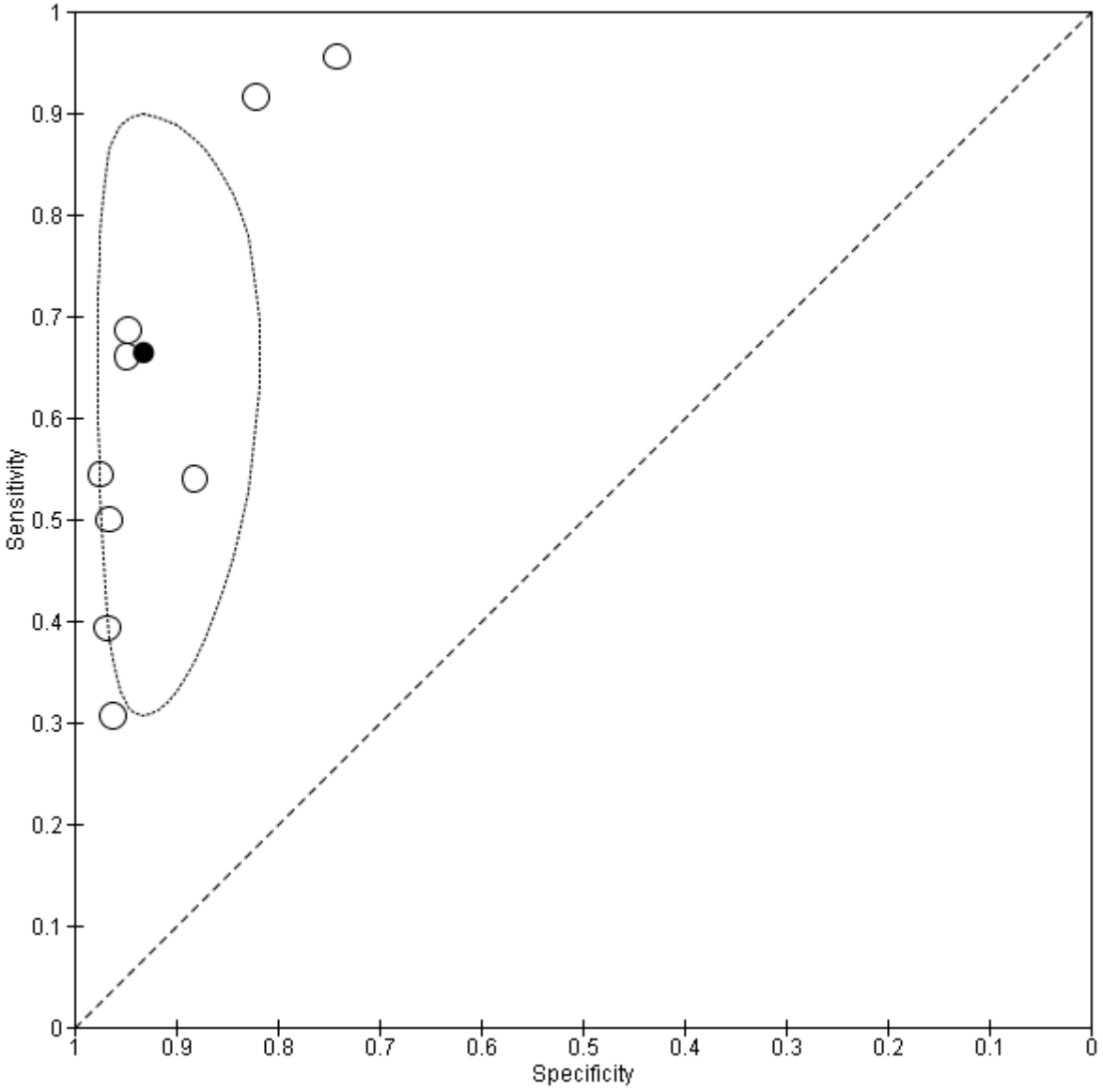
F.1.38 ATA ‘low suspicion’ or more



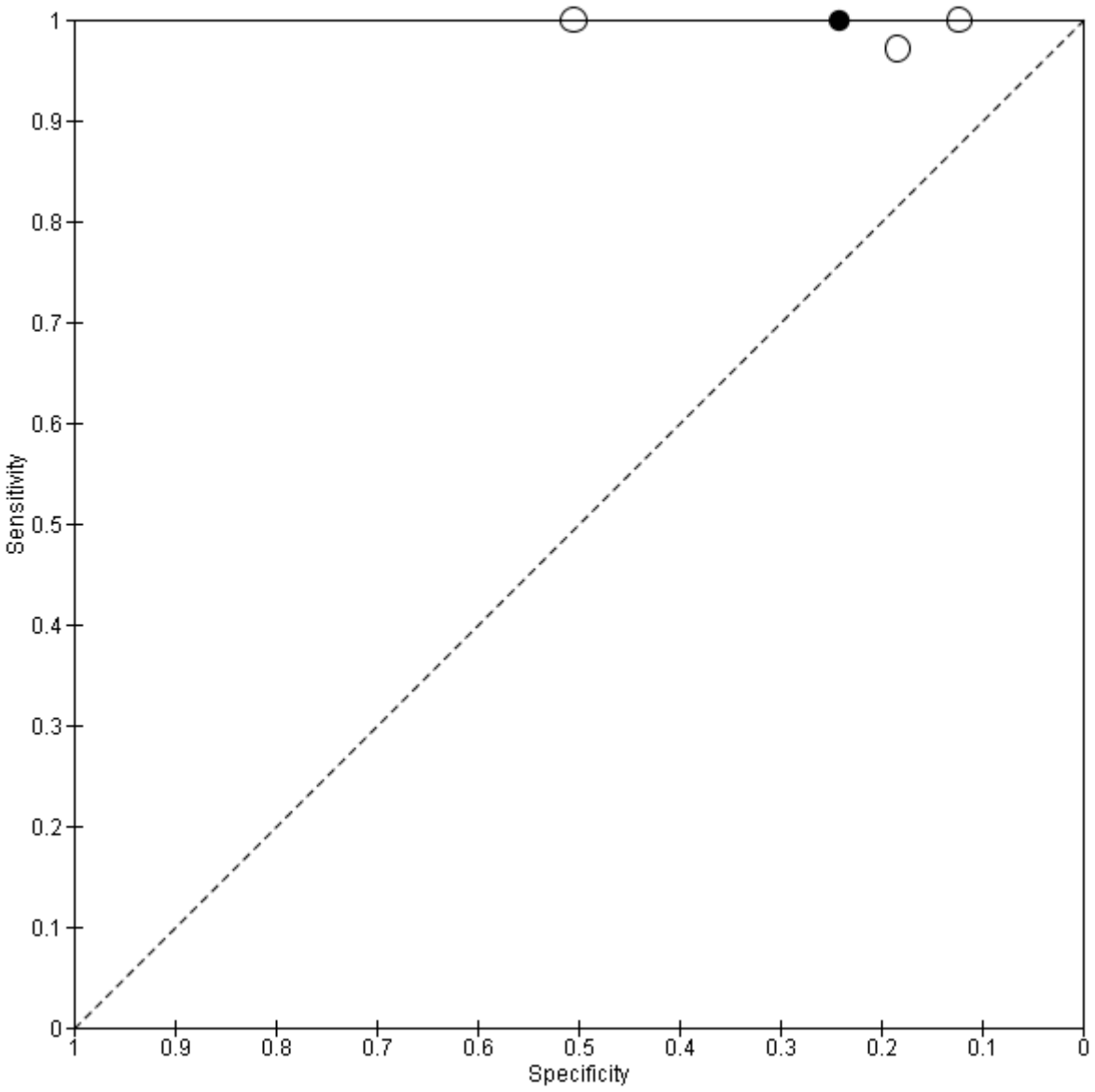
F.1.39 ATA 'intermediate suspicion' or more



