National Institute for Health and Care Excellence

Final

Thyroid cancer: assessment and management

[D] Evidence review for diagnostic accuracy of fine needle aspiration cytology

NICE guideline NG230

Evidence reviews underpinning recommendations 1.2.8 to 1.2.10 in the NICE guideline

December 2022

Final

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 All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN: 978-1-4731-4863-5

Contents

1 Diagnosis of thyroid nodule malignancies

1.1 Review question

1.1.1 For people with thyroid nodules that require further investigation following ultrasound, what is the diagnostic accuracy of fine needle aspiration cytology (FNAC) with rapid on-site evaluation, FNAC without rapid on-site evaluation or core biopsy for diagnosing thyroid cancer?

1.1.2 Introduction

Fine needle aspiration cytology (FNAC) and core biopsy are highly valuable diagnostic methods for analysing the nature of a thyroid nodule and assess the need for surgical management. FNAC with rapid on site evaluation (ROSE) also known as rapid on-site assessment (ROSA) helps to provide an assessment of adequacy on-site, however, requires adequate staffing support and can limit the type of cytological preparation used (direct smear vs cytospin and cell block). Cellular cell block preparations form suitable material for immunohistochemistry and cytogenetic testing using fluorescence in-situ hybridisation (FISH). Core biopsy, whilst a more invasive procedure than FNAC, provides a tissue biopsy which can be used for diagnosis, potentially reduces the inadequacy rates and can be suitable material to perform thyroid fusion gene panel testing in addition to immunohistochemistry and FISH testing when required.

Current practice in the UK is to classify thyroid cytology using the RCPath modification of BTA classification which maps over to the Bethesda classification system. The different Thy categories has an expected positive predictive value for malignancy and the guidance also suggests accepted inadequacy rate (Thy1 category). This review seeks to determine the accuracy of FNAC and core biopsy for detecting thyroid cancer in people identified on ultrasound as needing further assessment.

1.1.3 Summary of the protocol

For full details see the review protocol in Appendix A.

Table 1: PIRO characteristics of review question

Population	Inclusion: People aged 16 or over suspected of thyroid cancer with potentially malignant nodules on ultrasound. Exclusion: Children and young people under 16 years. Population strata: 1) papers containing people selected for FNAC with prior US; 2) papers where people were given FNAC without prior US (or where there was no report of prior US)
Target conditions	nodules with thyroid cancer malignancy
Index test	• Fine-needle aspiration cytology (FNAC) without rapid on-site evaluation (ROSE) with smear without cytospin and cellblock • Fine-needle aspiration cytology (FNAC) without ROSE with Cytospin and cell block, without smear. • Fine-needle aspiration cytology (FNAC) without ROSE with smear, cytospin and cell block • Fine-needle aspiration cytology (FNAC) with ROSE (by cytopathologist or technician) and with smear without cytospin and cell block • Fine-needle aspiration cytology (FNAC) with ROSE (by cytopathologist or technician) and with smear with cytospin and cell block

	• Core biopsy
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.

1.1.4 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.

1.1.5 Diagnostic evidence

1.1.5.1 Included studies

148 eligible studies were found and included in the review.[1-4, 6-9, 18, 19, 23-25, 29, 30, 32, 38, 41, 42, 47, 50, 51, 53, 55, 61, 67, 69, 70, 72, 80, 85, 88, 90, 91, 97, 98, 106, 108, 115, 123, 126, 127, 131, 133, 138, 144, 149, 150, 152, 153, 155, 159-161, 163, 166-168, 175, 182, 187, 188, 193, 194, 196, 199, 200, 204, 206-208, 210, 211, 217, 221, 223, 224, 226, 229, 233, 236, 237, 239-242, 252, 256-258, 260, 261, 266, 267, 269, 275-278, 282, 284-287, 289, 295, 296, 298-301, 307, 309-312, 315-317, 327, 329, 330, 332, 334, 339, 342-345, 347, 353-355, 360-365, 372, 377, 378, 381, 385, 389-392] 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 23.

Sensitivity and specificity were the outcomes used in this review. Sensitivity was identified as the primary measure in guiding decision-making. The committee therefore set clinical decision thresholds for sensitivity of 0.95, above which a test would be recommended, and 0.85, below which a test would be deemed of no clinical use. They also set clinical decision thresholds for specificity of 0.8, above which a test would be recommended, and 0.7, below which a test would be deemed of no clinical use.

Although the question specifies a population that has been selected for FNAC on the basis of prior US findings, this review contains two strata: one without evidence of prior US-based selection and one with evidence of US-based selection. This broadening of the scope of the review was carried out pre-hoc because the committee envisaged that many otherwise useful papers would exist where evidence of prior US-based selection was absent. This proved to be the case, and the evidence has been separated for the two strata.

Collection of a number of 'unsatisfactory' or 'inadequate' results, where an insufficient number of cells for adequate testing were collected in an aspiration, were a feature of many studies. This is a common problem with FNAC testing, and failure to allow for this in the analysis of results will ignore an important aspect of test accuracy performance. In some studies attempts were made to repeat unsatisfactory tests, even if these involved prolonged periods of waiting such as several days or weeks, and in all studies the data that has been analysed has been the fullest dataset available. However in most studies unsatisfactory results remained. Unfortunately, the vast majority of studies completely ignored the unsatisfactory results in their accuracy analyses. In this review the main analysis has attempted to adjust for this failing by using an adjusted analysis. [322] This adjusted analysis accounts for unsatisfactory findings by designating unsatisfactory FNAC findings that turn out

to be malignant on pathology as false negatives and unsatisfactory FNAC findings that turn out to be benign on histopathology as false positives. The rationale is that an unsatisfactory finding cannot definitively indicate malignancy or benignity – therefore in a patient who is shown by the gold standard to have a malignant nodule the unsatisfactory reading should be regarded as unsupportive of that finding and can therefore legitimately be seen as a false negative; likewise in a patient who is shown by the gold standard to have a benign nodule the unsatisfactory reading should be regarded as unsupportive of that finding and can therefore legitimately be seen as a false positive. As well as being a rational approach this strategy also allows this review to demonstrate any accuracy advantages of the 'ROSE' strategy, where rapid on-site evaluation may enable repeat measures to be made immediately. If the inadequate results are ignored in the analysis then this removes the very feature that would lead to differences in accuracy performance between the two approaches: it is the inadequate results that reduce accuracy and their removal would create equipoise. This would eliminate any purpose for comparing strategies with and without ROSE.

On the other hand, it could be argued that the adjustment strategy may be a somewhat harsh approach given that in the clinical setting an unsatisfactory reading may be satisfactorily repeated at a later date (albeit in many cases, if a ROSE approach is not employed, at a significantly later date), which would alleviate the diagnostic problem caused by an unsatisfactory reading. Therefore a 'raw analysis', where no correction has been made for unsatisfactory results, has also been performed as a sensitivity analysis.

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.

Data were combined on the basis of any established FNAC classification approach being used, such as the Bethesda or Royal College of Pathologists (RCPath) approaches (see Table 3). For example, all studies evaluating the Bethesda approach were combined within their respective strata. However, many studies did not use established approaches and tended to use four broad generic classification types, which were not named. The first type has been classified as 'two way', where the study authors simply classified FNAC findings as malignant or benign (or with suitable synonyms such as positive and negative). The second type has been classified as 'three way', where findings were classified as malignant, suspicious and benign. The middle category might be described in different ways, but there were always three categories. The third type has been classified as 'four way' and findings would usually be classified as malignant, suspicious, indeterminate and benign. The final type has been designated 'five way' and findings would be classified as malignant, suspicious, with two grades of indeterminate and benign. This could be regarded as roughly equivalent to Bethesda grades VI, V, IV, III and II respectively. These four different types were combined separately. The rationale for keeping the types separate is explained as follows. If everyone can be classified as either malignant or benign in type one then this means that the same terms must differ in meaning in the other types (two, three and four) because everyone cannot be classified as solely malignant or benign in the other types. This means that some people who would be classified as, for example, malignant in the '2 way' type would not be so classified in the 3-way type. Because the terms have different meanings across types they must be analysed separately.

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.

1.1.5.2 Excluded studies

See the excluded studies list in Appendix I.

1.1.6 Summary of studies aiming to detect nodule malignancy

Table 2: Summary of studies included in the evidence review

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
Abboud, 2003[1]	Lebanon	46	Patients undergoing thyroidectomy who also had FNAC	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Abou-Foul, 2021[2]	UK	471	All patients who had thyroid resection (total or hemithyroidectomy) and FNAC	If final histology reported incidental malignant lesions that were not sampled during the FNAC, these reports were excluded from the analysis	U	Y	Fine needle aspiration cytology without ROSE, with smear only
Acar, 2017[3]	Turkey	226 nodules (pre-Bethesda) and 316 nodules (Bethesda)	Patients undergoing total thyroidectomy for thyroid nodules, with FNAC pre-Bethesda or post-Bethesda inception	Not reported	U	Y	Fine needle aspiration cytology without ROSE, with smear only
Afroze, 2002[4]	Pakistan	170	Patients undergoing FNAC of thyroid nodules and subsequent thyroid surgery	Patients without computerised records or operated on outside study hospital	U	Y	Fine needle aspiration cytology with ROSE, with smear + cytospin and cell block
Agcaoglu, 2013[6]	Turkey	730	Prior US, otherwise not reported	Non-diagnostic results	Y	Y	Fine needle aspiration cytology with ROSE, with smear only (cytopathologist attended in 77% of FNAB procedures)
Aggarwal, 1989[7]	Unclear	36	Patients with ultrasonographically solitary cold thyroid nodules given FNAC and subsequent surgery	Not reported	Y	U	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
Agrawal, 1995 #1093[8]	India	100	Patients for whom FNAC and post-surgical pathology were available	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Aguilar-Diosdado, 1997[9]	Spain	289	Patients undergoing resection for nodular goitre; carcinoma or suspicious on FNAC; thyroid nodule associated with lymphadenopathy; thyroid nodule associated with previous radiation exposure; enlargement of a thyroid mass despite L-thyroxine therapy; clinical symptoms of hoarseness or dysphagia in patients with thyroid nodules [despite specific FNAC findings being an indication for surgery, the fact that most people being sent to surgery had benign FNAC findings meant this paper was deemed acceptable for inclusion].	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear + cytospin + cell block
Al-Hureibi, 2003[18]	Yemen	199	Patients undergoing FNAC and subsequent thyroid surgery for thyroid nodules/swelling.	Not reported	U	N	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
Altavilla, 1990[23]	Italy	257	Not reported	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Al-Taweel, 1990[19]	Kuwait	91	Consecutive patients undergoing FNAC for solitary thyroid nodules with subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Ananthakrishnan, 1990[24]	India	150	consecutive patients with a single palpable nodule in thyroid for whom FNAC and histopathology were performed	No histopathology available	U	U	Fine needle aspiration cytology without ROSE, with smear only
Anderson, 1987[25]	UK	373	Not reported	Not reported	U	N	Fine needle aspiration cytology without ROSE, with smear only
Arul, 2015[29]	India	392	All the FNACs of thyroid lesions between July 2012 and January 2015 were retrieved retrospectively; surgical histopathology obtained; FNAC classified according to 6 tier TBSRTC	No histopathology results	U	U	Fine needle aspiration cytology without ROSE, with smear only
Aydogan, 2019[30]	Turkey	514	Patients undergoing thyroidectomy after FNAC; decision for surgery depended on nodule size, malignant or indeterminate cytology, compressive symptoms, Graves disease and multinodular goitre [adequate number of benign on FNAC	Not reported	U	Y	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			to allow inclusion to this review].				
Bahaj, 2021[32]	Saudi Arabia	314	Patients undergoing FNAC and thyroid surgery	Not reported	U	Y	Fine needle aspiration cytology without ROSE, with smear only
Bashier, 1996[38]	Sudan	89	Patients with a solitary or significantly dominant thyroid nodule, followed up by histopathological confirmation	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Belanger, 1983[41]	Canada	63	Presence of a solid or partially cystic cold nodule; informed consent for surgery regardless of cytological findings; no surgical contraindications	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Bellantone, 2004[42]	Italy	119	Patients undergoing UG FNAC and subsequent surgery because of suspicious or malignant cytology, persistently nondiagnostic cytology, cytology consistent with predominantly follicular lesion, incomplete cyst resolution, compressive symptoms and/or large nodular size	Not reported	U	Y	Fine needle aspiration cytology without ROSE, with smear + cytospin + cell block.
Biscotti, 1995[47]	USA	41	FNAC specimens from patients who also provided a histopathological sample at surgery	Not reported	U	U	1. Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
							2. Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block – Thin-prep
Bodo, 1979[50]	Hungary	131	Patients with diffuse enlargement of the thyroid gland, given FNAC and surgery. No reasons given for surgery, but most given surgery were negative on FNAC, so FNAC not the only criterion.	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Borman, 1995[51]	USA	27	Patients with thyroid nodules undergoing FNAC with subsequent surgery. Surgery was given if indicated by FNAC, or if there were compression symptoms, a recurrent cyst or other clinical suspicion in the presence of benign FNAC findings. [Because there were almost half of all cases made up of benign FNAC cases this study has been included in the review.]	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Brauer, 1984[53]	USA	134	Patients undergoing FNAC for thyroid nodules with subsequent surgery. Majority had	Not reported	N	Y	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			hypofunctioning solitary nodules. Initially surgery was given to all patients regardless of FNAC. As the study progressed benign findings were less likely to be referred. [However, overall the number of benign FNAC findings sent to surgery is sufficient for inclusion to this review]				
Bugis, 1986[55]	Canada	198	Patients presenting with a solitary nodule, with FNAC and subsequent surgery.	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Can, 2008[61]	Turkey	23 (USG) and 18 (non USG)	All consecutive patients who underwent FNAC of thyroid nodules, followed by surgery	No surgery performed (note that this is an exclusion criterion for the data included here but was not an exclusion criterion for the study that also looked at data from patients who did not have surgery)	U	USG for 23 and non-USG for 18	Fine needle aspiration cytology without ROSE, with smear only
Chang, 1997[67]	China	662	Patients undergoing FNAC and surgery for thyroid nodules. Surgery indicated for those with a malignant or indeterminate result. Those with a benign result only underwent surgery in cases of a rapidly	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			growing nodule, local compression or cosmetic reasons.				
Choden, 2021[69]	Bhutan	81	Patients undergoing FNAC who also underwent surgical resection	Patients with missing data	U	U	Fine needle aspiration cytology without ROSE, with smear only
Choe, 2018[70]	South Korea	705	Patients undergoing core needle biopsy, with subsequent surgery. Reasons for surgery not given. [Some going to surgery had benign CNB results so CNB results were not sole criterion].	Not reported	Y	N	Core biopsy
Chow, 1999[72]	Hong Kong	76	Patients with non-toxic solitary thyroid nodules or predominant nodules in non-toxic nodular goitre who underwent surgery with prior FNAC. Benign FNAC findings were not routinely sent for surgery unless they increased in size of the patients requested surgery – however most of those referred for surgery were benign on FNAC.	Not reported	N	N	Fine needle aspiration cytology without ROSE, with smear only
Cristallini, 1989 #1161[80]	Italy	41	Patients undergoing thyroidectomy with prior FNAC	Toxic nodules	U	N	Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block
Danese, 1998[85]	Italy	535	Consecutive patients with single	Not reported	U	USG and no USG	Fine needle aspiration cytology without ROSE,

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			or multiple thyroid nodules given either conventional or UG FNAC, followed by surgery.				with smear + cytospin and cell block.
Davidsohn, 1995[88]	USA	50	Patients having an FNAC for thyroid nodules with subsequent thyroidectomy. If FNAC was benign surgery would still be given because of large nodules, patient preference or for cosmetic reasons	Not reported	U	U	Fine needle aspiration cytology with ROSE, with smear + cytospin and cell block
de Roy van Zuidewijn, 1994[90]	Holland	265	Patients undergoing FNAC and thyroidectomy	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
de Vos tot Nederveen Cappel, 2001[91]	Holland	254	Patients with FNACs carried out for thyroid nodules followed by thyroid surgery. People benign on FNAC were eligible for surgery if they had a rapidly growing nodule causing local compression, or due to cosmetic reasons	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Dwarakanathan, 1989[97]	USA	63	Patients undergoing FNAC and subsequent surgery for single nodules or multinodular goitres with a dominant nodule. Most nodules were cold on scan. Surgery was given for benign FNAC	Not reported	U	N	Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			findings for reasons of patient preference, cosmetic considerations, large goitres, large nodules, and other clinically worrisome features such as the age of the patient or male sex (n=26). This ensured all of the FNAC categories were covered in the study.				
El Hag, 2021[98]	Saudi Arabia	323	All thyroid FNAs with histopathology follow up	Not reported	U	Y	Fine needle aspiration cytology with ROSE, with smear only
Ferrari, 1985[106]	Italy	68	Patients with cold nodules undergoing FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block
Fiorentino, 2021[108]	Italy	693	Patients with FNAC and surgical specimens	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Francis, 1999[115]	Kuwait	45	Patients attending thyroid unit for FNA	Not meeting criteria for FNAC; aspirated cervical lymph nodes	U	U	Fine needle aspiration cytology without ROSE, with smear only
Gardiner, 1986[123]	Canada	207	Patients given FNAC for diffuse thyroid enlargements, multinodular thyroids and thyroids with discrete nodules; subsequent surgery	Not reported	N	U	Fine needle aspiration cytology without ROSE, with smear only
Gershengorn, 1977[126]	USA	33	Fifty consecutive patients presenting with discrete usually single thyroid nodules	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			given FNAC and surgery				
Giansanti, 1989[127]	Italy	114	Patients with solid, cold, thyroid nodules, with FNAC and subsequent surgery.	Not reported	U	N	Fine needle aspiration cytology without ROSE, with smear only
Gossain, 1998[131]	USA	19	Patients with a single palpable nodule, undergoing FNAC followed by surgery	Not reported	U	N	Fine needle aspiration cytology without ROSE, with smear only
Gould, 1989[133]	USA	69	People with thyroid nodules with an FNAC, touch imprint and final histopathology	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Guo, 2015[138]	China	489	All thyroid FNAs that were followed by surgery; indications for FNAC were palpable nodules with US finding suggesting malignancy such as microcalcification, margin irregularity, intranodular vascularity or taller than wide shape	Not reported	Y	Y (for 79%)	Fine needle aspiration cytology without ROSE, with smear only
Hamming, 1990[149]	Holland	169	Patients with nodular thyroid disease given FNAC and subsequent surgery. Surgery performed to confirm or exclude a malignant neoplasm or to remove a nodular goitre for cosmetic	Not reported	U	N	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			or mechanical reasons.				
Haberal, 2009[144]	Turkey	260	Adequate FNAC followed by thyroidectomy or lobectomy for a dominant thyroid nodule	Not reported	U	Y	Fine needle aspiration cytology without ROSE, with smear only
Hamming, 1998[150]	Holland	240	Patients operated on for nodular thyroid disease with an evaluable FNAC	non-evaluable smears – insufficient material for cytodiagnosis.	U	U	Fine needle aspiration cytology without ROSE, with smear only
Hawkins, 1987[153]	Spain	415	Patients referred to endocrinology unit because of diffuse or nodular goitres, with or without symptoms; surgery (in patients with positive or suspicious FNAB cytology and/or suggestive clinical histories, and in patients with cold thyroid nodules and negative FNAB results that did not respond to 6 months of suppressive thyroxine therapy	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block. Unclear in description but stated that 'if fluid was drawn the centrifuged sediment was studied', indicating that at least cytospin was used in addition to smear.
Harsoulis, 1986[152]	Greece	213	Patients with a solitary or dominant thyroid nodule within either a multinodular or diffusely enlarged gland who were subsequently given surgery. Surgery was indicated by FNAC but also by the recent	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			appearance of a cold solid nodule, a history of recurrent cysts and for all male patients				
Heimann, 1964[155]	Unclear	23	Patients undergoing FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Hosokawa, 2019[159]	Japan	685	Patients undergoing FNAC and surgery on thyroid nodules	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Hougaard Chakera, 2003[160]	Denmark	67	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Huang, 2020[161]	China	392	1. Thyroid nodules with 1~4 of the following five suspicious ultrasonic features - “solid nodules, hypoechoic or extremely hypoechoic, irregular boundary, microcalcification, taller-than-wide shape” - based on the classification standard of TI-RADS proposed by Kwak et al; 2. Conventional thyroid ultrasonography, ultrasound elastography and FNAC performed before surgery; and 3. Cytologic results as well as a final diagnosis of the nodules based on	1. Surgery for hyperthyroidism; 2. Previous history of neck radiation or surgery; and 3. Thyroid nodules that do not meet the standard of KWAK-TIRADS.	Y	N	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			postoperative pathology.				
Hussain, 1993[163]	UK	108	Patients identified by radionuclide imaging as having a solitary cold thyroid nodule, who had FNAC followed by surgery; surgery carried out on all patients with a solitary cold nodule	Not reported	U	U	Fine needle aspiration cytology with ROSE, with smear + cytospin and cell block
Jalan, 2017[166]	India	40	All patients with complaints of thyroid swelling [for this review, surgery]	Not reported	U	USG and non-USG done in 22, but not the majority. Non-USG done in the other 18	Fine needle aspiration cytology without ROSE, with smear only
Jat, 2019[167]	Saudi Arabia	75	All patients came in OPD with clinically diagnosed as a solitary thyroid nodule having no hyper or hypothyroidism, irrespective of age and sex; thyroid surgery	patients presenting with extra-thyroid neck swelling; patients having toxic or non- toxic diffuse or multinodular goitre	U	Y	Fine needle aspiration cytology with ROSE, with smear only
Jayaram, 1999[168]	Malaysia	325	Patients with thyroid lesions given FNAC and thyroid surgery	Not reported	N	U	Fine needle aspiration cytology with ROSE, with smear only
Kelman, 2001[175]	USA	109	Patients presenting with a thyroid nodule, who were given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Kim, 2013[182]	South Korea	200	Patients with thyroid nodules with a >90% solid component with maximum diameter of 5mm; underwent FNAC and surgery	Not reported	U	Y	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
Kimoto, 1999[187]	Japan	61	Not reported	Not reported	Y	Y	Fine needle aspiration cytology without ROSE, with smear only
Kini, 1985[188]	USA	379	Patients with thyroid nodules undergoing FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Kojic Katovic, 2004[193]	Croatia	80	Patients with complete pre-operative investigations for thyroid nodules (US, IS, FNA) and subsequent histopathological diagnosis	Not reported	Y	Y	Fine needle aspiration cytology without ROSE, with smear only
Kolendorf, 1975[194]	Denmark	20	Patients admitted for thyroid disorders, given FNAC and open surgical biopsy	Not reported	N	N	Fine needle aspiration cytology without ROSE, with smear only
Kothari, 2019 #1269 [196]	India	53	Not reported	Not reported	U	U	Fine needle aspiration cytology with ROSE, with smear only
Kumar, 1992[199]	India	86	consecutive patients with solitary nodules undergoing FNAC and subsequent surgery	Not reported	U	N	Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block
La ROSE, 1991[200]	Italy	827	Cold thyroid nodules examined with FNAC that were given subsequent surgery. Surgery was offered to those to those that were malignant or highly suspicious on FNAC;	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			probable adenoma were suggested to undergo surgery. 'Benign' or 'inadequate' nodules were also given surgery if there was clinical suspicion or through patient choice. [Thus although there was some bias in the access to surgery, there was definite access from all FNAC categories, allowing a reasonably valid assessment of accuracy to be made].				
Leenhardt, 1999[204]	France	94	Consecutive patients with thyroid nodules referred for FNAC after US; non palpable nodules. Surgery provided for a histopathological diagnosis. Surgery was offered to those to those that were malignant or suspicious on FNAC; supracentrimetric or isolated cold nodules; simultaneous presence of a palpable nodule in a multinodular gland and miscellaneous	Not reported	Y	Y	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			reasons. [Thus, although there was some bias in the access to surgery, there was definite access from all FNAC categories, allowing a reasonably valid assessment of accuracy to be made].				
Li, 2021[207]	China	623	Patients having FNAC and thyroid surgery	No report on the sensation during puncture of the nodule – whether 'soft', 'hard' or 'hard with grittiness'	U	Y	Fine needle aspiration cytology without ROSE, with smear only
Li, 2013[206]	China	51	Patients with suspected solid thyroid nodules, later given US guided biopsy and a histopathological confirmation after, presumably, surgery.	Patients hyper-susceptible to SonoVue or with coagulation dysfunction were excluded	U	Y	Core biopsy with US guidance Core biopsy with CEUS guidance
Liel, 1985[208]	Israel	49	Patients with 'cold' or 'warm' thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block
Lioe, 1998 #1280[210]	UK	67	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Liu, 2009[211]	Taiwan	40	Patients with auto-immune thyroiditis; hypothyroidism or hyperthyroidism with thyroid nodules; given	Diffuse thyroid disorders	U	U	Fine needle aspiration cytology with ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			FNAC with subsequent surgery				
Lukitto, 1998[217]	Indonesia	167	Patients with thyroid nodules undergoing FNAC and surgery. Indications for surgery not provided. Out of 250, 167 went for thyroidectomy, and 162 of these were 'negative' on FNAC, so it seems that the decision was not based on FNAC. Therefore this study has been included.	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Mamoon, 1997[221]	Pakistan	176	Patients undergoing FNAC and subsequent surgery for thyroid nodules	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Mandal, 2011[223]	India	108	Patients with nodular thyroid disease given FNAC followed by surgery	Diffuse goitre, debilitated elderly, other comorbidities making the patient unfit for surgery	N	N	Fine needle aspiration cytology without ROSE, with smear only
Mandreker, 1995[224]	India	238	Patients presenting with a diffuse or nodular thyroid enlargement and solitary thyroid nodule; FNAC and subsequent surgery carried out	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Maruta, 2003[226]	Japan	304	Thyroid nodule aspirations from a database where people has also had thyroid surgery	Not reported	U	Y	Fine needle aspiration cytology without ROSE, with smear only
Mastorakis, 2014[229]	Greece	1000	Patients with thyroid nodules given FNAC and	Not reported	N	Y	Fine needle aspiration cytology without ROSE,

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			subsequent surgery; surgery given on basis of FNAC results but also regardless of cytology – upon basis of other criteria such as multinodular lesions, nodule size or a lack of response to treatment or patient decision.				with smear + cytospin and cell block
McElroy, 2014[233]	USA	28	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block
Mehrotra, 2006[236]	UK	450	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	USG for 102; no USG for 348	Fine needle aspiration cytology without ROSE, with smear only
Meko, 1995[237]	USA	90	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	Y	Fine needle aspiration cytology with ROSE, with smear + cytospin and cell block
Merchant, 1995[239]	UK	56	Patients with thyroid nodules or diffuse thyroid enlargement given FNAC and subsequent surgery; surgery given secondary to cytology, clinical signs or evidence from second line investigations.	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Mijovic, 2009[240]	Canada	115	Consecutive patients undergoing thyroidectomy for cytologically proven malignancy or nodules suspicious	Not reported	U	N	Fine needle aspiration cytology without ROSE, with smear only AND some (unspecified number) were:

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			for being malignant (e.g. history of radiation exposure, family history, size and so on); surgery also performed on patients with Graves disease, large goitres and compression symptoms with FNAC performed pre-op.				Fine needle aspiration cytology without ROSE, with smear + cell block. The paper stated that: 'all cases had at least a smear stained with Papanicolaou, and, if enough material was available, a smear stained with Diff quick and a cell block was performed'
Mikosch, 2000[241]	Austria	708	Patients with thyroid nodules given FNAC and subsequent surgery; FNAC indicated by patients with hypoechoicity, irregular margins. microcalcifications US, growth of the nodule during follow up or hypofunctional nodules on scintiscan; reasons for surgery included cytological findings or obstructive reasons	Not reported	Y	Y	Fine needle aspiration cytology without ROSE, with smear only
Miller, 1979[242]	USA	147	Patients with discrete thyroid nodules given FNAC and subsequent surgery	Functional nodules and cystic nodules without appreciable residual after aspiration of fluid	U	U	Fine needle aspiration cytology without ROSE, with smear only
Munn, 1988 #1322[252]	USA	49	Patients with palpable thyroid nodules given FNAC and subsequent surgery	History of radiation exposure; family history of medullary carcinoma	U	U	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
Nagarajan, 2015 #1326[256]	USA	1320	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block
Nart, 2010 #1327[257]	Turkey	291	Patients with FNAC followed up with surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Natarajan, 1994[258]	India	25	Patients with solitary cold thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Naz, 2014[260]	Pakistan	61	Patients presenting with thyroid swelling, undergoing FNAC. For this review only those sent for surgery were included, but no rationale for surgery given; however it appears that those sent for surgery represented all gradings of the FNAC.	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear + cell block.
Ng, 1988 #1330[261]	Singapore	46	Patients with solitary thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Okumura, 1999 #1334[266]	Japan	109	Patients with thyroid nodules that were given FNAC and surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Ongphiphadhanakul, 1992 #1335[267]	Thailand	129	Patients with solitary thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
Ozdemir, 2017[269]	Turkey	1810 nodules (pre Bethesda) and 5115 nodules (post-Bethesda)	Patients with thyroid nodules given FNAC and subsequent surgery	Age <16 years; previous history of thyroid surgery or percutaneous invasive procedures to thyroid nodules; radiotherapy to head and neck	Y	Y	Fine needle aspiration cytology without ROSE, with smear only
Pepper, 1989[275]	USA	21	Patients with thyroid nodules given FNAC and subsequent surgery; surgery given because of FNAC findings or because of personal choice or because of nodule growth despite levothyroxine treatment	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block
Petersen, 1984[276]	Denmark	189	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block
Piana, 2011[277]	Italy	2047	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	Y	Y	Fine needle aspiration cytology without ROSE, with smear only
Pisani, 2000[278]	Italy	42	Consecutive patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	USG for both FNAC and CNB	Fine needle aspiration cytology without ROSE, with smear only Core biopsy
Prinz, 1983[282]	USA	109	Patients with palpable nodules hypo-functioning on thyroid scintiscan; subsequent thyroidectomy	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
Radetic, 1984[284]	Croatia	2190	Patients with thyroid goitres given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Raina, 2011[285]	India	25	Patients with thyroid nodules receiving FNAC [in review, only those confirmed by histopathology were included, but in paper there were additionally also 71 not sent for surgery. Reasons not given but FNAC results not the only reasons as half sent for surgery were benign on FNA]	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Rammeh, 2019 #1349[286]	Tunisia	64	Patients with palpable thyroid nodules given FNAC and subsequent surgery	Not reported	U	N	Fine needle aspiration cytology without ROSE, with smear only
Rana, 2021[287]	India	445	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Rege, 1987[289]	India	182	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Rodriguez, 1994[295]	Spain	170	Patients with solitary or dominant thyroid nodules given FNAC and subsequent surgery	inadequate samples	U	U	Fine needle aspiration cytology without ROSE, with smear only
Rosen, 1993[296]	Canada	41	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	Y	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
Rosen, 1981[298]	Canada	153	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Roy, 2019[299]	India	112	Patients over 15 years; euthyroid state on blood examination; presenting with clinical evidence of thyroid disease and swelling	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Rubenfeld, 1982[300]	USA	30	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology with ROSE, with smear + cytospin and cell block
Russ, 1978[301]	USA	29	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Schmid, 1986 #1370[307]	Austria	2709	Patients with cold or multinodular thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Schoedel, 2008 #1372[309]	USA	46	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	Y	Fine needle aspiration cytology without ROSE, with smear only
Schwartz, 1982 #1373[310]	USA	102	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Sclabas, 2003[311]	USA	240	Patients undergoing FNAC with or without US guidance; thyroidectomy	Not reported	Y	U (USG for some but not a majority)	Fine needle aspiration cytology WITH ROSE, with smear + cytospin and cell block
Scurry, 2000[312]	Australia and Canada	109	Patients with thyroid nodules given direct smear or smear/cytospin	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only OR

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			FNAC and subsequent surgery				Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block [cell-block not mentioned]: cytospin preparations were made in cases that yielded cyst fluid.
Settakorn, 2001[316]	Thailand	415	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Seya, 1990[317]	Japan	26	Patients with thyroid nodule examined using FNAC and given surgery. 64 did not receive surgery but reasons not given - however out of those going to surgery half were benign on FNAC so it does not seem that FNAC result was the only criterion for surgery.	Not reported	U	N	Fine needle aspiration cytology without ROSE, with smear only
Silverman, 1986[327]	USA	8	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Sirpal, 1996[329]	India	128	Patients with thyroid nodules given FNAC and subsequent surgery. Surgery contemplated where FNAC showed malignancy, follicular or HC tumour, cosmetically unacceptable	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			cases, compression symptoms or cases non-responsive to therapy.				
Slowinska-Klencka, 2008[330]	Poland	1694	Patients referred from outpatients clinics for US and then FNAB and thyroidectomy	Not reported	Y	Y	Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block
Seok, 2018[315]	South Korea	457	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Son, 2014[332]	South Korea	694	Patients undergoing total or hemithyroidectomy and also FNA	Not reported	U	Y	Fine needle aspiration cytology without ROSE, with smear only
Spiliotis, 1992 #1394[334]	Greece	201	Patients with thyroid nodules given FNAC and subsequent surgery	Toxic nodules	U	U	Fine needle aspiration cytology without ROSE, with smear only
Sukumaran, 2014[339]	India	248	Series of cases of thyroid nodules with underwent FNAC followed by surgery	Those not given surgery [although the majority having surgery were malignant or suspicious on FNAC there were a sufficient number that were benign to ensure that category was represented]	U	U – USG done only in some (non majority)	Fine needle aspiration cytology without ROSE, with smear only
Tabain, 2004[342]	Croatia	457	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Tabaqchali, 2000[343]	UK	302	patients with a dominant thyroid nodule who had FNAC carried out in the 6 year period 1990-1995 and subsequent partial	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
			or complete thyroidectomy.				
Takashima, 1994[344]	Japan	133	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	USG and no USG	Fine needle aspiration cytology without ROSE, with smear only
Takashima, 1992[345]	Japan	41	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	USG and no USG	Fine needle aspiration cytology without ROSE, with smear only
Tal, 1992[347]	USA	30	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Theoharis, 2013 #1410[353]	USA	372 nodules (pre Bethesda) and 379 nodules (post Bethesda implementation)	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Theoharis, 2009 #1411[354]	USA	378	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	Y	Fine needle aspiration cytology without ROSE, with smear only
Thomas, 1998[355]	Nigeria	93	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Tsou, 1997 #1417[360]	Taiwan	61	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Varhaug, 1981 #1418[361]	Norway	264	Patients with thyroid nodules given FNAC and subsequent surgery	Diffuse goitre and toxic goitre	U	U	Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block
Vojvodich, 1994[362]	Canada, UK	98	Patients with solitary thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
Walsh, 1983[363]	Australia	76	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Wang, 2020[364]	China	274	Patients undergoing US, FNAC and thyroidectomy	History of thyroid surgery; thyroid metastasis; surgically removed nodules that were not one-to-one matched with the US findings	Y	Y	Fine needle aspiration cytology without ROSE, with smear only
Wei, 2016[365]	China	78	Patients with suspicious thyroid nodules, diagnosed with FNAC and given surgery	Not reported	U	Y	Fine needle aspiration cytology without ROSE, with smear, combined with thin-prep cytology test, which uses a filtration process and thin-layer deposition of cells [appears similar to cytospin].
Wu, 2006[372]	China	401	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Xiong, 2019[377]	China	578	Patients with thyroid nodules treated at Peking University First Hospital from January 2015 to December 2017 were reviewed. Cases of thyroid follicular lesions with both CNB and resected specimens were retrieved	Not reported	U	U	Core biopsy
Xu, 2014[378]	China	945	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	Y	Y	Fine needle aspiration cytology without ROSE, with smear only

Study	Country	Sample size	Inclusion criteria	Exclusion criteria	Prior US used to select patients?	Use of ultrasound guidance?	FNAC strategy
Yavuz, 2020 #1436[381]	Unclear	34	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	Y	Fine needle aspiration cytology without ROSE, with smear only
Yoder, 2006[385]	USA	200	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	Y (81%)	Fine needle aspiration cytology with ROSE, with smear only
Zajdela, 1987 #1442[389]	France	372	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block
Zbar, 2009[390]	Barbados	63	Patients with thyroid nodules given FNAC and subsequent surgery	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Zelmanovitz, 1998[391]	Brazil	11	FNAC and thyroidectomy	Not reported	U	U	Fine needle aspiration cytology without ROSE, with smear only
Zhang, 2015[392]	Unclear	78	Thyroid nodules undergoing FNAC and subsequent thyroidectomy	Not reported	U	Y	Fine needle aspiration cytology with ROSE, with smear only

See Appendix D for full evidence tables

1.1.7 FNAC scales used

Table 3: Summary of the types of established FNAC scales used.

Scale name	Description and scoring
Bethesda	I = non-diagnostic or inadequate; II = benign; III = atypia/follicular lesion of undetermined significance; IV = follicular neoplasm or suspicious for follicular neoplasm; V = suspicious for malignancy; VI = malignant
Aspiration Cytology Grade (AC)	AC0/1 = unsatisfactory; AC2 = non-neoplastic; AC3 = equivocal; AC4 = suspicious; AC5 = diagnostic of malignancy
British Thyroid Association (BTA)	THY1 = non diagnostic/cyst; THY2 = non-neoplastic; THY3 = follicular/ suspected follicular neoplasm; THY4 = suspicion of malignancy (non diagnostic); THY5 = malignancy (diagnostic)
Royal College of Pathologists	Thy 1/Thy 1= non-diagnostic for cytological diagnosis; Thy 2/Thy 2c= non-neoplastic; Thy3a/Thy3f = neoplasm possible; Thy4 = suspicious of malignancy; Thy5 = malignant
Piana C1-5	C1 = non diagnostic; C2= benign; C3 = indeterminate; C4= suspicious; C5 = malignant
De May	inadequate, non-malignant, non-malignant follicular proliferation, suspicious for malignancy, malignant

1.1.8 Summary of the evidence – adjusted evidence

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 to Table 13 provide results using an adjusted analysis. This adjusted analysis accounts for unsatisfactory findings (which are otherwise ignored by the majority of studies in their analyses) and designates unsatisfactory FNAC findings that turn out to be benign on histopathology as false positives and unsatisfactory FNAC findings that turn out to be malignant on pathology as false negatives. This follows the logic that an unsatisfactory finding cannot definitively indicate benignity or malignancy – therefore in a patient who is shown by the gold standard to have a benign nodule the unsatisfactory reading should be regarded as unsupportive of that finding and is therefore legitimately a false positive; likewise in a patient who is shown by the gold standard to have a malignant nodule the unsatisfactory reading should be regarded as unsupportive of that finding and is therefore legitimately a false negative.

Table 4: Summary of evidence relating to FNAC used without ROSE, with smear only, in the stratum where US was not used to select patients (adjusted analysis).

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Bethesda Grade III or above	13	5,950	Pooled sensitivity (95% credible intervals): 0.9288(0.888-0.957)	Pooled specificity (95% credible intervals): 0.6268(0.509-0.730)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
Bethesda Grade IV or above	13	6,434	Pooled sensitivity (95% credible intervals): 0.8559 (0.7855-0.9078	Pooled specificity (95% credible intervals): 0.7864 (0.6961-0.8567)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
Bethesda Grade V or above	16	7,082	Pooled sensitivity (95% credible intervals): 0.771 (0.6996-0.8299)	Pooled specificity (95% credible intervals): 0.9214(0.8797-0.9506)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
Bethesda Grade VI	12	5,748	Pooled sensitivity (95% credible intervals): 0.4927 (0.607-0.6462)	Pooled specificity (95% credible intervals): 0.93(0.8805-0.9618)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	none[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
BTA THY 3a or above	2	579	0.90 [0.73, 0.98] 0.50 [0.40, 0.59]	0.85 [0.75, 0.92] 0.46 [0.41, 0.52]	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
BTA THY 3f or above	1	471	0.38 [0.29, 0.47]	0.56 [0.51, 0.61]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
BTA THY 4 or above	1	471	0.20 [0.13, 0.29]	0.62 [0.56, 0.67]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
BTA THY 5	2	579	0.60 [0.41, 0.77] 0.06 [0.02, 0.12]	1.0 [0.95, 1.00] 0.62 [0.57, 0.67]	**Sensitivity**				
					Very serious[a]	serious[c]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[c]	serious[c]	very serious[d]	VERY LOW
	3	627			**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
AC 3 or above			Pooled sensitivity (95% credible intervals): 0.7798 (0.497-0.928)	Pooled specificity (95% credible intervals): 0.271(0.097-0.567)	Very serious[a]	serious[b]	none[d]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[d]	none[d]	VERY LOW
AC 4 or above	3	627	Pooled sensitivity (95% credible intervals): 0.396 (0.165-0.687)	Pooled specificity (95% credible intervals): 0.705(0.385-0.904)	**Sensitivity**				
					Very serious[a]	serious[b]	none[d]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[d]	very serious[d]	VERY LOW
2 way: malignant v benign	13	1,108	Pooled sensitivity (95% credible intervals): 0.8174 (0.6714-0.9132)	Pooled specificity (95% credible intervals): 0.9507(0.8961-0.98)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
2 way: malignant v benign - sub-grouped for ultrasound guided	4	464	Pooled sensitivity (95% credible intervals): 0.9221 (0.728-0.9887)	Pooled specificity (95% credible intervals): 0.892(0.733-0.973)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c,e]	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
2 way: malignant v benign - sub-grouped for non-ultrasound guided	9	644	Pooled sensitivity (95% credible intervals): 0.7385 (0.5802-0.8848)	Pooled specificity (95% credible intervals): 0.9703 (0.919-0.991)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c,e]	none[d]	VERY LOW
3 way: suspicious or malignant (negative =benign)	52	11,387	Pooled sensitivity (95% credible intervals): 0.860 (0.8196-0.895)	Pooled specificity (95% credible intervals): 0.734(0.666-0.793)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[d]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[d]	serious[d]	VERY LOW
3 way: malignant (negative = suspicious or benign)	45	10,456	Pooled sensitivity (95% credible intervals): 0.589 (0.524-0.652)	Pooled specificity (95% credible intervals): 0.941(0.916-0.961)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[d]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[d]	none[d]	VERY LOW
4 way: malignant or suspicious or indeterminate (negative = benign)	12	2,255	Pooled sensitivity (95% credible intervals): 0.852 (0.720-0.933)	Pooled specificity (95% credible intervals): 0.606(0.404-0.778)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[d]	serious[b]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[d]	serious[b]	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
4 way: malignant or suspicious (negative = benign or indeterminate)	14	2,253	Pooled sensitivity (95% credible intervals): 0.6697 (0.492-0.816)	Pooled specificity (95% credible intervals): 0.874(0.798-0.927)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[d]	none[b]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[d]	serious[b]	VERY LOW
4 way: malignant (negative = benign or indeterminate or suspicious)	12	2,244	Pooled sensitivity (95% credible intervals): 0.3975 (0.224-0.589)	Pooled specificity (95% credible intervals): 0.970(0.930-0.990)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[d]	none[b]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[d]	none[b]	VERY LOW
5 way: malignant or suspicious or two grades of indeterminate (negative = benign)	6	2,063	Pooled sensitivity (95% credible intervals): 0.8762 (0.739-0.948)	Pooled specificity (95% credible intervals): 0.433(0.310-0.567)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[d]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[d]	none[b]	VERY LOW
5 way: malignant or suspicious or one grade of indeterminate (negative = lower grade of	5	1,954	Pooled sensitivity (95% credible intervals): 0.799 (0.6338- 0.9009)	Pooled specificity (95% credible intervals): 0.656(0.3815-0.864)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[d]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[d]	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
indeterminate or benign)									
5 way: malignant (negative = suspicious or two grades of indeterminate or benign)	6	2,071	Pooled sensitivity (95% credible intervals): 0.5631 (0.4037-0.7079)	Pooled specificity (95% credible intervals): 0.8313(0.6173-0.9403)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[d]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[d]	very serious[d]	VERY LOW
1 or more inclusions	1	70	0.54 [0.33, 0.74]	0.98 [0.88, 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
1 or more grooves	1	69	0.96 [0.78, 1.00]	0.41 [0.27, 0.57]	**Sensitivity**				
					Very serious[a]	serious[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]	serious[b]	NA[c]	none[d]	VERY LOW
2 or more grooves	1	69	0.78 [0.56, 0.93]	0.83 [0.69, 0.92]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	very serious[d]	VERY LOW
3 or more grooves	1	69	0.48 [0.27, 0.69]	1.00 [0.92, 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

(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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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.

(e) Sub-grouping resolved heterogeneity for specificity (neither the USG nor non-USG sub-groups demonstrated heterogeneity), but not sensitivity, where heterogeneity remained within the sub-groups. .

Table 5: Summary of evidence relating to FNAC used without ROSE, with smear only, in the stratum where US *was* used to select patients (adjusted analysis).

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Bethesda Grade III or above	3	5,781	Pooled sensitivity (95% credible intervals): 0.8997 (0.4552-0.9906)	Pooled specificity (95% credible intervals):0.4545(0.1294-0.8261)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
Bethesda Grade IV or above	3	5,781	Pooled sensitivity (95% credible intervals): 0.7431 (0.2181-0.9712)	Pooled specificity (95% credible intervals):0.7751(0.5099-0.9202)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
Bethesda Grade V or above	3	5,781	Pooled sensitivity (95% credible intervals): 0.5342 (0.2474-0.8006)	Pooled specificity (95% credible intervals):0.8877(0.4689-0.9885)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
Bethesda Grade VI or above	3	5,781	Pooled sensitivity (95% credible intervals): 0.1661 (0.03444-0.5315)	Pooled specificity (95% credible intervals):0.9231(0.477-0.9935)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[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
2 way: malignant versus benign	1	945	0.87 [0.84, 0.89]	0.83 [0.78, 0.87]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
3 way: suspicious or malignant (negative = benign)	1	94	0.80 [0.56, 0.94]	0.55 [0.43, 0.67]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
3 way: malignant (negative = suspicious or benign)	1	94	0.45 [0.23, 0.68]	0.78 [0.67, 0.87]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	very serious[d]	VERY LOW
4 way De May classification: malignant, suspicious, non malignant follicular proliferation (negative = benign)	1	708	0.92 [0.84, 0.97]	0.48 [0.44, 0.52]	**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

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
4 way De May classification: malignant, suspicious (negative = benign, non malignant follicular proliferation)	1	708	0.84 [0.74, 0.92]	0.75 [0.71, 0.78]	**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
4 way De May classification: malignant (negative = benign, non malignant follicular proliferation, suspicious)	1	708	0.70 [0.59, 0.80]	0.94 [0.92, 0.96]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
4 way Piana classification: C3 or more	1	708	0.88 [0.86, 0.91]	0.50 [0.47, 0.53]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	none[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
4 way Piana classification: C4 or more	1	708	0.66 [0.63, 0.69]	0.93 [0.91, 0.94]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
4 way Piana classification: C5 or more	1	708	0.49 [0.46, 0.53]	0.94 [0.92, 0.95]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
4 way generic: malignant, suspicious, indeterminate (benign = negative)	2	1,846	1.00 [0.79, 1.00] 0.68 [0.61, 0.74]	0.75 [0.51, 0.91] 0.70 [0.68, 0.71]	**Sensitivity**				
					very serious[a]	serious[b]	serious[c]	very serious [d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	none[c]	very serious [d]	VERY LOW
4 way generic: malignant, suspicious, (indeterminate, benign = negative)	2	1,871	0.89 [0.75, 0.96] 0.46 [0.39, 0.53]	0.76 [0.50, 0.93] 0.79 [0.77, 0.81]	**Sensitivity**				
					very serious[a]	serious[b]	serious[c]	very serious [d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	none[c]	very 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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 FNAC used without ROSE, with smear, cytospin and/or cell-block, in the stratum where US was <u>not</u> used to select patients (adjusted analysis).

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Bethesda Grade III or above	5	1,143	Pooled sensitivity (95% credible intervals): 0.9035 (0.731-0.970)	Pooled specificity (95% credible intervals): 0.763(0.532-0.897)	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
Bethesda Grade IV or above	5	1,143	Pooled sensitivity (95% credible intervals): 0.8008 (0.535-0.925)	Pooled specificity (95% credible intervals): 0.899(0.770-0.957)	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	serious[d]	VERY LOW
	5	1,143			**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
Bethesda Grade V or above			Pooled sensitivity (95% credible intervals): 0.732 (0.402-0.914)	Pooled specificity (95% credible intervals): 0.938(0.822-0.984)	Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
Bethesda Grade V or above	5	1,143	Pooled sensitivity (95% credible intervals): 0.507 (0.229-0.759)	Pooled specificity (95% credible intervals): 0.947(0.853-0.984)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
2 way: malignant v benign	1	76	0.91 [0.71, 0.99]	0.98 [0.90, 1.00]	**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
3 way: malignant or suspicious (negative = benign)	13	2,360	Pooled sensitivity (95% credible intervals): 0.9108 (0.8485-0.9551)	Pooled specificity (95% credible intervals): 0.6863(0.5762-0.776)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	Very serious [d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	serious [d]	VERY LOW
3 way: malignant (negative = benign or suspicious)	10	2,120	Pooled sensitivity (95% credible intervals): 0.6437 (0.5049-0.7711)	Pooled specificity (95% credible intervals): 0.973(0.944-0.989)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[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
4 way: malignant, suspicious, indeterminate (negative = benign)	5	639	Pooled sensitivity (95% credible intervals): 0.801 (0.644-0.904)	Pooled specificity (95% credible intervals): 0.321(0.102-0.641)	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
4 way: malignant, suspicious (negative = benign, indeterminate)	6	1,054	Pooled sensitivity (95% credible intervals): 0.639 (0.415-0.821)	Pooled specificity (95% credible intervals): 0.747(0.476-0.909)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
4 way: malignant (negative = benign, indeterminate, suspicious)	5	939	Pooled sensitivity (95% credible intervals): 0.323 (0.0999-0.6435)	Pooled specificity (95% credible intervals): 0.879(0.561-0.9776)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	Very serious[d]	VERY LOW
5 way: malignant, suspicious, 2 grades of indeterminate (negative = benign)	1	76	0.75 [0.43, 0.95]	0.44 [0.20, 0.70]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 FNAC used without ROSE, with smear, cytospin and/or cell-block, in the stratum where US was used to select patients (adjusted analysis).

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Bethesda Grade III or above	1	489	0.94 [0.91, 0.96]	0.44 [0.31, 0.57]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
Bethesda Grade IV or above	1	489	0.90 [0.87, 0.93]	0.64 [0.51, 0.76]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
	1	487	0.90 [0.87, 0.93]	0.72 [0.59, 0.82]	**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
Bethesda Grade V or above					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	very serious[d]	VERY LOW
Bethesda Grade VI	1	487	0.68 [0.64, 0.73]	0.92 [0.83, 0.97]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
Benign or above	1	1,694	0.72 [0.63, 0.80]	0.84 [0.83, 0.86]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 FNAC used with ROSE, with smear only, in the stratum where US was not used to select patients (adjusted analysis).

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Bethesda Grade III or above	1	323	0.88 [0.81, 0.94]	0.73 [0.67, 0.79]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
Bethesda Grade IV or above	1	323	0.72 [0.63, 0.80]	0.90 [0.85, 0.93]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
Bethesda Grade V or above	1	323	0.53 [0.43, 0.62]	0.98 [0.95, 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
Bethesda Grade VI	2	376	0.36 [0.27, 0.45]; 0.67 [0.09, 0.99]	0.76 [0.70, 0.82]; 1.00 [0.93, 1.00]	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
	3	193			**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
3 way: malignant and suspicious (negative = benign)			Pooled sensitivity (95% credible intervals): 0.888 (0.442-0.989)	Pooled specificity (95% credible intervals): 0.572(0.262-0.842)	Very serious[a]	serious[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	very serious[d]	VERY LOW
3 way: malignant (negative = benign and suspicious)	2	153	0.40 [0.12, 0.74] 0.70 [0.50, 0.86]	0.97 [0.89, 1.00] 0.82 [0.69, 0.92]	**Sensitivity**				
					Very serious[a]	none[d]	none[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	none[d]	none[c]	Very serious[d]	VERY LOW
4 way: malignant, suspicious, indeterminate (negative = benign)	2	525	0.89 [0.79, 0.95] 0.89 [0.79, 0.96]	0.72 [0.66, 0.77] 0.42 [0.33, 0.51]	**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
4 way: malignant, suspicious (negative = benign, indeterminate)	2	525	0.55 [0.42, 0.67] 0.67 [0.54, 0.78]	0.95 [0.92, 0.97] 0.92 [0.86, 0.96]]	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
4 way: malignant (negative = benign, indeterminate, suspicious)	2	525	0.50 [0.37, 0.63] 0.50 [0.37, 0.63]	0.96 [0.93, 0.98] 0.96 [0.92, 0.99]	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 FNAC used with ROSE, with smear only, in the stratum where US was used to select patients (adjusted analysis).

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
intermediate or malignant	1	730	0.75 [0.70, 0.79]	0.89 [0.86, 0.92]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW

(e) 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.

(f) 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.

(g) 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.

(h) 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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 FNAC used with ROSE, with smear, cytospin and/or cell-block, in the stratum where US was not used to select patients (adjusted analysis).

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
3 way: suspicious or malignant (negative = benign)	2	198	0.86 [0.42, 1.00] 0.68 [0.43, 0.87]	0.71 [0.61, 0.80] 0.55 [0.43, 0.67]	**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
3 way: malignant (negative = suspicious or benign)	1	108	0.57 [0.18, 0.90]	0.79 [0.70, 0.87]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
4 way: malignant, suspicious, indeterminate (negative = benign)	1	44	1.00 [0.78, 1.00]	0.41 [0.24, 0.61]	**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
4 way: malignant, suspicious (negative = benign, indeterminate)	1	44	0.67 [0.38, 0.88]	1.0 [0.88, 1.00]	**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
5 way: malignant, suspicious, 2 grades of indeterminate (negative = benign)	1	170	0.77 [0.55, 0.92]	0.75 [0.62, 0.82]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	very serious[d]	VERY LOW
5 way: malignant, suspicious (negative = 2 grades of indeterminate, benign)	1	170	0.77 [0.55, 0.92]	0.82 [0.75, 0.88]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
5 way: malignant, suspicious (negative = suspicious, lower grade of indeterminate, benign)	1	170	0.73 [0.50, 0.89]	0.95 [0.90, 0.98]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
5 way: malignant (negative = suspicious, 2 grades of indeterminate, benign)	1	170	0.59 [0.36, 0.79]	0.97 [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

(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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 11: Summary of evidence relating to FNAC used with ROSE, with smear, cytospin and/or cell-block, in the stratum where US was used to select patients (adjusted analysis).

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
indeterminate follicular, indeterminate Hurtle, Suspicious for malignancy, or positive	1	240	0.97 [0.92, 0.99]	0.37 [0.29, 0.46]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
Suspicious for malignancy, or indeterminate follicular or positive	1	240	0.95 [0.89, 0.98]	0.43 [0.35, 0.52]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
Suspicious for malignancy, or positive	1	240	0.84 [0.76, 0.91]	0.88 [0.82, 0.93]	**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
Positive for malignancy	1	240	0.71 [0.61, 0.79]	0.91 [0.84, 0.95]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 12: Summary of evidence relating to core biopsy, in the stratum where US was not used to select patients (adjusted analysis).

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
carcinoma or neoplasm (versus benign)	1	31	0.56 [0.21, 0.86]	0.41 [0.21, 0.64]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
carcinoma (versus benign/indeterminate)	2	35	0.33 [0.07, 0.70]; 0.00 [0.00, 0.97]	0.55 [0.32, 0.76]; 1.00 [0.29, 1.00]	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	very serious[d]	VERY LOW
CB grades V and VI	1	578	0.90 [0.88, 0.93]	0.97 [0.86, 1.00]	**Sensitivity**				
					Serious[a]	serious[b]	NA[c]	none[d]	LOW
					Specificity				
					Serious[a]	serious[b]	NA[c]	none[d]	LOW
CB grades III, V and VI	1	578	0.96 [0.94, 0.97]	0.95 [0.82, 0.99]	**Sensitivity**				
					Serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Serious[a]	serious[b]	NA[c]	none[d]	LOW
positive (versus negative) with CEUS guidance	1	310	0.83 [0.78, 0.87]	0.81 [0.70, 0.90]	**Sensitivity**				
					Very 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				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
positive (versus negative) with US guidance	1	310	0.48 [0.42, 0.55]	0.84 [0.74, 0.92]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 13: Summary of evidence relating to core biopsy, in the stratum where US was used to select patients (adjusted analysis).

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
indeterminate, follicular neoplasm, suspicious for malignancy, or malignant	1	705	0.99 [0.98, 1.00]	0.28 [0.22, 0.36]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
follicular neoplasm, suspicious for malignancy, or malignant	1	705	0.91 [0.88, 0.93]	0.66 [0.59, 0.73]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
suspicious for malignancy, or malignant	1	705	0.77 [0.73, 0.81]	0.98 [0.95, 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

(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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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

1.1.9 Summary of the evidence – raw-data evidence

It could be argued that the adjusted strategy may be a somewhat harsh approach given that in the clinical setting an unsatisfactory reading may be repeated, albeit in many cases (if a ROSE approach is not employed) at a significantly later date, and that the unsatisfactory readings may eventually be remedied. Therefore Table 14 to Table 23 also provide the evidence where no correction has been made for unsatisfactory results (essentially the raw data provided in the papers, where unsatisfactory data are completely ignored). 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 14: Summary of evidence relating to FNAC used without ROSE, with smear only, in the stratum where US was not used to select patients ('raw data analysis').

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Bethesda Grade III or above	13	5,639	Pooled sensitivity (95% credible intervals): 0.951 (0.9169-0.9727)	Pooled specificity (95% credible intervals): 0.6851(0.571-0.7813)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
Bethesda Grade IV or above	13	6,123	Pooled sensitivity (95% credible intervals): 0.8745(0.8093-0.9213)	Pooled specificity (95% credible intervals): 0.8586(0.7807-0.9131)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
Bethesda Grade V or above	16	6,777	Pooled sensitivity (95% credible intervals): 0.783 (0.7165-0.8388)	Pooled specificity (95% credible intervals): 0.9761(0.9621-0.986)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	none[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
Bethesda Grade VI	12	5,437	Pooled sensitivity (95% credible intervals): 0.5084(0.3744-0.6409)	Pooled specificity (95% credible intervals): 0.9969(0.9934-0.9987)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
BTA THY 3a or above	2	414	0.68 [0.57, 0.77] 0.90 [0.73, 0.98]	0.74 [0.68, 0.80] 0.85 [0.75, 0.92]	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	Very serious[d]	VERY LOW
BTA THY 3f or above	1	306	0.52 [0.41, 0.63]	0.90 [0.85, 0.94]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
BTA THY 4 or above	1	306	0.28 [0.19, 0.38]	0.99 [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
BTA THY 5	2	414	0.08 [0.03, 0.16] 0.60 [0.41, 0.77]	1.00 [0.98, 1.00] 1.00 [0.95, 1.00]	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	none[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
AC 3 or above	3	455	Pooled sensitivity (95% credible intervals): 0.926 (0.735-0.984)	Pooled specificity (95% credible intervals): 0.380(0.123-0.717)	**Sensitivity**				
					Very serious[a]	serious[b]	none[d]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[d]	serious[d]	VERY LOW
AC 4 or above	3	455	Pooled sensitivity (95% credible intervals): 0.470 (0.202-0.753)	Pooled specificity (95% credible intervals): 0.957(0.859-0.989)	**Sensitivity**				
					Very serious[a]	serious[b]	none[d]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[d]	none[d]	VERY LOW
2 way: malignant v benign	13	1,055	Pooled sensitivity (95% credible intervals): 0.8491 (0.7056-0.9315)	Pooled specificity (95% credible intervals): 0.9644(0.9261-0.9849)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
3 way: suspicious or malignant (negative =benign)	52	11,025	Pooled sensitivity (95% credible intervals): 0.881 (0.844-0.913)	Pooled specificity (95% credible intervals): 0.789(0.723-0.845)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[d]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[d]	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
3 way: malignant (negative = suspicious or benign)	45	10,134	Pooled sensitivity (95% credible intervals): 0.6042 (0.542-0.664)	Pooled specificity (95% credible intervals): 0.985(0.976-0.992)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[d]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[d]	none[d]	VERY LOW
4 way: malignant or suspicious or indeterminate (negative = benign)	12	2,176	Pooled sensitivity (95% credible intervals): 0.866 (0.747-0.938)	Pooled specificity (95% credible intervals): 0.645(0.445-0.801)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[d]	serious[b]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[d]	very serious[b]	VERY LOW
4 way: malignant or suspicious (negative = benign or indeterminate)	14	2,174	Pooled sensitivity (95% credible intervals): 0.670 (0.501-0.811)	Pooled specificity (95% credible intervals): 0.911(0.854-0.950)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[d]	none[b]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[d]	none[b]	VERY LOW
4 way: malignant (negative = benign or indeterminate or suspicious)	12	2169	Pooled sensitivity (95% credible intervals): 0.4053(0.2348-0.5934)	Pooled specificity (95% credible intervals): 0.989(0.977-0.996)	**Sensitivity**				
					Very serious[a]	serious[b]	none[b]	none[b]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[b]	none[b]	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
5 way: malignant or suspicious or two grades of indeterminate (negative = benign)	6	1,734	Pooled sensitivity (95% credible intervals): 0.9438 (0.883-0.9741)	Pooled specificity (95% credible intervals): 0.5409(0.4327-0.6871)	**Sensitivity**				
					Very serious[a]	serious[b]	none[b]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[d]	none[b]	VERY LOW
5 way: malignant or suspicious or one grade of indeterminate (negative = lower grade of indeterminate or benign)	5	1.656	Pooled sensitivity (95% credible intervals): 0.872 (0.755-0.937)	Pooled specificity (95% credible intervals): 0.819(0.549-0.963)	**Sensitivity**				
					Very serious[a]	serious[b]	none[b]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[d]	very serious[d]	VERY LOW
5 way: malignant (negative = suspicious or two grades of indeterminate or benign)	6	1,742	Pooled sensitivity (95% credible intervals): 0.621 (0.478-0.741)	Pooled specificity (95% credible intervals): 0.993(0.981-0.998)	**Sensitivity**				
					Very serious[a]	serious[b]	none[d]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[d]	none[d]	VERY LOW
1 or more inclusions	1	70	0.54 [0.33, 0.74]	0.98 [0.88, 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

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
1 or more grooves	1	69	0.96 [0.78, 1.00]	0.41 [0.27, 0.57]	**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
2 or more grooves	1	69	0.78 [0.56, 0.93]	0.83 [0.69, 0.92]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	very serious[d]	VERY LOW
3 or more grooves	1	69	0.48 [0.27, 0.69]	1.00 [0.92, 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

(f) 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.

(g) 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.

(h) 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.

(i) 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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 15: Summary of evidence relating to FNAC used without ROSE, with smear only, in the stratum where US *was* used to select patients ('raw data analysis').

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Bethesda Grade III or above	3	4,416	Pooled sensitivity (95% credible intervals): 0.961 (0.4931-0.998)	Pooled specificity (95% credible intervals): 0.5643(0.1249-0.9483)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
Bethesda Grade IV or above	3	4,416	Pooled sensitivity (95% credible intervals): 0.7946 (0.2439-0.9812)	Pooled specificity (95% credible intervals): 0.9139(0.5431-0.9885)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
Bethesda Grade V or above	3	4,416	Pooled sensitivity (95% credible intervals): 0.583 (0.2799-0.8368)	Pooled specificity (95% credible intervals): 0.9798(0.8353-0.9982)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
Bethesda Grade VI or above	3	4,416	Pooled sensitivity (95% credible	Pooled specificity (95% credible intervals):	**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
			intervals): 0.1834 (0.035-0.6009)	0.9978(0.9858-0.9997)	**Specificity**				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
2 way: malignant versus benign	1	945	0.87 [0.84, 0.89]	0.83 [0.78, 0.87]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
3 way: suspicious or malignant (negative = benign)	1	82	0.94 [0.71, 1.00]	0.63 [0.50, 0.75]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
3 way: malignant (negative = suspicious or benign)	1	82	0.53 [0.28, 0.77]	0.89 [0.79, 0.96]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
4 way De May classification: malignant, suspicious, non malignant	1	674	0.96 [0.89, 0.99]	0.50 [0.46, 0.54]	**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
follicular proliferation (negative = benign)					very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
4 way De May classification: malignant, suspicious (negative = benign, non malignant follicular proliferation)	1	674	0.88 [0.78, 0.94]	0.79 [0.75, 0.82]	**Sensitivity**				
					very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
4 way De May classification: malignant (negative = benign, non malignant follicular proliferation, suspicious)	1	674	0.73 [0.61, 0.83]	0.99 [0.98, 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
4 way Piana classification: C3 or more	1	1,951	0.91 [0.89, 0.93]	0.53 [0.50, 0.56]	**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
4 way Piana classification: C4 or more	1	1,951	0.68 [0.65, 0.71]	0.99 [0.98, 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
4 way Piana classification: C5 or more	1	1,951	0.51 [0.47, 0.54]	1.00 [1.00, 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
4 way generic: malignant, suspicious, indeterminate (benign = negative)	2	1,506	1.00 [0.79, 1.00] 0.79 [0.72, 0.85]	0.75 [0.51, 0.91] 0.87 [0.85, 0.88]	**Sensitivity**				
					very serious[a]	serious[b]	serious[c]	very serious [d]	VERY LOW
					Specificity				
					very serious[a]	serious[b]	NA[c]	very serious [d]	VERY LOW
4 way generic: malignant, suspicious, indeterminate	2	1,528	0.93 [0.81, 0.99] 0.54 [0.46, 0.61]	0.81 [0.54, 0.96] 0.98 [0.97, 0.98]	**Sensitivity**				
					very serious[a]	serious[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
(benign = negative)					very serious[a]	serious[b]	NA[c]	very serious [d]	VERY LOW

(e) 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.

(f) 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.

(g) 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.

(h) 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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 16: Summary of evidence relating to FNAC used without ROSE, with smear, cytospin and/or cell-block, in the stratum where US was not used to select patients ('raw data analysis').

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Bethesda Grade III or above	5	1,093	Pooled sensitivity (95% credible intervals): 0.937 (0.798-0.985)	Pooled specificity (95% credible intervals): 0.825(0.611-0.931)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
Bethesda Grade IV or above	5	1,093	Pooled sensitivity (95% credible intervals): 0.8403 (0.608-0.942)	Pooled specificity (95% credible intervals): 0.959(0.895-0.984)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
Bethesda Grade V or above	5	1,093	Pooled sensitivity (95% credible intervals): 0.768 (0.442-0.926)	Pooled specificity (95% credible intervals): 0.989(0.962-0.998)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
Bethesda Grade VI or above	5	1,093	Pooled sensitivity (95% credible intervals): 0.535 (0.249-0.779)	Pooled specificity (95% credible intervals): 0.996(0.980-0.999)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
	1	76	0.91 [0.71, 0.99]	0.98 [0.90, 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
2 way: malignant v benign					Very serious[a]	serious[b]	NA[c]	very serious [d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
3 way: malignant or suspicious (negative = benign)	13	2,264	Pooled sensitivity (95% credible intervals): 0.9322 (0.877-0.9699)	Pooled specificity (95% credible intervals): 0.7208(0.6166-0.8017)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	very serious [d]	VERY LOW
3 way: malignant (negative = benign or suspicious)	10	2,065	Pooled sensitivity (95% credible intervals): 0.664 (0.524-0.796)	Pooled specificity (95% credible intervals): 0.992(0.982-0.997)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
4 way: malignant, suspicious, indeterminate (negative = benign)	5	537	Pooled sensitivity (95% credible intervals): 0.890 (0.777-0.952)	Pooled specificity (95% credible intervals): 0.414(0.144-0.732)	**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
	6	952			**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
4 way: malignant, suspicious (negative = benign, indeterminate)			Pooled sensitivity (95% credible intervals): 0.707 (0.491-0.866)	Pooled specificity (95% credible intervals): 0.899(0.702-0.973)	Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	serious[d]	VERY LOW
4 way: malignant (negative = benign, indeterminate, suspicious)	5	846	Pooled sensitivity (95% credible intervals): 0.360 (0.124-0.669)	Pooled specificity (95% credible intervals): 0.993(0.975-0.999)	**Sensitivity**				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
5 way: malignant, suspicious, 2 grades of indeterminate (negative = benign)	1	25	0.82 [0.48, 0.98]	0.50 [0.23, 0.77]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW

(e) 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.

(f) 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.

(g) 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.

(h) 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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 17: Summary of evidence relating to FNAC used without ROSE, with smear, cytospin and/or cell-block, in the stratum where US was used to select patients ('raw data analysis').

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Bethesda Grade III or above	1	479	0.95 [0.92, 0.97]	0.47 [0.34, 0.61]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
Bethesda Grade IV or above	1	479	0.91 [0.88, 0.94]	0.69 [0.56, 0.81]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	very serious[d]	VERY LOW
Bethesda Grade V or above	1	477	0.91 [0.88, 0.94]	0.78 [0.65, 0.88]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	very serious[d]	VERY LOW
Bethesda Grade VI	1	477	0.69 [0.64, 0.74]	1.00 [0.94, 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
	1	1,656			**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
Benign or above			0.72 [0.63, 0.80]	0.86 [0.85, 0.88]	Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW

(e) 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.

(f) 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.

(g) 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.

(h) 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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 18: Summary of evidence relating to FNAC used with ROSE, with smear only, in the stratum where US was not used to select patients ('raw data analysis').

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
Bethesda Grade III or above	1	323	0.88 [0.81, 0.94]	0.73 [0.67, 0.79]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
Bethesda Grade IV or above	1	323	0.72 [0.63, 0.80]	0.90 [0.85, 0.93]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
Bethesda Grade V or above	1	323	0.53 [0.43, 0.62]	0.98 [0.95, 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
Bethesda Grade VI	2	376	0.36 [0.27, 0.45]; 0.67 [0.09, 0.99]	0.76 [0.70, 0.82]; 1.00 [0.93, 1.00]	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	very serious[d]	VERY LOW
	3	183			**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
3 way: malignant and suspicious (negative = benign)			Pooled sensitivity (95% credible intervals): 0.9076 (0.4968-0.9932)	Pooled specificity (95% credible intervals): 0.6237(0.3218-0.863)	Very serious[a]	serious[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	very serious[d]	VERY LOW
3 way: malignant (negative = benign and suspicious)	2	146	0.40 [0.12, 0.74] 0.70 [0.50, 0.86]	0.97 [0.89, 1.0] 0.95 [0.85, 0.99]	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
4 way: malignant, suspicious, indeterminate (negative = benign)	2	503	0.93 [0.84, 0.98] 0.95 [0.87, 0.99]	0.75 [0.69, 0.80] 0.43 [0.35, 0.52]	**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
4 way: malignant, suspicious (negative = benign, indeterminate)	2	503	0.57 [0.44, 0.70] 0.71 [0.58, 0.82]	0.99 [0.97, 1.00] 0.95 [0.90, 0.98]	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	none[c]	none[d]	VERY LOW
	2	503	0.52 [0.39, 0.65]	1.00 [0.99, 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
4 way: malignant (negative = benign, indeterminate, suspicious)			0.53 [0.40, 0.66]	1.00 [0.97, 1.00]	Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW

(i) 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.

(j) 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.

(k) 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.

(l) 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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 19: Summary of evidence relating to FNAC used with ROSE, with smear only, in the stratum where US was used to select patients ('raw data analysis').

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
intermediate or malignant	1	730	0.75 [0.70, 0.79]	0.89 [0.86, 0.92]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 20: Summary of evidence relating to FNAC used with ROSE, with smear, cytospin and/or cell-block, in the stratum where US was not used to select patients ('raw data analysis').

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
	2	174			**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
3 way: suspicious or malignant (negative = benign)			0.86 [0.42, 1.00] 0.72 [0.47, 0.90]	0.90 [0.81, 0.96] 0.57 [0.44, 0.68]	Very serious[a]	serious[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	serious[c]	none[d]	VERY LOW
3 way: malignant (negative = suspicious or benign)	1	87	0.57 [0.18, 0.90]	1.00 [0.95, 1.00]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
4 way: malignant, suspicious, indeterminate (negative = benign)	1	44	1.00 [0.78, 1.00]	0.41 [0.24, 0.61]	**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
4 way: malignant, suspicious (negative = benign, indeterminate)	1	44	0.67 [0.38, 0.88]	1.0 [0.88, 1.00]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
5 way: malignant, suspicious, 2 grades of	1	166	0.81 [0.58, 0.95]	0.77 [0.69, 0.83]	**Sensitivity**				
					Very 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
indeterminate (negative = benign)					**Specificity**				
					Very serious[a]	serious[b]	NA[c]	very serious[d]	VERY LOW
5 way: malignant, suspicious (negative = 2 grades of indeterminate, benign)	1	166	0.81 [0.58, 0.95]	0.84 [0.77, 0.90]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
5 way: malignant, suspicious (negative = suspicious, lower grade of indeterminate, benign)	1	166	0.76 [0.54, 0.92]	0.97 [0.92, 0.99]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
5 way: malignant (negative = suspicious, 2 grades of indeterminate, benign)	1	166	0.62 [0.38, 0.82]	0.99 [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

(e) 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.

(f) 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.

(g) 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.

(h) 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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 21: Summary of evidence relating to FNAC used with ROSE, with smear, cytospin and/or cell-block, in the stratum where US was used to select patients ('raw data analysis').

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
indeterminate follicular, indeterminate Hurtle, Suspicious for malignancy, or positive	1	229	0.98 [0.93, 1.00]	0.40 [0.32, 0.49]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
Suspicious for malignancy, or indeterminate follicular or positive	1	229	0.96 [0.90, 0.99]	0.46 [0.38, 0.56]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
Suspicious for malignancy, or positive	1	229	0.85 [0.77, 0.92]	0.95 [0.90, 0.98]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
Positive for malignancy	1	229	0.72 [0.62, 0.80]	0.98 [0.93, 1.0]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW

(e) 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.
(f) 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.
(g) 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.
(h) 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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 22: Summary of evidence relating to core biopsy, in the stratum where US was not used to select patients ('raw data analysis').

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
carcinoma or neoplasm (versus benign)	1	17	1.0 [0.48, 1.00]	0.75 [0.43, 0.95]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	serious [d]	VERY LOW
carcinoma (versus benign/indeterminate)	2	20	0.60 [0.15, 0.95]; not estimable	1.00 [0.74, 1.00]; 1.00 [0.29, 1.00]	**Sensitivity**				
					Very serious[a]	serious[b]	none[c]	very serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[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
CB grades V and VI	1	577	0.90 [0.88, 0.93]	1.00 [0.90, 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
CB grades III, V and VI	1	577	0.96 [0.94, 0.97]	0.97 [0.85, 1.00]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
positive (versus negative) with CEUS guidance	1	310	0.83 [0.78, 0.87]	0.81 [0.70, 0.90]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
positive (versus negative) with US guidance	1	310	0.48 [0.42, 0.55]	0.84 [0.74, 0.92]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW

(e) 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.
(f) 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.
(g) 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.
(h) 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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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 23: Summary of evidence relating to core biopsy, in the stratum where US was used to select patients ('raw data analysis').

Index Test (Definition of a POSITIVE test)	Number of studies	n	Sensitivity (95% CI)	Specificity (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	GRADE
indeterminate, follicular neoplasm, suspicious for malignancy, or malignant	1	701	0.99 [0.98, 1.00]	0.29 [0.22, 0.36]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
follicular neoplasm, suspicious for malignancy, or malignant	1	701	0.91 [0.88, 0.93]	0.68 [0.60, 0.75]	**Sensitivity**				
					Very serious[a]	serious[b]	NA[c]	none[d]	VERY LOW
					Specificity				
					Very serious[a]	serious[b]	NA[c]	serious[d]	VERY LOW
suspicious for malignancy, or malignant	1	701	0.77 [0.73, 0.81]	1.00 [0.98, 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

(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 (0.95 or 0.85 for sensitivity and 0.7 or 0.8 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

1.1.10 Economic evidence

1.1.10.1 Included studies

Two health economic studies with the relevant comparison were included in this review [54] [105]. This is summarised in the health economic evidence profile below (Table 24) and the health economic evidence table in Appendix H.

1.1.10.2 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 G.

1.1.11 Summary of included economic evidence

Table 24: Health economic evidence profile: FNAC with rapid on-site evaluation (ROSE) vs FNAC without ROSE

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Breeze 2014[54] (UK)	Partially applicable[(b)]	Potentially serious limitations [(a)]	• Cross-sectional diagnostic study • Cost-effectiveness analysis • Population: Adults with suspected thyroid cancer who underwent ultrasound-guided FNA cytology • Comparators: 1. FNAC without ROSE 2. FNAC with ROSE • Follow-up: NR	£52.05	FNAC with ROSE gives 14% more adequate samples than FNAC without ROSE FNAC with ROSE lasts 6 minutes longer than FNAC without ROSE FNAC with ROSE reduces the number of people who could receive FNAC during a day by 3	FNAC with ROSE costs £378 more for each additional satisfactory sample	Probability Intervention 3 cost effective (£20/30k threshold): NA Uncertainty: NR
Feletti 2021 [105] (Italy)	Partially applicable[(d)]	Potentially serious limitations [(e)]	• Decision tree model • Cost-effectiveness analysis	£15 [(f)]	Cytopathologist assistance prevents 5%	FNAC with ROSE costs £300 more for	Probability Intervention 3 cost effective (£20/30k threshold): NA

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
			• Population: people with suspected thyroid cancer who underwent ultrasound guided FNA with and without the assistance of a cytopathology • Comparators: 1. US-guided FNAC without cytopathologist assistance 2. US-guided FNAC with cytopathologist assistance • Time horizon: 1 year		of non-diagnostical Thy1 cytologies	each additional satisfactory sample	Uncertainty: NR

Abbreviations: FNAC = fine needle aspiration cytology; NA = not applicable; NR = not reported; ROSE= Rapid on-site evaluation

(a) FNAC costs were based on a French source. The additional cost assumed for ROSE likely overestimates the cost per hour of a cytopathologist in the UK

(b) Time horizon or duration over which clinic visits took place was not reported. The estimation of the cost of ROSE is not clear and was not explained. Cost and consequences of surgery or further testing if the second FNAC is inadequate (e.g. diagnostic thyroid lobectomy) were not included, potentially underestimating the impact of improved sampling associated with rapid onsite evaluation by biomedical scientist. Resource use was obtained from single centre study of unclear generalizability to wider UK context. Sensitivity analyses were not reported. Potential conflicts of interests were not declared. Funding source was not reported.

(c) 2012 UK pounds. Cost components incorporated: Ultrasound-guided FNA of suspicious nodules, repeated FNAC for inadequate samples, assessment by a biomedical scientist (BMS).

(d) Italian NHS

(e) No analysis of uncertainty. Cytology assistance in this analysis is not limited to on-site evaluation (ROSE) but includes the presence of the cytopathologist during the entire procedure. Baseline inadequate rates come from a single Italian centre with an excellent performance and small room for improvement. Relative treatment effects were estimated from a single centre and it is unclear whether they can be generalised to other centres. Cost and consequences of surgery or further testing if the second FNAC is inadequate (e.g. diagnostic thyroid lobectomy) were not included, potentially underestimating the impact of improved sampling associated with rapid onsite evaluation by biomedical scientist. Resource use and unit costs were obtained from a single Italian centre of unclear generalisability to UK context.

(f) 2021 Euro converted to UK pounds[268]. Cost components incorporated: Ultrasound-guided FNA of suspicious nodules, repeated FNAC for inadequate samples, cytopathologist assistance

1.1.12 Economic model

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

1.1.13 Cost comparison analysis

Although published literature found ROSE to increase the overall cost of FNAC, most of the studies included in the clinical and economic literature review agree that ROSE improves sample accuracy and therefore reduces the need of taking additional sampling. As it is unclear whether implementing ROSE in some or most centres would be beneficial for the NHS, a simple cost-comparison analysis was undertaken using UK unit costs and a NHS perspective.

The analysis assumed that every FNAC with an inadequate sample (Thy1) would require a repeat sampling with CNB. The hypothesis is that, although adding ROSE would make FNAC more expensive, a lower rate of inadequate samples would require less repeat tests, thus potentially saving money for the NHS and increasing NHS capacity.

The cost of a US-guided was estimated through the NHS Reference Costs 2019-2020. The additional cost of FNAC with ROSE was calculated assuming that a cytopathologist or a biomedical scientist (BMS) would be required for 44.4 minutes of their time to provide ROSE and interpret the results. This is based on a study[201] which measured the time the operators left the office to the time they returned after the aspiration procedure and interpretation. Although ROSE could be effectively performed by an adequately trained BMS, the interpretation of the results and the final diagnosis always require a consultant cytopathologist. Current practice in England shows that in many centres a consultant cythopathologist undertakes the whole procedure as well. A 2020 survey on cytopathology practice in the UK done by the Royal College of Pathologists[273] found a equal split between BMS and pathologist among those undertaking ROSE. This figure was used to estimate the average cost of ROSE in England using 50% the hourly cost of a BMS band 5 and 50% of a consultant cytopathologist's hourly cost. The resulting cost of £70 is in line with the estimations made by other UK studies[54, 280].

The committee noted that after an inadequate FNAC, an outpatient visit is often required for the clinicians to explain the results to the patient and discuss the follow-up test. The cost of an outpatient visit was estimated through NHS Reference Costs 2019-2020 by averaging the cost of a face-to-face and non face-to-face outpatient endocrinology follow-up attendance. The repeat test of choice was assumed to be a Core Needle Biopsy (CNB) as recommended for the management of a Thy1 non-diagnostic cytology. All unit costs are presented in table 25.

Table 25: Unit costs

Resource	Unit costs	Source
US-guided FNAC	£299	NHS Reference Costs 2019-2020[263]
US-guided FNAC with ROSE	£369	NHS Reference Costs 2019-2020[263] PSSRU 2020[84] Layfield 2001[201]
Core Needle Biopsy (CNB)	£429	NHS Reference Costs 2019-2020[263]
Endocrinology outpatient visit	£91	NHS Reference Costs 2019-2020[263]

The baseline inadequacy rate without ROSE in the UK was estimated from an evidence-based review looking at rates of Thy1 FNAC using RCPath Thy terminology[279]. This gives a baseline rate of 18.5% including cystic lesion Thy1c. The relative treatment effect of adding ROSE was obtained from the meta-analysis conducted from the clinical review. This gives a relative risk of 0.44 of non-diagnostic with FNAC ROSE versus FNAC without ROSE. This

estimation is supported by published evidence which found the same relative risk of 0.44 when comparing FNAC with ROSE and without ROSE[368]. Baseline inadequacy rates and relative treatment effect of ROSE are shown in table 26.

Table 26: Baseline inadequacy rate and ROSE relative treatment effect

Parameter	Value	Source
Inadequacy rate with no ROSE	18.5%	Poller 2020[279]
Relative risk of inadequacy with ROSE vs no ROSE	0.44	Clinical review Witt 2013[368]

The results of the analysis are illustrated in table 27.

Table 27: Cost analysis results (cohort of 1000 people)

Strategy	N° of inadequate samples	Cost per patient
FNAC with ROSE	81	£412
FNAC without ROSE	185	£395
Difference (ROSE – no ROSE)	- 104	£17

Scenario 1: repeat FNAC requires an outpatient visit before the test; Scenario 2: a repeat FNAC does not require a visit before.

The results showed that FNAC with ROSE reduces the number of inadequate sample (and therefore of repeat sampling) by 0.1 for every FNAC with ROSE performed. The cost analysis demonstrates that adding ROSE to a centre or an individual clinician with a baseline inadequacy rate of 18.5% would cost £17 more per patient. The results are mostly driven by the high cost of ROSE which is assumed to be undertaken by a consultant cytopathologists in half of the cases. Were ROSE to be undertaken solely by a BMS with the pathologist help only for the diagnosis, ROSE would be cost-saving at the baseline threshold.

It is uncertain whether offering ROSE would increase the capacity of the NHS. The analysis showed that for every ROSE, 0.1 less repeat FNAC are avoided but UK evidence suggests that ROSE increases the time of a FNAC by around 6 minutes. Hence, NHS capacity would improve only if the average time required for a FNAC exceeds 60 minutes which seem to be very unrealistic in the UK.

However, improving the adequacy rates of FNAC may have other benefits that this analysis is not capturing. A more efficient diagnostic pathway would translate into less burden to the patients who would not be required to repeat the same test twice while being concerned of having a cancer not yet diagnosed. It has also been suggested that lower inadequacy rates of FNAC could be associated with less unnecessary surgeries for people with benign nodules, which represent a high non cost-effective use of NHS resource and a potential harm for people[54].

At the current cost of ROSE, the threshold analysis shows that ROSE would become cost-saving only when the baseline risk goes above 24%. This is above the expected average rate of Thy1/Thy1c estimated by the Royal College of Pathologists ranging between 18% and 22%[82], suggesting that ROSE would become cost-effective when targeted to fewer centres or to individuals with a concerningly high inadequate rate. It is worth noting that both the clinical review and health economic analysis' definition of Thy1 includes non-diagnostic cystic lesions Thy1c. ROSE is not particularly helpful after a Thy1c as this is not operator-dependent. The only large review reporting rates of Thy1c[279] shows that Thy1c usually ranges between 5 to 10% of all samples. If we subtract this figure from the threshold of 24% estimated in the cost-comparison analysis, the new Thy1-exclusive threshold would range between 14%-19% which is in line with the threshold of 15% (excluding Thy1c) identified by the Royal College of Pathologist as a matter of concerns[279]. Therefore, If ROSE is implemented in centres or for individuals with a concerningly high rate of Thy1 (excluding

Thy1c) as defined by the Royal College of Pathologists, the intervention would likely be cost-effective, if not cost-saving, in the UK.

1.1.14 Economic evidence statements

Two cost-effectiveness analyses found FNAC with ROSE to cost, respectively, £300 and £378 more for each additional satisfactory cytology (different than the non-diagnostic category Thy1). Both studies were assessed as partially applicable with potentially serious limitations.

One original comparison analysis found that FNAC with ROSE cost £17 more per patient compared to FNAC without ROSE. The analysis was assessed as directly applicable with minor limitations.

1.1.15 The committee's discussion and interpretation of the evidence

1.1.15.1 The outcomes that matter most

Sensitivity and specificity were the outcomes used in this review. Sensitivity was identified as the primary measure in guiding decision-making. This was because the harms of false negatives (the proportion of which determine the level of sensitivity) are likely to be greater than the harms of false positives (the proportion of which determine the level of specificity). False negatives lead to people with a malignancy being missed by the index test, and therefore remaining undiagnosed and untreated, which can have very serious consequences. On the other hand, false positives may lead to people without malignancy being given unnecessary surgery. Whilst carrying the risk of serious harms, these were regarded as less serious harms than those posed by false negatives. The committee therefore set clinical decision thresholds for sensitivity of 0.95 and above for recommending a test, and 0.85, below which a test would be deemed of no clinical use. They also set clinical decision thresholds for specificity of 0.8 and above for recommending a test, and 0.7, below which a test would be deemed of no clinical use.

These figures were developed in the context of FNAC being used as a second line test after ultrasound has been used as the initial filter test to select people for FNAC testing (people positive on ultrasound). As the definitive second test, FNAC must be both highly sensitive and specific. In particular it needs to be highly sensitive, even more sensitive than the previous filter test. The previous filter test itself must be highly sensitive to ensure that people with actual malignancy are not missed at the first hurdle, but if the second test – FNAC – is not even *more* sensitive than this then it may lead to people that have been fed through from ultrasound testing with true malignancy being erroneously classified as benign at this second step. Therefore, FNAC used as a second definitive test ideally needs almost perfect sensitivity, and certainly needs to have a higher sensitivity than the recommended US test. It also needs to have a superior specificity as well, as the chief function of the second test is to 'mop-up' the many people who were positive on ultrasound who will actually have been false positive. In other words, FNAC will need to be able to accurately differentiate these people into those that are truly positive and those that are not. However, perfect specificity, although desirable, is not as essential as very high sensitivity, as the harms of some people being referred for surgery when they do not have malignancy are less critical than the harms of missing a positive diagnosis.