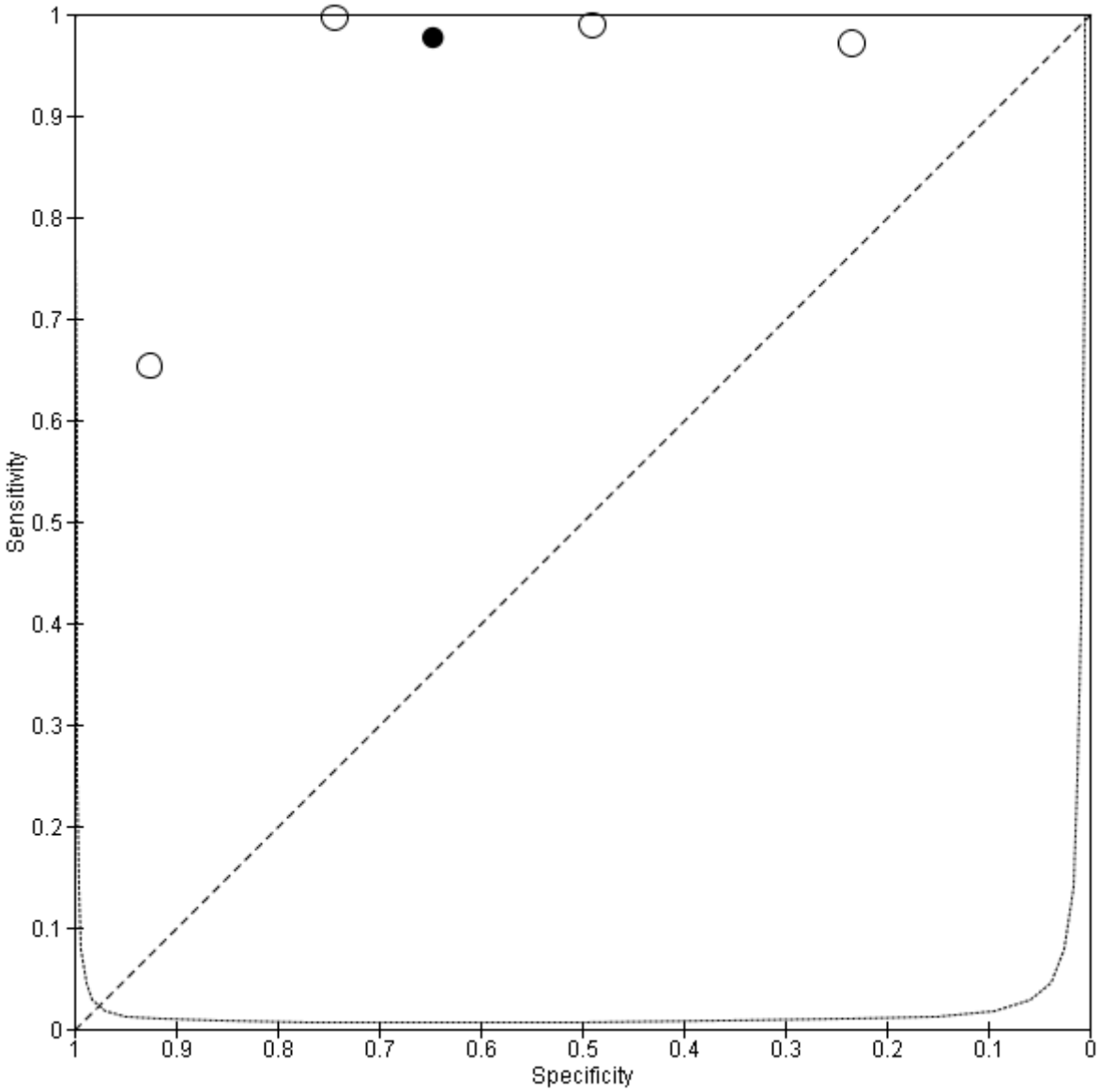
F.1.40 ATA high suspicion



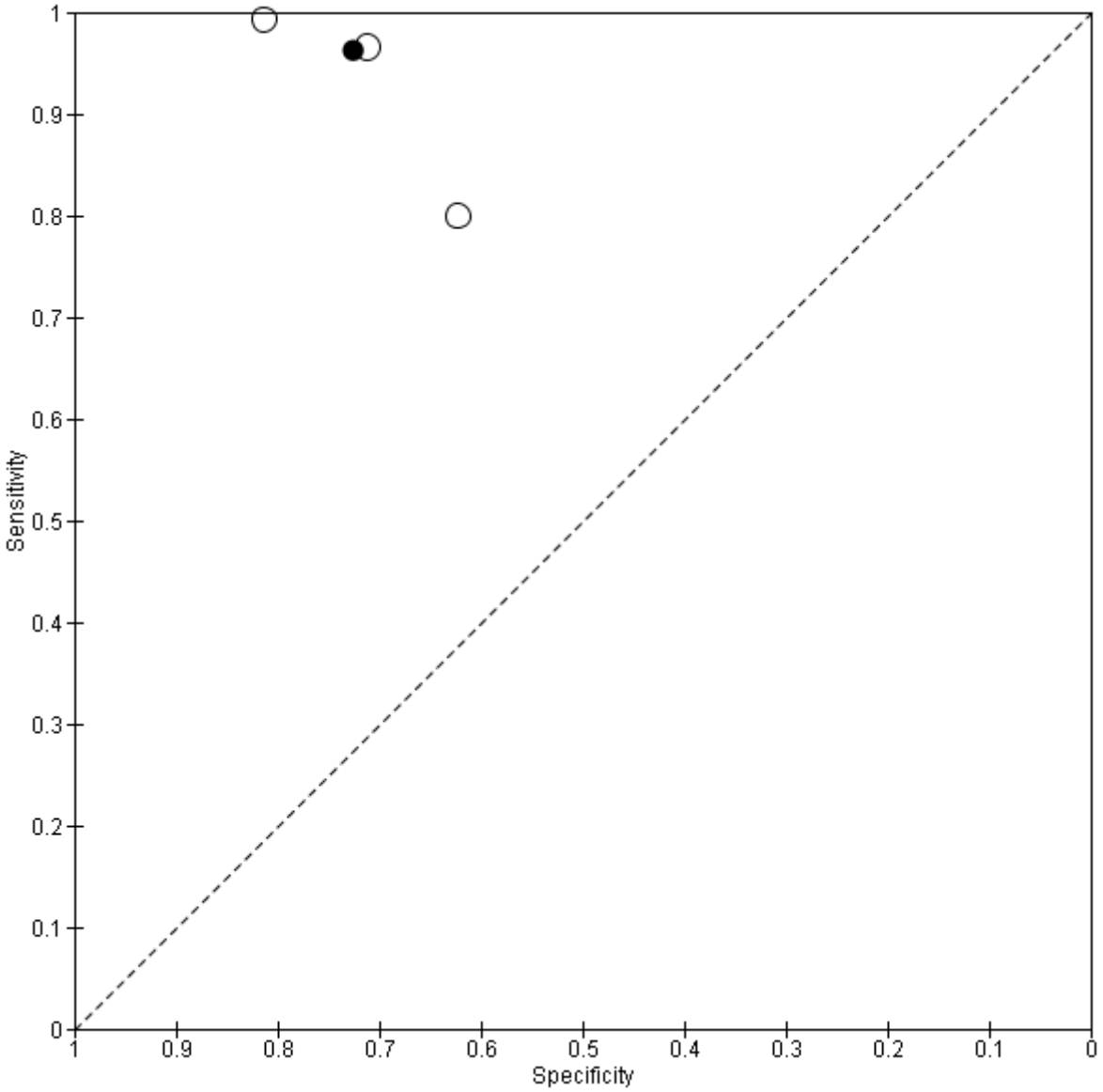
F.1.41 Horvath TIRADS 3 or more



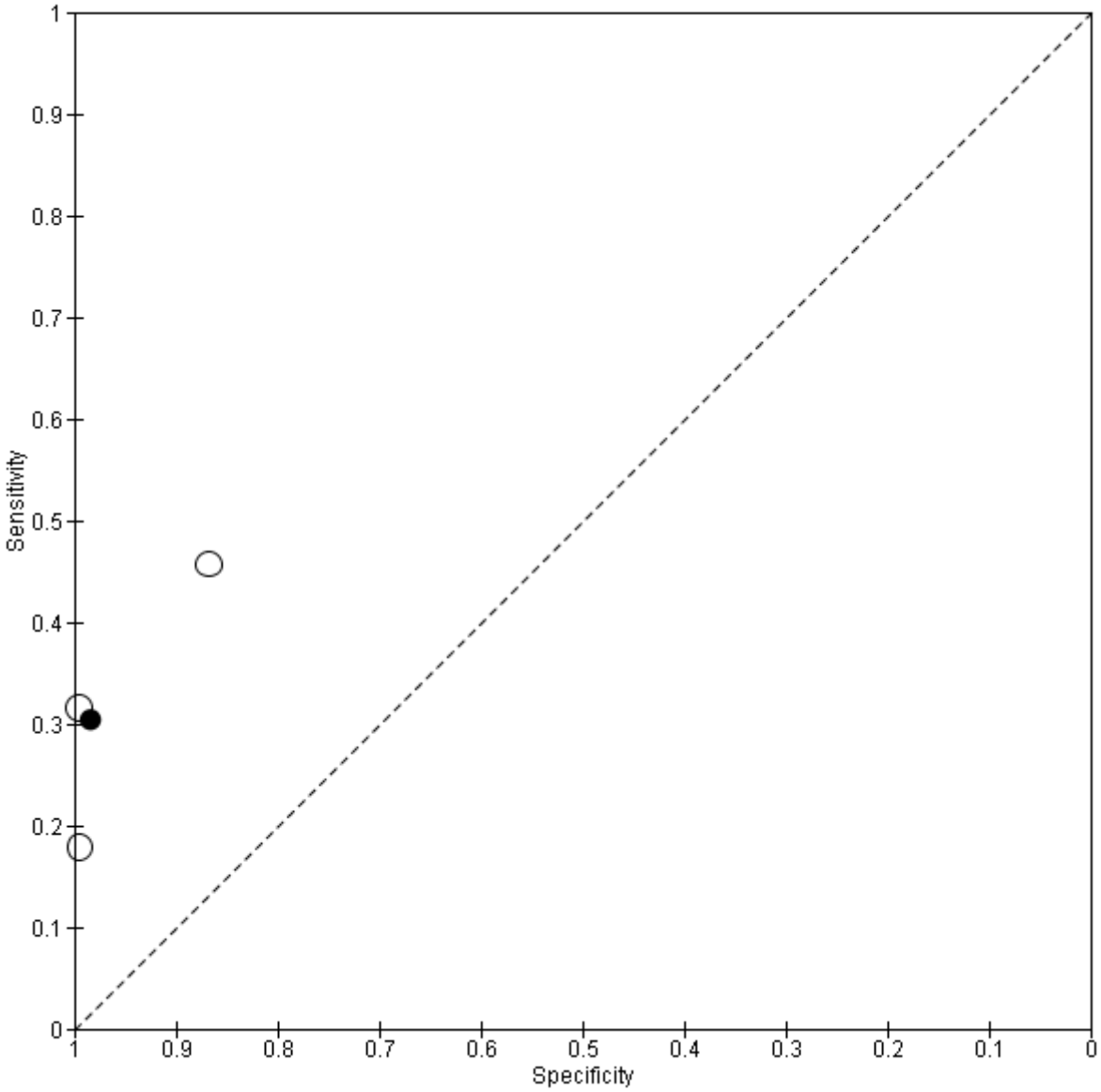
F.1.42 Horvath TIRADS 4a or more



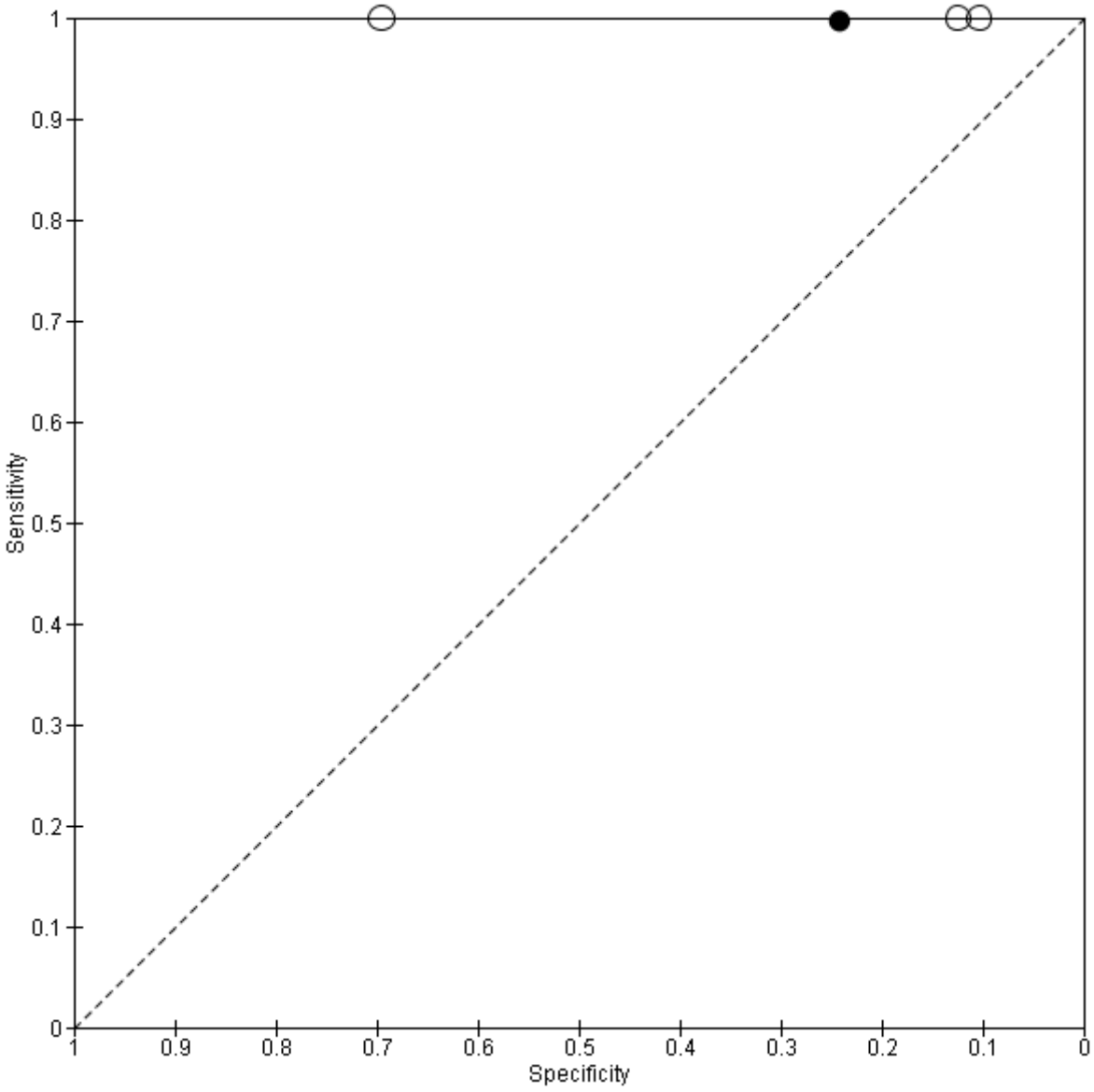
F.1.43 Horvath TIRADS 4b or more



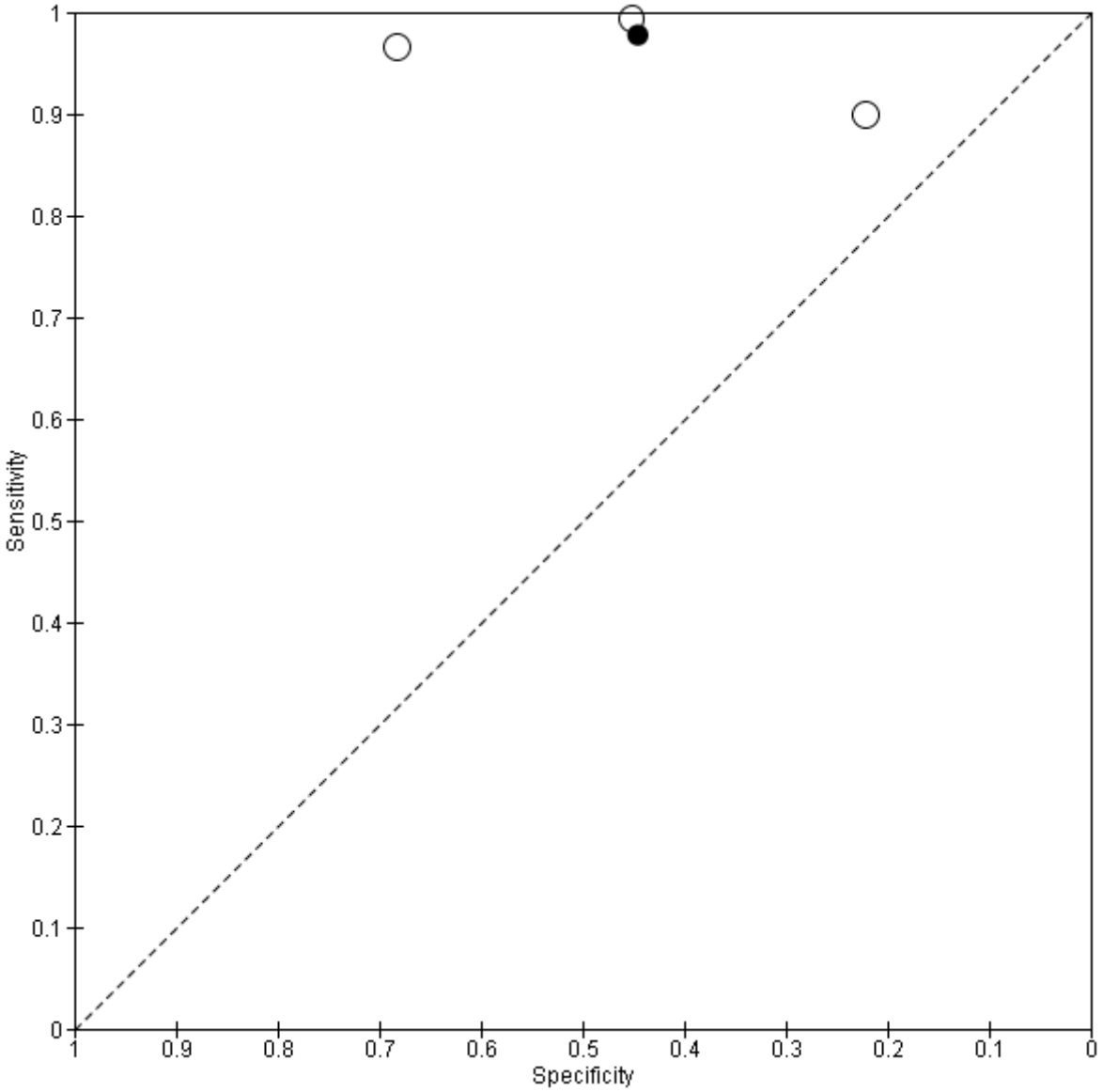
F.1.44 Horvath TIRADS 5



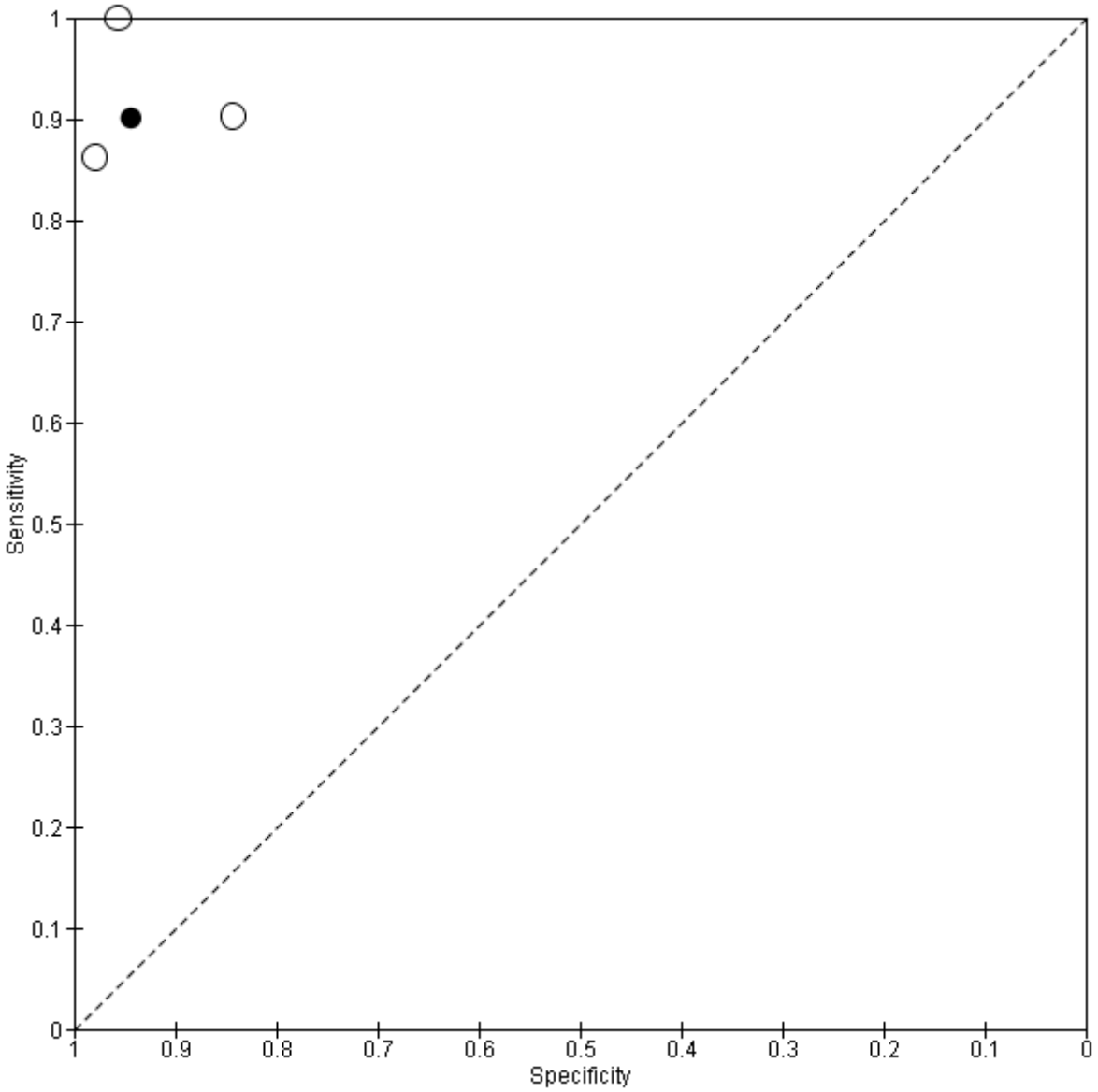
F.2.45 Russ TIRADS 3 and more



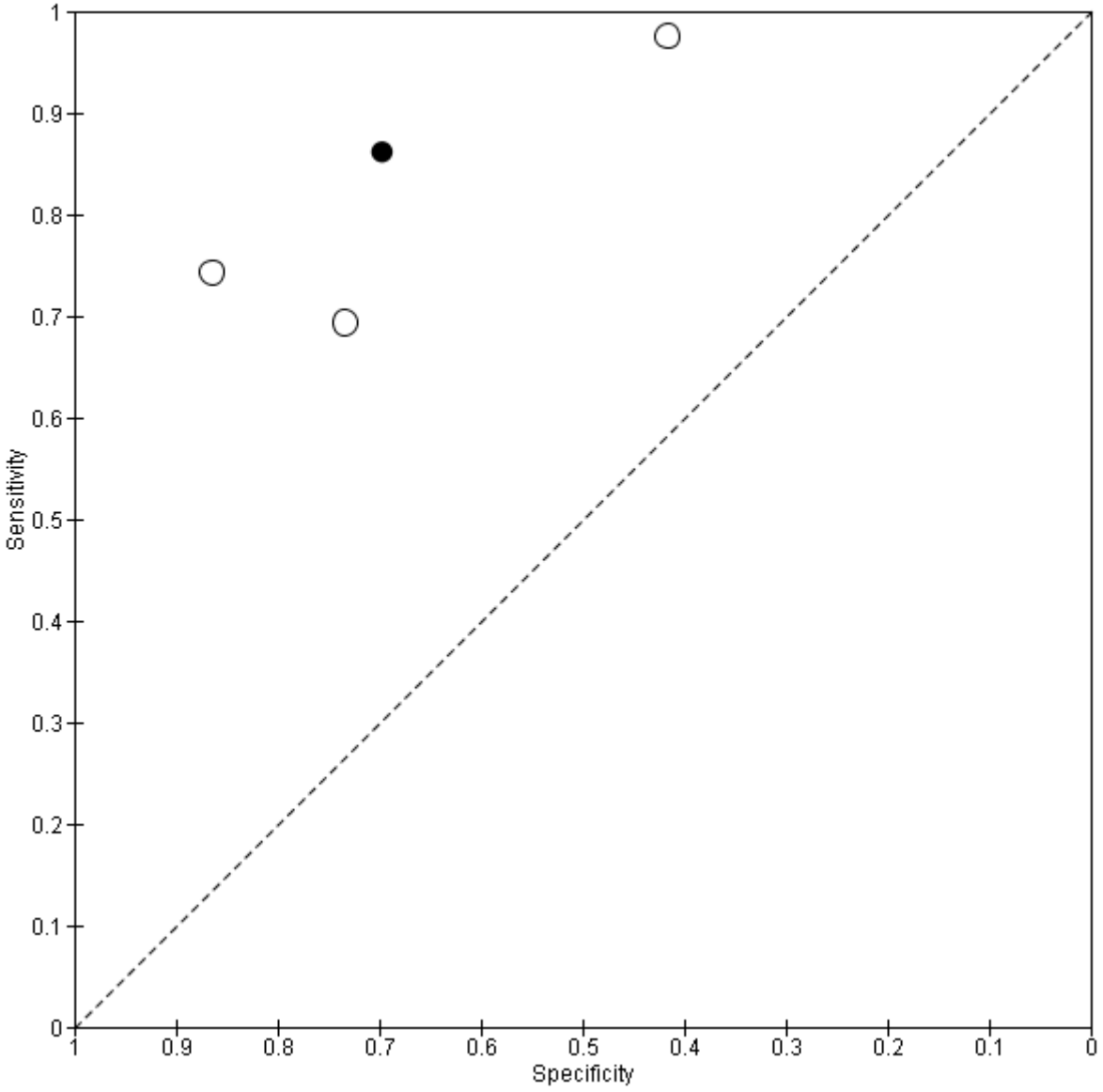