It should be noted that the target sensitivity value of 0.95 is comparable to that achieved by the best evidence identified from a first line US test, that is, using the threshold for a positive test of an EU TIRADS score of 4 or more. This follows, because if FNAC were to have a much lower sensitivity than the first line test, it would mean that some of the true positives fed through to FNAC might be erroneously deemed as negatives by FNAC. In addition, the target specificity value of 0.8 is considerably more than that achieved by the best evidence identified from a first line US test, that is, using the threshold for a positive test of an EU

TIRADS score of 4 or more. This was important to ensure that FNAC was better able to differentiate between the many false and negative positives fed through from ultrasound.

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.

1.1.15.2 The quality of the evidence

The quality of the evidence was graded as very low across all outcomes except three outcomes of low quality. The main reasons for this were risk of bias (as determined by QUADAS 2) which was very serious in the majority of outcomes. This is a mix of poor research or poor reporting and that research in this area is difficult. Most of studies do not describe whether the index and reference tests have been interpreted without knowledge of the other. Also, the time interval between the tests is unclear in most studies which indicates poor research as methods are not clearly described or not done. Most of the studies were also downgraded for patient selection as it is unclear if an appropriate inclusion/exclusion criterion have been considered with consecutive or random samples. The majority of studies are retrospective which would have made this difficult as these details may not have been recorded in patient records when selecting from databases. While some of the studies were old the committee agreed that the data would still be relevant to current practice.

GRADE ratings were also downgraded due to indirectness in outcomes where the majority of studies were retrospective. Retrospective data are collected before research is considered so are collected in a purely clinical context without concern for ensuring patients achieve diagnostic gold standards. Hence the tendency may be for less people to go to surgery unless clinically indicated by a worse FNAC – so lower FNAC gradings may be less represented. On the other hand, in a prospective study the context is not wholly clinical – the emphasis on research, and therefore ensuring that as many people as possible have gold standard measures, may mean that more are sent for surgery from lower FNAC grades. Having fewer people in lower FNAC grades can skew accuracy considerably, spuriously increasing sensitivity and reducing specificity.

Use of ultrasound guidance 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 ultrasound guidance was used to sub-group studies. Many meta-analyses demonstrated some degree of heterogeneity but sub-grouping for the use of ultrasound guidance resolved the heterogeneity within the sub-groups in one analysis only (the ‘2 way’ malignant/benign [FNAC without ROSE and direct smear only, without prior US, using adjusted approach] analysis). This indicated that ultrasound guidance was not an important factor influencing the variability in accuracy between studies for the other meta-analyses. Therefore, the other meta-analyses with heterogeneity were not sub-grouped and were downgraded for heterogeneity.

Poor reporting was a feature of many of the included studies. Classification into the different index test types was carried out on the basis of the information provided, which was often fairly sketchy. Several papers were excluded where no description of the FNAC description was given at all, as this made it impossible to place the paper into any of the index test categories.

Finally, many outcomes were downgraded for imprecision, partly because of small study sizes.

1.1.15.3 Benefits and harms

Two sets of data had been presented in the review: a) the raw data, which did not include consideration of the inadequate readings, and b) the adjusted data, which incorporated any inadequate data by classifying any inadequate FNAC results from gold-standard positive nodules as false negatives and classifying any inadequate FNAC results from gold-standard negative nodules as false positives. The latter approach follows the rationale that because the inadequate results cannot possibly demonstrate malignancy, they cannot ever be true positives in people with GS-proven malignancy (thus they must by exclusion be false negatives). Equally, because the inadequate results don't depict benignancy, they cannot ever be true negatives in people with GS-proven benignancy (thus they must by exclusion be false positives). The committee considered both types of data but favoured the former approach using the raw data. This was because clinically it is often possible to repeat an initially unsuccessful test successfully, and the time delay does not cause significant clinical harm. The committee also gave the opinion that there is no association between inadequacy and malignancy. Thus, inadequate results may be safe to ignore when considering diagnostic accuracy of FNAC. Therefore, all evidence used by the committee to form recommendations were the raw data.

The committee noted that Cytospin is a proprietary trade mark and agreed that 'liquid based cytology' is a generic term that includes 'Cytospin and cell block' and is therefore more appropriate to use in a guideline recommendation. When considering the raw diagnostic accuracy evidence from the review, the committee noted that only one FNAC meta-analysis yielded sensitivity and specificity values that were sufficiently close to the targets for recommendation. This was for the analysis in studies where neither ROSE nor prior US selection had been carried out but where studies had used direct smear and liquid based cytology (as required). This analysis, based on 5 studies and over 1000 participants, demonstrated a sensitivity of 0.937 and a specificity of 0.825 when using the threshold for a positive test of Bethesda grade III and above. In relation to this, the committee discussed how although much of the evidence in the review is based on the Bethesda grading scheme, the Bethesda classification scheme is not commonly used in the UK. The committee therefore recommended that a Bethesda-equivalent scheme widely used in the UK called the RC PATH modification of the BTA (RC PATH BTA) should be used instead. This uses qualitatively similar grades, whilst the main difference is fairly superficial, based on the labelling of each grade. RC PATH BTA grades Thy 1, 2, 3a, 3f, 4 and 5 are equivalent to Bethesda grades I, II, III, IV, V and VI respectively. Overall, they thought the result suggested liquid based cytology or direct smear should be used when processing FNAC samples.

The issue of Rapid Onsite evaluation was discussed. Data from the diagnostic accuracy review (please see cost-comparison analysis section 1.1.13) showed that ROSE reduced non-diagnostic results by 55%. After hearing the health economic evidence (please see section below) the committee agreed that certain sites, where inadequacy rates were poor, might benefit from rapid on site evaluation.

1.1.15.4 Cost effectiveness and resource use

Two health economics studies were included both being cost-effectiveness analyses looking at the impact of adding rapid on-site evaluation (ROSE) by a cytopathologist.

The first study was assessed to be partially applicable as, although conducted in the UK, it used unit costs estimated in other countries. The cost of FNAC was taken from a French cost analysis whereas the additional cost of ROSE was estimated using US literature, where the cost per hour of a cytopathologist is expected to be considerably higher than in the UK. Furthermore, the study was assessed to have potentially serious limitations as the sample size was small, resource use was estimated from a single hospital with unclear generalizability, estimation of cost was unclear and possibly not reflecting UK settings and the study failed to include relevant outcomes such as surgeries. The study found that at an

additional cost of £78 per patient, ROSE increases the adequate sample rates by 14% and the duration of the visit by 6 minutes. In other words, introducing ROSE would cost £378 for each additional satisfactory sample.

The second study retrospectively assessed a series of FNAC performed with and without cytopathologist assistance in an Italian centre and conducted alongside a cost-effectiveness analysis using unit costs estimated from the same centre. The analysis has some limitations as no analysis of uncertainty was conducted and the intervention presumably includes more than just ROSE as the cytopathology assisted the radiologist with the selection of the site of the nodule to take the sample from. Moreover, this specific Italian centre had exceptionally high performance in terms of diagnostic rates which may underestimate the effectiveness of the intervention, as ROSE is known to be more effective when there is large room for improvement. Relative treatment effects were estimated from a single centre and unit costs and resource use were obtained from an Italian institution hardly generalisable to the UK context. The analysis found that at an additional cost of £12, cytopathologist assistance prevents 5% of non-diagnostic results. In other words, introducing ROSE would cost £300 for each additional satisfactory sample.

Given the lack of a reliable UK studies as the only British study included made extensive use of non-UK sources, an original cost-comparison analysis was conducted to shed light on the advantage of introducing ROSE in UK centres. The meta-analysis conducted for the clinical review showed that ROSE reduces the number of non-diagnostic samples (Thy1/Thy1c) by 55%. This is in line with the published literature which reported a relative risk of inadequacy with ROSE versus without ROSE of 0.44. This relative risk was used in the analysis and applied to the baseline Thy1/Thy1c rate reported in the literature (18%). The analysis assumed that every non-diagnostic FNAC would require a further core-needle biopsy (CNB). The committee noted that before repeat sampling, an outpatient visit is often required as the clinician needs to review the results of the biopsy with the patient and discuss any follow-up test. The cost of an US-guided FNAC was collected from the NHS Reference Costs 2019-2020. The additional cost of ROSE was estimated to be £70, which are equivalent to 44 minutes of the hourly cost of a BMS band 5 or consultant cytopathologist in England. An equal split between BMS and cythopathologists was assumed as reported in a recent survey on cytopathology practice in the UK. The analysis found that FNAC with ROSE costs £17 more per patient compared to FNAC without ROSE and reduces the number of repeat tests by 0.1 for every FNAC with ROSE performed.

The committee recognised that cytopathologists and trained BMS are not widely available in the UK and that, in some small centres where only a few FNACs are performed every day, implementing ROSE would hardly be a cost-effective use of NHS resource. The committee also acknowledged that, although a consultant cytopathologist is always required for the final diagnosis, ROSE could be effectively and entirely undertaken by an adequately trained BMS. Although recent surveys suggest that in many cases a consultant cytopathologist undertakes the whole procedure, ROSE could become cheaper and thus more advantageous if, in the future, plans to train and rely on BMS more often are adopted.

A threshold analysis found that ROSE would become cost-saving when the baseline non-diagnostic rate (Thy1, Thy1c) is above 24%. The committee noted that ROSE is not useful in centres with a high rate of Thy1c as Thy1c describes a non-diagnostic cyst and is not operator- nor technique-dependent. The committee noted that the threshold estimated in the cost-comparison analysis, when excluding Thy1c, would be relatively similar to the Thy1c-exclusive threshold identified by the Royal College of Pathologists (>15%). Hence, the committee made a recommendation to implement ROSE when the non-diagnostic Thy1 rate is above 15% (excluding Thy1c). This could apply to both centres or individual clinicians with a high non-diagnostic rate. This targeted approach that prioritises centres and clinicians which would most benefit from ROSE is likely to be cost-effective, if not cost-saving, in the UK and would likely improve the diagnostic efficiency of the NHS in the long-term.

The committee recommended to use liquid-based cytology or direct smear when processing FNAC samples. Some centres also do both as part of a quality assurance process to get better results. Overall, this reflects current practice where liquid based cytology, direct smear or both are used and, as such, it is not expected to require additional NHS resources.

1.1.15.5 Other factors the committee took into account

The committee discussed how in practice that FNAC grades would not always be used as a blunt decision tool, but would usually also be used in conjunction with other information, such as the initial US findings. Given that people fed through to FNAC with a range of US findings in FNAC candidates, from mild hypoechoicity but no suspicious features to several suspicious features. It was discussed how an indeterminate FNAC finding combined with 3 suspicious features on US might be considered more indicated for surgery than an indeterminate FNAC finding combined with mild hypoechoicity and no suspicious features on US. However, it was agreed that there was no evidence from the current review to back up this view, and the committee agreed that any such decisions should be based on clinical expertise.

1.1.16 Recommendations supported by this evidence review

This evidence review supports recommendations 1.2.8 to 1.2.10.

References

1. Abboud B, Allam S, Chacra LA, Ingea H, Tohme C, Farah P. Use of fine-needle aspiration cytology and frozen section in the management of nodular goiters. Head and Neck. 2003; 25(1):32-36

2. Abou-Foul AK, Muzaffar J, Diakos E, Best JE, Momtahan N, Jayaram S. Correlation between thyroid fine needle aspiration cytology and postoperative histology: A 10-year single-centre experience. Cureus. 2021; 13(4):15

3. Acar Y, Dogan L, Guven HE, Aksel B, Karaman N, Ozaslan C et al. Bethesda made it clearer: A review of 542 patients in a single institution. Oncology Research and Treatment. 2017; 40(5):277-280

4. Afroze N, Kayani N, Hasan SH. Role of fine needle aspiration cytology in the diagnosis of palpable thyroid lesions. Indian Journal of Pathology and Microbiology. 2002; 45(3):241-246

5. Aftab MF, Manan A, Ullah S. Role of F.N.A.C. USG and scintigraphy in the evaluation of solitary thyroid nodule. Medical Forum Monthly. 2005; 16(5):6-11

6. Agcaoglu O, Aksakal N, Ozcinar B, Sarici IS, Ercan G, Kucukyilmaz M et al. Factors that affect the false-negative outcomes of fine-needle aspiration biopsy in thyroid nodules. International Journal of Endocrinology Print. 2013; 2013:126084

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. Agrawal S. Diagnostic accuracy and role of fine needle aspiration cytology in management of thyroid nodules. Journal of Surgical Oncology. 1995; 58(3):168-172

9. Aguilar-Diosdado M, Contreras A, Gavilan I, Escobar-Jimenez L, Giron JA, Escribano JC et al. Thyroid nodules. Role of fine needle aspiration and intraoperative frozen section examination. Acta Cytologica. 1997; 41(3):677-682

10. Ahari AA, Vajari MAM, Moghadam NK, Hashemi H, Parvin M, Khaleghi M. Comparison on the use of spinal (Stylet) needle and simple needle in ultrasound guided thyroid nodule fna; does the needle affect thyroid fna result? Iranian Journal of Radiology. 2020; 17(2):e98754

11. Ahn HS, Youn I, Na DG, Kim SJ, Lee MY. Diagnostic performance of core needle biopsy as a first-line diagnostic tool for thyroid nodules according to ultrasound patterns: Comparison with fine needle aspiration using propensity score matching analysis. Clinical Endocrinology. 2021; 94(3):494-503

12. Ahn SS, Kim EK, Kang DR, Lim SK, Kwak JY, Kim MJ. Biopsy of thyroid nodules: comparison of three sets of guidelines. AJR American Journal of Roentgenology. 2010; 194(1):31-37

13. Akerman M, Tennvall J, Biorklund A, Martensson H, Moller T. Sensitivity and specificity of fine needle aspiration cytology in the diagnosis of tumors of the thyroid gland. Acta Cytologica. 1985; 29(5):850-855

14. 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

15. Akhtar S, Awan MS. Role of fine needle aspiration and frozen section in determining the extent of thyroidectomy. European Archives of Oto-Rhino-Laryngology. 2007; 264(9):1075-1079

16. 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

17. Al-Dbahri S, Al-Sebeih K, Hier MP, Black MJ. An aggressive approach to the surgical management of suspicious thyroid nodules. Journal of Otolaryngology. 2001; 30(4):203-207

18. Al-Hureibi KA, Al-Hureibi AA, Abdulmughni YA, Aulaqi SM, Salman MS, Al-Zooba EM. The diagnostic value of fine needle aspiration cytology in thyroid swellings in a university hospital, Yemen. Saudi Medical Journal. 2003; 24(5):499-503

19. Al-Taweel AZ, Dashti H, Behbehani A, Olszewski W, Atia SO, El-Naqeeb N. Value of clinical examination, scintigraphy and fine-needle aspiration cytology in the diagnosis of solitary thyroid nodules. Medical Principles and Practice. 1990; 2(3-4):167-171

20. Alalawi Y, Moharram LM. Thyroid fine-needle aspiration: Histologic correlation of the diagnostic categories of the bethesda system with emphasis on "atypia of undetermined significance": A 5-year single-institution experience. World Journal of Endocrine Surgery. 2019; 11(3):76-79

21. Alhashem MH, Alabidi A, Aly MG. The bethesda system for reporting thyroid cytopathology: A retrospective review of its diagnostic utility at johns hopkins aramco healthcare, saudi arabia. American Journal of Otolaryngology Head and Neck Medicine and Surgery. 2021; 42(6):103088

22. Alshaikh S, Harb Z, Aljufairi E, Almahari SA. Classification of thyroid fine-needle aspiration cytology into Bethesda categories: An institutional experience and review of the literature. CytoJournal. 2018; 15 (1) (no pagination)(4)

23. Altavilla G, Pascale M, Nenci I. Fine needle aspiration cytology of thyroid gland diseases. Acta Cytologica. 1990; 34(2):251-256

24. Ananthakrishnan N, Rao KM, Narasimhan R, Veliath AJ. Problems and limitations with fine needle aspiration cytology of solitary thyroid nodules. Australian and New Zealand Journal of Surgery. 1990; 60(1):35-39

25. Anderson JB, Webb AJ. Fine-needle aspiration biopsy and the diagnosis of thyroid cancer. British Journal of Surgery. 1987; 74(4):292-296

26. Anderson TJ, Atalay MK, Grand DJ, Baird GL, Cronan JJ, Beland MD. Management of nodules with initially nondiagnostic results of thyroid fine-needle aspiration: can we avoid repeat biopsy? Radiology. 2014; 272(3):777-784

27. Archondakis S, Georgoulakis J, Stamataki M, Anninos D, Skagias L, Panayiotides I et al. Telecytology: a tool for quality assessment and improvement in the evaluation of thyroid fine-needle aspiration specimens. Telemedicine Journal and e-Health. 2009; 15(7):713-717

28. Arena S, Latina A, Stornello M, Saraceno G, Benvenga S. Intranuclear cytoplasmic inclusions in cytologically suspicious or malignant thyroid nodules: identification and correlation with echogenicity and size of the nodules. Endocrine. 2014; 46(1):114-122

29. Arul P, Akshatha C, Masilamani S. A study of malignancy rates in different diagnostic categories of the Bethesda system for reporting thyroid cytopathology: An institutional experience. Biomedical Journal. 2015; 38(6):517-522

30. Aydogan BI, Sahin M, Ceyhan K, Deniz O, Demir O, Emral R et al. The influence of thyroid nodule size on the diagnostic efficacy and accuracy of ultrasound guided fine-needle aspiration cytology. Diagnostic Cytopathology. 2019; 47(7):682-687

31. Aysan E, Kiran T, Idiz UO, Guler B, Akbulut H, Kunduz E et al. The diagnostic ability of core needle biopsy in nodular thyroid disease. Annals of the Royal College of Surgeons of England. 2017; 99(3):233-236

32. Bahaj AS, Alkaff HH, Melebari BN, Melebari AN, Sayed SI, Mujtaba SS et al. Role of fine-needle aspiration cytology in evaluating thyroid nodules. A retrospective study from a tertiary care center of Western region, Saudi Arabia. Saudi Medical Journal. 2021; 41(10):1098-1103

33. Bahar G, Braslavsky D, Shpitzer T, Feinmesser R, Avidan S, Popovtzer A et al. The cytological and clinical value of the thyroid "follicular lesion". American Journal of Otolaryngology. 2003; 24(4):217-220

34. Bajaj Y, De M, Thompson A. Fine needle aspiration cytology in diagnosis and management of thyroid disease. Journal of Laryngology and Otology. 2006; 120(6):467-469

35. Balas EA, Merei J. On statistical comparison of two diagnostic tests. Computers and Biomedical Research. 1985; 18(6):497-501

36. Bapat RD, Shah SH, Relekar RG, Pandit A, Bhandarkar SD. Analysis of 105 uninodular goitres. Journal of Postgraduate Medicine. 1992; 38(2):60-61

37. Basharat R, Bukhari MH, Saeed S, Hamid T. Comparison of fine needle aspiration cytology and thyroid scan in solitary thyroid nodule. Pathology Research International. 2011; 2011:754041

38. Bashier AH, Abdin I, Elhassan M, Sanhouri M, Ahmed ME. Solitary thyroid nodule in Khartoum. East African Medical Journal. 1996; 73(10):694-696

39. Baskin HJ, Guarda LA. Influence of needle biopsy on management of thyroid nodules: reasons to expand its use. Southern Medical Journal. 1987; 80(6):702-705

40. Beecham JE, Alibutud MF, Burke M. Fine-needle aspiration biopsy for the routine screening of Saudi patients with thyroid nodules. Annals of Saudi Medicine. 1988; 8(4):252-256

41. Belanger R, Guillet F, Matte R, Havrankova J, d'Amour P. The thyroid nodule: evaluation of fine-needle biopsy. Journal of Otolaryngology. 1983; 12(2):109-111

42. Bellantone R, Lombardi CP, Raffaelli M, Traini E, De Crea C, Rossi ED et al. Management of cystic or predominantly cystic thyroid nodules: the role of ultrasound-guided fine-needle aspiration biopsy. Thyroid. 2004; 14(1):43-47

43. Bernante P, Toniato A, Piotto A, Gemo G, Pagetta C, Bernardi C et al. Follicular neoplasms of the thyroid: diagnostic and operative management. Journal of Experimental and Clinical Cancer Research. 1998; 17(1):125-126

44. Bhartiya R, Mallik M, Kumari N, Prasad BN. Evaluation of thyroid lesions by fine-needle aspiration cytology based on Bethesda system for reporting thyroid cytopathology classification among the population of South Bihar. Indian journal of

medical and paediatric oncology : official journal of Indian Society of Medical & Paediatric Oncology. 2016; 37(4):265-270

45. Bhatki AM, Brewer B, Robinson-Smith T, Nikiforov Y, Steward DL. Adequacy of surgeon-performed ultrasound-guided thyroid fine-needle aspiration biopsy. Otolaryngology - Head & Neck Surgery. 2008; 139(1):27-31

46. Bhatti SUZ, Malook MSU, Zulqarnain MA. Diagnostic accuracy of fine Needle Aspiration Cytology in thyroid nodules. Pakistan Journal of Medical and Health Sciences. 2010; 4(3):245-247

47. Biscotti CV, Hollow JA, Toddy SM, Easley KA. ThinPrep versus conventional smear cytologic preparations in the analysis of thyroid fine-needle aspiration specimens. American Journal of Clinical Pathology. 1995; 104(2):150-153

48. Bisi H, De Camargo RYA, Filho AL. Role of fine-needle aspiration cytology in the management of thyroid nodules: Review of experience with 1,925 cases. Diagnostic Cytopathology. 1992; 8(5):504-510

49. Blumenfeld W, Nair R, Mir R. Diagnostic significance of papillary structures and intranuclear inclusions in Hurthle-cell neoplasms of the thyroid. Diagnostic Cytopathology. 1999; 20(4):185-189

50. Bodo M, Dobrossy L, Sinkovics I, Tarjan G, Daubner K. Fine-needle biopsy of thyroid gland. Journal of Surgical Oncology. 1979; 12(4):288-297

51. Borman KR, Hume AT. Credibility and clinical utility of thyroid fine-needle aspiration biopsy in a teaching hospital. American Journal of Surgery. 1995; 170(6):638-641; discussion 641-632

52. Bozbiyik O, Ozturk S, Unver M, Erol V, Bayol U, Aydin C. Reliability of fine needle aspiration biopsy in large thyroid nodules. Turkish Journal of Surgery. 2017; 33(1):10-13

53. Brauer RJ, Silver CE. Needle aspiration biopsy of thyroid nodules. Laryngoscope. 1984; 94(1):38-42

54. Breeze J, Poller DN, Gibson D, Tilley EA, Cooke L, Soar E et al. Rapid on-site assessment of specimens by biomedical scientists improves the quality of head and neck fine needle aspiration cytology. Cytopathology. 2014; 25(5):316-321

55. Bugis SP, Young JE, Archibald SD, Chen VS. Diagnostic accuracy of fine-needle aspiration biopsy versus frozen section in solitary thyroid nodules. American Journal of Surgery. 1986; 152(4):411-416

56. Burch HB, Burman KD, Reed HL, Buckner L, Raber T, Ownbey JL. Fine needle aspiration of thyroid nodules. Determinants of insufficiency rate and malignancy yield at thyroidectomy. Acta Cytologica. 1996; 40(6):1176-1183

57. Buzdar MU, Asim I, Hussain F, Qaisrani UK. Diagnostic accuracy of FNAC in cases of thyroid nodules while taking histopathology as gold standard. Pakistan Journal of Medical and Health Sciences. 2016; 14(4):1024-1026

58. Caleo A, Landolfi L, Vitale M, Di Crescenzo V, Vatrella A, De Rosa G et al. The diagnostic accuracy of fine-needle cytology of Hurthle cell lesions; A comprehensive cytological, clinical and ultrasonographic experience. International Journal of Surgery. 2016; 28(Suppl 1):S65-69

59. Camargo RY, Tomimori EK, Knobel M, Medeiros-Neto G. Preoperative assessment of thyroid nodules: role of ultrasonography and fine needle aspiration biopsy followed by cytology. Clinics (Sao Paulo, Brazil). 2007; 62(4):411-418

60. Can AS. Cost-effectiveness comparison between palpation- and ultrasound-guided thyroid fine-needle aspiration biopsies. BMC Endocrine Disorders. 2009; 9:14

61. Can AS, Peker K. Comparison of palpation-versus ultrasound-guided fine-needle aspiration biopsies in the evaluation of thyroid nodules. BMC Research Notes. 2008; 1:12

62. Cappelli C. Should we use stylet needles for aspiration cytology of thyroid nodules? Nature Clinical Practice Endocrinology & Metabolism. 2009; 5(2):84-85

63. Caraci P, Aversa S, Mussa A, Pancani G, Ondolo C, Conticello S. Role of fine-needle aspiration biopsy and frozen-section evaluation in the surgical management of thyroid nodules. British Journal of Surgery. 2002; 89(6):797-801

64. Carpi A, Ferrari E, De Gaudio C, Sagripanti A, Nicolini A, Di Coscio G. The value of aspiration needle biopsy in evaluating thyroid nodules. Thyroidology. 1994; 6(1):5-9

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

66. Chakravarthy NS, Chandramohan A, Prabhu AJ, Gowri M, Mannam P, Shyamkumar NK et al. Ultrasound-guided Fine-needle Aspiration Cytology along with Clinical and Radiological Features in Predicting Thyroid Malignancy in Nodules >=1 cm. Indian Journal of Endocrinology and Metabolism. 2018; 22(5):597-604

67. Chang HY, Lin JD, Chen JF, Huang BY, Hsueh C, Jeng LB et al. Correlation of fine needle aspiration cytology and frozen section biopsies in the diagnosis of thyroid nodules. Journal of Clinical Pathology. 1997; 50(12):1005-1009

68. Chen H, Nicol TL, Zeiger MA, Dooley WC, Ladenson PW, Cooper DS et al. Hurthle cell neoplasms of the thyroid: are there factors predictive of malignancy? Annals of Surgery. 1998; 227(4):542-546

69. Choden S, Wangmo C, Maharjan S. Application of the Bethesda system for reporting thyroid cytopathology for classification of thyroid nodules: A clinical and cytopathological characteristics in Bhutanese population. Diagnostic Cytopathology. 2021; 49(11):1179-1187

70. Choe JY, Kwak Y, Kim M, Chung YR, Kim HJ, Kim YK et al. Utility of a formatted pathologic reporting system in thyroid core needle biopsy: A validation study of 1998 consecutive cases. Clinical Endocrinology. 2018; 88(1):96-104

71. Choi JS, Choi Y, Kim EK, Yoon JH, Youk JH, Han KH et al. A risk-adapted approach using US features and FNA results in the management of thyroid incidentalomas identified by 18F-FDG PET. Ultraschall in der Medizin. 2014; 35(1):51-58

72. Chow TL, Venu V, Kwok SP. Use of fine-needle aspiration cytology and frozen section examination in diagnosis of thyroid nodules. Australian and New Zealand Journal of Surgery. 1999; 69(2):131-133

73. Chowdhury J, Das S, Maji D. A study on thyroid nodules: diagnostic correlation between fine needle aspiration cytology and histopathology. Journal of the Indian Medical Association. 2008; 106(6):389-390

74. Christ ML, Haja J. Intranuclear cytoplasmic inclusions (invaginations) in thyroid aspirations. Frequency and specificity. Acta Cytologica. 1979; 23(4):327-331

75. Chu EW, Hanson TA, Goldman JM, Robbins J. Study of cells in fine needle aspirations of the thyroid gland. Acta Cytologica. 1979; 23(4):308-314

76. Ciatti S, Bartolozzi C, Cicchi P, Lucarelli E. The role of ultrasonography and ultrasound guided biopsy in the management of patients with cold nodules of the thyroid. Ultrasound in Medicine and Biology. 1983; (Suppl 2):387-391

77. Ciobanu D, Caruntu ID, Vulpoi C, Florea N, Giusca SE. Morphometric parameters and silver stain used in diagnosis of thyroid follicular diseases. Romanian Journal of Morphology and Embryology. 2006; 47(4):323-330

78. Clary KM, Condel JL, Liu Y, Johnson DR, Grzybicki DM, Raab SS. Interobserver variability in the fine needle aspiration biopsy diagnosis of follicular lesions of the thyroid gland. Acta Cytologica. 2005; 49(4):378-382

79. Colacchio TA, LoGerfo P, Feind CR. Fine needle cytologic diagnosis of thyroid nodules. Review and report of 300 cases. American Journal of Surgery. 1980; 140(4):568-571

80. Cristallini EG, Bolis GB. Fine-needle aspiration biopsy in the preoperative diagnosis of solitary thyroid nodules. Applied Pathology. 1989; 7(3):149-153

81. Cristo AP, Goldstein HF, Faccin CS, Maia AL, Graudenz MS. Increasing diagnostic effectiveness of thyroid nodule evaluation by implementation of cell block preparation in routine US-FNA analysis. Archives of Endocrinology & Metabolism. 2016; 60(4):367-373

82. Cross P, Chandra A, Giles T, Johnson S, Kocjan G, Poller D et al. Guidance on the reporting of thyroid cytology specimens. London. The Royal College of Pathologists, 2016. Available from: https://www.rcpath.org/uploads/assets/7d693ce4-0091-4621-97f79e2a0d1034d6/g089_guidance_on_reporting_of_thyroid_cytology_specimens.pdf

83. Crowe A, Linder A, Hameed O, Salih C, Roberson J, Gidley J et al. The impact of implementation of the Bethesda System for Reporting Thyroid Cytopathology on the quality of reporting, "risk" of malignancy, surgical rate, and rate of frozen sections requested for thyroid lesions. Cancer Cytopathology. 2011; 119(5):315-321

84. Curtis L, Burns A. Unit costs of health and social care 2020. Canterbury. Personal Social Services Research Unit University of Kent, 2020. Available from: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2020/

85. Danese D, Sciacchitano S, Farsetti A, Andreoli M, Pontecorvi A. Diagnostic accuracy of conventional versus sonography-guided fine-needle aspiration biopsy of thyroid nodules. Thyroid. 1998; 8(1):15-21

86. Daskalakis A, Kostopoulos S, Spyridonos P, Glotsos D, Ravazoula P, Kardari M et al. Design of a multi-classifier system for discriminating benign from malignant thyroid nodules using routinely H&E-stained cytological images. Computers in Biology and Medicine. 2008; 38(2):196-203

87. Davidov T, Trooskin SZ, Shanker BA, Yip D, Eng O, Crystal J et al. Routine second-opinion cytopathology review of thyroid fine needle aspiration biopsies reduces diagnostic thyroidectomy. Surgery. 2010; 148(6):1294-1299; discussion 1299-1301

88. Davidsohn C, Cho C, Colandrea J, Lawrence G. Correlation of thyroid histopathology with fine needle aspiration of thyroid nodules: the St. Agnes Hospital experience. Maryland medical journal (Baltimore, Md. 1995; 44(1):36-38

89. Davoudi MM, Yeh KA, Wei JP. Utility of fine-needle aspiration cytology and frozen-section examination in the operative management of thyroid nodules. American Surgeon. 1997; 63(12):1084-1089; discussion 1089-1090

90. de Roy van Zuidewijn DB, Songun I, Hamming J, Kievit J, van de Velde CJ, Veselic M. Preoperative diagnostic tests for operable thyroid disease. World Journal of Surgery. 1994; 18(4):506-510; discussion 510-501

91. de Vos tot Nederveen Cappel RJ, Bouvy ND, Bonjer HJ, van Muiswinkel JM, Chadha S. Fine needle aspiration cytology of thyroid nodules: how accurate is it and what are the causes of discrepant cases? Cytopathology. 2001; 12(6):399-405

92. Dellal FD, Topaloglu O, Baser H, Dirikoc A, Alkan A, Altinboga AA et al. Are clinicopathological features of the isthmic thyroid nodule different from nodules in thyroid lobes? A single center experience. Archives of Endocrinology & Metabolism. 2021; 65(3):277-288

93. Deshpande V, Kapila K, Sai KS, Verma K. Follicular neoplasms of the thyroid. Decision tree approach using morphologic and morphometric parameters. Acta Cytologica. 1997; 41(2):369-376

94. Di Benedetto G, Fabozzi A, Rinaldi C. Clinical management of thyroid nodules with indeterminate cytology: our institutional experience using SIAPEC cytological criteria and V600-BRAF test. Pathologica. 2013; 105(1):1-4

95. Duek SD, Goldenberg D, Linn S, Krausz MM, Hershko DD. The role of fine-needle aspiration and intraoperative frozen section in the surgical management of solitary thyroid nodules. Surgery Today. 2002; 32(10):857-861

96. Dumitriu L, Mogos I, Calin E. Fine-needle aspiration biopsy of the thyroid correlated with clinical scintigraphic, echographic and pathologic data in nodular and diffuse goiter. Endocrinologie. 1984; 22(4):261-268

97. Dwarakanathan AA, Ryan WG, Staren ED, Martirano M, Economou SG. Fine-needle aspiration biopsy of the thyroid. Diagnostic accuracy when performing a moderate number of such procedures. Archives of Internal Medicine. 1989; 149(9):2007-2009

98. El Hag IA, Johnston J, Alessa E, Al Shammari M. Revised bethesda system for reporting thyroid cytology: Lessons learned from an appraisal of 5 years of experience in a central hospital. Cytopathology. 2021; 32(4):482-492

99. El Hag IA, Kollur SM, Chiedozi LC. The role of FNA in the initial management of thyroid lesions: 7-year experience in a district general hospital. Cytopathology. 2003; 14(3):126-130

100. Erdogan MF, Kamel N, Aras D, Akdogan A, Baskal N, Erdogan G. Value of re-aspirations in benign nodular thyroid disease. Thyroid. 1998; 8(12):1087-1090

101. Ersoz S, Mungan S, Sagnak Z, Ersoz HO, Odaci E. Nuclear morphometry for the differentiation of benign or malignant lesions with the diagnosis of atypia of undetermined significance/follicular lesion of undetermined significance in fine needle aspiration biopsy. Analytical and Quantitative Cytopathology and Histopathology. 2016; 38(5):288-294

102. Essex-Sorlie D, Penning CL, Freund GG. Impact of qualified (indeterminate) diagnoses on the accuracy of renal, thyroid, and breast fine-needle aspiration biopsy. American Journal of Clinical Pathology. 2000; 114(4):571-575

103. F IW, Fawad Khan S, Ur Rehman H, Ahmad Khan I. Role of fine needle aspiration cytology in diagnosis of solitary thyroid nodules. Iranian journal of otorhinolaryngology. 2011; 23(65):111-118

104. Fadda G, Minimo C, Rabitti C, Balsamo G, Verzi A, Gullotta G et al. Role of planimetric analysis in diagnosing thyroid follicular lesions on fine needle aspiration biopsies: a study with histologic correlation. Analytical and Quantitative Cytology and Histology. 1998; 20(3):192-198

105. Feletti F, Mellini L, Pironi F, Carnevale A, Parenti GC. Role of the cytopathologist during the procedure of fine-needle aspiration biopsy of thyroid nodules. Insights Into Imaging. 2021; 12(1):111

106. Ferrari E, Vailati A, Marelli G. The nodular goitre: A diagnostic approach. Medecine Biologie Environnement. 1985; 13(1):569-577

107. Ferraz de Oliveira AC, Destefani C, De Brot L, Lacerda D, Moreira FA, Pinto C et al. The usefulness of fine-needle aspirates for detection of recurrent carcinoma in the thyroid bed. Journal of the American Society of Cytopathology JASC. 2019; 8(1):34-38

108. Fiorentino V, Dell' Aquila M, Musarra T, Martini M, Capodimonti S, Fadda G et al. The role of cytology in the diagnosis of subcentimeter thyroid lesions. Diagnostics. 2021; 11(6)

109. Flanagan MB, Ohori NP, Carty SE, Hunt JL. Repeat thyroid nodule fine-needle aspiration in patients with initial benign cytologic results. American Journal of Clinical Pathology. 2006; 125(5):698-702

110. Fon LJ, Deans GT, Lioe TF, Lawson JT, Briggs K, Spence RA. An audit of thyroid surgery in a general surgical unit. Annals of the Royal College of Surgeons of England. 1996; 78(3 (Pt 1)):192-196

111. Frable MA, Frable WJ. Fine-needle aspiration biopsy revisited. Laryngoscope. 1982; 92(12):1414-1418

112. Frable MA, Frable WJ. Thin needle aspiration biopsy of the thyroid gland. Laryngoscope. 1980; 90(10 Pt 1):1619-1625

113. Frable WJ. The treatment of thyroid cancer. The role of fine-needle aspiration cytology. Archives of Otolaryngology -- Head & Neck Surgery. 1986; 112(11):1200-1203

114. Frable WJ, Frable MA. Thin-needle aspiration biopsy: the diagnosis of head and neck tumors revisited. Cancer. 1979; 43(4):1541-1548

115. Francis IM, Das DK. Role of fine needle aspiration, intraoperative imprint cytology and frozen section in the diagnosis of breast lumps and thyroid lesions. Medical Principles and Practice. 1999; 8(3):173-182

116. Franklyn JA, Daykin J, Young J, Oates GD, Sheppard MC. Fine needle aspiration cytology in diffuse or multinodular goitre compared with solitary thyroid nodules. BMJ. 1993; 307(6898):240

117. Franklyn JA, Fitzgerald MG, Oates GD, Sheppard MC. Fine needle aspiration cytology in the management of euthyroid goitre. Quarterly Journal of Medicine. 1987; 65(248):997-1003

118. Friedman M, Shimaoka K, Getaz P. Needle aspiration of 310 thyroid lesions. Acta Cytologica. 1979; 23(3):194-203

119. Frost AR, Sidawy MK, Ferfelli M, Tabbara SO, Bronner NA, Brosky KR et al. Utility of thin-layer preparations in thyroid fine-needle aspiration: diagnostic accuracy, cytomorphology, and optimal sample preparation. Cancer. 1998; 84(1):17-25

120. Fulciniti F, Benincasa G, Vetrani A, Palombini L. Follicular variant of papillary carcinoma: cytologic findings on FNAB samples-experience with 16 cases. Diagnostic Cytopathology. 2001; 25(2):86-93

121. Furlan JC, Bedard YC, Rosen IB. Single versus sequential fine-needle aspiration biopsy in the management of thyroid nodular disease. Canadian Journal of Surgery. 2005; 48(1):12-18

122. Galimberti A, Vitri P, De Pasquale L, Gobbi G, Bastagli A. Utility of fine needle aspiration and frozen section in the diagnosis of uncommon thyroid malignancies. Journal of Experimental and Clinical Cancer Research. 1997; 16(4):425-426

123. Gardiner GW, de Souza FM, Carydis B, Seemann C. Fine-needle aspiration biopsy of the thyroid gland: results of a five-year experience and discussion of its clinical limitations. Journal of Otolaryngology. 1986; 15(3):161-165

124. 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

125. Garg S, Desai NJ, Mehta D, Vaishnav M. To establish bethesda system for diagnosis of thyroid nodules on the basis of fnac with histopathological correlation. Journal of Clinical and Diagnostic Research JCDR. 2015; 9(12):EC17-21

126. Gershengorn MC, McClung MR, Chu EW, Hanson TA, Weintraub BD, Robbins J. Fine-needle aspiration cytology in the preoperative diagnosis of thyroid nodules. Annals of Internal Medicine. 1977; 87(3):265-269

127. Giansanti M, Monico S, Fugiani P. Fine-needle aspiration cytodiagnosis of the "cold" thyroid nodule. Tumori. 1989; 75(5):475-477

128. Gibb GK, Pasieka JL. Assessing the need for frozen sections: still a valuable tool in thyroid surgery. Surgery. 1995; 118(6):1005-1009; discussion 1009-1010

129. Godinho-Matos L, Kocjan G, Kurtz A. Contribution of fine needle aspiration cytology to diagnosis and management of thyroid disease. Journal of Clinical Pathology. 1992; 45(5):391-395

130. Goldfarb WB, Bigos TS, Eastman RC, Johnston H, Nishyama RH. Needle biopsy in the assessment and management of hypofunctioning thyroid nodules. American Journal of Surgery. 1982; 143(4):409-412

131. Gossain VV, Charnas J, Carella MJ, Rovner DR, Calaca WM. Evaluation of "solitary" thyroid nodules in a community practice: a managed care approach. American Journal of Managed Care. 1998; 4(5):679-684

132. Goulart APFE, Batista ERM, Figueira MG, Magalhaes PKR, Maciel LMZ. Evaluation of thyroid nodules in the Brazilian Public Health Care System, Supplementary Health System, and Private Health System in the northeastern region of the State of Sao Paulo. Archives of endocrinology and metabolism. 2021; 64(6):779-786

133. Gould E, Watzak L, Chamizo W, Albores-Saavedra J. Nuclear grooves in cytologic preparations. A study of the utility of this feature in the diagnosis of papillary carcinoma. Acta Cytologica. 1989; 33(1):16-20

134. Granados-Garcia M, Cortes-Flores AO, del Carmen Gonzalez-Ramirez I, Cano-Valdez AM, Flores-Hernandez L, Aguilar-Ponce JL. Follicular neoplasms of the thyroid: importance of clinical and cytological correlation. Cirugía y Cirujanos. 2010; 78(6):473-478

135. Greenblatt DY, Woltman T, Harter J, Starling J, Mack E, Chen H. Fine-needle aspiration optimizes surgical management in patients with thyroid cancer. Annals of Surgical Oncology. 2006; 13(6):859-863

136. Guadagni S, Amicucci G, Mariani G, Pietroletti R, Catarci M, Di Felice S et al. Diagnostic value of fine needle aspiration cytology in 98 patients with 'cold' thyroid nodules. Journal of Experimental and Clinical Cancer Research. 1988; 7(4):217-222

137. Gunes P, Demirturk P, Aker F, Tanriover O, Gonultas A, Akkaynak S. Evaluation of fine-needle aspiration of thyroid nodules in a series of 1,100 patients: Correlation between cytology and histopathology original article. Indian Journal of Surgery. 2015; 77(Suppl 3):990-995

138. Guo HQ, Zhang ZH, Zhao H, Niu LJ, Chang Q, Pan QJ. Factors influencing the reliability of thyroid fine-needle aspiration: Analysis of thyroid nodule size, guidance mode for aspiration and preparation method. Acta Cytologica. 2015; 59(2):169-174

139. Gupta Y, Bist SS, Agrawal V, Mishra S. Study of validity and reliability of fine needle aspiration cytology and Tc99m scintigraphy in thyroid swelling. World Journal of Endocrine Surgery. 2016; 8(2):143-146

140. H ZH. Early diagnosis of thyroid micro carcinoma based on needle aspiration cytology. Acta Medica Mediterranea. 2019; 35(4):1747-1751

141. Ha EJ, Na DG, Baek JH, Sung JY, Kim JH, Kang SY. Us fine-needle aspiration biopsy for thyroid malignancy: Diagnostic performance of seven society guidelines applied to 2000 thyroid nodules. Radiology. 2018; 287(3):893-900

142. Ha SM, Baek JH, Na DG, Jung CK, Suh CH, Shong YK et al. Assessing the diagnostic performance of thyroid biopsy with recommendations for appropriate interpretation. Ultrasonography. 2021; 40(2):228-236

143. Haas S, Trujillo A, Kunstle J. Fine needle aspiration of thyroid nodules in a rural setting. American Journal of Medicine. 1993; 94(4):357-361

144. Haberal AN, Toru S, Ozen O, Arat Z, Bilezikci B. Diagnostic pitfalls in the evaluation of fine needle aspiration cytology of the thyroid: correlation with histopathology in 260 cases. Cytopathology. 2009; 20(2):103-108

145. Haider AS, Rakha EA, Dunkley C, Zaitoun AM. The impact of using defined criteria for adequacy of fine needle aspiration cytology of the thyroid in routine practice. Diagnostic Cytopathology. 2011; 39(2):81-86

146. Hamaker RC, Singer MI, DeRossi RV, Shockley WW. Role of needle biopsy in thyroid nodules. Archives of Otolaryngology. 1983; 109(4):225-228

147. Hamburger JI. Needle aspiration for thyroid nodules. Skip ultrasound--do initial assessment in the office. Postgraduate Medicine. 1988; 84(8):61-66

148. Hamburger JI, Hamburger SW. Declining role of frozen section in surgical planning for thyroid nodules. Surgery. 1985; 98(2):307-312

149. Hamming JF, Goslings BM, van Steenis GJ, van Ravenswaay Claasen H, Hermans J, van de Velde CJ. The value of fine-needle aspiration biopsy in patients with nodular thyroid disease divided into groups of suspicion of malignant neoplasms on clinical grounds. Archives of Internal Medicine. 1990; 150(1):113-116

150. Hamming JF, Vriens MR, Goslings BM, Songun I, Fleuren GJ, van de Velde CJ. Role of fine-needle aspiration biopsy and frozen section examination in determining the extent of thyroidectomy. World Journal of Surgery. 1998; 22(6):575-579; discussion 579-580

151. Harach HR. Usefulness of fine needle aspiration of the thyroid in an endemic goiter region. Acta Cytologica. 1989; 33(1):31-35

152. Harsoulis P, Leontsini M, Economou A, Gerasimidis T, Smbarounis C. Fine needle aspiration biopsy cytology in the diagnosis of thyroid cancer: comparative study of 213 operated patients. British Journal of Surgery. 1986; 73(6):461-464

153. Hawkins F, Bellido D, Bernal C, Rigopoulou D, Ruiz Valdepenas MP, Lazaro E et al. Fine needle aspiration biopsy in the diagnosis of thyroid cancer and thyroid disease. Cancer. 1987; 59(6):1206-1209

154. Hawkins SP, Jamieson SG, Coomarasamy CN, Low IC. The global epidemic of thyroid cancer overdiagnosis illustrated using 18 months of consecutive nodule biopsy correlating clinical priority, ACR-TIRADS and Bethesda scoring. [Review]. Journal of Medical Imaging and Radiation Oncology. 2021; 65(3):309-316

155. Heimann P, Schnuerer LB. Needle biopsy of the thyroid gland. A report of 117 cases, with special reference to the diagnostic accuracy of the method in benign thyroid disorders. Acta Chirurgica Scandinavica. 1964; 128:85-93

156. Hirokawa M, Suzuki A, Hashimoto Y, Satoh S, Canberk S, Jhuang JY et al. Prevalence and diagnostic challenges of thyroid lymphoma: a multi-institutional study in non-Western countries. Endocrine Journal. 2020; 67(11):1085-1091

157. Hoffman WH. Diagnostic accuracy of fine needle aspiration biopsy in the diagnosis of thyroid malignancy. Pathologist. 1986; 40(2):9-14

158. Hong MJ, Na DG, Lee H. Diagnostic efficacy and safety of core needle biopsy as a first-line diagnostic method for thyroid nodules: A prospective cohort study. Thyroid. 2020; 30(8):1141-1149

159. Hosokawa S, Takebayashi S, Sasaki Y, Nakamura Y, Shinmura K, Takahashi G et al. Clinical analysis of false-negative fine needle aspiration cytology of head and neck cancers. Postgraduate Medicine. 2019; 131(2):151-155

160. Hougaard Chakera A, Stangerup SE, Andreassen UK, Christensen NR. Fine needle cytology in thyroid tumours, with and without aspiration a comparative study. Indian Journal of Otology. 2003; 9(2):9-13

161. 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

162. Hurtado-Lopez LM, Arellano-Montano S, Torres-Acosta EM, Zaldivar-Ramirez FR, Duarte-Torres RM, Alonso-De-Ruiz P et al. Combined use of fine-needle aspiration biopsy, MIBI scans and frozen section biopsy offers the best diagnostic accuracy in the assessment of the hypofunctioning solitary thyroid nodule. European Journal of Nuclear Medicine and Molecular Imaging. 2004; 31(9):1273-1279

163. Hussain ST, Beeby I, Missan A, Buxton-Thomas MS. Use of fine needle aspiration cytology in the management of the solitary cold thyroid nodule. Nuclear Medicine Communications. 1993; 14(4):335-338

164. Irish JC, van Nostrand AW, Asa SL, Gullane P, Rotstein L. Accuracy of pathologic diagnosis in thyroid lesions. Archives of Otolaryngology -- Head & Neck Surgery. 1992; 118(9):918-922

165. Irkorucu O, Tascilar O, Cakmak GK, Emre AU, Ucan HB, Kemal K et al. Frozen section and fine needle aspiration biopsy in thyroid surgery - needles and sections. Indian Journal of Surgery. 2007; 69(4):140-144

166. 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

167. Jat MA. Comparison of surgeon-performed ultrasound-guided fine needle aspiration cytology with histopathological diagnosis of thyroid nodules. Pakistan Journal of Medical Sciences. 2019; 35(4):1003-1007

168. Jayaram G, Razak A, Gan SK, Alhady SF. Fine needle aspiration cytology of the thyroid--a review of experience in 1853 cases. Malaysian Journal of Pathology. 1999; 21(1):17-27

169. Jing X, Knoepp SM, Roh MH, Hookim K, Placido J, Davenport R et al. Group consensus review minimizes the diagnosis of "follicular lesion of undetermined significance" and improves cytohistologic concordance. Diagnostic Cytopathology. 2012; 40(12):1037-1042

170. Kakudo K, Kameyama K, Hirokawa M, Katoh R, Nakamura H. Subclassification of follicular neoplasms recommended by the Japan thyroid association reporting system of thyroid cytology. International Journal of Endocrinology Print. 2015; 2015:938305

171. Karadeniz E, Yur M, Temiz A, Akcay MN. Malignancy risk for thyroid nodules larger than 4 cm and diagnostic reliability of ultrasound-guided FNAB results. Turkish Journal of Surgery. 2019; 35(1):13-18

172. Karstrup S, Balslev E, Juul N, Eskildsen PC, Baumbach L. US-guided fine needle aspiration versus coarse needle biopsy of thyroid nodules. European Journal of Ultrasound. 2001; 13(1):1-5

173. Katagiri M, Harada T, Kiyono T. Diagnosis of thyroid carcinoma by ultrasonic examination: comparison with diagnosis by fine needle aspiration cytology. Thyroidology. 1994; 6(1):21-26

174. Kawai T, Nishihara E, Kudo T, Ota H, Morita S, Kobayashi K et al. Histopathological diagnoses of "accessory" thyroid nodules diagnosed as benign by fine-needle aspiration cytology and ultrasonography. Thyroid. 2012; 22(3):299-303

175. Kelman AS, Rathan A, Leibowitz J, Burstein DE, Haber RS. Thyroid cytology and the risk of malignancy in thyroid nodules: importance of nuclear atypia in indeterminate specimens. Thyroid. 2001; 11(3):271-277

176. Kendall CH. Fine needle aspiration of thyroid nodules: three years' experience. Journal of Clinical Pathology. 1989; 42(1):23-27

177. Khan AR, Yasin SB, Makhdoomi R, Bhat SA. Fine needle aspiration cytology of "thyroid neoplasms" a ten year (1993-2003) study based on 400 neoplasms. JK Practitioner. 2004; 11(2):135-143

178. Khan DM, Srividhya VVL, Manimaran D, Ramakrishna BA. Pattern of thyroid neoplasms in nellore area - a clinicopathological correlation. International Journal of Pharma and Bio Sciences. 2013; 4(4):B1344-B1351

179. Khan EM, Pandey R. Differential diagnosis of fine needle aspiration smears of thyroid nodules. Cytologic features and AgNORs. Acta Cytologica. 1996; 40(5):959-962

180. Kikuchi S, Perrier ND, Ituarte PH, Treseler PA, Siperstein AE, Duh QY et al. Accuracy of fine-needle aspiration cytology in patients with radiation-induced thyroid neoplasms. British Journal of Surgery. 2003; 90(6):755-758

181. Kim DL, Song KH, Kim SK. High prevalence of carcinoma in ultrasonography-guided fine needle aspiration cytology of thyroid nodules. Endocrine Journal. 2008; 55(1):135-142

182. Kim DW, Jung SJ, In HS, Eom JW, Ryu JH, Kim YW. Ultrasound-guided fine-needle aspiration of thyroid nodules measuring less than 5 mm: Effects on specimen adequacy and diagnosis. Acta Cytologica. 2013; 57(1):38-44

183. Kim HK, Kim SY, Lee YS, Soh EY, Chang HS, Park CS. Suspicious thyroid nodules >=4 cm require diagnostic lobectomy regardless of their benign fine needle aspiration results. Asian journal of surgery. 2021; Epub

184. Kim JH, Kim NK, Oh YL, Kim HJ, Kim SY, Chung JH et al. The validity of ultrasonography-guided fine needle aspiration biopsy in thyroid nodules 4 cm or larger depends on ultrasonography characteristics. Endocrinology and Metabolism. 2014; 29(4):545-552

185. Kim JY, Chang S, Kwon AY, Park EY, Kim TH, Choi S et al. Core needle biopsy and ultrasonography are superior to fine needle aspiration in the management of follicular variant papillary thyroid carcinomas. Endocrine. 2022; 75(2):437-446

186. Kim SJ, Kim EK, Park CS, Chung WY, Oh KK, Yoo HS. Ultrasound-guided fine-needle aspiration biopsy in nonpalpable thyroid nodules: is it useful in infracentimetric nodules? Yonsei Medical Journal. 2003; 44(4):635-640

187. Kimoto T, Suemitsu K, Eda I, Shimizu T, Ohtani M, Nabika T. The efficiency of performing ultrasound-guided fine-needle aspiration biopsy following mass screening for thyroid tumors to avoid unnecessary surgery. Surgery Today. 1999; 29(9):880-883

188. Kini SR, Miller JM, Hamburger JI, Smith-Purslow MJ. Cytopathology of follicular lesions of the thyroid gland. Diagnostic Cytopathology. 1985; 1(2):123-132

189. Kini SR, Miller JM, Hamburger JI, Smith MJ. Cytopathology of papillary carcinoma of the thyroid by fine needle aspiration. Acta Cytologica. 1980; 24(6):511-521

190. Kizilkaya MC, Erozgen F, Akinci M, Kaplan R, Tuzun S, Citlak G. The predictive value of elastography in thyroid nodules and its comparison with fine-needle aspiration biopsy results. Turkish Journal of Surgery. 2014; 30(3):147-152

191. Kline TS, Neal HS. Needle biopsy. A pilot study. JAMA. 1973; 224(8):1143-1146

192. Knezevic-Usaj S, Eri Z, Panjkovic M, Klem I, Petrovic T, Ivkovic-Kapicl T et al. Diagnostic relevance of fine needle aspiration cytology in nodular thyroid lesions. Vojnosanitetski Pregled. 2012; 69(7):555-561

193. Kojic Katovic S, Halbauer M, Tomic-Brzac H. Importance of FNAC in the detection of tumours within multinodular goitre of the thyroid. Cytopathology. 2004; 15(4):206-211

194. Kolendorf K, Hansen JB, Engberg L, Friis T, Lindenberg J. Fine needle and open biopsy in thyroid disorders. Acta Chirurgica Scandinavica. 1975; 141(1):20-23

195. Kollur SM, El Sayed S, El Hag IA. Follicular thyroid lesions coexisting with Hashimoto's thyroiditis: incidence and possible sources of diagnostic errors. Diagnostic Cytopathology. 2003; 28(1):35-38

196. Kothari K, Tummidi S, Agnihotri M, Sathe P, Naik L. This 'rose' has no thorns-diagnostic utility of 'rapid on-site evaluation' (rose) in fine needle aspiration cytology. Indian Journal of Surgical Oncology. 2019; 10(4):688-698

197. Krishnappa P, Ramakrishnappa S, Kulkarni MH. Comparison of free hand versus ultrasound-guided fine needle aspiration of thyroid with histopathological correlation. Journal of Environmental Pathology, Toxicology and Oncology. 2013; 32(2):149-155

198. Kulstad R. Do all thyroid nodules >4 cm need to be removed? An evaluation of thyroid fine-needle aspiration biopsy in large thyroid nodules. Endocrine Practice. 2016; 22(7):791-798

199. Kumar A, Ahuja MM, Chattopadhyay TK, Padhy AK, Gupta AK, Kapila K et al. Fine needle aspiration cytology, sonography and radionuclide scanning in solitary thyroid nodule. Journal of the Association of Physicians of India. 1992; 40(5):302-306

200. La Rosa GL, Belfiore A, Giuffrida D, Sicurella C, Ippolito O, Russo G et al. Evaluation of the fine needle aspiration biopsy in the preoperative selection of cold thyroid nodules. Cancer. 1991; 67(8):2137-2141

201. Layfield LJ, Bentz JS, Gopez EV. Immediate on-site interpretation of fine-needle aspiration smears: a cost and compensation analysis. Cancer. 2001; 93(5):319-322

202. Lee J, Lee SY, Cha SH, Cho BS, Kang MH, Lee OJ. Fine-needle aspiration of thyroid nodules with macrocalcification. Thyroid. 2013; 23(9):1106-1112

203. Lee TI, Yang HJ, Lin SY, Lee MT, Lin HD, Braverman LE et al. The accuracy of fine-needle aspiration biopsy and frozen section in patients with thyroid cancer. Thyroid. 2002; 12(7):619-626

204. Leenhardt L, Hejblum G, Franc B, Fediaevsky LD, Delbot T, Le Guillouzic D et al. Indications and limits of ultrasound-guided cytology in the management of nonpalpable thyroid nodules. Journal of Clinical Endocrinology and Metabolism. 1999; 84(1):24-28

205. Lewis CM, Chang KP, Pitman M, Faquin WC, Randolph GW. Thyroid fine-needle aspiration biopsy: variability in reporting. Thyroid. 2009; 19(7):717-723

206. Li F, Luo H. Comparative study of thyroid puncture biopsy guided by contrast-enhanced ultrasonography and conventional ultrasound. Experimental and Therapeutic Medicine. 2013; 5(5):1381-1384

207. Li L, Chen X, Li P, Liu Y, Ma X, Ye YQ. The value of ultrasound-guided fine-needle aspiration cytology combined with puncture feeling in the diagnosis of thyroid nodules. Acta Cytologica. 2021; 65(5):368-376

208. Liel Y, Zirkin HJ, Sobel RJ. Fine needle aspiration of the thyroid. Five years' experience with 183 patients. Israel Journal of Medical Sciences. 1985; 21(9):719-721

209. Linhares SM, Handelsman R, Picado O, Farra JC, Lew JI. Fine needle aspiration and the Bethesda system: Correlation with histopathology in 1,228 surgical patients. Surgery. 2021; 170(5):1364-1368

210. Lioe TF, Elliott H, Allen DC, Spence RA. A 3-year audit of thyroid fine needle aspirates. Cytopathology. 1998; 9(3):188-192

211. Liu FH, Hsueh C, Chang HY, Liou MJ, Huang BY, Lin JD. Sonography and fine-needle aspiration biopsy in the diagnosis of benign versus malignant nodules in patients with autoimmune thyroiditis. Journal of Clinical Ultrasound. 2009; 37(9):487-492

212. Liu Z, Han R, Zhou W, Zhang J, Li H, Wan Z et al. Cytology versus calcitonin assay in fine-needle aspiration biopsy wash-out fluid (FNAB-CT) in diagnosis of medullary thyroid microcarcinoma. Endocrine. 2021; 74(2):340-348

213. Lo Gerfo P, Colacchio T, Caushaj F, Weber C, Feind C. Comparison of fine-needle and coarse-needle biopsies in evaluating thyroid nodules. Surgery. 1982; 92(5):835-838

214. Lobo C, McQueen A, Beale T, Kocjan G. The UK Royal College of Pathologists thyroid fine-needle aspiration diagnostic classification is a robust tool for the clinical management of abnormal thyroid nodules. Acta Cytologica. 2011; 55(6):499-506

215. Lodewijk L, Vriens MR, Vorselaars WM, van der Meij NT, Kist JW, Barentsz MW et al. Same-day fine-needle aspiration cytology diagnosis for thyroid nodules achieves rapid anxiety decrease and high diagnostic accuracy. Endocrine Practice. 2016; 22(5):561-566

216. Lopez LH, Canto JA, Herrera MF, Gamboa-Dominguez A, Rivera R, Gonzalez O et al. Efficacy of fine-needle aspiration biopsy of thyroid nodules: experience of a Mexican institution. World Journal of Surgery. 1997; 21(4):408-411

217. Lukitto P. Evaluation of the results of fine needle aspiration biopsy cytology on thyroid nodules and breast tumors at the Hasan Sadikin Hospital Bandung. Medical Journal of Indonesia. 1998; 7(1):3-7

218. 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

219. Makes B. Accuracy of frozen-section combined with imprint and fine needle aspiration biopsy in thyroid nodules. Medical Journal of Indonesia. 2007; 16(2):89-93

220. Malberger E, Kraus M, Lemberg S. Diagnostic accuracy of thyroid aspirative cytology in view of cumulative experience. Israel Journal of Medical Sciences. 1985; 21(9):713-718

221. Mamoon N, Mushtaq S, Muzaffar M, Khan AH. The use of fine needle aspiration biopsy in the management of thyroid disease. JPMA - Journal of the Pakistan Medical Association. 1997; 47(10):255-258

222. Manchanda GS, Mohan A, Garg N, Thakral RK, Bharti S, Sharma VK et al. Comparative study of aspiration and non aspiration techniques in diagnosis of thyroid lesions. Indian Journal of Public Health Research and Development. 2018; 9(3):16-19

223. Mandal S, Barman D, Mukherjee A, Mukherjee D, Saha J, Sinhas R. Fine needle aspiration cytology of thyroid nodules--evaluation of its role in diagnosis and management. Journal of the Indian Medical Association. 2011; 109(4):258-261

224. Mandreker SRS, Nadkarni NS, Pinto RGW, Menezes S. Role of fine needle aspiration cytology as the initial modality in the investigation of thyroid lesions. Acta Cytologica. 1995; 39(5):898-904

225. Martinek A, Dvorackova J, Honka M, Horacek J, Klvana P. Importance of guided fine needle aspiration cytology (FNAC) for the diagnostics of thyroid nodules - own experience. Biomedical Papers of the Medical Faculty of Palacky University in Olomouc, Czech Republic. 2004; 148(1):45-50

226. Maruta J, Hashimoto H, Yamashita H, Yamashita H, Noguchi S. Quick aspiration cytology for thyroid nodules by modified Ultrafast Papanicolaou staining. Diagnostic Cytopathology. 2003; 28(1):45-48

227. Mary Lilly S, Ramamoorthy V. Comparison of fine needle aspiration cytology with histopathology in thyroid swellings & its diagnostic accuracy. Indian Journal of Public Health Research and Development. 2019; 10(11):4237-4240

228. Masatsugu T, Yamashita H, Noguchi S, Nishii R, Koga Y, Watanabe S et al. Thyroid evaluation in patients with primary hyperparathyroidism. Endocrine Journal. 2005; 52(2):177-182

229. Mastorakis E, Meristoudis C, Margari N, Pouliakis A, Leventakos K, Chroniaris N et al. Fine needle aspiration cytology of nodular thyroid lesions: a 2-year experience of the Bethesda system for reporting thyroid cytopathology in a large regional and a university hospital, with histological correlation. Cytopathology. 2014; 25(2):120-128

230. Mathur SR, Kapila K, Verma K. Role of fine needle aspiration cytology in the diagnosis of goiter. Indian Journal of Pathology and Microbiology. 2005; 48(2):166-169

231. Maxwell JG, Scallion RR, White WC, Kotwall CA, Pollock H, Covington DL et al. Fine-needle aspiration cytology and thyroid surgery in the community hospital. American Journal of Surgery. 1996; 172(5):529-534; discussion 534-525

232. McCoy KL, Jabbour N, Ogilvie JB, Ohori NP, Carty SE, Yim JH. The incidence of cancer and rate of false-negative cytology in thyroid nodules greater than or equal to 4 cm in size. Surgery. 2007; 142(6):837-844; discussion 844.e831-833

233. McElroy MK, Mahooti S, Hasteh F. A single institution experience with the new bethesda system for reporting thyroid cytopathology: correlation with existing cytologic, clinical, and histological data. Diagnostic Cytopathology. 2014; 42(7):564-569

234. McHenry CR, Thomas SR, Slusarczyk SJ, Khiyami A. Follicular or Hurthle cell neoplasm of the thyroid: can clinical factors be used to predict carcinoma and determine extent of thyroidectomy? Surgery. 1999; 126(4):798-802; discussion 802-794

235. McIvor NP, Freeman JL, Rosen I, Bedard YC. Value of fine-needle aspiration in the diagnosis of Hurthle cell neoplasms. Head and Neck. 1993; 15(4):335-341

236. Mehrotra P, Viswanathan H, Johnson SJ, Wadehra V, Richardson DL, Lennard TW. Ultrasound guidance improves the adequacy of our preoperative thyroid cytology but not its accuracy. Cytopathology. 2006; 17(3):137-144

237. Meko JB, Norton JA. Large cystic/solid thyroid nodules: a potential false-negative fine-needle aspiration. Surgery. 1995; 118(6):996-1003; discussion 1003-1004

238. Meng C, Hinkle LE, Wang W, Su D, Li X. Hashimoto's thyroiditis elicits decreased diagnostic efficacy of thyroid nodule ultrasound-guided fine needle aspiration. International Journal of Clinical and Experimental Pathology. 2019; 12(9):3474-3482

239. Merchant WJ, Thomas SM, Coppen MJ, Prentice MG. The role of thyroid fine needle aspiration (FNA) cytology in a District General Hospital setting. Cytopathology. 1995; 6(6):409-418

240. Mijovic T, Rochon L, Gologan O, Hier MP, Black MJ, Young J et al. Fine-needle aspiration biopsies in the management of indeterminate follicular and Hurthle cell thyroid lesions. Otolaryngology - Head & Neck Surgery. 2009; 140(5):715-719

241. Mikosch P, Gallowitsch HJ, Kresnik E, Jester J, Wurtz FG, Kerschbaumer K et al. Value of ultrasound-guided fine-needle aspiration biopsy of thyroid nodules in an endemic goitre area. European Journal of Nuclear Medicine. 2000; 27(1):62-69

242. Miller JM, Hamburger JI, Kini S. Diagnosis of thyroid nodules. Use of fine-needle aspiration and needle biopsy. JAMA. 1979; 241(5):481-484

243. Miller JM, Hamburger JI, Kini SR. The needle biopsy diagnosis of papillary thyroid carcinoma. Cancer. 1981; 48(4):989-993

244. Miller JM, Kini SR, Hamburger JI. The diagnosis of malignant follicular neoplasms of the thyroid by needle biopsy. Cancer. 1985; 55(12):2812-2817

245. Miller TR, Bottles K, Holly EA, Friend NF, Abele JS. A step-wise logistic regression analysis of papillary carcinoma of the thyroid. Acta Cytologica. 1986; 30(3):285-293

246. Mo HS, Li ZX, Wang SD, Liao XH, Liang M, Hao XY. Ultrasonic features of thyroid nodules related to the false negativity in ultrasound-guided fine-needle aspiration for suspicious malignant thyroid nodules. International Journal of Clinical and Experimental Medicine. 2017; 10(9):13473-13481

247. Montironi R, Alberti R, Sisti S, Braccischi A, Scarpelli M, Mariuzzi GM. Discrimination between follicular adenoma and follicular carcinoma of the thyroid: preoperative validity of cytometry on aspiration smears. Applied Pathology. 1989; 7(6):367-374

248. Montironi R, Braccischi A, Scarpelli M, Sisti S, Alberti R. Well differentiated follicular neoplasms of the thyroid: reproducibility and validity of a 'decision tree' classification based on nucleolar and karyometric features. Cytopathology. 1992; 3(4):209-222

249. Montironi R, Braccischi A, Scarpelli M, Sisti S, Matera G, Mariuzzi GM et al. The number of nucleoli in benign and malignant thyroid lesions: a useful diagnostic sign in cytological preparations. Cytopathology. 1990; 1(3):153-161

250. Mora-Guzman I, Munoz de Nova JL, Marin-Campos C, Jimenez-Heffernan JA, Cuesta Perez JJ, Lahera Vargas M et al. Efficiency of the bethesda system for thyroid cytopathology. Cirugia Espanola. 2018; 96(6):363-368

251. Morgan JL, Serpell JW, Cheng MS. Fine-needle aspiration cytology of thyroid nodules: how useful is it? ANZ Journal of Surgery. 2003; 73(7):480-483

252. Munn JS, Castelli M, Prinz RA, Walloch JL. Needle biopsy of nodular thyroid disease. American Surgeon. 1988; 54(7):438-443

253. Muratli A, Erdogan N, Sevim S, Unal I, Akyuz S. Diagnostic efficacy and importance of fine-needle aspiration cytology of thyroid nodules. Journal of Cytology. 2014; 31(2):73-78

254. Na DG, Kim JH, Sung JY, Baek JH, Jung KC, Lee H et al. Core-needle biopsy is more useful than repeat fine-needle aspiration in thyroid nodules read as nondiagnostic or atypia of undetermined significance by the Bethesda system for reporting thyroid cytopathology. Thyroid. 2012; 22(5):468-475

255. Na DG, Min HS, Lee H, Won JK, Seo HB, Kim JH. Role of core needle biopsy in the management of atypia/follicular lesion of undetermined significance thyroid nodules: Comparison with repeat fine-needle aspiration in subcategory nodules. European Thyroid Journal. 2015; 4(3):189-196

256. Nagarajan N, Schneider EB, Ali SZ, Zeiger MA, Olson MT. How do liquid-based preparations of thyroid fine-needle aspiration compare with conventional smears? An analysis of 5475 specimens. Thyroid. 2015; 25(3):308-313

257. Nart D, Ertan Y, Argon A, Sezak M, Veral A, Makay O et al. Role of fine needle aspiration cytology and intraoperative diagnosis in the diagnosis of thyroid nodules. Turk Patoloji Dergisi/Turkish Journal of Pathology. 2010; 26(1):48-54

258. Natarajan V, Jayaram G, Kakar A, Prakash R. Solitary cold thyroid nodules -- a correlation of fine needle aspiration cytology with pentavalent technetium DMSA scanning and radionuclide perfusion scanning. Malaysian Journal of Pathology. 1994; 16(2):127-135

259. 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

260. Naz S, Hashmi AA, Khurshid A, Faridi N, Edhi MM, Kamal A et al. Diagnostic accuracy of Bethesda system for reporting thyroid cytopathology: an institutional perspective. International Archives of Medicine. 2014; 7:46

261. Ng EH, Thomas A, Nambiar R. Solitary thyroid nodules for surgery--the role of fine-needle aspiration biopsy cytology. Annals of the Academy of Medicine, Singapore. 1988; 17(1):15-18

262. Ng SC, Lin JD, Huang BY, Chen CH, Hsueh C, Lee N et al. Diagnosis and management of 34 Hurthle cell tumors. Changgeng Yi Xue Za Zhi Chang Gung Medical Journal. 1999; 22(3):445-452

263. NHS England and NHS Improvement. National Cost Collection Data Publication 2019-2020. London. 2020. Available from: https://www.england.nhs.uk/wp-content/uploads/2021/06/National-Cost-Collection-2019-20-Report-FINAL.pdf

264. Nirmal AK, Singh H, Jha JK. Fine needle aspiration of follicular lesions of the thyroid: Cytohistologic correlation and accuracy at Hapur region. Indian Journal of Public Health Research and Development. 2017; 8(4):68-72

265. Norton LW, Wangensteen SL. Needle aspiration biopsy of thyroid nodules. Arizona Medicine. 1981; 38(5):378-379

266. Okumura Y, Takeda Y, Sato S, Komatsu M, Nakagawa T, Akaki S et al. Comparison of differential diagnostic capabilities of 201Tl scintigraphy and fine-needle aspiration of thyroid nodules. Journal of Nuclear Medicine. 1999; 40(12):1971-1977

267. Ongphiphadhanakul B, Rajatanavin R, Chiemchanya S, Chailurkit L, Kongsuksai A, Isarangkul Na Ayuthya WI. Systematic inclusion of clinical and laboratory data improves diagnostic accuracy of fine-needle aspiration biopsy in solitary thyroid nodules. Acta Endocrinologica. 1992; 126(3):233-237

268. Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). 2021. Available from: http://www.oecd.org/std/ppp Last accessed: 24/03/2022.

269. Ozdemir D, Bestepe N, Faki S, Kilicarslan A, Parlak O, Ersoy R et al. Comparison of thyroid fine needle aspiration biopsy results before and after implementation of Bethesda classification. Cytopathology. 2017; 28(5):400-406

270. Pan X, Wang L. Comparison of diagnostic values between ultrasound elastography and ultrasound-guided thyroid nodular puncture in thyroid nodules. Oncology Letters. 2018; 16(4):5209-5213

271. Pasha HA, Mughal A, Wasif M, Dhanani R, Haider SA, Abbas SA. The efficacy of bethesda system for prediction of thyroid malignancies - A 9 year experience from a tertiary center. Iranian journal of otorhinolaryngology. 2021; 33(4):209-215

272. Patel K, Patel A, Shah K, Patel N. FNAC is a primary diagnostic tool in thyroid swelling. Biosciences Biotechnology Research Asia. 2014; 11(3):1873-1876

273. Pathologists RCo. Diagnostic cytopathology in the UK. 2021, Issue

274. Pavithra P, Rashmi MV. Utility of cytodiagnosis in the management of thyroid lesions. International Journal of Pharma and Bio Sciences. 2014; 5(4):B1173-B1182

275. Pepper GM, Zwickler D, Rosen Y. Fine-needle aspiration biopsy of the thyroid nodule. Results of a start-up project in a general teaching hospital setting. Archives of Internal Medicine. 1989; 149(3):594-596

276. Petersen SV, Greisen O. Fine-needle aspiration biopsy of the thyroid gland. Otolaryngology - Head & Neck Surgery. 1984; 92(3):295-297

277. Piana S, Frasoldati A, Ferrari M, Valcavi R, Froio E, Barbieri V et al. Is a five-category reporting scheme for thyroid fine needle aspiration cytology accurate? Experience of over 18,000 FNAs reported at the same institution during 1998-2007. Cytopathology. 2011; 22(3):164-173

278. Pisani T, Bononi M, Nagar C, Angelini M, Bezzi M, Vecchione A. Fine needle aspiration and core needle biopsy techniques in the diagnosis of nodular thyroid pathologies. Anticancer Research. 2000; 20(5C):3843-3847

279. Poller DN, Doyle V, Trimboli P, Bongiovanni M. Rates of Thy 1-non-diagnostic thyroid fine needle aspiration using the UK Royal College of Pathologists Thy Terminology. A systematic review of the literature comparing patients who undergo rapid on-site evaluation and those who do not. Cytopathology. 2020; 31(6):502-508

280. Poller DN, Kandaswamy P. A simplified economic approach to thyroid FNA cytology and surgical intervention in thyroid nodules. Journal of Clinical Pathology. 2013; 66(7):583

281. Postma DS, Becker MO, Roberts A, Gilleon S, Soto J. Thyroidectomy in a community hospital: Findings of 100 consecutive cases. Ear, Nose and Throat Journal. 2009; 88(5):E30

282. Prinz RA, O'Morchoe PJ, Barbato AL, Braithwaite SS, Brooks MH, Emanuele MA et al. Fine needle aspiration biopsy of thyroid nodules. Annals of Surgery. 1983; 198(1):70-73

283. Raab S, Veronezigurwell A. Thyroid-nodules in the elderly - clinical management and incidence of malignancy as determined by fine-needle aspiration biopsy. Oncology Reports. 1995; 2(6):1151-1155

284. Radetic M, Kralj Z, Padovan I. Reliability of aspiration biopsy in thyroid nodes: study of 2190 operated patients. Tumori. 1984; 70(3):271-276

285. Raina B, Misri A, Kanotra JP, Suhail M, Khajuria A, Gupta RK. Profile of fine needle aspiration cytology of thyroid nodule and its histopathological correlation. JK Practitioner. 2011; 16(1-2):87-91

286. Rammeh S, Romdhane E, Sassi A, Belhajkacem L, Blel A, Ksentini M et al. Accuracy of fine-needle aspiration cytology of head and neck masses. Diagnostic Cytopathology. 2019; 47(5):394-399

287. Rana C, Singh KR, Ramakant P, Babu S, Mishra A. Impact of cytological pitfalls in the Bethesda System of Reporting Thyroid Cytopathology, on surgical decision-making of patients with thyroid nodules: Can these pitfalls be avoided? Cytopathology. 2021; 32(2):192-204

288. Rangaswamy M, Narendra K, Patel S, Gururajprasad C, Manjunath G. Insight to neoplastic thyroid lesions by fine needle aspiration cytology. Journal of Cytology. 2013; 30(1):23-26

289. Rege JD, Nath AR, Bijlani JC, Trivedi DR, Deshpande DV. Fine needle aspiration cytology in solitary cold nodules of thyroid. Journal of the Association of Physicians of India. 1987; 35(12):819-821

290. Renshaw AA. Accuracy of thyroid fine-needle aspiration using receiver operator characteristic curves. American Journal of Clinical Pathology. 2001; 116(4):477-482

291. Renshaw AA. Hurthle cell carcinoma is a better gold standard than Hurthle cell neoplasm for fine-needle aspiration of the thyroid: defining more consistent and specific cytologic criteria. Cancer. 2002; 96(5):261-266

292. Renshaw AA, Gould EW. Characteristics of false-negative thyroid fine-needle aspirates. Acta Cytologica. 2018; 62(1):12-18

293. Renshaw AA, Pinnar N. Comparison of thyroid fine-needle aspiration and core needle biopsy. American Journal of Clinical Pathology. 2007; 128(3):370-374

294. Reyaz N, Baloch MB, Butt ME, Siddique S, Malik SH, Noor J. Comparative study of Fine Needle Aspiration Cytology (F.N.A.C) vs tissue biopsy in thyroid glands: Following bethesda system. Medical Forum Monthly. 2020; 31(9):136-140

295. Rodriguez JM, Parrilla P, Sola J, Bas A, Aguilar J, Moreno A et al. Comparison between preoperative cytology and intraoperative frozen-section biopsy in the diagnosis of thyroid nodules. British Journal of Surgery. 1994; 81(8):1151-1154

296. Rosen IB, Azadian A, Walfish PG, Salem S, Lansdown E, Bedard YC. Ultrasound-guided fine-needle aspiration biopsy in the management of thyroid disease. American Journal of Surgery. 1993; 166(4):346-349

297. Rosen IB, Provias JP, Walfish PG. Pathologic nature of cystic thyroid nodules selected for surgery by needle aspiration biopsy. Surgery. 1986; 100(4):606-613

298. Rosen IB, Wallace C, Strawbridge HG, Walfish PG. Reevaluation of needle aspiration cytology in detection of thyroid cancer. Surgery. 1981; 90(4):747-756

299. Roy PK, Bandyopadhyay S, Dubey AB, Sengupta A. A comparative study on aspiration cytology and histopathology in diagnosis of thyroid nodule and its correlation. Indian Journal of Otolaryngology & Head & Neck Surgery. 2019; 71(Suppl 1):997-1001

300. Rubenfeld S, Wheeler TM, Spjut HJ. Fine-needle aspiration biopsy of thyroid nodules. Texas Medicine. 1982; 78(9):41-44

301. Russ JE, Scanlon EF, Christ MA. Aspiration cytology of head and neck masses. American Journal of Surgery. 1978; 136(3):342-347

302. Sabel MS, Staren ED, Gianakakis LM, Dwarakanathan S, Prinz RA. Use of fine-needle aspiration biopsy and frozen section in the management of the solitary thyroid nodule. Surgery. 1997; 122(6):1021-1026; discussion 1026-1027

303. Sahin M, Sengul A, Berki Z, Tutuncu NB, Guvener ND. Ultrasound-guided fine-needle aspiration biopsy and ultrasonographic features of infracentimetric nodules in patients with nodular goiter: correlation with pathological findings. Endocrine Pathology. 2006; 17(1):67-74

304. Sangalli G, Serio G, Zampatti C, Lomuscio G, Colombo L. Fine needle aspiration cytology of primary lymphoma of the thyroid: A report of 17 cases. Cytopathology. 2001; 12(4):257-263

305. Sarda AK, Gupta A, Jain PK, Prasad S. Management options for solitary thyroid nodules in an endemic goitrous area. Postgraduate Medical Journal. 1997; 73(863):560-564

306. Sarkis LM, Norlen O, Aniss A, Watson N, Delbridge LW, Sidhu SB et al. The Australian experience with the Bethesda classification system for thyroid fine needle aspiration biopsies. Pathology. 2014; 46(7):592-595

307. Schmid KW, Hofstadter F, Propst A, Jr., Ladurner D, Zechmann W. A fourteen year practice with the fine needle aspiration biopsy of the thyroid in an endemic area. Pathology, Research and Practice. 1986; 181(3):308-310

308. Schnurer LB, Widstrom A. Fine-needle biopsy of the thyroid gland: a cytohistological comparison in cases of goiter. Annals of Otology, Rhinology and Laryngology. 1978; 87(2 Pt 1):224-227

309. Schoedel KE, Tublin ME, Pealer K, Ohori NP. Ultrasound-guided biopsy of the thyroid: a comparison of technique with respect to diagnostic accuracy. Diagnostic Cytopathology. 2008; 36(11):787-789

310. Schwartz AE, Nieburgs HE, Davies TF, Gilbert PL, Friedman EW. The place of fine needle biopsy in the diagnosis of nodules of the thyroid. Surgery, Gynecology and Obstetrics. 1982; 155(1):54-58

311. Sclabas GM, Staerkel GA, Shapiro SE, Fornage BD, Sherman SI, Vassillopoulou-Sellin R et al. Fine-needle aspiration of the thyroid and correlation with histopathology in a contemporary series of 240 patients. American Journal of Surgery. 2003; 186(6):702-709; discussion 709-710

312. Scurry JP, Duggan MA. Thin layer compared to direct smear in thyroid fine needle aspiration. Cytopathology. 2000; 11(2):104-115

313. Seifman MA, Grodski SF, Bailey M, Yeung MJ, Serpell JW. Surgery in the setting of Hashimoto's thyroiditis. ANZ Journal of Surgery. 2011; 81(7-8):519-523

314. Sengul I, Sengul D, Egrioglu E, Ozturk T. Laterality of the thyroid nodules, anatomic and sonographic, as an estimator of thyroid malignancy and its neoplastic nature by comparing the Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) and histopathology. Journal of BUON. 2020; 25(2):1116-1121

315. Seok JY, An J, Cho HY. Improvement of diagnostic performance of pathologists by reducing the number of pathologists responsible for thyroid fine needle aspiration cytology: An institutional experience. Diagnostic Cytopathology. 2018; 46(7):561-567

316. Settakorn J, Chaiwun B, Thamprasert K, Wisedmongkol W, Rangdaeng S. Fine needle aspiration of the thyroid gland. Journal of the Medical Association of Thailand. 2001; 84(10):1401-1406

317. Seya A, Oeda T, Terano T, Omura M, Tahara K, Nishikawa T et al. Comparative studies on fine-needle aspiration cytology with ultrasound scanning in the assessment of thyroid nodule. Japanese Journal of Medicine. 1990; 29(5):478-480

318. Sharma C. An analysis of trends of incidence and cytohistological correlation of papillary carcinoma of the thyroid gland with evaluation of discordant cases. Journal of Cytology. 2016; 33(4):192-198

319. Sharma R, Verma N, Kaushal V, Sharma DR, Sharma D. Diagnostic accuracy of fine-needle aspiration cytology of thyroid gland lesions: A study of 200 cases in Himalayan belt. Journal of Cancer Research and Therapeutics. 2017; 13(3):451-455

320. 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

321. Sheahan J, Fitzgibbon J, O'Leary G, Lee G. Efficacy and pitfalls of fine needle aspiration in the diagnosis of neck masses. Surgeon. 2004; 2(3):152-156

322. Shinkins B, Thompson M, Mallett S, Perera R. Diagnostic accuracy studies: how to report and analyse inconclusive test results. BMJ. 2013; 346:f2778

323. Shirzad M, Larijani B, Hedayat A, Kamalian N, Baradar-Jalili R, Bandarian F et al. Diagnostic value of frozen section examination in thyroid nodule--surgery at the Shariati Hospital (1997-2000). Endocrine Pathology. 2003; 14(3):263-268

324. Shrestha M, Crothers BA, Burch HB. The impact of thyroid nodule size on the risk of malignancy and accuracy of fine-needle aspiration: a 10-year study from a single institution. Thyroid. 2012; 22(12):1251-1256

325. Sidawy MK, Del Vecchio DM, Knoll SM. Fine-needle aspiration of thyroid nodules: correlation between cytology and histology and evaluation of discrepant cases. Cancer. 1997; 81(4):253-259

326. Silver CE, Brauer RJ, Schreiber K. Cytologic evaluation of thyroid nodules: New criteria for surgery. New York State Journal of Medicine. 1984; 84(3):109-112

327. Silverman JF, West RL, Finley JL, Larkin EW, Park HK, Swanson MS et al. Fine-needle aspiration versus large-needle biopsy or cutting biopsy in evaluation of thyroid nodules. Diagnostic Cytopathology. 1986; 2(1):25-30

328. Silverman JF, West RL, Larkin EW, Park HK, Finley JL, Swanson MS et al. The role of fine-needle aspiration biopsy in the rapid diagnosis and management of thyroid neoplasm. Cancer. 1986; 57(6):1164-1170

329. Sirpal YM. Efficacy of fine needle aspiration cytology in the management of thyroid diseases. Indian Journal of Pathology and Microbiology. 1996; 39(3):173-178

330. Slowinska-Klencka D, Popowicz B, Lewinski A, Sporny S, Klencki M. The fine-needle aspiration biopsy efficacy of small thyroid nodules in the area of recently normalized iodine supply. European Journal of Endocrinology. 2008; 159(6):747-754

331. Smadi AA, Ajarmeh K, Wreikat F. Fine-needle aspiration of thyroid nodules has high sensitivity and specificity. Rawal Medical Journal. 2008; 33(2):221-224

332. Son JI, Rhee SY, Woo JT, Park WS, Byun JK, Kim YJ et al. Insufficient experience in thyroid fine-needle aspiration leads to misdiagnosis of thyroid cancer. Endocrinology and Metabolism. 2014; 29(3):293-299

333. Soreide O, Varhaug JE, Heimann P. Thyroid carcinoma: diagnosis and treatment in 106 patients. Acta Chirurgica Scandinavica. 1979; 145(3):137-141

334. Spiliotis J, Scopa CD, Chalmoukis A, Androulakis J, Vagenakis A. Thyroid nodules indeterminate by fine needle aspiration biopsy. Surgical Research Communications. 1992; 12(3):233-236

335. Stanek-Widera A, Biskup-Fruzynska M, Snietura M, Zembala-Nozynska E, Sroda M, Szczesny-Karczewska W et al. Correspondence of cytological and histopathological diagnoses in diagnostic category V of the Bethesda system: "suspicious for malignancy". Polish Journal of Pathology. 2016; 67(1):24-32

336. Stanek-Widera A, Biskup-Fruzynska M, Zembala-Nozynska E, Poltorak S, Snietura M, Lange D. Suspicious for follicular neoplasm or follicular neoplasm? The dilemma of a pathologist and a surgeon. Endokrynologia Polska. 2016; 67(1):17-22

337. Stavric GD, Karanfilski BT, Kalamaras AK, Serafimov NZ, Georgievska BS, Korubin VH. Early diagnosis and detection of clinically non-suspected thyroid neoplasia by the cytologic method: a critical review of 1536 aspiration biopsies. Cancer. 1980; 45(2):340-344

338. Suh CH, Choi YJ, Lee JJ, Shim WH, Baek JH, Chung HC et al. Comparison of core-needle biopsy and fine-needle aspiration for evaluating thyroid incidentalomas detected by 18f-fluorodeoxyglucose positron emission tomography/computed tomography: A propensity score analysis. Thyroid. 2017; 27(10):1258-1266

339. Sukumaran R, Kattoor J, Pillai KR, Ramadas PT, Nayak N, Somanathan T et al. Fine needle aspiration cytology of thyroid lesions and its correlation with histopathology in a series of 248 patients. Indian Journal of Surgical Oncology. 2014; 5(3):237-241

340. Sulejmanovic M, Cickusic AJ, Salkic S. The value of fine-needle aspiration biopsy (fnab) in differential diagnosis of scintigraphic cold thyroid nodule. Acta Informatica Medica. 2019; 27(2):114-118

341. Suwatthanarak T, Prasert W. Diagnostic accuracy of fine needle aspiration cytology in thyroid nodules in thammasat university hospital. Journal of the Medical Association of Thailand. 2021; 104(10):1667-1670

342. Tabain I, Matesa N, Kusic Z. Accuracy of ultrasound guided fine needle aspiration in patients with nodular thyroid disease. Acta Clinica Croatica. 2004; 43(1):21-26