F.1.46 Russ TIRADS 4a and more



F.1.47 Russ TIRADS 4b and more

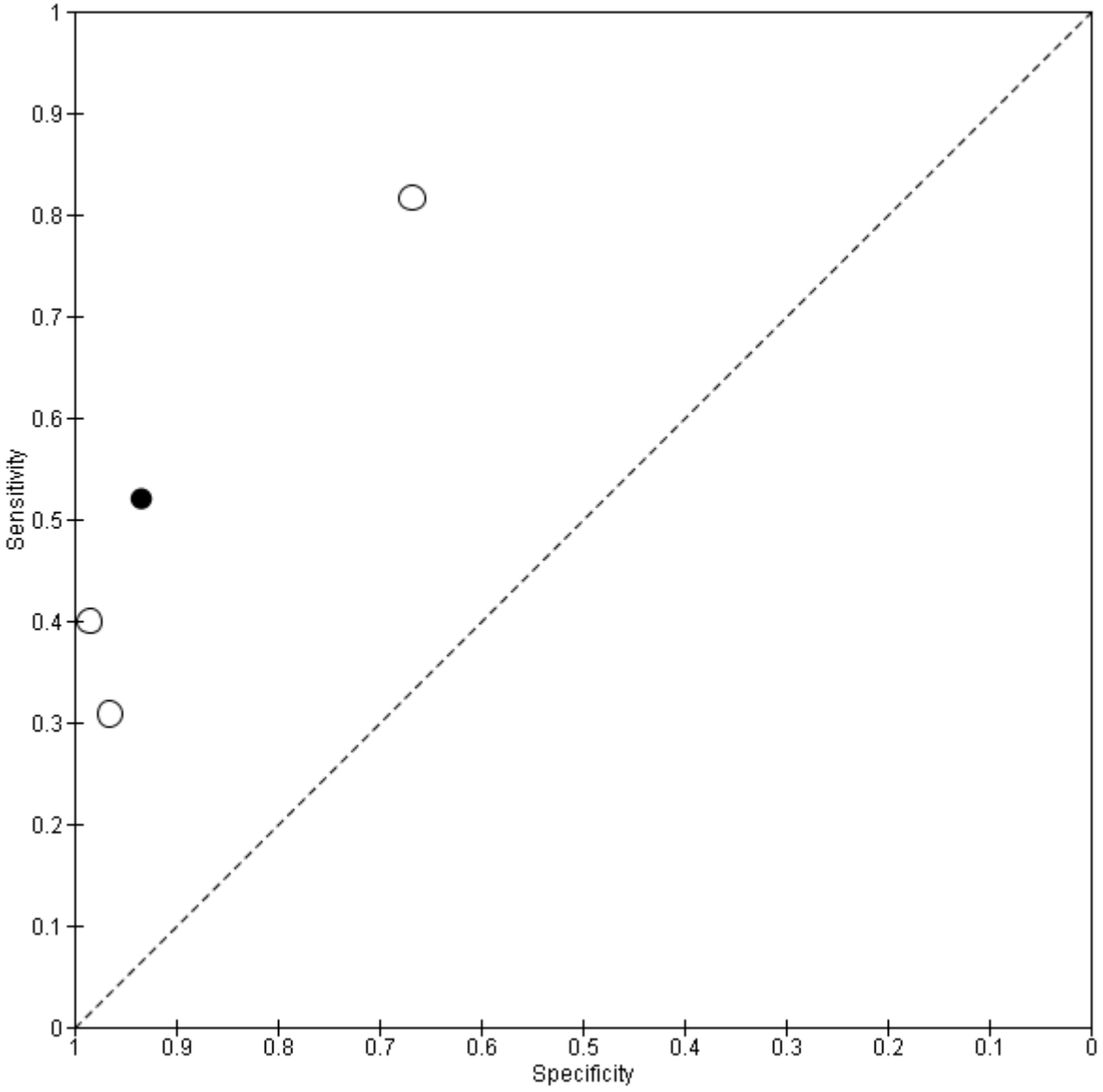


F.1.48 Korean TIRADS 4 and above



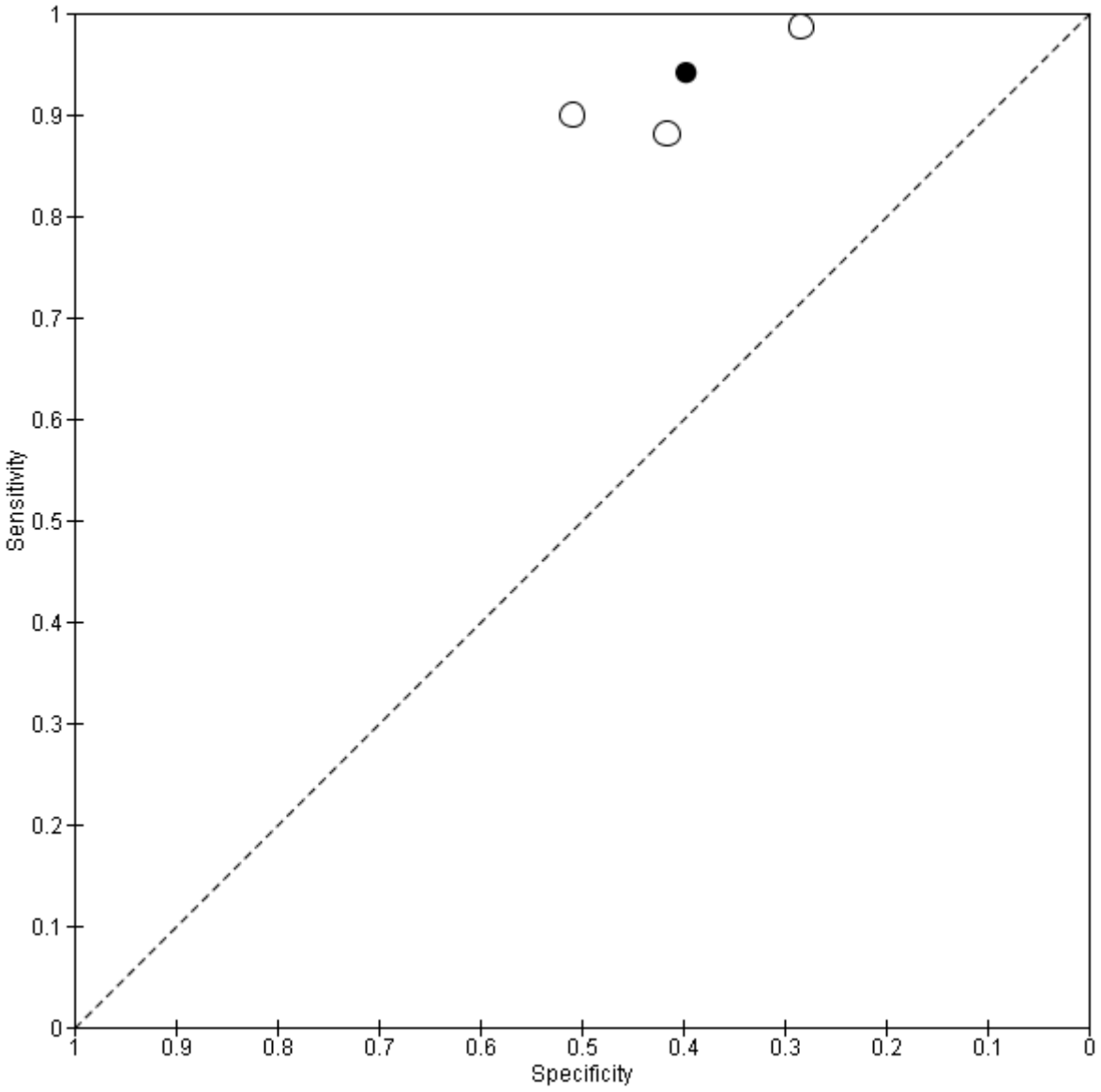
2

F.1.49 Korean TIRADS 5 and above

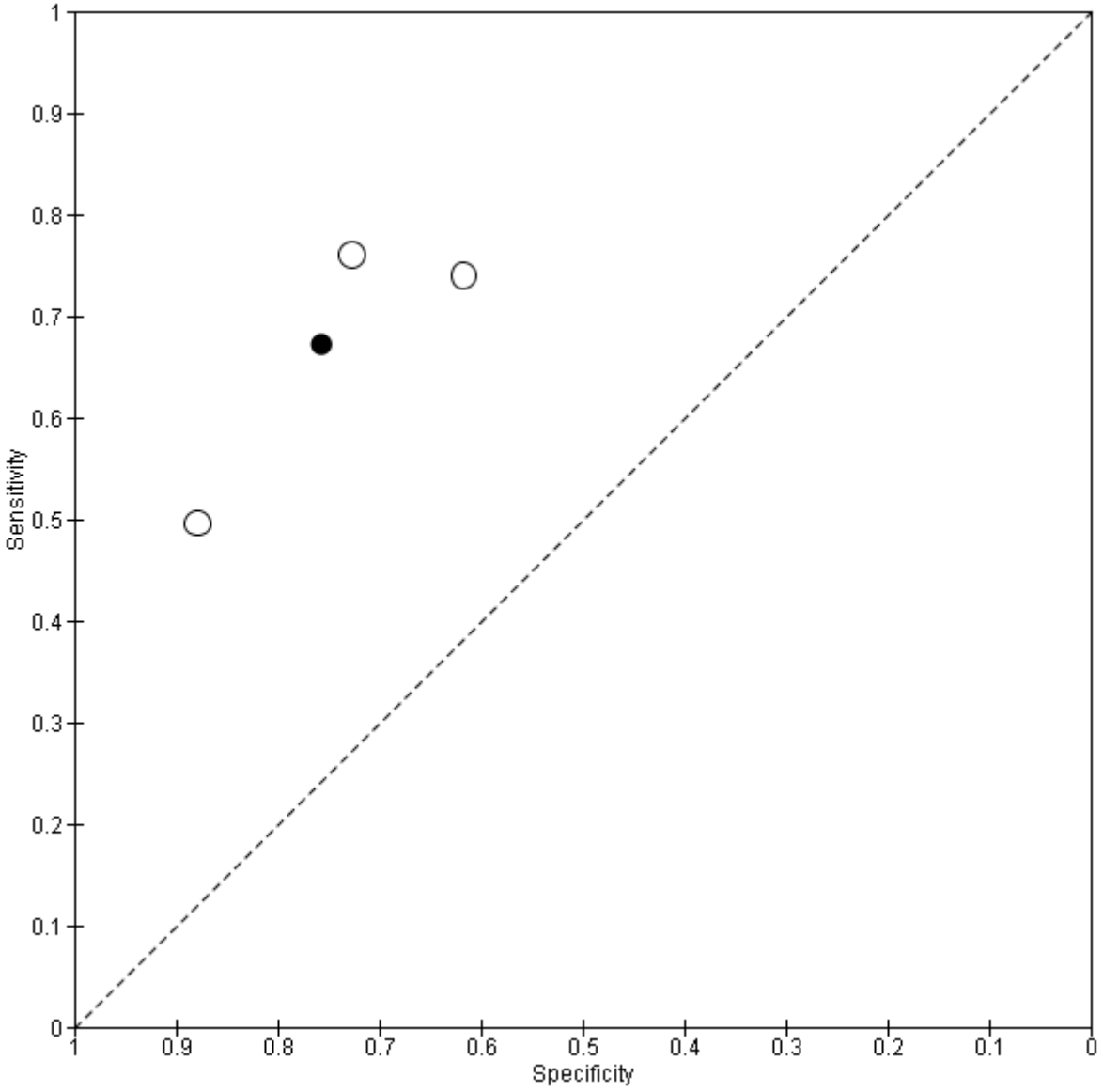


2

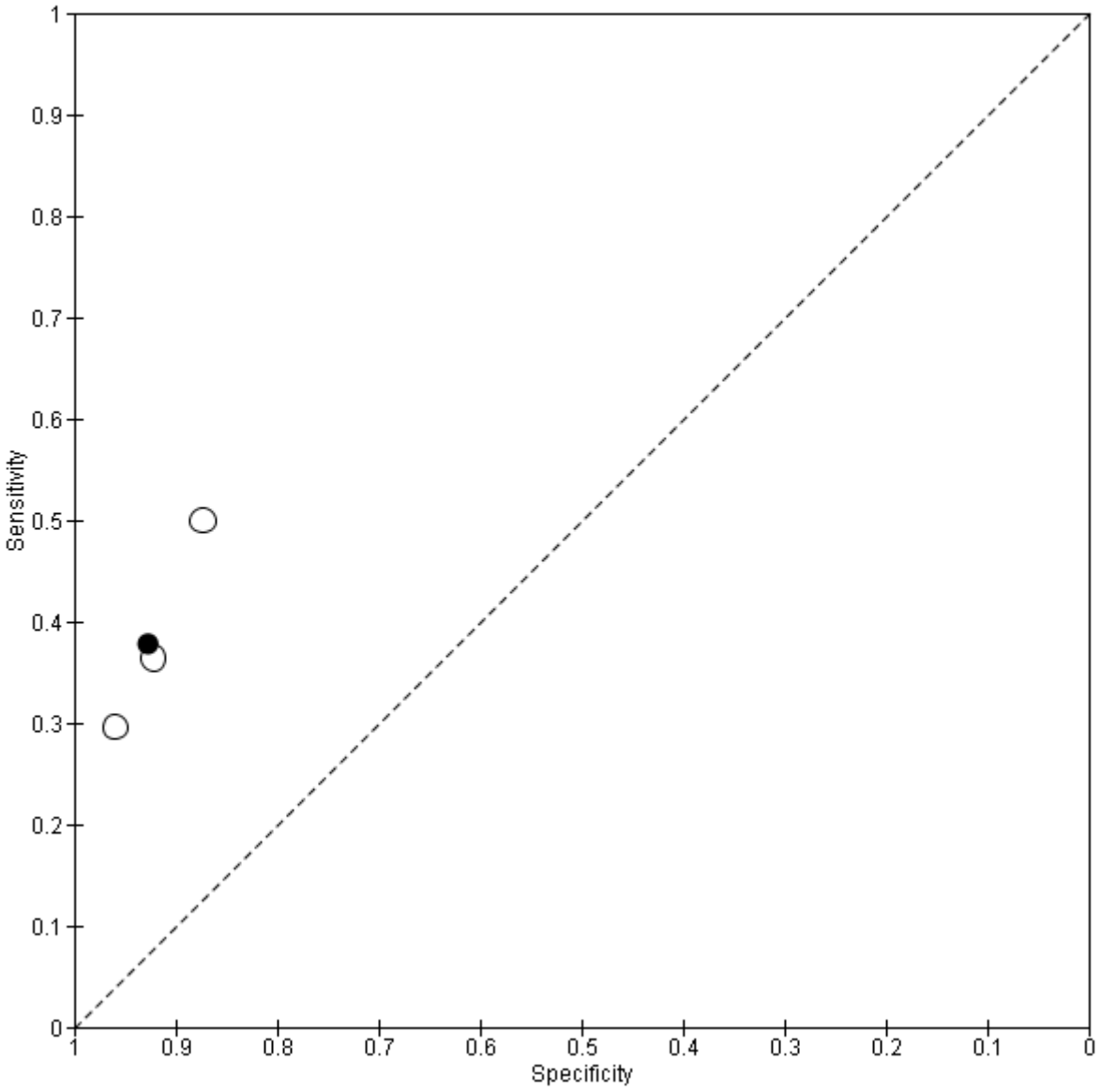
F.1.50 BTA 3 and above



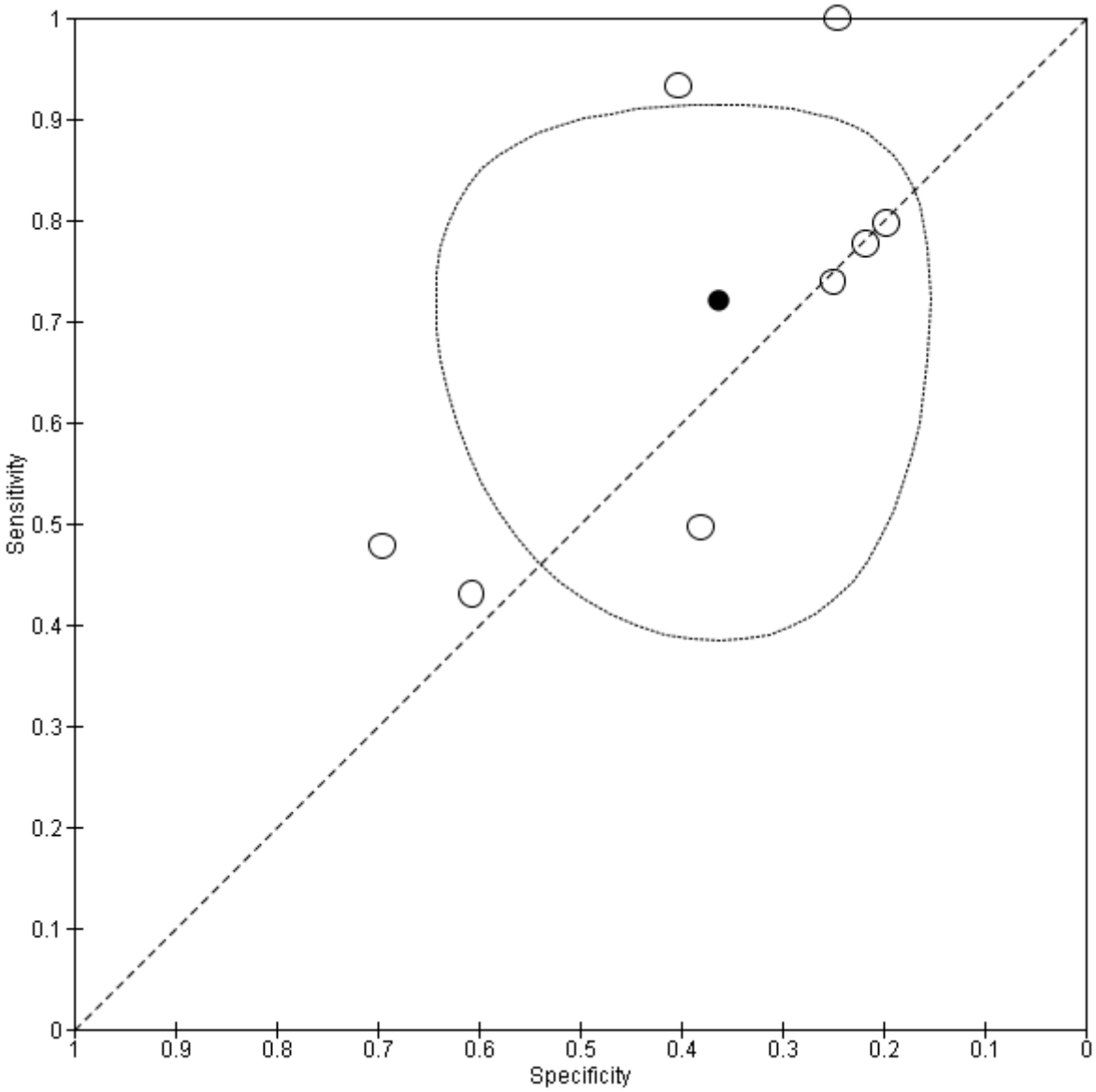
F.1.51 BTA 4 and above



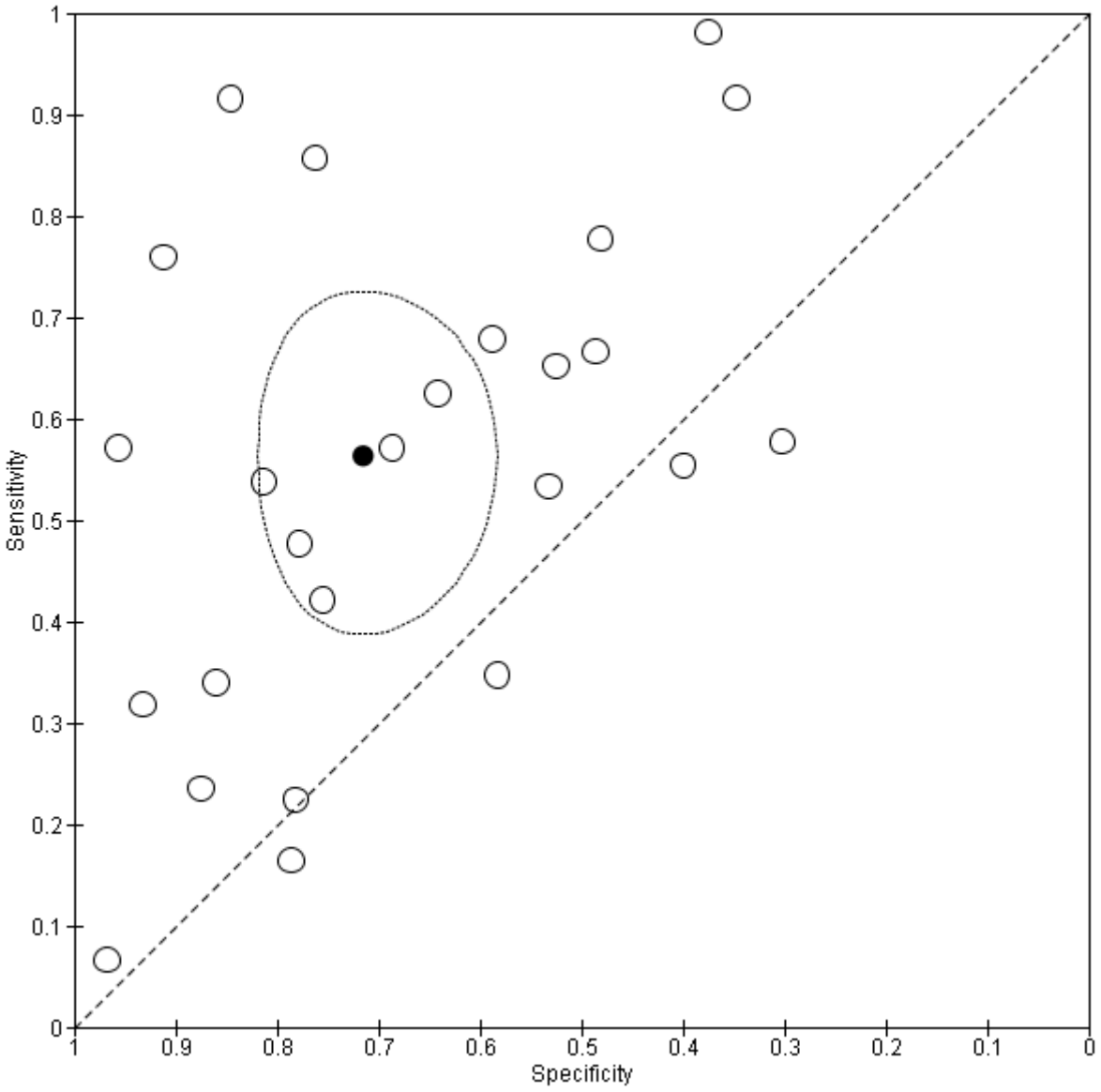
F.1.52 BTA 5



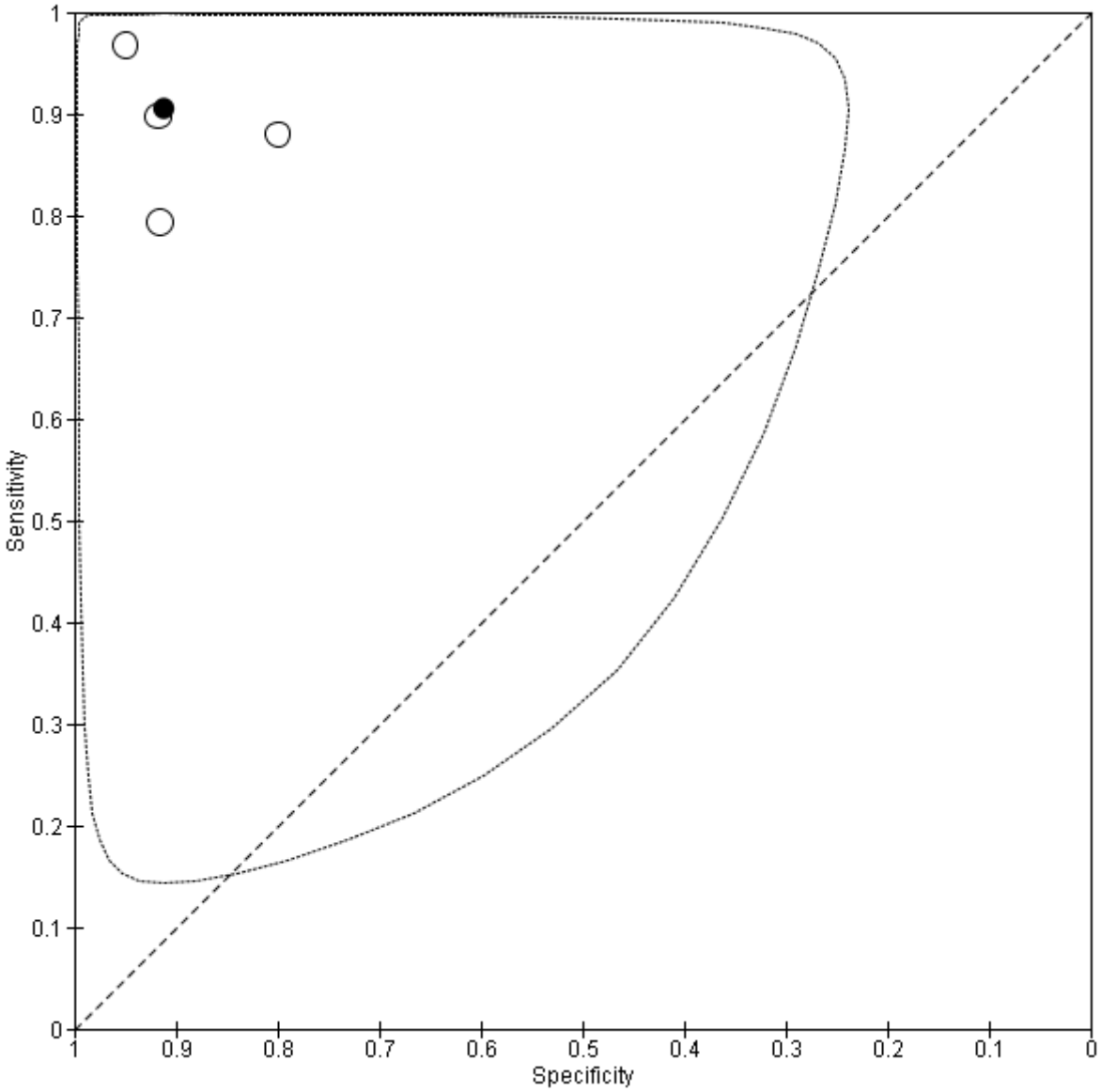
F.1.53 Any blood flow



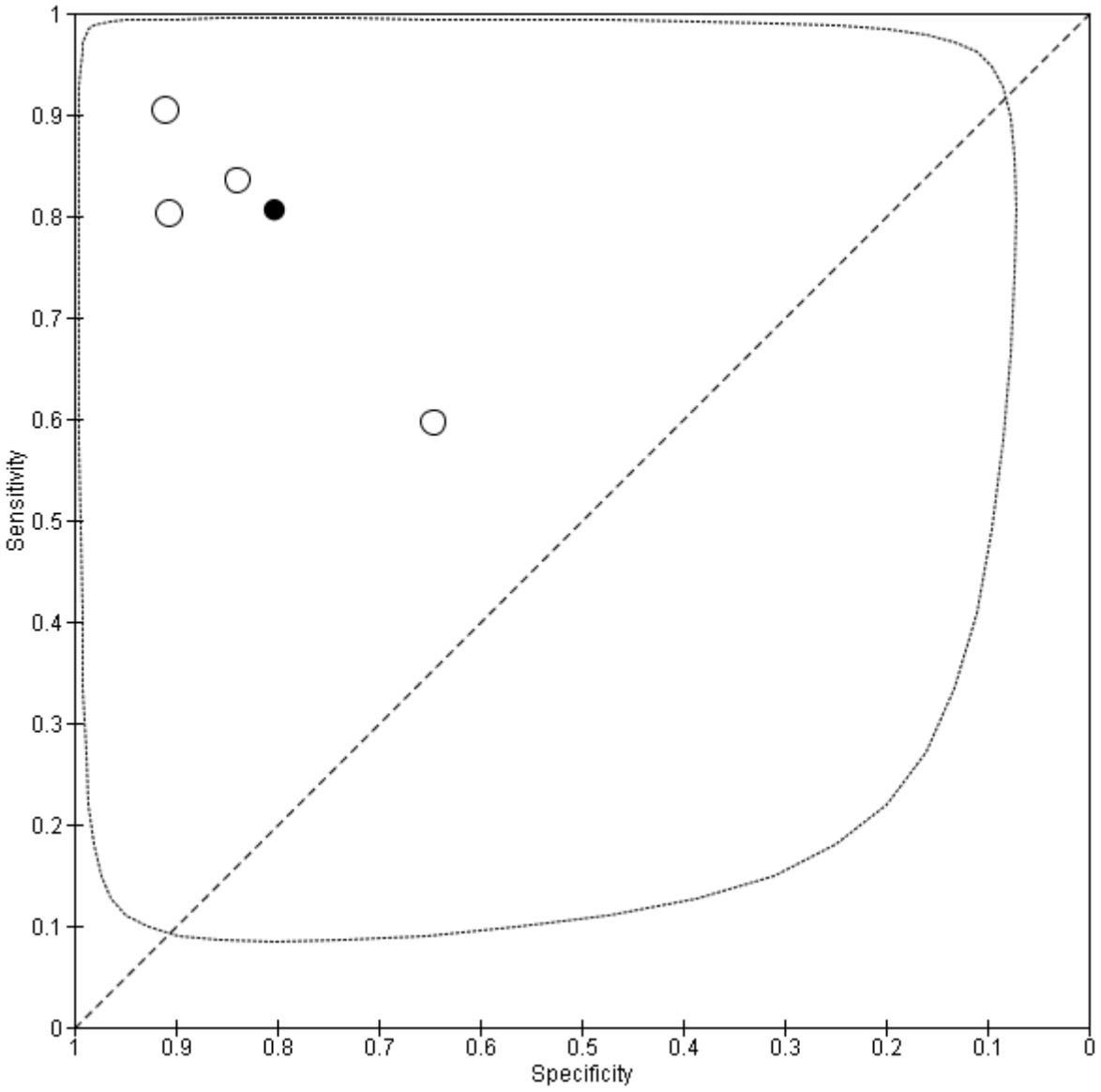
F.1.54 Central blood flow



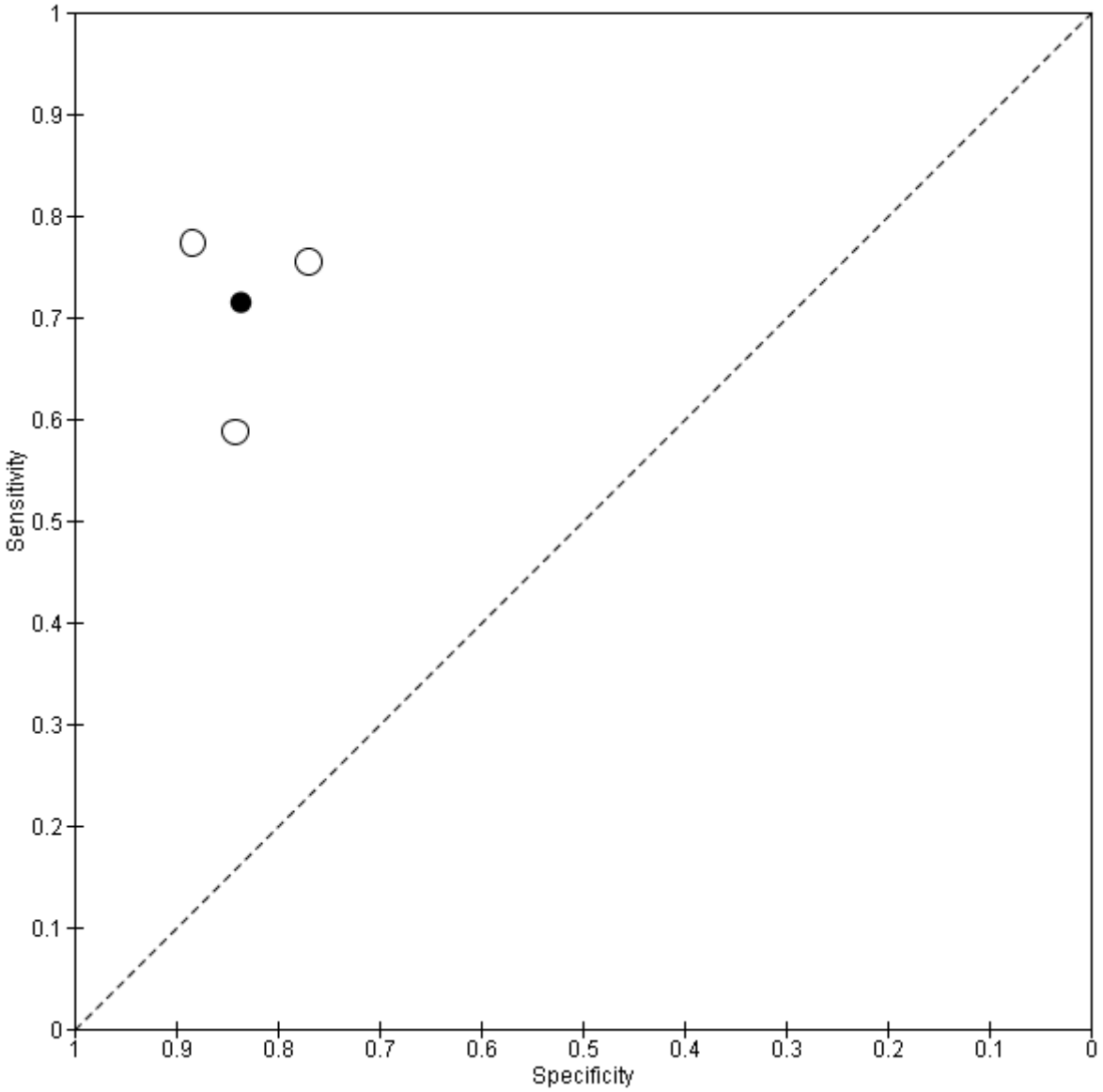
F.1.55 CEUS heterogeneous enhancement AND low enhancement pattern



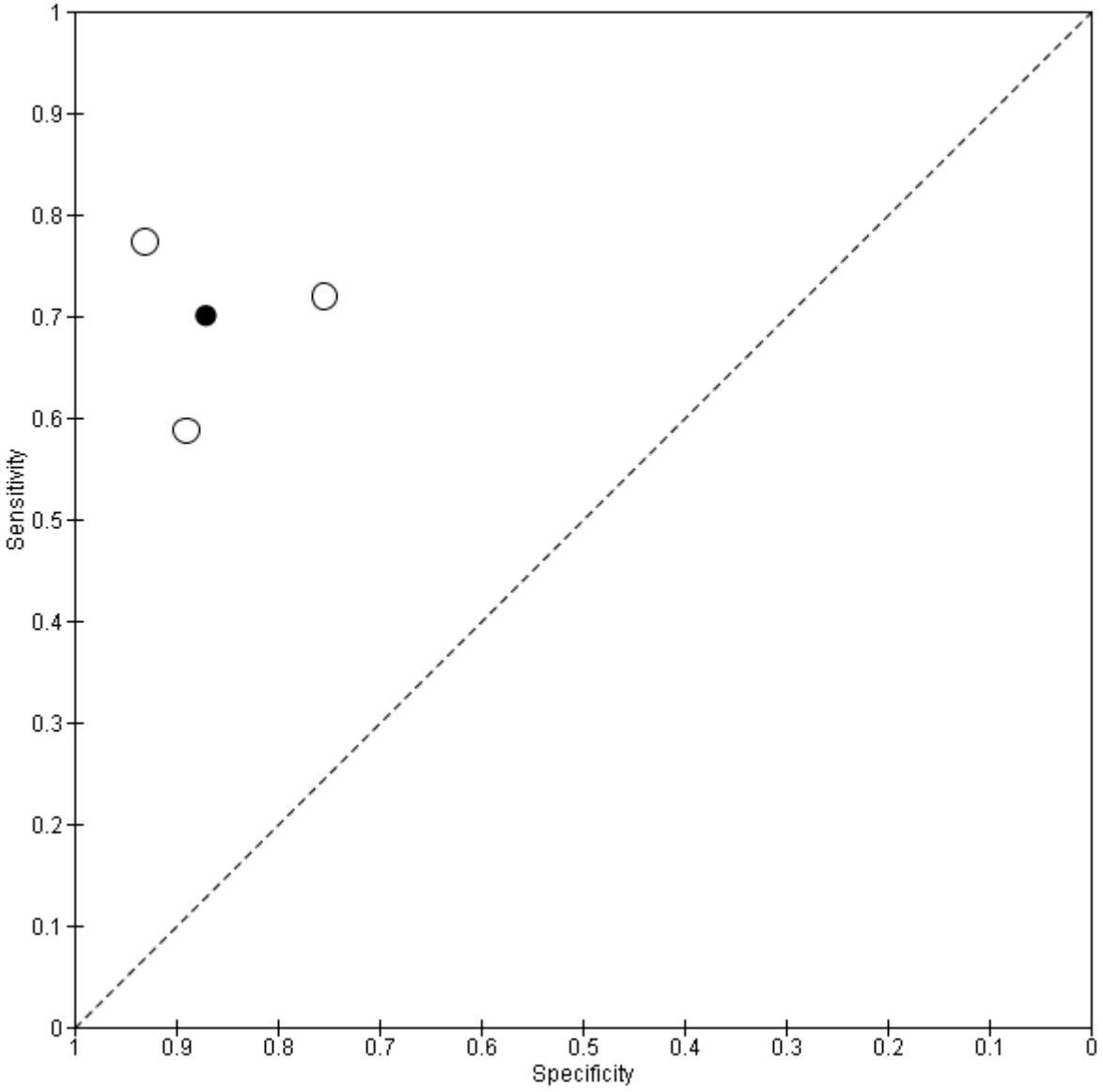
F.1.56 CEUS heterogeneous enhancement pattern



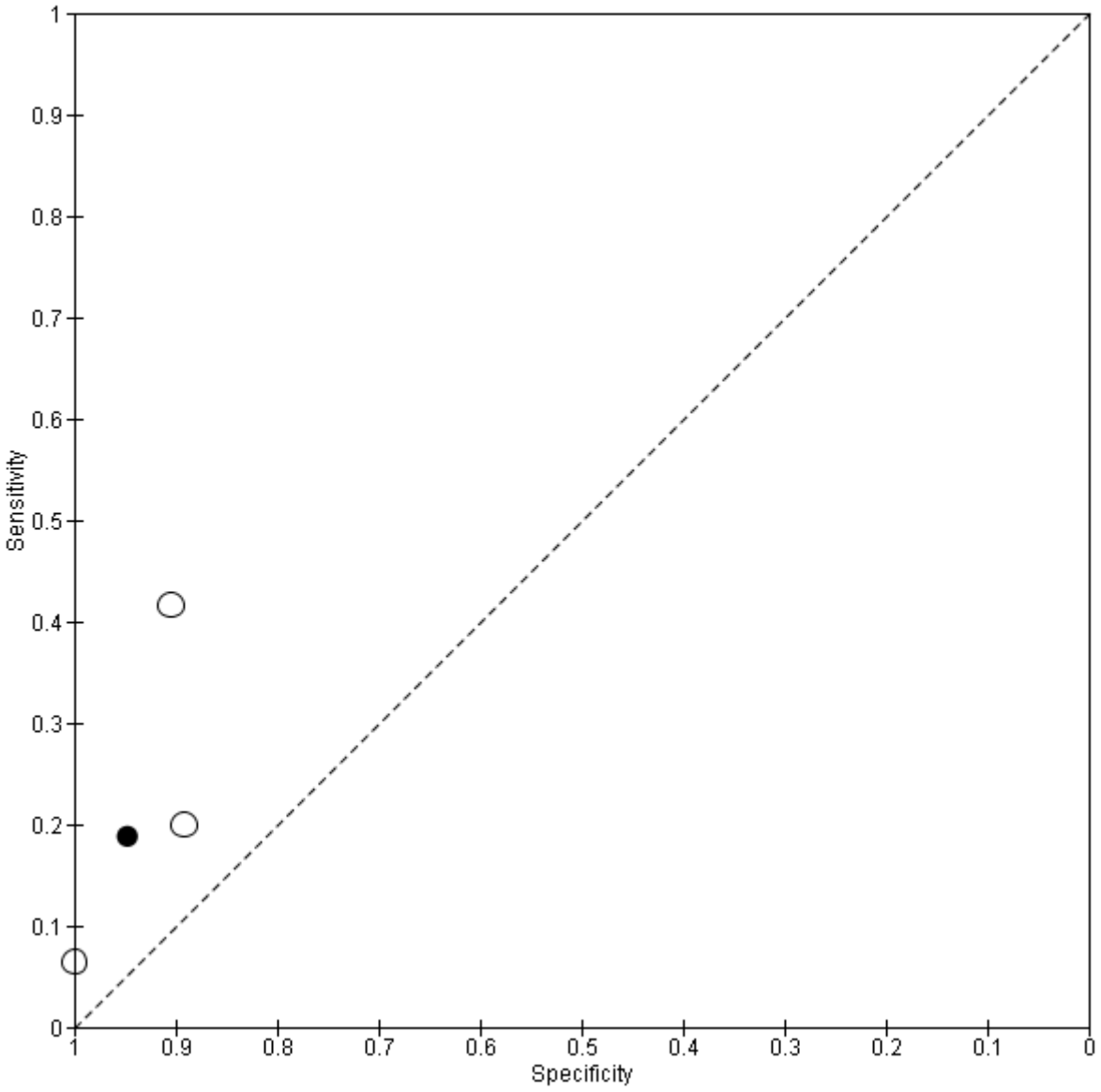
F.1.57 CEUS irregular shape



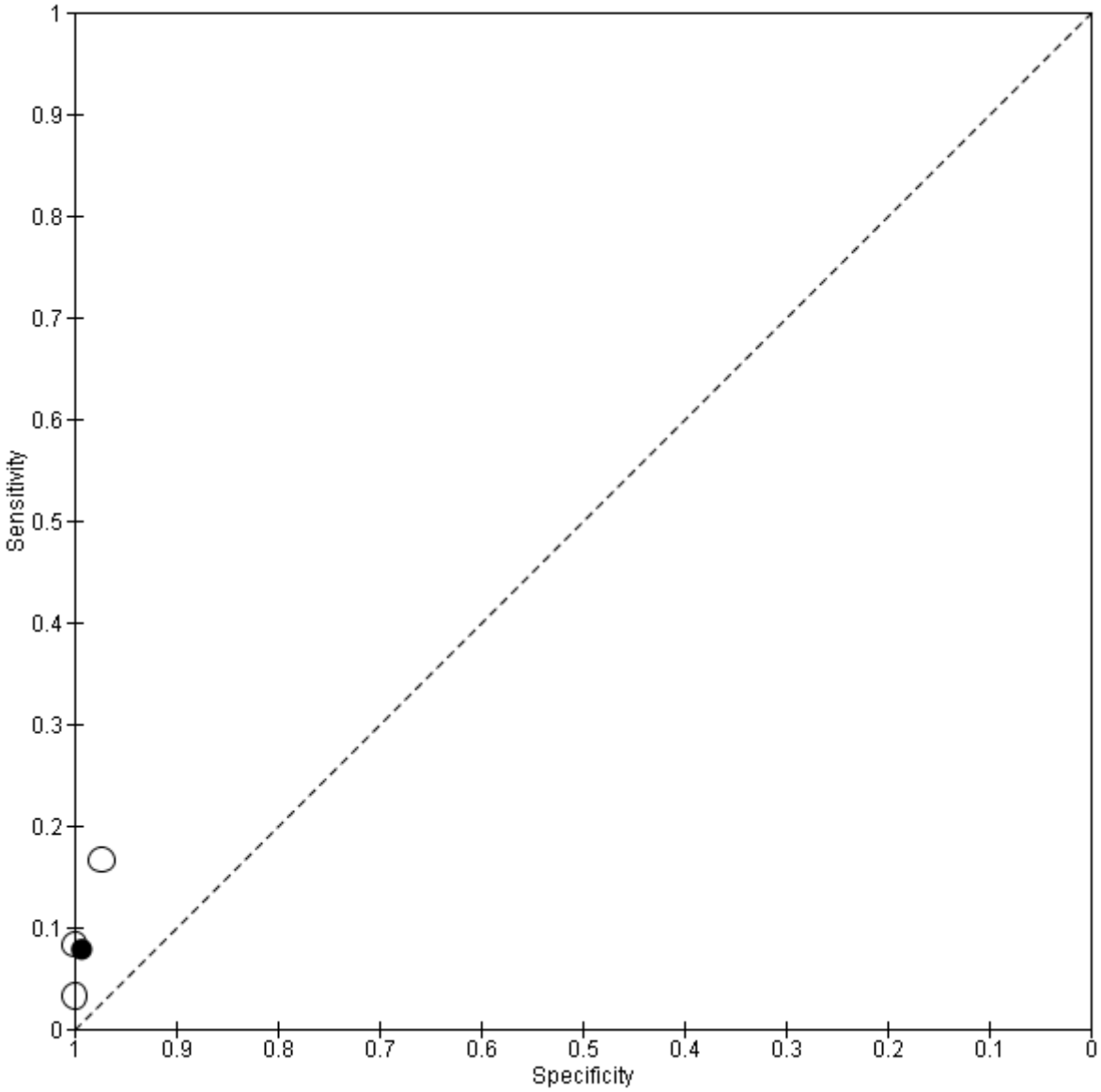
F.1.58 CEUS ill defined enhancement border