343. Tabaqchali MA, Hanson JM, Johnson SJ, Wadehra V, Lennard TW, Proud G. Thyroid aspiration cytology in Newcastle: a six year cytology/histology correlation study. Annals of the Royal College of Surgeons of England. 2000; 82(3):149-155

344. Takashima S, Fukuda H, Kobayashi T. Thyroid nodules: clinical effect of ultrasound-guided fine-needle aspiration biopsy. Journal of Clinical Ultrasound. 1994; 22(9):535-542

345. Takashima S, Matsuzuka F, Nagareda T, Tomiyama N, Kozuka T. Thyroid nodules associated with Hashimoto thyroiditis: assessment with US. Radiology. 1992; 185(1):125-130

346. Taki S, Kakuda K, Kakuma K, Annen Y, Katada S, Yamashita R et al. Thyroid nodules: evaluation with US-guided core biopsy with an automated biopsy gun. Radiology. 1997; 202(3):874-877

347. Tal A. Cold thyroid nodule: predictive value of fine needle aspiration biopsy at surgery. Journal of the Tennessee Medical Association. 1992; 85(8):369-371

348. Talpur KAH, Laghari AA, Malik AM, Khan SA. Role of FNAC versus histopathology in diagnosis of various body lumps. Journal of the Liaquat University of Medical and Health Sciences. 2007; 6(3):103-108

349. 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

350. Tao W, Qingjun Z, Wei Z, Fang Z, Lei Z, Yuanyuan N et al. Computed tomography versus ultrasound/fine needle aspiration biopsy in differential diagnosis of thyroid nodules: a retrospective analysis. Brazilian Journal of Otorhinolaryngology. 2021; 87(4):402-409

351. Tee YY, Lowe AJ, Brand CA, Judson RT. Fine-needle aspiration may miss a third of all malignancy in palpable thyroid nodules: a comprehensive literature review. Annals of Surgery. 2007; 246(5):714-720

352. Tele JS, Kadam RS, Hulwan AB, Pawar SJ, Patil MA. Utility of bethesda system for reporting thyroid fine needle aspirates. International Journal of Research in Pharmaceutical Sciences. 2020; 11(4):2356-2360

353. Theoharis C, Adeniran AJ, Roman S, Sosa JA, Chhieng D. The impact of implementing The Bethesda System for reporting of thyroid FNA at an academic center. Diagnostic Cytopathology. 2013; 41(10):858-863

354. Theoharis CG, Schofield KM, Hammers L, Udelsman R, Chhieng DC. The Bethesda thyroid fine-needle aspiration classification system: year 1 at an academic institution. Thyroid. 2009; 19(11):1215-1223

355. Thomas JO, Adeyi OA, Nwachokor FN, Olu-Eddo AO. Fine needle aspiration cytology in the management of thyroid enlargement: Ibadan experience. East African Medical Journal. 1998; 75(11):657-659

356. Thomas JO, Amanguno AU, Adeyi OA, Adesina AO. Fine needle aspiration (FNA) in the management of palpable masses in Ibadan: impact on the cost of care. Cytopathology. 1999; 10(3):206-210

357. Thomsen H, Andreassen JC, Bangsbo C. Fine-needle aspiration biopsy of tumors of head and neck. Journal of Laryngology and Otology. 1973; 87(12):1211-1216

358. Tilak V, Dhaded AV, Jain R. Fine needle aspiration cytology of head and neck masses. Indian Journal of Pathology and Microbiology. 2002; 45(1):23-29

359. Tomimori EK, Camargo RY, Bisi H, Medeiros-Neto G. Combined ultrasonographic and cytological studies in the diagnosis of thyroid nodules. Biochimie. 1999; 81(5):447-452

360. Tsou MH, Lin HH, Ko JS. Riu's stain and the cytologic diagnosis of thyroid fine-needle aspiration: a single cancer center experience. Diagnostic Cytopathology. 1997; 16(6):543-547

361. Varhaug JE, Segadal E, Heimann P. The utility of fine needle aspiration biopsy cytology in the management of thyroid tumors. World Journal of Surgery. 1981; 5(4):573-577

362. Vojvodich SM, Ballagh RH, Cramer H, Lampe HB. Accuracy of fine needle aspiration in the pre-operative diagnosis of thyroid neoplasia. Journal of Otolaryngology. 1994; 23(5):360-365

363. Walsh JF, Sarre R. Fine needle cytology of thyroid. Australian and New Zealand Journal of Surgery. 1983; 53(4):297-300

364. Wang Z, Zhu X, Yu X, Guan H, Zhao L, Zhang Y et al. The combination of ATA classification and FNA results can improve the diagnostic efficiency of malignant thyroid nodules. Endocrine Connections. 2020; 9(9):903-911

365. Wei Y, Lu Y, li C. Clinical application of ultrasound-guided thyroid fine needle aspiration biopsy and thinprep cytology test in diagnosis of thyroid disease. Asian Pacific Journal of Cancer Prevention: APJCP. 2016; 17(10):4689-4692

366. Werga P, Wallin G, Skoog L, Hamberger B. Expanding role of fine-needle aspiration cytology in thyroid diagnosis and management. World Journal of Surgery. 2000; 24(8):907-912

367. Williams BA, Bullock MJ, Trites JR, Taylor SM, Hart RD. Rates of thyroid malignancy by FNA diagnostic category. Journal of Otolaryngology: Head and Neck Surgery. 2013; 42:61

368. Witt BL, Schmidt RL. Rapid onsite evaluation improves the adequacy of fine-needle aspiration for thyroid lesions: a systematic review and meta-analysis. Thyroid. 2013; 23(4):428-435

369. Wong LQ, Baloch ZW. Analysis of the bethesda system for reporting thyroid cytopathology and similar precursor thyroid cytopathology reporting schemes. Advances in Anatomic Pathology. 2012; 19(5):313-319

370. Wong TH, Ong CL, Tan WT, Rauff A. The solitary thyroid nodule revisited. Annals of the Academy of Medicine, Singapore. 1993; 22(4):593-597

371. Wood MD, Huang Y, Bibbo M. Improving recognition of thyroid carcinoma in rapid-consultation specimens. Acta Cytologica. 2005; 49(3):291-296

372. Wu HH, Jones JN, Osman J. Fine-needle aspiration cytology of the thyroid: ten years experience in a community teaching hospital. Diagnostic Cytopathology. 2006; 34(2):93-96

373. Wu M. A correlation study between thyroid imaging report and data systems and the Bethesda system for reporting thyroid cytology with surgical follow-up - an ultrasound-trained cytopathologist's experience. Diagnostic Cytopathology. 2021; 49(4):494-499

374. Wu M, Choi Y, Zhang Z, Si Q, Salem F, Szporn A et al. Ultrasound guided FNA of thyroid performed by cytopathologists enhances Bethesda diagnostic value. Diagnostic Cytopathology. 2016; 44(10):787-791

375. Wu Q, Qu Y, Zang X, Li Y, Yi X, Wang Y et al. Preliminary study of confounding factors of elastography and the application of fine-needle aspiration in thyroid nodules with indeterminate elastography. Scientific Reports. 2017; 7(1):18005

376. Xavier-Junior JCC, Zogheib RJP, Camilo-Junior DJ, D'Avilla S CGP, Mattar NJ. An alternative method for smear preparation of fine-needle aspiration cytology of cystic

thyroid lesions: Evaluation of sample adequacy. Diagnostic Cytopathology. 2020; 48(11):1054-1057

377. Xiong Y, Yan L, Nong L, Zheng Y, Li T. Pathological diagnosis of thyroid nodules based on core needle biopsies: comparative study between core needle biopsies and resected specimens in 578 cases. Diagnostic Pathology. 2019; 14(1):10

378. Xu D, Xu HM, Li MK, Chen LY, Wang LJ. Feasibility of fine-needle aspiration biopsy and its applications in superficial cervical lesion biopsies. International Journal of Clinical and Experimental Pathology. 2014; 7(8):5165-5170

379. Yagmur Y, Akbulut S, Sakarya H, Sogutcu N, Gumus S. Assessment of the relationship between clinical and histopathological features in cases of thyroidectomy. Annali Italiani di Chirurgia. 2018; 89:199-205

380. Yassa L, Cibas ES, Benson CB, Frates MC, Doubilet PM, Gawande AA et al. Long-term assessment of a multidisciplinary approach to thyroid nodule diagnostic evaluation. Cancer. 2007; 111(6):508-516

381. Yavuz A, Akbudak I, Ucler R, Ozgokce M, Arslan H, Batur A. Comparison of efficiencies between shear wave elastography, fine-needle aspiration biopsy and american college of radiology thyroid imaging reporting and data system scoring system in determining the malignity potential of solid thyroid nodules. Ultrasound Quarterly. 2020; 37(2):155-160

382. Yildirim E, Akbas P, Erdogan KO, Bektas S, Gumuskaya PO, Er AM et al. The comparison of the histopathological results of the thyroid fine-needle aspiration biopsies in the 795 patients with thyroidectomy. Diagnostic Cytopathology. 2021; 49(6):671-676

383. Yilmaz N, Cansu GB, Toru S, Sari R, Ocak GG, Arici C et al. Cytopathology-histopathology correlation and the effect of nodule diameter on diagnostic performance in patients undergoing thyroid fine-needle aspiration biopsy. Journal of Cancer Research and Therapeutics. 2020; 16(Suppl 1):S53-S58

384. Ylagan LR, Farkas T, Dehner LP. Fine needle aspiration of the thyroid: a cytohistologic correlation and study of discrepant cases. Thyroid. 2004; 14(1):35-41

385. Yoder BJ, Redman R, Massoll NA. Validation of a five-tier cytodiagnostic system for thyroid fine needle aspiration biopsies using cytohistologic correlation. Thyroid. 2006; 16(8):781-786

386. Yokozawa T, Miyauchi A, Kuma K, Sugawara M. Accurate and simple method of diagnosing thyroid nodules by the modified technique of ultrasound-guided fine needle aspiration biopsy. Thyroid. 1995; 5(2):141-145

387. Yoo C, Choi HJ, Im S, Jung JH, Min K, Kang CS et al. Fine needle aspiration cytology of thyroid follicular neoplasm: cytohistologic correlation and accuracy. The Korean Journal of Pathology. 2013; 47(1):61-66

388. Zaidan S, Muftah M, El Rabty A, Naji N. The influence of fine needle aspiration results on the selection of surgical procedure for nodular goitre. Jamahiriya Medical Journal. 2010; 10(1):26-29

389. Zajdela A, Joly J, Gongora R. Fine needle cytology sampling. Practical value in diagnosing thyroid diseases. Acta Oto-Rhino-Laryngologica Belgica. 1987; 41(5):686-694

390. Zbar AP, Dafydd L, Samtani J, Alleyne W, Chiappa A, Jones SR et al. Fine-needle aspiration cytology of thyroid nodules: experience at the Queen Elizabeth Hospital, Barbados (1998-2002). International Surgery. 2009; 94(1):10-19

391. Zelmanovitz F, Gross JL. Cytopathological findings from fine-needle aspiration biopsy are accurate predictors of thyroid pathology in patients with functioning thyroid nodules. Journal of Endocrinological Investigation. 1998; 21(2):98-101

392. Zhang J, Chen Z, Anil G. Ultrasound-guided thyroid nodule biopsy: outcomes and correlation with imaging features. Clinical Imaging. 2015; 39(2):200-206

393. Zhang J, Wang J. The diagnostic evaluation of fine needle aspiration cytology of thyroid and its clinical application. Chinese-German Journal of Clinical Oncology. 2012; 11(6):317-319

394. Zhong LC, Lu F, Ma F, Xu HX, Li DD, Guo LH et al. Ultrasound-guided fine-needle aspiration of thyroid nodules: does the size limit its efficiency? International Journal of Clinical and Experimental Pathology. 2015; 8(3):3155-3159

395. Zosin I, Balas M. Clinical, ultrasonographical and histopathological aspects in Hashimoto's thyroiditis associated with malignant and benign thyroid nodules. Endokrynologia Polska. 2013; 64(4):255-262

396. Zoulias EA, Asvestas PA, Matsopoulos GK, Tseleni-Balafouta S. A decision support system for assisting fine needle aspiration diagnosis of thyroid malignancy. Analytical and Quantitative Cytology and Histology. 2011; 33(4):215-222

Appendices

Appendix A – Review protocols

A.1 Review protocol for accuracy of FNAC

Field	Content
PROSPERO registration number	CRD42021244440
Review title	The diagnostic accuracy of fine-needle aspiration cytology (FNAC) with rapid on-site assessment, FNAC without rapid on-site assessment or core biopsy for diagnosing thyroid cancer, for people with nodules shown by ultrasound* to require further investigation. *'positive' on US – that is, they had US characteristics that exceeded the chosen threshold.
Review question	For people with thyroid nodules that require further investigation following ultrasound, what is the diagnostic accuracy of FNAC with rapid on-site assessment, FNAC without rapid on-site assessment or core biopsy for diagnosing thyroid cancer?
Objective	To identify the most accurate methods of detecting thyroid cancer in this population of people identified at high risk.
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 Other searches: • None The searches may be re-run 6 weeks before the 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 suspected of thyroid cancer with potentially malignant nodules on ultrasound. Exclusion: Children and young people under 16 years.
Index Tests	• Fine-needle aspiration cytology (FNAC) with rapid on-site assessment of adequacy (by cytopathologist or technician) and with smear without cytospin and cell block • Fine-needle aspiration cytology (FNAC) with rapid on-site assessment of adequacy (by cytopathologist or technician) and with smear with cytospin and cell block

	• Fine-needle aspiration cytology (FNAC) without rapid on-site assessment with smear without cyptospin and cellblock • Fine-needle aspiration cytology (FNAC) without rapid on-site assessment with Cytospin and cell block, without smear. • Fine-needle aspiration cytology (FNAC) without rapid on-site assessment with smear, cytospin and cell block • Core biopsy
Reference standard	**Post-operative histopathological findings**
Types of study to be included	Cross-sectional/prospective/retrospective diagnostic studies, or any study containing a diagnostic accuracy analysis
Other exclusion criteria	Studies that do not report sensitivity and specificity, or insufficient data to derive these values. Non-English language studies.
Context	FNAC tends to be the second line test used in people who have suspicious US findings. FNAC can be performed in several different ways and it is important that the accuracy in detection of thyroid cancer cells is known for each of these methods so that the best method can be recommended. In addition, core biopsy may be used as an alternative and so it is important that the diagnostic accuracy of this is also known.
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).

Secondary outcomes (important outcomes)	NA
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. 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	Stratification: Prior US assessment / no prior US assessment 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. Is it US guided? Y/N
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: [NICE guideline webpage].
Other registration details	N/A
Reference/URL for published protocol	https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=244440
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

A.2 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).[259] **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 strategy

This literature search strategy was used for the following reviews:

- For people with thyroid nodules that require further investigation following ultrasound, what is the diagnostic accuracy of FNAC with rapid on-site evaluation, FNAC without rapid on-site evaluation or core biopsy for diagnosing thyroid cancer?

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 28: Database parameters, filters and limits applied

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

Database	Dates searched	Search filters and limits applied
	Cochrane Central Register of Controlled Trials to Issue 12 of 12, December 2021	
Epistemonikos (The Epistemonikos Foundation)	Inception – 13 January 2022	Systematic review Exclusions (Cochrane reviews) English language

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 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/
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.	exp Biopsy, Needle/
29.	((needle or core or puncture) adj3 (aspirat* or biops* or cytology)).ti,ab.
30.	(FNAC or FNA or FNAB or FNB or FNC or CNB).ti,ab.
31.	or/28-30
32.	27 and 31
33.	randomized controlled trial.pt.

34.	controlled clinical trial.pt.
35.	randomi#ed.ab.
36.	placebo.ab.
37.	randomly.ab.
38.	clinical trials as topic.sh.
39.	trial.ti.
40.	or/33-39
41.	Meta-Analysis/
42.	Meta-Analysis as Topic/
43.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
44.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
45.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
46.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
47.	(search* adj4 literature).ab.
48.	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
49.	cochrane.jw.
50.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
51.	or/41-50
52.	32 and (40 or 51)
53.	Epidemiologic studies/
54.	Observational study/
55.	exp Cohort studies/
56.	(cohort adj (study or studies or analys* or data)).ti,ab.
57.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
58.	((longitudinal or retrospective or prospective) and (study or studies or review or analys* or cohort* or data)).ti,ab.
59.	Controlled Before-After Studies/
60.	Historically Controlled Study/
61.	Interrupted Time Series Analysis/
62.	(before adj2 after adj2 (study or studies or data)).ti,ab.
63.	exp case control study/
64.	case control*.ti,ab.
65.	Cross-sectional studies/
66.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
67.	or/53-66
68.	32 and 67
69.	68 not 52
70.	exp "sensitivity and specificity"/
71.	(sensitivity or specificity).ti,ab.
72.	((pre test or pretest or post test) adj probability).ti,ab.
73.	(predictive value* or PPV or NPV).ti,ab.
74.	likelihood ratio*.ti,ab.
75.	likelihood function/
76.	((area under adj4 curve) or AUC).ti,ab.

77.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
78.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness)).ti,ab.
79.	gold standard.ab.
80.	exp Diagnostic errors/
81.	(false positiv* or false negativ*).tw.
82.	or/70-81
83.	32 and 82
84.	83 not (52 or 69)

Embase (Ovid) search terms

1.	exp Thyroid Cancer/
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 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.	exp Needle Biopsy/
28.	((needle or core or puncture) adj3 (aspirat* or biops* or cytology)).ti,ab.
29.	(FNAC or FNA or FNAB or FNB or FNC or CNB).ti,ab.
30.	or/27-29
31.	26 and 30
32.	random*.ti,ab.
33.	factorial*.ti,ab.
34.	(crossover* or cross over*).ti,ab.
35.	((doubl* or singl*) adj blind*).ti,ab.

36.	(assign* or allocat* or volunteer* or placebo*).ti,ab.
37.	crossover procedure/
38.	single blind procedure/
39.	randomized controlled trial/
40.	double blind procedure/
41.	or/32-40
42.	systematic review/
43.	Meta-Analysis/
44.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
45.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
46.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
47.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
48.	(search* adj4 literature).ab.
49.	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
50.	cochrane.jw.
51.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
52.	or/42-51
53.	31 and (41 or 52)
54.	Clinical study/
55.	Observational study/
56.	family study/
57.	longitudinal study/
58.	retrospective study/
59.	prospective study/
60.	cohort analysis/
61.	follow-up/
62.	cohort*.ti,ab.
63.	61 and 62
64.	(cohort adj (study or studies or analys* or data)).ti,ab.
65.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
66.	((longitudinal or retrospective or prospective) and (study or studies or review or analys* or cohort* or data)).ti,ab.
67.	(before adj2 after adj2 (study or studies or data)).ti,ab.
68.	exp case control study/
69.	case control*.ti,ab.
70.	cross-sectional study/
71.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
72.	or/54-60,63-71
73.	31 and 72
74.	73 not 53
75.	exp "sensitivity and specificity"/
76.	(sensitivity or specificity).ti,ab.
77.	((pre test or pretest or post test) adj probability).ti,ab.
78.	(predictive value* or PPV or NPV).ti,ab.

79.	likelihood ratio*.ti,ab.
80.	((area under adj4 curve) or AUC).ti,ab.
81.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
82.	diagnostic accuracy/
83.	diagnostic test accuracy study/
84.	gold standard.ab.
85.	exp diagnostic error/
86.	(false positiv* or false negativ*).ti,ab.
87.	differential diagnosis/
88.	(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.
89.	or/75-88
90.	31 and 89
91.	90 not (53 or 74)

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
#6.	MeSH descriptor: [Biopsy, Needle] explode all trees
#7.	(needle or core or puncture) near/3 (aspirat* or biops* or cytology):ti,ab
#8.	(FNAC or FNA or FNAB or FNB or FNC or CNB):ti,ab
#9.	#6 or #7 or #8
#10.	#5 and #9
#11.	conference:pt or (clinicaltrials or trialsearch):so
#12.	#10 not #11

Epistemonikos search terms

1.	(title:((title:(thyroid AND (cancer* OR neoplasm* OR nodule* OR carcinoma*)) OR abstract:(thyroid AND (cancer* OR neoplasm* OR nodule* OR carcinoma*))) AND (title:(needle OR puncture OR biops* OR aspirat*) OR abstract:(needle OR puncture OR biops* OR aspirat*))) OR abstract:((title:(thyroid AND (cancer* OR neoplasm* OR nodule* OR carcinoma*)) OR abstract:(thyroid AND (cancer* OR neoplasm* OR nodule* OR carcinoma*))) AND (title:(needle OR puncture OR biops* OR aspirat*) OR abstract:(needle OR puncture OR biops* OR aspirat*))))

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
	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

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

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

NHS EED and HTA (CRD) search terms

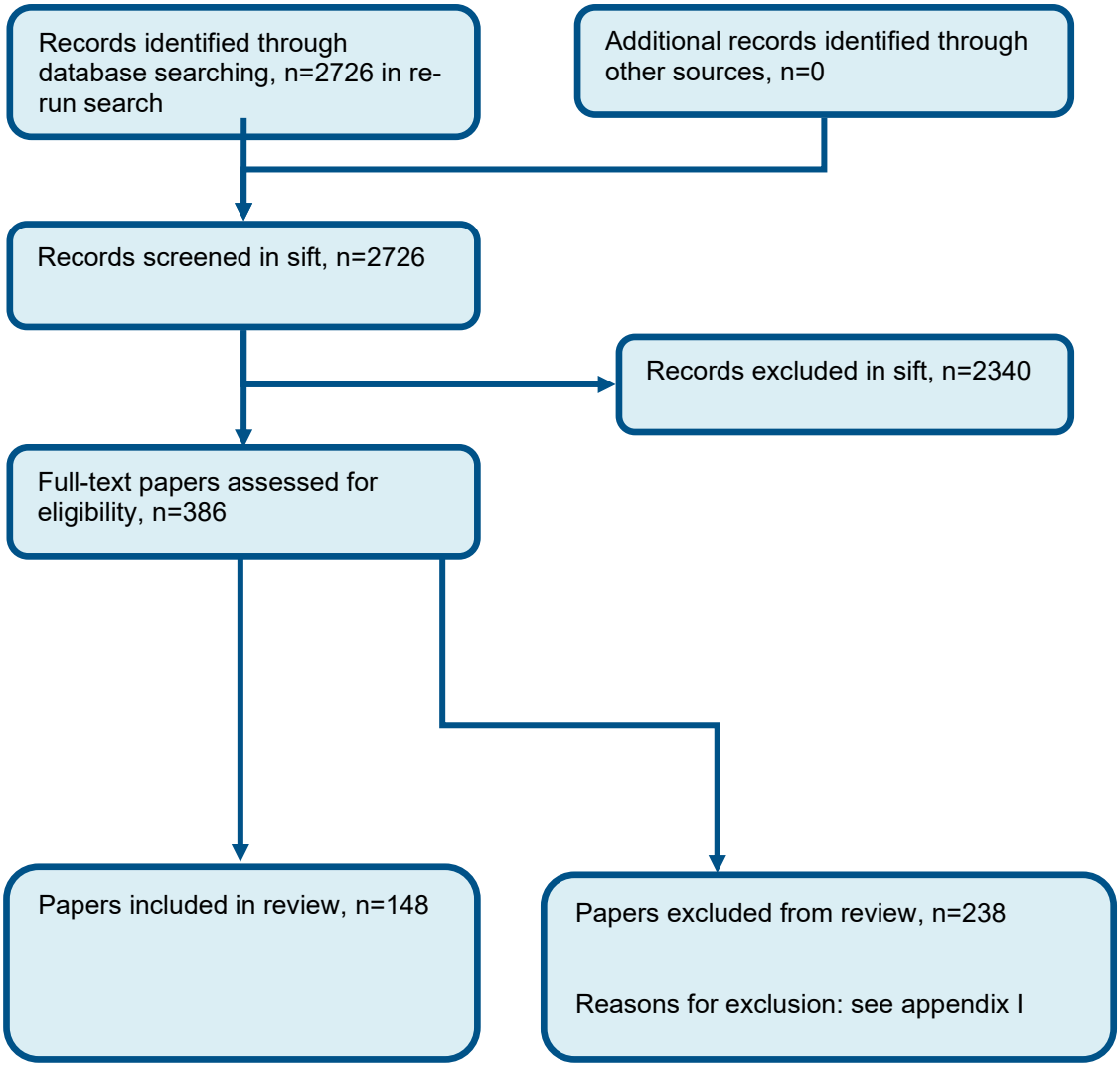
#1.	MeSH DESCRIPTOR Thyroid Neoplasms EXPLODE ALL TREES
#2.	((thyroid NEAR4 (cancer* or carcinom* or tumour* or tumor* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or papillar* or follicul* or lymphoma* or anaplastic)))
#3.	(((papillar* or follicul* or medullary or anaplastic)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

INHATA search terms

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

Appendix C – Diagnostic evidence study selection

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



Appendix D – Diagnostic accuracy evidence

NOTE: All data are calculated using adjusted approach – that is, any truly malignant unsatisfactory cytology taken as false negatives and any truly benign unsatisfactory cytology taken as false positives.

Reference	Agcaoglu, 2013[6]
Study type	Retrospective
Number of patients	n = 730 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Setting*: General Surgery Clinic *Country*: Turkey *Inclusion criteria*: Prior US, otherwise not reported *Exclusion criteria*: Non-diagnostic results *Stratum (prior US assessment / no prior US assessment): prior US* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology with ROSE, with smear only (cytopathologist attended in 77% of FNAB procedures) *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No

Reference	Agcaoglu, 2013[6]
	Blinding of gold standard test: No
Results	*Malignant nodules=320; benign nodules = 410* No data given for inadequate samples FNA grading: benign, indeterminate, malignant *FNAC rated indeterminate or malignant (+ve) [benign taken as -ve result]* TP: 239 FN: 81 FP: 45 TN: 365 ; *sensitivity*:0.747 , *specificity*: 0.890
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, so some bias possible in who was given surgery)*
Comments	

Reference	Anderson, 1987[25]
Study type	Retrospective
Number of patients	n = 373 nodules in 373 patients (solitary or dominant nodules only) – this was the sub-group with surgical histopathology eligible for this review
Patient characteristics	*Age, mean (SD): not reported for the sub-group with histopathological gold standard* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Setting*: Department of Surgery *Country*: UK *Inclusion criteria*: solitary nodule within the thyroid or a dominant nodule in a non-toxic goitre; submitted to partial or total thyroidectomy *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): unclear (some underwent US but unclear how many)*

Reference	Anderson, 1987[25]
	Sub-group (US-guided / not US guided): Non-USG
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings and autopsy in 4 cases *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* Yes
Results	*Malignant nodules=63; benign nodules = 310* No data given for inadequate samples *FNA grading: benign, suspicious, definitely malignant* *FNAC rated suspicious or definitely malignant (+ve) [benign taken as -ve result]* TP: 59 FN: 4 FP: 2 TN: 308 ; *sensitivity*: 0.937, *specificity*: 0.994
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, so some bias possible in who was given surgery)*
Comments	

Reference	Arul, 2015[29]
Study type	Retrospective
Number of patients	n = 392 nodules
Patient characteristics	*Age, mean (SD): Not reported* *Gender (female to male ratio): Not reported* *Ethnicity*: not reported *Setting*: University Hospital *Country*: India *Inclusion criteria*: all the FNACs of thyroid lesions between July 2012 and January 2015 were retrieved retrospectively; surgical histopathology obtained; FNAC classified according to 6 tier TBSRTC *Exclusion criteria*: No histopathology *Stratum (prior US assessment / no prior US assessment): unclear* *Sub-group (US-guided / not US guided): unclear*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No

Reference	Arul, 2015[29]
	Blinding of gold standard test: No
Results	*Malignant nodules=59; benign nodules = 333* *FNAC classification: Bethesda I-VI* Inadequate category: 0 malignant, 10 benign *FNAC 6 Tier Bethesda: atypia of undetermined significance/follicular lesions and above (+ve)* TP: 56 FN: 3 FP: 80 TN: 253 ; *sensitivity*:0.949, *specificity*: 0.760 *FNAC 6 Tier Bethesda: follicular neoplasms /suspicious for follicular neoplasms and above (+ve)* TP: 46 FN: 13 FP: 49 TN: 284 ; *sensitivity*: 0.779, *specificity*: 0.853 *FNAC 6 Tier Bethesda: suspicious for malignancy and above (+ve)* TP: 33 FN: 26 FP: 17 TN: 316 ; *sensitivity*: 0.559, *specificity*: 0.948 *FNAC 6 Tier Bethesda: malignant (+ve)* TP: 16 FN: 43 FP: 10 TN: 323 ; *sensitivity*: 0.271, *specificity*: 0.969
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, so some bias possible in who was given surgery)*
Comments	

Reference	Can, 2008[61]
Study type	retrospective
Number of patients	n = 23 nodules sent for surgery (USG) and 18 nodules sent for surgery (non-USG)

Reference	Can, 2008[61]
Patient characteristics	*Age, mean (SD): not available for those that had surgery* *Gender (female to male ratio): not available for those that had surgery* *Ethnicity*: not reported *Setting*: Outpatient endocrinology clinic *Country*: Turkey *Inclusion criteria*: All consecutive patients who underwent FNA of thyroid nodules, followed by surgery. *Exclusion criteria*: No surgery performed (note that this is an exclusion criterion for the data included here but was not an exclusion criterion for the study that also looked at data from patients who did not have surgery) *Stratum (prior US assessment / no prior US assessment): unclear* *Sub-group (US-guided / not US guided): USG for 23 and non-USG for 18*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Can, 2008[61]
Results	<u>*USG*</u> *FNA grading: benign, indeterminate (a pattern of follicular or Hurthle cell neoplasm or aspects of atypia suggestive, but not conclusive of the presence of a malignant neoplasm), malignant* <u>Inadequate category: 0 malignant, 1 benign</u> *FNAC rated indeterminate or malignant (+ve) [benign taken as -ve result]* TP: 8 FN: 0 FP: 4 TN: 11 ; *sensitivity*: 1.0, *specificity*: 0.733 <u>*Non-USG*</u> <u>Inadequate category: 0 malignant, 3 benign</u> *FNA grading: benign, indeterminate (a pattern of follicular or Hurthle cell neoplasm or aspects of atypia suggestive, but not conclusive of the presence of a malignant neoplasm), malignant* *FNAC rated indeterminate or malignant (+ve) [benign taken as -ve result]* TP: 2 FN: 0 FP: 4 TN: 12 ; *sensitivity*: 1.0, *specificity*: 0.75
Source of funding	<u>No funding stated</u>
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Chang, 1997[67]
Study type	Retrospective
Number of patients	n = 662 nodules from 662 patients
Patient characteristics	*Age, mean (SD): Not reported* *Gender (female to male ratio): Not reported*

Reference	Chang, 1997[67]
	Ethnicity: not reported *Setting*: Internal medicine Department *Country*: China *Inclusion criteria*: Patients undergoing FNA and surgery for thyroid nodules. Surgery indicated for those with a malignant or indeterminate result. Those with a benign result only underwent surgery in cases of a rapidly growing nodule, local compression or cosmetic reasons. *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): unclear* *Sub-group (US-guided / not US guided): not reported as USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Chang, 1997[67]
Results	Malignant=162; benign=500 Inadequate category: 6 malignant, 38 benign *FNAC rated indeterminate or malignant (+ve) [benign taken as -ve result]* TP: 139 FN: 23 FP: 161 TN: 339 ; *sensitivity*: 0.858, *specificity*: 0.678 *FNAC rated malignant (+ve) [indeterminate or benign taken as -ve result]* TP: 105 FN: 57 FP: 47 TN: 453 ; *sensitivity*: 0.648, *specificity*: 0.906
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, so some bias possible in who was given surgery)*
Comments	

Reference	Francis, 1999[115]
Study type	Retrospective
Number of patients	n = 45 patients
Patient characteristics	*Age, median (range): 37 (19-63)* *Gender (female to male ratio): 41:4* *Ethnicity*: not reported *Setting*: Cytology and Histopathology Units *Country*: Kuwait *Inclusion criteria*: Patients attending thyroid unit for FNA *Exclusion criteria*: Not meeting criteria for FNA; aspirated cervical lymph nodes *Stratum (prior US assessment / no prior US assessment): unclear*

Reference	Francis, 1999[115]
	Sub-group (US-guided / not US guided): not stated to be USG
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	*Malignant=20; benign=25* Fine needle aspiration cytology without ROSE, with smear only Inadequate category: 1 malignant, 3 benign *FNAC rated carcinoma or NHL or neoplasm or hyperplastic nodules (+ve) [benign taken as goitre, benign]* TP: 17 FN: 3 FP: 12 TN: 13 ; *sensitivity*: 0.85, *specificity*: 0.52 *FNAC rated carcinoma or NHL or hyperplastic nodules (+ve) [benign taken as neoplasm, goitre, benign]* TP: 14 FN: 6 FP: 3 TN: 22 ; *sensitivity*: 0.70, *specificity*: 0.88
Source of funding	No funding stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias*

Reference	Francis, 1999[115]
	Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)
Comments	

Reference	Haberal, 2009[144]
Study type	Retrospective - consecutive
Number of patients	n = 260 nodules in 260 patients
Patient characteristics	*Age, median (range): 46 (12-85)* *Gender (female to male ratio): 218:42* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Turkey *Inclusion criteria*: Adequate FNAC followed by thyroidectomy or lobectomy for a dominant thyroid nodule *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): unclear if prior US* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:*

Reference	Haberal, 2009[144]
	Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	*Malignant: 63; Benign: 197* <u>Inadequate category: not reported</u> *FNAC rated Hurtle Cell neoplasm/Follicular neoplasm, suspicious for neoplasm or malignant (+ve) [negative taken as negative]* TP: 59 FN: 4 FP: 31 TN: 166 ; *sensitivity*: 0.937, *specificity*: 0.843 *FNAC rated suspicious for neoplasm or malignant (+ve) [negative and Hurtle Cell neoplasm/Follicular neoplasm, taken as non-neoplasm]* TP: 53 FN: 10 FP: 18 TN: 179 ; *sensitivity*: 0.841, *specificity*: 0.909 *FNAC rated malignant only (+ve) [benign taken as Hurtle Cell neoplasm, Follicular neoplasm, suspicious for neoplasm or non-neoplasm]* TP: 41 FN: 22 FP: 1 TN: 196 ; *sensitivity*: 0.651, *specificity*: 0.995 These results are based on data in table in study and do not agree with reported sensitivity and specificity figures.
Source of funding	<u>No funding stated</u>
Limitations	*Risk of bias (QUADAS 2 – risk of bias): Very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Hamming, 1998[150]
Study type	Retrospective
Number of patients	n = 240 nodules
Patient characteristics	*Age, mean (range): 58 (14-81)*

Reference	Hamming, 1998[150]
	Gender (female to male ratio): 179:61 *Ethnicity*: not reported *Setting*: University Hospital *Country*: Holland *Inclusion criteria*: Patients operated on for nodular thyroid disease with an evaluable FNAC *Exclusion criteria*: non-evaluable smears – insufficient material for cytodiagnosis. *Stratum (prior US assessment / no prior US assessment): unclear if prior US* *Sub-group (US-guided / not US guided): not clear if USG used*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	*Malignant=72; benign=168* Inadequate category: not reported *FNAC rated indeterminate or malignant (+ve) [benign taken as -ve result]*

Reference	Hamming, 1998[150]
	TP: 67 FN: 5 FP: 69 TN: 99 ; *sensitivity*: 0.931, *specificity*: 0.589 *FNAC rated malignant (+ve) [benign or indeterminate taken as -ve result]* TP: 49 FN: 23 FP: 2 TN: 166 ; *sensitivity*: 0.6805, *specificity*: 0.988
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, so some bias possible in who was given surgery)*
Comments	

Reference	Hawkins, 1987[153]
Study type	Retrospective
Number of patients	n = 415 nodules
Patient characteristics	*Age, mean (SD): not provided for subset with surgery data* *Gender (female to male ratio): not available* *Ethnicity*: not reported *Setting*: Outpatient endocrinology unit *Country*: Spain *Inclusion criteria*: Patients referred to endocrinology unit because of diffuse or nodular goitres, with or without symptoms; surgery (in patients with positive or suspicious FNAB cytology and/or suggestive clinical histories, and in patients with cold thyroid nodules and negative FNAB results that did not respond to 6 months of suppressive thyroxine therapy *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): unclear if prior US* *Sub-group (US-guided / not US guided): unclear if USG*

Reference	Hawkins, 1987[153]
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block. Unclear in description but stated that 'if fluid was drawn the centrifuged sediment was studied', indicating that at least cytospin was used in addition to smear. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	*Malignant=73; benign=342* Inadequate category: not reported *FNAC rated 'positive' for carcinoma or suspicious follicular proliferative lesions (+ve) ['negative'(including non-malignant follicular proliferative lesions) taken as -ve result]* TP: 63 FN: 10 FP: 16 TN: 326 ; *sensitivity*:0.863, *specificity*: 0.953 *FNAC rated positive for carcinoma (+ve) ['negative' (including non-malignant follicular proliferative lesions) or suspicious follicular proliferative lesions taken as -ve result]* TP: 48 FN: 25 FP: 3 TN: 339 ; *sensitivity*: 0.658, *specificity*: 0.991
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, so some bias possible in who was given surgery)*
Comments	

Reference	Jat, 2019[167]
Study type	Prospective
Number of patients	n = 75 nodules
Patient characteristics	*Age, mean (SD): Not provided for surgical sub-set* *Gender (female to male ratio): Not provided for surgical sub-set* *Ethnicity*: not reported *Setting*: Outpatient department of surgery *Country*: Kingdom of Saudi Arabia *Inclusion criteria*: all patients came in OPD with clinically diagnosed as a solitary thyroid nodule having no hyper or hypothyroidism, irrespective of age and sex; thyroid surgery *Exclusion criteria*: patients presenting with extra-thyroid neck swelling; patients having toxic or non- toxic diffuse or multinodular goitre *Stratum (prior US assessment / no prior US assessment): prior US performed but not stated that the sample were selected through that* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology with ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	Jat, 2019[167]
	Blinding of index test: No *Blinding of gold standard test:* No
Results	*Malignant= 32; benign=43* Inadequate category: 10 inadequate results but no histopathology results available *FNA gradings: non-diagnostic, goitre, thyroiditis, follicular neoplasm/Hurthle cell neoplasm, malignancy* *FNAC rated follicular neoplasm/Hurthle cell neoplasm, malignancy (+ve) [goitre, thyroiditis taken as -ve result]* TP: 6 FN: 4 FP: 24 TN: 41 ; *sensitivity*: 0.60, *specificity*: 0.631 *FNAC rated malignancy (+ve) [follicular neoplasm/Hurthle cell neoplasm, goitre, thyroiditis taken as -ve result]* TP: 4 FN: 6 FP: 2 TN: 63 ; *sensitivity*: 0.40, *specificity*: 0.969
Source of funding	No funding stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): none*
Comments	

Reference	Kothari, 2019 #1269 [196]
Study type	Prospective
Number of patients	n = 53 nodules
Patient characteristics	*Age, mean (SD): 39 (not reported)* *Gender (female to male ratio): 3.8:1* *Ethnicity*: not reported *Setting*: Department of cytopathology *Country*: India

Reference	Kothari, 2019 #1269 [196]
	Inclusion criteria: Not reported; FNA with follow up histopathology *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): unclear if prior US* *Sub-group (US-guided / not US guided): not clear if USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology with ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	*Malignant= 3; benign=50 (somewhat unclear in paper)* Inadequate category: not reported *FNAC rated Bethesda VI (+ve) [benign taken as Bethesda II, III, IV result]* TP: 2 FN: 1 FP: 0 TN: 50 ; *sensitivity*: 0.667, *specificity*: 1.00
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, so some bias possible in who was given surgery)*
Comments	

Reference	La ROSE, 1991[200]
Study type	Retrospective
Number of patients	n = 827 nodules
Patient characteristics	*Age, mean (SD): Not reported* *Gender (female to male ratio): Not reported* *Ethnicity*: not reported *Setting*: Surgical/Endocrinology *Country*: Italy *Inclusion criteria*: Cold thyroid nodules examined with FNA that were given subsequent surgery. Surgery was offered to those to those that were malignant or highly suspicious on FNA; probable adenoma were suggested to undergo surgery. 'Benign' or 'inadequate' nodules were also given surgery if there was clinical suspicion or through patient choice. [Thus although there was some bias in the access to surgery, there was definite access from all FNA categories, allowing a reasonably valid assessment of accuracy to be made]. *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): prior US not reported* *Sub-group (US-guided / not US guided): No evidence of USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:*

Reference	La ROSE, 1991[200]
	Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	*Malignant=250; benign = 577* Inadequate category: 3 malignant, 19 benign *Used following scale; malignant, follicular lesion type I (suggestive of follicular carcinoma), follicular type II (probably malignant), follicular type III (suggestive of benign lesion), benign and inadequate.* *FNAC rated malignant, follicular lesion type I (suggestive of follicular carcinoma), follicular type II (probably malignant), follicular type III (suggestive of benign lesion) (+ve) [benign taken as -ve result]* TP: 241 FN: 9 FP: 320 TN: 257 ; *sensitivity*: 0.964, *specificity*: 0.445 *FNAC rated malignant, follicular lesion type I (suggestive of follicular carcinoma), follicular type II (probably malignant) (+ve) [benign and type III follicular lesions taken as -ve result]* TP: 215 FN: 35 FP: 87 TN: 490 ; *sensitivity*: 0.860, *specificity*: 0.849 *FNAC rated malignant, follicular lesion type I (suggestive of follicular carcinoma), (+ve) [benign and type III & II follicular lesions taken as -ve result]* TP: 200 FN: 50 FP: 25 TN: 552 ; *sensitivity*:0.800, *specificity*: 0.957 *FNAC rated type malignant (+ve) [benign and type III & II & I follicular lesions taken as -ve result]* TP: 179 FN: 79 FP: 23 TN: 554 ; *sensitivity*: 0.694, *specificity*: 0.960
Source of funding	No funding stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): none*
Comments	

Reference	Leenhardt, 1999[204]
Study type	Retrospective - consecutive
Number of patients	n = 94 nodules undergoing surgery
Patient characteristics	*Age, mean (SD): Not reported for those undergoing surgery* *Gender (female to male ratio): not reported for those undergoing surgery* *Ethnicity*: not reported *Setting*: Surgery/Endocrinology Unit *Country*: France *Inclusion criteria*: Consecutive patients with thyroid nodules referred for FNA after US; non palpable nodules. Surgery provided for a histopathological diagnosis. Surgery was offered to those to those that were malignant or suspicious on FNA; supracentrimetric or isolated cold nodules; simultaneous presence of a palpable nodule in a multinodular gland and miscellaneous reasons. [Thus, although there was some bias in the access to surgery, there was definite access from all FNA categories, allowing a reasonably valid assessment of accuracy to be made]. *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): prior US* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only. If repeated FNA, only the result of the last used in this analysis *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	Leenhardt, 1999[204]
	Blinding of index test: No *Blinding of gold standard test:* No
Results	*Malignant: 20; benign: 74* Inadequate category: 3 malignant, 9 benign *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 16 FN: 4 FP: 33 TN: 41 ; *sensitivity*: 0.8, *specificity*: 0.554 *FNAC rated malignant (+ve) [suspicious, benign taken as -ve result]* TP: 9 FN: 11 FP: 16 TN: 58 ; *sensitivity*: 0.45, *specificity*: 0.784
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, so some bias possible in who was given surgery)*
Comments	

Reference	Li, 2013[206]
Study type	Prospective
Number of patients	n = 51 nodules in 48 patients
Patient characteristics	*Age, mean (SD): 47.2(5.7)* *Gender (female to male ratio): 35:13* *Ethnicity*: not reported *Setting*: University Hospital *Country*: China *Inclusion criteria*: Patients with suspected solid thyroid nodules, later given US guided biopsy and a histopathological confirmation after, presumably, surgery.

Reference	Li, 2013[206]
	Exclusion criteria: Patients hyper-susceptible to SonoVue or with coagulation dysfunction were excluded *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Core biopsy with US guidance Core biopsy with CEUS guidance *Reference (gold) standard:* Surgical histopathological findings (though unclear) *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	*Gold standard results: PTC detected at puncture points: 240; No PTC detected at puncture points 70 [note unit of analysis is biopsy puncture points not nodules]* Inadequate category: 0 malignant, 0 benign *Biopsy with US guidance rated positive (+ve) [negative taken as -ve result]* TP: 116 FN: 124 FP: 11 TN: 59 ; *sensitivity*:0.483, *specificity*:0.843 *Biopsy with CEUS guidance rated positive (+ve) [negative taken as -ve result]* TP: 199 FN: 41 FP: 13 TN: 57 ; *sensitivity*:0.829, *specificity*:0.814
Source of funding	No funding stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): none*

Reference	Li, 2013[206]
Comments	

Reference	Lukitto, 1998[217]
Study type	Retrospective
Number of patients	n = 167 nodules in 167 patients
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Setting*: Division of surgical oncology *Country*: Indonesia *Inclusion criteria*: Patients with thyroid nodules undergoing FNAC and surgery. Indications for surgery not provided. Out of 250, 167 went for thyroidectomy, and 162 of these were 'negative' on FNA, so it seems that the decision was not based on FNAC. Therefore this study has been included. *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): not reported to be prior US* *Sub-group (US-guided / not US guided): Not reported to be USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings

Reference	Lukitto, 1998[217]
	Time between measurement of index test and reference standard: Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	*Malignant=16; benign=151* Inadequate category: not reported *FNAC rated positive (+ve) [negative taken as -ve result]* TP: 4 FN: 12 FP: 1 TN: 150 ; *sensitivity*: 0.25, *specificity*: 0.993
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, so some bias possible in who was given surgery)*
Comments	

Reference	Mijovic, 2009[240]
Study type	Retrospective - consecutive
Number of patients	n = 115 nodules from 115 patients
Patient characteristics	*Age, median (range): 51 (23-83)* *Gender (female to male ratio): 90:25* *Ethnicity*: not reported *Setting*: Teaching Hospital *Country*: Canada

Reference	Mijovic, 2009[240]
	Inclusion criteria: Consecutive patients undergoing thyroidectomy for cytologically proven malignancy or nodules suspicious for being malignant (e.g. history of radiation exposure, family history, size and so on); surgery also performed on patients with Graves disease, large goitres and compression symptoms with FNA performed pre-op. *Exclusion criteria*: *Stratum (prior US assessment / no prior US assessment): prior US not reported* *Sub-group (US-guided / not US guided): NO USG USED*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only AND some (unspecified number) were: Fine needle aspiration cytology without ROSE, with smear + cell block. The paper stated that: 'all cases had at least a smear stained with Papanicolaou, and, if enough material was available, a smear stained with Diff quick and a cell block was performed' *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	*Malignant: 73; benign 42* Inadequate category: 4 malignant, 5 benign *FNAC rated positive/suspicion of malignancy or indeterminate (+ve) [benign taken as -ve result]* TP: 63 FN: 10 FP: 28 TN:14 ; *sensitivity*: 0.863, *specificity*: 0.333

Reference	Mijovic, 2009[240]
	FNAC rated positive/suspicion of malignancy (+ve) [benign or indeterminate taken as -ve result] TP: 39 FN: 34 FP: 6 TN:36 ; *sensitivity*: 0.534, *specificity*: 0.857
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, so some bias possible in who was given surgery)*
Comments	

Reference	Nart, 2010 #1327[257]
Study type	Retrospective
Number of patients	n = 291 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Turkey *Inclusion criteria*: Patients with FNA followed up with surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test*

Reference	Nart, 2010 #1327[257]
	Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	*Malignant= 114; benign=177* Inadequate category: 9 malignant, 13 benign *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 45 FN: 69 FP: 24 TN: 153 ; *sensitivity*: 0.395, *specificity*: 0.864 *FNAC rated malignant (+ve) [benign or suspicious taken as -ve result]* TP: 25 FN: 89 FP: 13 TN: 164 ; *sensitivity*: 0.219, *specificity*: 0.927
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, so some bias possible in who was given surgery)*
Comments	

Reference	Naz, 2014[260]
Study type	Retrospective
Number of patients	n = 61 nodules
Patient characteristics	*Age, mean (SD): not reported for those sent to surgery* *Gender (female to male ratio): not reported for those sent to surgery*

Reference	Naz, 2014[260]
	Ethnicity: not reported *Setting*: Histopathology Department *Country*: Pakistan *Inclusion criteria*: Patients presenting with thyroid swelling, undergoing FNA. For this review only those sent for surgery were included, but no rationale for surgery given; however it appears that those sent for surgery represented all gradings of the FNAC. *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): No report of prior US* *Sub-group (US-guided / not US guided): Not reported to be USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear + cell block. Repeat aspiration performed for inadequate smears *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	*Gold standard results: Malignant=14; benign=47* Inadequate category: unclear *FNAC rated Bethesda 3 or above (+ve) [benign taken as Bethesda 2]*

Reference	Naz, 2014[260]
	TP: 9 FN: 5 FP: 7 TN: 40 ; *sensitivity*: 0.643, *specificity*: 0.851 *FNAC rated Bethesda 4 or above (+ve) [benign taken as Bethesda 2 or 3]* TP: 7 FN: 7 FP: 3 TN: 44 ; *sensitivity*: 0.50, *specificity*: 0.936 *FNAC rated Bethesda 5 or above (+ve) [benign taken as Bethesda 2 -4]* TP: 6 FN: 8 FP: 0 TN: 47 ; *sensitivity*: 0.429, *specificity*: 1.0
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, so some bias possible in who was given surgery)*
Comments	

Reference	Okumura, 1999 #1334[266]
Study type	Prospective
Number of patients	n = 109 nodules from 107 patients
Patient characteristics	*Age, mean (SD): 54.8(15.5)* *Gender (female to male ratio): 89: 18* *Ethnicity*: not reported *Setting*: Teaching hospital *Country*: Japan *Inclusion criteria*: Patients with thyroid nodules that were given FNA and surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): No prior US reported* *Sub-group (US-guided / not US guided): USG not reported*

Reference	Okumura, 1999 #1334[266]
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	*Gold standard results: Malignancy=50; benign=59* Inadequate category: unclear *FNAC scale: Class I= normal; class II abnormal; class III possible malignant; class IV probably malignant; class V definitely malignant.* *FNAC rated class II or above (+ve) [Class I taken as -ve result]* TP: 46 FN: 4 FP: 49 TN: 10 ; *sensitivity*: 0.92, *specificity*: 0.169 *FNAC rated class III or above (+ve) [class I or II taken as -ve result]* TP: 25 FN: 25 FP: 9 TN: 50 ; *sensitivity*: 0.50, *specificity*: 0.847 *FNAC rated class IV or above (+ve) [class I or II or III taken as -ve result]* TP: 18 FN: 32 FP: 2 TN: 57 ; *sensitivity*: 0.36, *specificity*: 0.966 *FNAC rated class V or above (+ve) [class I or II or III or IV taken as -ve result]* TP: 10 FN: 40 FP: 0 TN: 59 ; *sensitivity*: 0.20, *specificity*: 1.00

Reference	Okumura, 1999 #1334[266]
Source of funding	No funding stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): none*
Comments	

Reference	Prinz, 1983[282]
Study type	Retrospective, but unclear
Number of patients	n = 109 patients with 109 nodules
Patient characteristics	*Age, mean (SD):* *Gender (female to male ratio):* *Ethnicity*: not reported *Setting*: University hospital *Country*: USA *Inclusion criteria*: Patients with palpable nodules hypo-functioning on thyroid scintiscan; subsequent thyroidectomy *Exclusion criteria*: None reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only At least 6 groups of epithelial cells required for adequate cytological evaluation, unless there was obvious atypical changes in the existing cells. *Reference (gold) standard:* Surgical histopathological findings

Reference	Prinz, 1983[282]
	Time between measurement of index test and reference standard: Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=20 ;benign=89 <u>Inadequate category: 2 malignant, 29 benign</u> *FNAC rated carcinoma or lymphoma or follicular or hurtle cell neoplasm (+ve) [benign nodular goitre, thyroiditis taken as -ve result]* TP: 17 FN: 3 FP: 51 TN: 38 ; *sensitivity*: 0.85, *specificity*: 0.427 *FNAC rated carcinoma or lymphoma (+ve) [benign nodular goitre, thyroiditis, follicular or hurtle cell neoplasm taken as -ve result]* TP: 10 FN: 10 FP: 31 TN: 58 ; *sensitivity*: 0.50, *specificity*: 0.652
Source of funding	<u>No funding stated</u>
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Roy, 2019[299]
Study type	Prospective
Number of patients	n = 112 nodules in 112 patients
Patient characteristics	*Age, mean (SD): Not reported* *Gender (female to male ratio): 89-23* *Ethnicity*: not reported *Setting*: ENT department *Country*: India

Reference	Roy, 2019[299]
	Inclusion criteria: Patients over 15 years; euthyroid state on blood examination; presenting with clinical evidence of thyroid disease and swelling *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): prior US, but not stated that a certain level was a criterion for inclusion* *Sub-group (US-guided / not US guided): No USG reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 27; benign= 85 Inadequate category: unclear *FNAC rated papillary carcinoma, anaplastic carcinoma, follicular neoplasm, medullary carcinoma (positive) (+ve) [colloid/nodular goitre, adenomatoid goitre, Hashimoto's thyroiditis, and benign cystic lesion taken as -ve result]* TP: 22 FN: 5 FP: 4 TN: 81 ; *sensitivity*: 0.815, *specificity*: 0.953
Source of funding	No funding stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): none*
Comments	

Reference	Sclabas, 2003[311]
Study type	Retrospective - consecutive
Number of patients	n = 240 nodules in 240 patients
Patient characteristics	*Age, median (range): 46 (5-96)* *Gender (female to male ratio): 180:60* *Ethnicity*: not reported *Setting*: Department of surgical oncology *Country*: USA *Inclusion criteria*: Patients undergoing FNA with or without US guidance; thyroidectomy *Exclusion criteria*: not reported *Stratum (prior US assessment / no prior US assessment): prior US for majority* *Sub-group (US-guided / not US guided): USG for some (not majority)*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology WITH ROSE?, with smear + cytospin and cell block *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Sclabas, 2003[311]
Results	Gold standard results: malignant= 103 ;benign= 137 Inadequate category: 1 malignant, 10 benign *FNAC rated indeterminate follicular, indeterminate Hurtle, Suspicious for malignancy, or positive (+ve) [negative taken as -ve result]* TP: 100 FN: 3 FP: 86 TN: 51 ; *sensitivity*: 0.971, *specificity*: 0.372 *FNAC rated Suspicious for malignancy, or indeterminate follicular, or positive (+ve) [negative or indeterminate Hurtle, taken as -ve result]* TP: 98 FN: 5 FP: 78 TN: 59 ; *sensitivity*: 0.951, *specificity*:0.431 *FNAC rated Suspicious for malignancy, or positive (+ve) [negative or indeterminate follicular or indeterminate Hurtle, taken as -ve result]* TP: 87 FN: 16 FP: 16 TN: 121 ; *sensitivity*: 0.845, *specificity*: 0.883 *FNAC rated positive (+ve) [suspicious or negative or indeterminate follicular or indeterminate Hurtle, taken as -ve result]* TP: 73 FN: 30 FP: 13 TN: 124 ; *sensitivity* 0.709, *specificity*: 0.905
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, so some bias possible in who was given surgery)*
Comments	

Reference	Seya, 1990[317]
Study type	Retrospective
Number of patients	n = 26 nodules in 26 patients
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Japan

Reference	Seya, 1990[317]
	Inclusion criteria: Patients with thyroid nodule examined using FNA and given surgery. 64 did not receive surgery but reasons not given =- however out of those going to surgery half were benign on FNA so it does not seem that FNA result was the only criterion for surgery. *Exclusion criteria*: *Stratum (prior US assessment / no prior US assessment): prior US but this did not determine who had FNA* *Sub-group (US-guided / not US guided): No USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=13 ;benign=13 Inadequate category: not reported *FNAC rated malignant (+ve) [benign taken as -ve result]* TP: 11 FN: 2 FP: 0 TN: 13 ; *sensitivity*: 0.846, *specificity*: 1.0
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, so some bias possible in who was given surgery)*
Comments	

Reference	Slowinska-Klencka, 2008[330]
Study type	Retrospective
Number of patients	n = 1694 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): 1525:169* *Ethnicity*: not reported *Setting*: Clinical Endocrinology *Country*: Poland *Inclusion criteria*: Patients referred from outpatients clinics for US and then FNAB and thyroidectomy *Exclusion criteria*: Not stated *Stratum (prior US assessment / no prior US assessment): prior US* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* 1 year maximum *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Slowinska-Klencka, 2008[330]
Results	Gold standard results: malignant= 120 ;benign=1574 Inadequate category: 1 malignant, 37 benign *FNAC rated malignant or suspected follicular neoplasm/tumour or suspected oxyphilic neoplasm/tumour or unclassified suspected lesion (+ve) [benign taken as -ve result]* TP: 86 FN: 34 FP: 245 TN: 1329 ; *sensitivity*: 0.717, *specificity*: 0.844
Source of funding	Medical University of Lodz
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Son, 2014[332]
Study type	Retrospective
Number of patients	n = 694 nodules from 469 patients
Patient characteristics	*Age, mean (SD): skilled group 53.3(11.9); non-skilled group 51.6(12.6)* *Gender (female to male ratio): skilled 112:18; non-skilled 289:50* *Ethnicity*: not reported *Setting*: University Hospital *Country*: South Korea *Inclusion criteria*: Patients undergoing total or hemithyroidectomy and also FNA *Exclusion criteria*: Patients undergoing FNA in another hospital *Stratum (prior US assessment / no prior US assessment): prior US but not used to determine whether FNA was given* *Sub-group (US-guided / not US guided): USG*

Reference	Son, 2014[332]
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 450 ;benign=244 Inadequate category: 7 malignant, 23 benign *FNAC rated positive for malignancy and suspicious for malignancy and follicular neoplasm and AUS (+ve) [benign taken as -ve result]* TP: 414 FN: 36 FP: 57 TN: 187 ; *sensitivity*: 0.920, *specificity*: 0.766 *FNAC rated positive for malignancy and suspicious for malignancy and AUS (+ve) [benign or follicular neoplasm taken as -ve result]* TP: 409 FN: 41 FP: 53 TN: 191 ; *sensitivity*: 0.909, *specificity*: 0.783 *FNAC rated positive for malignancy and suspicious for malignancy (+ve) [benign or follicular neoplasm or AUS taken as -ve result]* TP: 348 FN: 102 FP: 31 TN: 213 ; *sensitivity*: 0.773, *specificity*: 0.873
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, so some bias possible in who was given surgery)*
Comments	

Reference	Sukumaran, 2014[339]
Study type	Retrospective

Reference	Sukumaran, 2014[339]
Number of patients	n = 248 nodules
Patient characteristics	*Age, range: 11-79* *Gender (female to male ratio): 179:69* *Ethnicity*: not reported *Setting*: Regional cancer centre *Country*: India *Inclusion criteria*: Series of cases of thyroid nodules with underwent FNA followed by surgery *Exclusion criteria*: Those not given surgery [although the majority having surgery were malignant or suspicious on FNA there were a sufficient number that were benign to ensure that category was represented] *Stratum (prior US assessment / no prior US assessment): prior US performed but no evidence that this influenced decision to go for FNA* *Sub-group (US-guided / not US guided): USG done only in some (non-majority)*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Sukumaran, 2014[339]
Results	Gold standard results: malignant= 198 ;benign= 50 Inadequate category: 1 malignant, 14 benign *FNAC rated FN/SFN or FLUS or suspicious or malignant (+ve) [benign taken as -ve result]* TP: 193 FN: 5 FP: 23 TN:27 ; *sensitivity*: 0.975, *specificity*: 0.54 *FNAC rated FN/SFN or suspicious or malignant (+ve) [FLUS or benign taken as -ve result]* TP: 187 FN: 11 FP: 18 TN:32 ; *sensitivity*: 0.944, *specificity*: 0.64 *FNAC rated suspicious or malignant (+ve) [FN/SFN or FLUS or benign taken as -ve result]* TP: 158 FN: 40 FP: 14 TN:36 ; *sensitivity*: 0.798, *specificity*: 0.72
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, so some bias possible in who was given surgery)*
Comments	

Reference	Tabaqchali, 2000[343]
Study type	Retrospective
Number of patients	n = 239 patients with 302 FNAs on single or dominant nodules (including 63 repeats aspirations on 45 patients)
Patient characteristics	*Age, mean (range): 48(8.5-85)* *Gender (female to male ratio): 213:26* *Ethnicity*: not reported *Setting*: Endocrine Surgery *Country*: UK *Inclusion criteria*: patients with a dominant thyroid nodule who had FNAC carried out in the 6 year period 1990-1995 and subsequent partial or complete thyroidectomy. *Exclusion criteria*: Not reported

Reference	Tabaqchali, 2000[343]
	Stratum (prior US assessment / no prior US assessment): no prior US reported *Sub-group (US-guided / not US guided): no USG reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only. In those having repeats the highest grade reported was used for diagnostic accuracy analysis. Cytologically inadequate samples were excluded. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 34 ; benign = 205 Inadequate category: 6 malignant, 70 benign *FNAC rated AC3 and above (+ve) [AC2 taken as -ve result]* TP: 25 FN: 9 FP: 136 TN: 69; *sensitivity*: 0.735, *specificity*: 0.337 *FNAC rated AC4 and above (+ve) [AC2-3 taken as -ve result]* TP: 13 FN: 21 FP: 77 TN: 128; *sensitivity*: 0.382, *specificity*: 0.624
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, so some bias possible in who was given surgery)*
Comments	

Reference	Wang, 2020[364]
Study type	Retrospective
Number of patients	n = 274 nodules in 196 patients
Patient characteristics	*Age, mean (SD): 47.24 (12.15)* *Gender (female to male ratio): 168:28* *Ethnicity*: not reported *Setting*: Teaching hospital *Country*: China *Inclusion criteria*: Patients undergoing US, FNA and thyroidectomy *Exclusion criteria*: History of thyroid surgery; thyroid metastasis; surgically removed nodules that were not one-to-one matched with the US findings *Stratum (prior US assessment / no prior US assessment): prior US used as indication for FNA (1 suspicious US characteristic)* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Wang, 2020[364]
Results	Gold standard results: malignant= 114 ;benign= 160 *BSRTC rating used I: DN/UNS; II: benign; III: AUS/FLUS; IV: FN/SFN; V: SFM; VI: Malignant* Inadequate category: 9 malignant, 9 benign *FNAC rated III or above (+ve) [II taken as -ve result]* TP: 99 FN: 15 FP: 67 TN: 93 ; *sensitivity*: 0.868, *specificity*: 0.581 *FNAC rated IV or above (+ve) [II-III taken as -ve result]* TP: 74 FN: 40 FP: 29 TN: 131 ; *sensitivity*: 0.649, *specificity*: 0.819 *FNAC rated V or above (+ve) [II-IV taken as -ve result]* TP: 73 FN: 41 FP: 22 TN: 138 ; *sensitivity*: 0.640, *specificity*: 0.863 *FNAC rated VI (+ve) [II-V taken as -ve result]* TP: 29 FN: 85 FP: 10 TN: 150 ; *sensitivity*: 0.254, *specificity*: 0.938
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, so some bias possible in who was given surgery)*
Comments	

Reference	Wei, 2016[365]
Study type	Retrospective/prospective
Number of patients	n = 78 nodules
Patient characteristics	*Age, mean (range): 47.6(33-64)* *Gender (female to male ratio): 44:34* *Ethnicity*: not reported *Setting*: General Hospital *Country*: China

Reference	Wei, 2016[365]
	Inclusion criteria: Patients with suspicious thyroid nodules, diagnosed with FNA and given surgery *Exclusion criteria*: *Stratum (prior US assessment / no prior US assessment): prior US but did not appear to be an indication for FNA* *Sub-group (US-guided / not US guided): USG used*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear, combined with thin-prep cytology test, which uses a filtration process and thin-layer deposition of cells [appears similar to cytospin]. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=22 ;benign=54 Non diagnostic were excluded from study (n=2) and so could not be included in analysis *FNAC rated malignant (+ve) [benign taken as -ve result]* TP: 20 FN: 2 FP: 1 TN: 53 ; *sensitivity*: 0.909, *specificity*: 0.981
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, so some bias possible in who was given surgery)*
Comments	

Reference	Xiong, 2019[377]
Study type	Retrospective/prospective
Number of patients	n = 578 nodules
Patient characteristics	*Age, median (range): 38(20-81)* *Gender (female to male ratio): 432:146* *Ethnicity*: not reported *Setting*: Teaching hospital *Country*: China *Inclusion criteria*: Patients with thyroid nodules treated at Peking University First Hospital from January 2015 to December 2017 were reviewed. Cases of thyroid follicular lesions with both CNB and resected specimens were retrieved *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Core biopsy *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* Yes *Blinding of gold standard test:* Yes

Reference	Xiong, 2019[377]
Results	Gold standard results: malignant= 541 ;benign=37 Inadequate category: 0 malignant, 1 benign *Used Gradings of the Korean Endocrine Pathology Thyroid Core needle Biopsy Study Group: 1: non-diagnostic or unsatisfactory; II: benign lesion; III: indeterminate lesion; IV follicular neoplasm or suspicious for a follicular neoplasm; V: suspicious for malignancy; VI: malignant* *Core biopsy grades V and VI (+ve) [Grades II, III, IV taken as -ve result]* TP: 489 FN: 52 FP: 1 TN: 36 ; *sensitivity*: 0.904, *specificity*: 0.973 *Core biopsy grades III, V and VI (+ve) [Grades II, IV taken as -ve result]* TP: 519 FN: 22 FP: 2 TN: 35 ; *sensitivity*: 0.959, *specificity*: 0.946
Source of funding	No funding stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias): Serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Zelmanovitz, 1998[391]
Study type	Retrospective
Number of patients	n = 11 nodules
Patient characteristics	*Age, range: 19-47* *Gender (female to male ratio): 11:0* *Ethnicity*: not reported *Setting*: Nuclear Medicine Department *Country*: Brazil *Inclusion criteria*: FNA and thyroidectomy *Exclusion criteria*: None reported

Reference	Zelmanovitz, 1998[391]
	Stratum (prior US assessment / no prior US assessment): no prior US reported *Sub-group (US-guided / not US guided): no USG reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 1 ;benign= 10 Inadequate category: not reported *FNAC rated malignant or indeterminate (+ve) [colloid goitre taken as -ve result]* TP: 1 FN: 0 FP: 1 TN: 9 ; *sensitivity*:1.0, *specificity*: 0.90 *FNAC rated malignant (+ve) [indeterminate or colloid goitre taken as -ve result]* TP: 1 FN: 0 FP: 0 TN: 10 ; *sensitivity*:1.0, *specificity*:1.0
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, so some bias possible in who was given surgery)*
Comments	

Reference	Zhang, 2015[392]
Study type	Retrospective
Number of patients	n = 78 nodules
Patient characteristics	*Age, mean (SD): not reported for those having surgery* *Gender (female to male ratio): not reported for those having surgery* *Ethnicity*: not reported *Setting*: Unclear *Country*: Unclear *Inclusion criteria*: Thyroid nodules undergoing FNA and subsequent thyroidectomy *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): prior US but results not an indication for FNA* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology with ROSE, with smear only Up to a maximum of 4 passes were routinely made if the aspirate was deemed inadequate or unsatisfactory *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Zhang, 2015[392]
Results	Gold standard results: malignant=27 ;benign=51 *FNAC ratings were benign (colloid nodules, hyperplastic nodules and thyroiditis), malignant, suspicious for malignancy, and indeterminate (including follicular or Hurtle cell neoplasm, atypia, or follicular lesion of undetermined significance)* Inadequate category: 0 malignant, 7 benign *FNAC rated indeterminate or malignant/suspicious for malignancy (+ve) [benign taken as -ve result]* TP: 26 FN: 1 FP: 27 TN: 24 ; *sensitivity*: 0.963, *specificity*: 0.471 *FNAC rated malignant/suspicious for malignancy (+ve) [benign or indeterminate taken as -ve result]* TP: 19 FN: 8 FP: 9 TN: 42 ; *sensitivity*: 0.703, *specificity*: 0.824
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, so some bias possible in who was given surgery)*
Comments	

Reference	Raina, 2011[285]
Study type	Retrospective
Number of patients	n = 25 nodules
Patient characteristics	*Age, mean (SD): not reported for those having surgery* *Gender (female to male ratio): not reported for those having surgery* *Ethnicity*: not reported *Setting*: Department of Surgery and ENT *Country*: India *Inclusion criteria*: Patients with thyroid nodules receiving FNA [in review, only those confirmed by histopathology were included, but in paper there were additionally also 71 not sent for surgery. Reasons not given but FNA results not the only reasons as half sent for surgery were benign on FNA]

Reference	Raina, 2011[285]
	Exclusion criteria: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=7; benign=18 Inadequate category: not reported *FNAC rated papillary carcinoma, medullary carcinoma, suspected malignancy (+ve) [follicular neoplasm, multinodular goitre and benign cystic lesion taken as -ve result]* TP: 5 FN: 2 FP: 1 TN: 17 ; *sensitivity*: 0.714, *specificity*: 0.944
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, so some bias possible in who was given surgery)*
Comments	

Reference	Huang, 2020[161]
Study type	Prospective

Reference	Huang, 2020[161]
Number of patients	n = 392 nodules
Patient characteristics	*Age, mean (range): 45.5 (24-77)* *Gender (female to male ratio): 280:112* *Ethnicity*: not reported *Setting*: Teaching Hospital *Country*: China *Inclusion criteria*: 1. Thyroid nodules with 1~4 of the following five suspicious ultrasonic features -"solid nodules, hypoechoic or extremely hypoechoic, irregular boundary, microcalcification, taller-than-wide shape" - based on the classification standard of TI-RADS proposed by Kwak et al; 2. Conventional thyroid ultrasonography, ultrasound elastography and FNAC performed before surgery; and 3. Cytologic results as well as a final diagnosis of the nodules based on postoperative pathology. *Exclusion criteria*: The exclusion criteria were as follows: 1. Surgery for hyperthyroidism; 2. Previous history of neck radiation or surgery; and 3. Thyroid nodules that do not meet the standard of KWAK-TIRADS. *Stratum (prior US assessment / no prior US assessment): prior US – Kwak TIRADs used to indicate FNA* *Sub-group (US-guided / not US guided): Not USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Huang, 2020[161]
Results	Gold standard results: malignant= 233 ;benign= 159 *Bethesda classification used.* Inadequate category: 4 malignant, 3 benign *FNAC rated BSRTC level III or higher (+ve) [level II taken as -ve result]* TP: 228 FN: 5 FP: 124 TN: 35 ; *sensitivity*: 0.979, *specificity*: 0.220 *FNAC rated BSRTC level IV or higher (+ve) [level II-III taken as -ve result]* TP: 218 FN: 15 FP: 33 TN: 126 ; *sensitivity*:0.936, *specificity*:0.792 *FNAC rated BSRTC level V or higher (+ve) [level II-IV taken as -ve result]* TP: 123 FN: 110 FP: 4 TN: 155 ; *sensitivity*: 0.528, *specificity*: 0.975 *FNAC rated BSRTC level VI (+ve) [level II-V taken as -ve result]* TP: 15 FN: 218 FP: 3 TN: 156 ; *sensitivity*:0.064, *specificity*: 0.981
Source of funding	No funding stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): none*
Comments	

Reference	Jalan, 2017[166]
Study type	Prospective
Number of patients	n = 40 nodules
Patient characteristics	*Age, range: 8-71* *Gender (female to male ratio):* *Ethnicity*: not reported *Setting*: Departments of pathology and radiology

Reference	Jalan, 2017[166]
	Country: India *Inclusion criteria*: All patients with complaints of thyroid swelling [for this review, surgery] *Exclusion criteria*: None *Stratum (prior US assessment / no prior US assessment): prior US not reported (US done concurrently)* *Sub-group (US-guided / not US guided): USG and non-USG done in 22, but not the majority. Non-USG done in the other 18*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=11 ;benign=29 Inadequate category: not reported per histological group *FNAC rated follicular neoplasm or malignant (+ve) [non-neoplastic taken as -ve result]* TP: 10 FN: 1 FP: 6 TN: 23 ; *sensitivity*:0.909, *specificity*: 0.793 *FNAC rated malignant (+ve) [follicular neoplasm or non-neoplastic taken as -ve result]* TP: 9 FN: 2 FP: 0 TN: 29 ; *sensitivity*:0.818, *specificity*: 1.0 Note in study the results were separated for conventional FNA and conventional FBNA + USG FNA. Because the latter group were not ALL done with USG FNA it was not deemed appropriate to analyses separately. Hence all have been analysed together.

Reference	Jalan, 2017[166]
Source of funding	No funding stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): none*
Comments	

Reference	Abboud, 2003[1]
Study type	Retrospective
Number of patients	n = 46 nodules
Patient characteristics	*Age, mean (SD): not reported for those having FNAC* *Gender (female to male ratio): Not reported for those having FNAC* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Lebanon *Inclusion criteria*: Patients undergoing thyroidectomy who also had FNAC *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): prior US not reported* *Sub-group (US-guided / not US guided): not specified as USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	Abboud, 2003[1]
	Blinding of index test: No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=15 ;benign=31 *FNAC classification: 1. Benign, 2 Malignant, 3 indeterminate (including atypical features or follicular/Hurthle cell neoplasm), 4 non-diagnostic.* The 3 non-diagnostic cases could not be included in the analysis below as the paper did not report the GS designation for these 3 cases *FNAC rated indeterminate or malignant (+ve) [benign taken as -ve result]* TP: 15 FN: 0 FP: 23 TN: 8 ; *sensitivity*: 1.0, *specificity*: 0.258 *FNAC rated malignant (+ve) [benign or indeterminate taken as -ve result]* TP: 11 FN: 4 FP: 2 TN: 29 ; *sensitivity*: 0.7333, *specificity*: 0.935 *Splitting indeterminate up between suspect/atypical and follicular neoplasm:* *FNAC rated malignant or suspect/atypical indeterminate (+ve) [benign or follicular neoplasm indeterminate taken as -ve result]* TP: 13 FN: 2 FP: 7 TN: 24 ; *sensitivity*: 0.867, *specificity*: 0.774
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, so some bias possible in who was given surgery)*
Comments	

Reference	Acar, 2017[3]
Study type	Retrospective/prospective
Number of patients	n = 226 nodules (pre-Bethesda) and 316 nodules (Bethesda)
Patient characteristics	*Age, mean (SD): 45.4(12.25) (pre-Bethesda) and 47(11.2) (Bethesda)* *Gender (female to male ratio): 79:21 (pre-Bethesda) and 80:20 (Bethesda)*

Reference	Acar, 2017[3]
	Ethnicity: not reported *Setting*: General Surgery *Country*: Turkey *Inclusion criteria*: Patients undergoing total thyroidectomy for thyroid nodules, with FNAC pre-Bethesda or post-Bethesda inception. *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): prior US reported but did not appear to be an indication for FNA provision* *Sub-group (US-guided / not US guided): USG for both groups routinely*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only Aspiration performed twice for each nodule. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	PRE-BETHESDA DATA Gold standard results: malignant=27 ;benign=199 Inadequate category: 1 malignant, 36 benign

Reference	Acar, 2017[3]
	Pre-Bethesda scale: non-diagnostic, benign, follicular lesion, follicular neoplasia, Hurthle cell neoplasia, suspicious for malignancy, and malignant *FNAC rated Follicular lesion, Follicular neoplasia, Hurthle cell neoplasia, suspicious or malignant (+ve) [benign taken as -ve result]* TP: 23 FN: 4 FP: 100 TN: 99 ; *sensitivity*:0.852, *specificity*:0.498 *FNAC rated Follicular neoplasia, Hurthle cell neoplasia, suspicious or malignant (+ve) [Follicular lesion, benign* TP: 23 FN: 4 FP: 93 TN: 106 ; *sensitivity*:0.852, *specificity*: 0.533 *FNAC rated Hurthle cell neoplasia, suspicious or malignant (+ve) [Follicular neoplasia, Follicular lesion, benign taken as -ve result]* TP: 21 FN: 6 FP: 57 TN: 142 ; *sensitivity*:0.778, *specificity*: 0.714 *FNAC rated suspicious or malignant (+ve) [Hurthle cell neoplasia, Follicular neoplasia, Follicular lesion, benign taken as -ve result]* TP: 19 FN: 8 FP: 49 TN: 150 ; *sensitivity*: 0.704, *specificity*: 0.754 *FNAC rated malignant (+ve) [Suspicious, Hurthle cell neoplasia, Follicular neoplasia, Follicular lesion, benign taken as -ve result]* TP: 15 FN: 12 FP: 36 TN: 163 ; *sensitivity*:0.556, *specificity*: 0.819 <u>BETHESDA DATA</u> Gold standard results: malignant=92 ;benign=224 *Bethesda scale: The standard 6 Bethesda groups* <u>Inadequate category: 2 malignant, 13 benign</u> *FNAC rated III or above (+ve) [II rated as -ve result]* TP: 87 FN: 5 FP: 123 TN: 101 ; *sensitivity*: 0.946 , *specificity*: 0.451 *FNAC rated IV or above (+ve) [II-III rated as -ve result]* TP: 82 FN: 10 FP: 59 TN: 164 ; *sensitivity*: 0.891, *specificity*: 0.735 *FNAC rated V or above (+ve) [II-IV rated as -ve result]* TP: 75 FN: 17 FP: 22 TN: 202 ; *sensitivity*: 0.815, *specificity*: 0.902 *FNAC rated VI (+ve) [II-V rated as -ve result]*

Reference	Acar, 2017[3]
	TP: 28 FN: 64 FP: 14 TN: 210 ; *sensitivity*:0.304, *specificity*: 0.938
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, so some bias possible in who was given surgery)*
Comments	

Reference	Afroze, 2002[4]
Study type	Retrospective
Number of patients	n = 170 nodules
Patient characteristics	*Age, range: 16-78* *Gender (female to male ratio): 122-48* *Ethnicity*: not reported *Setting*: Department of pathology *Country*: Pakistan *Inclusion criteria*: Patients undergoing FNAC of thyroid nodules and subsequent thyroid surgery *Exclusion criteria*: Patients without computerised records or operated on outside study hospital *Stratum (prior US assessment / no prior US assessment): no report of any prior US* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology with ROSE, with smear + cytospin and cell block

Reference	Afroze, 2002[4]
	With larger nodules the aspiration was repeated 2 or 3 times from different areas of the gland. Two smears prepared from each aspirate. Patient made to wait 20 minutes and if aspirate inadequate a repeat aspiration made again. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=22 ;benign=148 *FNAC classified as: benign, follicular lesion/neoplasm, suspicious, malignant, insufficient* Inadequate category: 1 malignant, 3 benign *FNAC rated follicular lesion, follicular neoplasm, suspicious, malignant (+ve) [benign taken as -ve result]* TP: 17 FN: 5 FP: 37 TN: 111 ; *sensitivity*: 0.773, *specificity*: 0.75 *FNAC rated follicular neoplasm, suspicious, malignant (+ve) [follicular lesion, and benign taken as -ve result]* TP: 17 FN: 5 FP: 26 TN: 122 ; *sensitivity*: 0.773, *specificity*: 0.824 *FNAC rated suspicious, malignant (+ve) [follicular neoplasm, follicular lesion, and benign taken as -ve result]* TP: 16 FN: 6 FP: 8 TN: 140 ; *sensitivity*: 0.727, *specificity*: 0.946 *FNAC rated malignant (+ve) [follicular neoplasm, follicular lesion, suspicious and benign taken as -ve result]* TP: 13 FN: 9 FP: 4 TN: 144 ; *sensitivity*: 0.591, *specificity*: 0.973
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, so some bias possible in who was given surgery)*
Comments	

Reference	Agrawal, 1995 #1093[8]
Study type	Retrospective
Number of patients	n = 100 nodules
Patient characteristics	*Age, range: 17-70* *Gender (female to male ratio): 74:26* *Ethnicity*: not reported *Setting*: Department of surgery *Country*: India *Inclusion criteria*: Patients for whom FNAC and post-surgical pathology were available *Exclusion criteria*: not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Agrawal, 1995 #1093[8]
Results	Gold standard results: malignant=34 ;benign=66 *FNAC classified as: category I: benign; category II thyroiditis; category III suspicious; category IV malignant; category V: inadequate* Inadequate category: 4 malignant, 7 benign *FNAC rated Thyroiditis, suspicious or malignant (+ve) [benign taken as -ve result]* TP: 26 FN: 8 FP: 21 TN: 45 ; *sensitivity*:0.765, *specificity*: 0.682 *FNAC rated suspicious or malignant (+ve) [benign, Thyroiditis taken as -ve result]* TP: 26 FN: 8 FP: 19 TN: 47 ; *sensitivity*: 0.765, *specificity*: 0.712 *FNAC rated malignant (+ve) [benign, Thyroiditis, suspicious taken as -ve result]* TP: 13 FN: 21 FP: 9 TN: 57 ; *sensitivity*: 0.382, *specificity*: 0.864
Source of funding	
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Aguilar-Diosdado, 1997[9]
Study type	Retrospective/prospective
Number of patients	n = 289 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Setting*: Secondary care *Country*: Spain *Inclusion criteria*: Patients undergoing resection for nodular goitre; carcinoma or suspicious on FNA; thyroid nodule associated with lymphadenopathy; thyroid nodule associated with previous radiation exposure; enlargement of a thyroid mass despite L-thyroxine therapy; clinical symptoms of hoarseness or dysphagia in patients with thyroid nodules [despite specific FNA findings being an

Reference	Aguilar-Diosdado, 1997[9]
	indication for surgery, the fact that most people being sent to surgery had benign FNA findings meant this paper was deemed acceptable for inclusion]. *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): prior US but not used as criterion for FNA* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear + cytospin + cell block Suggestion of cytospin: 'in the case of a cystic lesion all fluid was aspirated, centrifuged and processed for cytologic analysis. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=65 ;benign=224 *FNAC classification: benign (goitre, thyroiditis, thyroid cyst), follicular proliferation (follicular tumour, hyperplastic nodular goitre and HC tumour), malignancy, unsatisfactory specimen* Inadequate category: 3 malignant, 24 benign *FNAC rated follicular proliferation or malignant (+ve) [benign taken as -ve result]* TP: 43 FN: 22 FP: 57 TN: 167 ; *sensitivity*:0.661, *specificity*: 0.746 *FNAC rated malignant (+ve) [benign or follicular proliferation taken as -ve result]*

Reference	Aguilar-Diosdado, 1997[9]
	TP: 24 FN: 41 FP: 29 TN: 195 ; *sensitivity*: 0.369, *specificity*: 0.871
Source of funding	Institute of Health of Spain grant FIS 93/1318
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Al-Hureibi, 2003[18]
Study type	Retrospective
Number of patients	n = 199 nodules
Patient characteristics	*Age, mean (SD): 36.36 (11.95)* *Gender (female to male ratio): 219:24* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Yemen *Inclusion criteria*: Patients undergoing FNA and subsequent thyroid surgery for thyroid nodules/swelling. *Exclusion criteria*: none reported *Stratum (prior US assessment / no prior US assessment): prior US not reported* *Sub-group (US-guided / not US guided): No USG used*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only

Reference	Al-Hureibi, 2003[18]
	Reference (gold) standard: Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=38 ;benign=161 *FNAC classified as benign, thyroiditis, follicular neoplasm, suspicious, malignant* Inadequate category: 1 malignant, 2 benign *FNAC rated malignant or suspicious or follicular neoplasm or thyroiditis (+ve) [benign taken as -ve result]* TP: 15 FN: 23 FP: 32 TN: 129 ; *sensitivity*: 0.395, *specificity*: 0.801 *FNAC rated malignant or suspicious or follicular neoplasm (+ve) [benign or thyroiditis taken as -ve result]* TP: 15 FN: 23 FP: 26 TN: 135 ; *sensitivity*:0.395, *specificity*: 0.839 *FNAC rated malignant or suspicious (+ve) [benign or thyroiditis or follicular neoplasm taken as -ve result]* TP: 6 FN: 32 FP: 4 TN: 157 ; *sensitivity*: 0.158, *specificity*: 0.975
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, so some bias possible in who was given surgery)*
Comments	

Reference	Altavilla, 1990[23]
Study type	Retrospective
Number of patients	n = 257 nodules
Patient characteristics	*Age, mean (SD): Not reported* *Gender (female to male ratio): Not reported*

Reference	Altavilla, 1990[23]
	Ethnicity: not reported *Setting*: Institute of Pathology, University Hospital *Country*: Italy *Inclusion criteria*: Not reported *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=49 ;benign=208 *FNAC classification: benign, thyroiditis, suspect, malignant, inadequate.* Inadequate category: 3 malignant, 21 benign *FNAC rated thyroiditis, suspect or malignant (+ve) [benign taken as -ve result]*

Reference	Altavilla, 1990[23]
	TP: 39 FN: 10 FP: 60 TN: 148 ; *sensitivity*: 0.796, *specificity*: 0.711 *FNAC rated suspect or malignant (+ve) [thyroiditis, benign taken as -ve result]* TP: 38 FN: 11 FP: 56 TN: 152 ; *sensitivity*: 0.776, *specificity*: 0.731 *FNAC rated malignant (+ve) [suspect or thyroiditis, benign taken as -ve result]* TP: 20 FN: 29 FP: 21 TN: 187 ; *sensitivity*: 0.408, *specificity*: 0.899
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, so some bias possible in who was given surgery)*
Comments	

Reference	Al-Taweel, 1990[19]
Study type	Retrospective
Number of patients	n = 91 nodules
Patient characteristics	*Age, range: 18-85* *Gender (female to male ratio): 64:24* *Ethnicity*: not reported *Setting*: Department of Surgery *Country*: Kuwait *Inclusion criteria*: Consecutive patients undergoing FNAC for solitary thyroid nodules with subsequent surgery *Exclusion criteria*: none reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy

Reference	Al-Taweel, 1990[19]
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=17 ;benign=74 *FNAC classification: negative, positive, suspicious, inconclusive(unsatisfactory)* Inadequate category: 0 malignant, 3 benign *FNAC rated positive or suspicious (+ve) [negative taken as -ve result]* TP: 16 FN: 1 FP: 23 TN: 51 ; *sensitivity*: 0.941, *specificity*: 0.689 *FNAC rated positive (+ve) [negative or suspicious taken as -ve result]* TP: 12 FN: 5 FP: 3 TN: 71 ; *sensitivity*: 0.706, *specificity*: 0.959
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, so some bias possible in who was given surgery)*
Comments	

Reference	Ananthakrishnan, 1990[24]
Study type	Retrospective/prospective
Number of patients	n = 150 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported*

Reference	Ananthakrishnan, 1990[24]
	Ethnicity: not reported *Setting*: Department of surgery and pathology *Country*: India *Inclusion criteria*: consecutive patients with a single palpable nodule in thyroid for whom FNAC and histopathology were performed *Exclusion criteria*: No histopathology *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* Yes *Blinding of gold standard test:* Yes
Results	Gold standard results: malignant=21 ;benign=129 FNAC classifications: colloid nodule, thyroiditis, follicular neoplasm, malignant, inadequate Inadequate category: 6 malignant, 28 benign *FNAC rated malignant, follicular neoplasm or thyroiditis (+ve) [colloid nodule taken as -ve result]* TP: 13 FN: 8 FP: 79 TN: 50 ; *sensitivity*:0.619, *specificity*: 0.388

Reference	Ananthakrishnan, 1990[24]
	FNAC rated malignant, follicular neoplasm (+ve) [colloid nodule or thyroiditis taken as -ve result] TP: 12 FN: 9 FP: 78 TN: 51 ; *sensitivity*: 0.571, *specificity*: 0.395 *FNAC rated malignant (+ve) [colloid nodule or thyroiditis or follicular neoplasm taken as -ve result]* TP: 5 FN: 16 FP: 31 TN: 98 ; *sensitivity*: 0.238, *specificity*: 0.760
Source of funding	No funding stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias): No serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Aydogan, 2019[30]
Study type	Retrospective
Number of patients	n = 514 nodules from 371 patients
Patient characteristics	*Age, mean (SD): 50.9(13.4)* *Gender (female to male ratio): 294: 77* *Ethnicity*: not reported *Setting*: Teaching hospital *Country*: Turkey *Inclusion criteria*: Patients undergoing thyroidectomy after FNAC; decision for surgery depended on nodule size, malignant or indeterminate cytology, compressive symptoms, Graves disease and multinodular goitre [adequate number of benign on FNA to allow inclusion to this review]. *Exclusion criteria*: none reported *Stratum (prior US assessment / no prior US assessment): prior US, but did not appear to be an indication for FNA* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy

Reference	Aydogan, 2019[30]
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=161 ;benign= 355 *FNAC classifications were by Bethesda: non-diagnostic, benign, AUS/FLUS, FN/SFN, SFM, malignant* Inadequate category: 19 malignant, 32 benign *FNAC rated malignant, SFM, FN/SFN or AUS/FLUS (+ve) [benign taken as -ve result]* TP: 124 FN: 37 FP: 80 TN: 275 ; *sensitivity*: 0.7790, *specificity*: 0.775 *FNAC rated malignant or SFM or FN/SFN (+ve) [benign or AUS/FLUS taken as -ve result]* TP: 110 FN: 51 FP: 49 TN: 306 ; *sensitivity*: 0.683, *specificity*: 0.862 *FNAC rated malignant or SFM (+ve) [benign or AUS/FLUS* or *FN/SFN taken as -ve result]* TP: 95 FN: 66 FP: 34 TN: 321 ; *sensitivity*: 0.590, *specificity*: 0.904 *FNAC rated malignant (+ve) [benign or AUS/FLUS* or *FN/SFN or SFM taken as -ve result]* TP: 74 FN: 87 FP: 32 TN: 323 ; *sensitivity*: 0.460, *specificity*: 0.910
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, so some bias possible in who was given surgery)*
Comments	

Reference	Bashier, 1996[38]
Study type	Prospective
Number of patients	n = 89 nodules
Patient characteristics	*Age, mean (range): 47 (15-80)* *Gender (female to male ratio): 76:13* *Ethnicity*: not reported *Setting*: Teaching Hospital *Country*: Sudan *Inclusion criteria*: Patients with a solitary or significantly dominant thyroid nodule, followed up by histopathological confirmation *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): prior US but was not a criterion for selection to FNA* *Sub-group (US-guided / not US guided): No report of USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Bashier, 1996[38]
Results	Gold standard results: malignant=12; benign=77 *FNAC classification: not suspicious= nodular goitre; highly suspicious=follicular neoplasm and papillary or anaplastic carcinoma.* *FNAC rated highly suspicious (+ve) [not suspicious taken as -ve result]* Inadequate category: not reported TP: 11 FN: 1 FP: 12 TN: 65 ; *sensitivity*: 0.92, *specificity*: 0.846
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, so some bias possible in who was given surgery)*
Comments	

Reference	Belanger, 1983[41]
Study type	Prospective
Number of patients	n = 63 nodules
Patient characteristics	*Age, mean: 39.7* *Gender (female to male ratio): 55:8* *Ethnicity*: not reported *Setting*: Endocrine unit *Country*: Canada *Inclusion criteria*: Presence of a solid or partially cystic cold nodule; informed consent for surgery regardless of cytological findings; no surgical contraindications *Exclusion criteria*: none reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): no USG reported*

Reference	Belanger, 1983[41]
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=13 ;benign=50 *FNAC categories: benign, suspicious, malignant, inadequate* Inadequate category: 1 malignant, 5 benign *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 11 FN: 2 FP: 8 TN: 42 ; *sensitivity*:0.846, *specificity*: 0.84 *FNAC rated malignant (+ve) [benign or suspicious taken as -ve result]* TP: 9 FN: 4 FP: 6 TN: 44 ; *sensitivity*: 0.692, *specificity*: 0.88
Source of funding	No funding stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): none*
Comments	

Reference	Bellantone, 2004[42]
Study type	Retrospective
Number of patients	n = 119 nodules

Reference	Bellantone, 2004[42]
Patient characteristics	*Age, mean (SD): 46.6(12.8)* *Gender (female to male ratio): 88:31* *Ethnicity*: not reported *Setting*: Division of Endocrine surgery *Country*: Italy *Inclusion criteria*: Patients undergoing UG FNAC and subsequent surgery because of suspicious or malignant cytology, persistently nondiagnostic cytology, cytology consistent with predominantly follicular lesion, incomplete cyst resolution, compressive symptoms and/or large nodular size *Exclusion criteria*: not reported *Stratum (prior US assessment / no prior US assessment): prior US not reported as an indicator of FNA* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear + cytospin + cell block. Some (not a majority) appeared to be exposed to cytospin. Two aspirations done per patient, and for each aspiration 4 glass slides are made *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Bellantone, 2004[42]
Results	Gold standard results: malignant=21 ;benign=98 FNAC classification: benign, thyrocyte hyperplasia without nuclear atypia (THWNA), predominantly follicular lesion (PFL), suspicious (follicular lesion with nuclear pleomorphism), carcinoma, non-diagnostic Inadequate category: 2 malignant, 9 benign *FNAC rated carcinoma, suspicious, PFL or THWNA (+ve) [benign taken as -ve result]* TP: 17 FN: 4 FP: 70 TN: 28 ; *sensitivity*: 0.809, *specificity*: 0.286 *FNAC rated carcinoma, suspicious, or PFL (+ve) [benign or THWNA taken as -ve result]* TP: 16 FN: 5 FP: 59 TN: 39 ; *sensitivity*: 0.762, *specificity*: 0.398 *FNAC rated carcinoma, or suspicious (+ve) [benign or THWNA or PFL taken as -ve result]* TP: 11 FN: 10 FP: 14 TN: 84 ; *sensitivity*: 0.524, *specificity*: 0.857
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, so some bias possible in who was given surgery)*
Comments	

Reference	Biscotti, 1995[47]
Study type	Retrospective
Number of patients	n = 41 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Setting*: Department of anatomic pathology *Country*: USA *Inclusion criteria*: FNA specimens from patients who also provided a histopathological sample at surgery

Reference	Biscotti, 1995[47]
	Exclusion criteria: not reported *Stratum (prior US assessment / no prior US assessment): prior US not reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* *Each patient was given two passes. The first pass was used to prepare two direct smears. The second was rinsed onto Cyto:Lyt solution and then centrifuged and after discarding the supernatant the cell pellet was resuspended and a sample transferred to a second methanol-based preservative* 1. Fine needle aspiration cytology without ROSE, with smear only 2. Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block – Thin-prep *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=8 ;benign=33 FNAC classification: negative, colloid nodule, cyst, Graves, Hashimoto's thyroiditis, Hypercellular follicular nodule possibly malignant (HCFN), papillary carcinoma *STANDARD SMEAR* Inadequate category: not reported *FNAC using rated papillary carcinoma, HCFN, (+ve) [Colloid, cyst, negative, graves, Hashimoto's thyroiditis taken as -ve result]* TP: 8 FN: 0 FP: 5 TN: 28 ; *sensitivity*: 1.0, *specificity*: 0.848

Reference	Biscotti, 1995[47]
	FNAC using rated papillary carcinoma (+ve) [Colloid, cyst, negative, graves, Hashimoto's thyroiditis or HCFN taken as -ve result] TP: 5 FN: 3 FP: 0 TN: 33 ; *sensitivity*: 0.625, *specificity*: 1.0 *THIN-PREP SMEAR* Inadequate category: not reported *FNAC using rated papillary carcinoma, HCFN, (+ve) [Colloid, cyst, negative, graves, Hashimoto's thyroiditis taken as -ve result]* *TP: 8 FN: 0 FP: 7 TN: 26 ; sensitivity: 1.0, specificity: 0.788* *FNAC using rated papillary carcinoma (+ve) [Colloid, cyst, negative, graves, Hashimoto's thyroiditis or HCFN taken as -ve result]* *TP: 5 FN: 3 FP: 0 TN: 33 ; sensitivity: 0.625, specificity: 1.0*
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, so some bias possible in who was given surgery)*
Comments	

Reference	Bodo, 1979[50]
Study type	Retrospective
Number of patients	n = 131 nodules
Patient characteristics	*Age, mean (SD):* *Gender (female to male ratio):* *Ethnicity*: not reported *Setting*: National Oncological Institute *Country*: Hungary *Inclusion criteria*: Patients with diffuse enlargement of the thyroid gland, given FNA and surgery. No reasons given for surgery, but most given surgery were negative on FNA, so FNA not the only criterion. *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported*

Reference	Bodo, 1979[50]
	Sub-group (US-guided / not US guided): unclear
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=49 ;benign=82 FNAC classification: negative, suspect or positive Inadequate category: not reported *FNAC rated suspect or positive (+ve) [negative taken as -ve result]* TP: 42 FN: 7 FP: 8 TN: 74 ; *sensitivity*: 0.857, *specificity*: 0.902 *FNAC rated positive (+ve) [negative or suspect taken as -ve result]* TP: 39 FN: 10 FP: 4 TN: 78 ; *sensitivity*: 0.796, *specificity*: 0.951
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, so some bias possible in who was given surgery)*
Comments	

Reference	Borman, 1995[51]
Study type	Retrospective

Reference	Borman, 1995[51]
Number of patients	n = 27 nodules
Patient characteristics	*Age, mean (SD): Not reported for those given surgery* *Gender (female to male ratio): Not reported for those given surgery* *Ethnicity*: not reported *Setting*: Teaching Hospital *Country*: USA *Inclusion criteria*: Patients with thyroid nodules undergoing FNA with subsequent surgery. Surgery was given if indicated by FNA, or if there were compression symptoms, a recurrent cyst or other clinical suspicion in the presence of benign FNA findings. [Because there were almost half of all cases made up of benign FNA cases this study has been included in the review.] *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): prior US not reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Borman, 1995[51]
Results	Gold standard results: malignant=13 ;benign=14 FNAC classification: follicular neoplasm (FN), papillary carcinoma, benign Inadequate category: 0 malignant, 2 benign *FNAC rated FN or carcinoma (+ve) [benign taken as -ve result]* TP: 13 FN: 0 FP: 4 TN: 10 ; *sensitivity*: 1.0 , *specificity*: 0.714 *FNAC rated carcinoma (+ve) [benign or FN taken as -ve result]* TP: 6 FN: 7 FP: 2 TN: 12 ; *sensitivity*: 0.461 , *specificity*: 0.857
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, so some bias possible in who was given surgery)*
Comments	

Reference	Brauer, 1984[53]
Study type	Retrospective
Number of patients	n = 134 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): 105:29* *Ethnicity*: not reported *Setting*: Head and Neck service, surgical division *Country*: USA *Inclusion criteria*: Patients undergoing FNA for thyroid nodules with subsequent surgery. Majority had hypofunctioning solitary nodules. Initially surgery was given to all patients regardless of FNA. As the study progressed benign findings were less likely to be referred. [However, overall the number of benign FNA findings sent to surgery is sufficient for inclusion to this review] *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US*

Reference	Brauer, 1984[53]
	Sub-group (US-guided / not US guided): USG
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only Negative and inadequate aspirations were repeated when feasible and as often as deemed necessary. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 42 ;benign=92 FNAC classification: positive, questionable, negative Inadequate category: not reported *FNAC rated positive or questionable (+ve) [negative taken as -ve result]* TP: 39 FN: 3 FP: 54 TN: 38 ; *sensitivity*: 0.929, *specificity*: 0.413 *FNAC rated positive (+ve) [negative or questionable taken as -ve result]* TP: 23 FN: 19 FP: 1 TN: 91 ; *sensitivity*: 0.548, *specificity*: 0.989
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, so some bias possible in who was given surgery)*
Comments	

Reference	Bugis, 1986[55]
Study type	Retrospective
Number of patients	n = 198 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): Not reported* *Ethnicity*: not reported *Setting*: Head and Neck Service, General Hospital *Country*: Canada *Inclusion criteria*: Patients presenting with a solitary nodule, with FNA and subsequent surgery. *Exclusion criteria*: None reported *Stratum (prior US assessment / no prior US assessment): No prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Bugis, 1986[55]
Results	Gold standard results: malignant= 30 ;benign=168 FNAC classification: Positive, other (atypical follicular cells or suspicion of papillary formation), negative (benign cyst, adenomatous hyperplasia, colloid nodule, follicular neoplasm or thyroiditis), no reading (inadequate material) Inadequate category: malignant 0, benign 6 *FNAC rated positive or other (+ve) [negative taken as -ve result]* TP: 22 FN: 8 FP: 55 TN: 113 ; *sensitivity*:0.733, *specificity*: 0.673 *FNAC rated positive (+ve) [negative or other taken as -ve result]* TP: 13 FN: 17 FP: 9 TN: 159 ; *sensitivity*: 0.433, *specificity*: 0.946
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, so some bias possible in who was given surgery)*
Comments	

Reference	Choe, 2018[70]
Study type	Retrospective (consecutive)
Number of patients	n = 705 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Setting*: Secondary care *Country*: South Korea *Inclusion criteria*: Patients undergoing core needle biopsy, with subsequent surgery. Reasons for surgery not given. [Some going to surgery had benign CNB results so CNB results were not sole criterion]. *Exclusion criteria*: Not reported

Reference	Choe, 2018[70]
	Stratum (prior US assessment / no prior US assessment): prior US performed and used as criterion for CNB (any one of the standard US abnormal signs) *Sub-group (US-guided / not US guided): not USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Core biopsy *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=532 ;benign=173 CNB classification: non diagnostic, benign, indeterminate, follicular neoplasm, suspicious for malignancy, malignant Inadequate category: malignant 1, benign 3 *CNB rated indeterminate, follicular neoplasm, suspicious for malignancy, or malignant (+ve) [benign taken as -ve result]* TP: 527 FN: 5 FP: 124 TN: 49 ; *sensitivity*:0.991, *specificity*: 0.283 *CNB rated follicular neoplasm, suspicious for malignancy, or malignant (+ve) [indeterminate, or benign taken as -ve result]* TP: 483 FN: 49 FP: 58 TN: 115 ; *sensitivity*: 0.908, *specificity*: 0.665 *CNB rated suspicious for malignancy, or malignant (+ve) [indeterminate, follicular neoplasm, or benign taken as -ve result]* TP: 410 FN: 122 FP: 3 TN: 170 ; *sensitivity*: 0.771, *specificity*: 0.983
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, so some bias possible in who was given surgery)*

Reference	Choe, 2018[70]
Comments	

Reference	Chow, 1999[72]
Study type	Retrospective
Number of patients	n = 76 nodules
Patient characteristics	*Age, mean (SD): 42 (15-72)* *Gender (female to male ratio): not reported for the 76 with FNAC* *Ethnicity*: not reported *Setting*: Department of surgery *Country*: Hong Kong *Inclusion criteria*: Patients with non-toxic solitary thyroid nodules or predominant nodules in non-toxic nodular goitre who underwent surgery with prior FNAC. Benign FNA findings were not routinely sent for surgery unless they increased in size of the patients requested surgery – however most of those referred for surgery were benign on FNAC. *Exclusion criteria*: not reported *Stratum (prior US assessment / no prior US assessment): no prior US* *Sub-group (US-guided / not US guided): not USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	Chow, 1999[72]
	Blinding of index test: No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=12 ;benign=58 FNAC classification: inadequate, benign (colloid, histiocytes, chronic inflammatory cells, benign follicular cells), suspicious (abundant follicular cells in a background of absent or scanty colloid, but frank malignancy not seen), malignant (typical malignant cytological features present). Note that the paper did not report the histopathology for the 6 inadequate cases so these cannot be included in the analysis. Inadequate category: not reported *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 9 FN: 3 FP: 11 TN: 47 ; *sensitivity*: 0.75, *specificity*:0.810 *FNAC rated malignant (+ve) [benign or suspicious taken as -ve result]* TP: 7 FN: 5 FP: 3 TN: 55 ; *sensitivity*: 0.583, *specificity*: 0.948
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, so some bias possible in who was given surgery)*
Comments	

Reference	Cristallini, 1989 #1161[80]
Study type	Retrospective
Number of patients	n = 41 nodules
Patient characteristics	*Age, mean (range): 43.6 (16-84)* *Gender (female to male ratio): 33:8* *Ethnicity*: not reported *Setting*: Surgical centre

Reference	Cristallini, 1989 #1161[80]
	Country: Italy *Inclusion criteria*: Patients undergoing thyroidectomy with prior FNAC *Exclusion criteria*: Toxic nodules *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): no USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block . The residual material containing the smaller fragments was centrifuged and used for cytological smears. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 16 ;benign= 25 FNAC classification: malignant, follicular proliferative, benign, inadequate material Inadequate category: malignant 0, benign 2 *FNAC rated follicular proliferative or malignant (+ve) [benign taken as -ve result]* TP: 15 FN: 1 FP: 9 TN: 16 ; *sensitivity*: 0.938, *specificity*: 0.64 *FNAC rated malignant (+ve) [follicular proliferative or benign taken as -ve result]* TP: 15 FN: 1 FP: 2 TN: 23 ; *sensitivity*: 0.938, *specificity*: 0.92

Reference	Cristallini, 1989 #1161[80]
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, so some bias possible in who was given surgery)*
Comments	

Reference	Danese, 1998[85]
Study type	Retrospective
Number of patients	n = 535 (conventional FNA) + 540 (UG FNA) nodules
Patient characteristics	*Age, mean (SD): Not reported in those given surgery* *Gender (female to male ratio): Not reported in those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Italy *Inclusion criteria*: Consecutive patients with single or multiple thyroid nodules given either conventional or UG FNA, followed by surgery. *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG and no USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block.

Reference	Danese, 1998[85]
	Reference (gold) standard: Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	UG FNA Gold standard results: malignant= 103 ;benign= 437 FNAC classification: Inadequate, benign (colloid nodule, cyst, Hashimoto's or subacute thyroiditis), suspicious (indeterminate cytological pattern of follicular neoplasia), malignant (papillary/follicular carcinomas; medullary and anaplastic carcinomas) Inadequate category: malignant 1, benign 4 *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 99 FN: 4 FP: 130 TN: 307 ; *sensitivity*: 0.961, *specificity*: 0.703 *FNAC rated malignant (+ve) [suspicious or benign taken as -ve result]* TP: 70 FN: 33 FP: 4 TN: 433 ; *sensitivity*: 0.680, *specificity*: 0.991 Conventional FNA Gold standard results: malignant= 88 ;benign= 447 FNAC classification: Inadequate, benign (colloid nodule, cyst, Hashimoto's or subacute thyroiditis), suspicious (indeterminate cytological pattern of follicular neoplasia), malignant (papillary/follicular carcinomas; medullary and anaplastic carcinomas) Inadequate category: malignant 2, benign 11 *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 79 FN: 9 FP: 147 TN: 300 ; *sensitivity*: 0.898, *specificity*: 0.671

Reference	Danese, 1998[85]
	FNAC rated malignant (+ve) [suspicious or benign taken as -ve result] TP: 53 FN: 35 FP: 13 TN: 434 ; *sensitivity*: 0.602, *specificity*: 0.971
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, so some bias possible in who was given surgery)*
Comments	

Reference	Davidsohn, 1995[88]
Study type	Retrospective
Number of patients	n = 50 nodules
Patient characteristics	*Age, mean (range): 52 (27-77)* *Gender (female to male ratio): 47:3* *Ethnicity*: not reported *Setting*: Division of Endocrinology *Country*: USA *Inclusion criteria*: Patients having an FNA for thyroid nodules with subsequent thyroidectomy. If FNA was benign surgery would still be given because of large nodules, patient preference or for cosmetic reasons *Exclusion criteria*: None reported *Stratum (prior US assessment / no prior US assessment): prior US not reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology with ROSE, with smear + cytospin and cell block

Reference	Davidsohn, 1995[88]
	Several aspirations were performed and material was given to a cytotechnologist who was present during the procedure, Material from each pass was smeared on paired slides; one was air dried and the other was immediately alcohol fixed. The needle was rinsed in either normal saline or RPMI and cell block was prepared. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=15 ;benign=29 (note: no histopathology reported for the 6 with inadequate FNAC classification) FNAC classification: benign, malignant, suspicious or indeterminate (lesions with possible malignant potential), and inadequate Inadequate category: not reported *FNAC rated suspicious/indeterminate or malignant (+ve) [benign taken as -ve result]* TP: 15 FN: 0 FP: 17 TN: 12 ; *sensitivity*:1.0, *specificity*: 0.414 *FNAC rated malignant (+ve) [suspicious/indeterminate or benign taken as -ve result]* TP: 10 FN: 5 FP: 0 TN: 29 ; *sensitivity*: 0.667, *specificity*: 1.0
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, so some bias possible in who was given surgery)*
Comments	

Reference	de Roy van Zuidewijn, 1994[90]
Study type	Retrospective
Number of patients	n = 265 nodules

Reference	de Roy van Zuidewijn, 1994[90]
Patient characteristics	*Age, mean (SD): Not reported for those with FNA having surgery* *Gender (female to male ratio): Not reported for those with FNA having surgery* *Ethnicity*: not reported *Setting*: Departments of Surgery/Pathology *Country*: Holland *Inclusion criteria*: Patients undergoing FNA and thyroidectomy *Exclusion criteria*: none reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	de Roy van Zuidewijn, 1994[90]
Results	Gold standard results: malignant= 87 ;benign=178 FNAC classification: benign (class I), probably benign (class II), uncertain (class 3), probably malignant (class 4), malignant (class 5) and non-evaluable Inadequate category: malignant 1, benign 4 *FNAC rated class 3 or higher (+ve) [1-2 taken as -ve result]* TP: 80 FN: 7 FP: 63 TN: 115 ; *sensitivity*:0.920, *specificity*: 0.646 *FNAC rated class 4 or higher (+ve) [1-3 taken as -ve result]* TP: 68 FN: 19 FP: 19 TN: 159 ; *sensitivity*: 0.782, *specificity* 0.893 *FNAC rated class 5 (+ve) [1-4 taken as -ve result]* TP: 57 FN: 30 FP: 6 TN: 172 ; *sensitivity*: 0.655, *specificity* 0.966
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, so some bias possible in who was given surgery)*
Comments	

Reference	de Vos tot Nederveen Cappel, 2001[91]
Study type	Retrospective
Number of patients	n = 254 nodules in 231 patients
Patient characteristics	*Age, mean (range): 45 (12-82)* *Gender (female to male ratio): 183:48* *Ethnicity*: not reported *Setting*: Secondary care *Country*: Holland *Inclusion criteria*: Patients with FNACs carried out for thyroid nodules followed by thyroid surgery. People benign on FNA were eligible for surgery if they had a rapidly growing nodule causing local compression, or due to cosmetic reasons

Reference	de Vos tot Nederveen Cappel, 2001[91]
	Exclusion criteria: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* Yes *Blinding of gold standard test:* No
Results	Gold standard results: malignant=59 ;benign=195 FNAC classification: benign (smears with much colloid and few follicular cells), suspicious (follicular proliferation with minimal/no colloid and many follicular cells, and suggestive but not conclusive findings of malignancy), malignant, unsatisfactory, or inadequate Inadequate category: malignant 10, benign 40 *FNAC rated suspect or malignant (+ve) [benign taken as -ve result]* TP: 46 FN: 13 FP: 90 TN: 105 ; *sensitivity*: 0.780, *specificity*: 0.538 *FNAC rated malignant (+ve) [benign or suspect taken as -ve result]* TP: 33 FN: 26 FP: 41 TN: 154 ; *sensitivity*: 0.559, *specificity*: 0.790
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, so some bias possible in who was given surgery)*

Reference	de Vos tot Nederveen Cappel, 2001[91]
Comments	

Reference	Dwarakanathan, 1989[97]
Study type	Retrospective
Number of patients	n = 63 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Department of Internal Medicine *Country*: USA *Inclusion criteria*: Patients undergoing FNA and subsequent surgery for single nodules or multinodular goitres with a dominant nodule. Most nodules were cold on scan. Surgery was given for benign FNA findings for reasons of patient preference, cosmetic considerations, large goitres, large nodules, and other clinically worrisome features such as the age of the patient or male sex (n=26). This ensured all of the FNA categories were covered in the study. *Exclusion criteria*: not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): no USG used*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block The slides were prepared and stained by the standard Papanicolaou method. After this 1 mL of normal saline was aspirated into the syringe and the contents were subjected to cellblock examination. *Reference (gold) standard:* Surgical histopathological findings

Reference	Dwarakanathan, 1989[97]
	Time between measurement of index test and reference standard: Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=19 ;benign=44 FNAC classification: benign (class I and II) including colloid cells, thyroiditis, scanty degenerated cells, regular looking cells; possibly malignant (class III) including suspicious or atypical cells and increased follicular elements; probably malignant or malignant (class IV) including hyperchromasia, prominent nucleoli and mitoses. Papillary cancer features included cobble-stoning of nucleoli, nuclear vacuoles, psammoma bodies and papillary structures Inadequate category: not reported *FNAC rated III and above (+ve) [I and II taken as -ve result]* TP: 18 FN: 1 FP: 19 TN: 25 ; *sensitivity*: 0.947, *specificity*: 0.568 *FNAC rated IV (+ve) [I -III taken as -ve result]* TP: 15 FN: 4 FP: 1 TN: 43 ; *sensitivity*: 0.789, *specificity*: 0.977
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, so some bias possible in who was given surgery)*
Comments	

Reference	El Hag, 2021[98]
Study type	Retrospective
Number of patients	n = 323 nodules
Patient characteristics	*Age, mean (SD): Not reported* *Gender (female to male ratio): Not reported*

Reference	El Hag, 2021[98]
	Ethnicity: not reported *Setting*: Security Forces Hospital *Country*: Saudi Arabia *Inclusion criteria*: All thyroid FNAs with histopathology follow up *Exclusion criteria*: None reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology with ROSE, with smear only All FNAs were performed by a radiologist, under image guidance, and the specimens' adequacy was assessed on site. The FNA smears were stained by both diff quick and pap. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=112 (if including non-invasive follicular tumour with papillary-like nuclear features as malignant) ;benign=211 FNAC classification: Bethesda, using standard 6 categories: ND (1), benign (2), AUS (3), SFN (4), SFM (5), Malignant (6)

Reference	El Hag, 2021[98]
	Inadequate category: unclear *FNAC rated 3 or more (+ve) [2 taken as -ve result]* TP: 99 FN: 13 FP: 56 TN: 155 ; *sensitivity*: 0.884, *specificity*: 0.734 *FNAC rated 4 or more (+ve) [2-3 taken as -ve result]* TP: 81 FN: 31 FP: 22 TN: 189 ; *sensitivity*: 0.723, *specificity*: 0.895 *FNAC rated 5 or more (+ve) [2-4 taken as -ve result]* TP: 59 FN: 53 FP: 5 TN: 206 ; *sensitivity*: 0.527, *specificity*: 0.976 *FNAC rated 6 (+ve) [2-5 taken as -ve result]* TP: 40 FN: 72 FP: 50 TN: 161 ; *sensitivity*: 0.357, *specificity*: 0.763 Gold standard results: malignant=94 (if NOT including non-invasive follicular tumour with papillary-like nuclear features as malignant) ;benign=229 FNAC classification: Bethesda, using standard 6 categories: ND (1), benign (2), AUS (3), SFN (4), SFM (5), Malignant (6) Inadequate category: unclear *FNAC rated 3 or more (+ve) [2 taken as -ve result]* TP: 85 FN: 9 FP: 70 TN: 159 ; *sensitivity*: 0.904, *specificity*: 0.694 *FNAC rated 4 or more (+ve) [2-3 taken as -ve result]* TP: 74 FN: 20 FP: 29 TN: 200 ; *sensitivity*: 0.787, *specificity*: 0.873 *FNAC rated 5 or more (+ve) [2-4 taken as -ve result]* TP: 59 FN: 35 FP: 5 TN: 224 ; *sensitivity*: 0.628 , *specificity*: 0.978 *FNAC rated 6 (+ve) [2-5 taken as -ve result]* TP: 40 FN: 54 FP: 1 TN: 228 ; *sensitivity*: 0.426, *specificity*: 0.996
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, so some bias possible in who was given surgery)*

Reference	El Hag, 2021[98]
Comments	

Reference	Ferrari, 1985[106]
Study type	Retrospective
Number of patients	n = 68 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Department of Internal medicine *Country*: Italy *Inclusion criteria*: Patients with cold nodules undergoing FNA and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block The material obtained was smeared on a slide, fixed and stained. Cystic formations were completely emptied; the liquid obtained was centrifuged and treated as described above. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	Ferrari, 1985[106]
	Blinding of index test: No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=9 (including 1 Hodgkin's disease in the inadequate FNA category) ;benign=59 FNAC classification: inadequate, benign (cystic or colloid formations and thyroiditis), uncertain/suspicious (follicular proliferations and oncocytic adenomas) Inadequate category: malignant 2, benign 0 *FNAC rated uncertain/suspicious or malignant (+ve) [benign taken as -ve result]* TP: 7 FN: 2 FP: 16 TN:43 ; *sensitivity*:0.778, *specificity*: 0.729 *FNAC rated malignant (+ve) [uncertain/suspicious or benign taken as -ve result]* TP: 6 FN: 3 FP: 0 TN:59 ; *sensitivity*: 0.667, *specificity*: 1.0
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, so some bias possible in who was given surgery)*
Comments	

Reference	Gardiner, 1986[123]
Study type	Retrospective
Number of patients	n = 207 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Setting*: Secondary care *Country*: Canada

Reference	Gardiner, 1986[123]
	Inclusion criteria: Patients given FNAC for diffuse thyroid enlargements, multinodular thyroids and thyroids with discrete nodules; subsequent surgery *Exclusion criteria*: not reported *Stratum (prior US assessment / no prior US assessment): no prior US* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=45 ;benign=162 FNAC classification: unsatisfactory; benign (scant mixture of colloid and uniform follicular cells); atypical; malignant Inadequate category: malignant 2, benign 19 *FNAC rated atypical or malignant (+ve) [benign taken as -ve result]* TP: 28 FN: 17 FP: 46 TN:116 ; *sensitivity*: 0.622, *specificity*: 0.716 *FNAC rated malignant (+ve) [atypical or benign taken as -ve result]* TP: 11 FN: 34 FP: 19 TN:143 ; *sensitivity*: 0.244, *specificity*: 0.883
Source of funding	No funding stated

Reference	Gardiner, 1986[123]
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Gershengorn, 1977[126]
Study type	Retrospective/prospective
Number of patients	n = 33 nodules
Patient characteristics	*Age, mean (range): 39 (22-63)* *Gender (female to male ratio): 28:5* *Ethnicity*: not reported *Setting*: Clinical endocrinology *Country*: USA *Inclusion criteria*: Fifty consecutive patients presenting with discrete usually single thyroid nodules given FNA and surgery *Exclusion criteria*: none reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): No USG reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only Each nodule was aspirated twice. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:*

Reference	Gershengorn, 1977[126]
	Not clear *Blinding of index test:* Yes *Blinding of gold standard test:* No
Results	Gold standard results: malignant=12 ;benign=20 FNAC classification: inadequate, benign, suspicious (occasional epithelial cells showed marked cellular changes suggestive of malignancy or when cells were abundant but aggregated together in clumps preventing interpretation), malignant (large numbers of cohesive epithelial cells showed marked variation in size, shape and nuclear structure, often with enlarged, irregular and multiple nuclei. In the single inadequate case no histopathology was given, so it cannot be included in the analysis. <u>Inadequate category: not reported</u> *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 11 FN: 1 FP: 3 TN: 17 ; *sensitivity*: 0.917, *specificity*: 0.85 *FNAC rated malignant (+ve) [suspicious or benign taken as -ve result]* TP: 8 FN: 4 FP: 1 TN: 19 ; *sensitivity*: 0.667, *specificity*: 0.95
Source of funding	<u>No funding stated</u>
Limitations	*Risk of bias (QUADAS 2 – risk of bias): Serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Giansanti, 1989[127]
Study type	Retrospective/prospective
Number of patients	n = 114 nodules
Patient characteristics	*Age, mean (SD): not reported for those having surgery* *Gender (female to male ratio): not reported for those having surgery* *Ethnicity*: not reported

Reference	Giansanti, 1989[127]
	Setting: Centre for Nuclear Medicine *Country*: Italy *Inclusion criteria*: Patients with solid, cold, thyroid nodules, with FNA and subsequent surgery. *Exclusion criteria*: not reported *Stratum (prior US assessment / no prior US assessment): prior US but did not appear to be an indication for FNA* *Sub-group (US-guided / not US guided): no USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 25;benign=89 FNAC classification: positive: malignant neoplasm, follicular proliferative lesion (suspected neoplasm), Hurthle cell neoplasm; negative: inflammatory lesion, nonneoplastic lesion and unsuitable for diagnosis Inadequate category: not reported *FNAC rated indeterminate or malignant (+ve) [benign taken as -ve result]* TP: 20 FN: 5 FP: 27 TN: 62 ; *sensitivity*: 0.80, *specificity*: 0.697

Reference	Giansanti, 1989[127]
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, so some bias possible in who was given surgery)*
Comments	

Reference	Gossain, 1998[131]
Study type	Retrospective
Number of patients	n = 19 nodules
Patient characteristics	*Age, mean (SD): not reported for those having surgery* *Gender (female to male ratio): not reported for those having surgery* *Ethnicity*: not reported *Setting*: Division of Endocrinology and metabolism *Country*: USA *Inclusion criteria*: Patients with a single palpable nodule, undergoing FNA followed by surgery *Exclusion criteria*: None reported *Stratum (prior US assessment / no prior US assessment): US reported but not an indication for FNA* *Sub-group (US-guided / not US guided): no USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only

Reference	Gossain, 1998[131]
	Reference (gold) standard: Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 9;benign=10 FNAC classification: inadequate, benign (cellular architecture consistent with nodular goitre, lymphocytic thyroiditis or granulomatous thyroiditis), suggestive of malignancy (papillary clusters or follicular cells, Hurthle cells without evidence of lymphocytic thyroiditis, clear nuclear inclusions, or psammoma bodies), or malignant (architecture consistent with the corresponding malignant tumour) Inadequate category: malignant 0, benign 0 *FNAC rated suggestive of malignancy or malignant (+ve) [benign taken as -ve result]* TP: 7 FN: 2 FP: 1 TN: 9 ; *sensitivity*: 0.778, *specificity*: 0.9 *FNAC rated malignant (+ve) [suggestive of malignancy or benign taken as -ve result]* TP: 4 FN: 5 FP: 0 TN: 10 ; *sensitivity*: 0.444, *specificity*: 1.0
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, so some bias possible in who was given surgery)*
Comments	

Reference	Gould, 1989[133]
Study type	Retrospective
Number of patients	n = 69 nodules
Patient characteristics	*Age, mean (SD): Not reported*

Reference	Gould, 1989[133]
	Gender (female to male ratio): Not reported *Ethnicity*: not reported *Setting*: University Hospital *Country*: USA *Inclusion criteria*: People with thyroid nodules with an FNA, touch imprint and final histopathology *Exclusion criteria*: not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only The cytology preparations were examined for the presence of nuclear grooves and cytoplasmic and intranuclear inclusions. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Gould, 1989[133]
Results	Gold standard results: malignant=24 ;benign=46 INCLUSIONS FNAC classification: 0=no inclusions; 1=1 inclusion, 2= 2 inclusions, 3=3 or more inclusions Inadequate category: not reported *FNAC rated 1 or more inclusions (+ve) [0 inclusions taken as -ve result]* TP: 13 FN: 11 FP: 1 TN: 45 ; *sensitivity*: 0.542, *specificity*: 0.978 GROOVES Inadequate category: not reported FNAC classification: 0=no grooves; 1=1 groove, 2= 2 grooves, 3=3 or more grooves *FNAC rated 1 or more grooves (+ve) [0 grooves taken as -ve result]* TP: 22 FN: 1 FP: 27 TN: 19 ; *sensitivity*: 0.957, *specificity*: 0.413 *FNAC rated 2 or more grooves (+ve) [0-1 grooves taken as -ve result]* TP: 18 FN: 5 FP: 8 TN: 38 ; *sensitivity*: 0.783, *specificity*: 0.826 *FNAC rated 3 or more grooves (+ve) [0-2 grooves taken as -ve result]* TP: 11 FN: 12 FP: 0 TN: 46 ; *sensitivity*: 0.478, *specificity*: 1.0
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, so some bias possible in who was given surgery)*
Comments	

Reference	Guo, 2015[138]
Study type	Retrospective
Number of patients	n = 489 nodules
Patient characteristics	*Age, mean (SD): not reported for those having surgery* *Gender (female to male ratio): Not reported for those having surgery*

Reference	Guo, 2015[138]
	Ethnicity: not reported *Setting*: Departments of pathology and diagnostic radiology *Country*: China *Inclusion criteria*: All thyroid FNAs that were followed by surgery; indications for FNA were palpable nodules with US finding suggesting malignancy such as microcalcification, margin irregularity, intranodular vascularity or taller than wide shape *Exclusion criteria*: None reported *Stratum (prior US assessment / no prior US assessment): prior US* *Sub-group (US-guided / not US guided): USG for those using TP with non-palpable nodules: 79.3%)*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only Aspiration was performed at least 4-6 times. Biopsies were performed 1-2 times for every nodule. For palpable nodules, the cytopathologist prepared one conventional preparation and the residual specimens in the needle were rinsed in cytolyt for a ThinPrep (TP) slide. One TP slide was prepared for non-palpable nodules and the FNA was performed by a radiologist. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Guo, 2015[138]
Results	Gold standard results: malignant= 425 ;benign= 64 FNAC classification: Bethesda 1-6 (1=ND, 2=benign, 3=AUS/FLUS, 4=FN/SFN, 5=SM, 6=M) <u>Inadequate category: malignant 5, benign 5</u> *FNAC rated 3 or more (+ve) [2 taken as -ve result]* TP: 399 FN: 26 FP: 36 TN: 28 ; *sensitivity*: 0.939, *specificity*:0.438 *FNAC rated 4 or more (+ve) [2-3 taken as -ve result]* TP: 383 FN: 42 FP: 23 TN: 41 ; *sensitivity*: 0.901, *specificity*:0.641 *FNAC rated 5 or more (+ve) [2-4 taken as -ve result]* TP: 382 FN: 41 FP: 18 TN: 46 ; *sensitivity*: 0.899, *specificity*:0.719 *FNAC rated 6 (+ve) [2-5 taken as -ve result]* TP: 289 FN: 134 FP: 5 TN: 59 ; *sensitivity*: 0.68, *specificity*: 0.922
Source of funding	<u>No funding stated</u>
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Hamming, 1990[149]
Study type	Retrospective
Number of patients	n = 169 nodules
Patient characteristics	*Age, median (range): 58 (14-81)* *Gender (female to male ratio): 129: 40* *Ethnicity*: not reported *Setting*: Department of surgery *Country*: Holland

Reference	Hamming, 1990[149]
	Inclusion criteria: Patients with nodular thyroid disease given FNA and subsequent surgery. Surgery performed to confirm or exclude a malignant neoplasm or to remove a nodular goitre for cosmetic or mechanical reasons. *Exclusion criteria*: None reported *Stratum (prior US assessment / no prior US assessment): prior US but not used as indication for FNA* *Sub-group (US-guided / not US guided): Not USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only In 33 cases more than 1 biopsy was done because of an inadequate sample or doubt about the result and in these cases the last assessable sample was used for evaluation. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=39 ;benign=130 FNAC classification: not assessable, benign, uncertain, malignant Inadequate category: malignant 1, benign 4 *FNAC rated uncertain or malignant (+ve) [benign taken as -ve result]* TP: 35 FN: 4 FP: 41 TN: 89 ; *sensitivity*: 0.897, *specificity*: 0.685 *FNAC rated malignant (+ve) [uncertain or benign taken as -ve result]*

Reference	Hamming, 1990[149]
	TP: 29 FN: 10 FP: 6 TN: 124 ; *sensitivity*: 0.744, *specificity*: 0.954
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, so some bias possible in who was given surgery)*
Comments	

Reference	Harsoulis, 1986[152]
Study type	Retrospective/prospective
Number of patients	n = 213 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Setting*: Endocrine outpatient clinic *Country*: Greece *Inclusion criteria*: Patients with a solitary or dominant thyroid nodule within either a multinodular or diffusely enlarged gland who were subsequently given surgery. Surgery was indicated by FNA but also by the recent appearance of a cold solid nodule, a history of recurrent cysts and for all male patients *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): no USG reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only

Reference	Harsoulis, 1986[152]
	Reference (gold) standard: Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* Yes *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 37 ;benign= 176 FNAC classification: not assessable, benign, suspicious, malignant Inadequate category: 0 = malignant, 23 benign *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 33 FN: 4 FP: 30 TN: 146 ; *sensitivity*: 0.892, *specificity*: 0.685 Note that non assessable data has been incorporated in review analysis (but left out in original paper)
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, so some bias possible in who was given surgery)*
Comments	

Reference	Heimann, 1964[155]
Study type	Retrospective
Number of patients	n = 23 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Setting*: Unclear

Reference	Heimann, 1964[155]
	Country: Unclear *Inclusion criteria*: Patients undergoing FNA and subsequent surgery *Exclusion criteria*: not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=5; benign=18 FNAC classification: benign, suspicious or malignant Inadequate category: not reported *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 4 FN: 1 FP: 0 TN: 18 ; *sensitivity*: 0.80, *specificity*: 1.0
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, so some bias possible in who was given surgery)*

Reference	Heimann, 1964[155]
Comments	

Reference	Hosokawa, 2019[159]
Study type	Retrospective
Number of patients	n = 685 nodules
Patient characteristics	*Age, mean (SD): not reported for thyroid sub-group* *Gender (female to male ratio): not reported of thyroid sub-group* *Ethnicity*: not reported *Setting*: secondary care *Country*: Japan *Inclusion criteria*: Patients undergoing FNA and surgery on thyroid nodules *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No

Reference	Hosokawa, 2019[159]
	Blinding of gold standard test: No
Results	Gold standard results: malignant= 272 ;benign= 413 FNAC classification: Bethesda Inadequate category: used THY1 as negative and not possible to extricate *FNAC rated IV to VI (+ve) [benign taken as I-III]* TP: 222 FN: 50 FP: 21 TN: 392 ; *sensitivity*: 0.816, *specificity*: 0.949
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, so some bias possible in who was given surgery)*
Comments	

Reference	Hussain, 1993[163]
Study type	Retrospective
Number of patients	n = 108 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Setting*: District General Hospital *Country*: UK *Inclusion criteria*: Patients identified by radionuclide imaging as having a solitary cold thyroid nodule, who had FNA followed by surgery; surgery carried out on all patients with a solitary cold nodule *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): No prior US reported*

Reference	Hussain, 1993[163]
	Sub-group (US-guided / not US guided): USG not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology with ROSE, with smear + cytospin and cell block. The material obtained was transferred to a glass slide smeared and fixed with cytospray. If the aspirate was small then cytospin was added to the syringe. The aspirate was examined by the same cytologist. If the aspirate was deemed inadequate it was repeated at the same visit. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 7;benign=101 FNAC classification: benign (follicular adenoma, colloid nodule, non-specific), inadequate, suspicious (cannot exclude Ca), malignant (i.e., papillary or follicular Ca) Inadequate category: malignant 0, benign 21 *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 6 FN: 1 FP: 29 TN: 72 ; *sensitivity*: 0.857, *specificity*: 0.713 *FNAC rated malignant (+ve) [suspicious or benign taken as -ve result]* TP: 4 FN: 3 FP: 21 TN: 80 ; *sensitivity*: 0.571, *specificity*: 0.792
Source of funding	South East Thames Regional Health Authority Recent Medical Advances Fund

Reference	Hussain, 1993[163]
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Jayaram, 1999[168]
Study type	Retrospective
Number of patients	n = 325 nodules
Patient characteristics	*Age, mean (SD): Not reported* *Gender (female to male ratio): Not reported* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Malaysia *Inclusion criteria*: Patients with thyroid lesions given FNA and thyroid surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US* *Sub-group (US-guided / not US guided): no USG reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology with ROSE, with smear only In selected cases a Diff-Quik stain was done at the bedside on one smear and examined under a microscope. Based on the findings of the Diff-Quik stained smear, needling was repeated if required to obtain additional smears for any subsequent special or immune-staining techniques *Reference (gold) standard:* Surgical histopathological findings

Reference	Jayaram, 1999[168]
	Time between measurement of index test and reference standard: Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 64 ;benign= 261 FNAC classification: Carcinoma (including primitive neuroectodermal tumour), Hurthle cell tumour, follicular neoplasm/equivocal, no malignancy/nodular goitre, inadequate. Inadequate category: malignant 3, benign 10 *FNAC rated carcinoma, Hurthle cell tumour, follicular neoplasms/equivocal (+ve) [no malignancy/nodular goitre taken as -ve result]* TP: 57 FN: 7 FP: 73 TN: 188 ; *sensitivity*: 0.891, *specificity*: 0.720 *FNAC rated carcinoma, Hurthle cell tumour (+ve) [follicular neoplasms/equivocal, no malignancy/nodular goitre taken as -ve result]* TP: 35 FN: 29 FP: 13 TN: 248 ; *sensitivity*: 0.547, *specificity*: 0.950 *FNAC rated carcinoma (+ve) [follicular neoplasms/equivocal, no malignancy/nodular goitre* or *Hurthle cell tumour taken as -ve result]* TP: 32 FN: 32 FP: 10 TN: 251 ; *sensitivity*: 0.5, *specificity*: 0.962
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, so some bias possible in who was given surgery)*
Comments	

Reference	Kelman, 2001[175]
Study type	Retrospective
Number of patients	n = 109 nodules
Patient characteristics	*Age, mean (SD): Not reported for those having surgery* *Gender (female to male ratio): Not reported for those having surgery*

Reference	Kelman, 2001[175]
	Ethnicity: not reported *Setting*: Teaching Hospital *Country*: USA *Inclusion criteria*: Patients presenting with a thyroid nodule, who were given FNA and subsequent surgery *Exclusion criteria*: None *Stratum (prior US assessment / no prior US assessment): prior US not reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=134 ;benign=350 FNAC classification: inadequate, nodular goitre/chronic thyroiditis/microfollicles, atypia, hurthle cell neoplasm or malignant Inadequate category: malignant 37, benign 172 *FNAC rated atypia, microfollicles, hurthle cell neoplasm or malignant (+ve) [nodular goitre/chronic thyroiditis taken as -ve result]*

Reference	Kelman, 2001[175]
	TP: 91 FN: 43 FP: 246 TN: 104 ; *sensitivity*: 0.679, *specificity*: 0.297 *FNAC rated atypia, hurthle cell neoplasm or malignant (+ve) [nodular goitre/chronic thyroiditis/microfollicles taken as -ve result]* TP: 87 FN: 47 FP: 203 TN: 147 ; *sensitivity*: 0.649, *specificity*: 0.420
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, so some bias possible in who was given surgery)*
Comments	

Reference	Kim, 2013[182]
Study type	Retrospective
Number of patients	n = 200 nodules
Patient characteristics	*Age, mean (SD): not reported for those having surgery* *Gender (female to male ratio): not reported for those having surgery* *Ethnicity*: not reported *Setting*: Teaching Hospital *Country*: South Korea *Inclusion criteria*: Patients with thyroid nodules with a >90% solid component with maximum diameter of 5mm; underwent FNA and surgery *Exclusion criteria*: None reported *Stratum (prior US assessment / no prior US assessment): prior US not reported* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only

Reference	Kim, 2013[182]
	For each sample, a smear was prepared on 4-6 slides. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=142 ;benign=58 FNAC classification: Bethesda I-VI Inadequate category: not reported *FNAC rated III and above (+ve) [I-II taken as -ve result]* TP: 118 FN: 24 FP: 11 TN: 47 ; *sensitivity*: 0.831, *specificity*: 0.810 *FNAC rated V and above (+ve) [I-IV taken as -ve result]* TP: 103 FN: 39 FP: 4 TN: 54 ; *sensitivity*: 0.725, *specificity*: 0.931
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, so some bias possible in who was given surgery)*
Comments	

Reference	Kimoto, 1999[187]
Study type	Retrospective
Number of patients	n = 61 nodules
Patient characteristics	*Age, mean (SD): not reported for those having surgery* *Gender (female to male ratio): 61:0*

Reference	Kimoto, 1999[187]
	Ethnicity: not reported *Setting*: Department of Surgery *Country*: Japan *Inclusion criteria*: none reported *Exclusion criteria*: none reported *Stratum (prior US assessment / no prior US assessment): prior US used to decide who would have FNA: if US showed simple cysts, small cysts of <10mm with echogenic area, small homogenous solid areas <5mm with a regular margin and minute calcified lesions of <3mm in diameter then these would NOT be given FNA* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=44 ;benign=17 FNAC classification: class I – no atypical cells; class II – atypical cells without malignancy; class IIIa – atypical cells highly suspected of being benign; class IIIb – atypical cells highly suspected of being malignant; class IV - malignant

Reference	Kimoto, 1999[187]
	Inadequate category: malignant 2, benign 1 *FNAC rated IIIb or higher (+ve) [I-IIIa taken as -ve result]* TP: 39 FN: 5 FP: 4 TN: 13 ; *sensitivity*: 0.886, *specificity*: 0.765 Note that insufficient aspirates were included in the analysis in this review as -ve cytological findings, but not included in the analysis in the paper (though details of the histopathology for them was given)
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, so some bias possible in who was given surgery)*
Comments	

Reference	Kini, 1985[188]
Study type	Retrospective/prospective
Number of patients	n = 379 nodules
Patient characteristics	*Age, mean (SD): not reported for those having surgery* *Gender (female to male ratio): not reported for those having surgery* *Ethnicity*: not reported *Setting*: Secondary Care *Country*: USA *Inclusion criteria*: Patients with thyroid nodules undergoing FNA and subsequent surgery *Exclusion criteria*: None reported *Stratum (prior US assessment / no prior US assessment): No prior US reported* *Sub-group (US-guided / not US guided): No USG used*
Target condition(s)	Thyroid nodule malignancy

Reference	Kini, 1985[188]
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=99 ;benign=280 Inadequate category: not reported FNAC classification: nodular goitre, follicular adenoma, suspicious for follicular carcinoma, carcinoma [incorporating follicular carcinoma, suspicious for follicular variant papillary carcinoma, follicular variant papillary carcinoma] *FNAC rated follicular adenoma, suspicious for follicular carcinoma, carcinoma (+ve) [benign taken as nodular goitre]* TP: 93 FN: 6 FP: 179 TN: 101 ; *sensitivity*:0.939, *specificity*: 0.361 *FNAC rated suspicious for follicular carcinoma, carcinoma (+ve) [follicular adenoma, benign taken as nodular goitre]* TP: 64 FN: 35 FP: 50 TN: 230 ; *sensitivity*: 0.646, *specificity*: 0.821 *FNAC rated carcinoma (+ve) [suspicious for follicular carcinoma, follicular adenoma, benign taken as nodular goitre]* TP: 53 FN: 46 FP: 15 TN: 265 ; *sensitivity*: 0.535, *specificity*: 0.946
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, so some bias possible in who was given surgery)*
Comments	

Reference	Kojic Katovic, 2004[193]
Study type	Retrospective
Number of patients	n = 80 nodules
Patient characteristics	*Age, range: 12-73* *Gender (female to male ratio): 73:7* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Croatia *Inclusion criteria*: Patients with complete pre-operative investigations for thyroid nodules (US, IS, FNA) and subsequent histopathological diagnosis *Exclusion criteria*: None reported *Stratum (prior US assessment / no prior US assessment): prior US, and looks as though US was used as a filter (226 nodules given US and 185 nodules given FNAC) but details unclear* *Sub-group (US-guided / not US guided): USG used*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Kojic Katovic, 2004[193]
Results	Gold standard results: malignant=30 ;benign=71 FNAC classification: Goitre, follicular tumour, hurthle tumour, carcinoma [incorporating papillary, follicular, medullary and differentiated carcinoma] Inadequate category: not reported *FNAC rated follicular tumour, hurthle tumour, carcinoma (+ve) [goitre taken as -ve result]* TP: 30 FN: 0 FP: 56 TN: 15 ; *sensitivity*: 1.0, *specificity*: 0.211 *FNAC rated follicular tumour, carcinoma (+ve) [hurthle tumour, goitre taken as -ve result]* TP: 29 FN: 1 FP: 54 TN: 17 ; *sensitivity*: 0.967, *specificity*: 0.239 *FNAC rated carcinoma (+ve) [follicular tumour, hurthle tumour, goitre taken as -ve result]* TP: 24 FN: 6 FP: 9 TN: 62 ; *sensitivity*: 0.80 , *specificity*: 0.873 Note: results extracted from 2 separate tables in paper (1 and 2).
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, so some bias possible in who was given surgery)*
Comments	

Reference	Kolendorf, 1975[194]
Study type	Retrospective
Number of patients	n = 20 nodules
Patient characteristics	*Age, mean (SD): not reported for those having surgery* *Gender (female to male ratio): not reported for those having surgery* *Ethnicity*: not reported *Setting*: Surgical Department *Country*: Denmark *Inclusion criteria*: Patients admitted for thyroid disorders, given FNA and open surgical biopsy

Reference	Kolendorf, 1975[194]
	Exclusion criteria: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US* *Sub-group (US-guided / not US guided): USG not used*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=2 ;benign=18 FNAC classification: No signs of malignancy, malignancy suspected, malignant Inadequate category: not reported *FNAC rated malignancy suspected or malignant (+ve) [no signs taken as -ve result]* TP: 0 FN: 2 FP: 3 TN: 15 ; *sensitivity*: 0.00, *specificity*: 0.833 *FNAC rated malignant (+ve) [malignancy suspected or no signs taken as -ve result]* TP: 0 FN: 2 FP: 0 TN: 18 ; *sensitivity*: 0.00, *specificity*: 1.0
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, so some bias possible in who was given surgery)*

Reference	Kolendorf, 1975[194]
Comments	

Reference	Kumar, 1992[199]
Study type	Retrospective
Number of patients	n = 88 nodules
Patient characteristics	*Age, mean (SD): not reported for those having surgery* *Gender (female to male ratio): not reported for those having surgery* *Ethnicity*: not reported *Setting*: Departments of endocrinology and metabolism *Country*: India *Inclusion criteria*: consecutive patients with solitary nodules undergoing FNA and subsequent surgery *Exclusion criteria*: not reported *Stratum (prior US assessment / no prior US assessment): prior US performed but did not appear to be an indication for FNA* *Sub-group (US-guided / not US guided): No USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block Aspirated material was expelled as droplets onto slides. Two or more slides were prepared. In case fluid was aspirated, it was centrifuged and slides prepared with cellular deposits *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	Kumar, 1992[199]
	Blinding of index test: No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 13 ;benign= 73 FNAC classification: unsatisfactory, cystic degeneration, adenomatous goitre, hyperplasia, follicular neoplasm, carcinomas Inadequate category: 0 malignant, 6 benign *FNAC rated follicular neoplasm, carcinomas (+ve) [cystic degeneration, adenomatous goitre, hyperplasia taken as -ve result]* TP: 12 FN: 1 FP: 21 TN: 52 ; *sensitivity*: 0.923, *specificity*: 0.712 *FNAC rated carcinomas (+ve) [follicular neoplasm, cystic degeneration, adenomatous goitre, hyperplasia taken as -ve result]* TP: 8 FN: 5 FP: 7 TN: 66 ; *sensitivity*: 0.615, *specificity*: 0.904
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, so some bias possible in who was given surgery)*
Comments	

Reference	Liu, 2009[211]
Study type	Retrospective
Number of patients	n = 40 patients with 40 nodules
Patient characteristics	*Age, mean (SD): 43.7 (11.4)* *Gender (female to male ratio): 37:3* *Ethnicity*: not reported *Setting*: Secondary care *Country*: Taiwan

Reference	Liu, 2009[211]
	Inclusion criteria: Patients with auto-immune thyroiditis; hypothyroidism or hyperthyroidism with thyroid nodules; given FNAC with subsequent surgery *Exclusion criteria*: Diffuse thyroid disorders *Stratum (prior US assessment / no prior US assessment): prior US not reported as an indicator for FNA* *Sub-group (US-guided / not US guided): USG not used (unclear)*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology with ROSE, with smear only All smears were interpreted within 3 minutes of their presentation. An unsatisfactory smear led to a repeat FNA *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 24 ;benign=16 FNAC classification: non-diagnostic, benign, malignant (included indeterminate) Inadequate category: malignant 1, benign 2 *FNAC rated indeterminate or malignant (+ve) [benign taken as -ve result]* TP: 22 FN: 2 FP: 6 TN: 10 ; *sensitivity*: 0.917, *specificity*: 0.625
Source of funding	No funding stated

Reference	Liu, 2009[211]
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Mamoon, 1997[221]
Study type	Retrospective
Number of patients	n = 176 nodules
Patient characteristics	*Age, mean (SD): not reported for those with surgery* *Gender (female to male ratio): not reported for those with surgery* *Ethnicity*: not reported *Setting*: Army medical college *Country*: Pakistan *Inclusion criteria*: Patients undergoing FNA and subsequent surgery for thyroid nodules *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only Several passes were made on each aspiration. 2 -4 smears were made in each case. Cytospin and cell block preparations were not made routinely. *Reference (gold) standard:* Surgical histopathological findings

Reference	Mamoon, 1997[221]
	Time between measurement of index test and reference standard: Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=14 ;benign=162 FNAC classification: negative, suspicious, follicular neoplasm, positive Inadequate category: not reported *FNAC rated positive or follicular neoplasm or suspicious (+ve) [negative taken as -ve result]* TP: 13 FN: 1 FP: 16 TN: 146 ; *sensitivity*: 0.929, *specificity*: 0.901 *FNAC rated positive or suspicious (+ve) [negative or follicular neoplasm taken as -ve result]* TP: 11 FN: 3 FP: 8 TN: 154 ; *sensitivity*: 0.786, *specificity*: 0.951 *FNAC rated positive (+ve) [negative or follicular neoplasm or suspicious taken as -ve result]* TP: 6 FN: 8 FP: 2 TN: 160 ; *sensitivity*: 0.429, *specificity*: 0.988
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, so some bias possible in who was given surgery)*
Comments	

Reference	Mandal, 2011[223]
Study type	Prospective
Number of patients	n = 108 nodules
Patient characteristics	*Age, range: 15-71* *Gender (female to male ratio): 5:1* *Ethnicity*: not reported

Reference	Mandal, 2011[223]
	Setting: University Hospital *Country*: India *Inclusion criteria*: Patients with nodular thyroid disease given FNAC followed by surgery *Exclusion criteria*: Diffuse goitre, debilitated elderly, other comorbidities making the patient unfit for surgery *Stratum (prior US assessment / no prior US assessment): no prior US* *Sub-group (US-guided / not US guided): not USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only. At least 2 air-dried and 2 fixed smears made. Repetition of aspiration was done where the first aspiration was inadequate. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=30 ; benign=78 FNAC classification: BTA classification - THY1 non diagnostic, THY1 cyst, THY2 non-neoplastic, THY3 follicular lesion, suspected follicular neoplasm, THY4 suspicious but non diagnostic of malignancy, THY5 diagnostic of malignancy Inadequate category: not reported *FNAC rated suspicious (THY3/4) or malignant (THY 5) (+ve) [THY 2 taken as -ve result]*

Reference	Mandal, 2011[223]
	TP: 27 FN: 3 FP: 12 TN: 66 ; *sensitivity*: 0.90, *specificity*: 0.846 *FNAC rated malignant (THY 5) (+ve) [suspicious (THY3/4) or THY 2 taken as -ve result]* TP: 18 FN: 12 FP: 0 TN: 78 ; *sensitivity*: 0.60, *specificity*: 1.0
Source of funding	No funding stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): none*
Comments	

Reference	Mandreker, 1995[224]
Study type	Retrospective
Number of patients	n = 238 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Teaching Hospital *Country*: India *Inclusion criteria*: Patients presenting with a diffuse or nodular thyroid enlargement and solitary thyroid nodule; FNA and subsequent surgery carried out *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported to be used*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only

Reference	Mandreker, 1995[224]
	Reference (gold) standard: Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=31 ;benign=207 Inadequate category: malignant 1, benign 24 *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 21 FN: 10 FP: 53 TN: 154 ; *sensitivity*: 0.677, *specificity*: 0.744 *FNAC rated malignant (+ve) [suspicious or benign taken as -ve result]* TP: 11 FN: 20 FP: 25 TN: 182 ; *sensitivity*: 0.355, *specificity*: 0.879
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, so some bias possible in who was given surgery)*
Comments	

Reference	Maruta, 2003[226]
Study type	Retrospective
Number of patients	n = 304 nodules
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported*

Reference	Maruta, 2003[226]
	Ethnicity: not reported *Setting*: Department of Pathology *Country*: Japan *Inclusion criteria*: thyroid nodule spirations from a database where people has also had thyroid surgery *Exclusion criteria*: not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided):* USG
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 148 ;benign=156 FNAC classification: Benign, inadequate, malignant Inadequate category: malignant 28, benign 25 *FNAC rated malignant (+ve) [benign taken as -ve result]*

Reference	Maruta, 2003[226]
	TP: 112 FN: 36 FP: 28 TN: 128 ; *sensitivity*: 0.757, *specificity*: 0.821
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, so some bias possible in who was given surgery)*
Comments	

Reference	Mastorakis, 2014[229]
Study type	Retrospective/prospective
Number of patients	n = 500 + 500 nodules, from 2 centres
Patient characteristics	*Age, median (range): Gp A: 47.4(13-85; Gp B: 48.6 (12-83)* *Gender (female to male ratio): Gp A: 395:105; Gp B: 359:141* *Ethnicity*: not reported *Setting*: Two settings: large regional hospital in Crete and University Hospital in Athens *Country*: Greece *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery; surgery given on basis of FNA results but also regardless of cytology – upon basis of other criteria such as multinodular lesions, nodule size or a lack of response to treatment or patient decision. *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG used*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test*

Reference	Mastorakis, 2014[229]
	Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block Used ThinPrep method proprietary fixative and haemolytic cytolyt solution. Used a 21-guage needle which maximizes yield and offers possibility of cell block as supplement to ThinPrep, whereas the haemolysis provided by cytolyt offers a better quality material, unobscured by red cells. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Group A Gold standard results: malignant= 211; benign=289 FNAC classification: TBSRTC (Bethesda): ND/UNS, Benign, AUS/FLUS, FN/SFN, SFM, Malignant. Inadequate category: malignant 5, benign 10 *FNAC rated AUS/FLUS, FN/SFN, SFM, Malignant (+ve) [benign taken as -ve result]* *TP: 197 FN: 14 FP: 53 TN: 236 ; sensitivity: 0.934, specificity:0.817* *FNAC rated FN/SFN, SFM, Malignant (+ve) [AUS/FLUS, benign taken as -ve result]* *TP: 186 FN: 25 FP: 17 TN: 272 ; sensitivity: 0.882, specificity:0.941* *FNAC rated SFM, Malignant (+ve) [FN/SFN, AUS/FLUS, benign taken as -ve result]* *TP: 184 FN: 27 FP: 13 TN: 276 ; sensitivity: 0.872, specificity:0.955* Group B Gold standard results: malignant= 81; benign=419

Reference	Mastorakis, 2014[229]
	FNAC classification: TBSRTC (Bethesda): ND/UNS, Benign, AUS/FLUS, FN/SFN, SFM, Malignant. Inadequate category: malignant 1, benign 25 *FNAC rated AUS/FLUS, FN/SFN, SFM, Malignant (+ve) [benign taken as -ve result]* *TP: 77 FN: 4 FP: 61 TN: 358 ; sensitivity: 0.951, specificity:0.854* *FNAC rated FN/SFN, SFM, Malignant (+ve) [AUS/FLUS, benign taken as -ve result]* *TP: 75 FN: 6 FP: 38 TN: 381 ; sensitivity: 0.926, specificity:0.909* *FNAC rated SFM, Malignant (+ve) [FN/SFN, AUS/FLUS, benign taken as -ve result]* *TP: 75 FN: 6 FP: 27 TN: 392 ; sensitivity: 0.926, specificity:0.936*
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, so some bias possible in who was given surgery)*
Comments	

Reference	McElroy, 2014[233]
Study type	Retrospective
Number of patients	n = 28 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Department of pathology *Country*: USA *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported

Reference	McElroy, 2014[233]
	Stratum (prior US assessment / no prior US assessment): no prior US reported *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block Cytology cases included direct smear slides, but most cases also included one low cellular or acellular cell-block *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Using older system of FNA grading (2006) Gold standard results: malignant=12 ;benign=16 FNAC classification: unsatisfactory, benign, atypia, follicular lesion, follicular neoplasm, suspicious, malignant Inadequate category: malignant 1, benign 2 *FNAC rated atypia, follicular lesion, follicular neoplasm, suspicious, malignant (+ve) [benign taken as -ve result]* TP: 9 FN: 3 FP: 9 TN:7 ; *sensitivity*: 0.75, *specificity*: 0.438 Using Bethesda grading (regraded data from 2006) Gold standard results: malignant=12 ;benign=16

Reference	McElroy, 2014[233]
	FNAC classification: Bethesda Inadequate category: malignant 3, benign 4 *FNAC rated AUS/FLUS, FN/SFN, suspicious, malignant (+ve) [benign taken as -ve result]* TP: 9 FN: 3 FP: 6 TN:10 ; *sensitivity*: 0.75, *specificity*: 0.625
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, so some bias possible in who was given surgery)*
Comments	

Reference	Mehrotra, 2006[236]
Study type	Retrospective
Number of patients	n = 450 nodules (348 freehand and 102 USG)
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Secondary care *Country*: UK *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG for 102; no USG for 348*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test*

Reference	Mehrotra, 2006[236]
	Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Using NO USG (freehand) Gold standard results: malignant=61 ;benign=234 FNAC classification: AC0/1: unsatisfactory, AC2: non-neoplastic, AC3: equivocal, often a follicular lesion, AC4: suspicious of malignancy, AC5: diagnostic of malignancy Inadequate category: malignant 10, benign 74 *FNAC rated AC3, AC4/5 (+ve) [AC2 taken as -ve result]* TP: 48 FN: 13 FP: 167 TN:67; *sensitivity*: 0.787, *specificity*:0.286 *FNAC rated AC4/5 (+ve) [AC2 or AC3, taken as -ve result]* TP: 25 FN: 36 FP: 80 TN: 154; *sensitivity*: 0.410, *specificity*: 0.658 Using USG Gold standard results: malignant=25 ;benign=68 FNAC classification: AC0/1: unsatisfactory, AC2: non-neoplastic, AC3: equivocal, often a follicular lesion, AC4: suspicious of malignancy, AC5: diagnostic of malignancy Inadequate category: malignant 3, benign 9

Reference	Mehrotra, 2006[236]
	FNAC rated AC3, AC4/5 (+ve) [AC2 taken as -ve result] TP: 20 FN: 5 FP: 55 TN:13; *sensitivity*: 0.80, *specificity*:0.191 *FNAC rated AC4/5 (+ve) [AC2 or AC3, taken as -ve result]* TP: 10 FN: 15 FP: 12 TN: 56; *sensitivity*: 0.40, *specificity*: 0.823
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, so some bias possible in who was given surgery)*
Comments	

Reference	Meko, 1995[237]
Study type	Retrospective/prospective
Number of patients	n = 90 nodules
Patient characteristics	*Age, mean (range): 49 (15-86)* *Gender (female to male ratio): 79:11* *Ethnicity*: not reported *Setting*: Department of Surgery *Country*: USA *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test*

Reference	Meko, 1995[237]
	Fine needle aspiration cytology <u>with ROSE</u>, with smear + cytospin and cell block Note does not mention cell-block. <u>*Reference (gold) standard:*</u> Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=19 ;benign=71 FNAC classification: unsatisfactory, benign, suspicious, malignant <u>Inadequate category: malignant 1, benign 2</u> *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 13 FN: 6 FP: 32 TN: 39 ; *sensitivity*: 0.684, *specificity*: 0.549
Source of funding	<u>No funding stated</u>
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Merchant, 1995[239]
Study type	Retrospective
Number of patients	n = 56 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery*

Reference	Merchant, 1995[239]
	Ethnicity: not reported *Setting*: District General Hospital *Country*: UK *Inclusion criteria*: Patients with thyroid nodules or diffuse thyroid enlargement given FNAC and subsequent surgery; surgery given secondary to cytology, clinical signs or evidence from second line investigations. *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG used if nodule not palpable but numbers not given.*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=11 ;benign=45 FNAC classification: Insufficient, benign, suspicious, neoplasm Inadequate category: malignant 1, benign 6

Reference	Merchant, 1995[239]
	FNAC rated suspicious or neoplasm (+ve) [benign taken as -ve result] TP: 8 FN: 3 FP: 11 TN: 34 ; *sensitivity*: 0.727, *specificity*: 0.756 *FNAC rated neoplasm (+ve) [suspicious or benign taken as -ve result]* TP: 5 FN: 6 FP: 8 TN: 37 ; *sensitivity*: 0.455, *specificity*: 0.822
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, so some bias possible in who was given surgery)*
Comments	

Reference	Mikosch, 2000[241]
Study type	Retrospective
Number of patients	n = 708 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Outpatients *Country*: Austria *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery; FNA indicated by patients with hypoechoicity, irregular margins. microcalcifications US, growth of the nodule during follow up or hypofunctional nodules on scintiscan; reasons for surgery included cytological findings or obstructive reasons *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): prior US used to determine eligibility* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy

Reference	Mikosch, 2000[241]
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 77 ;benign=631 FNAC classification: inadequate, non-malignant, non-malignant follicular proliferation, suspicious for malignancy, malignant Inadequate category: malignant 3, benign 31 *FNAC rated non-malignant follicular proliferation, suspicious for malignancy, malignant (+ve) [non-malignant taken as -ve result]* *TP: 71 FN: 6 FP: 331 TN: 300 ; sensitivity: 0.922, specificity: 0.475* *FNAC rated suspicious for malignancy, malignant (+ve) [non-malignant follicular proliferation, non-malignant taken as -ve result]* *TP: 65 FN: 12 FP: 160 TN: 471 ; sensitivity: 0.844, specificity: 0.746* *FNAC rated malignant (+ve) [suspicious for malignancy, non-malignant follicular proliferation, non-malignant taken as -ve result]* *TP: 54 FN: 23 FP: 38 TN: 593 ; sensitivity: 0.701, specificity: 0.940*
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, so some bias possible in who was given surgery)*
Comments	

Reference	Miller, 1979[242]
Study type	Retrospective
Number of patients	n = 147 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Department of Endocrinology *Country*: USA *Inclusion criteria*: Patients with discrete thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Functional nodules and cystic nodules without appreciable residual after aspiration of fluid *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Miller, 1979[242]
Results	Gold standard results: malignant=45 ;benign=102 FNAC classification: low risk of malignancy, intermediate risk, high risk Inadequate category: not reported *FNAC rated intermediate risk or high risk (+ve) [low risk taken as -ve result]* TP: 43 FN: 2 FP: 54 TN: 48 ; *sensitivity*: 0.956, *specificity*: 0.471 *FNAC rated high risk (+ve) [intermediate risk or low risk taken as -ve result]* TP: 35 FN: 10 FP: 20 TN: 82 ; *sensitivity*: 0.778, *specificity*: 0.804
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, so some bias possible in who was given surgery)*
Comments	

Reference	Munn, 1988 #1322[252]
Study type	Retrospective
Number of patients	n = 49 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Teaching Hospital *Country*: USA *Inclusion criteria*: Patients with palpable thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: History of radiation exposure; family history of medullary carcinoma *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*

Reference	Munn, 1988 #1322[252]
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only Note- core biopsy evaluated in a small sub-set within this study, but unable to include in review as poorly reported – unclear how many had surgery and whether the diagnostic accuracy data are based on surgery as a gold standard. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=16 ;benign=33 FNAC classification: Benign (benign nodular goitre, thyroiditis), Follicular neoplasm, Carcinoma (including lymphoma, PC, medullary carcinoma, metastatic carcinoma) No data given for inadequate samples *FNAC rated follicular neoplasm or carcinoma (+ve) [benign taken as -ve result]* TP: 14 FN: 2 FP: 21 TN: 12 ; *sensitivity*: 0.875, *specificity*: 0.364 *FNAC rated carcinoma (+ve) [follicular neoplasm or benign taken as -ve result]* TP: 12 FN: 4 FP: 3 TN: 30 ; *sensitivity*: 0.75, *specificity*: 0.909
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, so some bias possible in who was given surgery)*
Comments	

Reference	Nagarajan, 2015 #1326[256]
Study type	Retrospective
Number of patients	n = 1272 nodules (for standard smear) and 54 (for liquid based preparation)
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Departments of Surgery and Pathology *Country*: USA *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only AND Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block. It is assumed that this is equivalent to liquid based preparation. *Reference (gold) standard:* Surgical histopathological findings

Reference	Nagarajan, 2015 #1326[256]
	Time between measurement of index test and reference standard: Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Conventional smear Gold standard results: malignant=467 ;benign=805 FNAC classification: Bethesda I-VI scale Inadequate category: malignant 8, benign 101 *FNAC rated III-VI (+ve) [II (benign) taken as -ve result]* TP: 438 FN: 29 FP: 345 TN: 460 ; *sensitivity*: 0.938, *specificity*: 0.571 *FNAC rated IV-VI (+ve) [II (benign)-III taken as -ve result]* TP: 354 FN: 113 FP: 205 TN: 600 ; *sensitivity*: 0.758, *specificity*: 0.745 *FNAC rated V-VI (+ve) [II (benign)-IV taken as -ve result]* TP: 321 FN: 146 FP: 122 TN: 683 ; *sensitivity*: 0.687, *specificity*: 0.848 *FNAC rated VI (+ve) [II (benign)-V taken as -ve result]* TP: 242 FN: 225 FP: 103 TN: 702 ; *sensitivity*: 0.518, *specificity*: 0.872 Liquid based preparation Gold standard results: malignant=26 ;benign=28 FNAC classification: Bethesda I-VI scale Inadequate category: malignant 0, benign 2 *FNAC rated III-VI (+ve) [II (benign) taken as -ve result]*

Reference	Nagarajan, 2015 #1326[256]
	TP: 25 FN: 1 FP: 15 TN: 13 ; *sensitivity*: 0.962, *specificity*: 0.464 *FNAC rated IV-VI (+ve) [II (benign)-III taken as -ve result]* TP: 21 FN: 5 FP: 4 TN: 24 ; *sensitivity*: 0.808, *specificity*: 0.857 *FNAC rated V-VI (+ve) [II (benign)-IV taken as -ve result]* TP: 17 FN: 9 FP: 3 TN: 25 ; *sensitivity*: 0.654, *specificity*: 0.893 *FNAC rated VI (+ve) [II (benign)-V taken as -ve result]* TP: 12 FN: 14 FP: 2 TN: 26 ; *sensitivity*: 0.462, *specificity*: 0.929
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, so some bias possible in who was given surgery)*
Comments	

Reference	Natarajan, 1994[258]
Study type	Retrospective
Number of patients	n = 25 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Teaching Hospital *Country*: India *Inclusion criteria*: Patients with solitary cold thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported*

Reference	Natarajan, 1994[258]
	Sub-group (US-guided / not US guided): USG not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 13;benign=12 FNAC classification: non-neoplastic (colloid goitre, Hashimoto's thyroiditis), equivocal (suspected extrathyroidal malignancy, suspected neoplasm), malignant (medullary, anaplastic, follicular or papillary tumour) No data given for inadequate samples *FNAC rated equivocal or malignant (+ve) [non-neoplastic taken as -ve result]* TP: 13 FN: 0 FP: 5 TN: 7 ; *sensitivity*: 1.0, *specificity*: 0.583 *FNAC rated malignant (+ve) [equivocal or non-neoplastic taken as -ve result]* TP: 11 FN: 2 FP: 0 TN: 12 ; *sensitivity*: 0.846, *specificity*: 1.0
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, so some bias possible in who was given surgery)*
Comments	

Reference	Ng, 1988 #1330[261]
Study type	Retrospective
Number of patients	n = 46 nodules
Patient characteristics	*Age, mean (SD): 39.4 (14.9)* *Gender (female to male ratio): 5.2:1* *Ethnicity*: not reported *Setting*: General Hospital *Country*: Singapore *Inclusion criteria*: Patients with solitary thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Ng, 1988 #1330[261]
Results	Gold standard results: malignant=10 ;benign=36 FNAC classification: benign, suspicious, malignant, inadequate Inadequate category: malignant 0, benign 4 *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 6 FN: 4 FP: 4 TN: 32 ; *sensitivity*: 0.6, *specificity*: 0.889
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, so some bias possible in who was given surgery)*
Comments	

Reference	Ongphiphadhanakul, 1992 #1335[267]
Study type	Retrospective/prospective
Number of patients	n = 129 nodules
Patient characteristics	*Age, mean (SD): 40.7(1.2)* *Gender (female to male ratio): 105:24* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Thailand *Inclusion criteria*: Patients with solitary thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*

Reference	Ongphiphadhanakul, 1992 #1335[267]
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=28 ;benign=101 FNAC classification: malignant, suspected malignant, benign No data given for inadequate samples *FNAC rated suspected or malignant (+ve) [benign taken as -ve result]* TP: 20 FN: 8 FP: 15 TN: 86 ; *sensitivity*: 0.714, *specificity*: 0.851 *FNAC rated malignant (+ve) [suspected or benign taken as -ve result]* TP: 14 FN: 14 FP: 4 TN: 97 ; *sensitivity*: 0.5, *specificity*: 0.960
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, so some bias possible in who was given surgery)*
Comments	

Reference	Ozdemir, 2017[269]
Study type	Retrospective/prospective
Number of patients	n = 1810 nodules (pre Bethesda) and 5115 nodules (post-Bethesda)

Reference	Ozdemir, 2017[269]
Patient characteristics	*Age, mean (SD): 51.98(12.07) pre-Bethesda; 49.46 (11.98) post-Bethesda* *Gender (female to male ratio): 78.6:21.4 pre-Bethesda; 77.8:22.2* *Ethnicity*: not reported *Setting*: Department of Endocrinology *Country*: Turkey *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Age <16 years; previous history of thyroid surgery or percutaneous invasive procedures to thyroid nodules; radiotherapy to head and neck *Stratum (prior US assessment / no prior US assessment): prior US reported – only nodules >1cm OR <1cm with one or more suspicious US features were given FNA* *Sub-group (US-guided / not US guided): USG used*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Ozdemir, 2017[269]
Results	PRE-BETHESDA Gold standard results: malignant=193 ;benign=1617 FNAC classification: Non-diagnostic, benign, indeterminate, suspicious, malignant Inadequate category: malignant 27, benign 313 *FNAC rated indeterminate, suspicious, malignant (+ve) [benign taken as -ve result]* TP: 131 FN: 62 FP: 488 TN: 1129 ; *sensitivity*: 0.679, *specificity*: 0.698 *FNAC rated suspicious, malignant (+ve) [indeterminate or benign taken as -ve result]* TP: 89 FN: 104 FP: 336 TN: 1281 ; *sensitivity*: 0.461, *specificity*: 0.792 POST-BETHESDA Gold standard results: malignant=466 ;benign=4649 FNAC classification: Bethesda - ND, Benign, AUS/FLUS, FN/SFN, SFM, Malignant (I-VI) Inadequate category: malignant 66, benign 1274 *FNAC rated AUS/FLUS, FN/SFN, SFM, Malignant (+ve) [benign taken as -ve result]* TP: 339 FN: 127 FP: 1899 TN: 2750 ; *sensitivity*: 0.727, *specificity*: 0.592 *FNAC rated FN/SFN, SFM, Malignant (+ve) [AUS/FLUS, benign taken as -ve result]* TP: 223 FN: 243 FP: 1358 TN: 3291 ; *sensitivity*: 0.479, *specificity*: 0.708 *FNAC rated SFM, Malignant (+ve) [FN/SFN, AUS/FLUS, benign taken as -ve result]* TP: 204 FN: 262 FP: 1311 TN: 3338 ; *sensitivity*: 0.438, *specificity*: 0.718 *FNAC rated Malignant (+ve) [SFM, FN/SFN, AUS/FLUS, benign taken as -ve result]* TP: 116 FN: 350 FP: 1280 TN: 3369 ; *sensitivity*: 0.249, *specificity*: 0.725
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, so some bias possible in who was given surgery)*
Comments	

Reference	Pepper, 1989[275]
Study type	Retrospective
Number of patients	n = 21 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Teaching Hospital *Country*: USA *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery; surgery given because of FNA findings or because of personal choice or because of nodule growth despite levothyroxine treatment *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): prior US reported but did not appear to be used to define who should have FNA* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block - *In the event that a cystic lesion was entered, all the fluid was drained and placed into alcohol. Smears were obtained from the sediment obtained by centrifugation.* *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No

Reference	Pepper, 1989[275]
	Blinding of gold standard test: No
Results	Gold standard results: malignant= 6 ;benign= 15 No data given for inadequate samples FNAC classification: malignant, suspicious (numerous follicular cells with clear nuclear intrusions; oxyphilic cells without lymphocytic thyroiditis; psammoma antibodies; atypical follicular cells; papillary clusters of follicular cells; hypercellularity) and benign Inadequate category: not reported *FNAC rated malignant or suspicious (+ve) [benign taken as -ve result]* TP: 5 FN: 1 FP: 8 TN: 7 ; *sensitivity*: 0.833, *specificity*: 0.467
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, so some bias possible in who was given surgery)*
Comments	

Reference	Petersen, 1984[276]
Study type	Retrospective
Number of patients	n = 189 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Secondary Care *Country*: Denmark *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery

Reference	Petersen, 1984[276]
	Exclusion criteria: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block Fluid from cells is fixed in parts with alcohol and centrifuged. The sediment is spread out on a glass slide and stained. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=21 ;benign=168 FNAC classification: Neoplasia, benign (cyst/diffuse benign lesion), inconclusive Inadequate category: malignant 1, benign 40 *FNAC rated neoplasia (+ve) [benign taken as -ve result]* TP: 19 FN: 2 FP: 84 TN: 84 ; *sensitivity*: 0.905, *specificity*: 0.50
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, so some bias possible in who was given surgery)*

Reference	Petersen, 1984[276]
Comments	

Reference	Piana, 2011[277]
Study type	Retrospective
Number of patients	n = 2047 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Department of Pathology *Country*: Italy *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): prior US used to select patients for FNA* *Sub-group (US-guided / not US guided): USG used*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No

Reference	Piana, 2011[277]
	Blinding of gold standard test: No
Results	Gold standard results: malignant=840 ;benign=1207 FNAC classification: C1-C5: C1=non diagnostic, C2=benign, C3=indeterminate, C4=suspicious, C5=malignant Inadequate category: malignant 23, benign 73 *FNAC rated C3-C5 (+ve) [benign (C2) taken as -ve result]* TP: 743 FN: 97 FP: 607 TN: 600 ; *sensitivity*:0.885, *specificity*: 0.497 *FNAC rated C4-C5 (+ve) [C3 and benign taken as -ve result]* TP: 555 FN: 285 FP: 84 TN: 1123 ; *sensitivity*:0.661, *specificity*: 0.930 *FNAC rated C5 (+ve) [C3, C4 and benign taken as -ve result]* TP: 415 FN: 425 FP: 73 TN: 1134 ; *sensitivity*: 0.494, *specificity*: 0.939
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, so some bias possible in who was given surgery)*
Comments	

Reference	Pisani, 2000[278]
Study type	Retrospective
Number of patients	n = 42 nodules (for FNA) and 29 nodules (for core biopsy)
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Italy

Reference	Pisani, 2000[278]
	Inclusion criteria: Consecutive patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG for both FNA and CNB*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only Core biopsy *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	FNA Gold standard results: malignant=13 ;benign=29 No inconclusive results FNAC classification: carcinoma, neoplasm, colloid goitre Inadequate category: Malignant 0, benign 0 *FNAC rated carcinoma or neoplasm (+ve) [colloid goitre taken as -ve result]* TP: 13 FN: 0 FP: 12 TN: 17 ; *sensitivity*:1.0, *specificity*: 0.586

Reference	Pisani, 2000[278]
	FNAC rated carcinoma (+ve) [colloid goitre or neoplasm taken as -ve result] TP: 10 FN: 3 FP: 0 TN: 29 ; *sensitivity*: 0.769, *specificity*: 1.0 CNB Gold standard results: malignant=9 ;benign=22 Inadequate category: Malignant 4, benign 10 FNAC classification: non-diagnostic, carcinoma, neoplasm, colloid goitre *FNAC rated carcinoma or neoplasm (+ve) [colloid goitre taken as -ve result]* TP: 5 FN: 4 FP: 13 TN: 9 ; *sensitivity*:0.556, *specificity*: 0.409 *FNAC rated carcinoma (+ve) [colloid goitre or neoplasm taken as -ve result]* TP: 3 FN: 6 FP: 10 TN: 12 ; *sensitivity*:0.333, *specificity*: 0.545
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, so some bias possible in who was given surgery)*
Comments	

Reference	Radetic, 1984[284]
Study type	Retrospective
Number of patients	n = 2190 nodules
Patient characteristics	*Age, mean: 45.7* *Gender (female to male ratio): 1975:215* *Ethnicity*: not reported *Setting*: General Hospital *Country*: Croatia (was Yugoslavia at time of paper)

Reference	Radetic, 1984[284]
	Inclusion criteria: Patients with thyroid goitres given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=248; benign=1942 FNAC classification: Negative, suspicious, positive Inadequate category: not reported *FNAC rated suspicious or positive (+ve) [negative taken as -ve result]* TP: 170 FN: 78 FP: 179 TN: 1763 ; *sensitivity*: 0.685, *specificity*: 0.908 *FNAC rated positive (+ve) [suspicious or negative taken as -ve result]* TP: 88 FN: 160 FP: 9 TN: 1933 ; *sensitivity*: 0.355, *specificity*: 0.995
Source of funding	No funding stated

Reference	Radetic, 1984[284]
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Rammeh, 2019 #1349[286]
Study type	Retrospective/prospective
Number of patients	n = 64 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Secondary care *Country*: Tunisia *Inclusion criteria*: Patients with palpable thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): No USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:*

Reference	Rammeh, 2019 #1349[286]
	Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=24 ;benign=40 FNAC classification: Bethesda I-VI Inadequate category: not reported *FNAC rated V or VI (+ve) [II to IV taken as -ve result (unclear if I included)]* TP: 20 FN: 4 FP: 6 TN: 34 ; *sensitivity*: 0.833, *specificity*: 0.85 :
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, so some bias possible in who was given surgery)*
Comments	

Reference	Rana, 2021[287]
Study type	Retrospective
Number of patients	n = 445 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: India

Reference	Rana, 2021[287]
	Inclusion criteria: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=105 ;benign=340 FNAC classification: Bethesda I-VI Non-diagnostic cases were expressly excluded by study authors and not included in analysis; insufficient information to impute them. *FNAC rated V or VI (+ve) [II to IV taken as -ve result]* TP: 89 FN: 16 FP: 3 TN:337 ; *sensitivity*: 0.847, *specificity*: 0.991 Note that the sensitivity and specificity data differ from those in the paper. The results given here reflect the numbers with histopathological malignancy and benign findings (table 4 in paper) and the raw FN and FP data provided by the paper. It was assumed that the probability of error in calculated results was greater than that in the raw data.
Source of funding	No funding stated

Reference	Rana, 2021[287]
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Rege, 1987[289]
Study type	Retrospective
Number of patients	n = 182 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Thyroid clinic *Country*: India *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear

Reference	Rege, 1987[289]
	Blinding of index test: No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=15 ;benign=170 FNAC classification: Benign, malignant (no further information provided) Out of the original 215 cases, 30 people provided non diagnostic/unsatisfactory samples. The histopathology of these people were not provided and so cannot be imputed into the analysis *FNAC rated malignant (+ve) [benign taken as -ve result]* TP: 13 FN: 2 FP: 0 TN: 170 ; *sensitivity*: 0.867, *specificity*: 1.0 Note: data unclearly reported in the paper and the data reported here is the best interpretation.
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, so some bias possible in who was given surgery)*
Comments	

Reference	Rodriguez, 1994[295]
Study type	Retrospective
Number of patients	n = 170 nodules
Patient characteristics	*Age, mean (SD): 41(3)* *Gender (female to male ratio): 154:16* *Ethnicity*: not reported *Setting*: General Surgery *Country*: Spain *Inclusion criteria*: Patients with solitary or dominant thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: inadequate samples

Reference	Rodriguez, 1994[295]
	Stratum (prior US assessment / no prior US assessment): no prior US reported *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=27 ;benign=143 FNAC classification: benign (colloid nodule), suspicious (follicular proliferation), malignant (medullary, papillary or follicular carcinoma) Non-diagnostic cytology was excluded by study authors and so we were unable to impute this in analysis *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 26 FN: 1 FP: 67 TN: 76 ; *sensitivity*: 0.963, *specificity*: 0.531 *FNAC rated malignant (+ve) [suspicious or benign taken as -ve result]* TP: 17 FN: 10 FP: 0 TN: 143 ; *sensitivity*: 0.630, *specificity*: 1.00
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, so some bias possible in who was given surgery)*

Reference	Rodriguez, 1994[295]
Comments	

Reference	Rosen, 1993[296]
Study type	Retrospective
Number of patients	n = 41 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Canada *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Rosen, 1993[296]
Results	Gold standard results: malignant=16 ;benign=25 FNAC classification: Inadequate, benign (cyst, colloid, thyroiditis), follicular lesion, cancer Inadequate aspirates: 1 malignant, 10 benign on histopathology. *FNAC rated follicular lesion or cancer (+ve) [benign taken as -ve result]* TP: 13 FN: 3 FP: 23 TN: 2 ; *sensitivity*:0.8125, *specificity*:0.08 *FNAC rated cancer (+ve) [follicular lesion or benign taken as -ve result]* TP: 9 FN: 7 FP: 10 TN: 15 ; *sensitivity*: 0.563, *specificity*: 0.60
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, so some bias possible in who was given surgery)*
Comments	

Reference	Rosen, 1981[298]
Study type	Retrospective
Number of patients	n = 153 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Surgery and Endocrinology *Country*: Canada *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported

Reference	Rosen, 1981[298]
	Stratum (prior US assessment / no prior US assessment): no prior US reported *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 40 ;benign= 113 FNAC classification: Inadequate, Benign (cyst, colloid or thyroiditis), adenoma, carcinoma Inadequate aspirates: 1 malignant, 8 benign on histopathology. *FNAC rated adenoma or carcinoma (+ve) [benign taken as -ve result]* TP: 34 FN: 6 FP: 87 TN: 26 ; *sensitivity*: 0.85, *specificity*: 0.230 *FNAC rated carcinoma (+ve) [adenoma or benign taken as -ve result]* TP: 16 FN: 24 FP: 10 TN: 103 ; *sensitivity*: 0.40, *specificity*: 0.911
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, so some bias possible in who was given surgery)*
Comments	

Reference	Rubenfeld, 1982[300]
Study type	Retrospective
Number of patients	n = 30 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Secondary care *Country*: USA *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology with ROSE, with smear + cytospin and cell block. If the nodule was cystic as much of the fluid as possible was aspirated as smears prepared after centrifugation and/or filtration. A biopsy was performed on any mass remaining after aspiration after a cystic lesion. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No

Reference	Rubenfeld, 1982[300]
	Blinding of gold standard test: No
Results	Gold standard results: malignant= 15;benign=15 FNAC classification: unsatisfactory, negative, suspicious (suggestive but not confirmatory of malignancy), positive. Inadequate samples included in the analysis in paper – as a negative cytoscopic finding; unable to use as WCS strategy as do not know the number of unsatisfactory (only that total number of benign and unsatisfactory = 4). *FNAC rated indeterminate or malignant (+ve) [benign/unsatisfactory taken as -ve result]* TP: 15 FN: 0 FP: 11 TN: 4; *sensitivity*: 1.0, *specificity*: 0.267
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, so some bias possible in who was given surgery)*
Comments	

Reference	Russ, 1978[301]
Study type	Retrospective
Number of patients	n = 29 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Secondary care *Country*: USA *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported*

Reference	Russ, 1978[301]
	Sub-group (US-guided / not US guided): USG not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 11 ;benign=18 FNAC classification: benign (including indeterminate such as adenoma), malignant (carcinoma) Inadequate samples not reported and so could not be imputed *FNAC rated malignant (+ve) [benign taken as -ve result]* TP: 8 FN: 3 FP: 0 TN: 18 ; *sensitivity*: 0.727, *specificity*: 1.0
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, so some bias possible in who was given surgery)*
Comments	

Reference	Schmid, 1986 #1370[307]
Study type	Retrospective/prospective
Number of patients	n = 2709 nodules

Reference	Schmid, 1986 #1370[307]
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Setting*: Institute of pathology *Country*: Austria *Inclusion criteria*: Patients with cold or multinodular thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Schmid, 1986 #1370[307]
Results	Gold standard results: malignant=357 ;benign=2352 FNAC classification: negative, suspect, positive, unsatisfactory Non-diagnostic findings: 17.7% overall but no breakdown given per histological findings *FNAC rated suspect or positive (+ve) [negative taken as -ve result]* TP: 302 FN: 55 FP: 499 TN: 1852 ; *sensitivity*: 0.846, *specificity*: 0.787 *FNAC rated positive (+ve) [suspect or negative taken as -ve result]* TP: 255 FN: 102 FP: 207 TN: 2145 ; *sensitivity*: 0.714, *specificity*: 0.912
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, so some bias possible in who was given surgery)*
Comments	

Reference	Schoedel, 2008 #1372[309]
Study type	Prospective
Number of patients	n = 46 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Department of pathology *Country*: USA *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported*

Reference	Schoedel, 2008 #1372[309]
	Sub-group (US-guided / not US guided): USG used
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only. Both capillary and aspiration methods were tested separately but results have been combined for this review. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 21 ;benign=25 FNAC classification: positive, suspected, atypical, follicular neoplasm, negative, non-diagnostic Non diagnostic findings: malignant 1, benign 3. *FNAC rated positive, suspected, atypical/follicular neoplasm (+ve) [negative taken as -ve result]* TP: 14 FN: 7 FP: 7 TN:18 ; *sensitivity*: 0.667, *specificity*: 0.720 *FNAC rated positive, suspected (+ve) [atypical/follicular neoplasm or negative taken as -ve result]* TP: 8 FN: 13 FP: 3 TN: 22 ; *sensitivity*: 0.381, *specificity*: 0.88
Source of funding	No funding stated
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): none*
Comments	

Reference	Schwartz, 1982 #1373[310]
Study type	Retrospective
Number of patients	n = 102 nodules
Patient characteristics	*Age, mean (range): 44(21-89)* *Gender (female to male ratio): 86:16* *Ethnicity*: not reported *Setting*: Head and neck service, secondary care *Country*: USA *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only [note: core biopsy also studied but data insufficient for analysis in this review] *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Schwartz, 1982 #1373[310]
Results	Gold standard results: malignant=11 ;benign=81 FNAC classification: malignant and benign Non-diagnostic findings: 10 patients but histologic findings not given so cannot be imputed *FNAC rated malignant (+ve) [benign taken as -ve result]* TP: 5 FN: 6 FP: 3 TN: 78 ; *sensitivity*:0/455, *specificity*: 0.963
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, so some bias possible in who was given surgery)*
Comments	

Reference	Scurry, 2000[312]
Study type	Retrospective
Number of patients	n = 109 nodules (standard smear), 92 nodules (cytospin)
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: secondary care *Country*: Australia and Canada *Inclusion criteria*: Patients with thyroid nodules given direct smear or smear/cytospin FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*

Reference	Scurry, 2000[312]
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only OR Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block [cell-block not mentioned]: cytospin preparations were made in cases that yielded cyst fluid. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Standard smear Gold standard results: malignant=37 ;benign=73 FNAC classification: negative, intermediate (includes follicular neoplasm and atypia), suspicious, malignant, non-diagnostic Non-diagnostic: 7 malignant, 33 benign *FNAC rated indeterminate, suspicious or malignant (+ve) [negative taken as -ve result]* TP: 23 FN: 14 FP: 60 TN:13 ; *sensitivity*:0.622, *specificity*: 0.178 *FNAC rated suspicious or malignant (+ve) [negative or indeterminate taken as -ve result]* TP: 10 FN: 27 FP: 36 TN:37 ; *sensitivity*:0.270, *specificity*: 0.507 Cytospin Gold standard results: malignant=32 ;benign=60

Reference	Scurry, 2000[312]
	FNAC classification: negative, intermediate (includes follicular neoplasm and atypia), suspicious, malignant, non-diagnostic Non-diagnostic: 6 malignant, 25 benign *FNAC rated indeterminate, suspicious or malignant (+ve) [negative taken as -ve result]* TP: 22 FN: 10 FP: 57 TN: 3 ; *sensitivity*:0.688, *specificity*: 0.005 *FNAC rated suspicious or malignant (+ve) [negative or indeterminate taken as -ve result]* TP: 10 FN: 22 FP: 28 TN:32 ; *sensitivity*:0.455, *specificity*: 0.533
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, so some bias possible in who was given surgery)*
Comments	

Reference	Settakorn, 2001[316]
Study type	Retrospective/prospective
Number of patients	n = 415 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Thailand *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported*

Reference	Settakorn, 2001[316]
	Sub-group (US-guided / not US guided): USG not reported
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=48 ;benign=182 FNAC classification: Benign (goitre, diffuse thyroid hyperplasia), suspicious (follicular or Hurthle cell neoplasm), malignant Non-diagnostic: 185 unsatisfactory, but histological details not given so cannot be imputed. Inclusion of these data would have changed results significantly. *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 44 FN: 4 FP: 28 TN: 154 ; *sensitivity*:0.917, *specificity*: 0.846 *FNAC rated malignant (+ve) [suspicious or benign taken as -ve result]* TP: 37 FN: 11 FP: 4 TN: 178 ; *sensitivity*:0.771, *specificity*: 0.978
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, so some bias possible in who was given surgery)*
Comments	

Reference	Silverman, 1986[327]
Study type	Retrospective
Number of patients	n = 8 nodules (FNA) and 4 nodules (CNB)
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: USA *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only AND Core biopsy *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No

Reference	Silverman, 1986[327]
	Blinding of gold standard test: No
Results	<u>FNA</u> Gold standard results: malignant=2 ;benign=6 FNAC classification: Benign (follicular adenoma, benign nodular goitre), malignant (papillary carcinoma, etc) <u>Non-diagnostic findings: malignant 0, benign 0</u> *FNAC rated malignant (+ve) [benign taken as -ve result]* TP: 1 FN: 1 FP: 0 TN: 6 ; *sensitivity*: 0.5, *specificity*: 1.0 <u>CB</u> Gold standard results: malignant=1 ;benign=3 FNAC classification: Benign (follicular adenoma, benign nodular goitre), malignant (papillary carcinoma, etc) <u>Non-diagnostic findings: malignant 1, benign 0</u> *FNAC rated malignant (+ve) [benign taken as -ve result]* TP: 0 FN: 1 FP: 0 TN: 3 ; *sensitivity*: 0.0, *specificity*: 1.0
Source of funding	<u>No funding stated</u>
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Sirpal, 1996[329]
Study type	Retrospective
Number of patients	n = 128 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery*

Reference	Sirpal, 1996[329]
	Ethnicity: not reported *Setting*: Army Hospital *Country*: India *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery. Surgery contemplated where FNA showed malignancy, follicular or HC tumour, cosmetically unacceptable cases, compression symptoms or cases non-responsive to therapy. *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=14 ;benign=114 FNAC classification: Benign (cystic degeneration, colloid/adenomatous goitre, Hashitoxicosis), suspicious (HCA, FN), malignant, unsatisfactory Non-diagnostic findings: 0 malignant, 4 benign

Reference	Sirpal, 1996[329]
	FNAC rated malignant or suspicious (+ve) [benign taken as -ve result] TP: 13 FN: 1 FP: 17 TN: 97 ; *sensitivity*: 0.929, *specificity*: 0.851 *FNAC rated malignant (+ve) [benign or suspicious taken as -ve result]* TP: 12 FN: 2 FP: 4 TN: 110 ; *sensitivity*: 0.857, *specificity*: 0.965
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, so some bias possible in who was given surgery)*
Comments	

Reference	Spiliotis, 1992 #1394[334]
Study type	Retrospective
Number of patients	n = 201 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Greece *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Toxic nodules *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy

Reference	Spiliotis, 1992 #1394[334]
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant= 31;benign=170 FNAC classification: benign, indeterminate, suspicious, malignant, unsatisfactory Non-diagnostic findings: 0 malignant, 10 benign *FNAC rated indeterminate, suspicious, malignant (+ve) [benign taken as -ve result]* TP: 28 FN: 3 FP: 42 TN: 128 ; *sensitivity*: 0.903, *specificity*: 0.753 *FNAC rated suspicious, malignant (+ve) [benign or indeterminate taken as -ve result]* TP: 25 FN: 6 FP: 30 TN: 140 ; *sensitivity*: 0.806, *specificity*: 0.824
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, so some bias possible in who was given surgery)*
Comments	

Reference	Tabain, 2004[342]
Study type	Retrospective
Number of patients	n = 457 nodules

Reference	Tabain, 2004[342]
Patient characteristics	*Age, mean (SD): 47.7 (13.2)* *Gender (female to male ratio): 378: 79* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Croatia *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Tabain, 2004[342]
Results	Gold standard results: malignant=93 ;benign=364 FNAC classification: Benign (nodular goitre, thyroiditis), indeterminate (cellular Follicular lesion, suspicious follicular neoplasm), malignant (unequivocal evidence of carcinoma), non-diagnostic Non-diagnostic findings: 0 malignant, 8 benign *FNAC rated malignant or indeterminate (+ve) [benign taken as -ve result]* TP: 92 FN: 1 FP: 158 TN: 206 ; *sensitivity*: 0.989, *specificity*: 0.566 *FNAC rated malignant (+ve) [benign or indeterminate taken as -ve result]* TP: 67 FN: 26 FP: 17 TN: 347 ; *sensitivity*: 0.720, *specificity*: 0.953
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, so some bias possible in who was given surgery)*
Comments	

Reference	Takashima, 1994[344]
Study type	Retrospective
Number of patients	n = 99 nodules (UG) and 34 nodules (palpation)
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Japan *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery

Reference	Takashima, 1994[344]
	Exclusion criteria: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG and no USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	USG-FNA Gold standard results: malignant= 67;benign=32 FNAC classification: malignant, benign Non-diagnostic findings: not reported for histologic categories so cannot be imputed *FNAC rated malignant (+ve) [benign taken as -ve result]* TP: 64 FN: 3 FP: 3 TN: 29 ; *sensitivity*: 0.955, *specificity*: 0.906 Non-USG-FNA Gold standard results: malignant= 24; benign=10 FNAC classification: malignant, benign Non-diagnostic findings: not reported for histologic categories so cannot be imputed

Reference	Takashima, 1994[344]
	FNAC rated malignant (+ve) [benign taken as -ve result] TP: 21 FN: 3 FP: 1 TN: 9 ; *sensitivity*: 0.875, *specificity*: 0.900
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, so some bias possible in who was given surgery)*
Comments	

Reference	Takashima, 1992[345]
Study type	Retrospective
Number of patients	n = 27 nodules (UG) and 14 nodules (palpation)
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Japan *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG and no USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only

Reference	Takashima, 1992[345]
	Reference (gold) standard: Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	USG-FNA Gold standard results: malignant= 16;benign=11 FNAC classification: malignant, benign Non-diagnostic findings: not reported for histologic categories so cannot be imputed *FNAC rated malignant (+ve) [benign taken as -ve result]* TP: 16 FN: 0 FP: 0 TN: 11 ; *sensitivity*: 1.0, *specificity*: 1.0 Non-USG-FNA Gold standard results: malignant= 8; benign=6 FNAC classification: malignant, benign Non-diagnostic findings: not reported for histologic categories so cannot be imputed *FNAC rated malignant (+ve) [benign taken as -ve result]* TP: 6 FN: 2 FP: 0 TN: 6 ; *sensitivity*: 0.75, *specificity*: 1.0
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, so some bias possible in who was given surgery)*
Comments	

Reference	Tal, 1992[347]
Study type	Retrospective
Number of patients	n = 30 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: General Hospital *Country*: USA *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No

Reference	Tal, 1992[347]
	Blinding of gold standard test: No
Results	Gold standard results: malignant=8 ;benign=22 FNAC classification: negative, suspicious (cells suggestive of malignancy, or Hurthle cells), positive, inadequate Non-diagnostic findings: not reported *FNAC rated positive or suspicious (+ve) [negative taken as -ve result]* TP: 7 FN: 1 FP: 5 TN: 17 ; *sensitivity*: 0.875, *specificity*: 0.773
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, so some bias possible in who was given surgery)*
Comments	

Reference	Theoharis, 2013 #1410[353]
Study type	Retrospective
Number of patients	n = 372 nodules (pre Bethesda) and 379 nodules (post Bethesda implementation)
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Department of Pathology *Country*: USA *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery

Reference	Theoharis, 2013 #1410[353]
	Exclusion criteria: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Pre-Bethesda Gold standard results: malignant=188 ;benign=184 FNAC classification: unsatisfactory, benign, indeterminate, follicular neoplasm, suspicious, positive Non-diagnostic findings: 8 malignant, 18 benign *FNAC rated indeterminate, follicular neoplasm, suspicious, positive (+ve) [benign taken as -ve result]* TP: 168 FN: 20 FP: 99 TN: 85 ; *sensitivity*: 0.894, *specificity*: 0.462 *FNAC rated follicular neoplasm, suspicious, positive (+ve) [indeterminate, benign taken as -ve result]* TP: 160 FN: 28 FP: 90 TN: 94 ; *sensitivity*: 0.851, *specificity*: 0.511 *FNAC rated suspicious, positive (+ve) [follicular neoplasm, indeterminate, benign taken as -ve result]* TP: 136 FN: 52 FP: 21 TN: 163 ; *sensitivity*: 0.723, *specificity*: 0.886

Reference	Theoharis, 2013 #1410[353]
	Post-Bethesda implementation Gold standard results: malignant=199 ;benign=180 FNAC classification: Bethesda - non-diagnostic, benign, indeterminate, follicular neoplasm, suspicious, positive Non-diagnostic findings: 6 malignant, 10 benign *FNAC rated indeterminate, follicular neoplasm, suspicious, positive (+ve) [benign taken as -ve result]* TP: 177 FN: 22 FP: 79 TN: 101 ; *sensitivity*: 0.889, *specificity*: 0.561 *FNAC rated follicular neoplasm, suspicious, positive (+ve) [indeterminate, benign taken as -ve result]* TP: 169 FN: 30 FP: 68 TN: 112 ; *sensitivity*: 0.849, *specificity*: 0.622 *FNAC rated suspicious, positive (+ve) [follicular neoplasm, indeterminate, benign taken as -ve result]* TP: 144 FN: 55 FP: 14 TN: 166 ; *sensitivity*: 0.724, *specificity*: 0.922
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, so some bias possible in who was given surgery)*
Comments	

Reference	Theoharis, 2009 #1411[354]
Study type	Retrospective
Number of patients	n = 378 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: USA

Reference	Theoharis, 2009 #1411[354]
	Inclusion criteria: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG (majority)*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Theoharis, 2009 #1411[354]
Results	Gold standard results: malignant=202 ;benign=176 FNAC classification: unsatisfactory, benign, indeterminate, FN/HCN, SFM, Malignant Non-diagnostic findings: 8 malignant, 17 benign *FNAC rated indeterminate, FN/HCN, SFM, Malignant (+ve) [benign taken as -ve result]* TP: 186 FN: 16 FP: 102 TN: 74 ; *sensitivity*: 0.921, *specificity*: 0.420 *FNAC rated FN/HCN, SFM, Malignant (+ve) [indeterminate, benign taken as -ve result]* TP: 173 FN: 29 FP: 88 TN: 88 ; *sensitivity*: 0.856, *specificity*: 0.500 *FNAC rated SFM, Malignant (+ve) FN/HCN, [FN/HCN, indeterminate, benign taken as -ve result]* TP: 138 FN: 64 FP: 21 TN: 155 ; *sensitivity*: 0.683, *specificity*: 0.881
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, so some bias possible in who was given surgery)*
Comments	

Reference	Thomas, 1998[355]
Study type	Retrospective
Number of patients	n = 93 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Nigeria *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery

Reference	Thomas, 1998[355]
	Exclusion criteria: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=18 ;benign=75 FNAC classification: benign, indeterminate, malignant Non-diagnostic findings: not reported *FNAC rated malignant or indeterminate (+ve) [benign taken as -ve result]* TP: 15 FN: 3 FP: 15 TN: 60 ; *sensitivity*: 0.833, *specificity*: 0.80 *FNAC rated malignant (+ve) [benign or indeterminate taken as -ve result]* TP: 12 FN: 6 FP: 3 TN: 72 ; *sensitivity*: 0.667, *specificity*: 0.96
Source of funding	No funding stated

Reference	Thomas, 1998[355]
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Reference	Tsou, 1997 #1417[360]
Study type	Retrospective
Number of patients	n = 61 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Cancer centre *Country*: Taiwan *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG / non USG – unclear if one of them was >75%*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only (Riu's stain) *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:*

Reference	Tsou, 1997 #1417[360]
	Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	FNA Gold standard results: malignant=40 ;benign=21 FNAC classification: Benign, suspicious, malignant Non-diagnostic findings: none in the surgical cohort *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 38 FN: 2 FP: 10 TN: 11 ; *sensitivity*: 0.95, *specificity*: 0.524 *FNAC rated malignant (+ve) [suspicious or benign taken as -ve result]* TP: 29 FN: 11 FP: 0 TN: 21 ; *sensitivity*: 0.725, *specificity*: 1.0
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, so some bias possible in who was given surgery)*
Comments	

Reference	Varhaug, 1981 #1418[361]
Study type	Retrospective
Number of patients	n = 264 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported

Reference	Varhaug, 1981 #1418[361]
	Setting: University Hospital *Country*: Norway *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Diffuse goitre and toxic goitre *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block – cystic fluid was centrifuged before making smears *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	FNA Gold standard results: malignant=68 ;benign=196 FNAC classification: malignant, suspected, follicular neoplasia, benign, non-diagnostic Non-diagnostic findings: 7 malignant, 36 benign *FNAC rated malignant, suspected, follicular neoplasia (+ve) [benign taken as -ve result]*

Reference	Varhaug, 1981 #1418[361]
	TP: 52 FN: 16 FP: 84 TN: 112 ; *sensitivity*: 0.765, *specificity*: 0.571 *FNAC rated malignant, suspected (+ve) [benign, follicular neoplasia taken as -ve result]* TP: 42 FN: 26 FP: 47 TN: 149 ; *sensitivity*: 0.618, *specificity*: 0.760
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, so some bias possible in who was given surgery)*
Comments	

Reference	Vojvodich, 1994[362]
Study type	Retrospective
Number of patients	n = 98 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Canada *Inclusion criteria*: Patients with solitary thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy

Reference	Vojvodich, 1994[362]
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block. If fluid was aspirated, cytospin preparations, rather than direct smears, were made. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	FNA Gold standard results: malignant= 35 ;benign= 50 FNAC classification: benign, suspicious for malignancy, diagnostic of malignancy, or unsatisfactory Non-diagnostic findings: 13 overall, but histological breakdown not provided so cannot be imputed into analysis *FNAC rated suspicious or malignant (+ve) [benign taken as -ve result]* TP: 29 FN: 6 FP: 6 TN: 44 ; *sensitivity*: 0.829, *specificity*: 0.88 *FNAC rated malignant (+ve) [suspicious or benign taken as -ve result]* TP: 14 FN: 21 FP: 0 TN: 50 ; *sensitivity*: 0.40, *specificity*: 1.0
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, so some bias possible in who was given surgery)*
Comments	

Reference	Walsh, 1983[363]
Study type	Retrospective
Number of patients	n = 76 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: secondary care *Country*: Australia *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Walsh, 1983[363]
Results	FNA Gold standard results: malignant=9 ; benign=67 FNAC classification: benign, suspicious, malignant, unsatisfactory Non-diagnostic findings: 1 malignant, 9 benign *FNAC rated malignant or suspicious (+ve) [benign taken as -ve result]* TP: 7 FN: 2 FP: 14 TN: 53 ; *sensitivity*: 0.778, *specificity*: 0.791 *FNAC rated malignant (+ve) [benign or suspicious taken as -ve result]* TP: 2 FN: 7 FP: 9 TN: 58 ; *sensitivity*: 0.222, *specificity*: 0.866
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, so some bias possible in who was given surgery)*
Comments	

Reference	Wu, 2006[372]
Study type	Retrospective
Number of patients	n = 401 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery

Reference	Wu, 2006[372]
	Exclusion criteria: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	Gold standard results: malignant=112 ; benign=289 FNAC classification: benign (goitre/colloid/thyroiditis), suspicious (nuclear features and cellular features suggestive of malignancy but inadequate cellularity prohibits definitive diagnosis), malignant, atypical (nuclear atypia such as nuclear enlargement, grooves, pseudo inclusions and prominent nucleoli), follicular neoplasm, follicular lesion, inadequate (8-10 cluster on 2 slides) Non-diagnostic findings: 2 malignant, 15 benign *FNAC rated malignant, suspicious, FN, atypia, FL (+ve) [benign taken as -ve result]* TP: 99 FN: 13 FP: 141 TN: 148 ; *sensitivity*: 0.884, *specificity*: 0.512 *FNAC rated malignant, suspicious, FN, atypia (+ve) [benign, FL taken as -ve result]* TP: 92 FN: 20 FP: 97 TN: 192 ; *sensitivity*: 0.821, *specificity*: 0.664 *FNAC rated malignant, suspicious, FN (+ve) [benign, FL, atypia taken as -ve result]* TP: 76 FN: 36 FP: 80 TN: 209 ; *sensitivity*: 0.679, *specificity*: 0.723

Reference	Wu, 2006[372]
	FNAC rated malignant, suspicious (+ve) [benign, FL, atypia, FN taken as -ve result] TP: 47 FN: 65 FP: 21 TN: 268 ; *sensitivity*: 0.419, *specificity*: 0.927
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, so some bias possible in who was given surgery)*
Comments	

Reference	Yavuz, 2020 #1436[381]
Study type	Retrospective
Number of patients	n = 34 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test*

Reference	Yavuz, 2020 #1436[381]
	Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	FNA Gold standard results: malignant=18 ;benign=16 FNAC classification: positive, negative Non-diagnostic findings: not reported *FNAC rated malignant (+ve) [benign taken as -ve result]* TP: 17 FN: 1 FP: 2 TN: 14; *sensitivity*: 0.944, *specificity*: 0.875
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, so some bias possible in who was given surgery)*
Comments	

Reference	Yoder, 2006[385]
Study type	Retrospective
Number of patients	n = 200 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery*

Reference	Yoder, 2006[385]
	Ethnicity: not reported *Setting*: University Hospital *Country*: USA *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG for 81%*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology with ROSE, with smear only. On site cytotechnologist for adequacy. *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Yoder, 2006[385]
Results	Gold standard results: malignant=66 ;benign=134 FNAC classification: unsatisfactory, benign, indeterminate, suspicious, malignant. Non-diagnostic findings: 4 malignant, 5 benign *FNAC rated indeterminate, suspicious, malignant (+ve) [benign taken as -ve result]* TP: 59 FN: 7 FP: 78 TN: 56 ; *sensitivity*: 0.894, *specificity*: 0.418 *FNAC rated suspicious, malignant (+ve) [indeterminate, benign taken as -ve result]* TP: 44 FN: 22 FP: 11 TN: 123 ; *sensitivity*: 0.666, *specificity*: 0.918
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, so some bias possible in who was given surgery)*
Comments	

Reference	Zajdela, 1987 #1442[389]
Study type	Retrospective
Number of patients	n = 372 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: unclear *Country*: France *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery

Reference	Zajdela, 1987 #1442[389]
	Exclusion criteria: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block. In the event of a liquid sample the centrifugation pellet is spread, fixed and stained *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	FNA Gold standard results: malignant=126 ;benign=246 FNAC classification: malignant, suspicious, benign Non-diagnostic findings not reported *FNAC rated malignant or suspicious (+ve) [benign taken as -ve result]* TP: 116 FN: 10 FP: 31 TN: 215 ; *sensitivity*: 0.921, *specificity*: 0.874 *FNAC rated malignant (+ve) [benign or suspicious taken as -ve result]* TP: 94 FN: 32 FP: 3 TN: 243 ; *sensitivity*: 0.746, *specificity*: 0.988

Reference	Zajdela, 1987 #1442[389]
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, so some bias possible in who was given surgery)*
Comments	

Reference	Zbar, 2009[390]
Study type	Retrospective
Number of patients	n = 63 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Barbados *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings

Reference	Zbar, 2009[390]
	Time between measurement of index test and reference standard: Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	FNA Gold standard results: malignant=8 ;benign=55 FNAC classification: benign, follicular neoplasm, suspicious for PTC, PTC. Non-diagnostic findings: not clearly reported *FNAC rated follicular neoplasm, suspicious for PTC, PTC (+ve) [benign taken as -ve result]* TP: 3 FN: 5 FP: 10 TN: 45 ; *sensitivity*: 0.375, *specificity*: 0.818 *FNAC rated suspicious for PTC, PTC (+ve) [follicular neoplasm, benign taken as -ve result]* TP: 3 FN: 5 FP: 3 TN: 52 ; *sensitivity*: 0.375, *specificity*: 0.945
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, so some bias possible in who was given surgery)*
Comments	

Reference	Xu, 2014[378]
Study type	Retrospective
Number of patients	n = 945 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported

Reference	Xu, 2014[378]
	Setting: Cancer Hospital *Country*: China *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): prior US reported and appears to have been used as an indication for FNA* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	FNA Gold standard results: malignant=659 ;benign=286 FNAC classification: positive, negative Non-diagnostic findings: not reported *FNAC rated positive (+ve) [negative taken as -ve result]* TP: 572 FN: 87 FP: 49 TN: 237 ; *sensitivity*: 0.868, *specificity*: 0.829

Reference	Xu, 2014[378]
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, so some bias possible in who was given surgery)*
Comments	

Reference	Liel, 1985[208]
Study type	Retrospective
Number of patients	n = 49 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: Israel *Inclusion criteria*: Patients with 'cold' or 'warm' thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test*

Reference	Liel, 1985[208]
	Fine needle aspiration cytology without ROSE, with smear + cytospin and cell block. Whenever enough solid mass was left, aspiration of the cyst wall was performed. The fluid was centrifuged and examined after fixation and preparation as a cell block *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	FNA Gold standard results: malignant=13 ;benign=36 FNAC classification: Inadequate, benign, follicular neoplasm, suspicious, malignant Non-diagnostic findings: 1 malignant, 7 benign *FNAC rated follicular neoplasm, suspicious, malignant (+ve) [benign taken as -ve result]* TP: 11 FN: 2 FP: 16 TN: 20 ; *sensitivity*: 0.846, *specificity*: 0.555 *FNAC rated suspicious, malignant (+ve) [follicular neoplasm, benign taken as -ve result]* TP: 9 FN: 4 FP: 11 TN: 25 ; *sensitivity*: 0.692, *specificity*: 0.694
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, so some bias possible in who was given surgery)*
Comments	

Reference	Lioe, 1998 #1280[210]
Study type	Retrospective

Reference	Lioe, 1998 #1280[210]
Number of patients	n = 67 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: Departments of histo/cytopathology and surgery *Country*: UK *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	Lioe, 1998 #1280[210]
Results	FNA Gold standard results: malignant=13 ;benign=54 FNAC classification: unsatisfactory, non-neoplastic, reactive vs neoplastic, neoplastic Non-diagnostic findings: 2 malignant, 10 benign *FNAC rated reactive vs neoplastic, neoplastic (+ve) [non-neoplastic taken as -ve result]* TP: 11 FN: 2 FP: 37 TN: 17 ; *sensitivity*: 0.846, *specificity*: 0.315 *FNAC rated neoplastic (+ve) [reactive vs neoplastic, non-neoplastic taken as -ve result]* TP: 9 FN: 4 FP: 23 TN: 31 ; *sensitivity*: 0.692, *specificity*: 0.574
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, so some bias possible in who was given surgery)*
Comments	

Reference	Aggarwal, 1989[7]
Study type	Retrospective
Number of patients	n = 36 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: *Inclusion criteria*: Patients with ultrasonographically solitary cold thyroid nodules given FNAC and subsequent surgery

Reference	Aggarwal, 1989[7]
	Exclusion criteria: Not reported *Stratum (prior US assessment / no prior US assessment): US used to select FNA cases on basis of solitary nodules* *Sub-group (US-guided / not US guided): USG in some but not others (not precisely defined)*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	FNA Gold standard results: malignant=16 ;benign=20 FNAC classification: colloid goitre, follicular neoplasm, equivocal (indeterminate), carcinoma. Non-diagnostic findings: not reported *FNAC rated follicular neoplasm, equivocal (indeterminate), carcinoma (+ve) [colloid goitre taken as -ve result]* TP: 16 FN: 0 FP: 5 TN: 15 ; *sensitivity*: 1.0, *specificity*: 0.75 *FNAC rated equivocal (indeterminate), carcinoma (+ve) [follicular neoplasm, colloid goitre taken as -ve result]* TP: 16 FN: 0 FP: 2 TN: 18 ; *sensitivity*: 1.0, *specificity*: 0.90 *FNAC rated carcinoma (+ve) [equivocal (indeterminate), follicular neoplasm, colloid goitre taken as -ve result]* TP: 12 FN: 4 FP: 0 TN: 20 ; *sensitivity*: 0.75, *specificity*: 1.0

Reference	Aggarwal, 1989[7]
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, so some bias possible in who was given surgery)*
Comments	

Reference	Seok, 2018[315]
Study type	Retrospective
Number of patients	n = 457 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery* *Ethnicity*: not reported *Setting*: University Hospital *Country*: South Korea *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings

Reference	Seok, 2018[315]
	Time between measurement of index test and reference standard: Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	FNA Gold standard results: malignant=377 ;benign=80 FNAC classification: Bethesda I-VI Non-diagnostic findings: 10 malignant, 16 benign *FNAC rated III-VI (+ve) [II taken as -ve result]* TP: 364 FN: 13 FP: 60 TN: 20 ; *sensitivity*: 0.966, *specificity*: 0.25 *FNAC rated IV-VI (+ve) [II-III taken as -ve result]* TP: 319 FN: 58 FP: 20 TN: 60 ; *sensitivity*: 0.846, *specificity*: 0.75 *FNAC rated V-VI (+ve) [II-IV taken as -ve result]* TP: 316 FN: 61 FP: 16 TN: 64 ; *sensitivity*: 0.838, *specificity*: 0.80
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, so some bias possible in who was given surgery)*
Comments	

Reference	Hougaard Chakera, 2003[160]
Study type	Retrospective
Number of patients	n = 67 nodules
Patient characteristics	*Age, mean (SD): not reported for those given surgery* *Gender (female to male ratio): not reported for those given surgery*

Reference	Hougaard Chakera, 2003[160]
	Ethnicity: not reported *Setting*: unclear *Country*: Denmark *Inclusion criteria*: Patients with thyroid nodules given FNAC and subsequent surgery *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): no prior US reported* *Sub-group (US-guided / not US guided): USG not reported*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	FNA Gold standard results: malignant=10 ;benign=57 FNAC classification: malignant, suspicious, benign Non-diagnostic findings: not reported

Reference	Hougaard Chakera, 2003[160]
	FNAC rated malignant and suspicious (+ve) [benign taken as -ve result] TP: 6 FN: 4 FP: 7 TN: 50 ; *sensitivity*: 0.6, *specificity*: 0.877
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, so some bias possible in who was given surgery)*
Comments	

Reference	**Choden, 2021[69]**
Study type	Retrospective
Number of patients	n = 81 nodules
Patient characteristics	*Age, mean (SD): 46.51(15.9), though this was in overall sample, not in those with surgical resection* *Gender (female to male ratio): unclear in those with surgical resection* *Ethnicity*: not reported *Setting*: Secondary care *Country*: Bhutan *Inclusion criteria*: Patients undergoing FNA who also underwent surgical resection *Exclusion criteria*: Patients with missing data *Stratum (prior US assessment / no prior US assessment): Unclear - US mentioned but FNAC appeared to depend on other factors such as radiological and clinical findings too.* *Sub-group (US-guided / not US guided): FNA guidance not mentioned*
Target condition(s)	Thyroid nodule malignancy

Reference	Choden, 2021[69]
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	*Malignant nodules= 36 ; benign nodules = 45* Non diagnostic Bethesda I = 0,0 FNA grading: Bethesda rating *FNAC rated III or above (+ve) [II taken as -ve result]* TP: 34 FN: 2 FP: 16 TN:29 ; *sensitivity*: 0.944 , *specificity*: 0.644 *FNAC rated IV or above (+ve) [II-III taken as -ve result]* TP: 33 FN: 3 FP: 10 TN:35 ; *sensitivity*: 0.917 , *specificity*: 0.778 *FNAC rated V or above (+ve) [II-IV taken as -ve result]* TP: 28 FN: 8 FP: 1 TN:44 ; *sensitivity*: 0.778 , *specificity*: 0.978 *FNAC rated VI (+ve) [II-V taken as -ve result]* TP: 21 FN: 15 FP: 0 TN:45 ; *sensitivity*: 0.583 , *specificity*: 1.0
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, so some bias possible in who was given surgery)*
Comments	

Reference	**Abou-Foul, 2021[2]**
Study type	Retrospective
Number of patients	n = 471 nodules
Patient characteristics	*Age, mean (SD): not reported for analysed sub-group* *Gender (female to male ratio): not reported for analysed sub-group* *Ethnicity*: not reported *Setting*: Secondary care *Country*: UK *Inclusion criteria*: all patients who had thyroid resection (total or hemithyroidectomy) and FNAC *Exclusion criteria*: If final histology reported incidental malignant lesions that were not sampled during the FNAC, these reports were excluded from the analysis *Stratum (prior US assessment / no prior US assessment): unclear* *Sub-group (US-guided / not US guided): USG*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	**Abou-Foul, 2021[2]**
Results	*Malignant nodules= 119; benign nodules = 352* Thy1: 32 malignant, 133 benign FNA grading: RCPath Thy grading system: Thy1, 2, 3a, 3f, 4, and 5 (generally regarded as equivalent to Bethesda categories I to V respectively) *WCS results:* *Thy 3a and above (+ve) [Thy2 taken as -ve result]* TP: 59 FN:60 FP: 189 TN: 163 ; *sensitivity*: 0.496, *specificity*: 0.463 *Thy 3f and above (+ve) [Thy2-3a taken as -ve result]* TP: 45 FN:74 FP: 155 TN: 197 ; *sensitivity*: 0.378, *specificity*: 0.560 *Thy 4 and above (+ve) [Thy2-3f taken as -ve result]* TP: 24 FN:95 FP: 135 TN: 217 ; *sensitivity*: 0.202, *specificity*: 0.616 *Thy 5 (+ve) [Thy2-4 taken as -ve result]* TP: 7 FN: 112 FP: 133 TN: 219 ; *sensitivity*: 0.059, *specificity*: 0.622
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, so some bias possible in who was given surgery)*
Comments	

Reference	**Li, 2021[207]**
Study type	Retrospective
Number of patients	n = 623 nodules
Patient characteristics	*Age, mean (range): 47.3 (7-88)* *Gender (female to male ratio): 488:135* *Ethnicity*: not reported *Setting*: Secondary care *Country*: China *Inclusion criteria*: Patients having FNAC and thyroid surgery

Reference	**Li, 2021[207]**
	Exclusion criteria: No report on the sensation during puncture of the nodule – whether 'soft', 'hard' or 'hard with grittiness'. *Stratum (prior US assessment / no prior US assessment): Prior US assessment, but unclear if this was used as a criterion for FNAC* *Sub-group (US-guided / not US guided): USG.*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	*Malignant nodules= 508; benign nodules =115* No data given for inadequate samples FNA grading: Bethesda *Bethesda V or VI (+ve) [I to IV taken as -ve result]* TP: 452 FN: 56 FP: 8 TN: 107 ; *sensitivity*: 0.889 , *specificity*: 0.930
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, so some bias possible in who was given surgery)*
Comments	

Reference	**Fiorentino, 2021[108]**
Study type	Retrospective
Number of patients	n = 693 nodules (this study focussed on sub-centimetre nodules but also presented data for nodules >1cm. We have summed the data from both sub-groups because this review does not stratify for nodule size)
Patient characteristics	*Age, mean (SD): not reported* *Gender (female to male ratio): not reported* *Ethnicity*: not reported *Setting*: Secondary care *Country*: Italy *Inclusion criteria*: Patients with FNAC and surgical specimens *Exclusion criteria*: Not reported *Stratum (prior US assessment / no prior US assessment): US performed but unclear if used as a criterion for FNAC* *Sub-group (US-guided / not US guided): unclear*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No

Reference	**Fiorentino, 2021[108]**
Results	*Malignant nodules= 416; benign nodules =277* ND: 2 malignant, 4 benign FNA grading: Bethesda *WCS:* *FNAC III or higher (+ve) [II taken as -ve result]* TP: 408 FN: 8 FP: 91 TN: 186 ; *sensitivity*: 0.981 , *specificity*: 0.671 *FNAC IV or higher (+ve) [II - III taken as -ve result]* TP: 402 FN: 14 FP: 49 TN: 228 ; *sensitivity*: 0.966 , *specificity*: 0.823 *FNAC V or higher (+ve) [II - IV taken as -ve result]* TP: 387 FN: 29 FP: 6 TN: 271 ; *sensitivity*: 0.930 , *specificity*: 0.978 *FNAC VI (+ve) [II - V taken as -ve result]* TP: 250 FN: 166 FP: 4 TN: 273 ; *sensitivity*: 0.601 , *specificity*: 0.986
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, so some bias possible in who was given surgery)*
Comments	

Reference	**Bahaj, 2021[32]**
Study type	Retrospective
Number of patients	n = 314 nodules
Patient characteristics	*Age, mean (SD): 42.3(7.3)* *Gender (female to male ratio): 258:56* *Ethnicity*: not reported *Setting*: Secondary care *Country*: Saudi Arabia *Inclusion criteria*: Patients undergoing FNAC and thyroid surgery

Reference	**Bahaj, 2021[32]**
	Exclusion criteria: Not reported *Stratum (prior US assessment / no prior US assessment): US was used but unclear if used as a criterion for FNAC* *Sub-group (US-guided / not US guided): USG used*
Target condition(s)	Thyroid nodule malignancy
Index test(s) and reference standard	*Index test* Fine needle aspiration cytology without ROSE, with smear only *Reference (gold) standard:* Surgical histopathological findings *Time between measurement of index test and reference standard:* Not clear *Blinding of index test:* No *Blinding of gold standard test:* No
Results	*Malignant nodules=150; benign nodules = 164* Inadequate samples: 2 malignant, 6 benign FNA grading: Bethesda *Bethesda III or higher (+ve) [II taken as -ve result]* TP: 127 FN: 23 FP: 33 TN: 131 ; *sensitivity*: 0.847, *specificity*: 0.799 *Bethesda IV or higher (+ve) [II-III taken as -ve result]* TP: 92 FN: 58 FP: 17 TN: 147 ; *sensitivity*: 0.613, *specificity*: 0.896 *Bethesda V or higher (+ve) [II-IV taken as -ve result]* TP: 86 FN: 64 FP: 10 TN: 154 ; *sensitivity*: 0.573, *specificity*: 0.939 *Bethesda VI or higher (+ve) [II-V taken as -ve result]* TP: 17 FN: 133 FP: 6 TN: 158 ; *sensitivity*: 0.113, *specificity*: 0.963
Source of funding	No funding stated

Reference	**Bahaj, 2021[32]**
Limitations	*Risk of bias (QUADAS 2 – risk of bias): very serious risk of bias* *Indirectness (QUADAS 2 - applicability): serious (retrospective, so some bias possible in who was given surgery)*
Comments	

Appendix E – QUADAS2 risk of bias assessment

Table 29: QUADAS2 risk of bias assessment summary

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
Abboud, 200[31]	U	U	U	U	Very serious risk of bias
Abou-Foul, 2021[2]	U	U	U	U	Very serious risk of bias
Acar, 20173	U	U	U	U	Very serious risk of bias
Afroze, 20024	U	U	U	U	Very serious risk of bias
Agcaoglu, 2013[6]	U	U	U	U	Very serious risk of bias
Aggarwal, 1989[7]	U	U	U	U	Very serious risk of bias
Agrawal, 1995[8]	U	U	U	U	Very serious risk of bias
Aguilar-Diosdado, 199[79]	U	U	U	U	Very serious risk of bias
Al-Hureibi, 2003[18]	U	U	U	U	Very serious risk of bias
Altavilla, 1990[23]	U	U	U	U	Very serious risk of bias
Al-Taweel, 1990[19]	U	U	U	U	Very serious risk of bias
Ananthakrishnan, 1990[24]	L	Y	Y	U	No serious risk of bias
Anderson, 1987[25]	U	U	Y	U	Very serious risk of bias
Arul, 2015[29]	U	U	U	U	Very serious risk of bias
Aydogan, 2019[30]	U	U	U	U	Very serious risk of bias
Bahaj, 2021[32]	U	U	U	U	Very serious risk of bias
Bashier, 1996[38]	U	U	U	U	Very serious risk of bias
Belanger, 1983[41]	U	U	U	U	Very serious risk of bias
Bellantone, 2004[42]	U	U	U	U	Very serious risk of bias
Biscotti, 1995[47]	U	U	U	U	Very serious risk of bias
Bodo, 1979[50]	U	U	U	U	Very serious risk of bias
Borman, 1995[51]	U	U	U	U	Very serious risk of bias
Brauer, 1984[53]	U	U	U	U	Very serious risk of bias
Bugis, 1986[55]	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
Can, 2008[61]	U	U	U	U	Very serious risk of bias
Chang, 1997[67]	U	U	U	U	Very serious risk of bias
Choe, 2018[70]	U	U	U	U	Very serious risk of bias
Choden, 2021[69]	U	U	U	U	Very serious risk of bias
Chow, 1999[72]	U	U	U	U	Very serious risk of bias
Cristallini, 1989 #1161[80]	U	U	U	U	Very serious risk of bias
Danese, 1998[85]	U	U	U	U	Very serious risk of bias
Davidsohn, 1995[88]	U	U	U	U	Very serious risk of bias
de Roy van Zuidewijn, 1994[90]	U	U	U	U	Very serious risk of bias
de Vos tot Nederveen Cappel, 2001[91]	U	Y	U	U	Very serious risk of bias
Dwarakanathan, 1989[97]	U	U	U	U	Very serious risk of bias
El Hag, 2021[98]	U	U	U	U	Very serious risk of bias
Ferrari, 1985[106]	U	U	U	U	Very serious risk of bias
Fiorentino, 2021[108]	U	U	U	U	Very serious risk of bias
Francis, 1999[115]	U	U	U	U	Very serious risk of bias
Gardiner, 1986[123]	U	U	U	U	Very serious risk of bias
Gershengorn, 1977[126]	L	Y	U	U	Serious risk of bias
Giansanti, 1989[127]	U	U	U	U	Very serious risk of bias
Gossain, 1998[131]	U	U	U	U	Very serious risk of bias
Gould, 1989[133]	U	U	U	U	Very serious risk of bias
Guo, 2015[138]	U	U	U	U	Very serious risk of bias
Haberal, 2009[144]	U	U	U	U	Very serious risk of bias
Hamming, 1998[150]	U	U	U	U	Very serious risk of bias
Hamming, 1990[149]	U	U	U	U	Very serious risk of bias
Hawkins, 1987[153]	U	U	U	U	Very serious risk of bias
Harsoulis, 1986[152]	U	Y	U	U	Very serious risk of bias
Heimann, 1964[155]	U	U	U	U	Very serious risk of bias
Hosokawa, 2019[159]	U	U	U	U	Very serious risk of bias
Hougaard Chakera, 2003[160]	U	U	U	U	Very serious risk of bias
Huang, 2020[161]	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
Hussain, 1993[163]	U	U	U	U	Very serious risk of bias
Jalan, 2017[166]	U	U	U	U	Very serious risk of bias
Jat, 2019[167]	U	U	U	U	Very serious risk of bias
Jayaram, 1999[168]	U	U	U	U	Very serious risk of bias
Kelman, 2001[175]	U	U	U	U	Very serious risk of bias
Kim, 2013[182]	U	U	U	U	Very serious risk of bias
Kimoto, 1999[187]	U	U	U	U	Very serious risk of bias
Kini, 1985[188]	U	U	U	U	Very serious risk of bias
Kojic Katovic, 2004[193]	U	U	U	U	Very serious risk of bias
Kolendorf, 1975[194]	U	U	U	U	Very serious risk of bias
Kothari, 2019 #1269 [196]	U	U	U	U	Very serious risk of bias
Kumar, 1992[199]	L	U	U	U	Very serious risk of bias
La ROSE, 1991[200]	U	U	U	U	Very serious risk of bias
Leenhardt, 1999[204]	U	U	U	U	Very serious risk of bias
Li, 2013[206]	U	U	U	U	Very serious risk of bias
Li, 2021[207]	U	U	U	U	Very serious risk of bias
Liel, 1985[208]	U	U	U	U	Very serious risk of bias
Lioe, 1998 #1280[210]	U	U	U	U	Very serious risk of bias
Liu, 2009[211]	U	U	U	U	Very serious risk of bias
Lukitto, 1998[217]	U	U	U	U	Very serious risk of bias
Mamoon, 1997[221]	U	U	U	U	Very serious risk of bias
Mandal, 2011[223]	U	U	U	U	Very serious risk of bias
Mandreker, 1995[224]	U	U	U	U	Very serious risk of bias
Mastorakis, 2014[229]	U	U	U	U	Very serious risk of bias
McElroy, 2014[233]	U	U	U	U	Very serious risk of bias
Mehrotra, 2006[236]	U	U	U	U	Very serious risk of bias
Meko, 1995[237]	U	U	U	U	Very serious risk of bias
Merchant, 1995[239]	U	U	U	U	Very serious risk of bias
Mijovic, 2009[240]	L	U	U	U	Very serious risk of bias
Mikosch, 2000[241]	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
Miller, 1979[242]	U	U	U	U	Very serious risk of bias
Munn, 1988 #1322[252]	U	U	U	U	Very serious risk of bias
Nagarajan, 2015 #1326[256]	U	U	U	U	Very serious risk of bias
Natarajan, 1994[258]	U	U	U	U	Very serious risk of bias
Ng, 1988 #1330[261]	U	U	U	U	Very serious risk of bias
Nart, 2010 #1327[257]	U	U	U	U	Very serious risk of bias
Naz, 2014[260]	U	U	U	U	Very serious risk of bias
Okumura, 1999 #1334[266]	U	U	U	U	Very serious risk of bias
Ongphiphadhanakul, 1992 #1335[267]	U	U	Y	U	Very serious risk of bias
Ozdemir, 2017[269]	U	U	Y	U	Very serious risk of bias
Pepper, 1989[275]	U	U	U	U	Very serious risk of bias
Petersen, 1984[276]	U	U	U	U	Very serious risk of bias
Piana, 2011[277]	U	U	U	U	Very serious risk of bias
Pisani, 2000[278]	L	U	U	U	Very serious risk of bias
Prinz, 1983[282]	L	U	U	U	Very serious risk of bias
Radetic, 1984[284]	U	U	U	U	Very serious risk of bias
Raina, 2011[285]	U	U	U	U	Very serious risk of bias
Rammeh, 2019 #1349[286]	U	U	U	U	Very serious risk of bias
Rana, 2021[287]	U	U	U	U	Very serious risk of bias
Rege, 1987[289]	U	U	U	U	Very serious risk of bias
Rodriguez, 1994[295]	U	U	U	U	Very serious risk of bias
Rosen, 1993[296]	U	U	U	U	Very serious risk of bias
Rosen, 1981[298]	U	U	U	U	Very serious risk of bias
Roy, 2019[299]	L	U	U	U	Very serious risk of bias
Rubenfeld, 1982[300]	U	U	U	U	Very serious risk of bias
Russ, 1978[301]	U	U	U	U	Very serious risk of bias
Schmid, 1986 #1370[307]	U	U	U	U	Very serious risk of bias
Schoedel, 2008 #1372[309]	U	U	U	U	Very serious risk of bias
Schwartz, 1982 #1373[310]	U	U	U	U	Very serious risk of bias
Sclabas, 2003[311]	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
Scurry, 2000[312]	U	U	U	U	Very serious risk of bias
Settakorn, 2001[316]	U	U	U	U	Very serious risk of bias
Seya, 1990[317]	U	U	U	U	Very serious risk of bias
Silverman, 1986[327]	U	U	U	U	Very serious risk of bias
Sirpal, 1996[329]	U	U	U	U	Very serious risk of bias
Slowinska-Klencka, 2008[330]	U	U	U	N – 1 year	Very serious risk of bias
Seok, 2018[315]	U	U	U	U	Very serious risk of bias
Son, 2014[332]	U	U	U	U	Very serious risk of bias
Spiliotis, 1992 #1394[334]	U	U	U	U	Very serious risk of bias
Sukumaran, 2014[339]	U	U	U	U	Very serious risk of bias
Tabain, 2004[342]	U	U	U	U	Very serious risk of bias
Tabaqchali, 2000[343]	U	U	U	U	Very serious risk of bias
Takashima, 1994[344]	U	U	U	U	Very serious risk of bias
Takashima, 1992[345]	U	U	U	U	Very serious risk of bias
Tal, 1992[347]	U	U	U	U	Very serious risk of bias
Theoharis, 2013 #1410[353]	U	U	U	U	Very serious risk of bias
Theoharis, 2009 #1411[354]	U	U	U	U	Very serious risk of bias
Thomas, 1998[355]	U	U	U	U	Very serious risk of bias
Tsou, 1997 #1417[360]	U	U	U	U	Very serious risk of bias
Varhaug, 1981 #1418[361]	U	U	U	U	Very serious risk of bias
Vojvodich, 1994[362]	U	U	U	U	Very serious risk of bias
Walsh, 1983[363]	U	U	U	U	Very serious risk of bias
Wang, 2020[364]	U	U	U	U	Very serious risk of bias
Wei, 2016[365]	U	U	U	U	Very serious risk of bias
Wu, 2006[372]	U	U	U	U	Very serious risk of bias
Xiong, 2019[377]	U	Y	Y	U	Serious risk of bias
Xu, 2014[378]	U	U	U	U	Very serious risk of bias
Yavuz, 2020 #1436[381]	U	U	U	U	Very serious risk of bias
Yoder, 2006[385]	U	U	U	U	Very serious risk of bias
Zajdela, 1987 #1442[389]	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
Zbar, 2009[390]	U	U	U	U	Very serious risk of bias
Zelmanovitz, 1998[391]	U	U	U	U	Very serious risk of bias
Zhang, 2015[392]	U	U	U	U	Very serious risk of bias

L=low risk, H=high risk, Y=Yes, N=No, U=unclear, which counts as 'No'

Appendix F – Forest plots

F.1 Coupled sensitivity and specificity forest plots

Adjusted analysis

FNAC, no ROSE, smear only, without prior US

Figure 2: Bethesda Grade III or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Acar, 2017 #1088	87	123	5	101	0.95 [0.88, 0.98]	0.45 [0.38, 0.52]
Arul, 2015 #1113	56	80	3	253	0.95 [0.86, 0.99]	0.76 [0.71, 0.80]
Aydogan, 2019 #1114	124	80	37	275	0.77 [0.70, 0.83]	0.77 [0.73, 0.82]
Bahaj, 2021 #1873	127	33	23	131	0.85 [0.78, 0.90]	0.80 [0.73, 0.86]
Choden, 2021 #1855	34	16	2	29	0.94 [0.81, 0.99]	0.64 [0.49, 0.78]
Fiorentino, 2021 #1857	408	91	8	186	0.98 [0.96, 0.99]	0.67 [0.61, 0.73]
Kim, 2013 #1257	118	11	24	47	0.83 [0.76, 0.89]	0.81 [0.69, 0.90]
Nagarajan, 2015 #1326	438	345	29	460	0.94 [0.91, 0.96]	0.57 [0.54, 0.61]
Seok, 2018 #1377	364	60	13	20	0.97 [0.94, 0.98]	0.25 [0.16, 0.36]
Son, 2014 #1392	414	57	36	187	0.92 [0.89, 0.94]	0.77 [0.71, 0.82]
Sukumaran, 2014 #1399	193	23	5	27	0.97 [0.94, 0.99]	0.54 [0.39, 0.68]
Theoharis, 2009 #1411	186	112	16	74	0.92 [0.87, 0.95]	0.40 [0.33, 0.47]
Theoharis, 2013 #1410	177	79	22	101	0.89 [0.84, 0.93]	0.56 [0.49, 0.63]

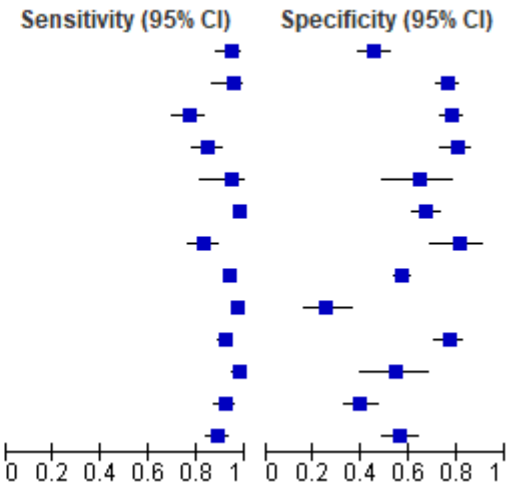


Figure 3: Bethesda Grade IV or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Acar, 2017 #1088	82	59	10	164	0.89 [0.81, 0.95]	0.74 [0.67, 0.79]
Arul, 2015 #1113	46	49	13	284	0.78 [0.65, 0.88]	0.85 [0.81, 0.89]
Aydogan, 2019 #1114	110	49	51	306	0.68 [0.61, 0.75]	0.86 [0.82, 0.90]
Bahaj, 2021 #1873	92	17	58	147	0.61 [0.53, 0.69]	0.90 [0.84, 0.94]
Choden, 2021 #1855	33	10	3	35	0.92 [0.78, 0.98]	0.78 [0.63, 0.89]
Fiorentino, 2021 #1857	402	49	14	228	0.97 [0.94, 0.98]	0.82 [0.77, 0.87]
Hosokawa, 2019 #1234	222	21	50	392	0.82 [0.76, 0.86]	0.95 [0.92, 0.97]
Nagarajan, 2015 #1326	354	205	113	600	0.76 [0.72, 0.80]	0.75 [0.71, 0.78]
Seok, 2018 #1377	319	20	58	60	0.85 [0.81, 0.88]	0.75 [0.64, 0.84]
Son, 2014 #1392	409	53	41	191	0.91 [0.88, 0.93]	0.78 [0.73, 0.83]
Sukumaran, 2014 #1399	187	18	11	32	0.94 [0.90, 0.97]	0.64 [0.49, 0.77]
Theoharis, 2009 #1411	173	98	29	88	0.86 [0.80, 0.90]	0.47 [0.40, 0.55]
Theoharis, 2013 #1410	169	68	30	112	0.85 [0.79, 0.90]	0.62 [0.55, 0.69]

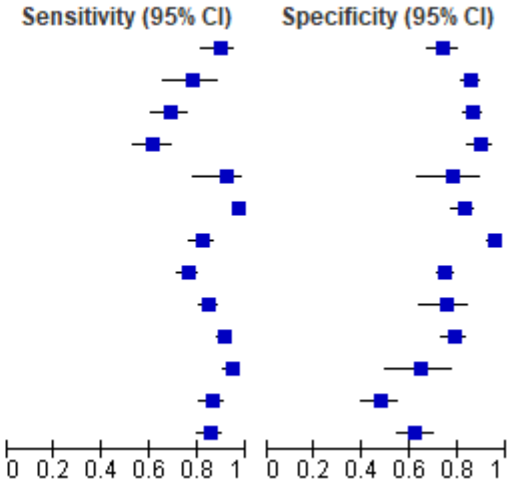


Figure 4: Bethesda Grade V or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Acar, 2017 #1088	75	22	17	202	0.82 [0.72, 0.89]	0.90 [0.86, 0.94]
Arul, 2015 #1113	33	17	26	316	0.56 [0.42, 0.69]	0.95 [0.92, 0.97]
Aydogan, 2019 #1114	95	34	66	321	0.59 [0.51, 0.67]	0.90 [0.87, 0.93]
Bahaj, 2021 #1873	86	10	64	154	0.57 [0.49, 0.65]	0.94 [0.89, 0.97]
Choden, 2021 #1855	28	1	8	44	0.78 [0.61, 0.90]	0.98 [0.88, 1.00]
Fiorentino, 2021 #1857	387	6	29	271	0.93 [0.90, 0.95]	0.98 [0.95, 0.99]
Kim, 2013 #1257	103	4	39	54	0.73 [0.64, 0.80]	0.93 [0.83, 0.98]
Li, 2021 #1865	452	8	56	107	0.89 [0.86, 0.92]	0.93 [0.87, 0.97]
Nagarajan, 2015 #1326	321	122	146	683	0.69 [0.64, 0.73]	0.85 [0.82, 0.87]
Rammeh, 2019 #1349	20	6	4	34	0.83 [0.63, 0.95]	0.85 [0.70, 0.94]
Rana, 2021 #1350	89	3	16	337	0.85 [0.76, 0.91]	0.99 [0.97, 1.00]
Seok, 2018 #1377	316	16	61	64	0.84 [0.80, 0.87]	0.80 [0.70, 0.88]
Son, 2014 #1392	348	31	102	213	0.77 [0.73, 0.81]	0.87 [0.82, 0.91]
Sukumaran, 2014 #1399	158	14	40	36	0.80 [0.74, 0.85]	0.72 [0.58, 0.84]
Theoharis, 2009 #1411	138	21	64	165	0.68 [0.61, 0.75]	0.89 [0.83, 0.93]
Theoharis, 2013 #1410	144	14	55	166	0.72 [0.66, 0.78]	0.92 [0.87, 0.96]

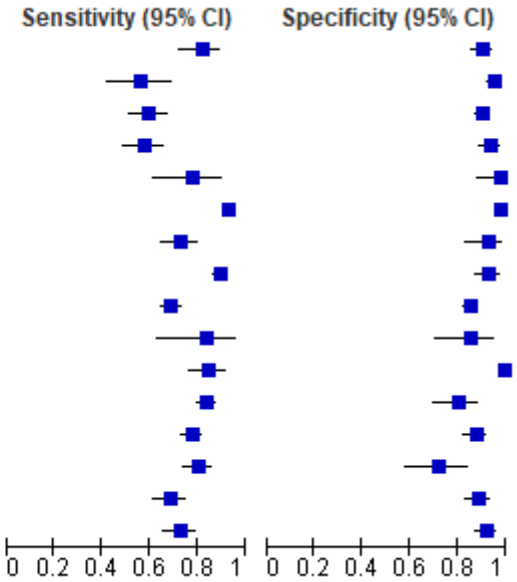


Figure 5: Bethesda Grade VI

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Acar, 2017 #1088	28	14	64	210	0.30 [0.21, 0.41]	0.94 [0.90, 0.97]
Arul, 2015 #1113	16	10	43	323	0.27 [0.16, 0.40]	0.97 [0.95, 0.99]
Aydogan, 2019 #1114	74	32	87	323	0.46 [0.38, 0.54]	0.91 [0.88, 0.94]
Bahaj, 2021 #1873	17	6	133	158	0.11 [0.07, 0.18]	0.96 [0.92, 0.99]
Choden, 2021 #1855	21	0	15	45	0.58 [0.41, 0.74]	1.00 [0.92, 1.00]
Fiorentino, 2021 #1857	250	4	166	273	0.60 [0.55, 0.65]	0.99 [0.96, 1.00]
Nagarajan, 2015 #1326	242	103	225	702	0.52 [0.47, 0.56]	0.87 [0.85, 0.89]
Seok, 2018 #1377	246	16	131	64	0.65 [0.60, 0.70]	0.80 [0.70, 0.88]
Son, 2014 #1392	273	25	177	217	0.61 [0.56, 0.65]	0.90 [0.85, 0.93]
Sukumaran, 2014 #1399	148	14	50	36	0.75 [0.68, 0.81]	0.72 [0.58, 0.84]
Theoharis, 2009 #1411	112	17	90	169	0.55 [0.48, 0.62]	0.91 [0.86, 0.95]
Theoharis, 2013 #1410	118	10	81	170	0.59 [0.52, 0.66]	0.94 [0.90, 0.97]

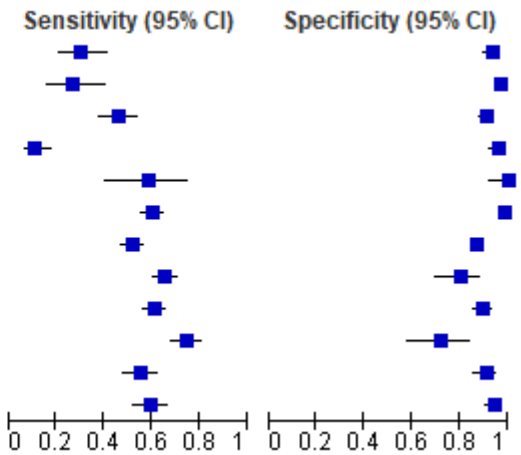


Figure 6: BTA THY 3a or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abou-Foul, 2021 #1872	59	189	60	163	0.50 [0.40, 0.59]	0.46 [0.41, 0.52]
Mandal, 2011 #1293	27	12	3	66	0.90 [0.73, 0.98]	0.85 [0.75, 0.92]

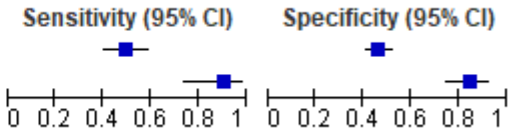


Figure 7: BTA THY 3f or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abou-Foul, 2021 #1872	45	155	74	197	0.38 [0.29, 0.47]	0.56 [0.51, 0.61]

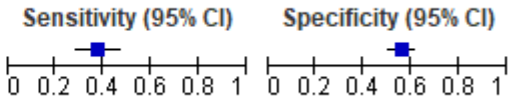


Figure 8: BTA THY 4 or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abou-Foul, 2021 #1872	24	135	95	217	0.20 [0.13, 0.29]	0.62 [0.56, 0.67]

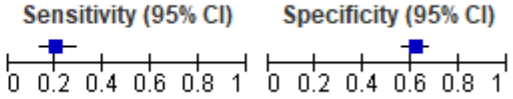


Figure 9: BTA THY 5

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abou-Foul, 2021 #1872	7	133	112	219	0.06 [0.02, 0.12]	0.62 [0.57, 0.67]
Mandal, 2011 #1293	18	0	12	78	0.60 [0.41, 0.77]	1.00 [0.95, 1.00]

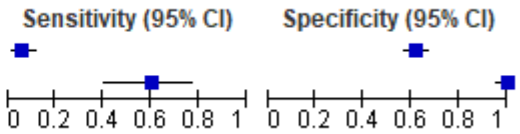


Figure 10: AC 3 or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Mehrotra, 2006 #1306	48	167	13	67	0.79 [0.66, 0.88]	0.29 [0.23, 0.35]
Mehrotra, 2006 #1306b	20	55	5	13	0.80 [0.59, 0.93]	0.19 [0.11, 0.30]
Tabaqchali, 2000 #1402	25	136	9	69	0.74 [0.56, 0.87]	0.34 [0.27, 0.41]

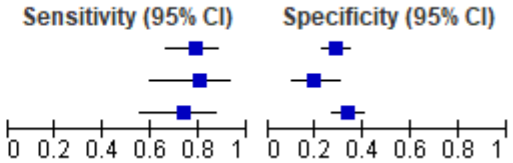


Figure 11: AC 4 or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Mehrotra, 2006 #1306	25	80	36	154	0.41 [0.29, 0.54]	0.66 [0.59, 0.72]
Mehrotra, 2006 #1306b	10	12	15	56	0.40 [0.21, 0.61]	0.82 [0.71, 0.91]
Tabaqchali, 2000 #1402	13	77	21	128	0.38 [0.22, 0.56]	0.62 [0.55, 0.69]

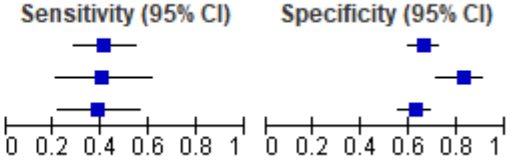


Figure 12: 2 way: malignant v benign

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bashier, 1996 #1121	11	12	1	65	0.92 [0.62, 1.00]	0.84 [0.74, 0.92]
Lukitto, 1998 #1286	4	1	12	150	0.25 [0.07, 0.52]	0.99 [0.96, 1.00]
Maruta, 2003 #1296	112	28	36	128	0.76 [0.68, 0.82]	0.82 [0.75, 0.88]
Rege, 1987 #1352	13	0	2	170	0.87 [0.60, 0.98]	1.00 [0.98, 1.00]
Russ, 1978 #1364	8	0	3	18	0.73 [0.39, 0.94]	1.00 [0.81, 1.00]
Schwartz, 1982 #1373	5	3	6	78	0.45 [0.17, 0.77]	0.96 [0.90, 0.99]
Seya, 1990 #1379	11	0	2	13	0.85 [0.55, 0.98]	1.00 [0.75, 1.00]
Silverman, 1986 #1387	1	0	1	6	0.50 [0.01, 0.99]	1.00 [0.54, 1.00]
Takashima, 1992 #1404	16	0	0	11	1.00 [0.79, 1.00]	1.00 [0.72, 1.00]
Takashima, 1994 #1403	64	3	3	29	0.96 [0.87, 0.99]	0.91 [0.75, 0.98]
Takashima, 1994 #1403b	21	1	3	9	0.88 [0.68, 0.97]	0.90 [0.55, 1.00]
Takashima,1992 #1404b	6	0	2	6	0.75 [0.35, 0.97]	1.00 [0.54, 1.00]
Yavuz, 2020 #1436	17	2	1	14	0.94 [0.73, 1.00]	0.88 [0.62, 0.98]

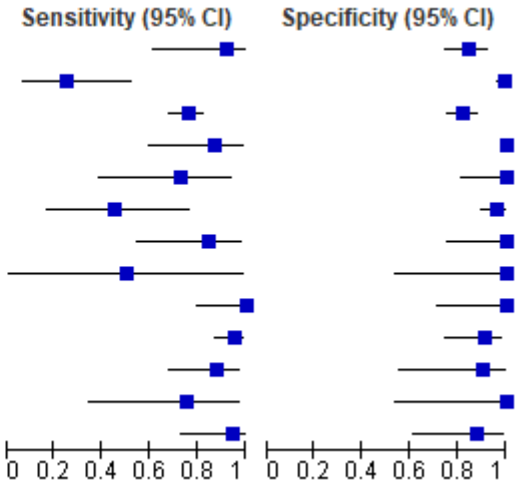
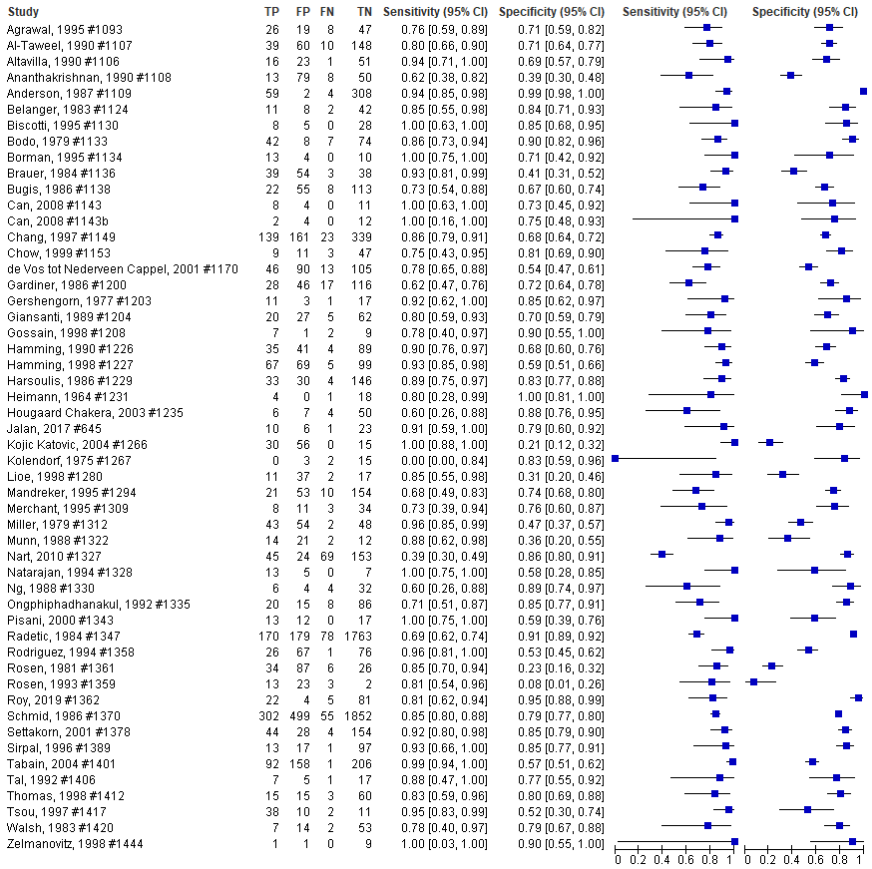
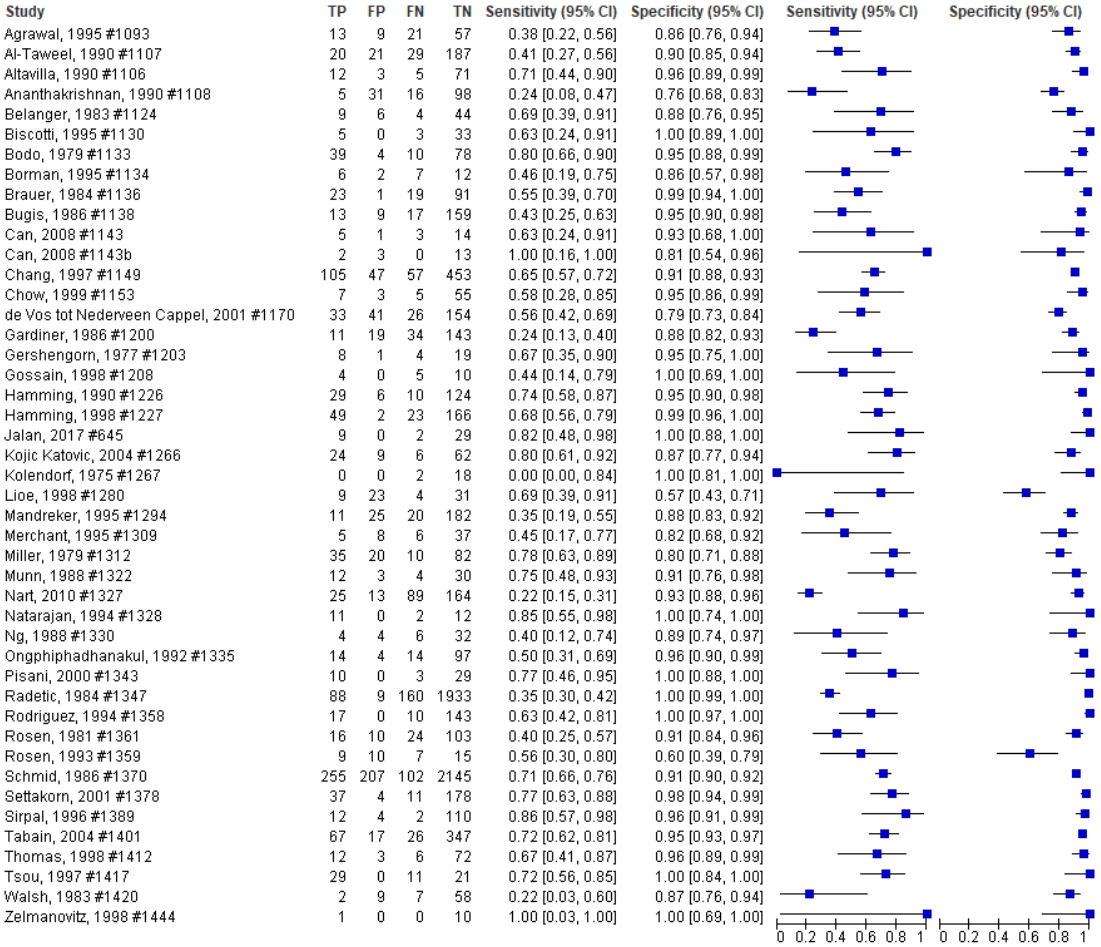


Figure 13: 3 way: suspicious or malignant (negative =benign)



Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Agrawal, 1995 #1093	26	19	8	47	0.76 [0.59, 0.89]	0.71 [0.59, 0.82]
Al-Taweel, 1990 #1107	39	60	10	148	0.80 [0.66, 0.90]	0.71 [0.64, 0.77]
Altavilla, 1990 #1106	16	23	1	51	0.94 [0.71, 1.00]	0.69 [0.57, 0.79]
Ananthakrishnan, 1990 #1108	13	79	8	50	0.62 [0.38, 0.82]	0.39 [0.30, 0.48]
Anderson, 1987 #1109	59	2	4	308	0.94 [0.85, 0.98]	0.99 [0.98, 1.00]
Belanger, 1983 #1124	11	8	2	42	0.85 [0.55, 0.98]	0.84 [0.71, 0.93]
Biscotti, 1995 #1130	8	5	0	28	1.00 [0.63, 1.00]	0.85 [0.68, 0.95]
Bodo, 1979 #1133	42	8	7	74	0.86 [0.73, 0.94]	0.90 [0.82, 0.96]
Borman, 1995 #1134	13	4	0	10	1.00 [0.75, 1.00]	0.71 [0.42, 0.92]
Brauer, 1984 #1136	39	54	3	38	0.93 [0.81, 0.99]	0.41 [0.31, 0.52]
Bugis, 1986 #1138	22	55	8	113	0.73 [0.54, 0.88]	0.67 [0.60, 0.74]
Can, 2008 #1143	8	4	0	11	1.00 [0.63, 1.00]	0.73 [0.45, 0.92]
Can, 2008 #1143b	2	4	0	12	1.00 [0.16, 1.00]	0.75 [0.48, 0.93]
Chang, 1997 #1149	139	161	23	339	0.86 [0.79, 0.91]	0.68 [0.64, 0.72]
Chow, 1999 #1153	9	11	3	47	0.75 [0.43, 0.95]	0.81 [0.69, 0.90]
de Vos tot Nederveen Cappel, 2001 #1170	46	90	13	105	0.78 [0.65, 0.88]	0.54 [0.47, 0.61]
Gardiner, 1986 #1200	28	46	17	116	0.62 [0.47, 0.76]	0.72 [0.64, 0.78]
Gershengorn, 1977 #1203	11	3	1	17	0.92 [0.62, 1.00]	0.85 [0.62, 0.97]
Giansanti, 1989 #1204	20	27	5	62	0.80 [0.59, 0.93]	0.70 [0.59, 0.79]
Gossain, 1998 #1208	7	1	2	9	0.78 [0.40, 0.97]	0.90 [0.55, 1.00]
Hamming, 1990 #1226	35	41	4	89	0.90 [0.76, 0.97]	0.68 [0.60, 0.76]
Hamming, 1998 #1227	67	69	5	99	0.93 [0.85, 0.98]	0.59 [0.51, 0.66]
Harsoulis, 1986 #1229	33	30	4	146	0.89 [0.75, 0.97]	0.83 [0.77, 0.88]
Heimann, 1964 #1231	4	0	1	18	0.80 [0.28, 0.99]	1.00 [0.81, 1.00]
Hougaard Chakera, 2003 #1235	6	7	4	50	0.60 [0.26, 0.88]	0.88 [0.76, 0.95]
Jalan, 2017 #645	10	6	1	23	0.91 [0.59, 1.00]	0.79 [0.60, 0.92]
Kojic Katovic, 2004 #1266	30	56	0	15	1.00 [0.88, 1.00]	0.21 [0.12, 0.32]
Kolendorf, 1975 #1267	0	3	2	15	0.00 [0.00, 0.84]	0.83 [0.59, 0.96]
Lioe, 1998 #1280	11	37	2	17	0.85 [0.55, 0.98]	0.31 [0.20, 0.46]
Mandreker, 1995 #1294	21	53	10	154	0.68 [0.49, 0.83]	0.74 [0.68, 0.80]
Merchant, 1995 #1309	8	11	3	34	0.73 [0.39, 0.94]	0.76 [0.60, 0.87]
Miller, 1979 #1312	43	54	2	48	0.96 [0.85, 0.99]	0.47 [0.37, 0.57]
Munn, 1988 #1322	14	21	2	12	0.88 [0.62, 0.98]	0.36 [0.20, 0.55]
Nart, 2010 #1327	45	24	69	153	0.39 [0.30, 0.49]	0.86 [0.80, 0.91]
Natarajan, 1994 #1328	13	5	0	7	1.00 [0.75, 1.00]	0.58 [0.28, 0.85]
Ng, 1988 #1330	6	4	4	32	0.60 [0.26, 0.88]	0.89 [0.74, 0.97]
Ongphiphadhanakul, 1992 #1335	20	15	8	86	0.71 [0.51, 0.87]	0.85 [0.77, 0.91]
Pisani, 2000 #1343	13	12	0	17	1.00 [0.75, 1.00]	0.59 [0.39, 0.76]
Radetic, 1984 #1347	170	179	78	1763	0.69 [0.62, 0.74]	0.91 [0.89, 0.92]
Rodriguez, 1994 #1358	26	67	1	76	0.96 [0.81, 1.00]	0.53 [0.45, 0.62]
Rosen, 1981 #1361	34	87	6	26	0.85 [0.70, 0.94]	0.23 [0.16, 0.32]
Rosen, 1993 #1359	13	23	3	2	0.81 [0.54, 0.96]	0.08 [0.01, 0.26]
Roy, 2019 #1362	22	4	5	81	0.81 [0.62, 0.94]	0.95 [0.88, 0.99]
Schmid, 1986 #1370	302	499	55	1852	0.85 [0.80, 0.88]	0.79 [0.77, 0.80]
Settakorn, 2001 #1378	44	28	4	154	0.92 [0.80, 0.98]	0.85 [0.79, 0.90]
Sirpal, 1996 #1389	13	17	1	97	0.93 [0.66, 1.00]	0.85 [0.77, 0.91]
Tabain, 2004 #1401	92	158	1	206	0.99 [0.94, 1.00]	0.57 [0.51, 0.62]
Tal, 1992 #1406	7	5	1	17	0.88 [0.47, 1.00]	0.77 [0.55, 0.92]
Thomas, 1998 #1412	15	15	3	60	0.83 [0.59, 0.96]	0.80 [0.69, 0.88]
Tsou, 1997 #1417	38	10	2	11	0.95 [0.83, 0.99]	0.52 [0.30, 0.74]
Walsh, 1983 #1420	7	14	2	53	0.78 [0.40, 0.97]	0.79 [0.67, 0.88]
Zelmanovitz, 1998 #1444	1	1	0	9	1.00 [0.03, 1.00]	0.90 [0.55, 1.00]

Figure 14: 3 way: malignant (negative = suspicious or benign)



Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Agrawal, 1995 #1093	13	9	21	57	0.38 [0.22, 0.56]	0.86 [0.76, 0.94]
Al-Taweel, 1990 #1107	20	21	29	187	0.41 [0.27, 0.56]	0.90 [0.85, 0.94]
Altavilla, 1990 #1106	12	3	5	71	0.71 [0.44, 0.90]	0.96 [0.89, 0.99]
Ananthakrishnan, 1990 #1108	5	31	16	98	0.24 [0.08, 0.47]	0.76 [0.68, 0.83]
Belanger, 1983 #1124	9	6	4	44	0.69 [0.39, 0.91]	0.88 [0.76, 0.95]
Biscotti, 1995 #1130	5	0	3	33	0.63 [0.24, 0.91]	1.00 [0.89, 1.00]
Bodo, 1979 #1133	39	4	10	78	0.80 [0.66, 0.90]	0.95 [0.88, 0.99]
Borman, 1995 #1134	6	2	7	12	0.46 [0.19, 0.75]	0.86 [0.57, 0.98]
Brauer, 1984 #1136	23	1	19	91	0.55 [0.39, 0.70]	0.99 [0.94, 1.00]
Bugis, 1986 #1138	13	9	17	159	0.43 [0.25, 0.63]	0.95 [0.90, 0.98]
Can, 2008 #1143	5	1	3	14	0.63 [0.24, 0.91]	0.93 [0.68, 1.00]
Can, 2008 #1143b	2	3	0	13	1.00 [0.16, 1.00]	0.81 [0.54, 0.96]
Chang, 1997 #1149	105	47	57	453	0.65 [0.57, 0.72]	0.91 [0.88, 0.93]
Chow, 1999 #1153	7	3	5	55	0.58 [0.28, 0.85]	0.95 [0.86, 0.99]
de Vos tot Nederveen Cappel, 2001 #1170	33	41	26	154	0.56 [0.42, 0.69]	0.79 [0.73, 0.84]
Gardiner, 1986 #1200	11	19	34	143	0.24 [0.13, 0.40]	0.88 [0.82, 0.93]
Gershengorn, 1977 #1203	8	1	4	19	0.67 [0.35, 0.90]	0.95 [0.75, 1.00]
Gossain, 1998 #1208	4	0	5	10	0.44 [0.14, 0.79]	1.00 [0.69, 1.00]
Hamming, 1990 #1226	29	6	10	124	0.74 [0.58, 0.87]	0.95 [0.90, 0.98]
Hamming, 1998 #1227	49	2	23	166	0.68 [0.56, 0.79]	0.99 [0.96, 1.00]
Jalan, 2017 #645	9	0	2	29	0.82 [0.48, 0.98]	1.00 [0.88, 1.00]
Kojic Katovic, 2004 #1266	24	9	6	62	0.80 [0.61, 0.92]	0.87 [0.77, 0.94]
Kolendorf, 1975 #1267	0	0	2	18	0.00 [0.00, 0.84]	1.00 [0.81, 1.00]
Lioe, 1998 #1280	9	23	4	31	0.69 [0.39, 0.91]	0.57 [0.43, 0.71]
Mandreker, 1995 #1294	11	25	20	182	0.35 [0.19, 0.55]	0.88 [0.83, 0.92]
Merchant, 1995 #1309	5	8	6	37	0.45 [0.17, 0.77]	0.82 [0.68, 0.92]
Miller, 1979 #1312	35	20	10	82	0.78 [0.63, 0.89]	0.80 [0.71, 0.88]
Munn, 1988 #1322	12	3	4	30	0.75 [0.48, 0.93]	0.91 [0.76, 0.98]
Nart, 2010 #1327	25	13	89	164	0.22 [0.15, 0.31]	0.93 [0.88, 0.96]
Natarajan, 1994 #1328	11	0	2	12	0.85 [0.55, 0.98]	1.00 [0.74, 1.00]
Ng, 1988 #1330	4	4	6	32	0.40 [0.12, 0.74]	0.89 [0.74, 0.97]
Ongphiphadhanakul, 1992 #1335	14	4	14	97	0.50 [0.31, 0.69]	0.96 [0.90, 0.99]
Pisani, 2000 #1343	10	0	3	29	0.77 [0.46, 0.95]	1.00 [0.88, 1.00]
Radetic, 1984 #1347	88	9	160	1933	0.35 [0.30, 0.42]	1.00 [0.99, 1.00]
Rodriguez, 1994 #1358	17	0	10	143	0.63 [0.42, 0.81]	1.00 [0.97, 1.00]
Rosen, 1981 #1361	16	10	24	103	0.40 [0.25, 0.57]	0.91 [0.84, 0.96]
Rosen, 1993 #1359	9	10	7	15	0.56 [0.30, 0.80]	0.60 [0.39, 0.79]
Schmid, 1986 #1370	255	207	102	2145	0.71 [0.66, 0.76]	0.91 [0.90, 0.92]
Settakorn, 2001 #1378	37	4	11	178	0.77 [0.63, 0.88]	0.98 [0.94, 0.99]
Sirpal, 1996 #1389	12	4	2	110	0.86 [0.57, 0.98]	0.96 [0.91, 0.99]
Tabain, 2004 #1401	67	17	26	347	0.72 [0.62, 0.81]	0.95 [0.93, 0.97]
Thomas, 1998 #1412	12	3	6	72	0.67 [0.41, 0.87]	0.96 [0.89, 0.99]
Tsou, 1997 #1417	29	0	11	21	0.72 [0.56, 0.85]	1.00 [0.84, 1.00]
Walsh, 1983 #1420	2	9	7	58	0.22 [0.03, 0.60]	0.87 [0.76, 0.94]
Zelmanovitz, 1998 #1444	1	0	0	10	1.00 [0.03, 1.00]	1.00 [0.69, 1.00]