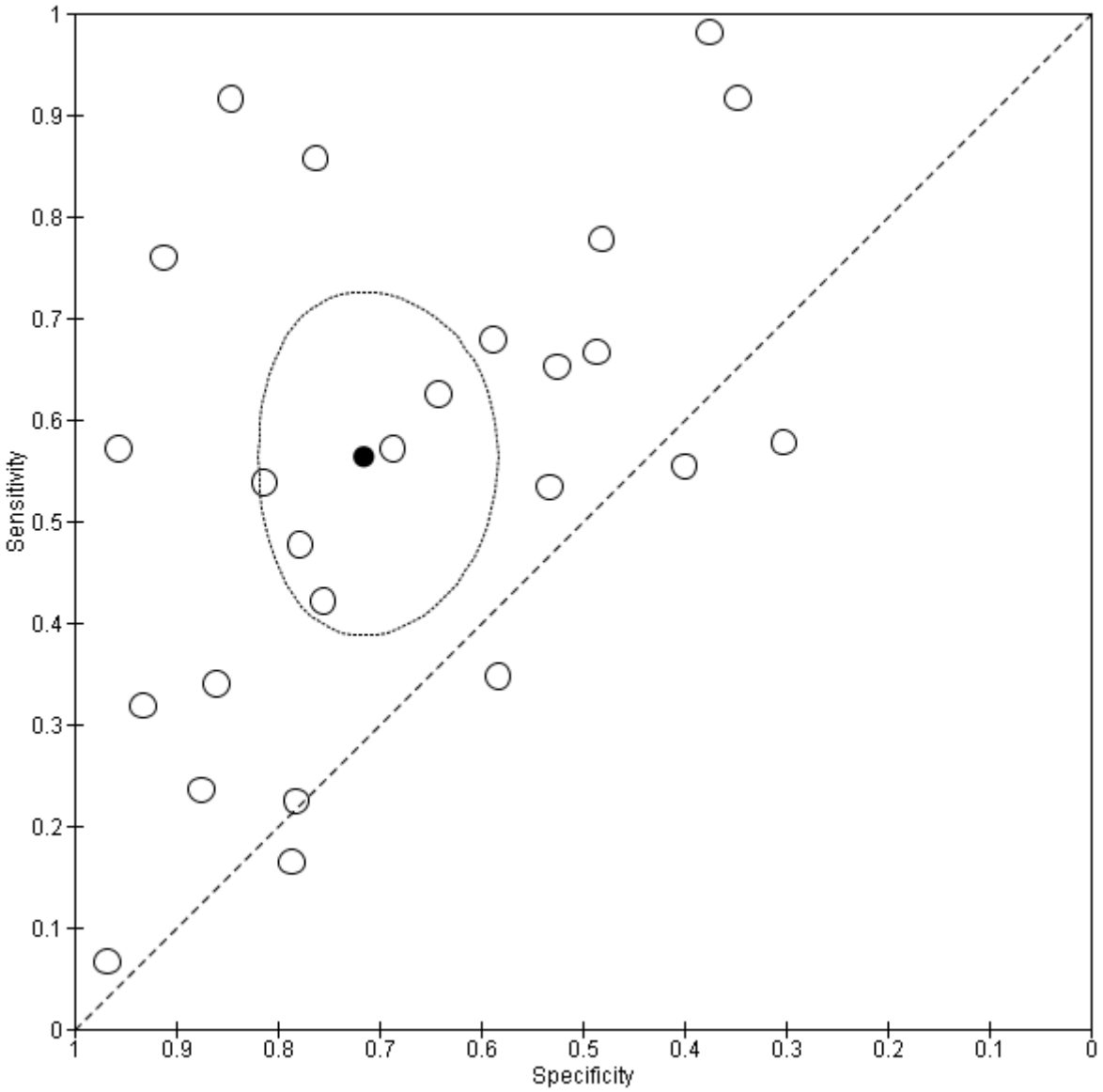


F.1.59 Microcalcifications AND hypoechoicity AND type III vascularisation

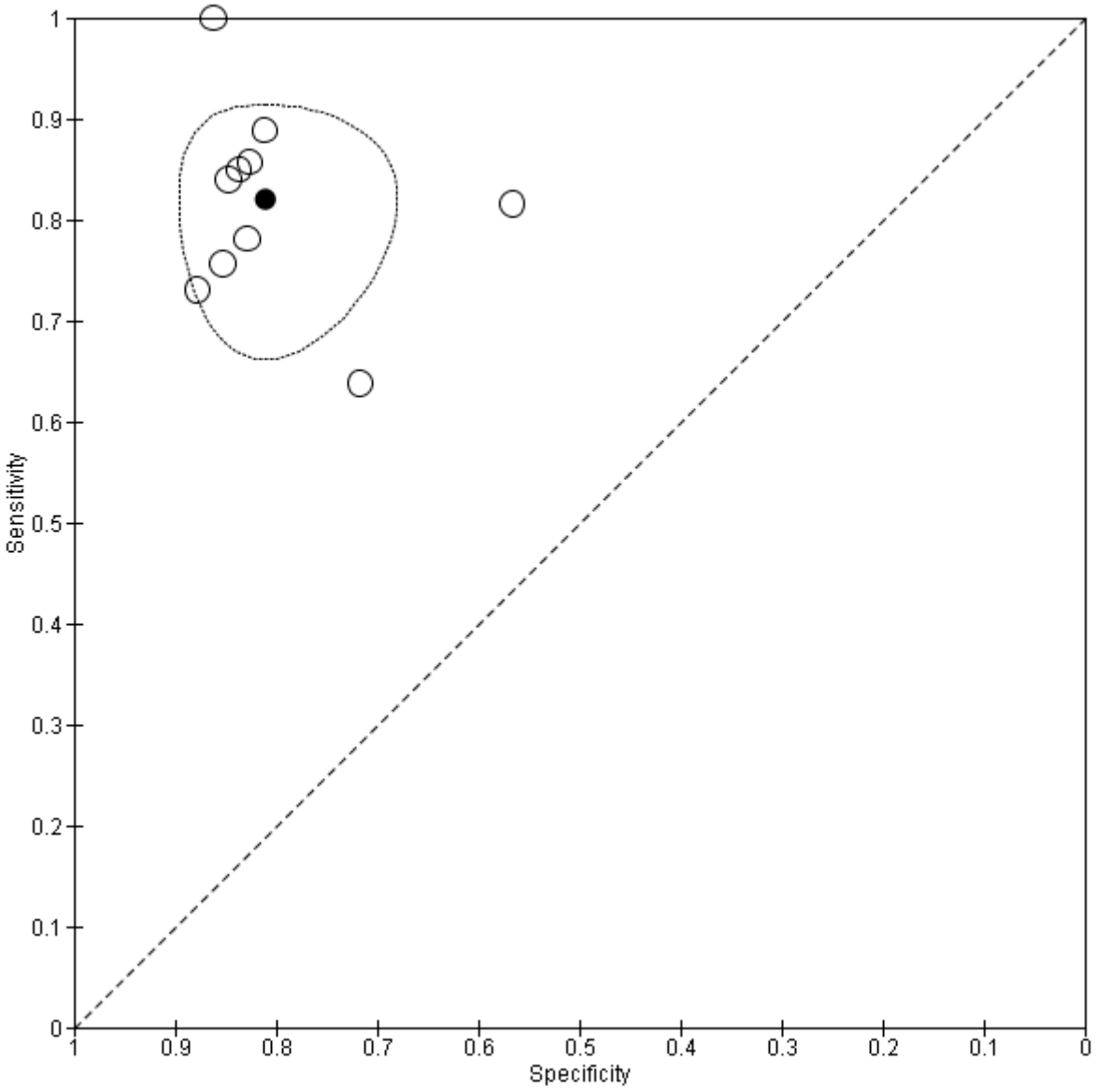


F.1.60 Absent halo AND hypoechoicity AND type III vascularisation

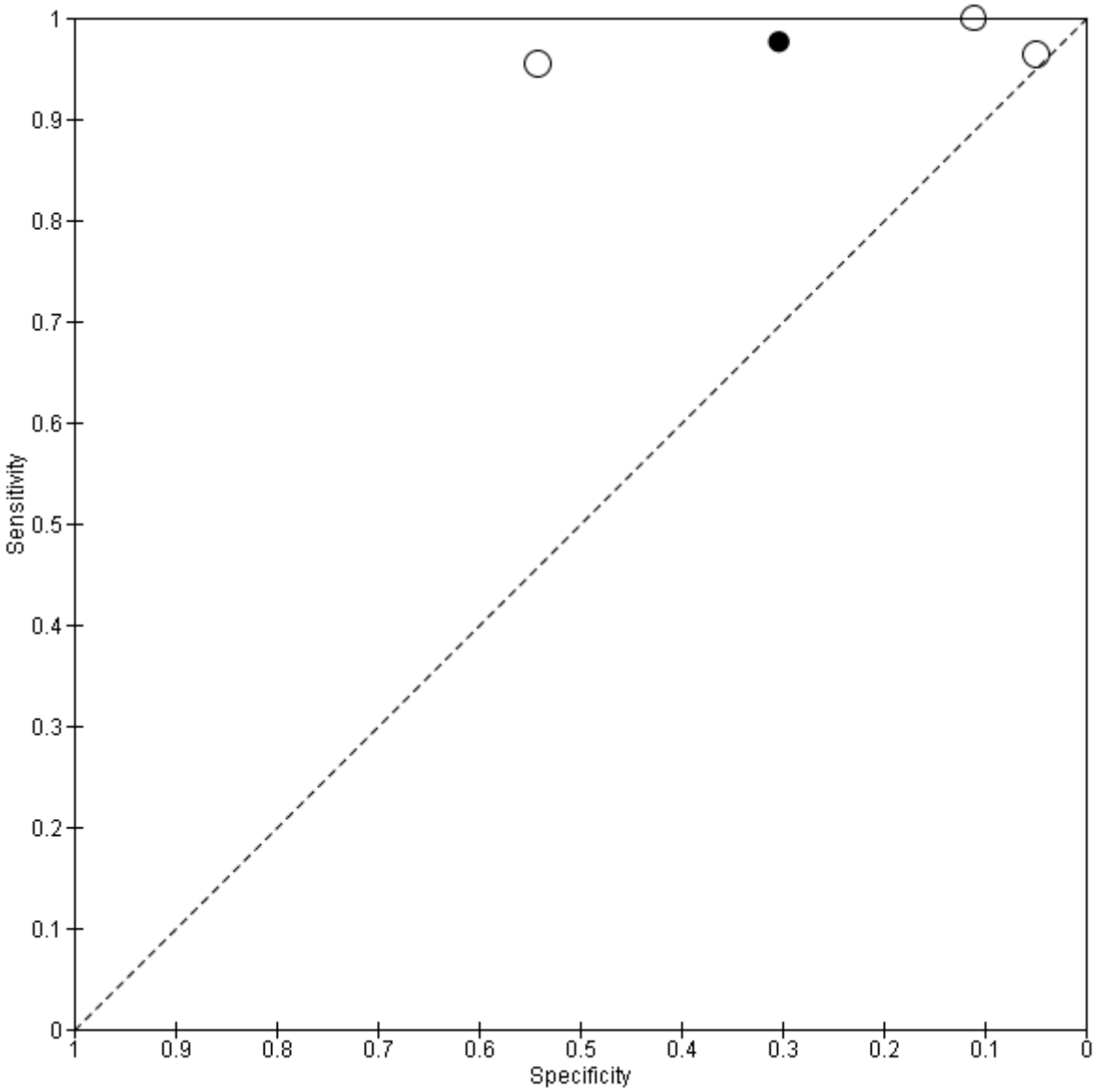



Sensitivity
Specificity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

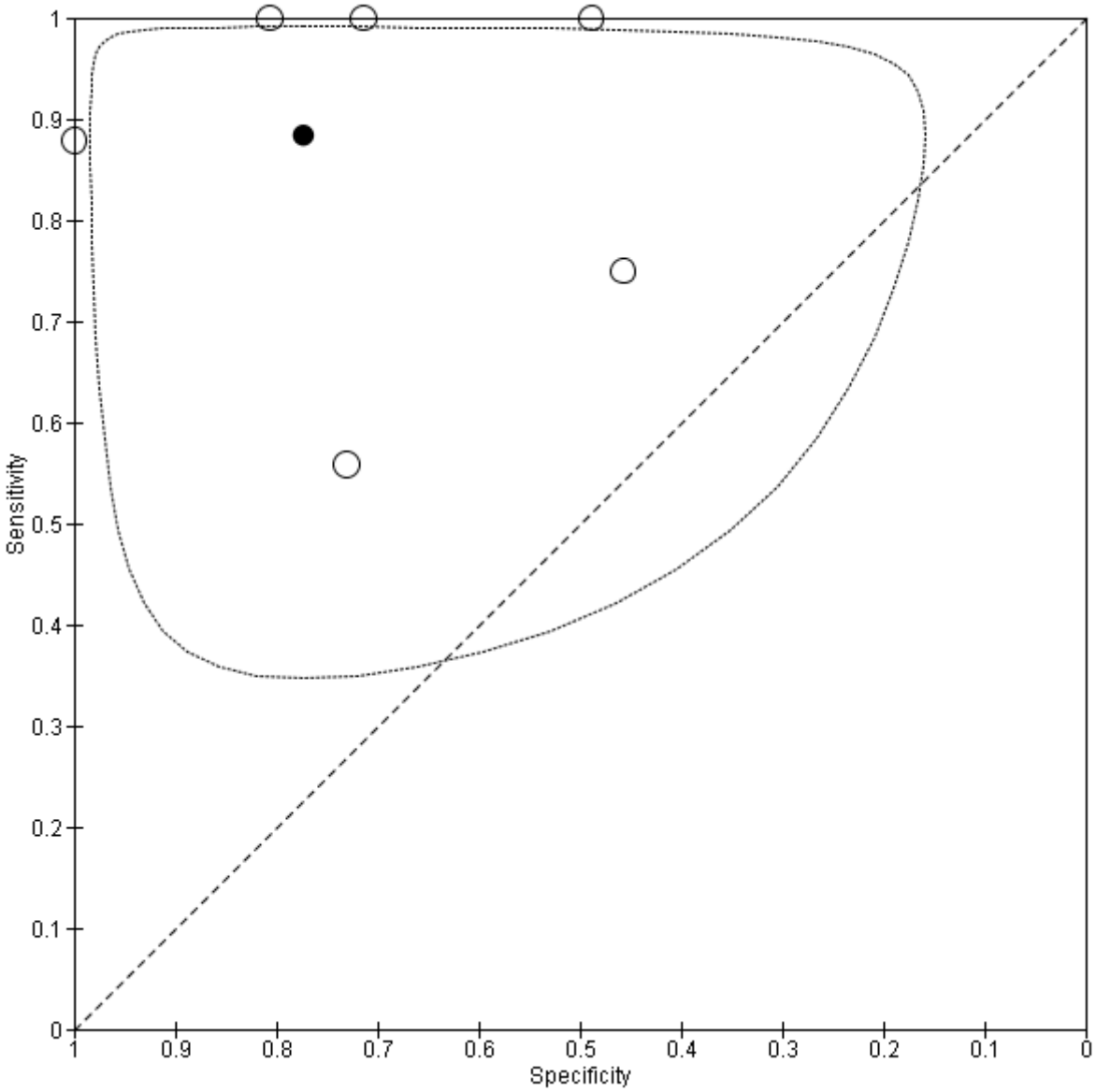
F.1.61 Elastography – Asteria 1-4 scale: 3 and above



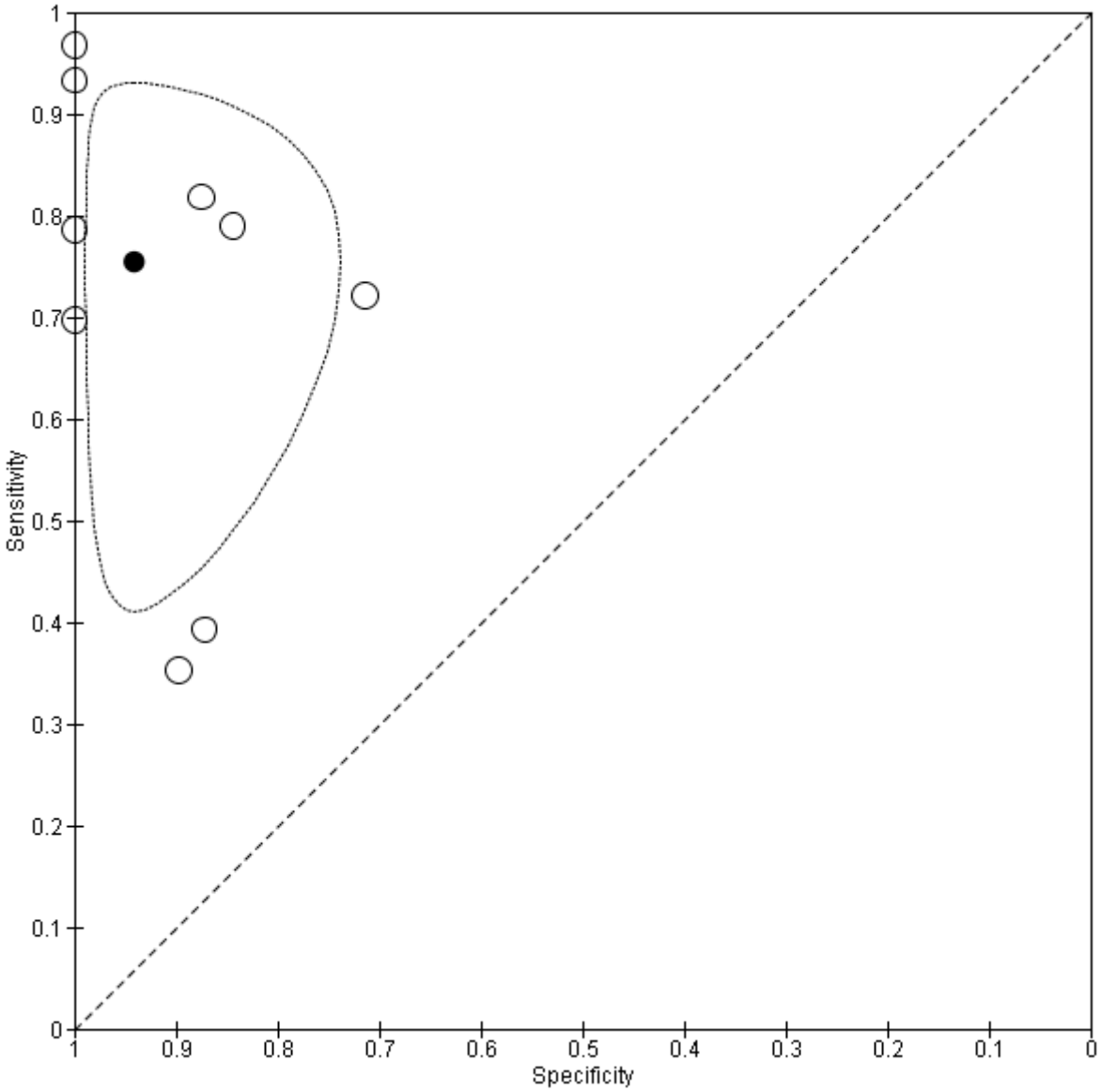
F.1.62 Elastography – Rago 1-5 scale: 2 and above



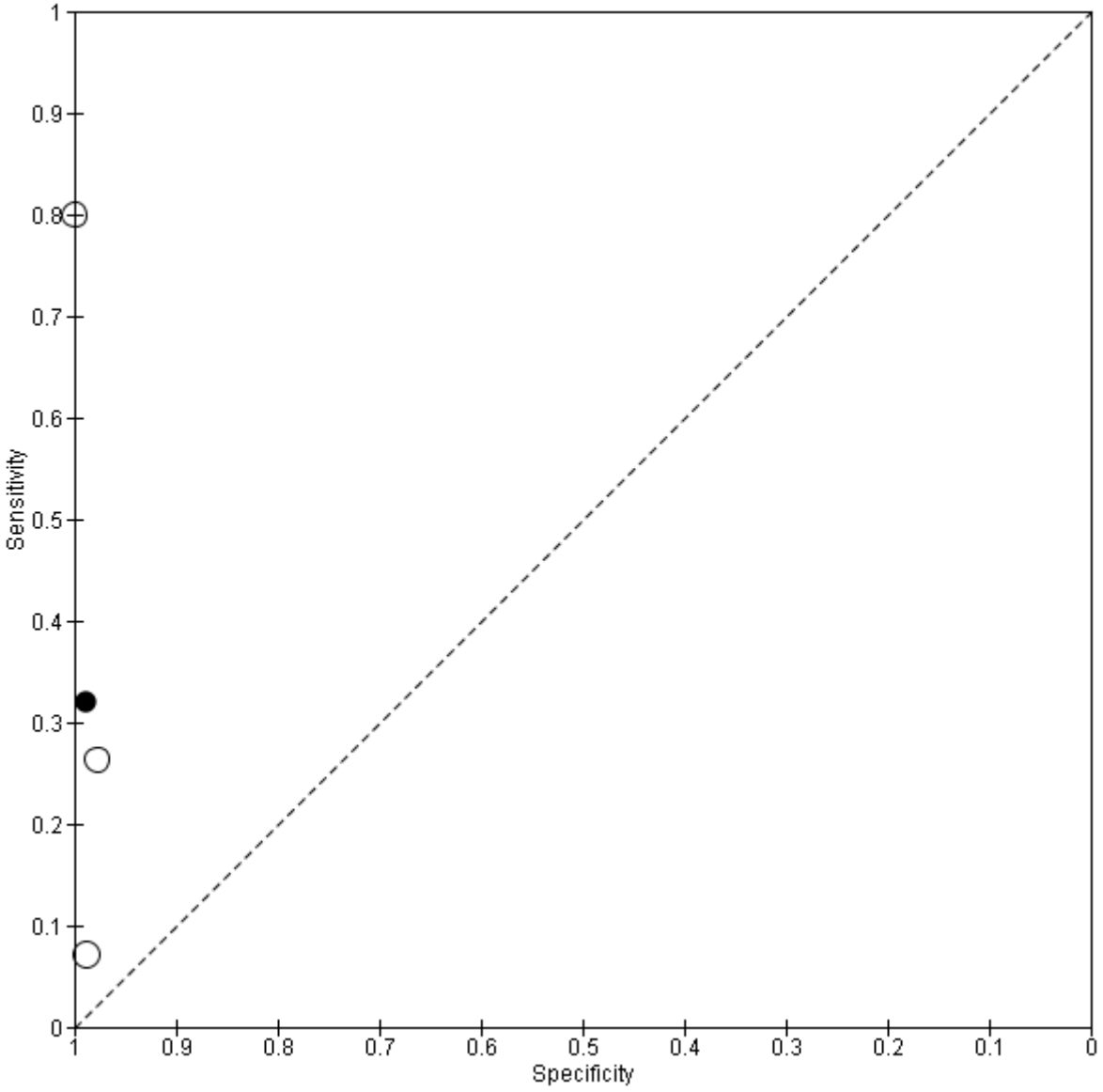
F.1.63 Elastography – Rago 1-5 scale: 3 and above



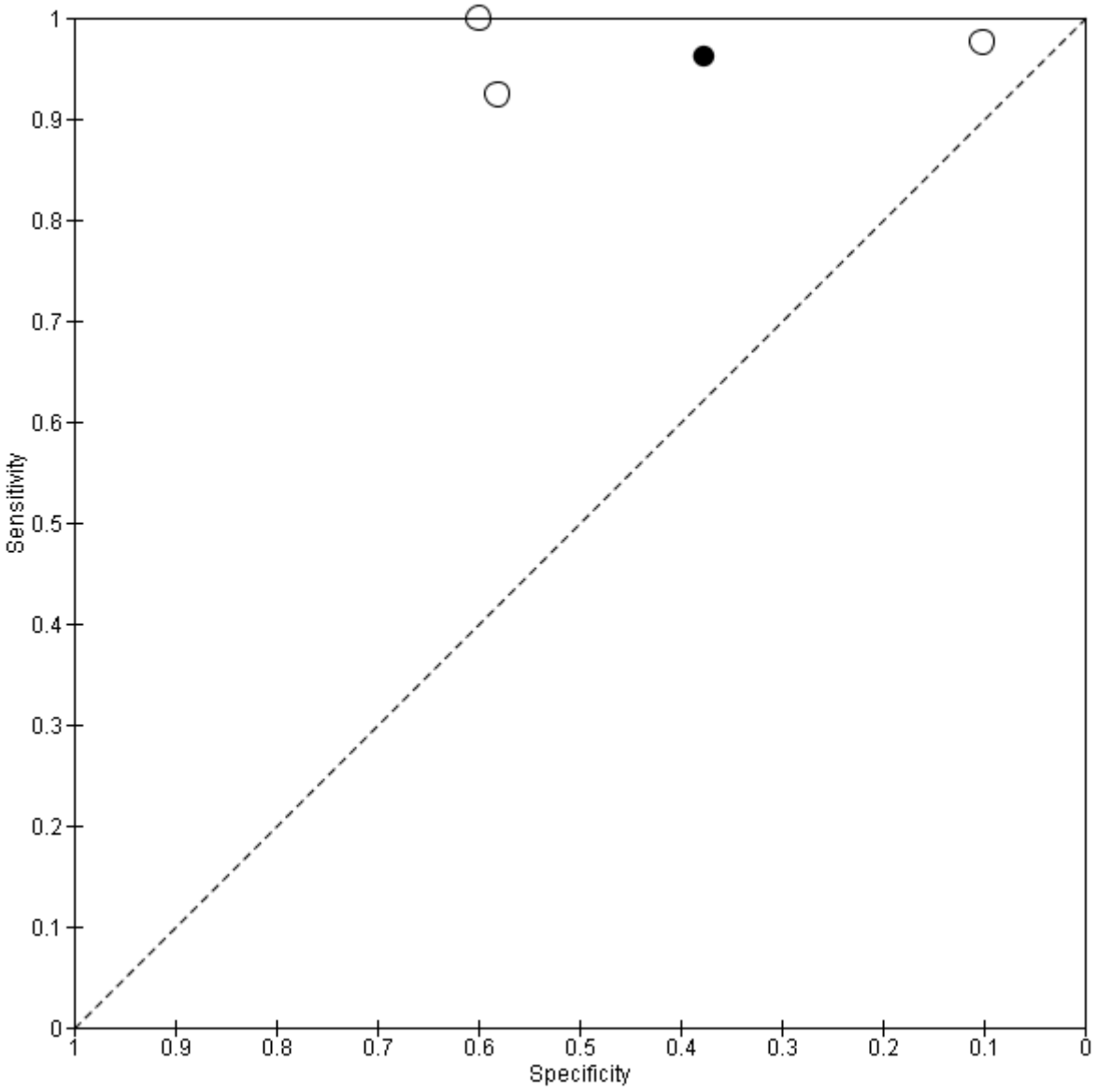
F.1.64 Elastography – Rago 1-5 scale: 4 and above



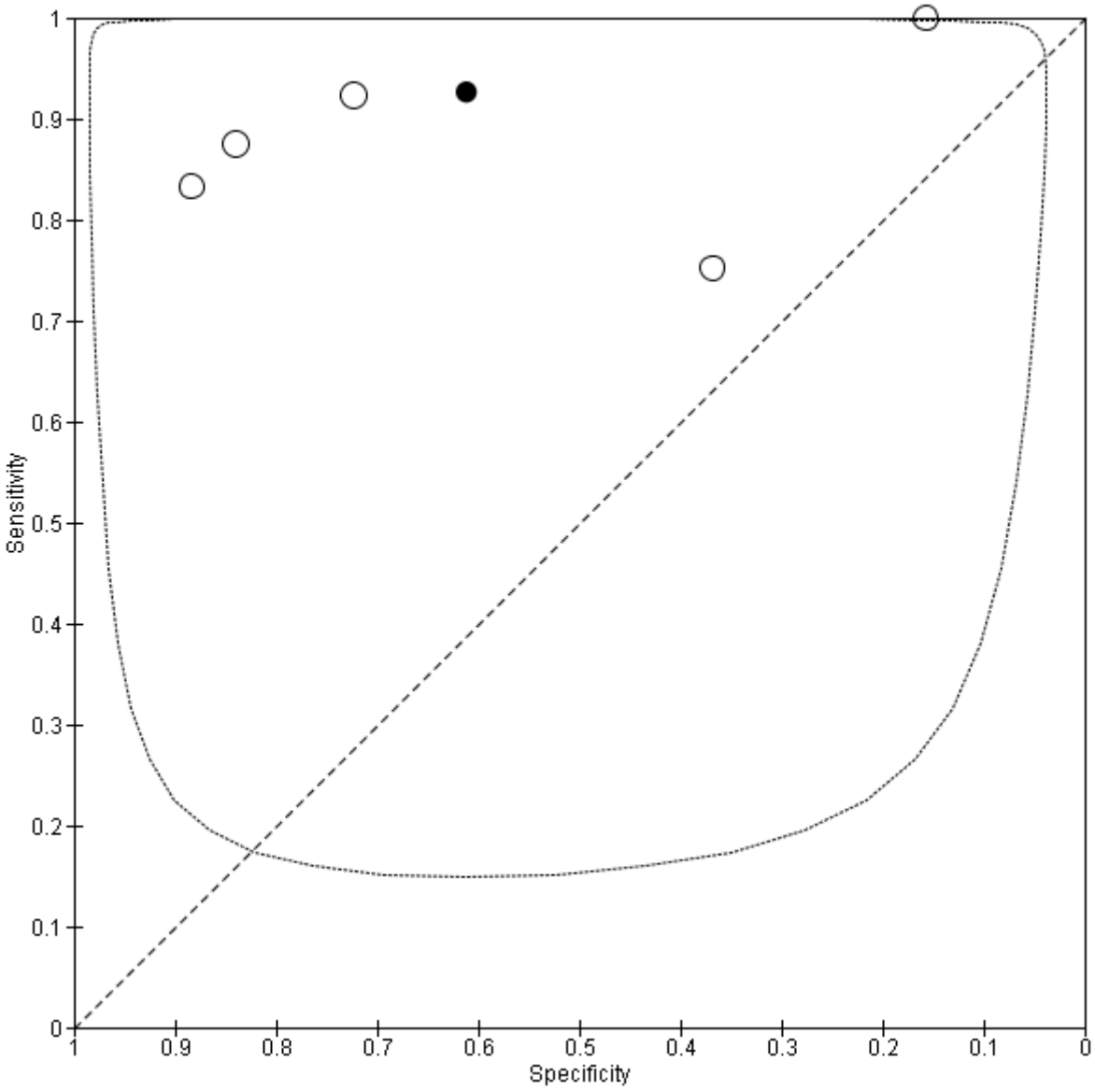
F.1.65 Elastography – Rago 1-5 scale: 5



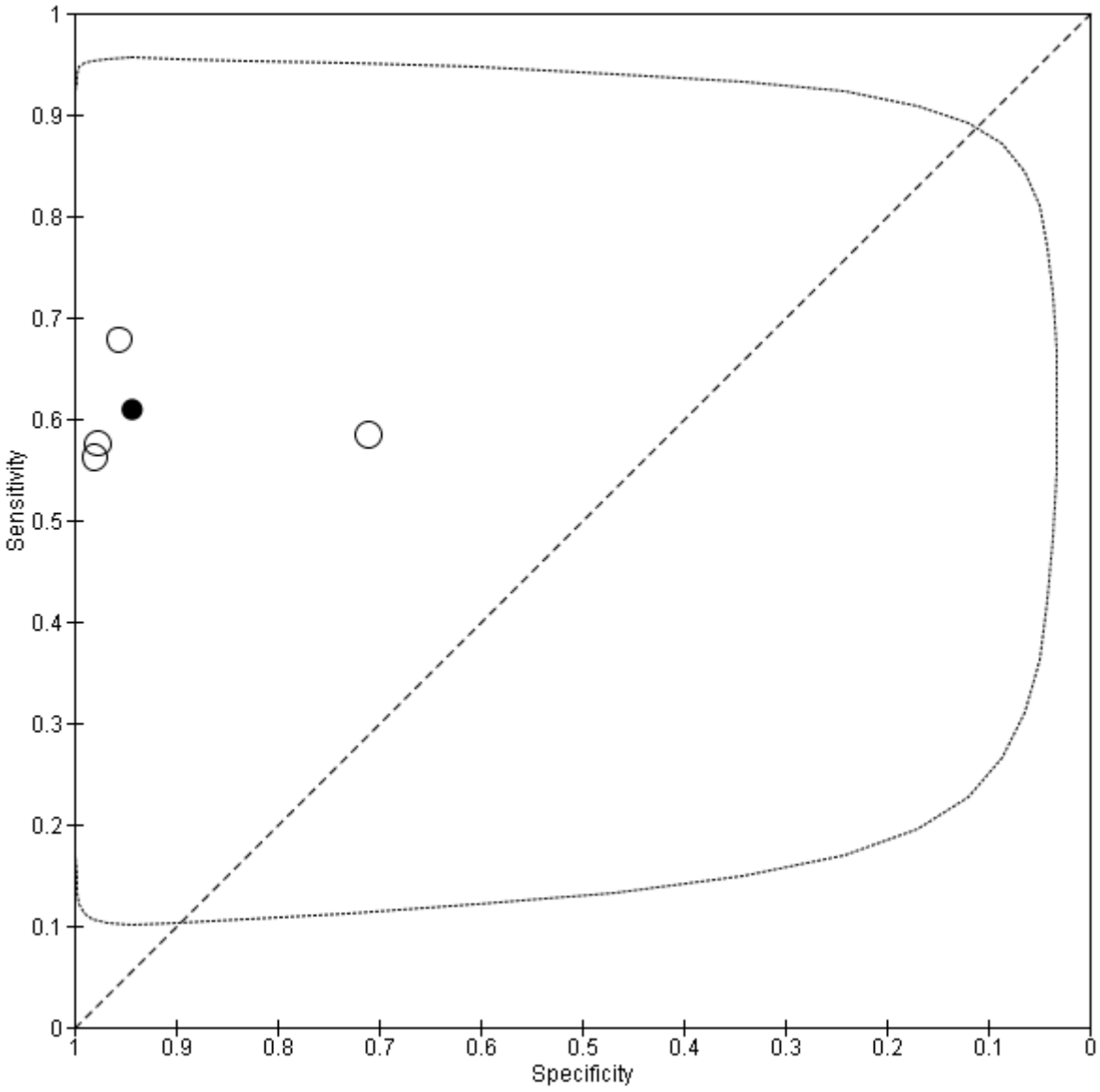
F.1.66 Elastography – ITOH 1-5 scale: 2 or more



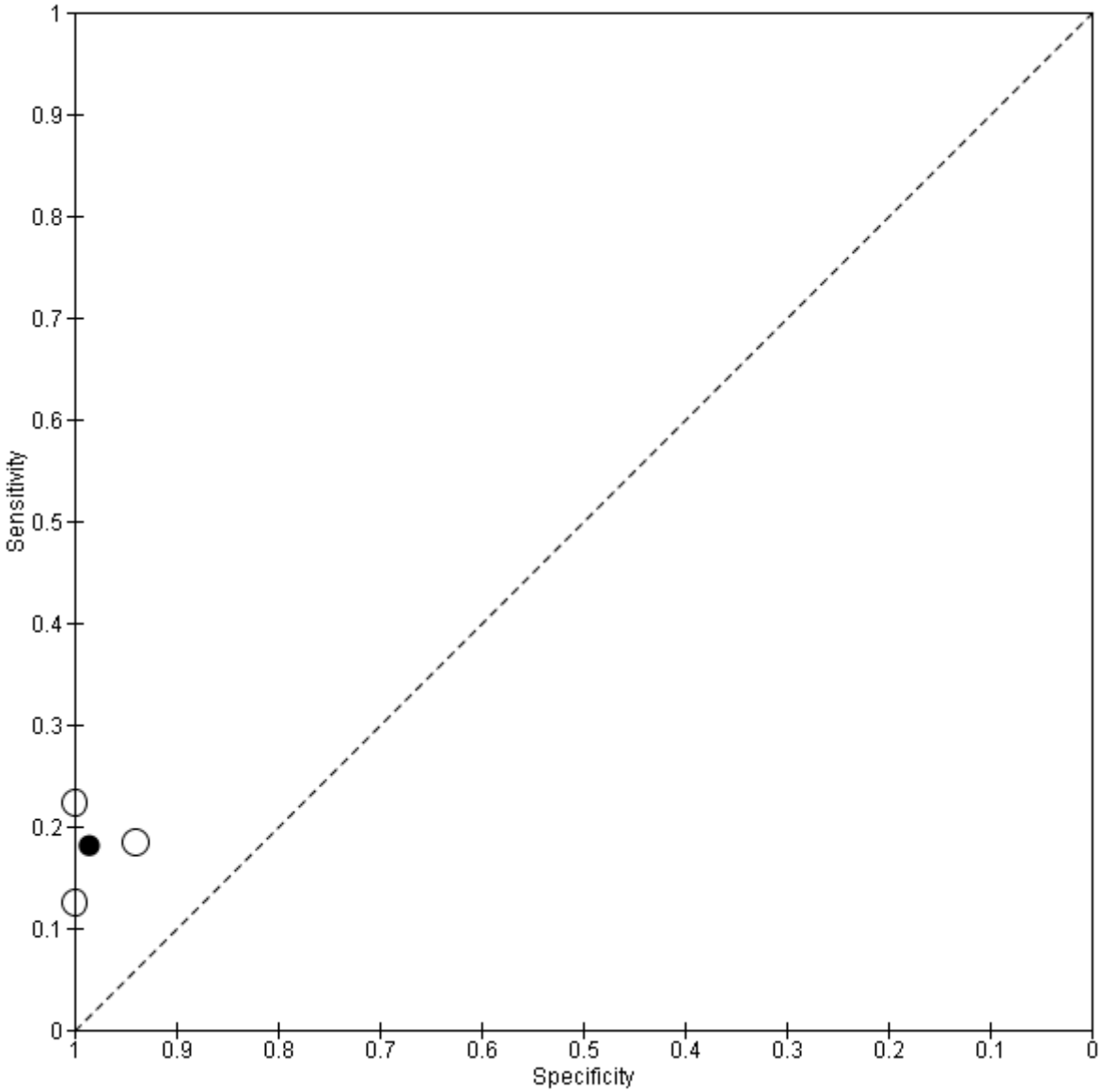
F.1.67 Elastography – ITOH 1-5 scale: 3 or more



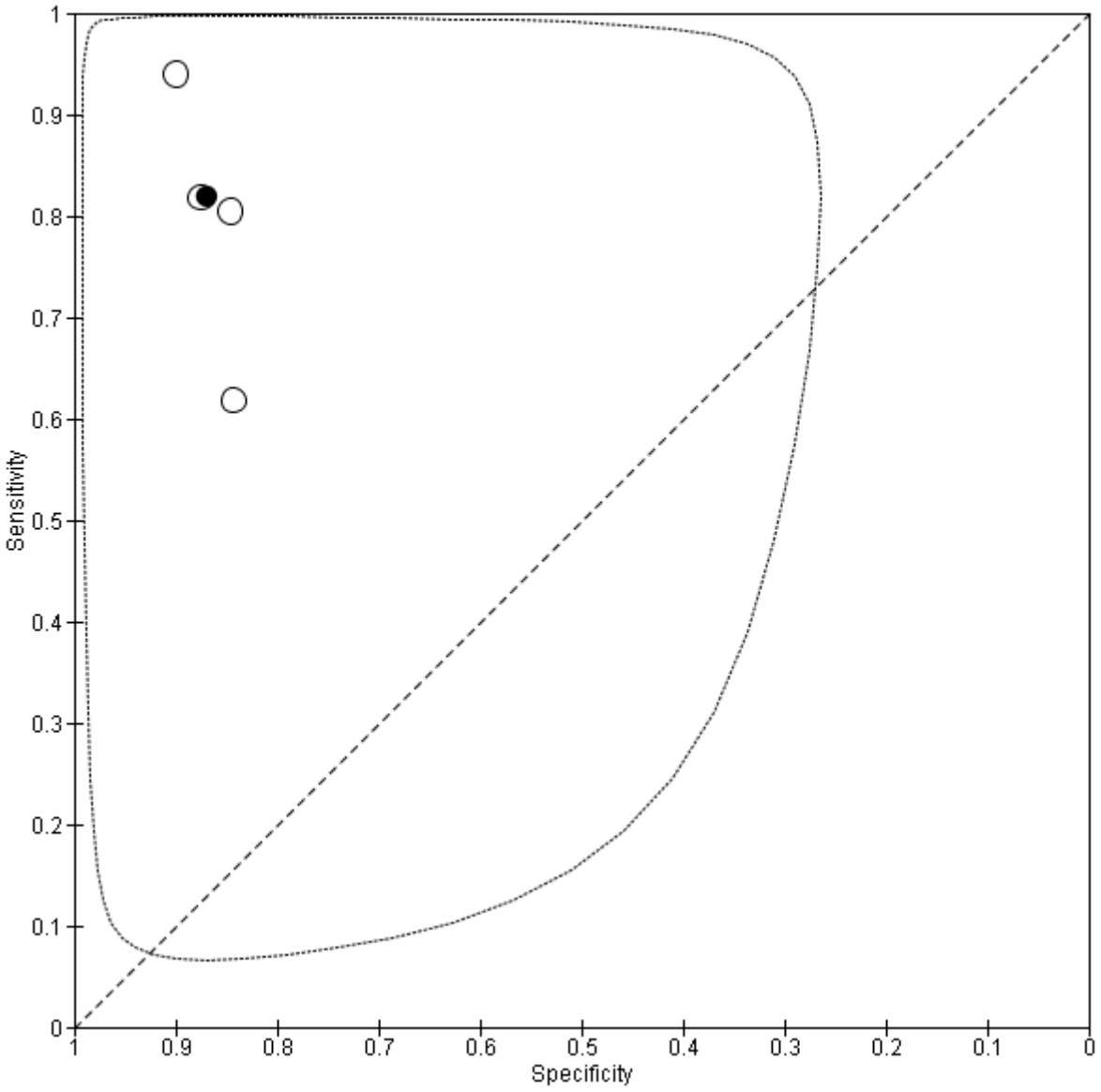
F.1.68 Elastography – ITOH 1-5 scale: 4 or more



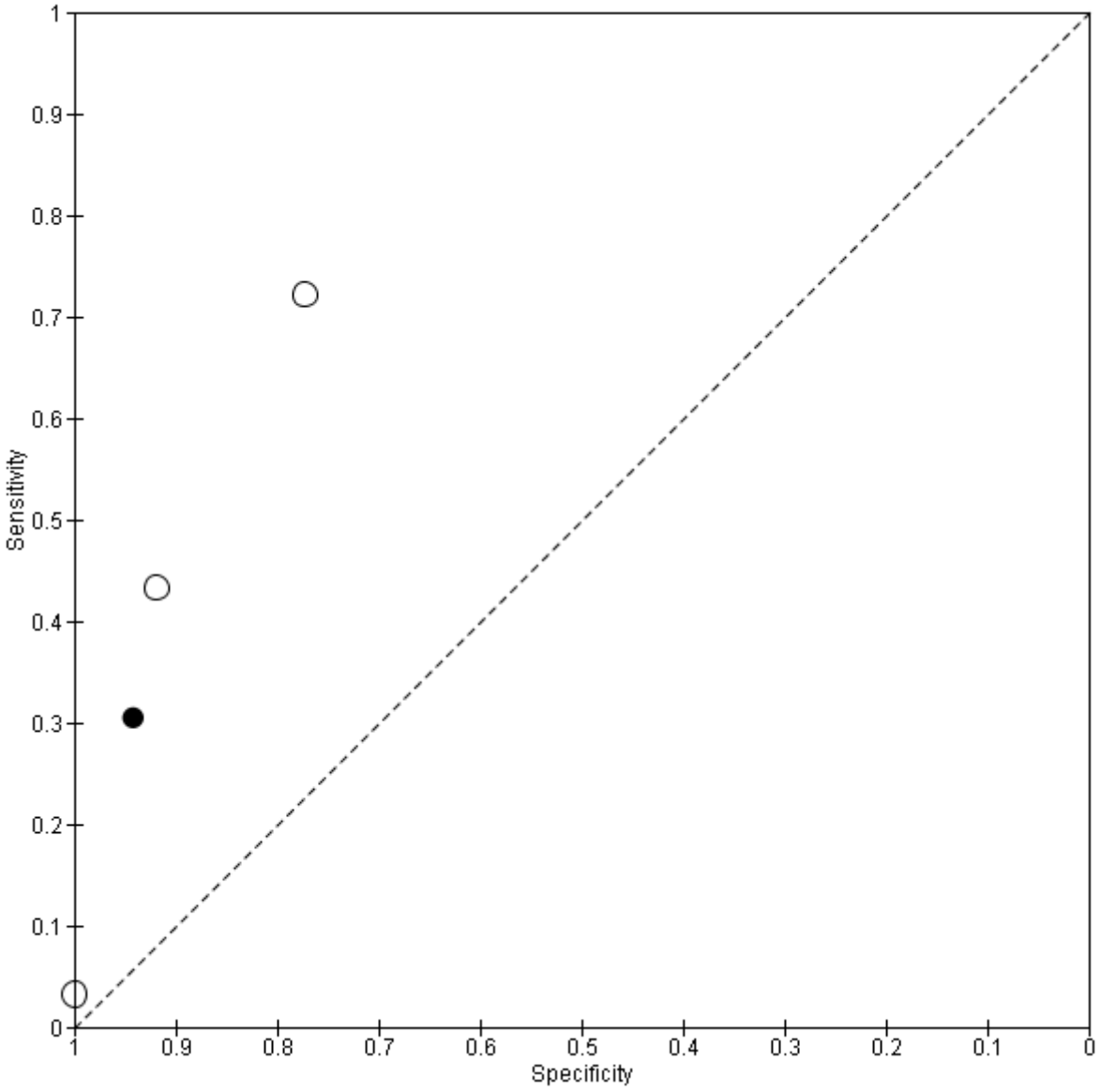
F.1.69 Elastography – ITOH 1-5 scale: 5



F.1.70 Elastography – RGB 0-4 scale: 3 or more



F.6.71 Microcalcifications AND absent halo AND type III vascularisation



F.2 Forest plots Threshold of size and classification of thyroid nodules.