Figure 15: 4 way: malignant or suspicious or indeterminate (negative = benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abboud, 2003 #1087	15	23	0	8	1.00 [0.78, 1.00]	0.26 [0.12, 0.45]
Al-Hureibi, 2003 #1104	15	26	23	135	0.39 [0.24, 0.57]	0.84 [0.77, 0.89]
de Roy van Zuidewijn, 1994 #1169	80	63	7	115	0.92 [0.84, 0.97]	0.65 [0.57, 0.72]
Haberal, 2009 #1220	59	31	4	166	0.94 [0.85, 0.98]	0.84 [0.78, 0.89]
Kini, 1985 #1262	93	179	6	101	0.94 [0.87, 0.98]	0.36 [0.30, 0.42]
Mamoon, 1997 #1290	13	16	1	146	0.93 [0.66, 1.00]	0.90 [0.84, 0.94]
Okumura, 1999 #1334	46	49	4	10	0.92 [0.81, 0.98]	0.17 [0.08, 0.29]
Schoedel, 2008 #1372	14	7	7	18	0.67 [0.43, 0.85]	0.72 [0.51, 0.88]
Scurry, 2000 #1375	23	60	14	13	0.62 [0.45, 0.78]	0.18 [0.10, 0.29]
Spiliotis, 1992 #1394	28	42	3	128	0.90 [0.74, 0.98]	0.75 [0.68, 0.82]
Wu, 2006 #1428	99	141	13	148	0.88 [0.81, 0.94]	0.51 [0.45, 0.57]
Zbar, 2009 #1443	3	10	5	45	0.38 [0.09, 0.76]	0.82 [0.69, 0.91]

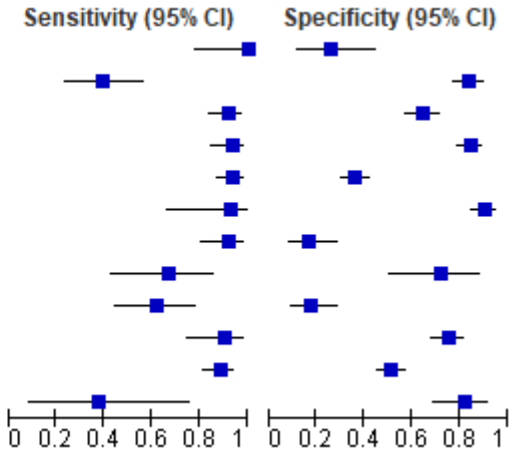


Figure 16: 4 way: malignant or suspicious (negative = benign or indeterminate)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abboud, 2003 #1087	13	7	2	24	0.87 [0.60, 0.98]	0.77 [0.59, 0.90]
Aggarwal, 1989 #693	16	2	0	18	1.00 [0.79, 1.00]	0.90 [0.68, 0.99]
Al-Hureibi, 2003 #1104	6	4	32	157	0.16 [0.06, 0.31]	0.98 [0.94, 0.99]
de Roy van Zuidewijn, 1994 #1169	68	19	19	159	0.78 [0.68, 0.86]	0.89 [0.84, 0.93]
Haberal, 2009 #1220	53	18	10	179	0.84 [0.73, 0.92]	0.91 [0.86, 0.94]
Kini, 1985 #1262	64	50	35	230	0.65 [0.54, 0.74]	0.82 [0.77, 0.86]
Mamoon, 1997 #1290	11	8	3	154	0.79 [0.49, 0.95]	0.95 [0.91, 0.98]
Okumura, 1999 #1334	25	9	25	50	0.50 [0.36, 0.64]	0.85 [0.73, 0.93]
Raina, 2011 #1348	5	1	2	17	0.71 [0.29, 0.96]	0.94 [0.73, 1.00]
Schoedel, 2008 #1372	8	3	13	22	0.38 [0.18, 0.62]	0.88 [0.69, 0.97]
Scurry, 2000 #1375	10	36	27	37	0.27 [0.14, 0.44]	0.51 [0.39, 0.63]
Spiliotis, 1992 #1394	25	30	6	140	0.81 [0.63, 0.93]	0.82 [0.76, 0.88]
Wu, 2006 #1428	92	97	20	192	0.82 [0.74, 0.89]	0.66 [0.61, 0.72]
Zbar, 2009 #1443	3	3	5	52	0.38 [0.09, 0.76]	0.95 [0.85, 0.99]

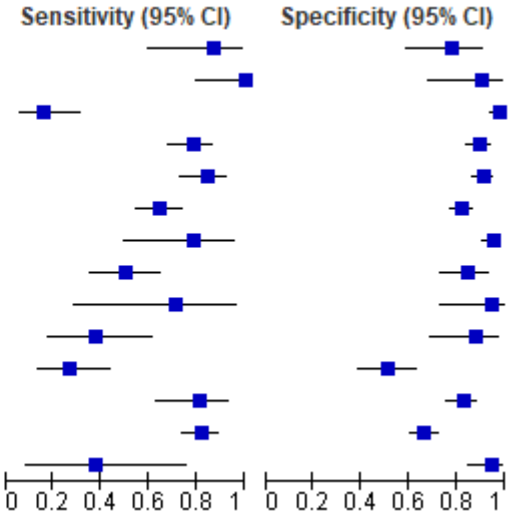


Figure 17: 4 way: malignant (negative = benign or indeterminate or suspicious)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abboud, 2003 #1087	11	2	4	29	0.73 [0.45, 0.92]	0.94 [0.79, 0.99]
Aggarwal, 1989 #693	12	0	4	20	0.75 [0.48, 0.93]	1.00 [0.83, 1.00]
Al-Hureibi, 2003 #1104	4	2	34	159	0.11 [0.03, 0.25]	0.99 [0.96, 1.00]
de Roy van Zuidewijn, 1994 #1169	57	6	30	172	0.66 [0.55, 0.75]	0.97 [0.93, 0.99]
Haberal, 2009 #1220	41	1	22	196	0.65 [0.52, 0.77]	0.99 [0.97, 1.00]
Kini, 1985 #1262	53	15	46	265	0.54 [0.43, 0.64]	0.95 [0.91, 0.97]
Mamoon, 1997 #1290	6	2	8	160	0.43 [0.18, 0.71]	0.99 [0.96, 1.00]
Okumura, 1999 #1334	10	0	40	59	0.20 [0.10, 0.34]	1.00 [0.94, 1.00]
Scurry, 2000 #1375	2	34	35	39	0.05 [0.01, 0.18]	0.53 [0.41, 0.65]
Spiliotis, 1992 #1394	17	10	13	160	0.57 [0.37, 0.75]	0.94 [0.89, 0.97]
Wu, 2006 #1428	34	15	78	274	0.30 [0.22, 0.40]	0.95 [0.92, 0.97]
Zbar, 2009 #1443	1	1	7	54	0.13 [0.00, 0.53]	0.98 [0.90, 1.00]

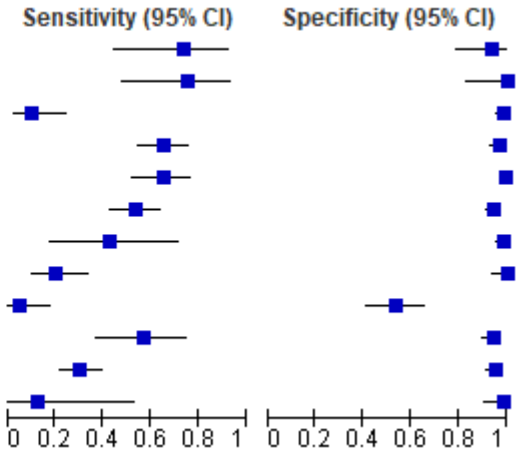


Figure 18: 5 way: malignant or suspicious or two grades of indeterminate (negative = benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Acar, 2017 #1088	23	100	4	99	0.85 [0.66, 0.96]	0.50 [0.43, 0.57]
Francis, 1999 #1192	17	12	3	13	0.85 [0.62, 0.97]	0.52 [0.31, 0.72]
Kelman, 2001 #1250	91	246	43	104	0.68 [0.59, 0.76]	0.30 [0.25, 0.35]
La Rosa, 1991 #1273	241	320	9	257	0.96 [0.93, 0.98]	0.45 [0.40, 0.49]
Prinz, 1983 #1345	17	51	3	38	0.85 [0.62, 0.97]	0.43 [0.32, 0.54]
Theoharis, 2013 #1410	168	99	20	85	0.89 [0.84, 0.93]	0.46 [0.39, 0.54]

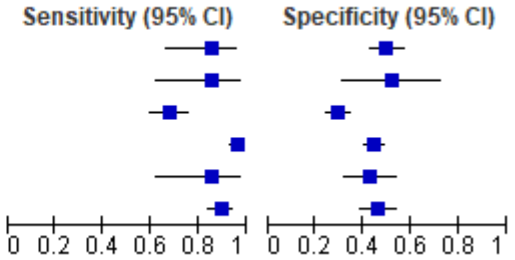


Figure 19: 5 way: malignant or suspicious or one grade of indeterminate (negative = lower grade of indeterminate or benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Acar, 2017 #1088	23	93	4	106	0.85 [0.66, 0.96]	0.53 [0.46, 0.60]
Francis, 1999 #1192	14	3	6	22	0.70 [0.46, 0.88]	0.88 [0.69, 0.97]
Kelman, 2001 #1250	87	203	47	147	0.65 [0.56, 0.73]	0.42 [0.37, 0.47]
La Rosa, 1991 #1273	215	87	35	490	0.86 [0.81, 0.90]	0.85 [0.82, 0.88]
Theoharis, 2013 #1410	160	90	28	94	0.85 [0.79, 0.90]	0.51 [0.44, 0.59]

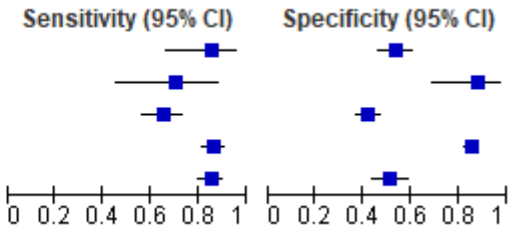


Figure 20: 5 way: malignant (negative = suspicious or two grades of indeterminate or benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Acar, 2017 #1088	15	36	12	163	0.56 [0.35, 0.75]	0.82 [0.76, 0.87]
Francis, 1999 #1192	12	3	8	22	0.60 [0.36, 0.81]	0.88 [0.69, 0.97]
Kelman, 2001 #1250	54	173	80	177	0.40 [0.32, 0.49]	0.51 [0.45, 0.56]
La Rosa, 1991 #1273	179	23	79	554	0.69 [0.63, 0.75]	0.96 [0.94, 0.97]
Prinz, 1983 #1345	10	31	10	58	0.50 [0.27, 0.73]	0.65 [0.54, 0.75]
Theoharis, 2013 #1410	115	18	73	166	0.61 [0.54, 0.68]	0.90 [0.85, 0.94]

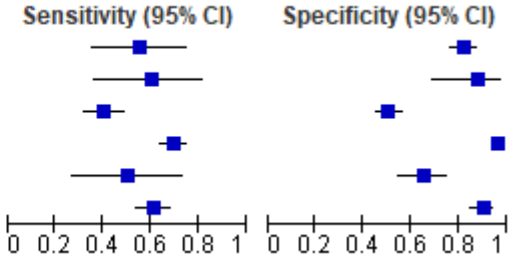


Figure 21: 1 or more inclusions

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Gould, 1989 #1210	13	1	11	45	0.54 [0.33, 0.74]	0.98 [0.88, 1.00]

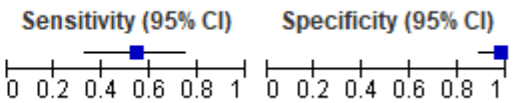


Figure 22: 1 or more grooves

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Gould, 1989 #1210	22	27	1	19	0.96 [0.78, 1.00]	0.41 [0.27, 0.57]

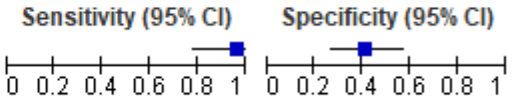


Figure 23: 2 or more grooves

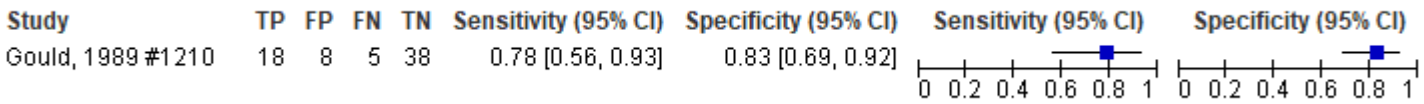
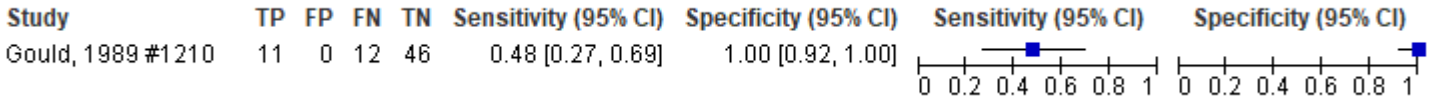


Figure 24: 3 or more grooves



FNAC, no ROSE, smear only, with prior US

Figure 25: Bethesda Grade III or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Huang, 2020 #796	228	124	5	35	0.98 [0.95, 0.99]	0.22 [0.16, 0.29]
Ozdemir, 2017 #1336	339	1899	127	2750	0.73 [0.68, 0.77]	0.59 [0.58, 0.61]
Wang, 2020 #1421	99	67	15	93	0.87 [0.79, 0.92]	0.58 [0.50, 0.66]

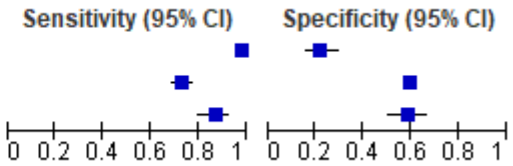


Figure 26: Bethesda Grade IV or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Huang, 2020 #796	218	33	15	126	0.94 [0.90, 0.96]	0.79 [0.72, 0.85]
Ozdemir, 2017 #1336	223	1358	243	3291	0.48 [0.43, 0.52]	0.71 [0.69, 0.72]
Wang, 2020 #1421	74	29	40	131	0.65 [0.55, 0.74]	0.82 [0.75, 0.88]

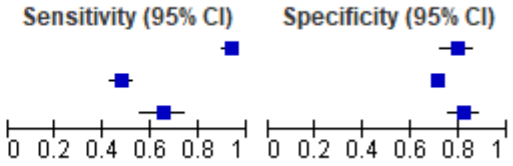


Figure 27: Bethesda Grade V or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Huang, 2020 #796	123	4	110	155	0.53 [0.46, 0.59]	0.97 [0.94, 0.99]
Ozdemir, 2017 #1336	204	1311	262	3338	0.44 [0.39, 0.48]	0.72 [0.70, 0.73]
Wang, 2020 #1421	73	22	41	138	0.64 [0.55, 0.73]	0.86 [0.80, 0.91]

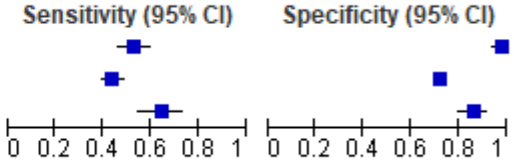


Figure 28: Bethesda Grade VI or above

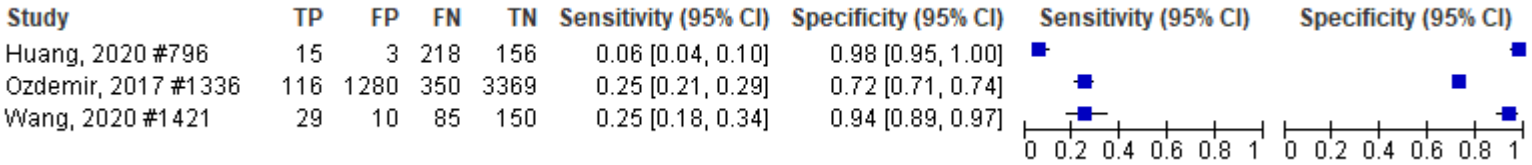


Figure 29: 2 way: malignant versus benign

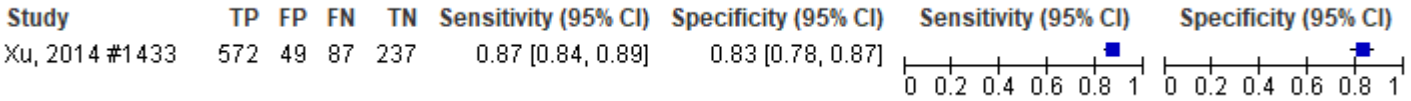


Figure 30: 3 way: suspicious or malignant (negative = benign)

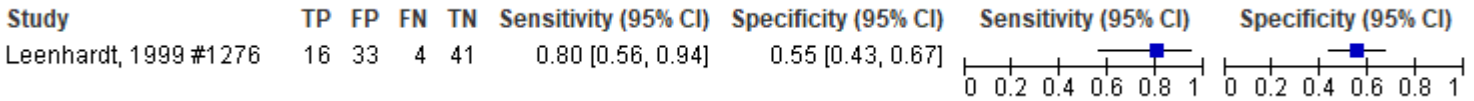


Figure 31: 3 way: malignant (negative = suspicious or benign)

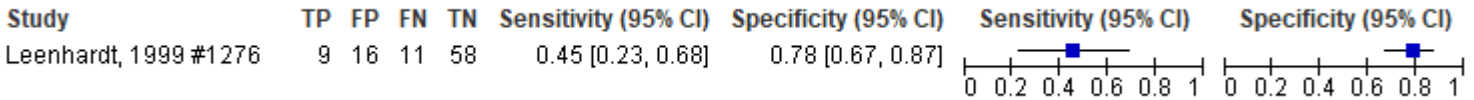


Figure 32: 4 way De May classification: malignant, suspicious, non malignant follicular proliferation (negative = benign)

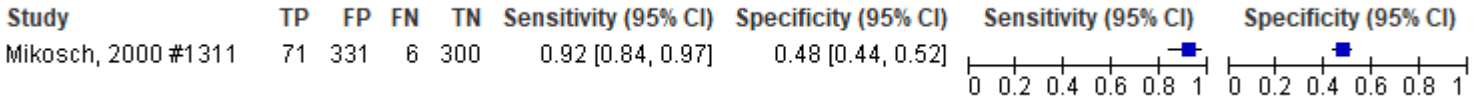


Figure 33: 4 way De May classification: malignant, suspicious (negative = benign, non malignant follicular proliferation)

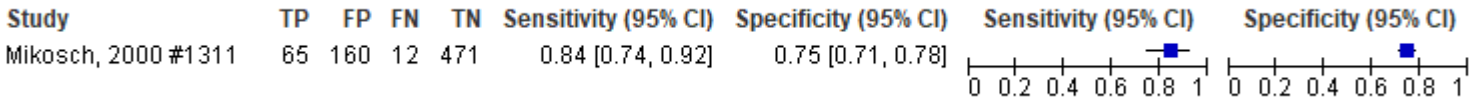


Figure 34: 4 way De May classification: malignant (negative = benign, non malignant follicular proliferation, suspicious)

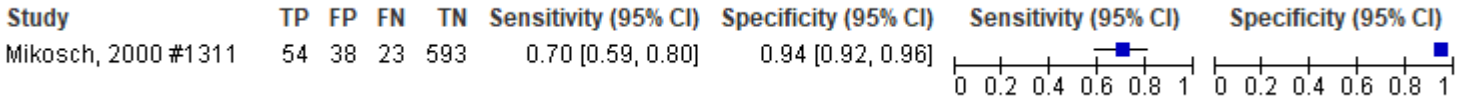


Figure 35: 4 way Piana classification: C3 or more

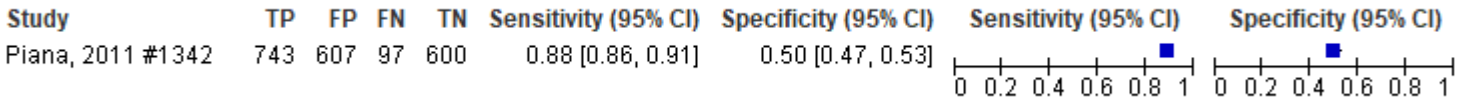


Figure 36: 4 way Piana classification: C4 or more

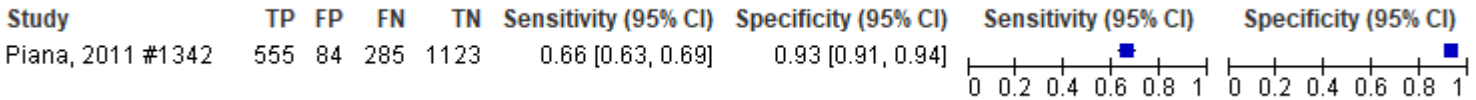


Figure 37: 4 way Piana classification: C5 or more

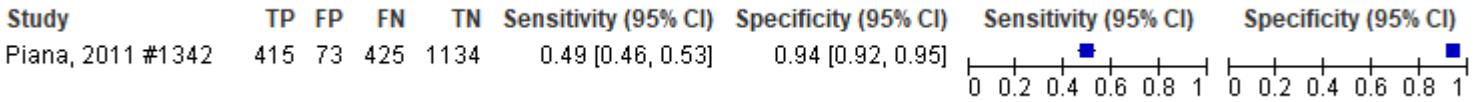


Figure 38: 4 way generic: malignant, suspicious, indeterminate (benign = negative)

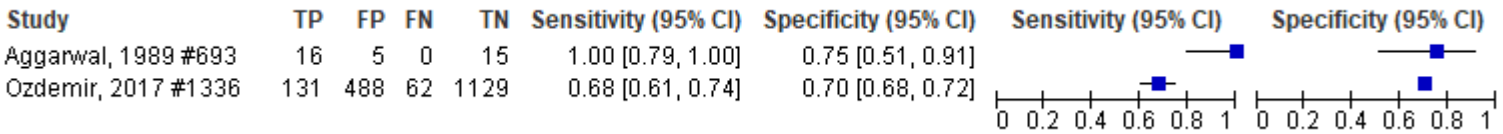
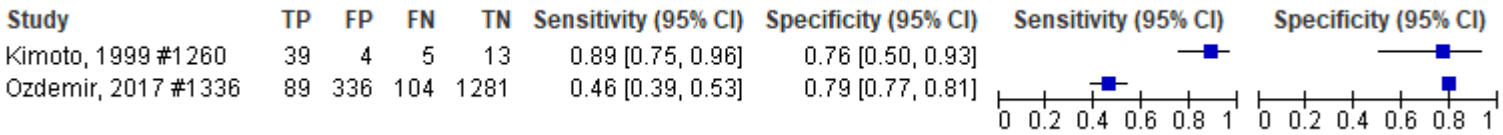


Figure 39: 4 way generic: malignant, suspicious(indeterminate, benign = negative)



FNAC, no ROSE, smear, with cytospin and/or cell-block, without prior US

Figure 40: Bethesda Grade III or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Mastorakis, 2014 #1299	197	53	14	236	0.93 [0.89, 0.96]	0.82 [0.77, 0.86]
Mastorakis, 2014 #1299b	77	61	4	358	0.95 [0.88, 0.99]	0.85 [0.82, 0.89]
McElroy, 2014 #1303	9	6	3	10	0.75 [0.43, 0.95]	0.63 [0.35, 0.85]
Nagarajan, 2015 #1326	25	15	1	13	0.96 [0.80, 1.00]	0.46 [0.28, 0.66]
Naz, 2014 #1329	9	7	5	40	0.64 [0.35, 0.87]	0.85 [0.72, 0.94]

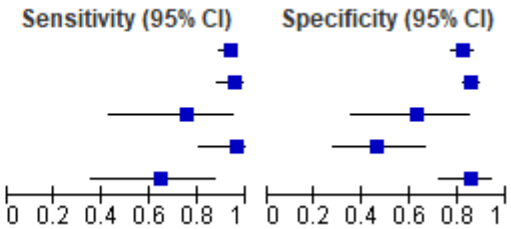


Figure 41: Bethesda Grade IV or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Mastorakis, 2014 #1299	186	17	25	272	0.88 [0.83, 0.92]	0.94 [0.91, 0.97]
Mastorakis, 2014 #1299b	75	38	6	381	0.93 [0.85, 0.97]	0.91 [0.88, 0.94]
McElroy, 2014 #1303	7	5	5	11	0.58 [0.28, 0.85]	0.69 [0.41, 0.89]
Nagarajan, 2015 #1326	21	4	5	24	0.81 [0.61, 0.93]	0.86 [0.67, 0.96]
Naz, 2014 #1329	7	3	7	44	0.50 [0.23, 0.77]	0.94 [0.82, 0.99]

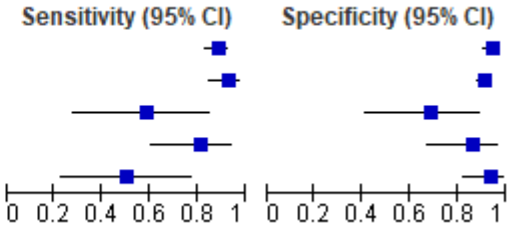


Figure 42: Bethesda Grade V or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Mastorakis, 2014 #1299	184	13	27	276	0.87 [0.82, 0.91]	0.96 [0.92, 0.98]
Mastorakis, 2014 #1299b	75	27	6	392	0.93 [0.85, 0.97]	0.94 [0.91, 0.96]
McElroy, 2014 #1303	5	4	7	12	0.42 [0.15, 0.72]	0.75 [0.48, 0.93]
Nagarajan, 2015 #1326	17	3	9	25	0.65 [0.44, 0.83]	0.89 [0.72, 0.98]
Naz, 2014 #1329	6	0	8	47	0.43 [0.18, 0.71]	1.00 [0.92, 1.00]

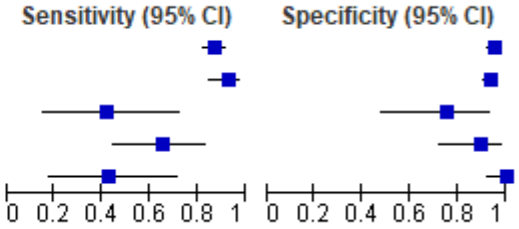


Figure 43: Bethesda Grade VI or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Mastorakis, 2014 #1299	133	11	78	278	0.63 [0.56, 0.70]	0.96 [0.93, 0.98]
Mastorakis, 2014 #1299b	61	25	20	394	0.75 [0.64, 0.84]	0.94 [0.91, 0.96]
McElroy, 2014 #1303	5	4	7	12	0.42 [0.15, 0.72]	0.75 [0.48, 0.93]
Nagarajan, 2015 #1326	12	2	14	26	0.46 [0.27, 0.67]	0.93 [0.76, 0.99]
Naz, 2014 #1329	2	0	12	47	0.14 [0.02, 0.43]	1.00 [0.92, 1.00]

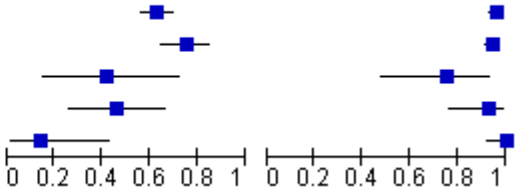


Figure 44: 2 way: malignant v benign

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Wei, 2016 #1422	20	1	2	53	0.91 [0.71, 0.99]	0.98 [0.90, 1.00]

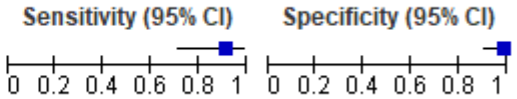


Figure 45: 3 way: malignant or suspicious (negative = benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Aguilar-Diosdado, 1997 #1094	43	57	22	167	0.66 [0.53, 0.77]	0.75 [0.68, 0.80]
Biscotti, 1995 #1130	8	7	0	26	1.00 [0.63, 1.00]	0.79 [0.61, 0.91]
Cristallini, 1989 #1161	15	9	1	16	0.94 [0.70, 1.00]	0.64 [0.43, 0.82]
Danese, 1998 #1164	99	130	4	307	0.96 [0.90, 0.99]	0.70 [0.66, 0.75]
Danese, 1998 #1164b	79	147	9	300	0.90 [0.81, 0.95]	0.67 [0.63, 0.71]
Dwarakanathan, 1989 #1176	18	19	1	25	0.95 [0.74, 1.00]	0.57 [0.41, 0.72]
Ferrari, 1985 #1184	7	16	2	43	0.78 [0.40, 0.97]	0.73 [0.60, 0.84]
Kumar, 1992 #1272	12	21	1	52	0.92 [0.64, 1.00]	0.71 [0.59, 0.81]
Pepper, 1989 #1340	5	8	1	7	0.83 [0.36, 1.00]	0.47 [0.21, 0.73]
Petersen, 1984 #1341	19	84	2	84	0.90 [0.70, 0.99]	0.50 [0.42, 0.58]
Rubenfeld, 1982 #1363	15	11	0	4	1.00 [0.78, 1.00]	0.27 [0.08, 0.55]
Vojvodich, 1994 #1419	29	6	6	44	0.83 [0.66, 0.93]	0.88 [0.76, 0.95]
Zajdela, 1987 #1442	116	31	10	215	0.92 [0.86, 0.96]	0.87 [0.83, 0.91]

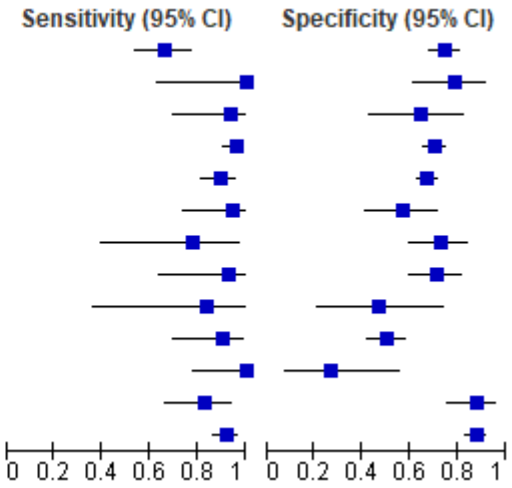


Figure 46: 3 way: malignant (negative = benign or suspicious)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Aguilar-Diosdado, 1997 #1094	24	29	41	195	0.37 [0.25, 0.50]	0.87 [0.82, 0.91]
Biscotti, 1995 #1130	5	0	3	33	0.63 [0.24, 0.91]	1.00 [0.89, 1.00]
Cristallini, 1989 #1161	15	2	1	23	0.94 [0.70, 1.00]	0.92 [0.74, 0.99]
Danese, 1998 #1164	70	4	33	433	0.68 [0.58, 0.77]	0.99 [0.98, 1.00]
Danese, 1998 #1164b	53	13	35	434	0.60 [0.49, 0.71]	0.97 [0.95, 0.98]
Dwarakanathan, 1989 #1176	15	1	4	43	0.79 [0.54, 0.94]	0.98 [0.88, 1.00]
Ferrari, 1985 #1184	6	0	3	59	0.67 [0.30, 0.93]	1.00 [0.94, 1.00]
Kumar, 1992 #1272	8	7	5	66	0.62 [0.32, 0.86]	0.90 [0.81, 0.96]
Vojvodich, 1994 #1419	14	0	21	50	0.40 [0.24, 0.58]	1.00 [0.93, 1.00]
Zajdela, 1987 #1442	94	3	32	243	0.75 [0.66, 0.82]	0.99 [0.96, 1.00]

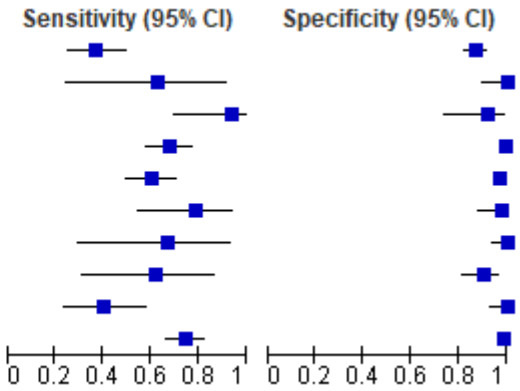


Figure 47: 4 way: malignant, suspicious, indeterminate (negative = benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bellantone, 2004 #1125	17	70	4	28	0.81 [0.58, 0.95]	0.29 [0.20, 0.39]
Liel, 1985 #1279	11	16	2	20	0.85 [0.55, 0.98]	0.56 [0.38, 0.72]
Mijovic, 2009 #1310	63	28	10	14	0.86 [0.76, 0.93]	0.33 [0.20, 0.50]
Scurry, 2000 #1375	22	57	10	3	0.69 [0.50, 0.84]	0.05 [0.01, 0.14]
Varhaug, 1981 #1418	52	84	16	112	0.76 [0.65, 0.86]	0.57 [0.50, 0.64]

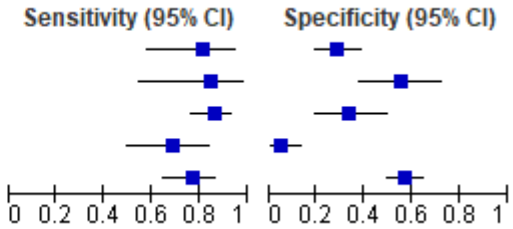


Figure 48: 4 way: malignant, suspicious (negative = benign, indeterminate)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bellantone, 2004 #1125	16	59	5	39	0.76 [0.53, 0.92]	0.40 [0.30, 0.50]
Hawkins, 1987 #1230	63	16	10	326	0.86 [0.76, 0.93]	0.95 [0.93, 0.97]
Liel, 1985 #1279	9	11	4	25	0.69 [0.39, 0.91]	0.69 [0.52, 0.84]
Mijovic, 2009 #1310	39	6	34	36	0.53 [0.41, 0.65]	0.86 [0.71, 0.95]
Scurry, 2000 #1375	10	28	22	32	0.31 [0.16, 0.50]	0.53 [0.40, 0.66]
Varhaug, 1981 #1418	42	47	26	149	0.62 [0.49, 0.73]	0.76 [0.69, 0.82]

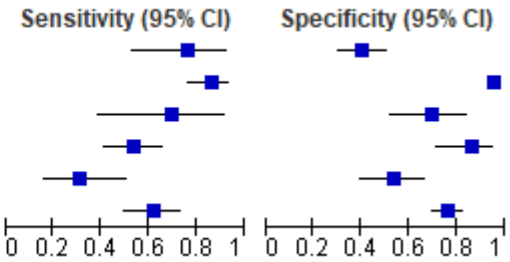


Figure 49: 4 way: malignant (negative = benign, indeterminate, suspicious)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bellantone, 2004 #1125	4	9	17	89	0.19 [0.05, 0.42]	0.91 [0.83, 0.96]
Hawkins, 1987 #1230	48	3	25	339	0.66 [0.54, 0.76]	0.99 [0.97, 1.00]
Liel, 1985 #1279	8	8	8	25	0.50 [0.25, 0.75]	0.76 [0.58, 0.89]
Scurry, 2000 #1375	2	25	30	35	0.06 [0.01, 0.21]	0.58 [0.45, 0.71]
Varhaug, 1981 #1418	26	36	42	160	0.38 [0.27, 0.51]	0.82 [0.75, 0.87]

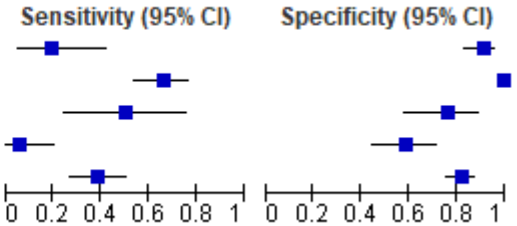
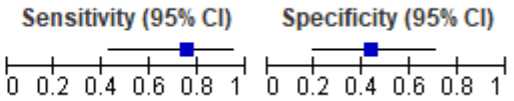


Figure 50: 5 way: malignant, suspicious, 2 grades of indeterminate (negative = benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
McElroy, 2014 #1303	9	9	3	7	0.75 [0.43, 0.95]	0.44 [0.20, 0.70]



FNAC, no ROSE, smear, with cytospin and/or cell-block, with prior US

Figure 51: Bethesda Grade III or above

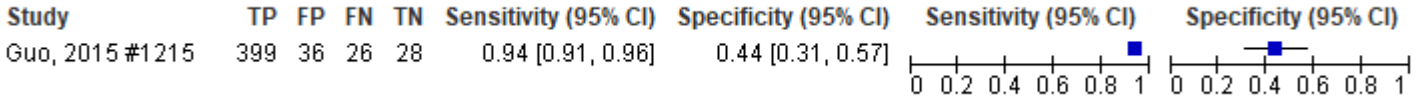


Figure 52: Bethesda Grade IV or above

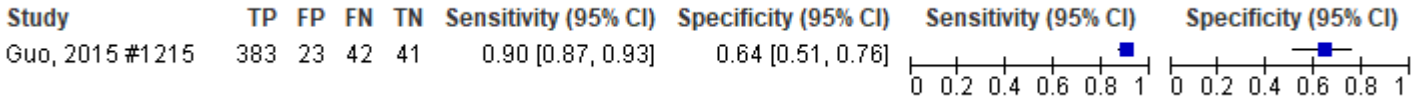


Figure 53: Bethesda Grade V or above

Figure 54: Bethesda Grade VI

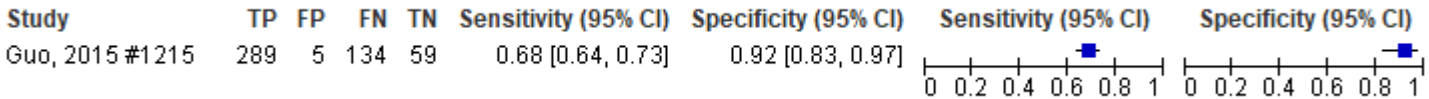
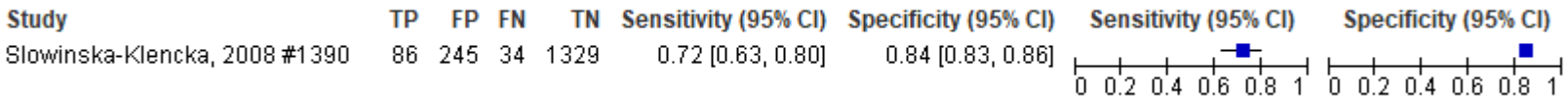


Figure 55: Benign or above



FNAC, with ROSE, smear only, without prior US

Figure 56: Bethesda Grade III or above

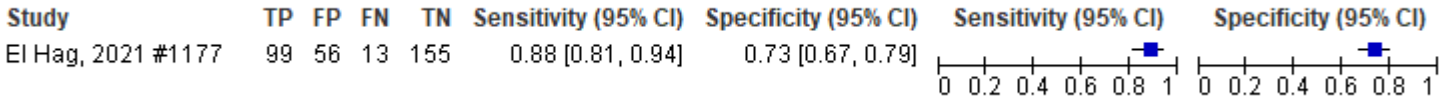


Figure 57: Bethesda Grade IV or above

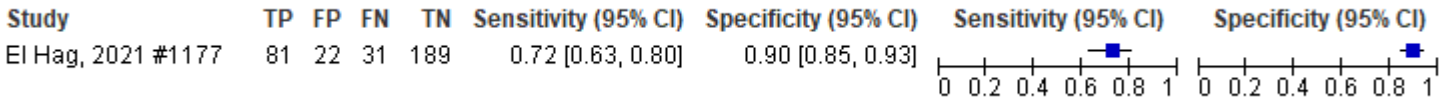


Figure 58: Bethesda Grade V or above

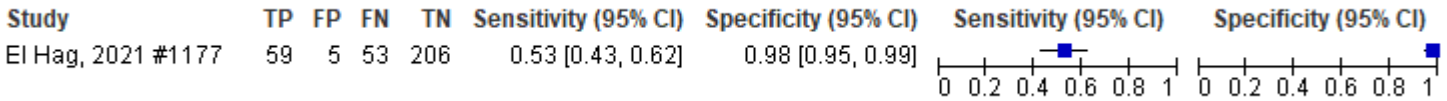


Figure 59: Bethesda Grade VI

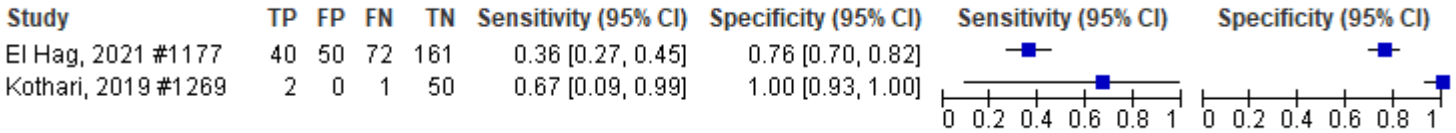


Figure 60: 3 way: malignant and suspicious (negative = benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Jat, 2019 #1242	6	24	4	41	0.60 [0.26, 0.88]	0.63 [0.50, 0.75]
Liu, 2009 #1281	22	6	2	10	0.92 [0.73, 0.99]	0.63 [0.35, 0.85]
Zhang, 2015 #1445	26	27	1	24	0.96 [0.81, 1.00]	0.47 [0.33, 0.62]

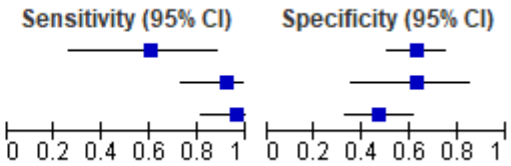


Figure 61: 3 way: malignant (negative = benign and suspicious)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Jat, 2019 #1242	4	2	6	63	0.40 [0.12, 0.74]	0.97 [0.89, 1.00]
Zhang, 2015 #1445	19	9	8	42	0.70 [0.50, 0.86]	0.82 [0.69, 0.92]

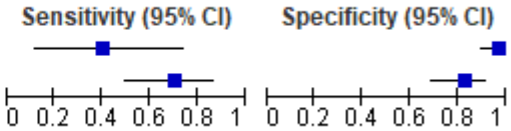


Figure 62: 4 way: malignant, suspicious, indeterminate (negative = benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Jayaram, 1999 #1243	57	73	7	188	0.89 [0.79, 0.95]	0.72 [0.66, 0.77]
Yoder, 2006 #1438	59	78	7	56	0.89 [0.79, 0.96]	0.42 [0.33, 0.51]

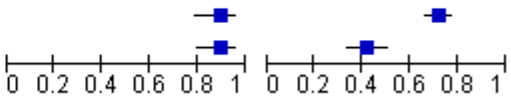


Figure 63: 4 way: malignant, suspicious (negative = benign, indeterminate)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Jayaram, 1999 #1243	35	13	29	248	0.55 [0.42, 0.67]	0.95 [0.92, 0.97]
Yoder, 2006 #1438	44	11	22	123	0.67 [0.54, 0.78]	0.92 [0.86, 0.96]

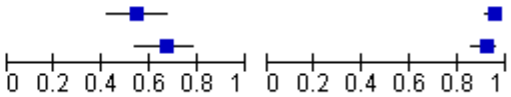
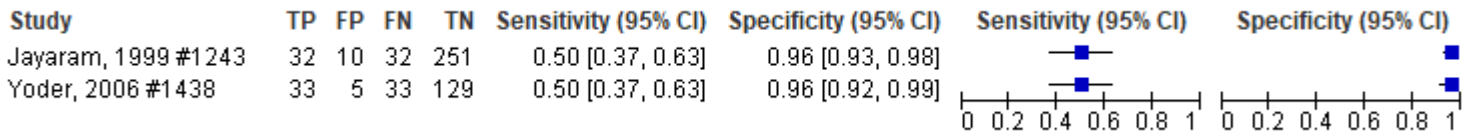
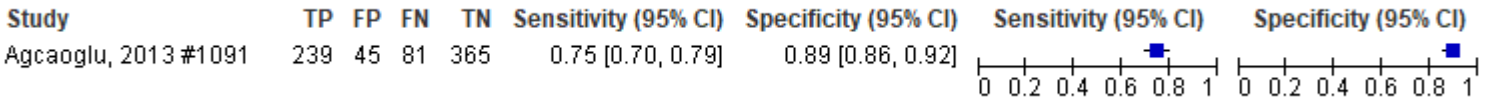


Figure 64: 4 way: malignant (negative = benign, indeterminate, suspicious)



FNAC, with ROSE, smear only, with prior US

Figure 65: intermediate or malignant



FNAC, with ROSE, smear, with cytospin and/or cell-block, without prior US

Figure 66: 3 way: suspicious or malignant (negative = benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Hussain, 1993 #1238	6	29	1	72	0.86 [0.42, 1.00]	0.71 [0.61, 0.80]
Meko, 1995 #1307	13	32	6	39	0.68 [0.43, 0.87]	0.55 [0.43, 0.67]

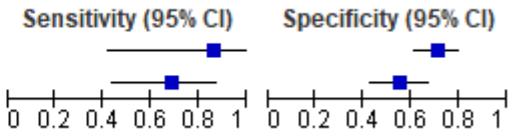


Figure 67: 3 way: malignant (negative = suspicious or benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Hussain, 1993 #1238	4	21	3	80	0.57 [0.18, 0.90]	0.79 [0.70, 0.87]

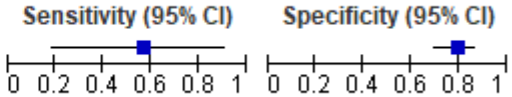


Figure 68: 4 way: malignant, suspicious, indeterminate (negative = benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Davidsohn, 1995 #1167	15	17	0	12	1.00 [0.78, 1.00]	0.41 [0.24, 0.61]

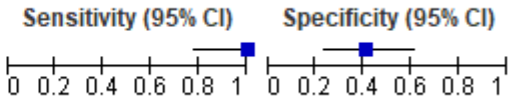


Figure 69: 4 way: malignant, suspicious (negative = benign, indeterminate)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Davidsohn, 1995 #1167	10	0	5	29	0.67 [0.38, 0.88]	1.00 [0.88, 1.00]

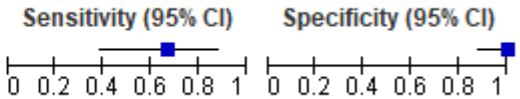


Figure 70: 5 way: malignant, suspicious, 2 grades of indeterminate (negative = benign)

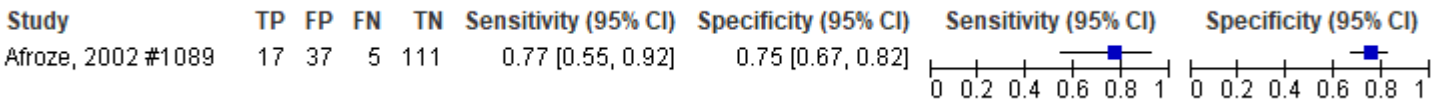


Figure 71: 5 way: malignant, suspicious (negative = 2 grades of indeterminate, benign)

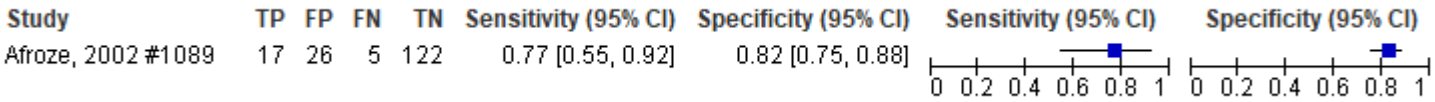


Figure 72: 5 way: malignant, suspicious (negative = suspicious, lower grade of indeterminate, benign)

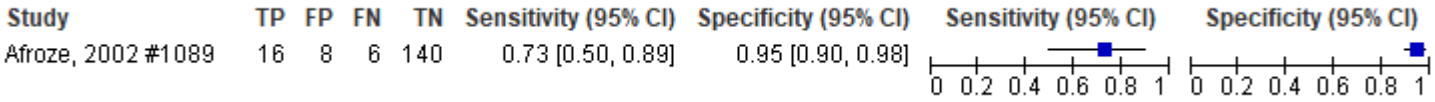
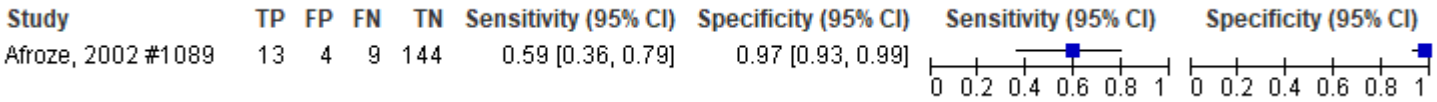


Figure 73: 5 way: malignant (negative = suspicious, 2 grades of indeterminate, benign)



FNAC, with ROSE, smear, with cytospin and/or cell-block, with prior US

Figure 74: indeterminate follicular, indeterminate Hurtle, Suspicious for malignancy, or positive

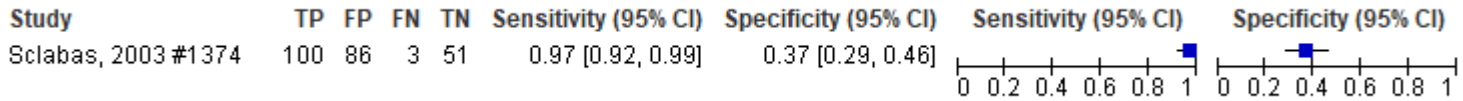


Figure 75: Suspicious for malignancy, or indeterminate follicular or positive

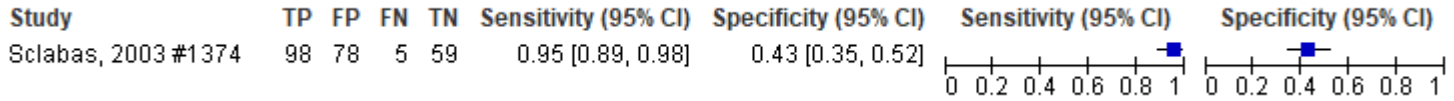


Figure 76: Suspicious for malignancy, or positive

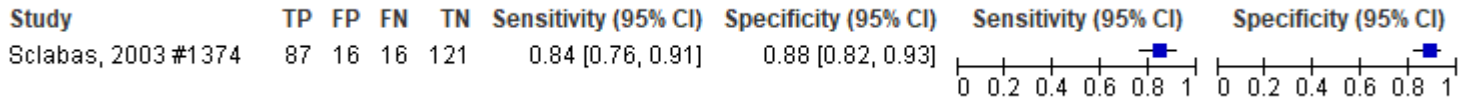
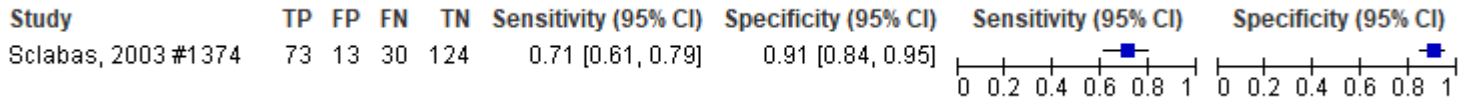


Figure 77: Positive for malignancy



Core biopsy, without prior US

Figure 78: carcinoma or neoplasm (versus benign)

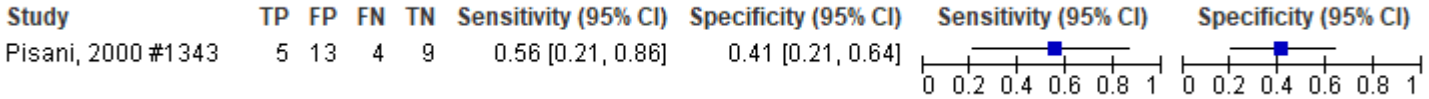


Figure 79: carcinoma (versus benign/indeterminate)

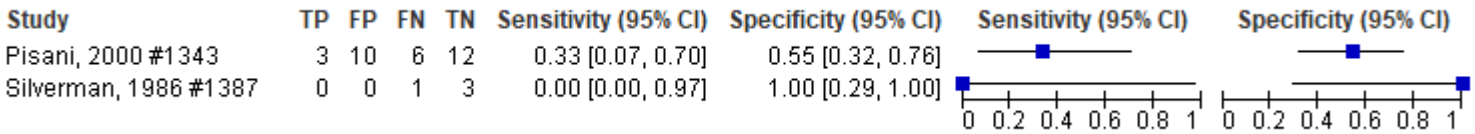


Figure 80: CB grades V and VI

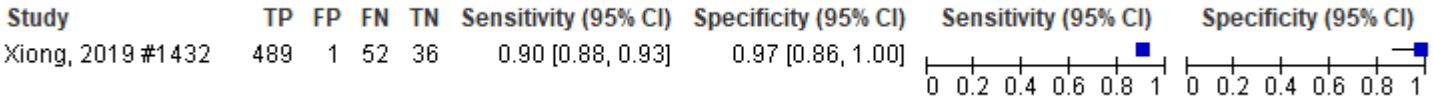


Figure 81: CB grades III, V and VI

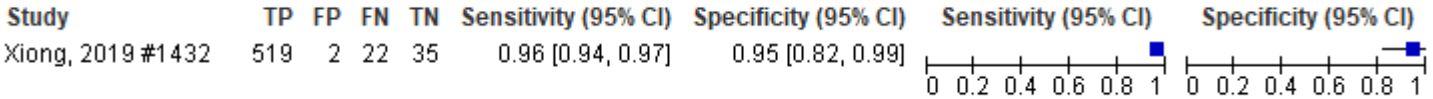


Figure 82: positive (versus negative) with CEUS guidance

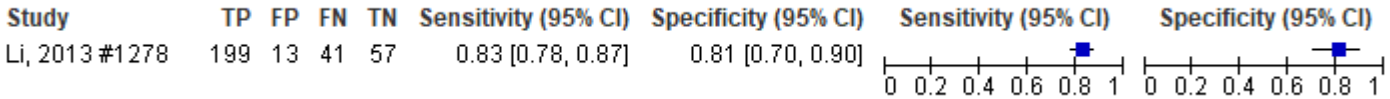
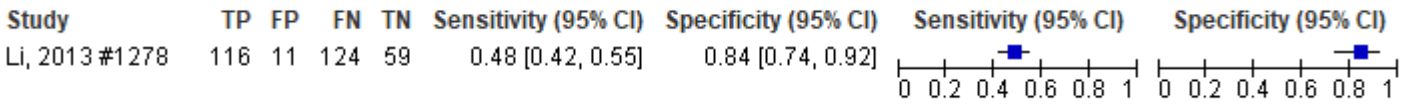


Figure 83: positive (versus negative) with US guidance



Core biopsy, with prior US

Figure 84: indeterminate, follicular neoplasm, suspicious for malignancy, or malignant

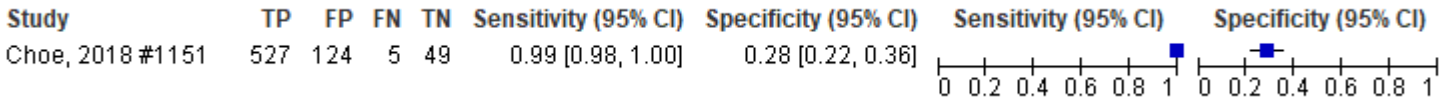


Figure 85: follicular neoplasm, suspicious for malignancy, or malignant

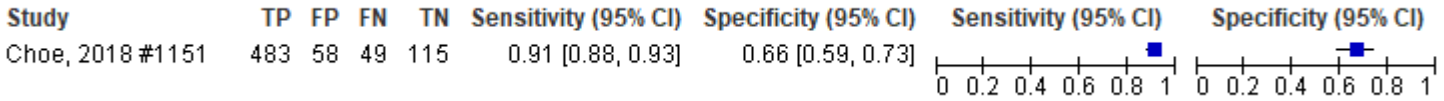
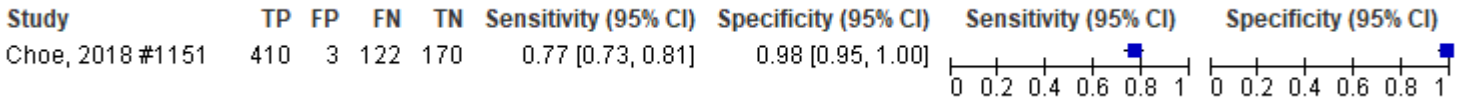


Figure 86: suspicious for malignancy, or malignant



Raw data analysis

FNAC, no ROSE, smear only, without prior US

Figure 87: Bethesda Grade III or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Acar, 2017 #1088	87	110	3	101	0.97 [0.91, 0.99]	0.48 [0.41, 0.55]
Arul, 2015 #1113	56	70	3	253	0.95 [0.86, 0.99]	0.78 [0.73, 0.83]
Aydogan, 2019 #1114	124	48	18	275	0.87 [0.81, 0.92]	0.85 [0.81, 0.89]
Bahaj, 2021 #1873	127	27	21	131	0.86 [0.79, 0.91]	0.83 [0.76, 0.88]
Choden, 2021 #1855	34	16	2	29	0.94 [0.81, 0.99]	0.64 [0.49, 0.78]
Fiorentino, 2021 #1857	408	87	6	186	0.99 [0.97, 0.99]	0.68 [0.62, 0.74]
Kim, 2013 #1257	118	11	24	47	0.83 [0.76, 0.89]	0.81 [0.69, 0.90]
Nagarajan, 2015 #1326	438	244	21	460	0.95 [0.93, 0.97]	0.65 [0.62, 0.69]
Seok, 2018 #1377	364	44	3	20	0.99 [0.98, 1.00]	0.31 [0.20, 0.44]
Son, 2014 #1392	414	34	29	187	0.93 [0.91, 0.96]	0.85 [0.79, 0.89]
Sukumaran, 2014 #1399	193	9	4	27	0.98 [0.95, 0.99]	0.75 [0.58, 0.88]
Theoharis, 2009 #1411	186	95	8	74	0.96 [0.92, 0.98]	0.44 [0.36, 0.52]
Theoharis, 2013 #1410	177	69	16	101	0.92 [0.87, 0.95]	0.59 [0.52, 0.67]

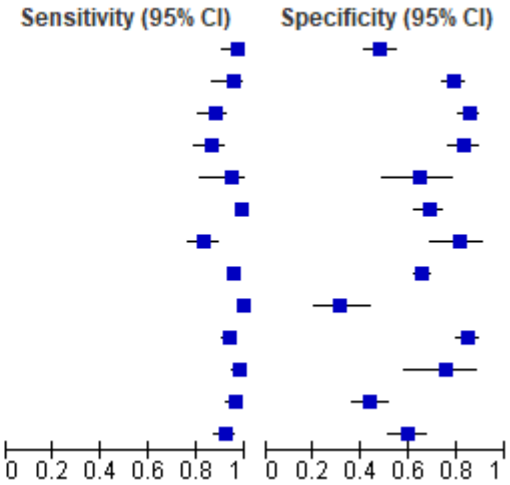


Figure 88: Bethesda Grade IV or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Acar, 2017 #1088	82	46	8	164	0.91 [0.83, 0.96]	0.78 [0.72, 0.83]
Arul, 2015 #1113	46	39	13	284	0.78 [0.65, 0.88]	0.88 [0.84, 0.91]
Aydogan, 2019 #1114	110	17	32	306	0.77 [0.70, 0.84]	0.95 [0.92, 0.97]
Bahaj, 2021 #1873	92	11	56	147	0.62 [0.54, 0.70]	0.93 [0.88, 0.96]
Choden, 2021 #1855	33	10	3	35	0.92 [0.78, 0.98]	0.78 [0.63, 0.89]
Fiorentino, 2021 #1857	402	45	12	228	0.97 [0.95, 0.98]	0.84 [0.79, 0.88]
Hosokawa, 2019 #1234	222	21	50	392	0.82 [0.76, 0.86]	0.95 [0.92, 0.97]
Nagarajan, 2015 #1326	354	104	105	600	0.77 [0.73, 0.81]	0.85 [0.82, 0.88]
Seok, 2018 #1377	319	4	48	60	0.87 [0.83, 0.90]	0.94 [0.85, 0.98]
Son, 2014 #1392	409	30	34	191	0.92 [0.89, 0.95]	0.86 [0.81, 0.91]
Sukumaran, 2014 #1399	187	4	10	32	0.95 [0.91, 0.98]	0.89 [0.74, 0.97]
Theoharis, 2009 #1411	173	81	21	88	0.89 [0.84, 0.93]	0.52 [0.44, 0.60]
Theoharis, 2013 #1410	169	58	24	112	0.88 [0.82, 0.92]	0.66 [0.58, 0.73]

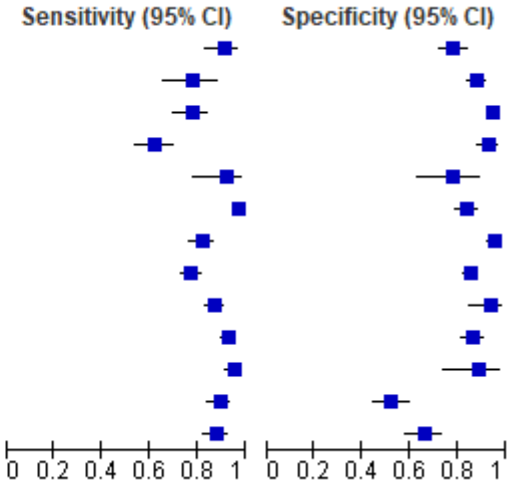


Figure 89: Bethesda Grade V or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Acar, 2017 #1088	75	9	15	202	0.83 [0.74, 0.90]	0.96 [0.92, 0.98]
Arul, 2015 #1113	33	7	26	316	0.56 [0.42, 0.69]	0.98 [0.96, 0.99]
Aydogan, 2019 #1114	95	2	47	321	0.67 [0.59, 0.75]	0.99 [0.98, 1.00]
Bahaj, 2021 #1873	86	4	62	154	0.58 [0.50, 0.66]	0.97 [0.94, 0.99]
Choden, 2021 #1855	28	1	8	44	0.78 [0.61, 0.90]	0.98 [0.88, 1.00]
Fiorentino, 2021 #1857	387	2	27	271	0.93 [0.91, 0.96]	0.99 [0.97, 1.00]
Kim, 2013 #1257	103	4	39	54	0.73 [0.64, 0.80]	0.93 [0.83, 0.98]
Li, 2021 #1865	452	8	56	107	0.89 [0.86, 0.92]	0.93 [0.87, 0.97]
Nagarajan, 2015 #1326	321	21	138	683	0.70 [0.66, 0.74]	0.97 [0.95, 0.98]
Rammeh, 2019 #1349	20	6	4	34	0.83 [0.63, 0.95]	0.85 [0.70, 0.94]
Rana, 2021 #1350	89	3	16	337	0.85 [0.76, 0.91]	0.99 [0.97, 1.00]
Seok, 2018 #1377	316	0	51	64	0.86 [0.82, 0.89]	1.00 [0.94, 1.00]
Son, 2014 #1392	348	8	95	213	0.79 [0.74, 0.82]	0.96 [0.93, 0.98]
Sukumaran, 2014 #1399	158	0	39	36	0.80 [0.74, 0.86]	1.00 [0.90, 1.00]
Theoharis, 2009 #1411	138	4	56	165	0.71 [0.64, 0.77]	0.98 [0.94, 0.99]
Theoharis, 2013 #1410	144	4	49	166	0.75 [0.68, 0.81]	0.98 [0.94, 0.99]

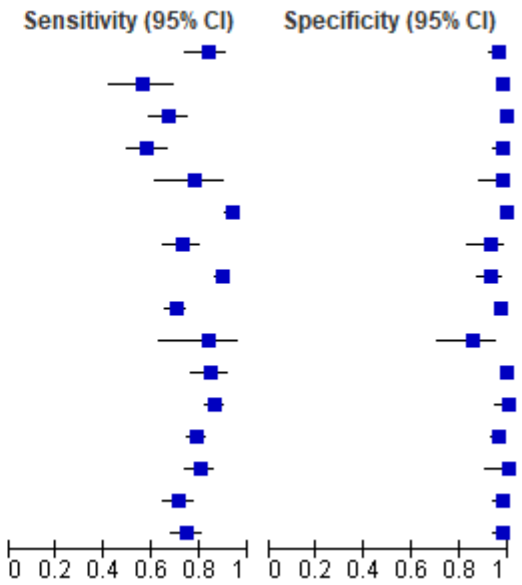


Figure 90: Bethesda Grade VI

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Acar, 2017 #1088	28	1	62	210	0.31 [0.22, 0.42]	1.00 [0.97, 1.00]
Arul, 2015 #1113	16	0	43	323	0.27 [0.16, 0.40]	1.00 [0.99, 1.00]
Aydogan, 2019 #1114	74	0	68	323	0.52 [0.44, 0.61]	1.00 [0.99, 1.00]
Bahaj, 2021 #1873	17	0	131	158	0.11 [0.07, 0.18]	1.00 [0.98, 1.00]
Choden, 2021 #1855	21	0	15	45	0.58 [0.41, 0.74]	1.00 [0.92, 1.00]
Fiorentino, 2021 #1857	250	0	164	273	0.60 [0.55, 0.65]	1.00 [0.99, 1.00]
Nagarajan, 2015 #1326	242	2	217	702	0.53 [0.48, 0.57]	1.00 [0.99, 1.00]
Seok, 2018 #1377	246	0	121	64	0.67 [0.62, 0.72]	1.00 [0.94, 1.00]
Son, 2014 #1392	273	2	170	217	0.62 [0.57, 0.66]	0.99 [0.97, 1.00]
Sukumaran, 2014 #1399	148	0	49	36	0.75 [0.68, 0.81]	1.00 [0.90, 1.00]
Theoharis, 2009 #1411	112	0	82	169	0.58 [0.50, 0.65]	1.00 [0.98, 1.00]
Theoharis, 2013 #1410	118	0	75	170	0.61 [0.54, 0.68]	1.00 [0.98, 1.00]

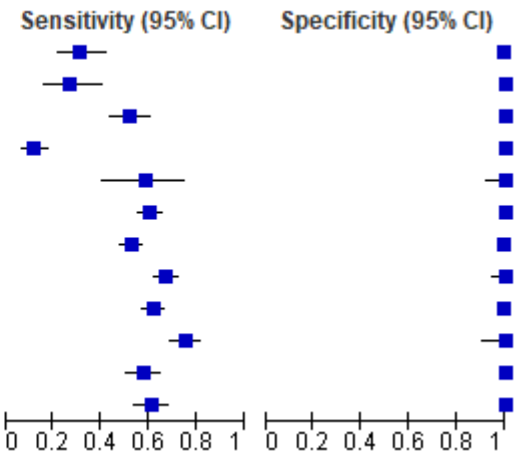


Figure 91: BTA THY 3a or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abou-foul, 2021 #1872	59	56	28	163	0.68 [0.57, 0.77]	0.74 [0.68, 0.80]
Mandal, 2011 #1293	27	12	3	66	0.90 [0.73, 0.98]	0.85 [0.75, 0.92]

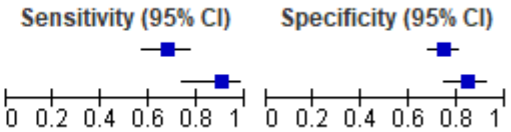


Figure 92: BTA THY 3f or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abou-foul, 2021 #1872	45	22	42	197	0.52 [0.41, 0.63]	0.90 [0.85, 0.94]

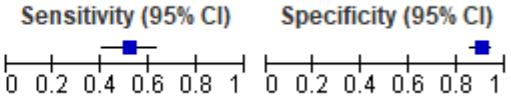


Figure 93: BTA THY 4 or above

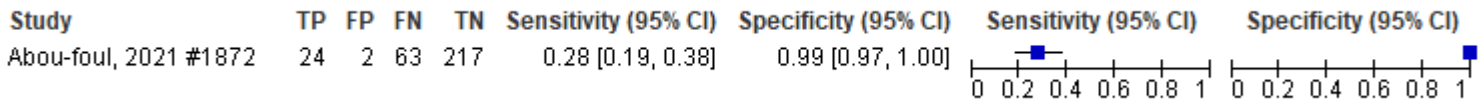


Figure 94: BTA THY 5

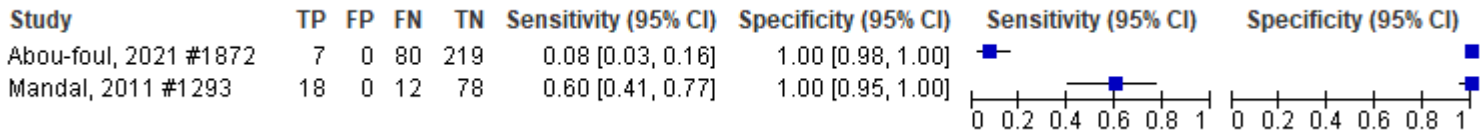


Figure 95: AC 3 or above

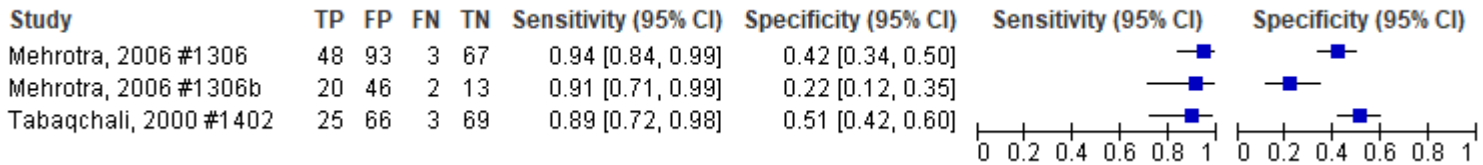


Figure 96: AC 4 or above

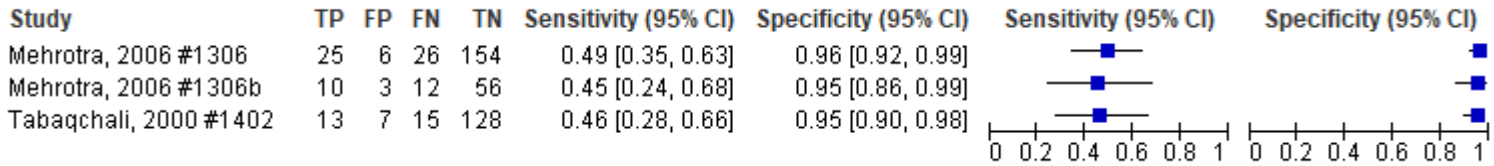


Figure 97: 2 way: malignant v benign

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bashier, 1996 #1121	11	12	1	65	0.92 [0.62, 1.00]	0.84 [0.74, 0.92]
Lukitto, 1998 #1286	4	1	12	150	0.25 [0.07, 0.52]	0.99 [0.96, 1.00]
Maruta, 2003 #1296	112	3	8	128	0.93 [0.87, 0.97]	0.98 [0.93, 1.00]
Rege, 1987 #1352	13	0	2	170	0.87 [0.60, 0.98]	1.00 [0.98, 1.00]
Russ, 1978 #1364	8	0	3	18	0.73 [0.39, 0.94]	1.00 [0.81, 1.00]
Schwartz, 1982 #1373	5	3	6	78	0.45 [0.17, 0.77]	0.96 [0.90, 0.99]
Seya, 1990 #1379	11	0	2	13	0.85 [0.55, 0.98]	1.00 [0.75, 1.00]
Silverman, 1986 #1387	1	0	1	6	0.50 [0.01, 0.99]	1.00 [0.54, 1.00]
Takashima, 1992 #1404	16	0	0	11	1.00 [0.79, 1.00]	1.00 [0.72, 1.00]
Takashima, 1994 #1403	64	3	3	29	0.96 [0.87, 0.99]	0.91 [0.75, 0.98]
Takashima, 1994 #1403b	21	1	3	9	0.88 [0.68, 0.97]	0.90 [0.55, 1.00]
Takashima,1992 #1404b	6	0	2	6	0.75 [0.35, 0.97]	1.00 [0.54, 1.00]
Yavuz, 2020 #1436	17	2	1	14	0.94 [0.73, 1.00]	0.88 [0.62, 0.98]

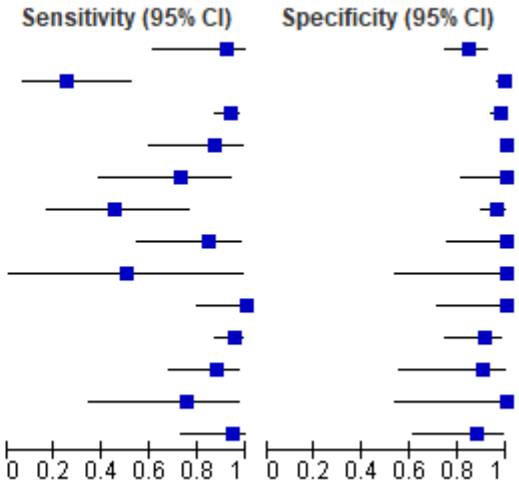


Figure 98: 3 way: suspicious or malignant (negative =benign)

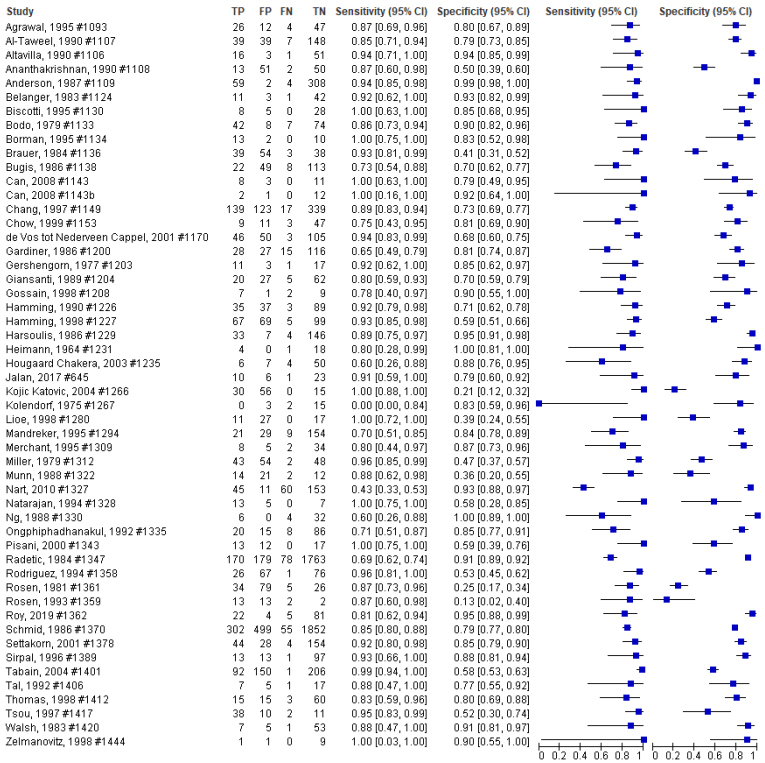


Figure 99: 3 way: malignant (negative = suspicious or benign)

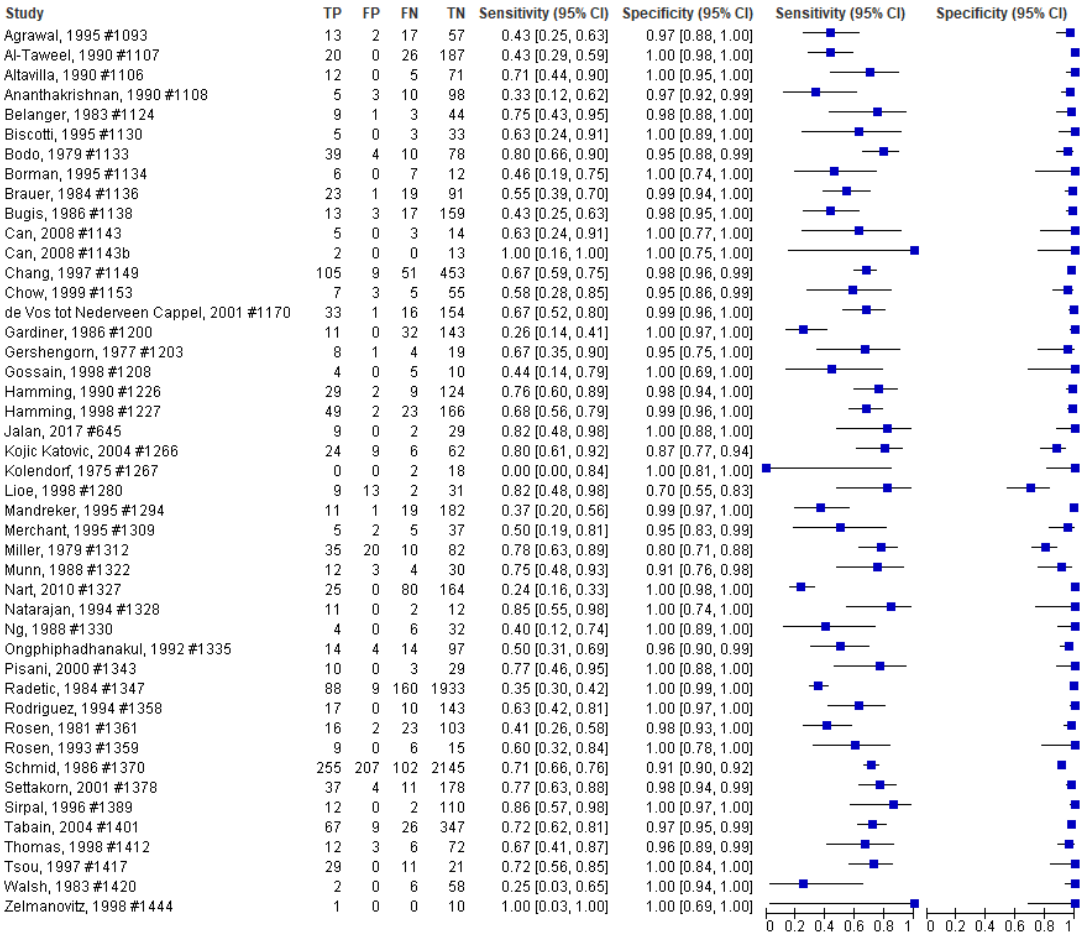


Figure 100: 4 way: malignant or suspicious or indeterminate (negative = benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abboud, 2003 #1087	15	23	0	8	1.00 [0.78, 1.00]	0.26 [0.12, 0.45]
Al-Hureibi, 2003 #1104	15	24	22	135	0.41 [0.25, 0.58]	0.85 [0.78, 0.90]
de Roy van Zuidewijn, 1994 #1169	80	59	6	115	0.93 [0.85, 0.97]	0.66 [0.59, 0.73]
Haberal, 2009 #1220	59	31	4	166	0.94 [0.85, 0.98]	0.84 [0.78, 0.89]
Kini, 1985 #1262	93	179	6	101	0.94 [0.87, 0.98]	0.36 [0.30, 0.42]
Mamoon, 1997 #1290	13	16	1	146	0.93 [0.66, 1.00]	0.90 [0.84, 0.94]
Okumura, 1999 #1334	46	49	4	10	0.92 [0.81, 0.98]	0.17 [0.08, 0.29]
Schoedel, 2008 #1372	14	4	6	18	0.70 [0.46, 0.88]	0.82 [0.60, 0.95]
Scurry, 2000 #1375	23	27	7	13	0.77 [0.58, 0.90]	0.33 [0.19, 0.49]
Spiliotis, 1992 #1394	28	32	3	128	0.90 [0.74, 0.98]	0.80 [0.73, 0.86]
Wu, 2006 #1428	99	126	11	148	0.90 [0.83, 0.95]	0.54 [0.48, 0.60]
Zbar, 2009 #1443	3	10	5	45	0.38 [0.09, 0.76]	0.82 [0.69, 0.91]

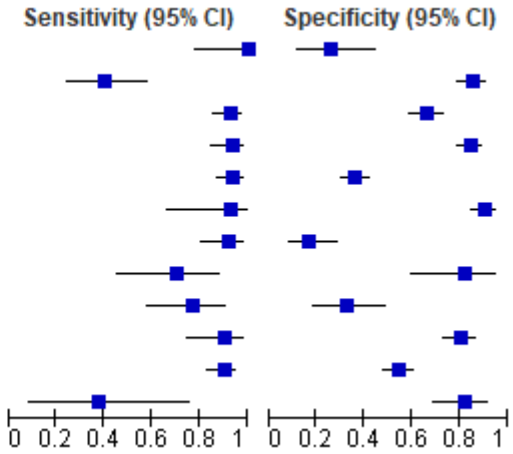


Figure 101: 4 way: malignant or suspicious (negative = benign or indeterminate)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abboud, 2003 #1087	13	7	2	24	0.87 [0.60, 0.98]	0.77 [0.59, 0.90]
Aggarwal, 1989 #693	16	2	0	18	1.00 [0.79, 1.00]	0.90 [0.68, 0.99]
Al-Hureibi, 2003 #1104	6	2	31	157	0.16 [0.06, 0.32]	0.99 [0.96, 1.00]
de Roy van Zuidewijn, 1994 #1169	68	15	18	159	0.79 [0.69, 0.87]	0.91 [0.86, 0.95]
Haberal, 2009 #1220	53	18	10	179	0.84 [0.73, 0.92]	0.91 [0.86, 0.94]
Kini, 1985 #1262	64	50	35	230	0.65 [0.54, 0.74]	0.82 [0.77, 0.86]
Mamoon, 1997 #1290	11	8	3	154	0.79 [0.49, 0.95]	0.95 [0.91, 0.98]
Okumura, 1999 #1334	25	9	25	50	0.50 [0.36, 0.64]	0.85 [0.73, 0.93]
Raina, 2011 #1348	5	1	2	17	0.71 [0.29, 0.96]	0.94 [0.73, 1.00]
Schoedel, 2008 #1372	8	0	12	22	0.40 [0.19, 0.64]	1.00 [0.85, 1.00]
Scurry, 2000 #1375	10	3	20	37	0.33 [0.17, 0.53]	0.93 [0.80, 0.98]
Spiliotis, 1992 #1394	25	20	6	140	0.81 [0.63, 0.93]	0.88 [0.81, 0.92]
Wu, 2006 #1428	92	82	18	192	0.84 [0.75, 0.90]	0.70 [0.64, 0.75]
Zbar, 2009 #1443	3	3	5	52	0.38 [0.09, 0.76]	0.95 [0.85, 0.99]

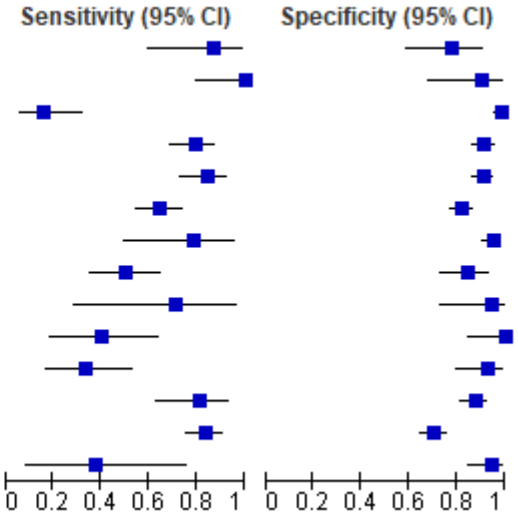


Figure 102: 4 way: malignant (negative = benign or indeterminate or suspicious)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abboud, 2003 #1087	11	2	4	29	0.73 [0.45, 0.92]	0.94 [0.79, 0.99]
Aggarwal, 1989 #693	12	0	4	20	0.75 [0.48, 0.93]	1.00 [0.83, 1.00]
Al-Hureibi, 2003 #1104	4	0	33	159	0.11 [0.03, 0.25]	1.00 [0.98, 1.00]
de Roy van Zuidewijn, 1994 #1169	57	2	29	172	0.66 [0.55, 0.76]	0.99 [0.96, 1.00]
Haberal, 2009 #1220	41	1	22	196	0.65 [0.52, 0.77]	0.99 [0.97, 1.00]
Kini, 1985 #1262	53	15	46	265	0.54 [0.43, 0.64]	0.95 [0.91, 0.97]
Mamoon, 1997 #1290	6	2	8	160	0.43 [0.18, 0.71]	0.99 [0.96, 1.00]
Okumura, 1999 #1334	10	0	40	59	0.20 [0.10, 0.34]	1.00 [0.94, 1.00]
Scurry, 2000 #1375	2	1	28	39	0.07 [0.01, 0.22]	0.97 [0.87, 1.00]
Spiliotis, 1992 #1394	17	0	13	160	0.57 [0.37, 0.75]	1.00 [0.98, 1.00]
Wu, 2006 #1428	34	0	76	274	0.31 [0.22, 0.40]	1.00 [0.99, 1.00]
Zbar, 2009 #1443	1	1	7	54	0.13 [0.00, 0.53]	0.98 [0.90, 1.00]

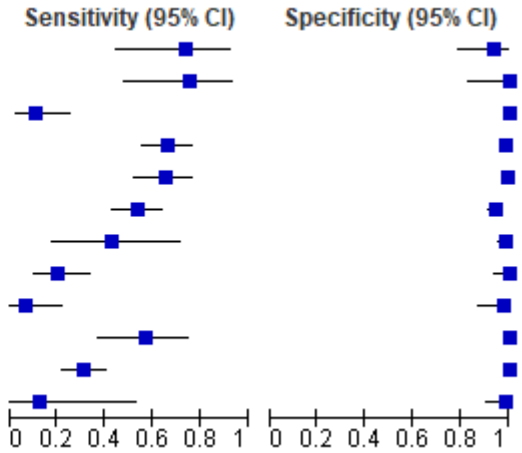


Figure 103: 5 way: malignant or suspicious or two grades of indeterminate (negative = benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Acar, 2017 #1088	23	64	3	99	0.88 [0.70, 0.98]	0.61 [0.53, 0.68]
Francis, 1999 #1192	17	9	2	13	0.89 [0.67, 0.99]	0.59 [0.36, 0.79]
Kelman, 2001 #1250	91	74	6	104	0.94 [0.87, 0.98]	0.58 [0.51, 0.66]
La Rosa, 1991 #1273	241	301	6	257	0.98 [0.95, 0.99]	0.46 [0.42, 0.50]
Prinz, 1983 #1345	17	22	1	38	0.94 [0.73, 1.00]	0.63 [0.50, 0.75]
Theoharis, 2013 #1410	168	81	12	85	0.93 [0.89, 0.97]	0.51 [0.43, 0.59]

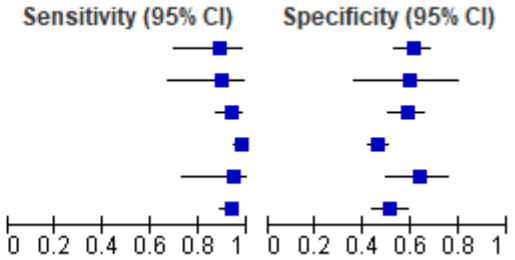


Figure 104: 5 way: malignant or suspicious or one grade of indeterminate (negative = lower grade of indeterminate or benign)

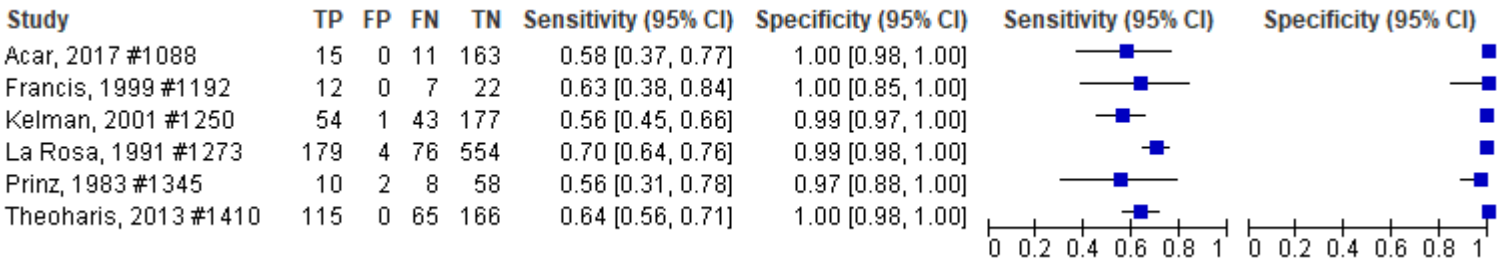


Figure 105: 5 way: malignant (negative = suspicious or two grades of indeterminate or benign)