F.2.1 Volumetric progression versus no volumetric progression

Figure 233: Baseline nodule volume

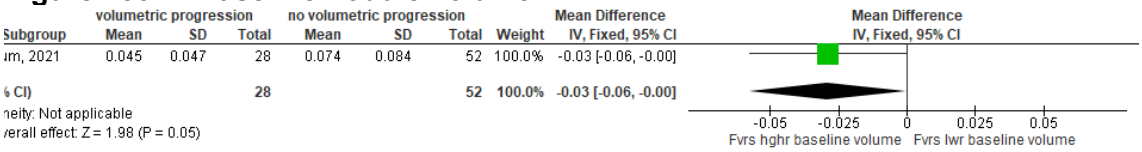


Figure 234:Baseline nodule diameter

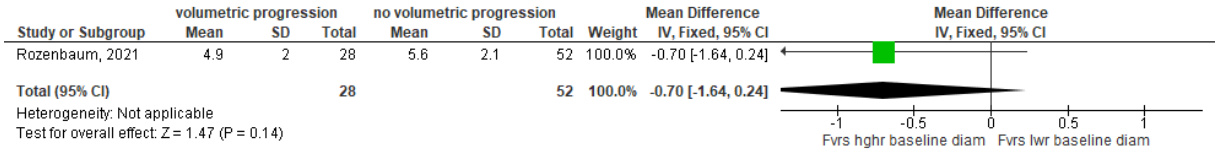


Figure 235:Microcalcifications

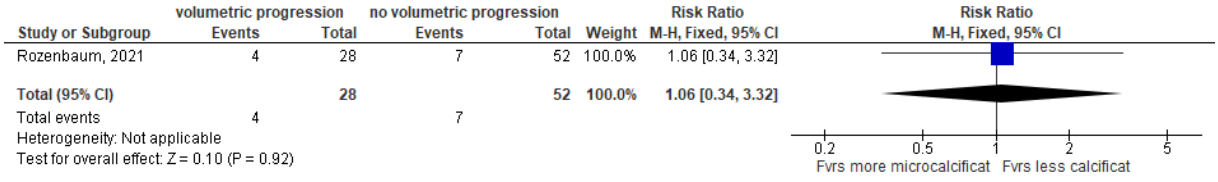


Figure 236:Hypoechogenecity

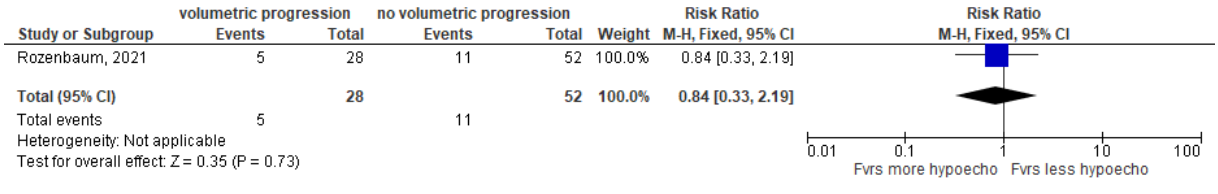


Figure 237:Irregular margins

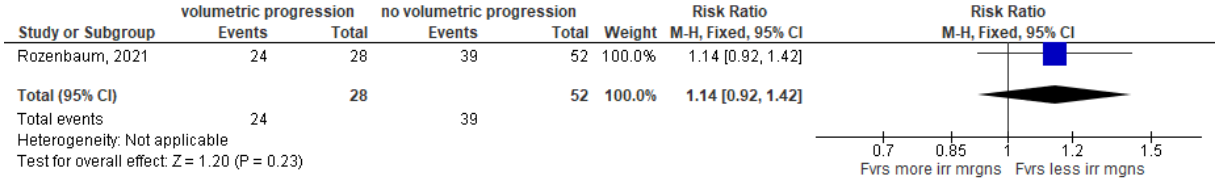


Figure 238:Irregular shape

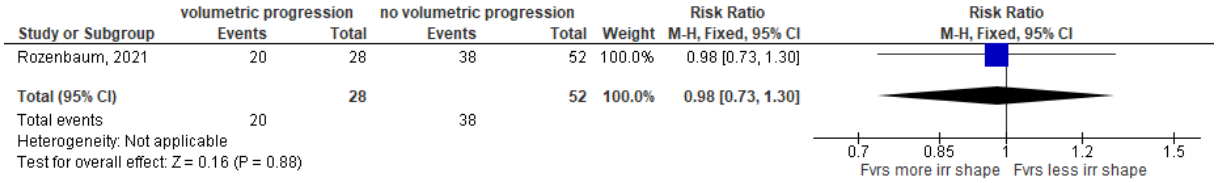


Figure 239:Two or more criteria on EU TIRADS 5

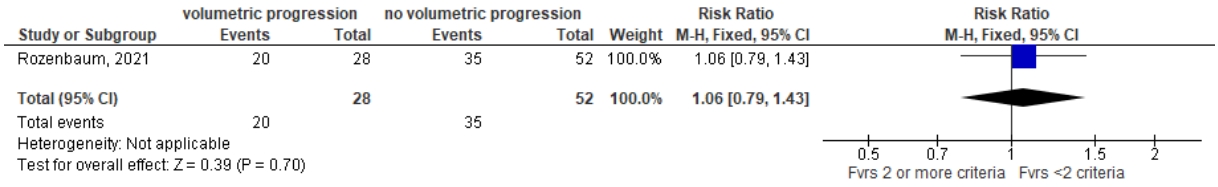


Figure 240:No vascularity

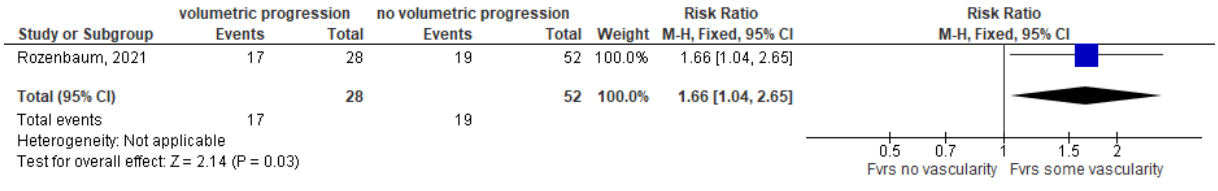


Figure 241:Peripheral vascularity

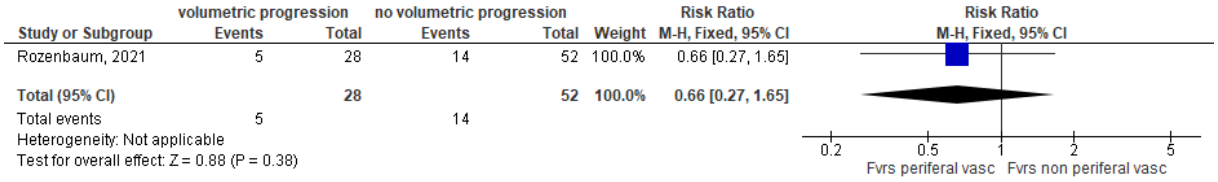
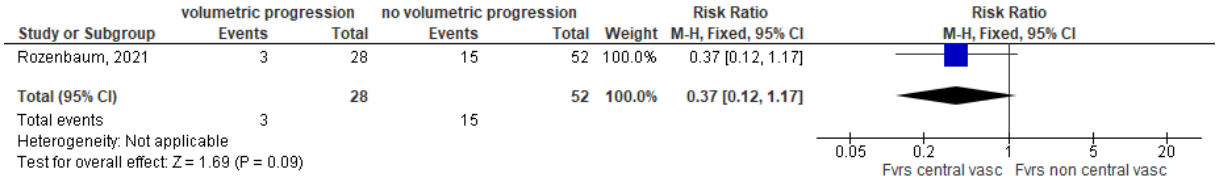


Figure 242:Central vascularity



Appendix G GRADE table

G.1 Diagnostic accuracy of ultrasounds

Summary of evidence found in Table 4 – Table 10.

G.2 Threshold of nodule size and classification

Table 17: Clinical evidence profile: Threshold of size and US characteristics for efficacy of active surveillance

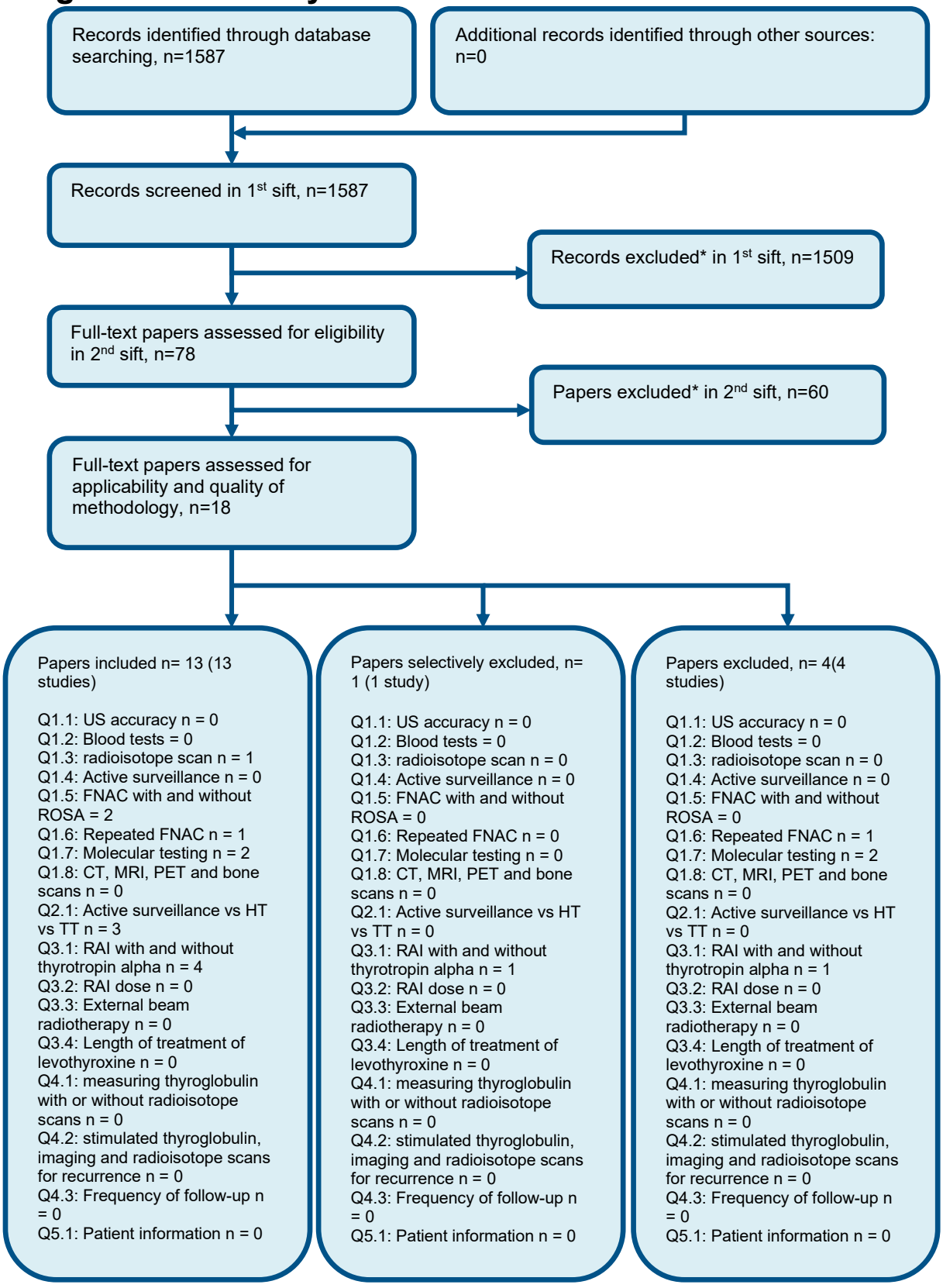
Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Overall	Control	Relative (95% CI)	Absolute		
baseline volume (Better indicated by lower values)												
1	non randomised study	Very serious[1]	NA	Serious[2]	Serious[3]	none	28	52	-	MD 0.03 lower (0.06 lower to 0 higher)	VERY LOW	CRITICAL
baseline diameter (Better indicated by lower values)												
1	non randomised study	Very serious[1]	NA	Serious[2]	Serious[3]	none	28	52	-	MD 0.7 lower (1.64 lower to 0.24 higher)	VERY LOW	CRITICAL
microcalcifications												
1	non randomised study	Very serious[1]	NA	Serious[2]	Very serious[3]	none	4/28 (14.3%)	7/52 (13.5%)	RR 1.06 (0.34 to 3.32)	8 more per 1000 (from 89 fewer to 312 more)	VERY LOW	CRITICAL
hypoechogenicity												
1	non randomised study	Very serious[1]	NA	Serious[2]	Very serious[3]	none	5/28 (17.9%)	11/52 (21.2%)	RR 0.84 (0.33 to 2.19)	34 fewer per 1000 (from 142 fewer to 252 more)	VERY LOW	CRITICAL
irregular margins												
1	non randomised study	Very serious[1]	NA	Serious[2]	Serious[3]	none	24/28 (85.7%)	39/52 (75%)	RR 1.14 (0.92 to 1.42)	105 more per 1000 (from 60 fewer to 315 more)	VERY LOW	CRITICAL
								75%		105 more per 1000 (from 60 fewer to 315 more)		
irregular shape												
1	non randomised study	Very serious[1]	NA	Serious[2]	Very serious[3]	none	20/28 (71.4%)	38/52 (73.1%)	RR 0.98 (0.73 to 1.3)	15 fewer per 1000 (from 197 fewer to 219 more)	VERY LOW	CRITICAL
2 or more criteria on EU TIRADS 5												
1	non randomised study	Very serious[1]	NA	Serious[2]	Very serious[3]	none	20/28 (71.4%)	35/52 (67.3%)	RR 1.06 (0.79 to 1.43)	40 more per 1000 (from 141 fewer to 289 more)	VERY LOW	CRITICAL

no vascularity											
1	non randomised study	Very serious[1]	NA	Serious[2]	Serious[3]	none	17/28 (60.7%)	19/52 (36.5%)	RR 1.66 (1.04 to 2.65)	241 more per 1000 (from 15 more to 603 more)	VERY LOW	CRITICAL
peripheral vascularity											
1	non randomised study	Very serious[1]	NA	Serious[2]	Very serious[3]	none	5/28 (17.9%)	14/52 (26.9%)	RR 0.66 (0.27 to 1.65)	92 fewer per 1000 (from 197 fewer to 175 more)	VERY LOW	CRITICAL
central vascularity											
1	non randomised study	Very serious[1]	NA	Serious[2]	Serious[3]	none	3/28 (10.7%)	15/52 (28.8%)	RR 0.37 (0.12 to 1.17)	182 fewer per 1000 (from 254 fewer to 49 more)	VERY LOW	CRITICAL

1. Risk of bias was assessed with the ROBINS tool. This yielded on an overall judgement of critical risk of bias, based on failure to adjust for any confounding.
2. Indirectness was deemed serious due to the outcome of nodular volumetric progression not necessarily relating to cancer progression
3. Imprecision was rated very serious if the 95% CIs crossed both MIDs and serious if they crossed one MID. MIDs were taken as + 0.5 x the standard deviation of the control group for continuous variables* and a RR of 0.8 and 1.25 for binary variables.
   *MID for nodule volume was 0.042, based on sd of 0.084 in control group, and MID for nodule diameter was 1.05, based on sd of 2.1 in control group

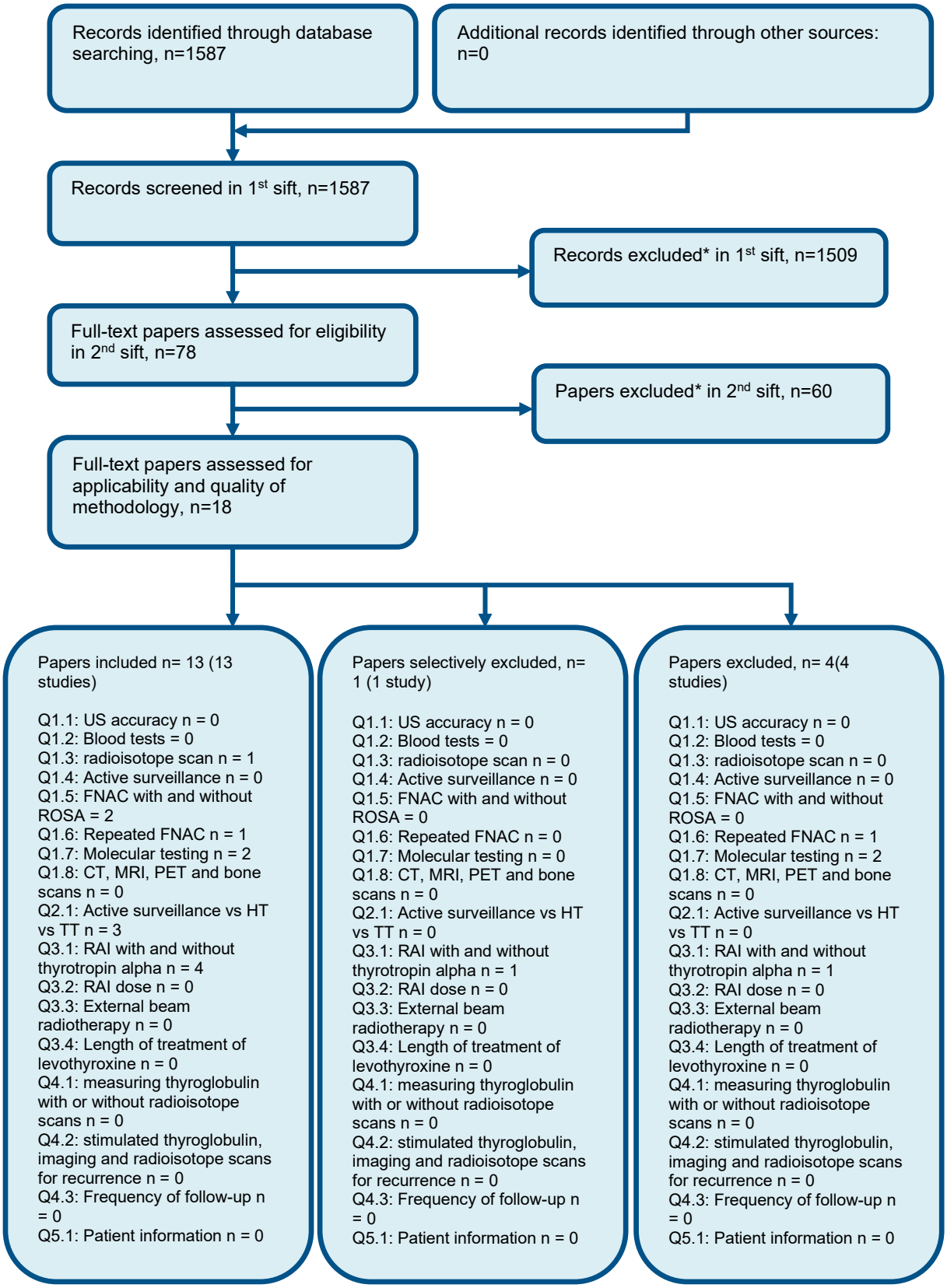
Appendix H Economic evidence study selection

H.1 Diagnostic accuracy of ultrasounds



* Non-relevant population, intervention, comparison, design or setting; non-English language

H.2 Threshold of nodule size and classification



* Non-relevant population, intervention, comparison, design or setting; non-English language

Appendix I Economic evidence tables

None.