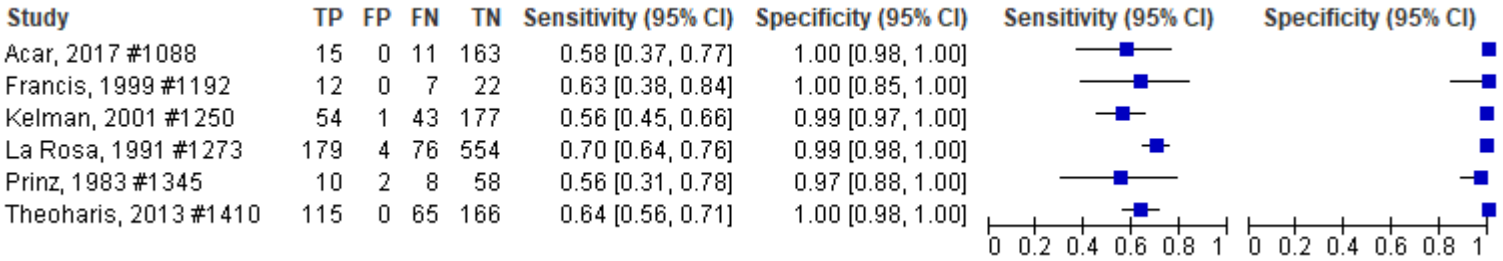


Figure 106: 1 or more inclusions

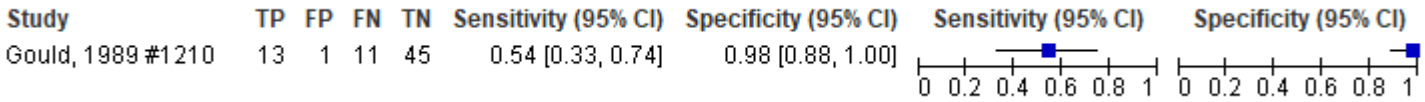


Figure 107: 1 or more grooves

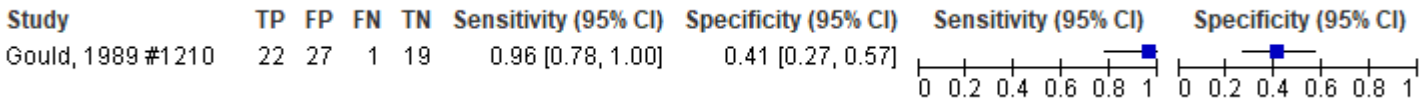


Figure 108: 2 or more grooves

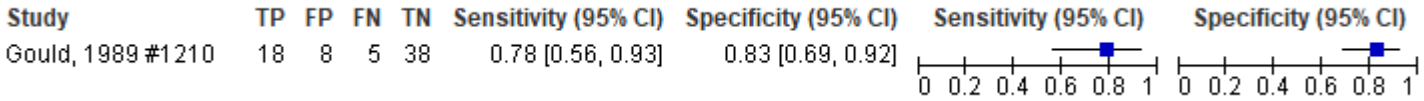
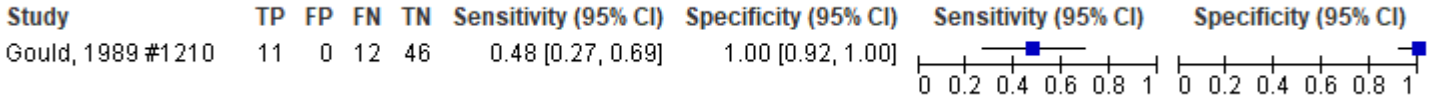


Figure 109: 3 or more grooves



FNAC, no ROSE, smear only, with prior US

Figure 110: Bethesda Grade III or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Huang, 2020 #796	228	121	1	35	1.00 [0.98, 1.00]	0.22 [0.16, 0.30]
Ozdemir, 2017 #1336	339	625	61	2750	0.85 [0.81, 0.88]	0.81 [0.80, 0.83]
Wang, 2020 #1421	99	58	6	93	0.94 [0.88, 0.98]	0.62 [0.53, 0.69]

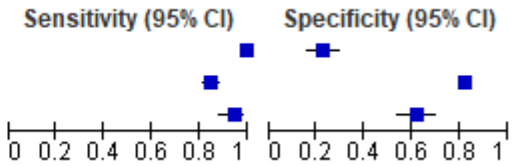


Figure 111: Bethesda Grade IV or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Huang, 2020 #796	218	30	11	126	0.95 [0.92, 0.98]	0.81 [0.74, 0.87]
Ozdemir, 2017 #1336	223	84	177	3291	0.56 [0.51, 0.61]	0.98 [0.97, 0.98]
Wang, 2020 #1421	74	20	31	131	0.70 [0.61, 0.79]	0.87 [0.80, 0.92]

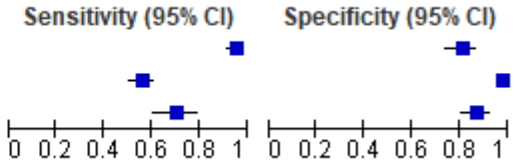


Figure 112: Bethesda Grade V or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Huang, 2020 #796	123	1	106	155	0.54 [0.47, 0.60]	0.99 [0.96, 1.00]
Ozdemir, 2017 #1336	204	37	196	3338	0.51 [0.46, 0.56]	0.99 [0.98, 0.99]
Wang, 2020 #1421	73	13	32	138	0.70 [0.60, 0.78]	0.91 [0.86, 0.95]

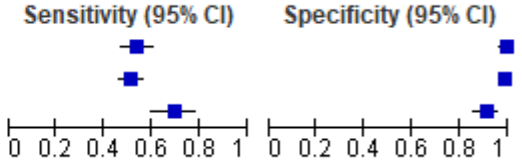


Figure 113: Bethesda Grade VI or above

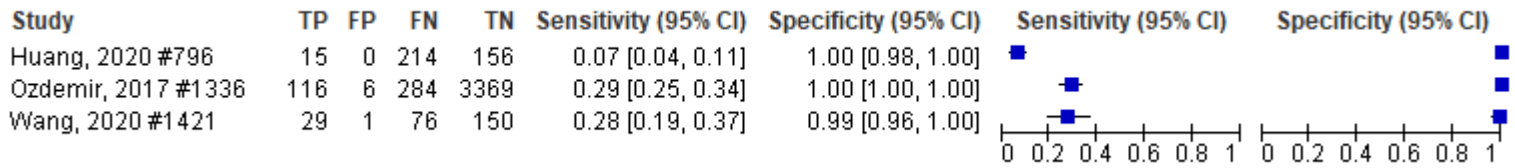


Figure 114: 2 way: malignant versus benign

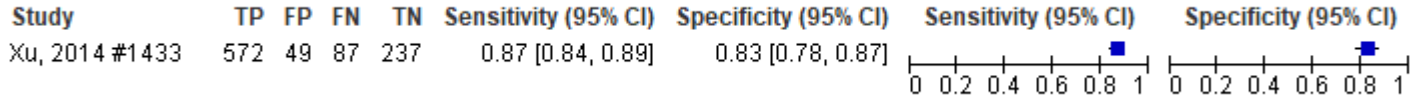


Figure 115: 3 way: suspicious or malignant (negative = benign)

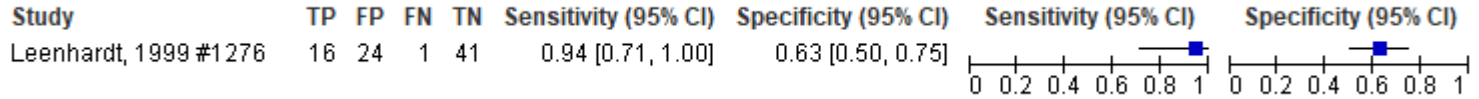


Figure 116: 3 way: malignant (negative = suspicious or benign)

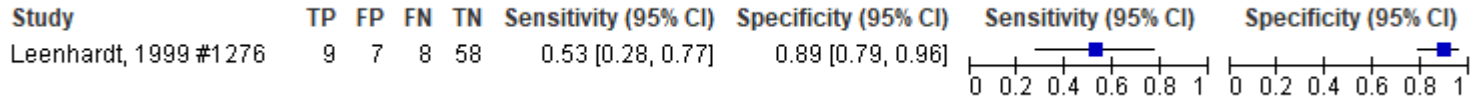


Figure 117: 4 way De May classification: malignant, suspicious, non malignant follicular proliferation (negative = benign)

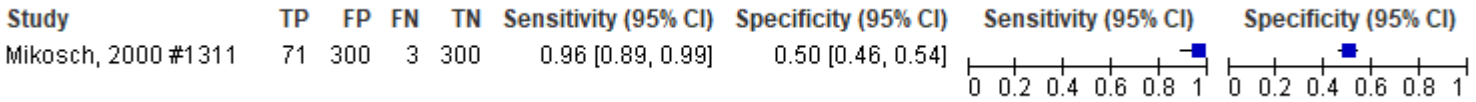


Figure 118: 4 way De May classification: malignant, suspicious (negative = benign, non malignant follicular proliferation)

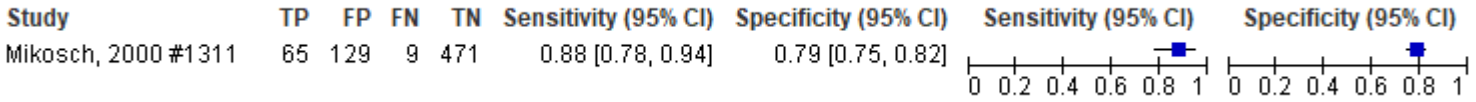


Figure 119: 4 way De May classification: malignant (negative = benign, non malignant follicular proliferation, suspicious)

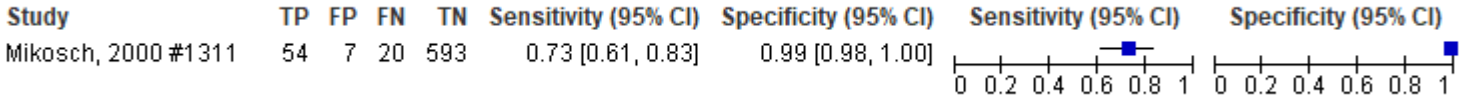


Figure 120: 4 way Piana classification: C3 or more

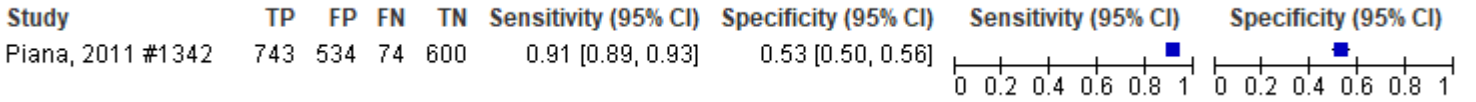


Figure 121: 4 way Piana classification: C4 or more

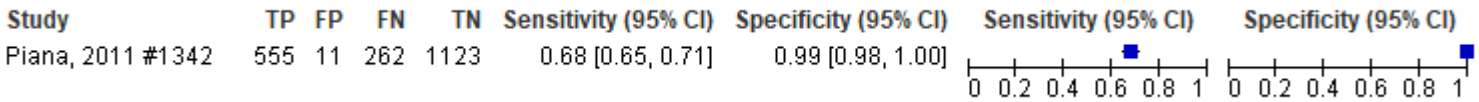


Figure 122: 4 way Piana classification: C5 or more

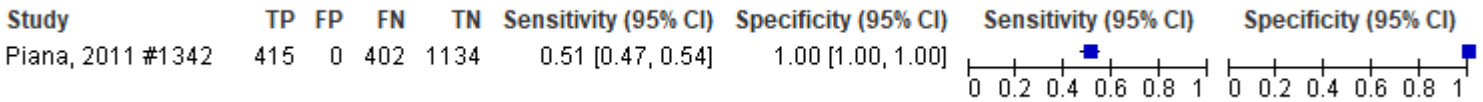


Figure 123: 4 way generic: malignant, suspicious, indeterminate (benign = negative)

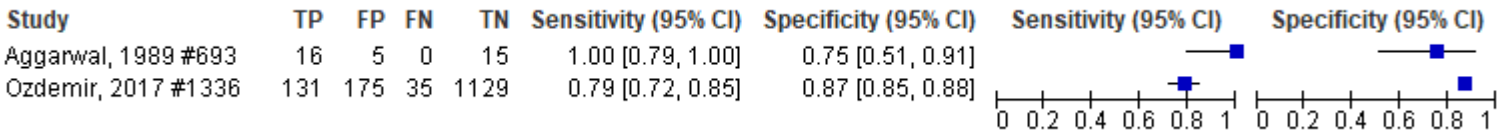
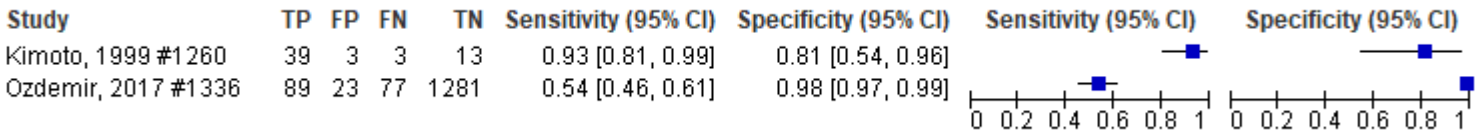


Figure 124: 4 way generic: malignant, suspicious, indeterminate (benign = negative)



FNAC, no ROSE, smear, with cytospin and/or cell-block, without prior US

Figure 125: Bethesda Grade III or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Mastorakis, 2014 #1299	197	43	9	236	0.96 [0.92, 0.98]	0.85 [0.80, 0.89]
Mastorakis, 2014 #1299b	77	36	3	358	0.96 [0.89, 0.99]	0.91 [0.88, 0.94]
McElroy, 2014 #1303	9	2	0	10	1.00 [0.66, 1.00]	0.83 [0.52, 0.98]
Nagarajan, 2015 #1326	25	13	1	13	0.96 [0.80, 1.00]	0.50 [0.30, 0.70]
Naz, 2014 #1329	9	7	5	40	0.64 [0.35, 0.87]	0.85 [0.72, 0.94]

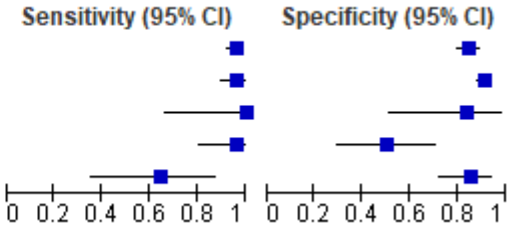


Figure 126: Bethesda Grade IV or above

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Mastorakis, 2014 #1299	186	7	20	272	0.90 [0.85, 0.94]	0.97 [0.95, 0.99]
Mastorakis, 2014 #1299b	75	13	5	381	0.94 [0.86, 0.98]	0.97 [0.94, 0.98]
McElroy, 2014 #1303	7	1	2	11	0.78 [0.40, 0.97]	0.92 [0.62, 1.00]
Nagarajan, 2015 #1326	21	2	5	24	0.81 [0.61, 0.93]	0.92 [0.75, 0.99]
Naz, 2014 #1329	7	3	7	44	0.50 [0.23, 0.77]	0.94 [0.82, 0.99]

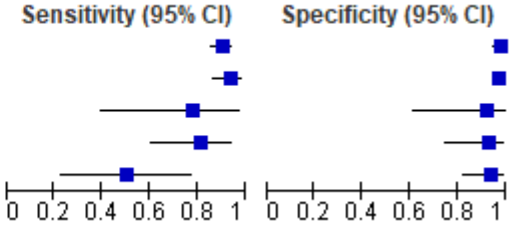


Figure 127: Bethesda Grade V or above

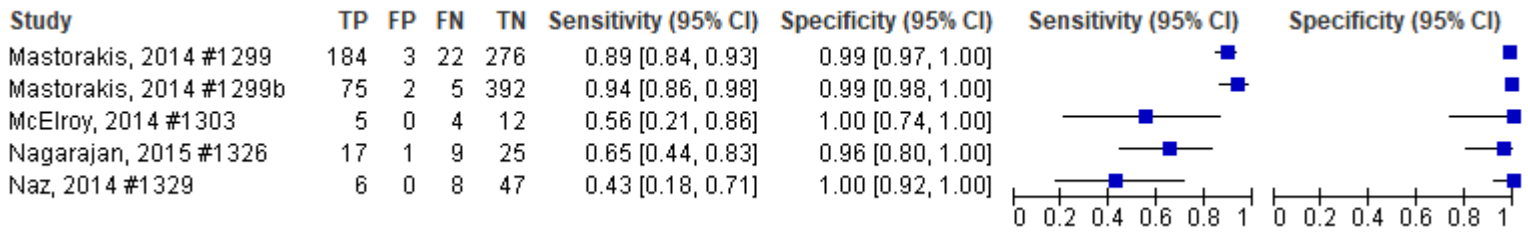


Figure 128: Bethesda Grade VI or above

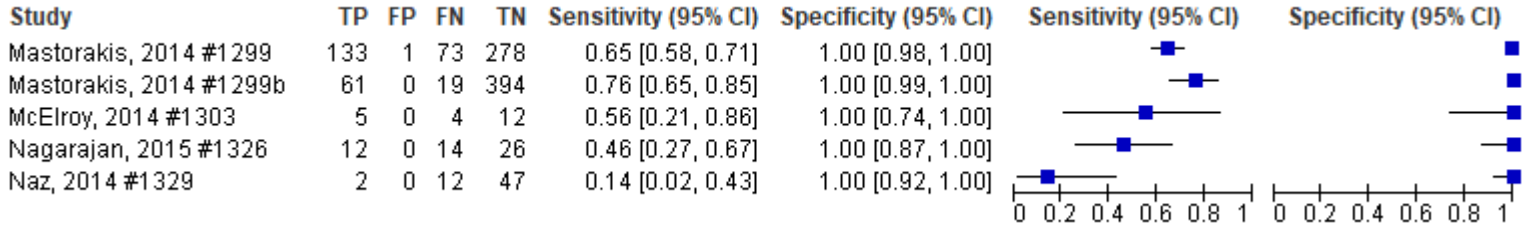


Figure 129: 2 way: malignant v benign

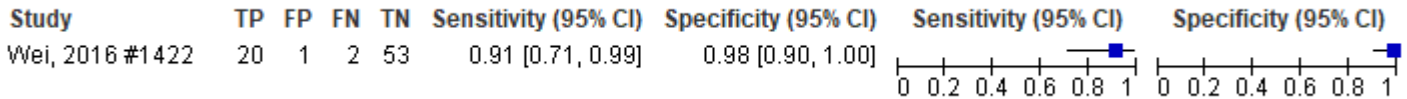


Figure 130: 3 way: malignant or suspicious (negative = benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Aguilar-Diosdado, 1997 #1094	43	33	19	167	0.69 [0.56, 0.80]	0.83 [0.78, 0.88]
Biscotti, 1995 #1130	8	7	0	26	1.00 [0.63, 1.00]	0.79 [0.61, 0.91]
Cristallini, 1989 #1161	15	7	1	16	0.94 [0.70, 1.00]	0.70 [0.47, 0.87]
Danese, 1998 #1164	99	126	3	307	0.97 [0.92, 0.99]	0.71 [0.66, 0.75]
Danese, 1998 #1164b	79	136	7	300	0.92 [0.84, 0.97]	0.69 [0.64, 0.73]
Dwarakanathan, 1989 #1176	18	19	1	25	0.95 [0.74, 1.00]	0.57 [0.41, 0.72]
Ferrari, 1985 #1184	7	16	0	43	1.00 [0.59, 1.00]	0.73 [0.60, 0.84]
Kumar, 1992 #1272	12	15	1	52	0.92 [0.64, 1.00]	0.78 [0.66, 0.87]
Pepper, 1989 #1340	5	8	1	7	0.83 [0.36, 1.00]	0.47 [0.21, 0.73]
Petersen, 1984 #1341	19	44	1	84	0.95 [0.75, 1.00]	0.66 [0.57, 0.74]
Rubenfeld, 1982 #1363	15	11	0	4	1.00 [0.78, 1.00]	0.27 [0.08, 0.55]
Vojvodich, 1994 #1419	29	6	6	44	0.83 [0.66, 0.93]	0.88 [0.76, 0.95]
Zajdela, 1987 #1442	116	31	10	215	0.92 [0.86, 0.96]	0.87 [0.83, 0.91]

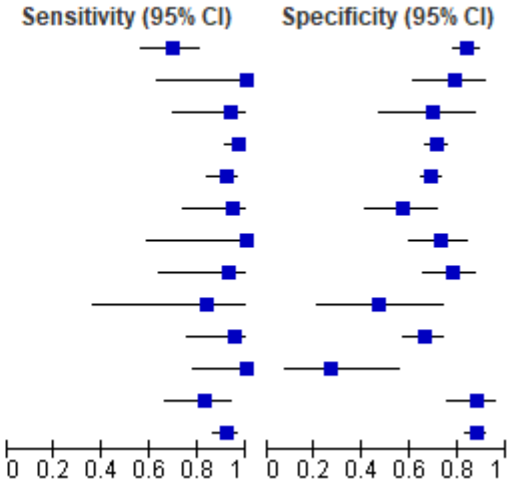


Figure 131: 3 way: malignant (negative = benign or suspicious)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Aguilar-Diosdado, 1997 #1094	24	5	38	195	0.39 [0.27, 0.52]	0.97 [0.94, 0.99]
Biscotti, 1995 #1130	5	0	3	33	0.63 [0.24, 0.91]	1.00 [0.89, 1.00]
Cristallini, 1989 #1161	15	0	1	23	0.94 [0.70, 1.00]	1.00 [0.85, 1.00]
Danese, 1998 #1164	70	0	32	433	0.69 [0.59, 0.77]	1.00 [0.99, 1.00]
Danese, 1998 #1164b	53	2	33	434	0.62 [0.51, 0.72]	1.00 [0.98, 1.00]
Dwarakanathan, 1989 #1176	15	1	4	43	0.79 [0.54, 0.94]	0.98 [0.88, 1.00]
Ferrari, 1985 #1184	6	0	1	59	0.86 [0.42, 1.00]	1.00 [0.94, 1.00]
Kumar, 1992 #1272	8	1	5	66	0.62 [0.32, 0.86]	0.99 [0.92, 1.00]
Vojvodich, 1994 #1419	14	0	21	50	0.40 [0.24, 0.58]	1.00 [0.93, 1.00]
Zajdela, 1987 #1442	94	3	32	243	0.75 [0.66, 0.82]	0.99 [0.96, 1.00]

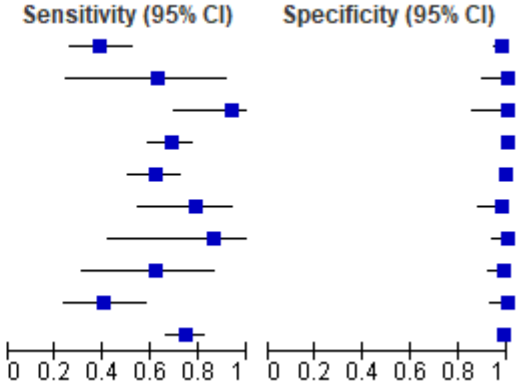


Figure 132: 4 way: malignant, suspicious, indeterminate (negative = benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bellantone, 2004 #1125	17	61	2	28	0.89 [0.67, 0.99]	0.31 [0.22, 0.42]
Liel, 1985 #1279	11	9	1	20	0.92 [0.62, 1.00]	0.69 [0.49, 0.85]
Mijovic, 2009 #1310	63	23	6	14	0.91 [0.82, 0.97]	0.38 [0.22, 0.55]
Scurry, 2000 #1375	22	32	4	3	0.85 [0.65, 0.96]	0.09 [0.02, 0.23]
Varhaug, 1981 #1418	52	48	9	112	0.85 [0.74, 0.93]	0.70 [0.62, 0.77]

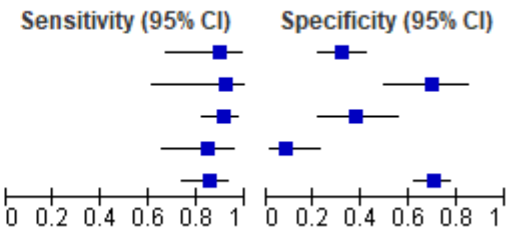


Figure 133: 4 way: malignant, suspicious (negative = benign, indeterminate)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bellantone, 2004 #1125	16	50	3	39	0.84 [0.60, 0.97]	0.44 [0.33, 0.55]
Hawkins, 1987 #1230	63	16	10	326	0.86 [0.76, 0.93]	0.95 [0.93, 0.97]
Liel, 1985 #1279	9	4	3	25	0.75 [0.43, 0.95]	0.86 [0.68, 0.96]
Mijovic, 2009 #1310	39	1	30	36	0.57 [0.44, 0.68]	0.97 [0.86, 1.00]
Scurry, 2000 #1375	10	3	16	32	0.38 [0.20, 0.59]	0.91 [0.77, 0.98]
Varhaug, 1981 #1418	42	11	19	149	0.69 [0.56, 0.80]	0.93 [0.88, 0.97]

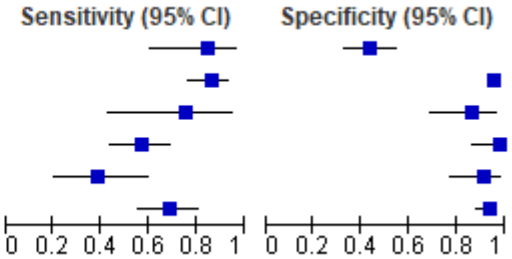


Figure 134: 4 way: malignant (negative = benign, indeterminate, suspicious)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bellantone, 2004 #1125	4	0	15	89	0.21 [0.06, 0.46]	1.00 [0.96, 1.00]
Hawkins, 1987 #1230	48	3	25	339	0.66 [0.54, 0.76]	0.99 [0.97, 1.00]
Liel, 1985 #1279	8	1	7	25	0.53 [0.27, 0.79]	0.96 [0.80, 1.00]
Scurry, 2000 #1375	2	0	24	35	0.08 [0.01, 0.25]	1.00 [0.90, 1.00]
Varhaug, 1981 #1418	26	0	35	160	0.43 [0.30, 0.56]	1.00 [0.98, 1.00]

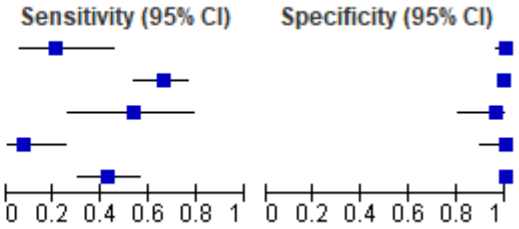
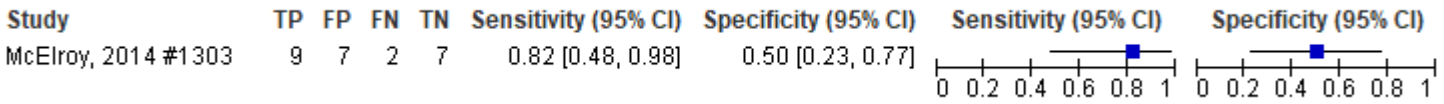


Figure 135: 5 way: malignant, suspicious, 2 grades of indeterminate (negative = benign)



FNAC, no ROSE, smear, with cytospin and/or cell-block, with prior US

Figure 136: Bethesda Grade III or above

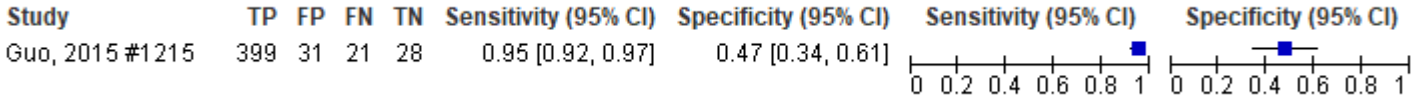


Figure 137: Bethesda Grade IV or above

Figure 138: Bethesda Grade V or above

Figure 139: Bethesda Grade VI

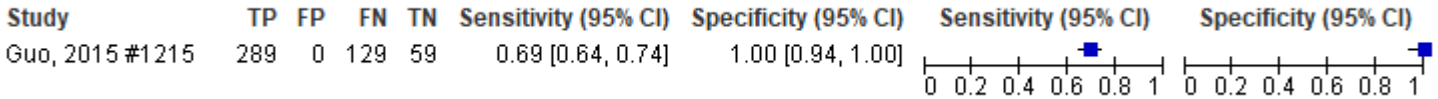
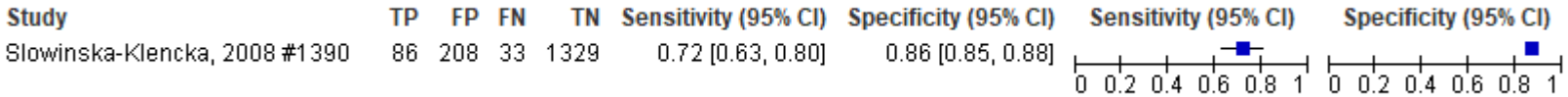


Figure 140: Benign or above



FNAC, with ROSE, smear only, without prior US

Figure 141: Bethesda Grade III or above

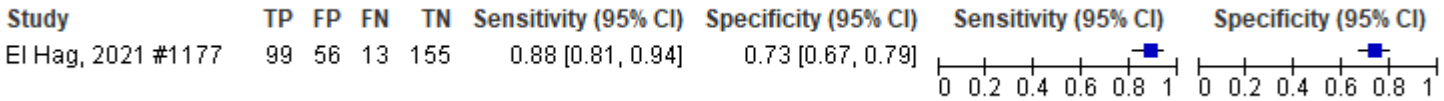


Figure 142: Bethesda Grade IV or above

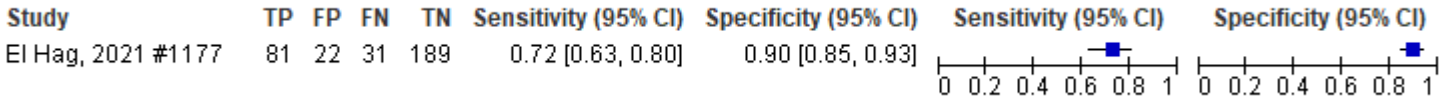


Figure 143: Bethesda Grade V or above

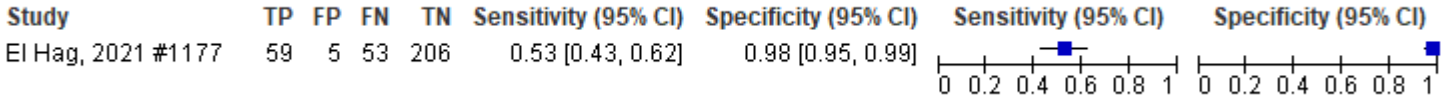


Figure 144: Bethesda Grade VI

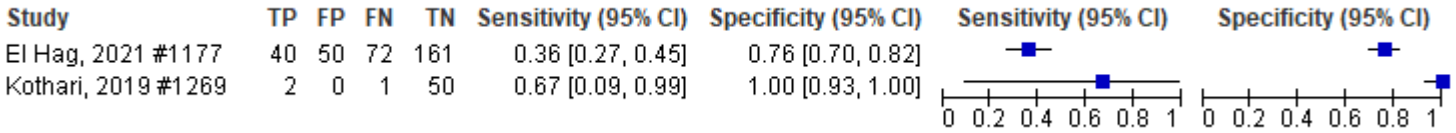


Figure 145: 3 way: malignant and suspicious (negative = benign)

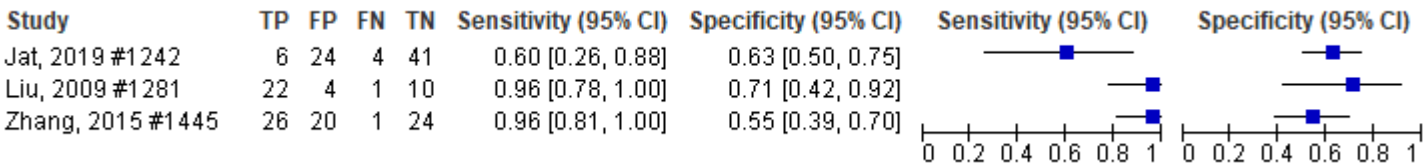


Figure 146: 3 way: malignant (negative = benign and suspicious)

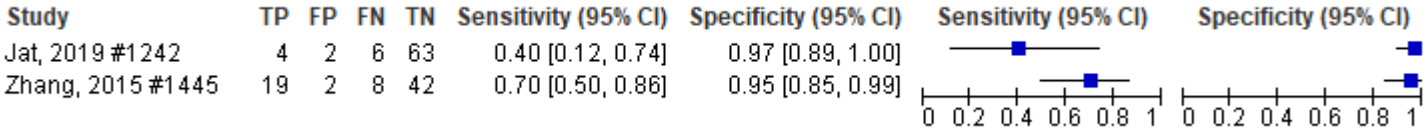


Figure 147: 4 way: malignant, suspicious, indeterminate (negative = benign)

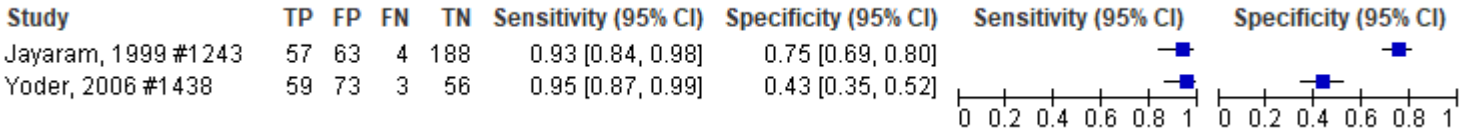


Figure 148: 4 way: malignant, suspicious (negative = benign, indeterminate)

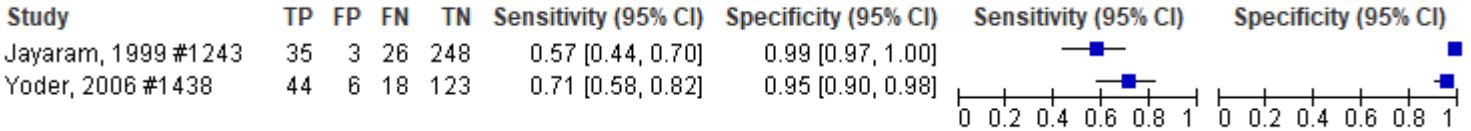
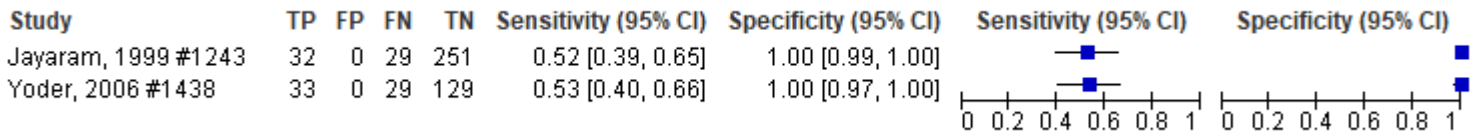
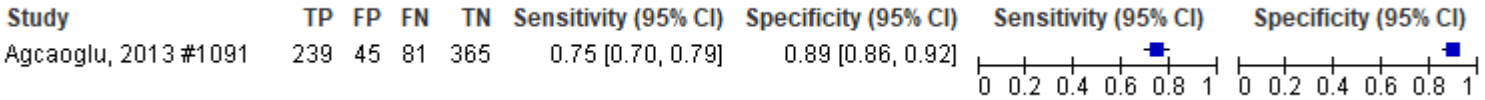


Figure 149: 4 way: malignant (negative = benign, indeterminate, suspicious)



FNAC, with ROSE, smear only, with prior US

Figure 150: intermediate or malignant



FNAC, with ROSE, smear, with cytospin and/or cell-block, without prior US

Figure 151: 3 way: suspicious or malignant (negative = benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Hussain, 1993 #1238	6	8	1	72	0.86 [0.42, 1.00]	0.90 [0.81, 0.96]
Meko, 1995 #1307	13	30	5	39	0.72 [0.47, 0.90]	0.57 [0.44, 0.68]

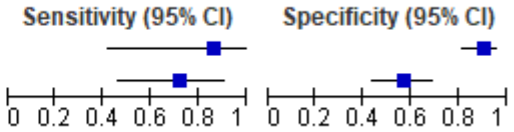


Figure 152: 3 way: malignant (negative = suspicious or benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Hussain, 1993 #1238	4	0	3	80	0.57 [0.18, 0.90]	1.00 [0.95, 1.00]

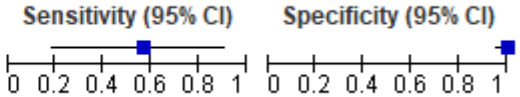


Figure 153: 4 way: malignant, suspicious, indeterminate (negative = benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Davidsohn, 1995 #1167	15	17	0	12	1.00 [0.78, 1.00]	0.41 [0.24, 0.61]

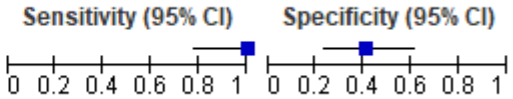


Figure 154: 4 way: malignant, suspicious (negative = benign, indeterminate)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Davidsohn, 1995 #1167	10	0	5	29	0.67 [0.38, 0.88]	1.00 [0.88, 1.00]

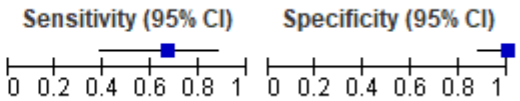


Figure 155: 5 way: malignant, suspicious, 2 grades of indeterminate (negative = benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Afroze, 2002 #1089	17	34	4	111	0.81 [0.58, 0.95]	0.77 [0.69, 0.83]		

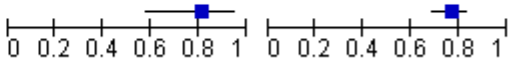


Figure 156: 5 way: malignant, suspicious (negative = 2 grades of indeterminate, benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Afroze, 2002 #1089	17	23	4	122	0.81 [0.58, 0.95]	0.84 [0.77, 0.90]		

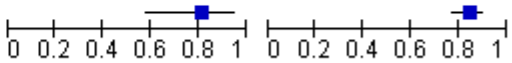


Figure 157: 5 way: malignant, suspicious (negative = suspicious, lower grade of indeterminate, benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Afroze, 2002 #1089	16	5	5	140	0.76 [0.53, 0.92]	0.97 [0.92, 0.99]		

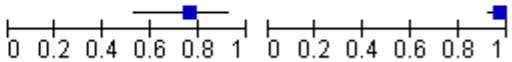
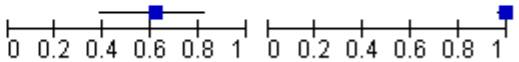


Figure 158: 5 way: malignant (negative = suspicious, 2 grades of indeterminate, benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Afroze, 2002 #1089	13	1	8	144	0.62 [0.38, 0.82]	0.99 [0.96, 1.00]		



FNAC, with ROSE, smear, with cytospin and/or cell-block, with prior US

Figure 159: indeterminate follicular, indeterminate Hurtle, Suspicious for malignancy, or positive

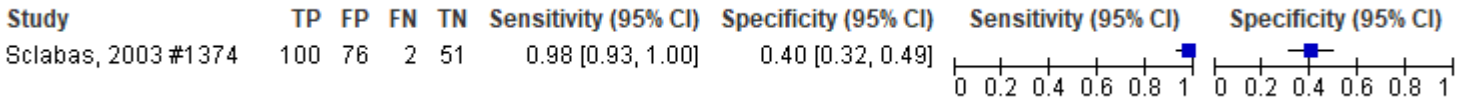


Figure 160: Suspicious for malignancy, or indeterminate follicular or positive

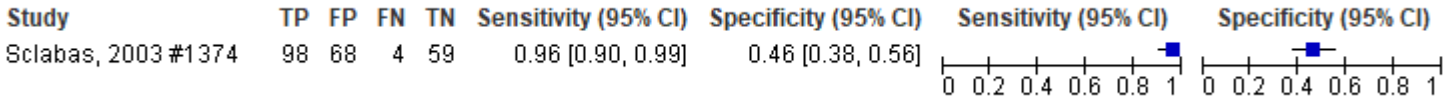


Figure 161: Suspicious for malignancy, or positive

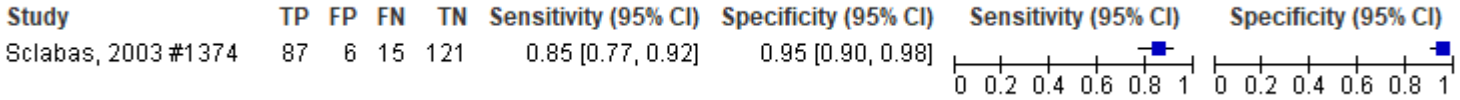
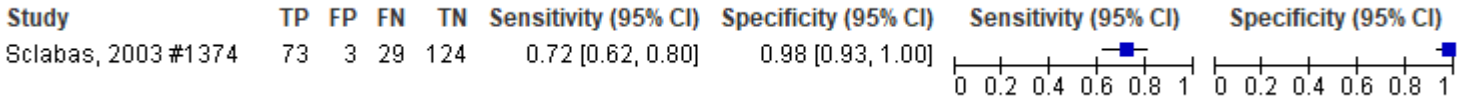


Figure 162: Positive for malignancy



Core biopsy, without prior US

Figure 163: carcinoma or neoplasm (versus benign)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Pisani, 2000 #1343	5	3	0	9	1.00 [0.48, 1.00]	0.75 [0.43, 0.95]

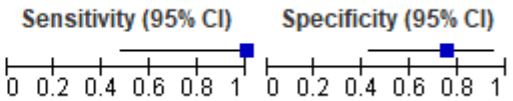


Figure 164: carcinoma (versus benign/indeterminate)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Pisani, 2000 #1343	3	0	2	12	0.60 [0.15, 0.95]	1.00 [0.74, 1.00]
Silverman, 1986 #1387	0	0	0	3	Not estimable	1.00 [0.29, 1.00]

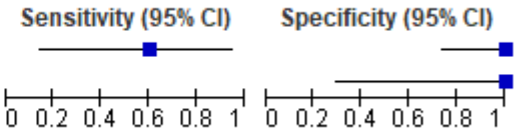


Figure 165: CB grades V and VI

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Xiong, 2019 #1432	489	0	52	36	0.90 [0.88, 0.93]	1.00 [0.90, 1.00]

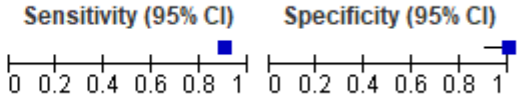


Figure 166: CB grades III, V and VI

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Xiong, 2019 #1432	519	1	22	35	0.96 [0.94, 0.97]	0.97 [0.85, 1.00]

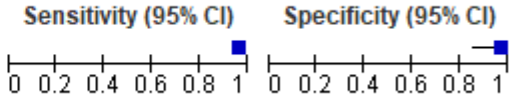


Figure 167: positive (versus negative) with CEUS guidance

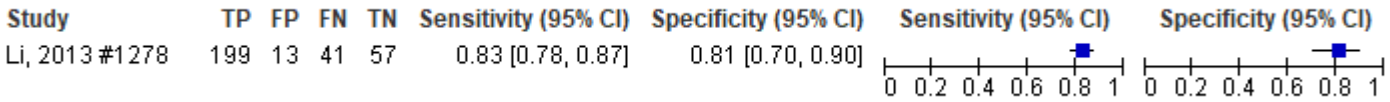
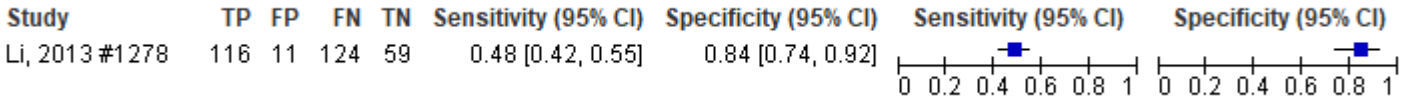


Figure 168: positive (versus negative) with US guidance



Core biopsy, with prior US

Figure 169: indeterminate, follicular neoplasm, suspicious for malignancy, or malignant

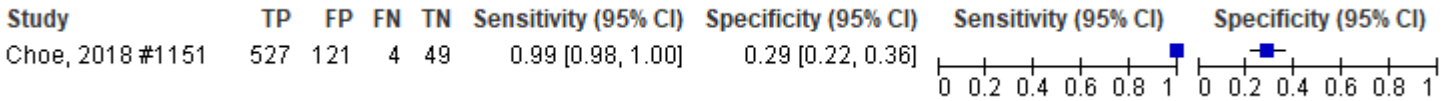


Figure 170: follicular neoplasm, suspicious for malignancy, or malignant

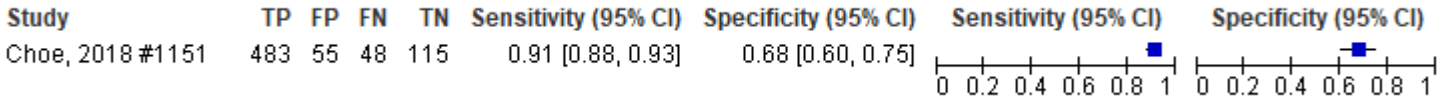
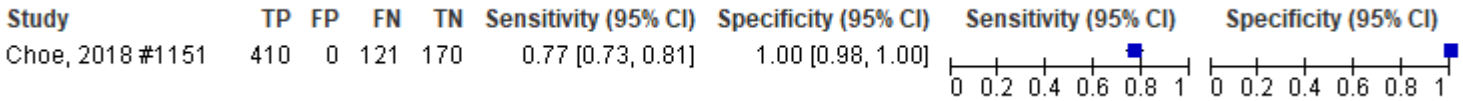


Figure 171: suspicious for malignancy, or malignant



F.2 Sensitivity / 1-specificity plots

In the plots below, the black dot represents the point estimate and the ellipse corresponds to the 95% confidence region around the pooled sensitivity and specificity.

Adjusted analysis

FNAC, no ROSE, smear only, without prior US

Figure 172: Bethesda Grade III or above

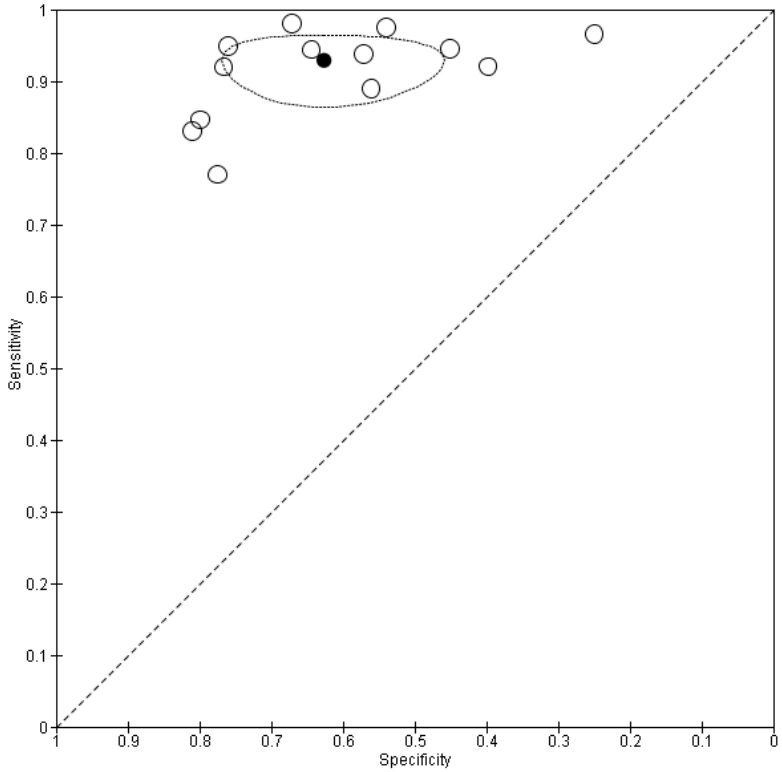


Figure 173: Bethesda Grade IV or above

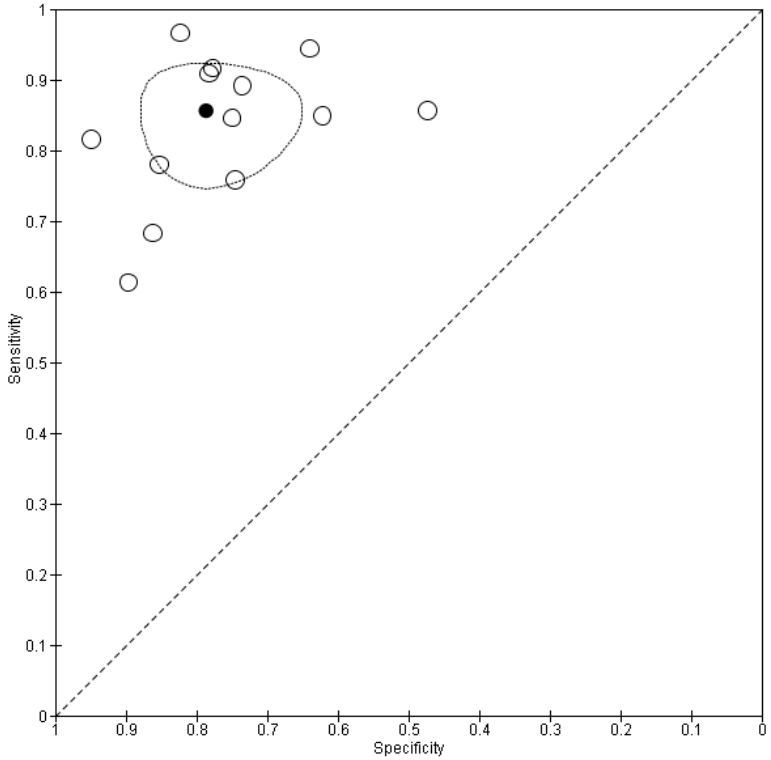


Figure 174: Bethesda Grade V or above

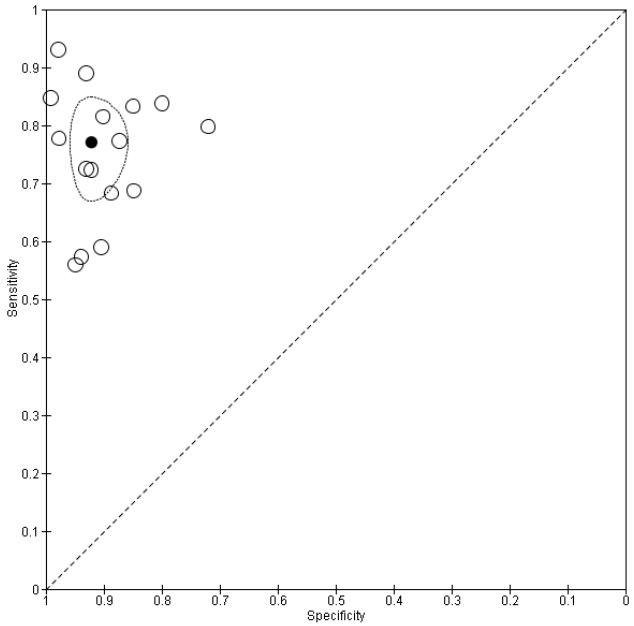


Figure 175: Bethesda Grade VI

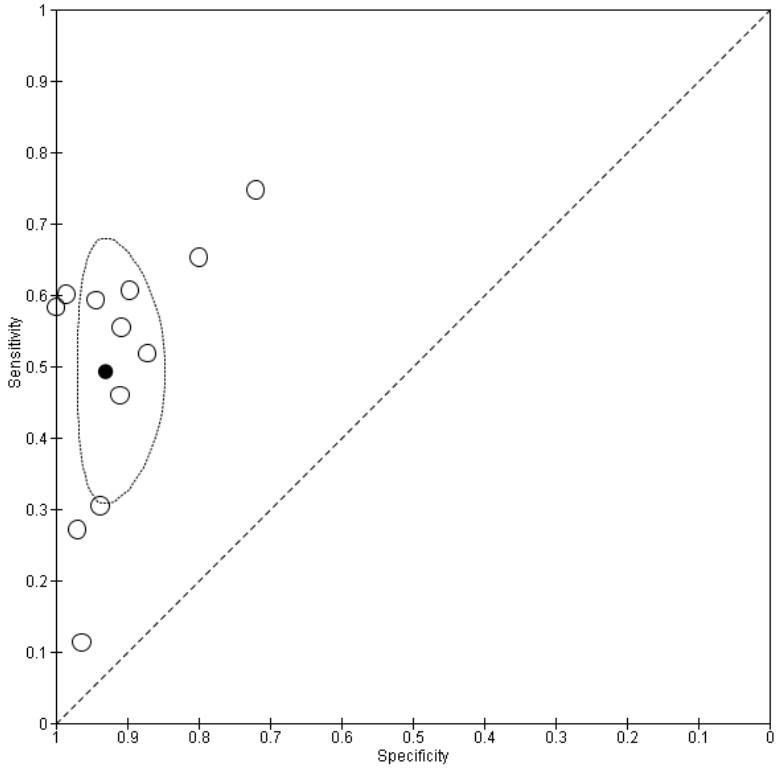


Figure 176: BTA THY 3a or above

No meta-analysis carried out as less than 3 studies

Figure 177: BTA THY 3f or above

No meta-analysis carried out as less than 3 studies

Figure 178: BTA THY 4 or above

No meta-analysis carried out as less than 3 studies

Figure 179: BTA THY 5

No meta-analysis carried out as less than 3 studies

Figure 180: AC 3 or above

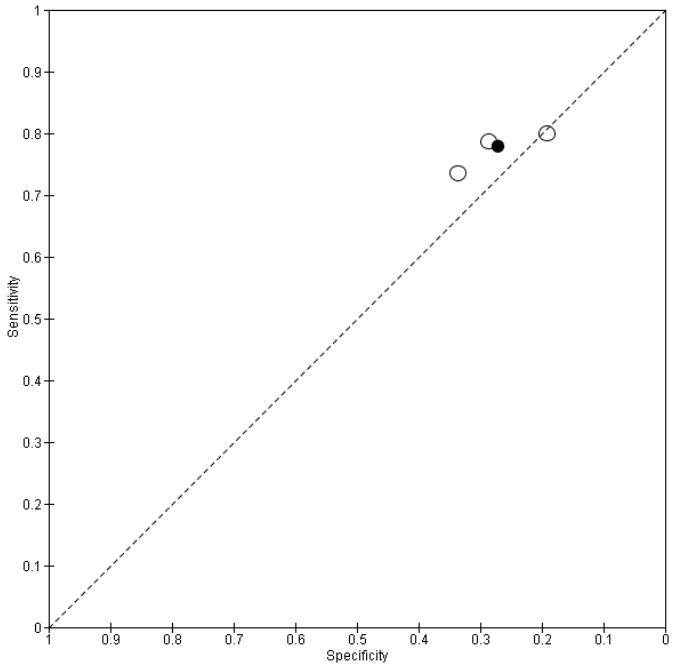


Figure 181: AC 4 or above

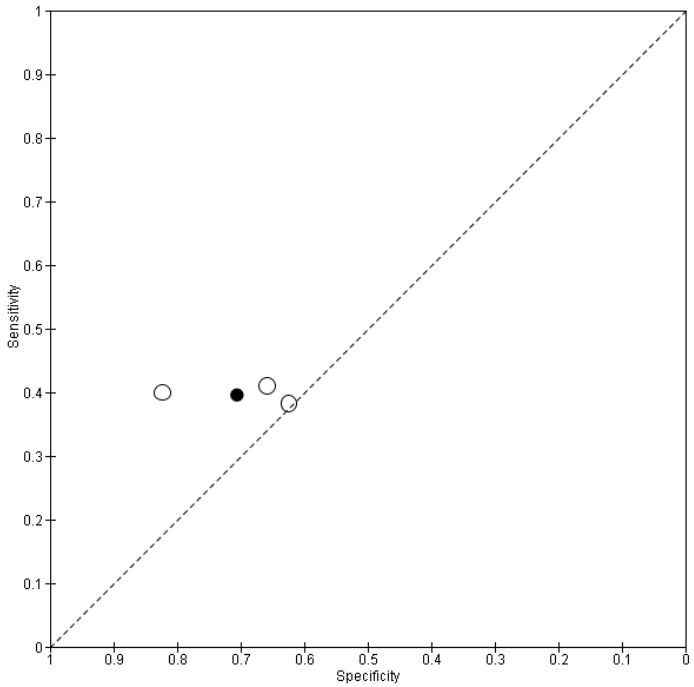


Figure 182: 2 way: malignant v benign

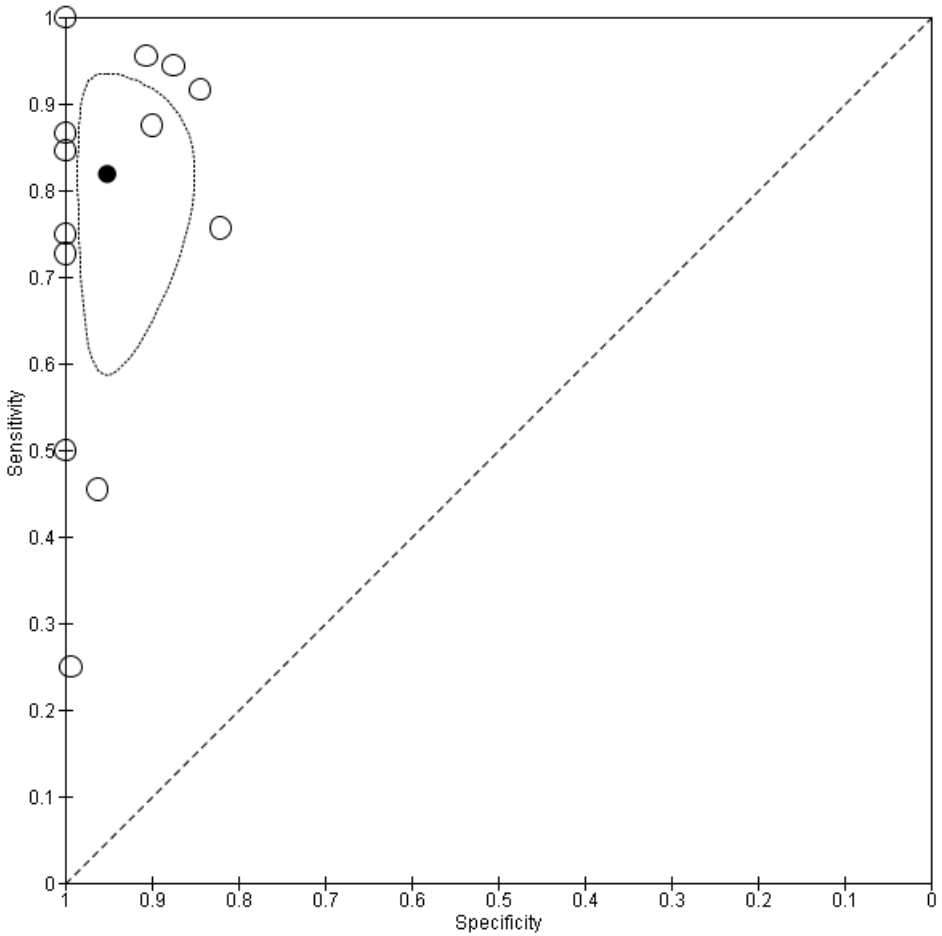


Figure 183: 3 way: suspicious or malignant (negative =benign)

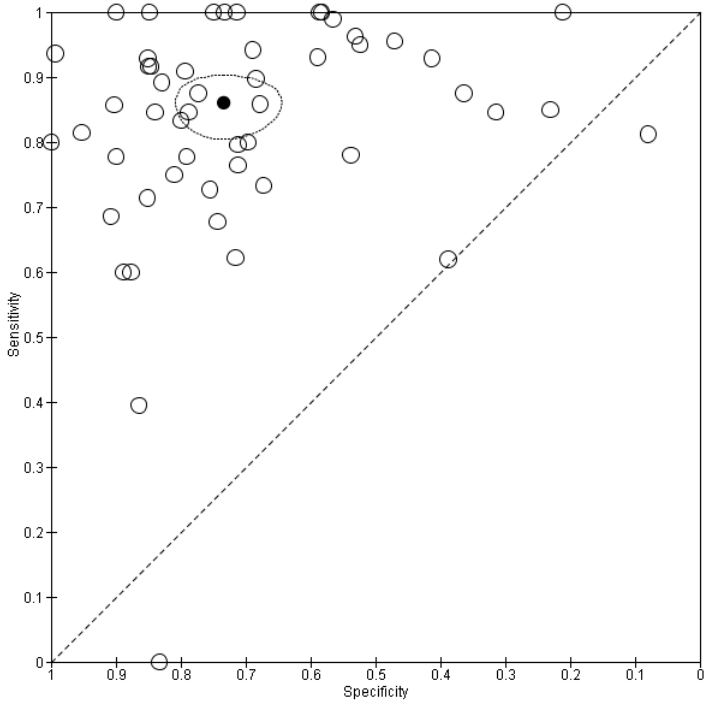


Figure 184: 3 way: malignant (negative = suspicious or benign)

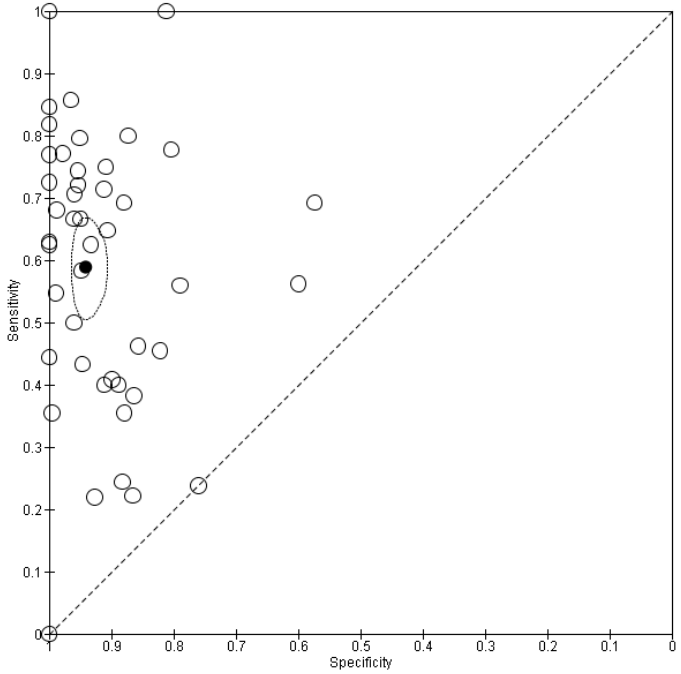


Figure 185: 4 way: malignant or suspicious or indeterminate (negative = benign)

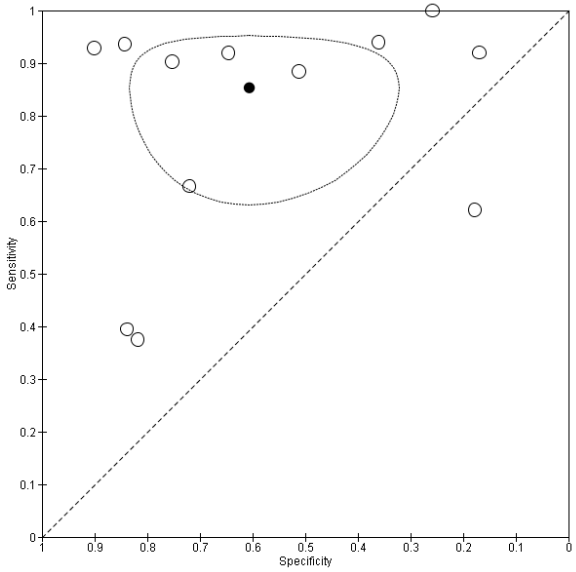


Figure 186: 4 way: malignant or suspicious (negative = benign or indeterminate)

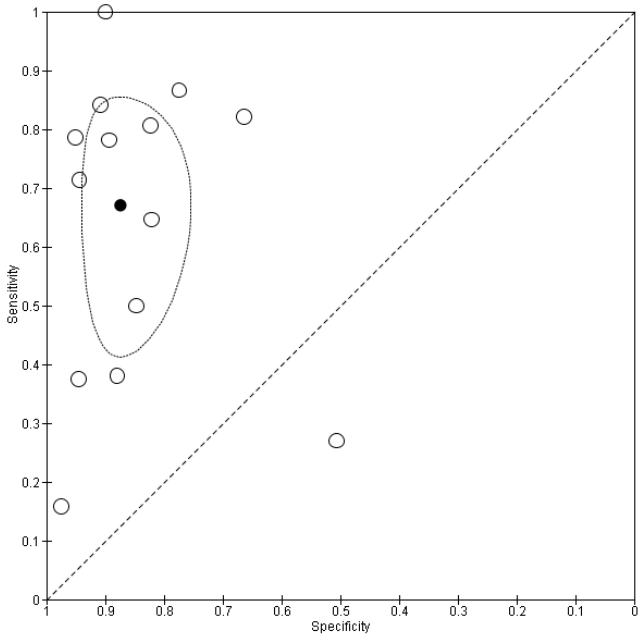


Figure 187: 4 way: malignant (negative = benign or indeterminate or suspicious)

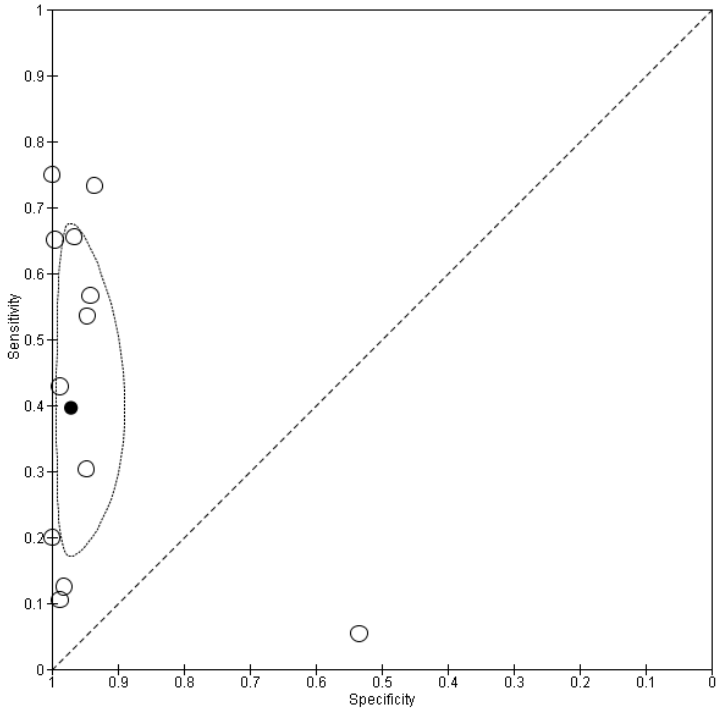

Sensitivity
Specificity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

Figure 188: 5 way: malignant or suspicious or two grades of indeterminate (negative = benign)

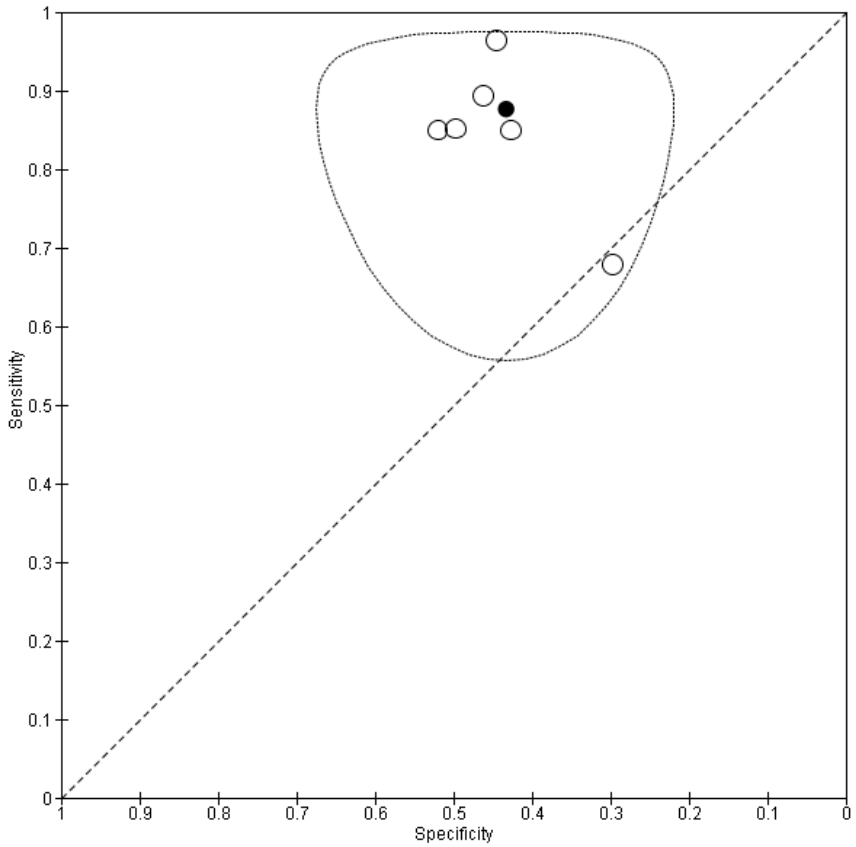


Figure 189: 5 way: malignant or suspicious or one grade of indeterminate (negative = lower grade of indeterminate or benign)

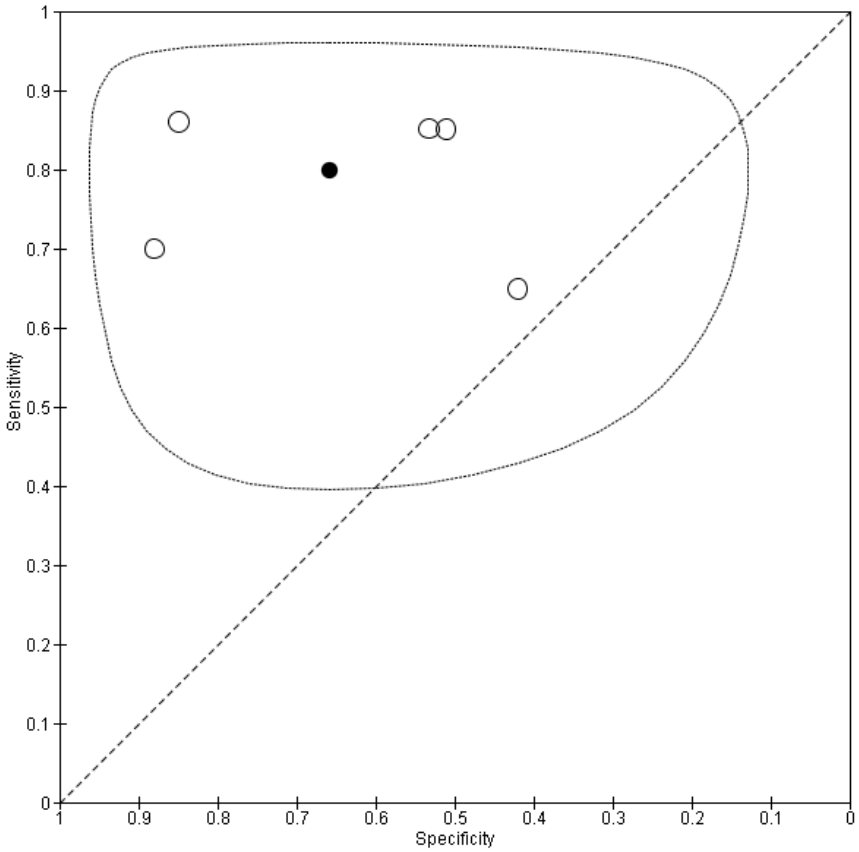


Figure 190: 5 way: malignant (negative = suspicious or two grades of indeterminate or benign)

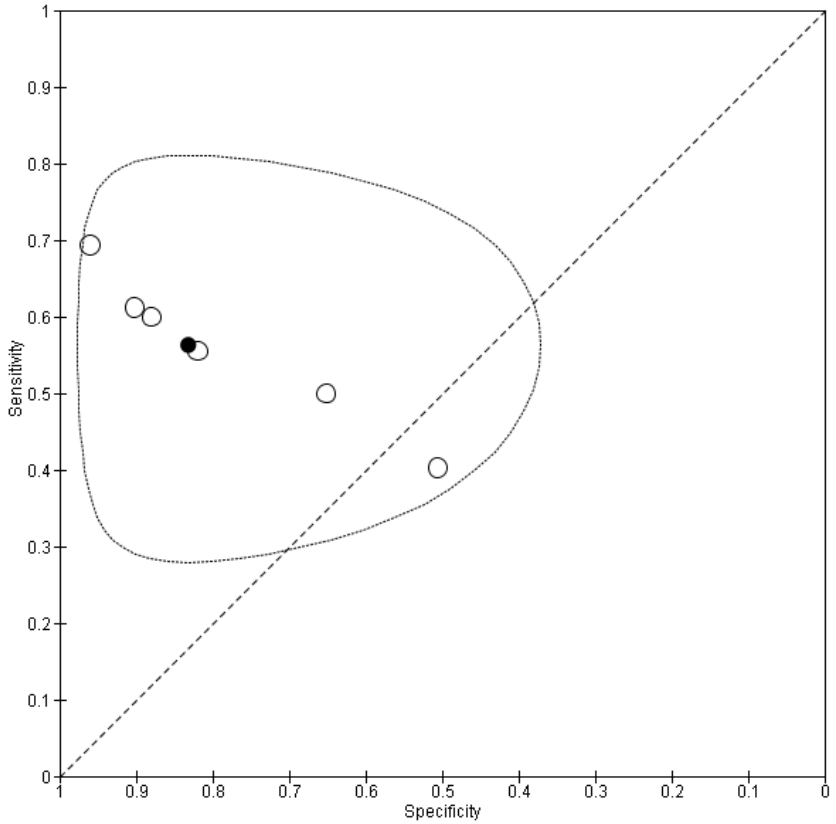


Figure 191: 1 or more inclusions

No meta-analysis carried out as less than 3 studies

Figure 192: 1 or more grooves

No meta-analysis carried out as less than 3 studies

Figure 193: 2 or more grooves

No meta-analysis carried out as less than 3 studies

Figure 194: 3 or more grooves

No meta-analysis carried out as less than 3 studies

FNAC, no ROSE, smear only, with prior US

Figure 195: Bethesda Grade III or above

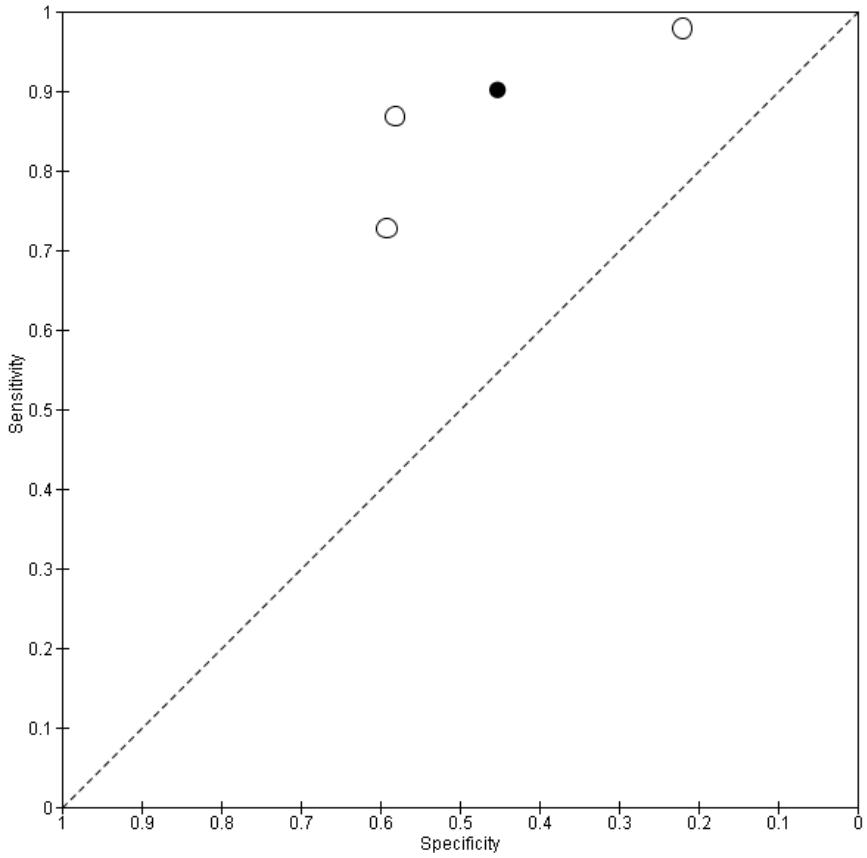


Figure 196: Bethesda Grade IV or above

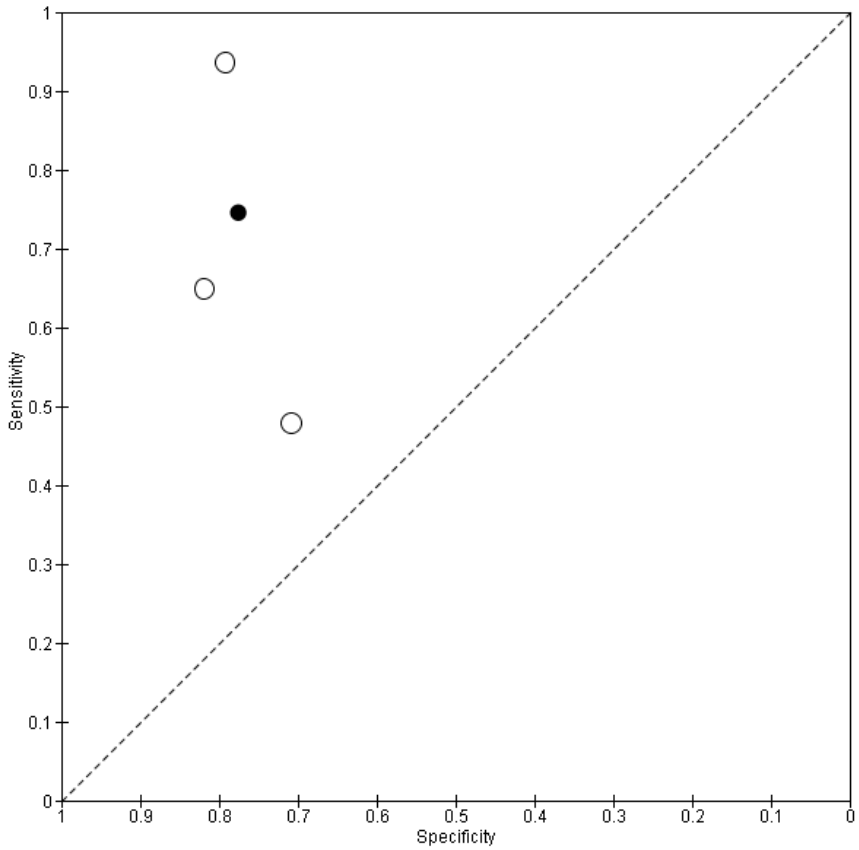


Figure 197: Bethesda Grade V or above

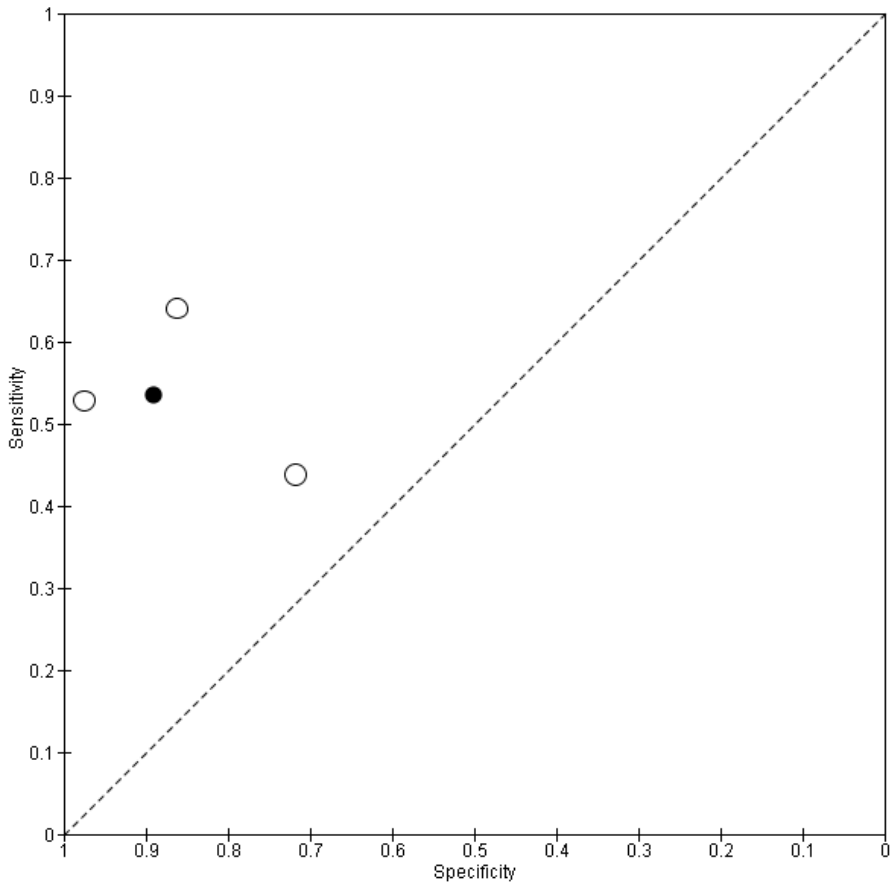


Figure 198: Bethesda Grade VI or above

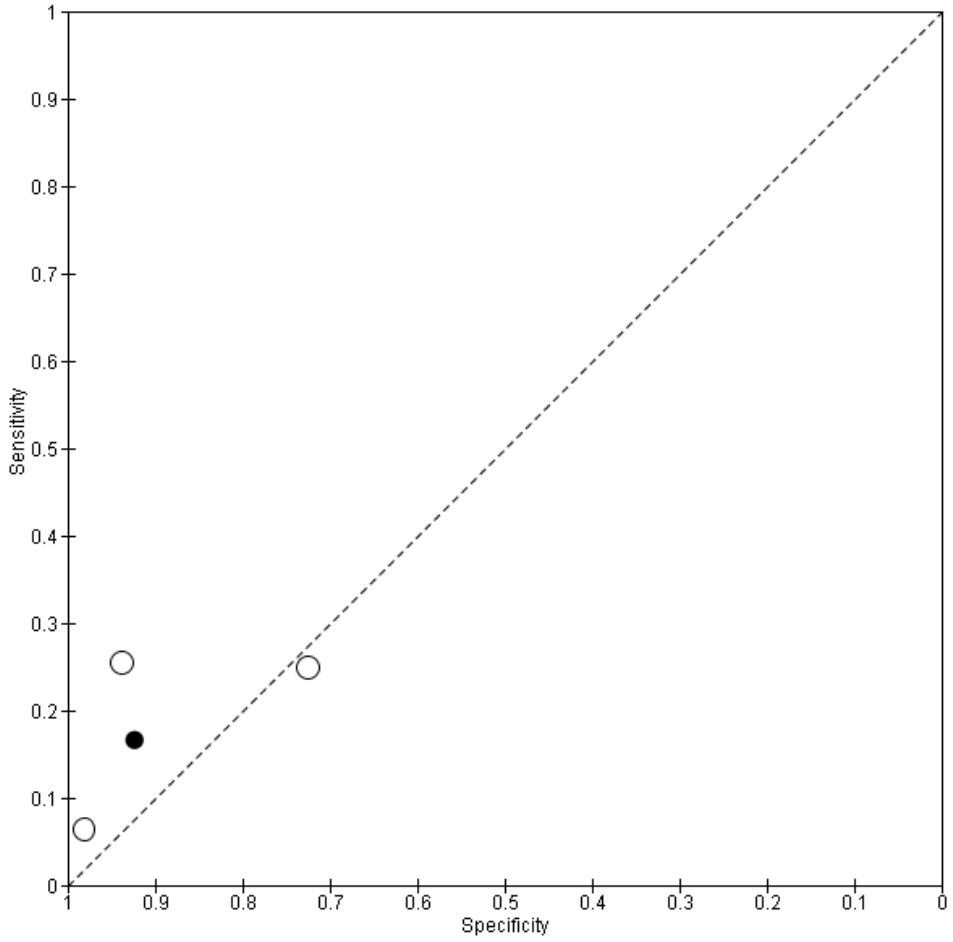


Figure 199: 2 way: malignant versus benign

No meta-analysis carried out as less than 3 studies

Figure 200: 3 way: suspicious or malignant (negative = benign)

No meta-analysis carried out as less than 3 studies

Figure 201: 3 way: malignant (negative = suspicious or benign)

No meta-analysis carried out as less than 3 studies

Figure 202: 4 way De May classification: malignant, suspicious, non malignant follicular proliferation (negative = benign)

No meta-analysis carried out as less than 3 studies

Figure 203: 4 way De May classification: malignant, suspicious (negative = benign, non malignant follicular proliferation)

No meta-analysis carried out as less than 3 studies

Figure 204: 4 way De May classification: malignant (negative = benign, non malignant follicular proliferation, suspicious)

No meta-analysis carried out as less than 3 studies

Figure 205: 4 way Piana classification: C3 or more

No meta-analysis carried out as less than 3 studies

Figure 206: 4 way Piana classification: C4 or more

No meta-analysis carried out as less than 3 studies

Figure 207: 4 way Piana classification: C5 or more

No meta-analysis carried out as less than 3 studies

Figure 208: 4 way generic: malignant, suspicious, indeterminate (benign = negative)

No meta-analysis carried out as less than 3 studies

Figure 209: 4 way generic: malignant, suspicious, indeterminate (benign = negative)