Appendix J Excluded studies

J.1 Clinical studies: Diagnostic Accuracy

Table 18: Studies excluded from the clinical review

Reference	Reason for exclusion
Abbasian Ardakani, 2015[2]	Unclear gold standard
Abdelrahman, 2015[4]	FNA gold standard for some
Afifi, 2017[5]	FNA gold standard for some
Aggarwal, 2017[6]	FNA gold standard
Aghaghazvini, 2020[8]	FNA gold standard in some
Ahmadi, 2019[9]	Patients already tested with FNAC - not relevant to our review population
Ahn, 2018[11]	FNA gold standard for some
Al-Chalabi, 2019[14]	FNA gold standard for some
Amin, 2021[15]	FNA gold standard for all
An, 2020[16]	Restricted to patients at TIRADS 4
Arambewela, 2020[19]	FNA gold standard in some
Arpana, 2018[20]	FNA gold standard
Aydin, 2014[22]	FNA gold standard
Azizi, 2013[23]	FNA gold standard for some
Bae, 2018[24]	FNA gold standard in some
Baek, 2021[25]	Study evaluated detection of diffuse thyroid disease, not malignancy
Bafaraj, 2020[26]	Unclear gold standard
Bas, 2022[28]	FNA gold standard for some
Baz, 2021[29]	FNA gold standard for some
Becker, 1997[30]	Nuclear scintigraphy was gold standard
Bederina, 2014[31]	FNA gold standard for some
Bonavita, 2009[34]	FNA gold standard for some
Bozbora, 2002[37]	Unclear gold standard
Brandenstein, 2021[38]	No diagnostic accuracy evaluation
Brandler, 2018[39]	Non-malignant target condition
Brunese, 2008[41]	Unable to source
Cakal, 2015[42]	FNA gold standard
Cao, 2021[46]	FNA gold standard for some
Castellana, 2020[47]	No mention of gold standard; all patients had malignancy
Celletti, 2021[48]	Restricted to a sample with indeterminate pathology
Cetin, 2015[49]	FNA gold standard for some

Reference	Reason for exclusion
Chanda, 2020[50]	Detection of superficial lymph node malignancy
Chang, 2016[51]	Biopsy gold standard (non-surgical)
Chen, 2010[56]	No diagnostic accuracy analysis
Chen, 2016[52]	FNA gold standard for some
Chen, 2019[53]	FNA gold standard for some
Chen, 2019[54]	FNA gold standard for some
Cheng, 2013[58]	FNA gold standard for some
Cheng, 2013[59]	Did not characterise target group as 'malignant' - merely 'non benign'. This group included follicular neoplasms and so was not a malignant group.
Cho, 2017[61]	FNA gold standard for some
Choi, 2010[62]	Incorrect gold standard for some (i.e. Tg, imaging, etc)
Choi, 2015[63]	Paper was focussed on use of computer detection software rather than US itself
Choi, 2017[64]	FNA gold standard for some
Chung, 2021[65]	FNA / CNB gold standard for most (91.4%)
Ciledag, 2012[66]	FNA gold standard for some
Clark, 2019[67]	FNA gold standard
Colakoglu, 2016[70]	FNA gold standard
Cordes, 2016[71]	Excluded - restricted to follicular neoplasms diagnosed by FNA
Creo, 2018[72]	FNA gold standard for some
da Silva, 2021[73]	Not available
Daniels, 2021[74]	FNA gold standard for most (62%)
Dawoud, 2017[75]	FNA gold standard for some
Deng, 2014[76]	FNA gold standard for some
Dhayalan, 2018[79]	FNA gold standard
Dighe, 2008[80]	FNA gold standard
Du, 2018[83]	FNA gold standard in some
Duan, 2016[84]	FNA gold standard for some
Dy, 2017[85]	FNA gold standard for some
Ebeed, 2017[86]	FNA gold standard for some
Elsayed, 2016[88]	FNA gold standard
Esfahanian, 2016[89]	FNA gold standard
Fang, 2019[90]	Predictive model including non-US variables
Farghadani, 2019[91]	FNA gold standard for some
Farihah, 2018[92]	FNA gold standard for some
Ferrari, 2008[93]	FNA gold standard for some
Freire da Silva, 2021[94]	Duplicate
Fresilli, 2020[95]	FNA gold standard for some
Friedrich-Rust, 2010[97]	FNA gold standard in some
Friedrich-Rust, 2012[96]	FNA gold standard for some
Friedrich-Rust, 2016[98]	FNA gold standard for some

Reference	Reason for exclusion
Fukuhara, 2018[99]	FNA gold standard for some
Gacayan, 2021[100]	FNA gold standard for all
Galimzianova, 2020[101]	Did not use surgical histopathology as gold standard
Gannon, 2018[102]	FNA gold standard for some
Gao, 2018[103]	Paper was focussed on use of computer detection software rather than US itself
Gao, 2019[104]	No positive or negative diagnostic assignation
Gitto, 2019[109]	FNA gold standard for some
Goel, 2020[110]	FNA gold standard for some
Gotzberger, 2016[113]	FNA gold standard for some
Goundan, 2021[114]	FNA gold standard for some
Grani, 2015[116]	FNA gold standard for some
Grani, 2020[115]	FNA gold standard for majority
Grazhdani, 2014[118]	FNA gold standard for some
Gregory, 2018[119]	FNA gold standard
Guan, 2019[122]	FNA gold standard for some
Gul, 2009[123]	FNA gold standard
Gulcelik, 2008[124]	Patients already tested with FNAC - not relevant to our review population
Guo, 2019[125]	FNA gold standard for some
Ha, 2015[126]	Some had only core needle biopsy (CNB) as the reference standard. Like FNAC, this is not as good as surgical histopathology, so shouldn't be used.
Ha, 2016[127]	FNA gold standard for some
Ha, 2017[130]	FNA gold standard for some
Ha, 2018[128]	FNA gold standard for some
Ha, 2019[131]	FNA gold standard for some
Ha, 2021[129]	FNA gold standard for some
Hachim, 2018[132]	Unclear gold standard
Hahn, 2018[133]	FNA gold standard for some
Hamidi, 2015[134]	Unclear gold standard (unlikely to be surgical histopathology)
Han, 2019[137]	FNA gold standard for some
Han, 2021[136]	Not aimed at diagnosing nodule malignancy - diagnosing ETEs and cervical lymph node metastases
Han, 2020[135]	All subjects receiving gold standard test had thyroid cancer. Accuracy of ultrasound was evaluated only for detection of extrathyroidal extension and LN mets.
Han, 2021[138]	Not available
Hayes, 2021[141]	unclearly reported - unclear how US results related to gold standard pathology results
He, 2017[143]	FNA gold standard for some
Hess, 2020[145]	FNA gold standard for some

Reference	Reason for exclusion
Hu, 2015[152]	No diagnostic accuracy analysis
Hu, 2020[151]	Prediction of ETE not malignancy
Hu, 2021[150]	FNA gold standard for some
Huang, 2019[158]	FNA gold standard for some
Huang, 2021[157]	FNA gold standard for some
Huaqun, 2019[159]	FNA gold standard for some
Huh, 2021[160]	FNA gold standard in some
Hussain, 2020[161]	Unclear gold standard (unlikely to be surgical histopathology)
Hwang, 2011[162]	Detection of lymph node metastases
Ito, 2007[164]	FNA gold standard for some
Ivanac, 2007[165]	FNA gold standard for some
Jainulabdeen, 2019[166]	Looking at staging not diagnosis
Jalan, 2017[167]	FNA gold standard
Jeong, 2019[168]	FNA gold standard for some
Jiang, 2015[170]	No diagnostic accuracy evaluation
Kagoya, 2010[174]	FNA gold standard for some
Kathuria, 2003[176]	Gold standard unclear
Kikovic, 2021[178]	FNA, clinical examination and echosonography used as gold standards and true surgical histopathology only one of several methods used.
Kim, 2010[190]	FNA gold standard for majority
Kim, 2013[182]	FNA gold standard for some
Kim, 2013[184]	FNA gold standard for some
Kim, 2014[183]	Detection of ETE in patients who all had PTC
Kim, 2014[187]	FNA gold standard
Kim, 2015[186]	FNA gold standard for some
Kim, 2015[191]	FNA gold standard for majority
Kim, 2019[185]	FNA gold standard
Kim, 2021[181]	FNA gold standard for some
Ko, 2014[194]	FNA gold standard for some
Ko, 2016[193]	FNA gold standard
Koh, 2018[196]	FNA gold standard for majority
Koike, 2001[197]	Data not possible to interpret
Kong, 2019[199]	Unclear if GS was surgical histopathology
Kuo, 2020[203]	Unclear if GS was surgical histopathology
Lampung, 2018[206]	FNA gold standard for some
Latif, 2021[208]	FNA gold standard for most (86%)
Leboulleux, 2007[209]	Detection of cervical lymph node malignancy
Lee, 2003[210]	FNA gold standard for some
Lee, 2009[214]	FNA gold standard for some
Lee, 2010[215]	No diagnostic accuracy evaluation
Lee, 2011[216]	FNA gold standard
Lee, 2017[213]	FNA gold standard in some

Reference	Reason for exclusion
Li, 2016[221]	Detection of calcifications not malignancy
Li, 2020[222]	Unclear gold standard - 'pathology and imaging'.
Li, 2020[224]	FNA gold standard for some
Li, 2020[226]	FNA gold standard for some
Liang, 2018[227]	Exclude - no sensitivity and specificity data available; only AUC data and ROC curves too low resolution to allow extrapolation of data
Liao, 2019[228]	FNA gold standard
Lim, 2008[231]	FNA gold standard for some
Lim-Dunham, 2017[229]	FNA gold standard for some
Lim-Dunham, 2019[230]	FNA gold standard for some
Lin, 1997[233]	No diagnostic accuracy evaluation for US
Lingam, 2013[235]	FNA gold standard for some
Lippolis, 2011[236]	Patients already tested with FNAC - not relevant to our review population
Liu, 2015[237]	FNA gold standard for some
Liu, 2016[245]	FNA gold standard for some
Liu, 2017[239]	FNA gold standard for some
Liu, 2017[246]	FNA gold standard for some
Liu, 2021[241]	Gold standard was core needle biopsy
Lu, 1994[247]	FNA gold standard for some
Lu, 2017[248]	Insufficient data to calculate sensitivity and specificity
Lu, 2019[249]	Paper was focussed on use of computer detection software rather than US itself
Luo, 2012[250]	FNA gold standard for some
Lyshchik, 2005[251]	FNA gold standard for some
Lyshchik, 2005[252]	Manuscript in Chinese
Macedo, 2018[257]	FNA gold standard for some
Maddaloni, 2021[258]	FNA gold standard for some
Magri, 2013[259]	FNA gold standard
Magri, 2015[261]	FNA gold standard
Maino, 2021[264]	FNA gold standard for some
Mallikarjunappa, 2014[265]	Unclear gold standard
Mansor, 2012[266]	FNA gold standard
Marturano, 2020[267]	Excluded - Restricted to Thy3b on FNA
Merhav, 2021[269]	FNA gold standard for most (94%)
Miao, 2020[270]	Unclear gold standard
Mohammadi, 2013[272]	FNA gold standard
Mohanapriya, 2018[273]	Did not specify thresholds - left to discretion of US operator
Mohebbi, 2019[274]	FNA gold standard
Moon, 2007[276]	FNA gold standard
Moon, 2008[279]	FNA gold standard for some
Moon, 2010[277]	FNA gold standard for some

Reference	Reason for exclusion
Moon, 2012[278]	FNA gold standard for some
Na, 2016[280]	FNA gold standard for some
Nabahati, 2019[281]	FNA gold standard
Nam, 2016[282]	FNA gold standard for some
Ning, 2012[287]	FNA gold standard in some
Okamoto, 1994[289]	No thresholds provided
Okasha, 2018[290]	Detection of lymph node malignancy
Okasha, 2021[291]	FNA gold standard for most (86%)
Oliveira, 2018[292]	FNA gold standard for some
Orhan Soylemez, 2021[293]	Only a small subset of patients had surgical histopathology gold standard - for these the sensitivities and specificities for the different scales were given but the thresholds for positive/negative were not reported.
Ozel, 2012[294]	FNA gold standard for some
Pandey, 2017[297]	FNA gold standard for some
Pang, 2017[298]	FNA gold standard for some
Paredes-Manjarrez, 2021[299]	FNA gold standard for all
Park, 2009[302]	Staging study and all had papillary carcinoma
Park, 2014[304]	FNA gold standard
Park, 2015[301]	FNA gold standard in some
Pei, 2020[308]	FNA gold standard for some
Persichetti, 2018[309]	FNA gold standard for some
Petrone, 2012[310]	FNA gold standard in some
Polat, 2019[312]	FNA gold standard for some
Raggiunti, 2011[316]	FNA gold standard
Ramundo, 2020[319]	Evaluation of extrathyroidal extension
Rosario, 2005[324]	Differentiation of benign and metastatic lymph nodes
Rosario, 2015[325]	FNA gold standard for some
Rossing, 2012[326]	FNA gold standard for some
Rubaltelli, 2009[328]	Review (non-systematic)
Sahin, 2021 [329]	FNA gold standard for all
Sahli, 2019[330]	Patients already tested with FNAC - not relevant to our review population
Samir, 2015[331]	Patients already tested with FNAC - not relevant to our review population
Schenke, 2019[335]	No diagnostic accuracy analysis
Sebag, 2010[338]	FNA gold standard for some
Seifert, 2021[339]	FNA gold standard for some
Sengul, 2020[341]	Did not provide relevant diagnostic accuracy data
Seo, 2012[342]	Detection of recurrent disease
Shi, 2020[348]	FNA gold standard for some
Shi, 2021[347]	SR - references checked
Shin, 2020[351]	Evaluating machine learning rather than US

Reference	Reason for exclusion
Song, 2015[358]	Biopsy was gold standard
Song, 2021[359]	FNA or biopsy used as gold standard not surgical pathology
Songsaeng, 2019[360]	Unable to source
Studeny, 2021[363]	No diagnostic accuracy evaluation
Sultan, 2015[365]	FNA gold standard
Sun, 2020[366]	FNA gold standard for some
Sych, 2021[368]	FNA gold standard for some
Szczepanek-Parulska, 2020[369]	FNA gold standard for some
Tae, 2007[371]	Manuscript in Chinese
Tahmasebi, 2016[372]	FNA gold standard
Tan, 2021[374]	FNA gold standard
Tatar, 2014[377]	FNA gold standard for some
Tian, 2015[379]	SR - references checked
Tian, 2018[378]	FNA gold standard for some
Unluturk, 2012[382]	FNA gold standard for some
Vidal-Casariego, 2012[384]	FNA gold standard in some
Vinayak, 2012[385]	FNA gold standard for some
Walsh, 2012[387]	Excluded as this is an ex vivo study i.e., ultrasound (elastography) was performed on tissues, not people.
Wang, 2012[399]	No diagnostic accuracy analysis
Wang, 2013[390]	Core biopsy gold standard
Wang, 2015[394]	Hashimoto thyroiditis cohort
Wang, 2016[389]	Sample restricted to those with TIRADs of 4 or 5
Wang, 2019[395]	FNA gold standard for some
Wang, 2020[393]	FNA gold standard for some
Wang, 2020[396]	FNA gold standard in some
Weller, 2020[401]	FNA gold standard for some
Wettasinghe, 2019[402]	FNA gold standard in some
Woon, 2020[403]	FNA gold standard for some
Wu, 2016[406]	Same data as Wu, 2016 #687
Xia, 2017[407]	Incorrect index test, extreme learning approach
Xia, 2018[408]	Not a diagnostic accuracy study
Xiao, 2020[409]	Unclear gold standard ('histology' was described as the GS but this could have been FNA)
Xu, 2019[414]	FNA gold standard for some
Xu, 2019[415]	FNA gold standard for some
Xu, 2020[411]	Unclear thresholds used for CEUS
Xue, 2016[416]	Core biopsy gold standard
Yang, 2020[418]	SR - references checked
Yoo, 2021[419]	FNA / CNB gold standard for all
Yoon, 2016[420]	FNA gold standard for some

Reference	Reason for exclusion
Young, 2011[421]	Not evaluating US
Zhang, 2015[426]	FNA gold standard for some
Zhang, 2015[432]	FNA gold standard for some
Zhang, 2017[430]	FNA gold standard for some
Zhang, 2018[427]	FNA gold standard for some
Zhang, 2020[422]	All had thyroid malignancy
Zhang, 2020[429]	FNA gold standard for some
Zhang, 2021[428]	FNA gold standard for some
Zhao, 2018[434]	FNA gold standard for some
Zhao, 2019[435]	FNA gold standard for some
Zhao, 2021[436]	FNA gold standard for some
Zhu, 2021[437]	FNA gold standard for some

J.2 Clinical studies: Threshold of nodule size and classification

Table 19: Studies excluded from the clinical review

Study	Exclusion reason
Abbasian ardakani 2019[1]	Incorrect study design (cross sectional so no follow up period); no relevant outcomes
Ajmal 2015[12]	People given biopsy/FNA
Angell 2015[17]	People given biopsy/FNA
Arambewela 2020[19]	Unclear if participants received active surveillance/were discharged
Berker 2008[32]	No relevant outcomes
Brito 2016[40]	People given biopsy/FNA
Cohen 2017[68]	People given biopsy/FNA
Cohen 2019[69]	People given biopsy/FNA
Deveci 2007[78]	People given biopsy/FNA
Dong 2018[82]	People given biopsy/FNA
Flam, 2021 #1919	Review
Ghai, 2021 [107]	Review
Griffin 2017[120]	Incorrect population (patients had previously had thyroid cancer diagnosis)
Hayes, 2021[141]	FNA/biopsy carried out.
Hong 2012[148]	No relevant outcomes
Hong 2018[146]	People given biopsy/FNA
Hu 2021[153]	Participants did not receive active surveillance
Ibrahim 2015[163]	People given biopsy/FNA
Kaya 2019[177]	People given biopsy/FNA
Kim 2016[192]	Incorrect intervention/comparators
Koseoglu atilla 2018[200]	People given biopsy/FNA
Kuma 1992[201]	People given biopsy/FNA
Kuma 1994[202]	People given biopsy/FNA
Lai 2016[205]	No relevant outcomes
Lampung 2018[206]	No relevant outcomes
Lang 2017[207]	Incorrect population (participants had undergone surgery at baseline)
Lee 2013[211]	Incorrect population (participants already had diagnosis and had undergone surgery)
Lee 2017[212]	People given biopsy/FNA
Lin 2020[234]	Incorrect population (patients had undergone RFA)

Study	Exclusion reason
Liu 2016[244]	Incorrect population (participants had proven malignancy at baseline)
Lyu 2019[254]	People given biopsy/FNA
Oh 2019[288]	Incorrect population (participant already had a diagnosis of PTC)
Pompili 2013[313]	Incorrect population (participants had undergone thyroidectomy)
Sanabria 2020[332]	People given biopsy/FNA
Sengul 2019[340]	No relevant outcomes
Wang 2015[397]	Incorrect population (participants had already undergone thyroidectomy)
Xia 2018[408]	Incorrect population (participants had PTMC at baseline)
Xu 2017[413]	People given biopsy/FNA

J.3 Health Economic studies

Published health economic studies that met the inclusion criteria (relevant population, comparators, economic study design, published 2005 or later and not from non-OECD country or USA) but that were excluded following appraisal of applicability and methodological quality are listed below. See the health economic protocol for more details.

None.