No meta-analysis carried out as less than 3 studies

FNAC, no ROSE, smear, with cytospin and/or cell-block, without prior US

Figure 210: Bethesda Grade III or above

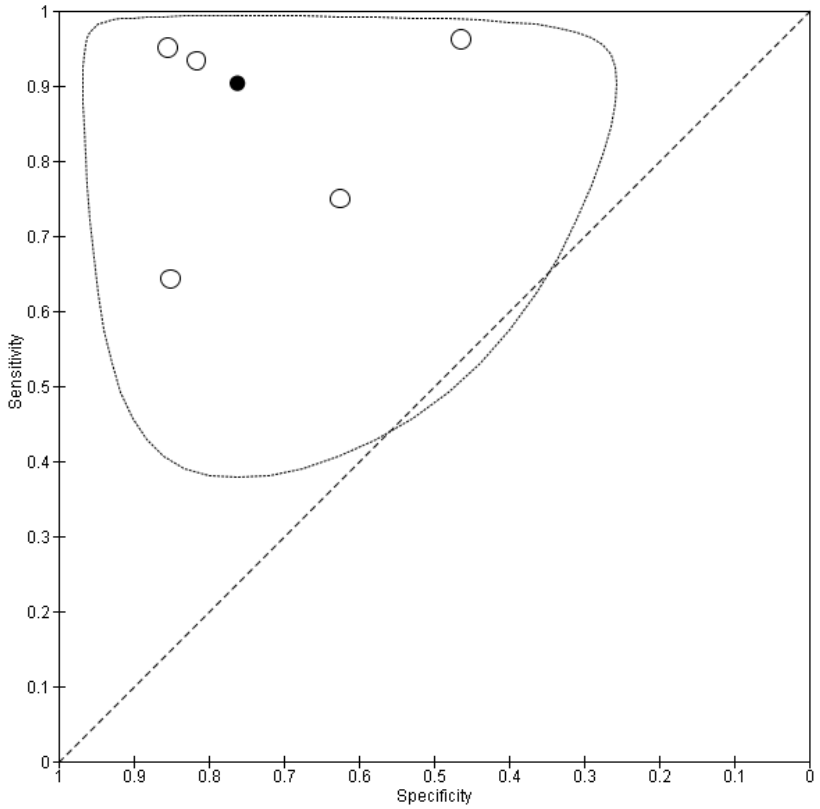


Figure 211: Bethesda Grade IV or above

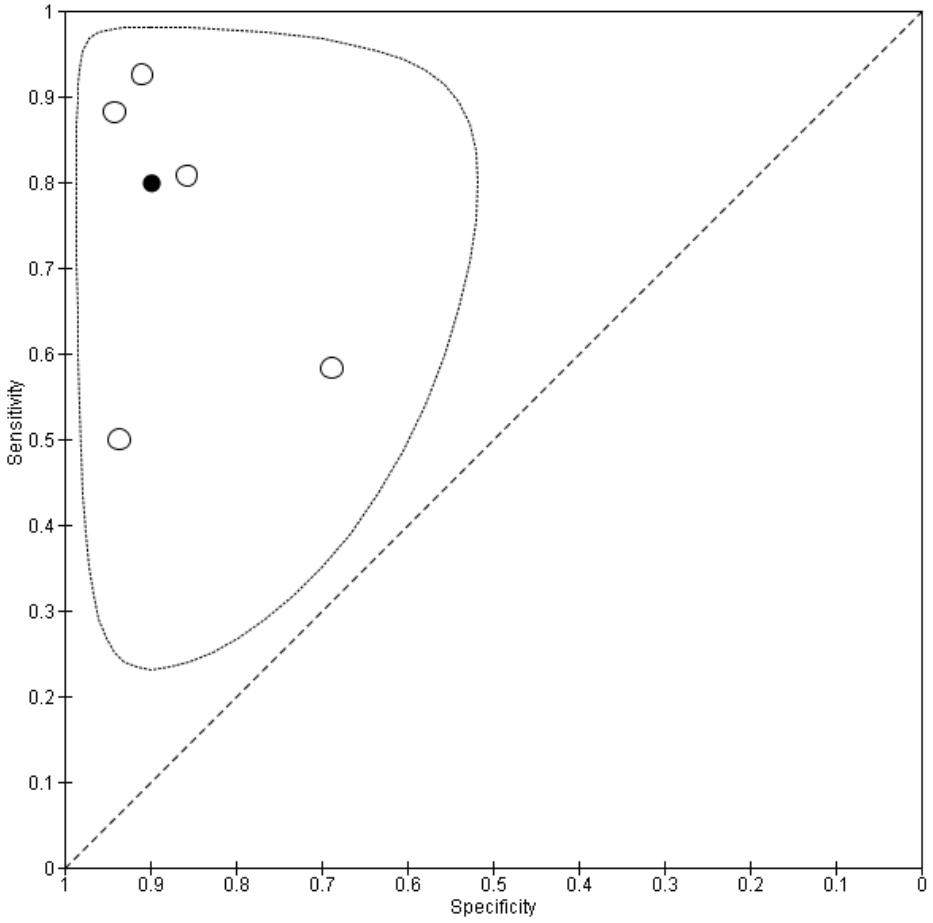


Figure 212: Bethesda Grade V or above

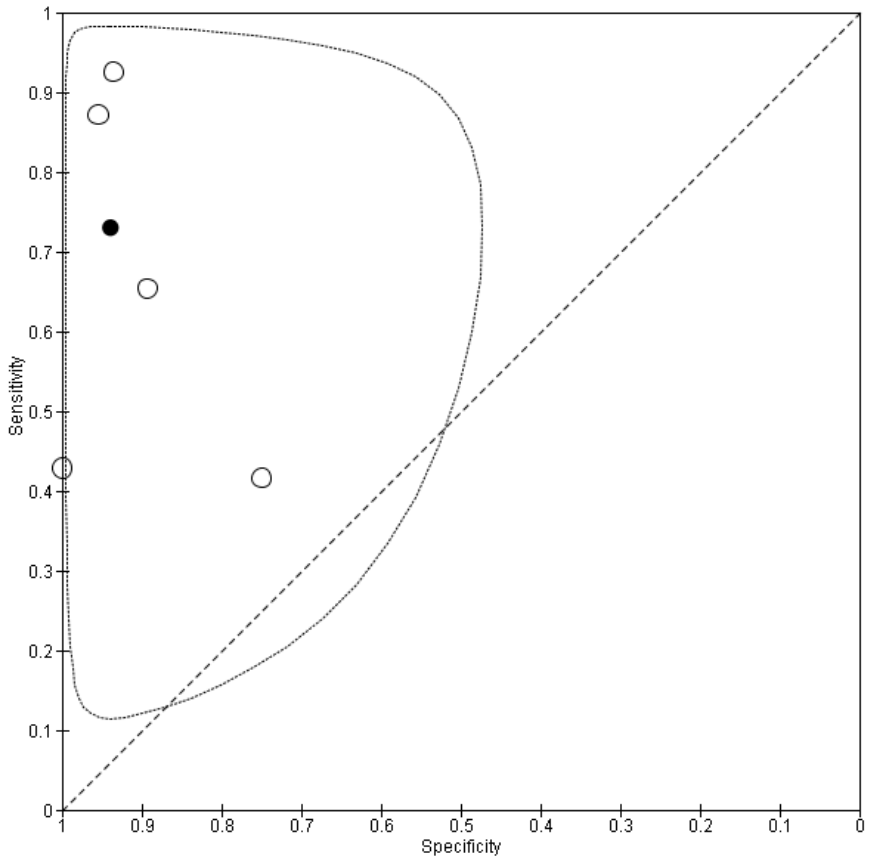


Figure 213: Bethesda Grade VI or above

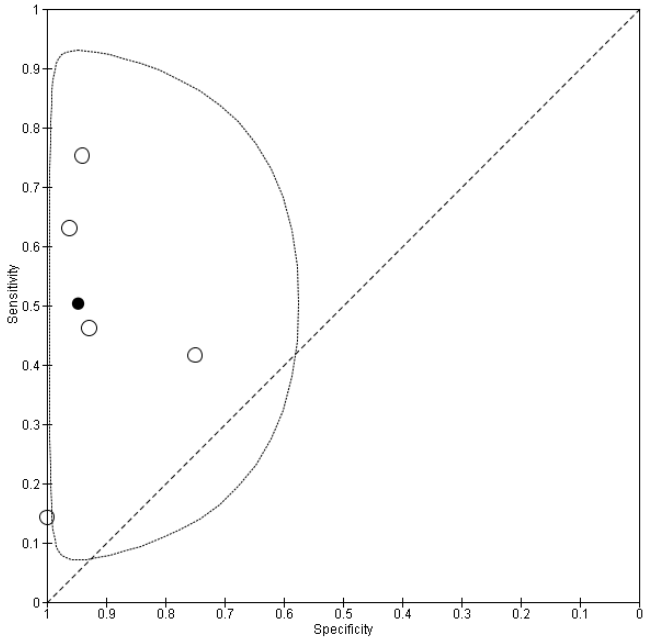


Figure 214: 2 way: malignant v benign

No meta-analysis carried out as less than 3 studies

Figure 215: 3 way: malignant or suspicious (negative = benign)

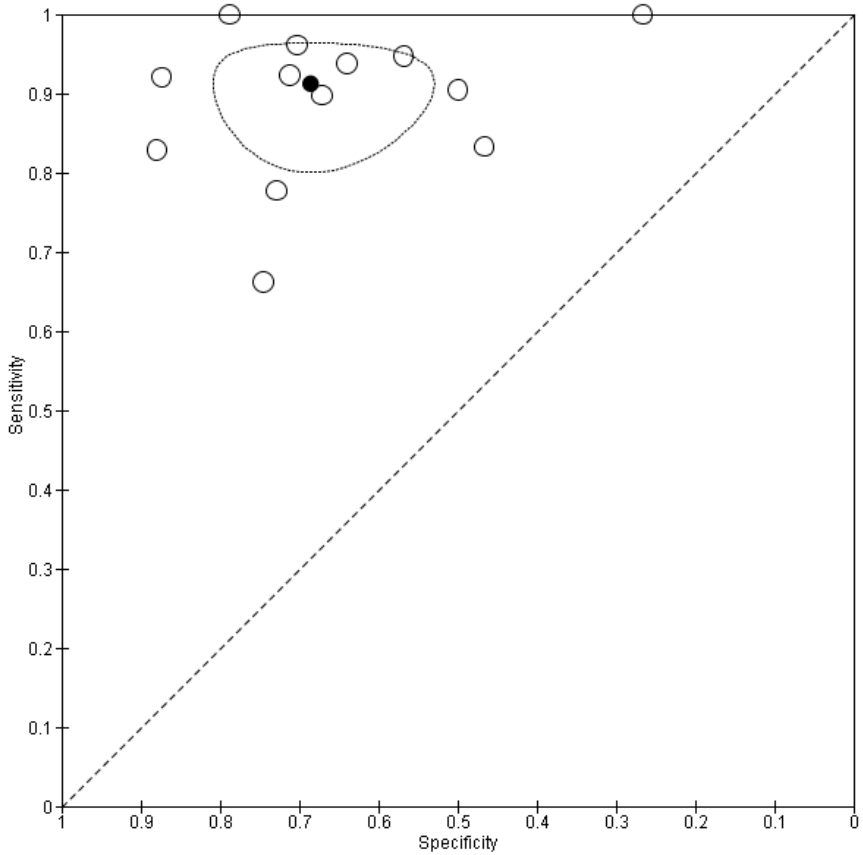


Figure 216: 3 way: malignant (negative = benign or suspicious)

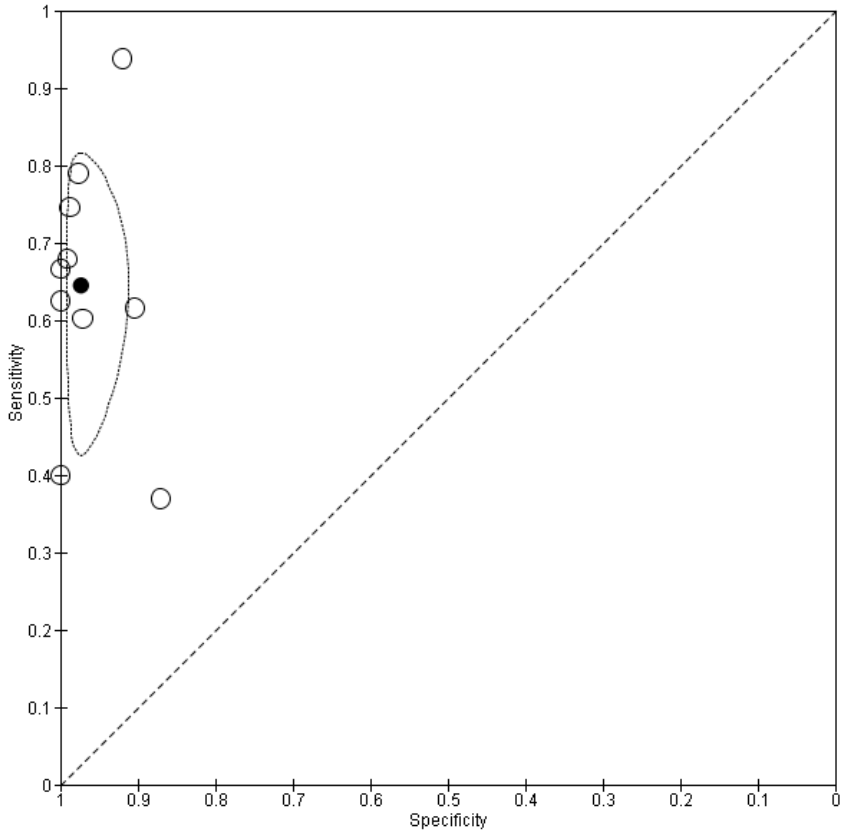


Figure 217: 4 way: malignant, suspicious, indeterminate (negative = benign)

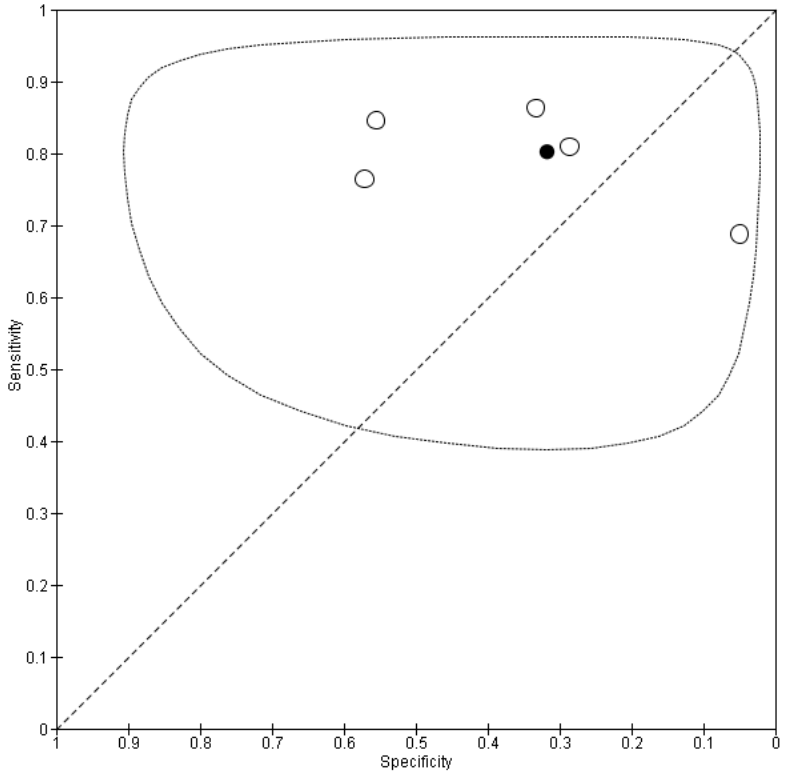


Figure 218: 4 way: malignant, suspicious (negative = benign, indeterminate)

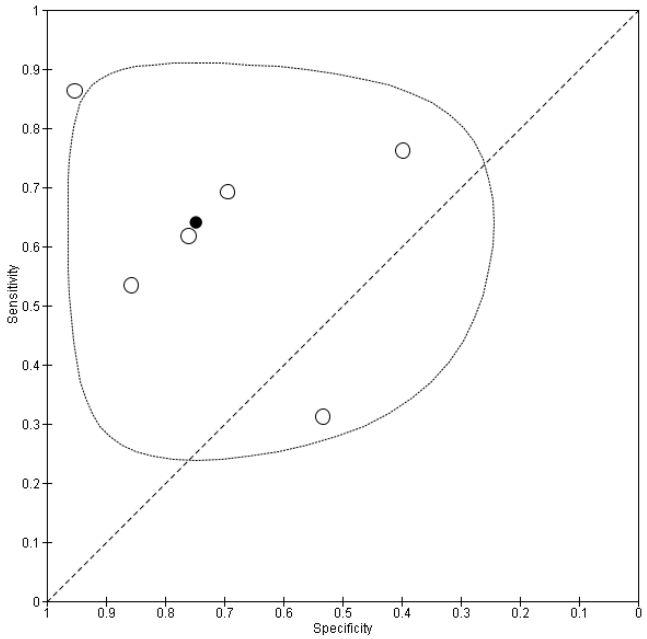


Figure 219: 4 way: malignant (negative = benign, indeterminate, suspicious)

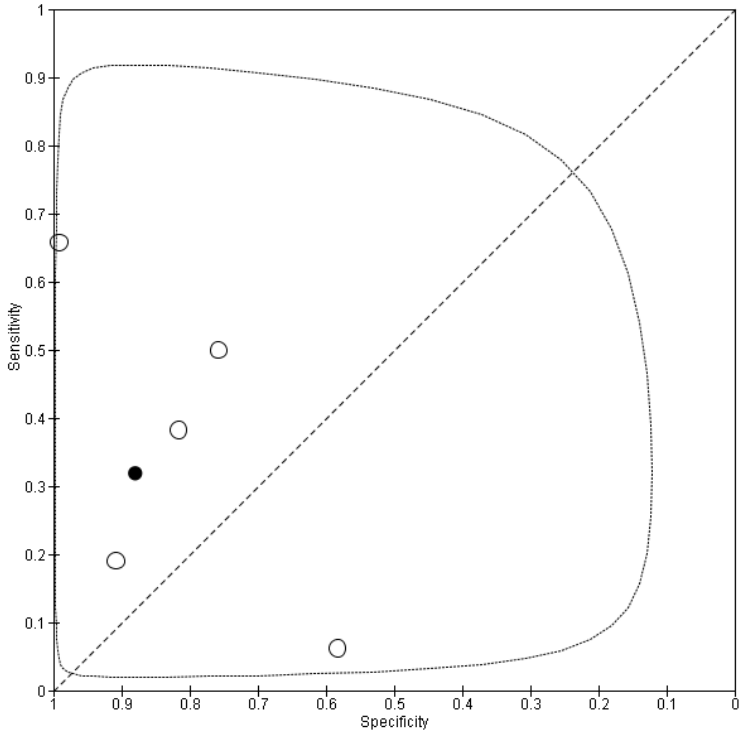


Figure 220: 5 way: malignant, suspicious, 2 grades of indeterminate (negative = benign)

No meta-analysis carried out as less than 3 studies

FNAC, no ROSE, smear, with cytospin and/or cell-block, with prior US

Figure 221: Bethesda Grade III or above

No meta-analysis carried out as less than 3 studies

Figure 222: Bethesda Grade IV or above

No meta-analysis carried out as less than 3 studies

Figure 223: Bethesda Grade V or above

No meta-analysis carried out as less than 3 studies

Figure 224: Bethesda Grade VI

No meta-analysis carried out as less than 3 studies

Figure 225: Benign or above

No meta-analysis carried out as less than 3 studies

FNAC, with ROSE, smear only, without prior US

Figure 226: Bethesda Grade III or above

No meta-analysis carried out as less than 3 studies

Figure 227: Bethesda Grade IV or above

No meta-analysis carried out as less than 3 studies

Figure 228: Bethesda Grade V or above

No meta-analysis carried out as less than 3 studies

Figure 229: Bethesda Grade VI

No meta-analysis carried out as less than 3 studies

Figure 230: 3 way: malignant and suspicious (negative = benign)

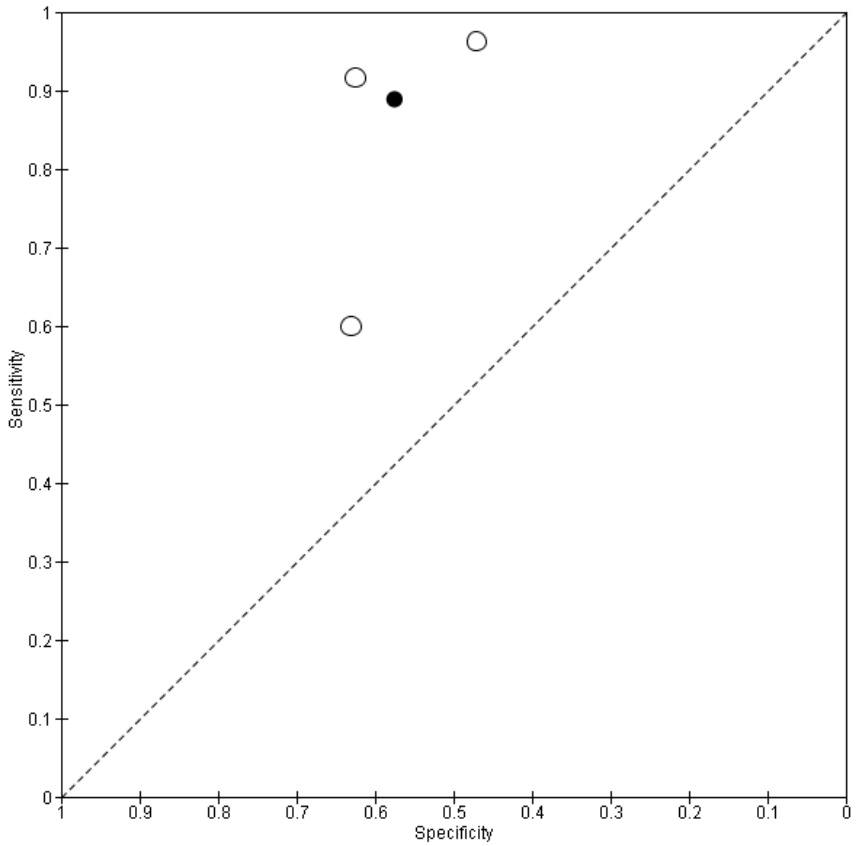


Figure 231: 3 way: malignant (negative = benign and suspicious)

No meta-analysis carried out as less than 3 studies

Figure 232: 4 way: malignant, suspicious, indeterminate (negative = benign)

No meta-analysis carried out as less than 3 studies

Figure 233: 4 way: malignant, suspicious (negative = benign, indeterminate)

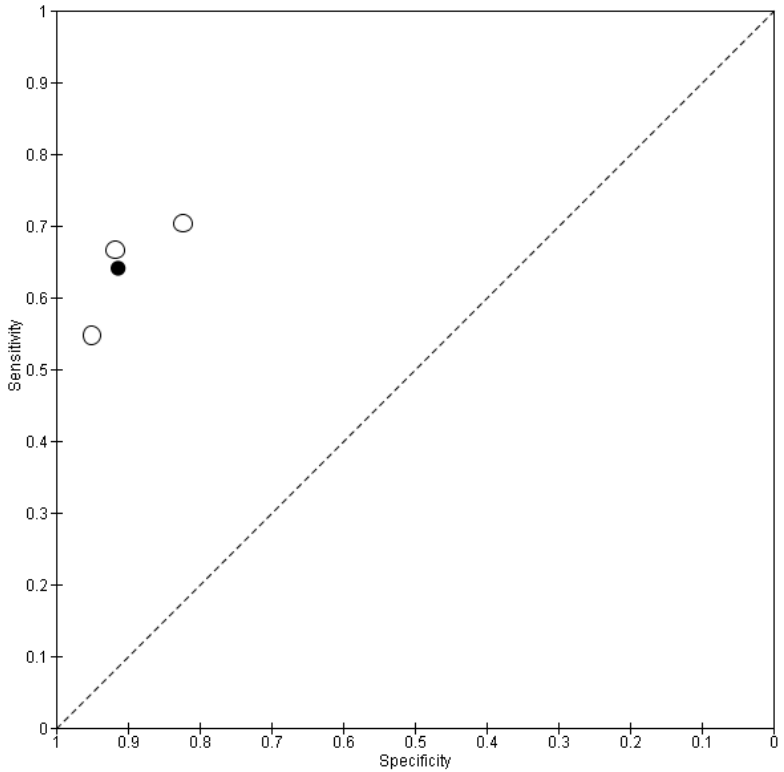


Figure 234: 4 way: malignant (negative = benign, indeterminate, suspicious)

No meta-analysis carried out as less than 3 studies

FNAC, with ROSE, smear only, with prior US

Figure 235: intermediate or malignant

No meta-analysis carried out as less than 3 studies

FNAC, with ROSE, smear, with cytospin and/or cell-block, without prior US

Figure 236: 3 way: suspicious or malignant (negative = benign)

No meta-analysis carried out as less than 3 studies

Figure 237: 3 way: malignant (negative = suspicious or benign)

No meta-analysis carried out as less than 3 studies

Figure 238: 4 way: malignant, suspicious, indeterminate (negative = benign)

No meta-analysis carried out as less than 3 studies

Figure 239: 4 way: malignant, suspicious (negative = benign, indeterminate)

No meta-analysis carried out as less than 3 studies

Figure 240: 5 way: malignant, suspicious, 2 grades of indeterminate (negative = benign)

No meta-analysis carried out as less than 3 studies

Figure 241: 5 way: malignant, suspicious (negative = 2 grades of indeterminate, benign)

No meta-analysis carried out as less than 3 studies

Figure 242: 5 way: malignant, suspicious (negative = suspicious, lower grade of indeterminate, benign)

No meta-analysis carried out as less than 3 studies

Figure 243: 5 way: malignant (negative = suspicious, 2 grades of indeterminate, benign)

No meta-analysis carried out as less than 3 studies

FNAC, with ROSE, smear, with cytospin and/or cell-block, with prior US

Figure 244: indeterminate follicular, indeterminate Hurtle, Suspicious for malignancy, or positive

No meta-analysis carried out as less than 3 studies

Figure 245: Suspicious for malignancy, or indeterminate follicular or positive

No meta-analysis carried out as less than 3 studies

Figure 246: Suspicious for malignancy, or positive

No meta-analysis carried out as less than 3 studies

Figure 247: Positive for malignancy

No meta-analysis carried out as less than 3 studies

Core biopsy, without prior US

Figure 248: carcinoma or neoplasm (versus benign)

No meta-analysis carried out as less than 3 studies

Figure 249: carcinoma (versus benign/indeterminate)

No meta-analysis carried out as less than 3 studies

Figure 250: CB grades V and VI

No meta-analysis carried out as less than 3 studies

Figure 251: CB grades III, V and VI

No meta-analysis carried out as less than 3 studies

Figure 252: positive (versus negative) with CEUS guidance

No meta-analysis carried out as less than 3 studies

Figure 253: positive (versus negative) with US guidance

No meta-analysis carried out as less than 3 studies

Core biopsy, with prior US

Figure 254: indeterminate, follicular neoplasm, suspicious for malignancy, or malignant

No meta-analysis carried out as less than 3 studies

Figure 255: follicular neoplasm, suspicious for malignancy, or malignant

No meta-analysis carried out as less than 3 studies

Figure 256: suspicious for malignancy, or malignant

No meta-analysis carried out as less than 3 studies

Raw data analysis

FNAC, no ROSE, smear only, without prior US

Figure 257: Bethesda Grade III or above

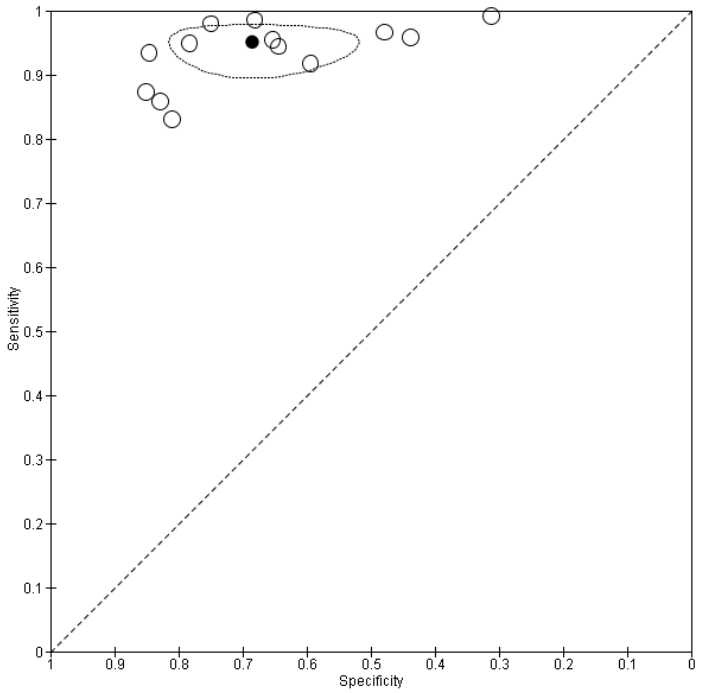


Figure 258: Bethesda Grade IV or above

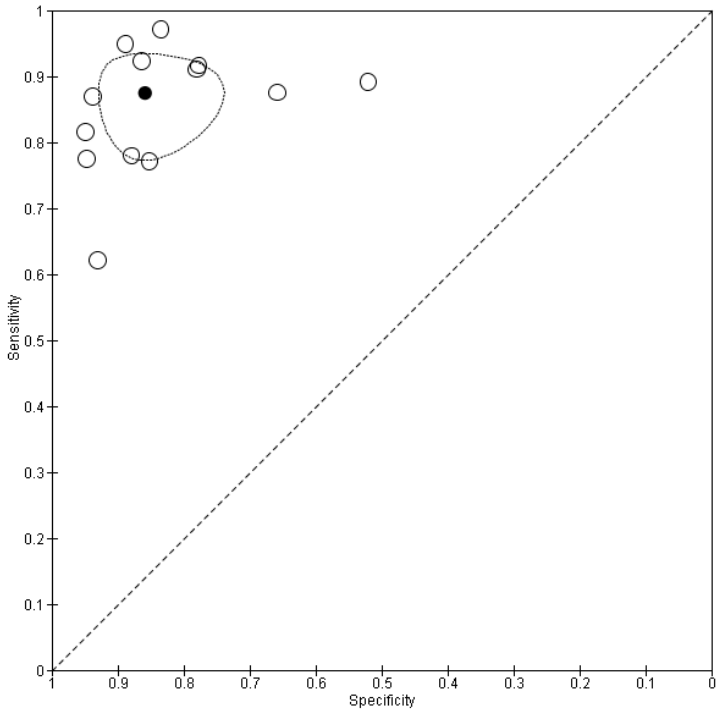


Figure 259: Bethesda Grade V or above

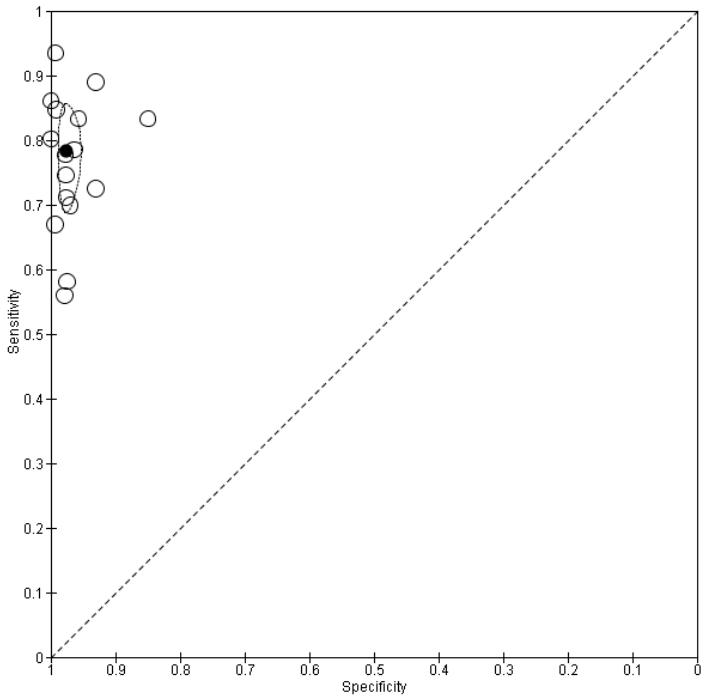


Figure 260: Bethesda Grade VI

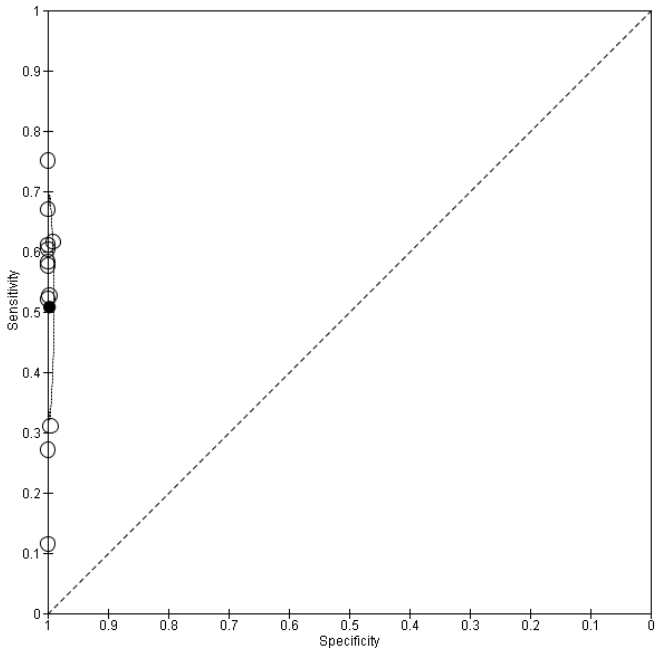


Figure 261: BTA THY 3a or above

No meta-analysis carried out as less than 3 studies

Figure 262: BTA THY 3f or above

No meta-analysis carried out as less than 3 studies

Figure 263: BTA THY 4 or above

No meta-analysis carried out as less than 3 studies

Figure 264: BTA THY 5

No meta-analysis carried out as less than 3 studies

Figure 265: AC 3 or above

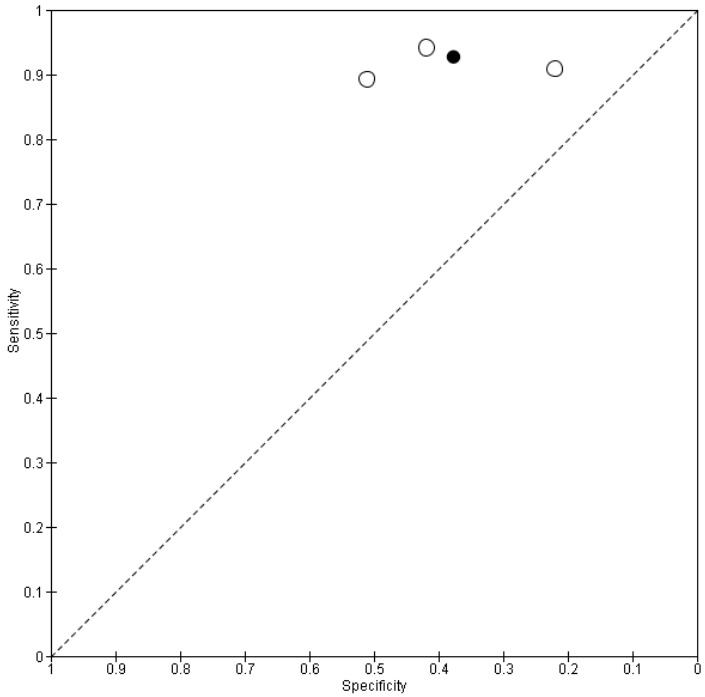


Figure 266: AC 4 or above

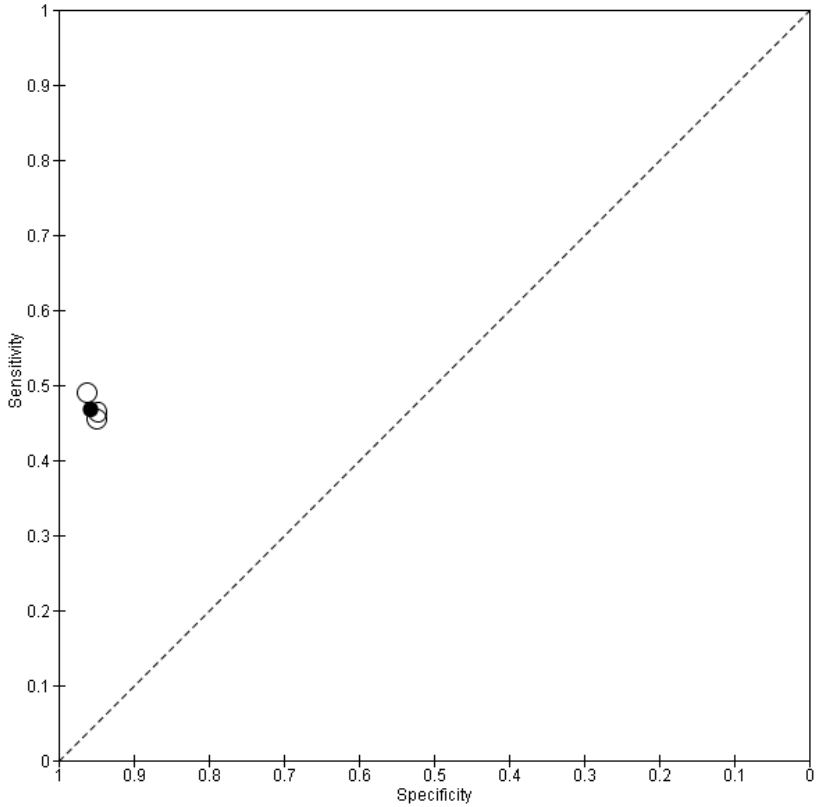


Figure 267: 2 way: malignant v benign

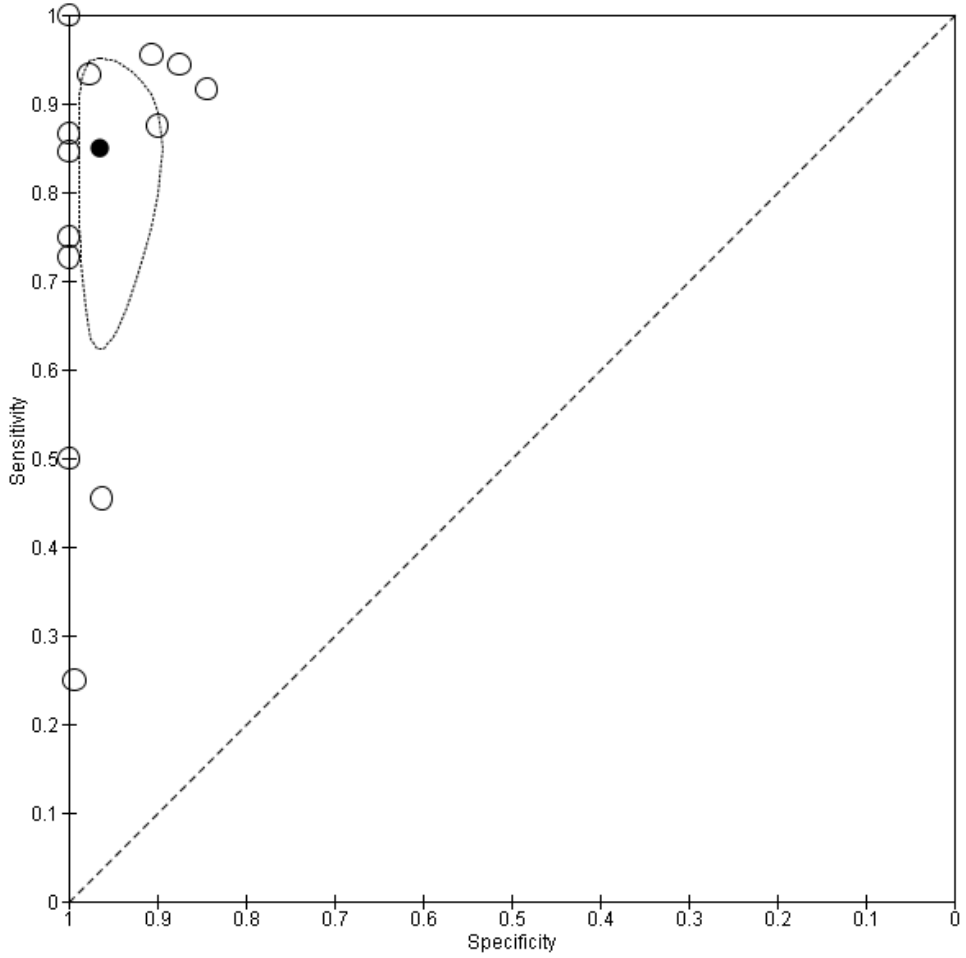


Figure 268: 3 way: suspicious or malignant (negative =benign)

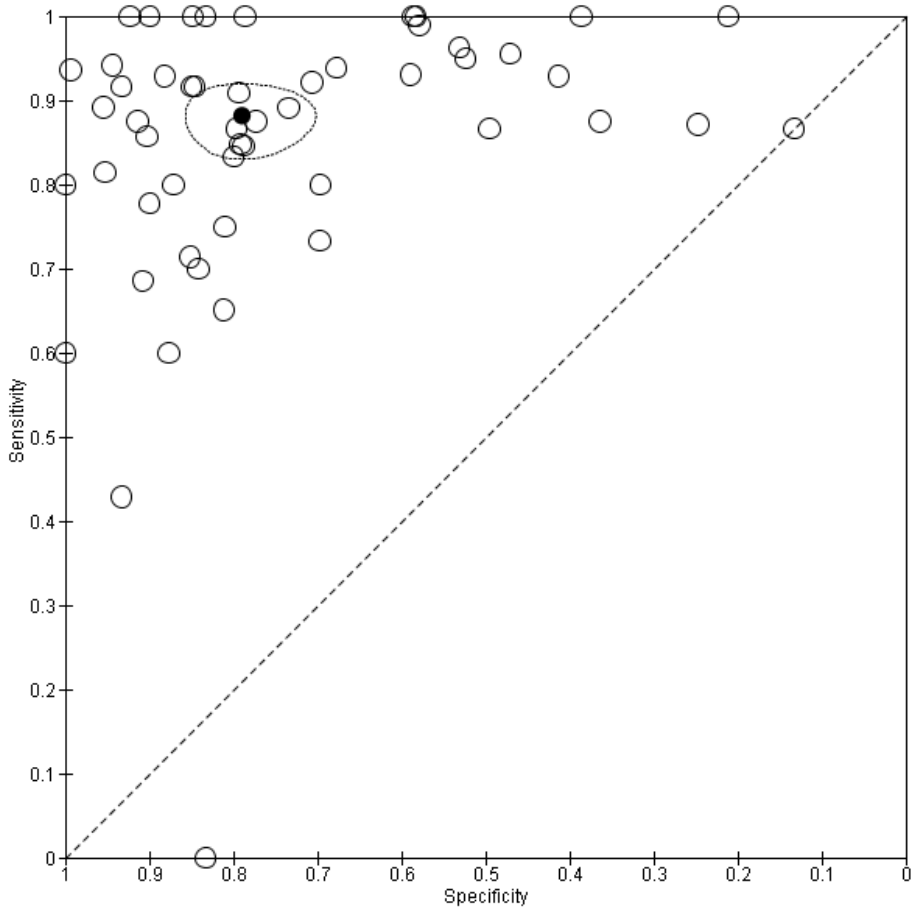


Figure 269: 3 way: malignant (negative = suspicious or benign)

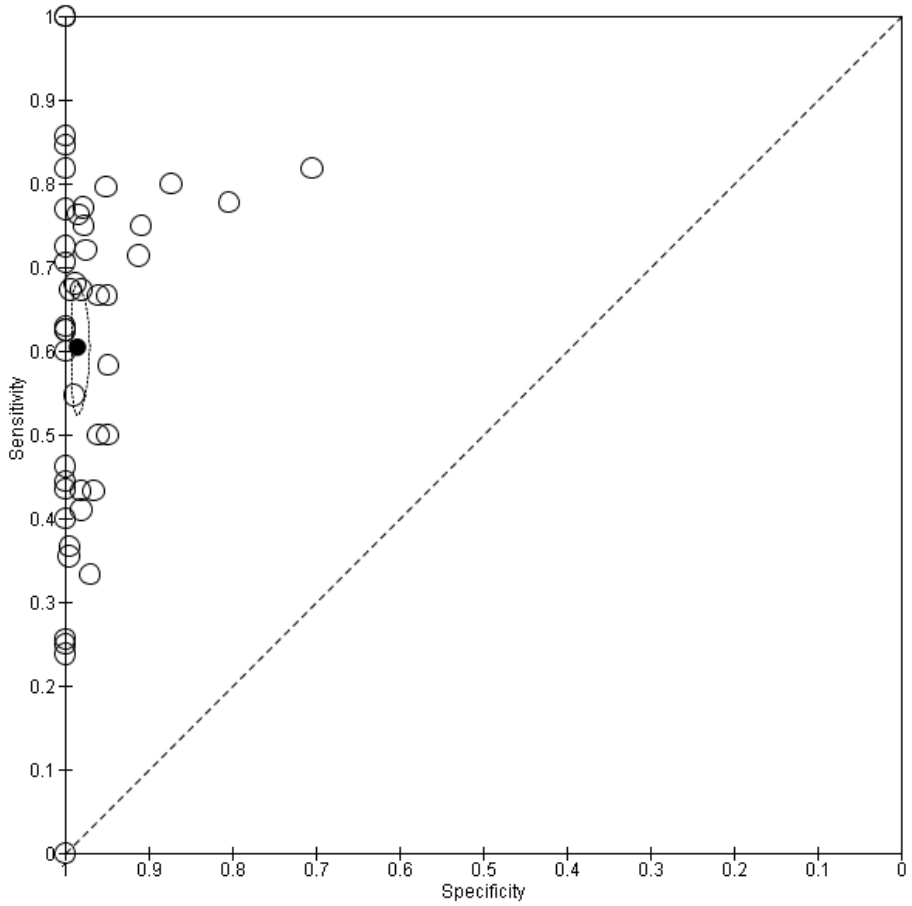


Figure 270: 4 way: malignant or suspicious or indeterminate (negative = benign)

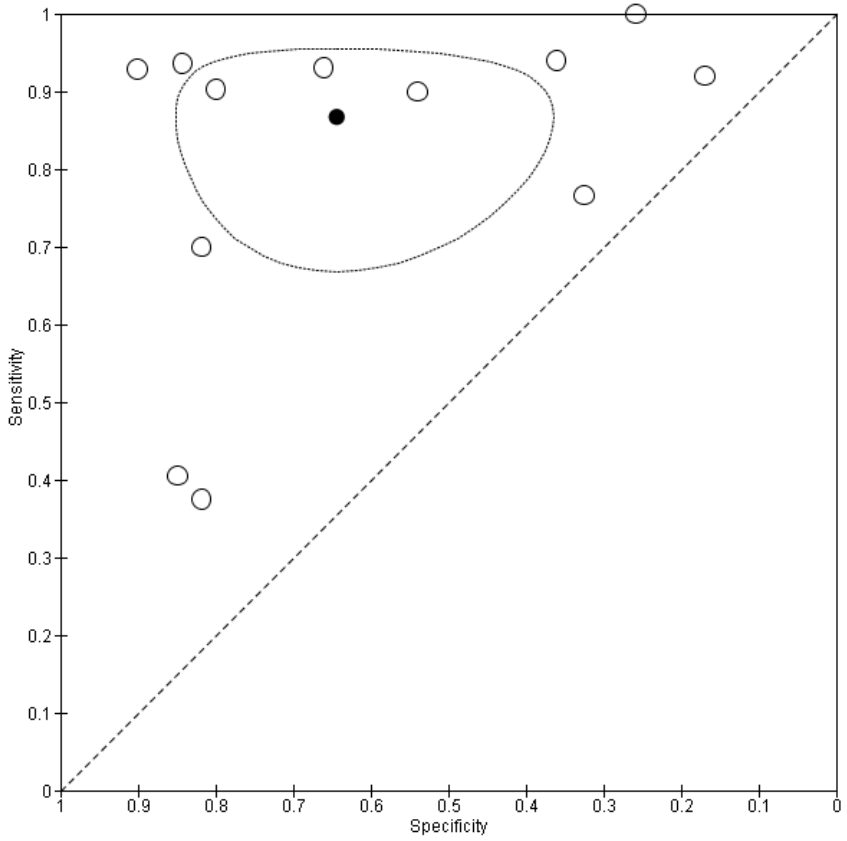


Figure 271: 4 way: malignant or suspicious (negative = benign or indeterminate)

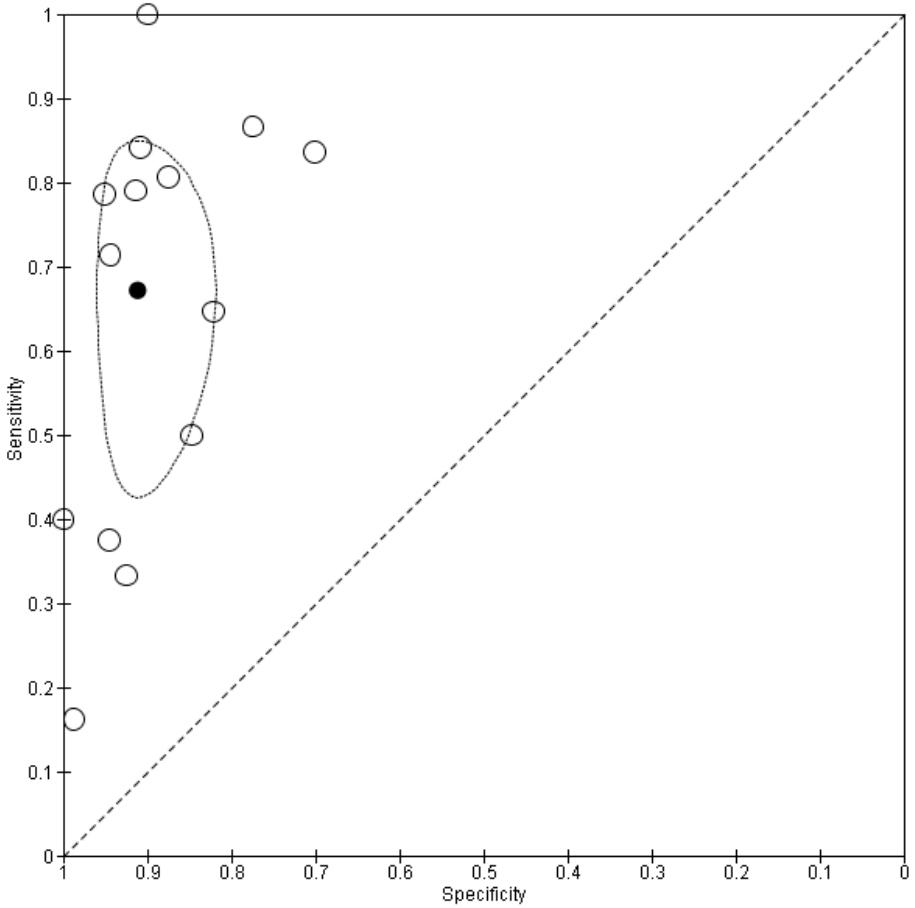


Figure 272: 4 way: malignant (negative = benign or indeterminate or suspicious)

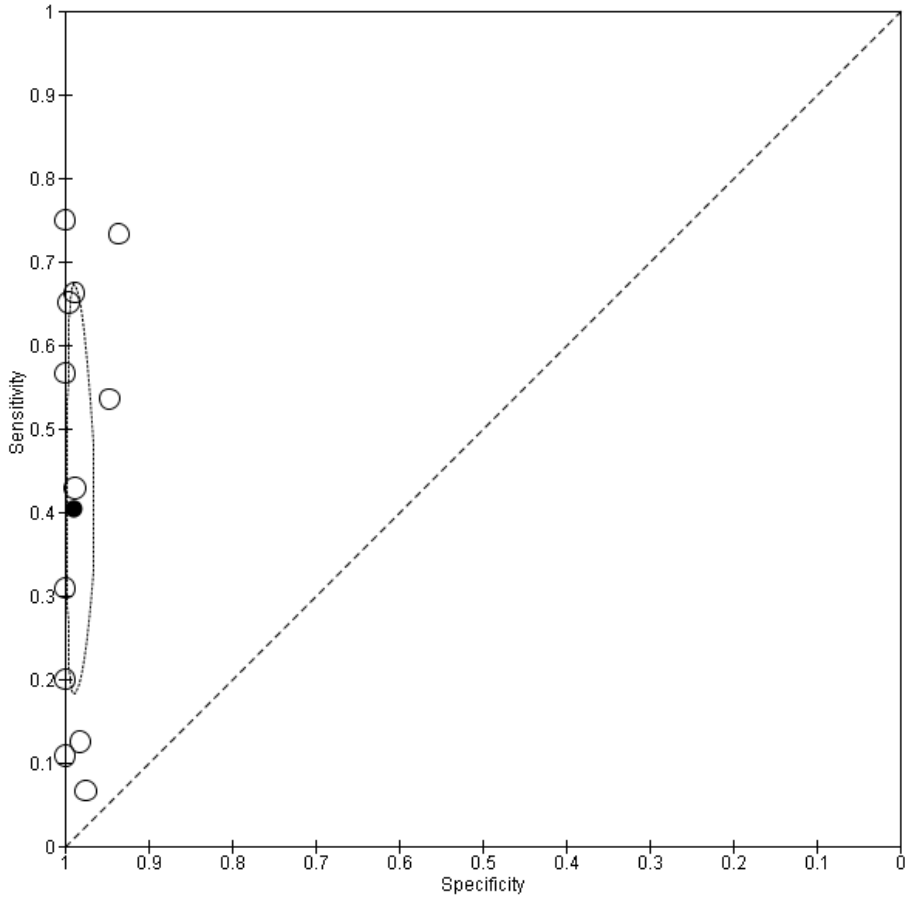


Figure 273: 5 way: malignant or suspicious or two grades of indeterminate (negative = benign)

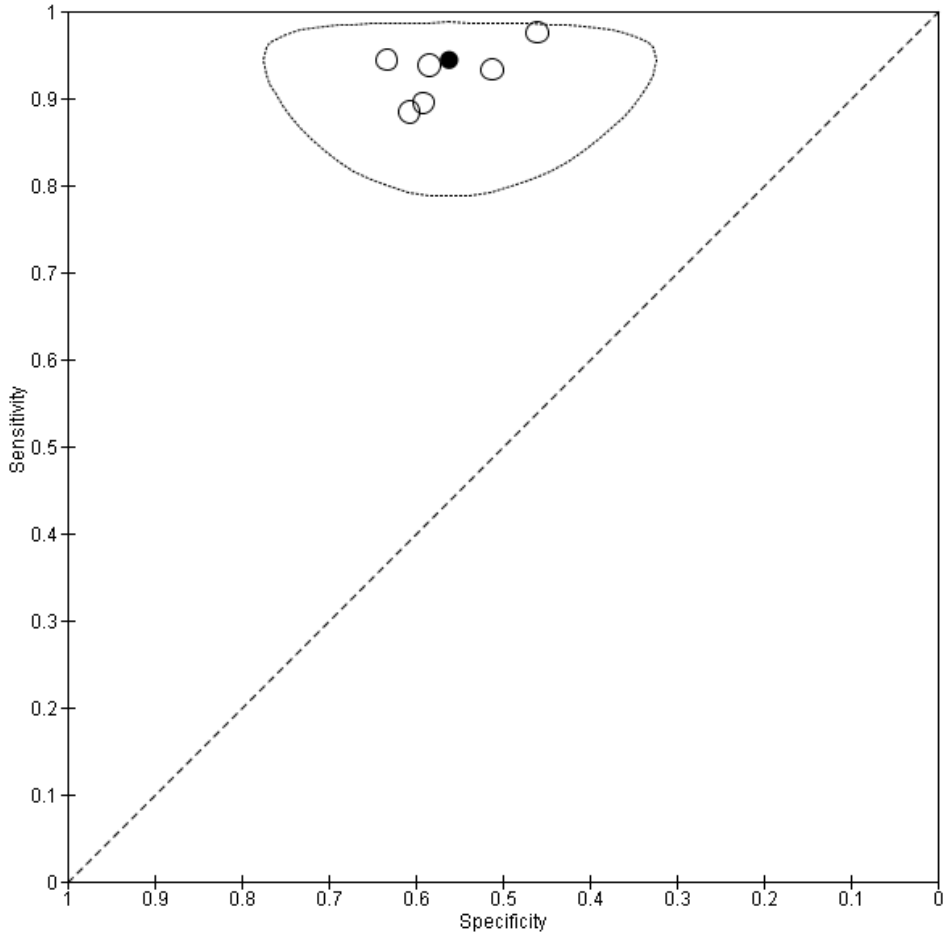

Sensitivity
Specificity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

Figure 274: 5 way: malignant or suspicious or one grade of indeterminate (negative = lower grade of indeterminate or benign)

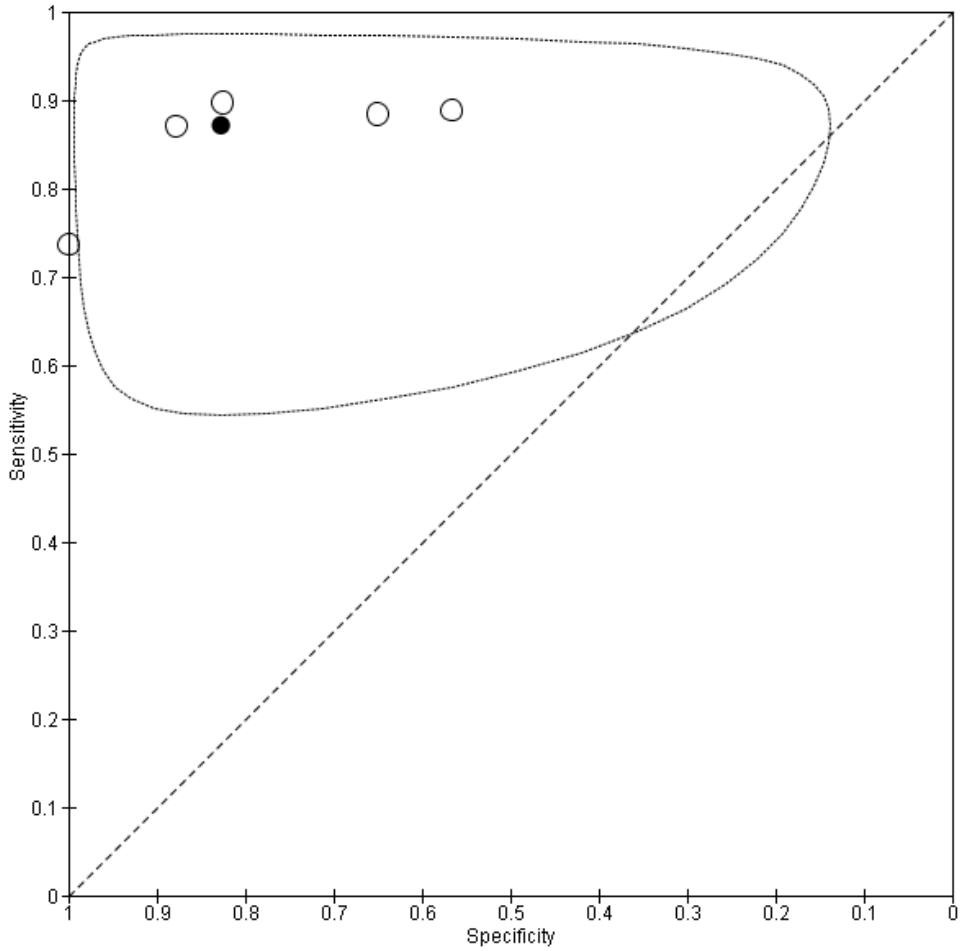


Figure 275: 5 way: malignant (negative = suspicious or two grades of indeterminate or benign)

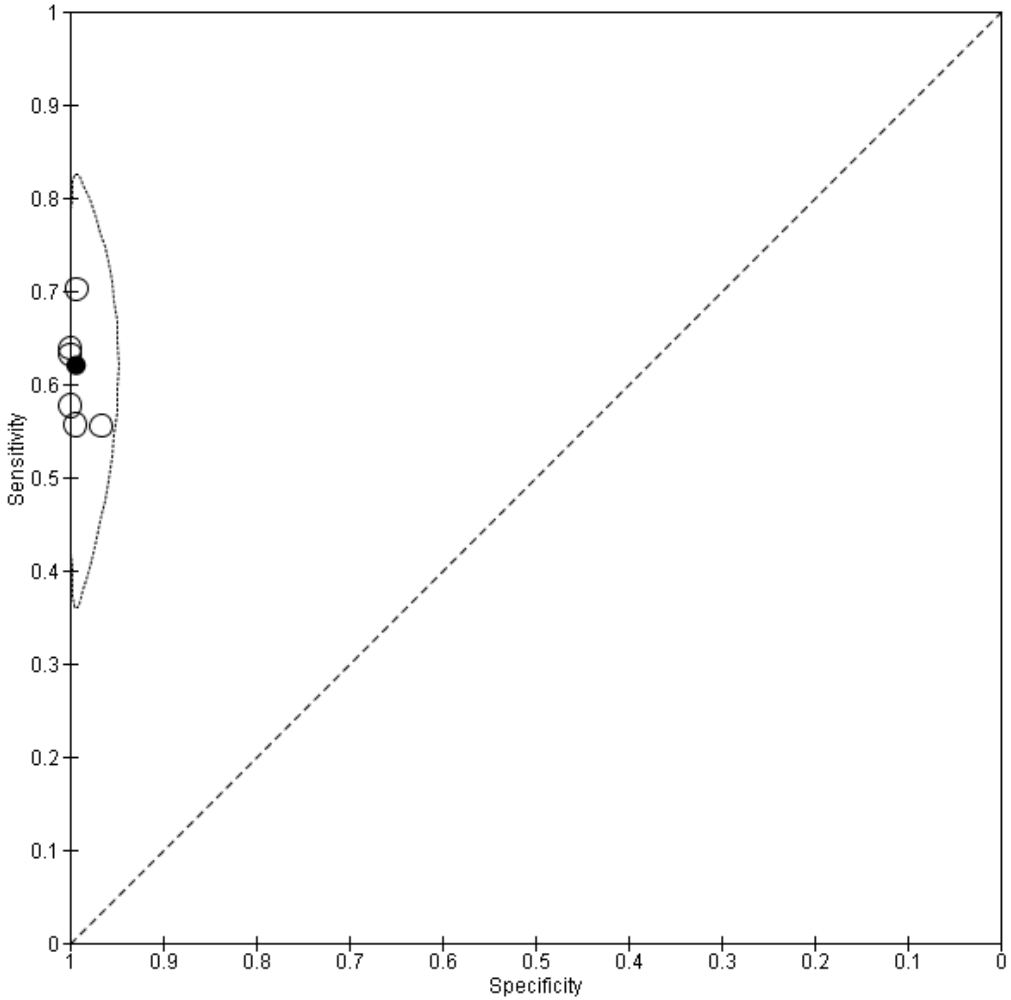

Sensitivity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Specificity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

Figure 276: 1 or more inclusions

No meta-analysis carried out as less than 3 studies

Figure 277: 1 or more grooves

No meta-analysis carried out as less than 3 studies

Figure 278: 2 or more grooves

No meta-analysis carried out as less than 3 studies

Figure 279: 3 or more grooves

No meta-analysis carried out as less than 3 studies

FNAC, no ROSE, smear only, with prior US

Figure 280: Bethesda Grade III or above

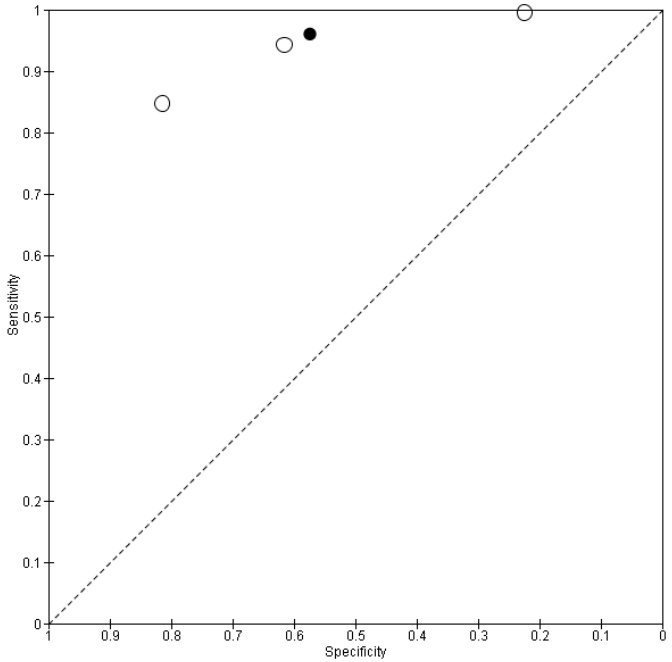


Figure 281: Bethesda Grade IV or above

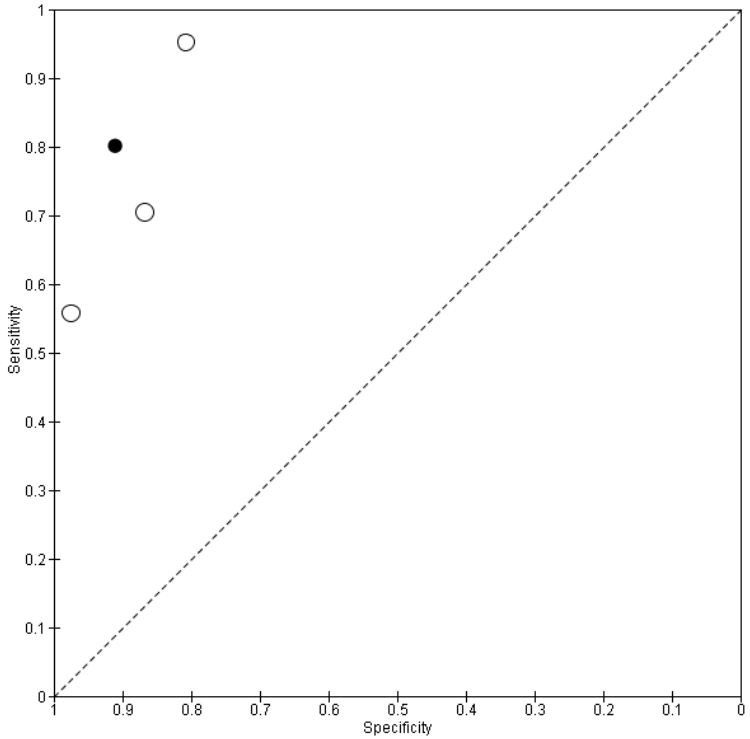


Figure 282: Bethesda Grade V or above

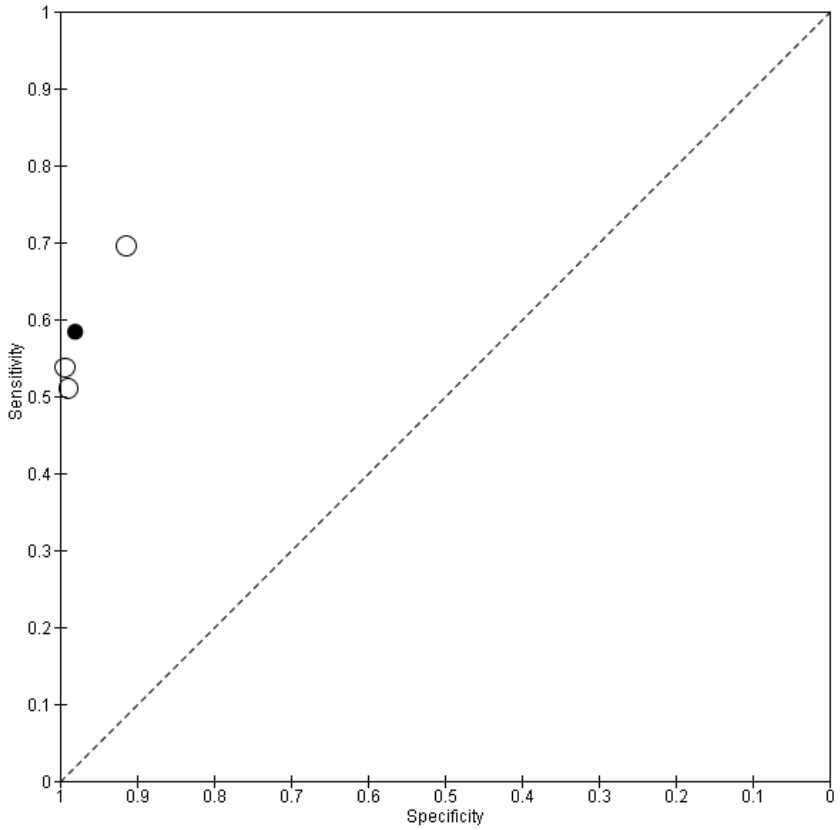


Figure 283: Bethesda Grade VI or above

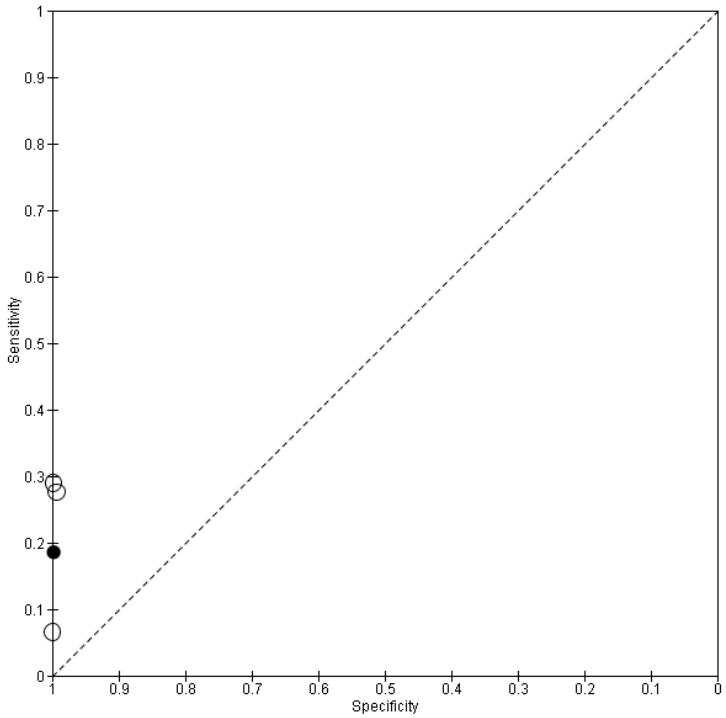


Figure 284: 2 way: malignant versus benign

No meta-analysis carried out as less than 3 studies

Figure 285: 3 way: suspicious or malignant (negative = benign)

No meta-analysis carried out as less than 3 studies

Figure 286: 3 way: malignant (negative = suspicious or benign)

No meta-analysis carried out as less than 3 studies

Figure 287: 4 way De May classification: malignant, suspicious, non malignant follicular proliferation (negative = benign)

No meta-analysis carried out as less than 3 studies

Figure 288: 4 way De May classification: malignant, suspicious (negative = benign, non malignant follicular proliferation)

No meta-analysis carried out as less than 3 studies

Figure 289: 4 way De May classification: malignant (negative = benign, non malignant follicular proliferation, suspicious)

No meta-analysis carried out as less than 3 studies

Figure 290: 4 way Piana classification: C3 or more

No meta-analysis carried out as less than 3 studies

Figure 291: 4 way Piana classification: C4 or more

No meta-analysis carried out as less than 3 studies

Figure 292: 4 way Piana classification: C5 or more

No meta-analysis carried out as less than 3 studies

Figure 293: 4 way generic: malignant, suspicious, indeterminate (benign = negative)

No meta-analysis carried out as less than 3 studies

Figure 294: 4 way generic: malignant, suspicious, indeterminate (benign = negative)

No meta-analysis carried out as less than 3 studies

FNAC, no ROSE, smear, with cytospin and/or cell-block, without prior US

Figure 295: Bethesda Grade III or above

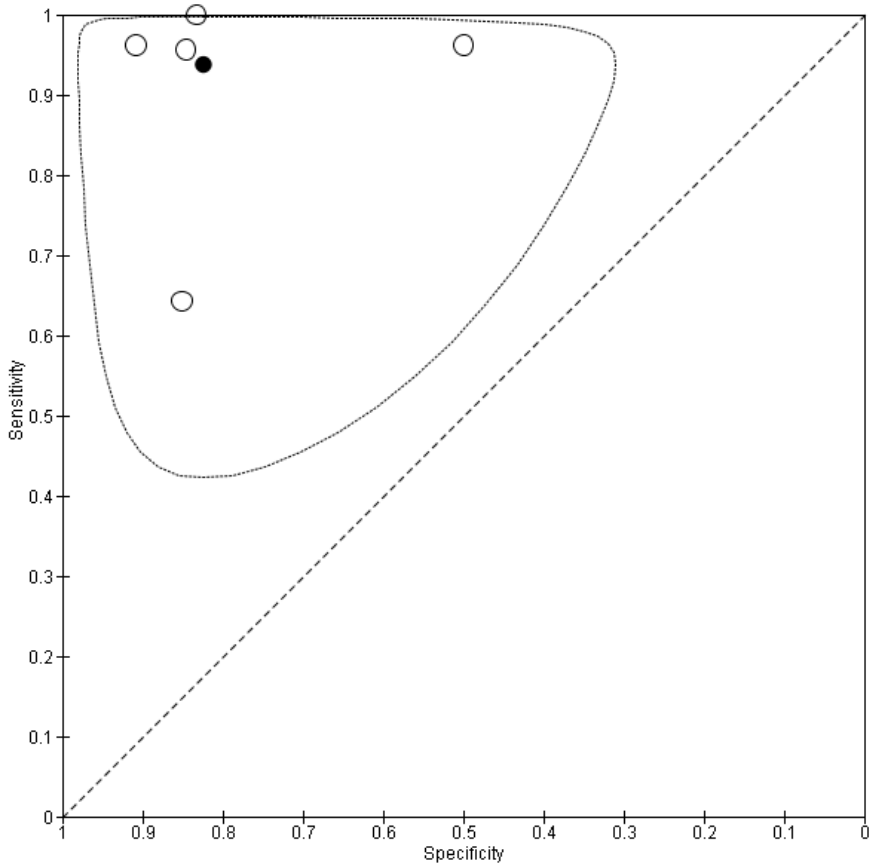


Figure 296: Bethesda Grade IV or above

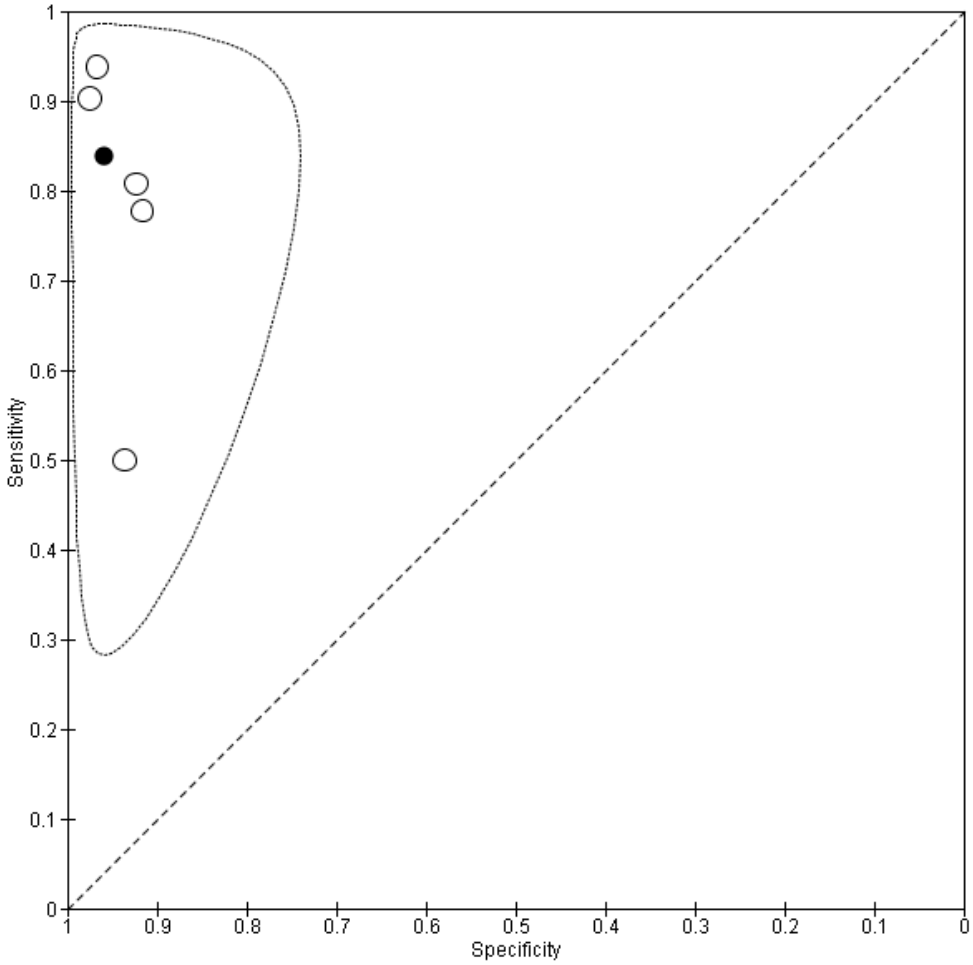


Figure 297: Bethesda Grade V or above

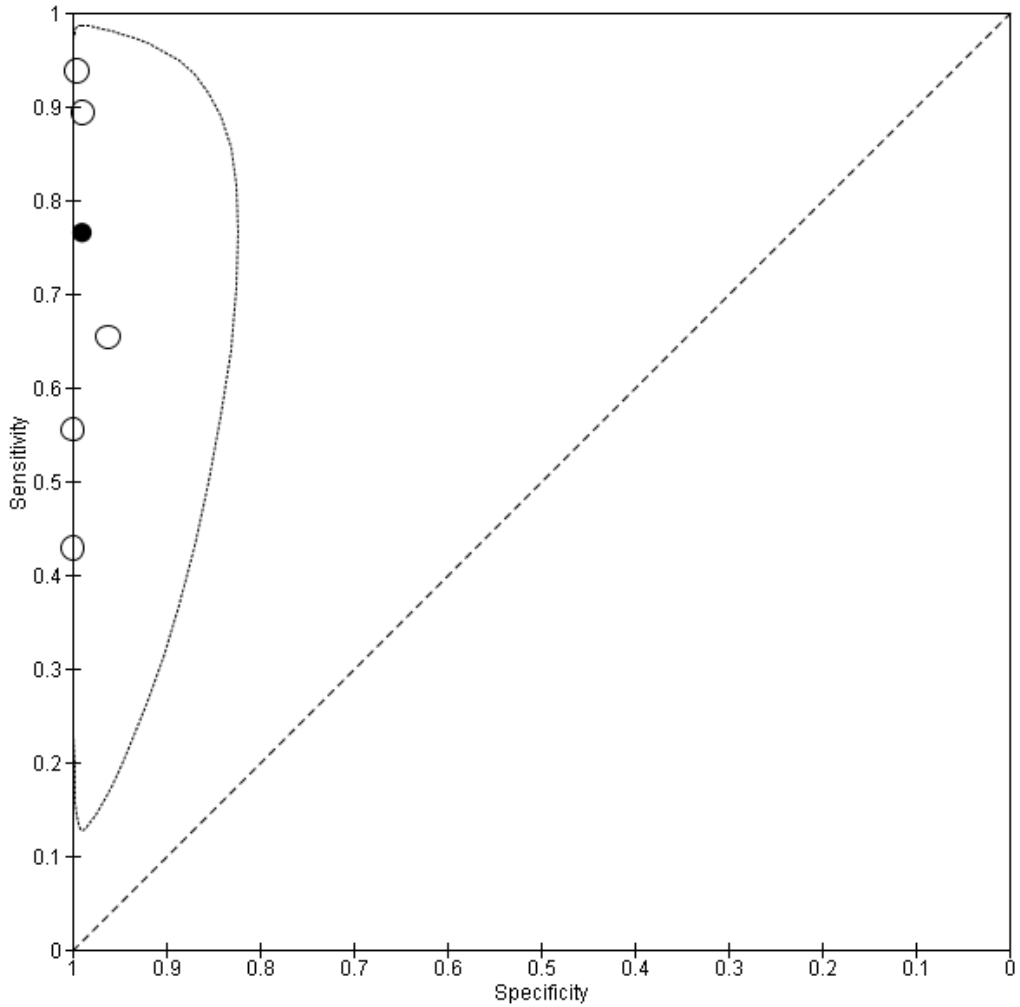


Figure 298: Bethesda Grade VI or above

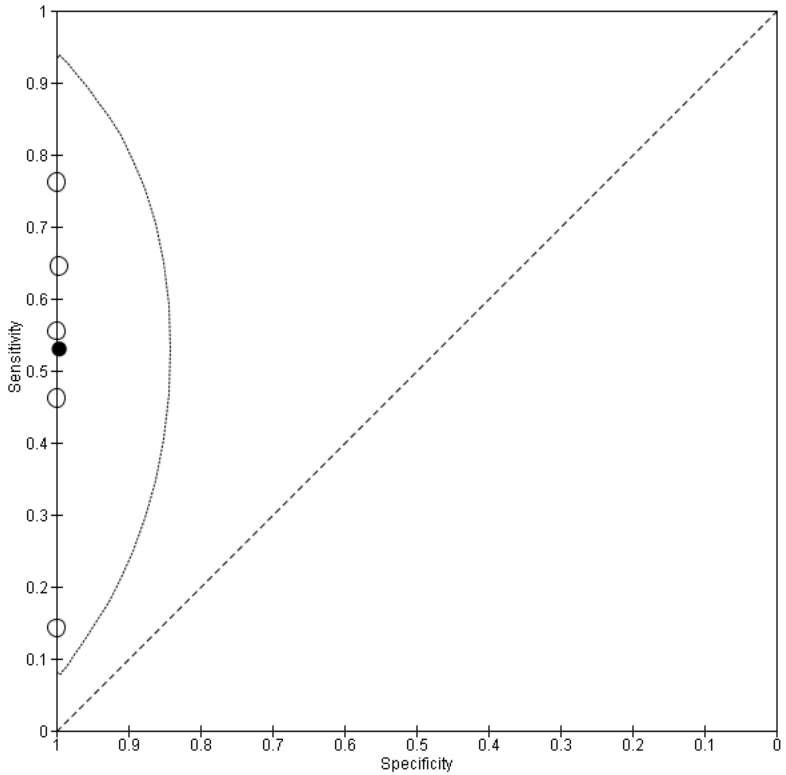


Figure 299: 2 way: malignant v benign

No meta-analysis carried out as less than 3 studies

Figure 300: 3 way: malignant or suspicious (negative = benign)

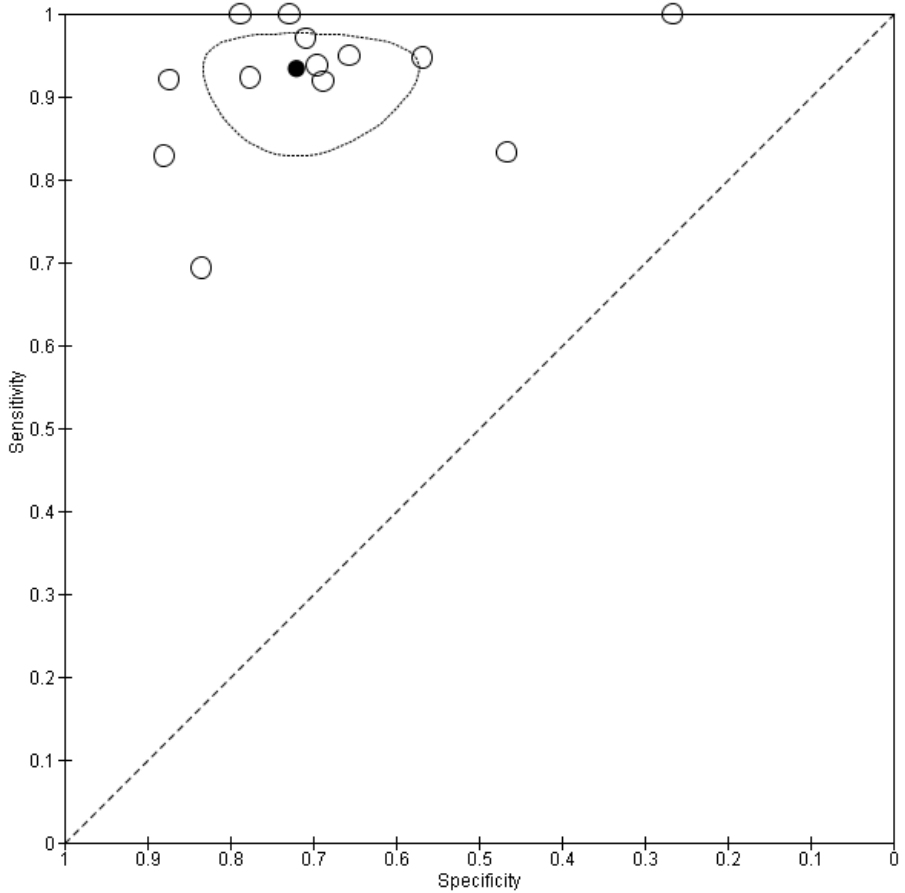


Figure 301: 3 way: malignant (negative = benign or suspicious)

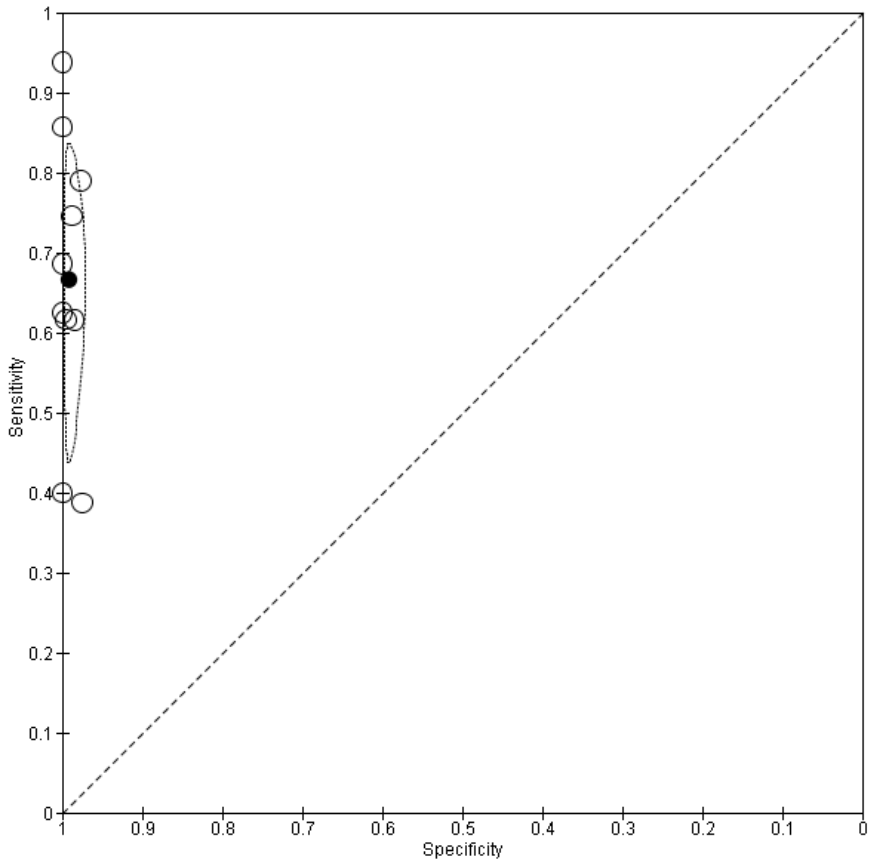


Figure 302: 4 way: malignant, suspicious, indeterminate (negative = benign)

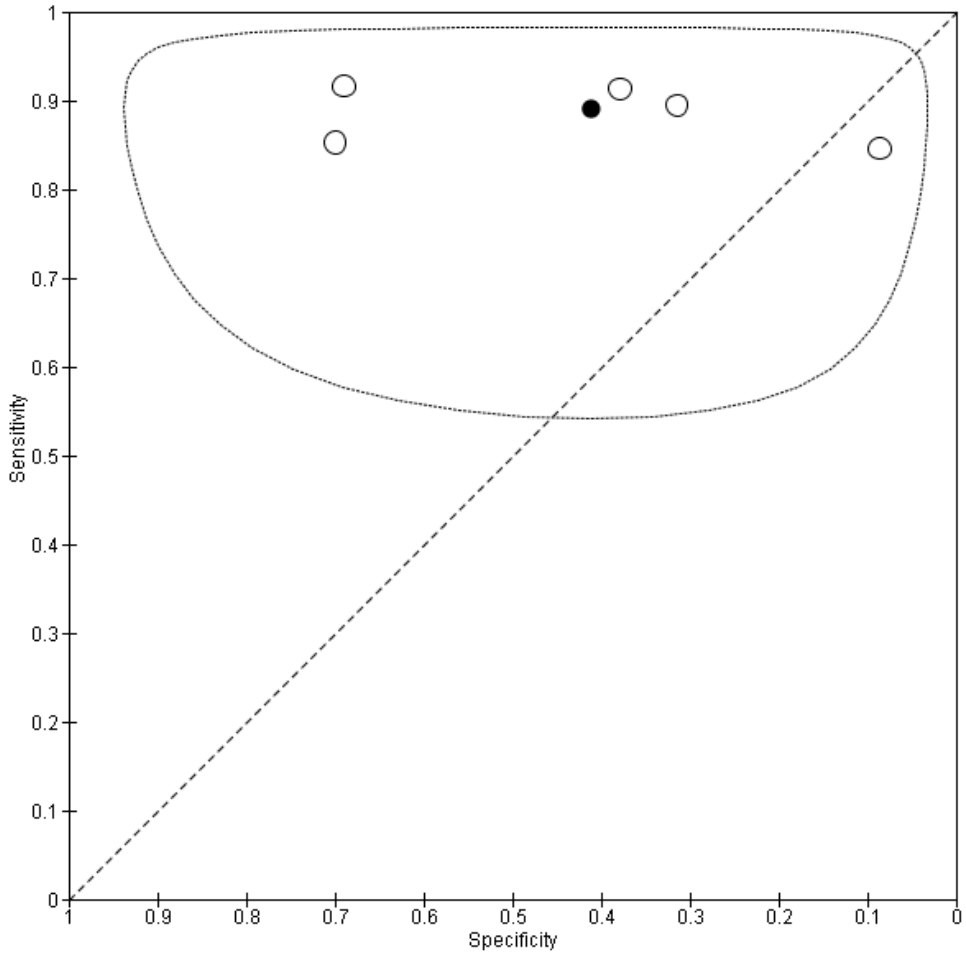


Figure 303: 4 way: malignant, suspicious (negative = benign, indeterminate)

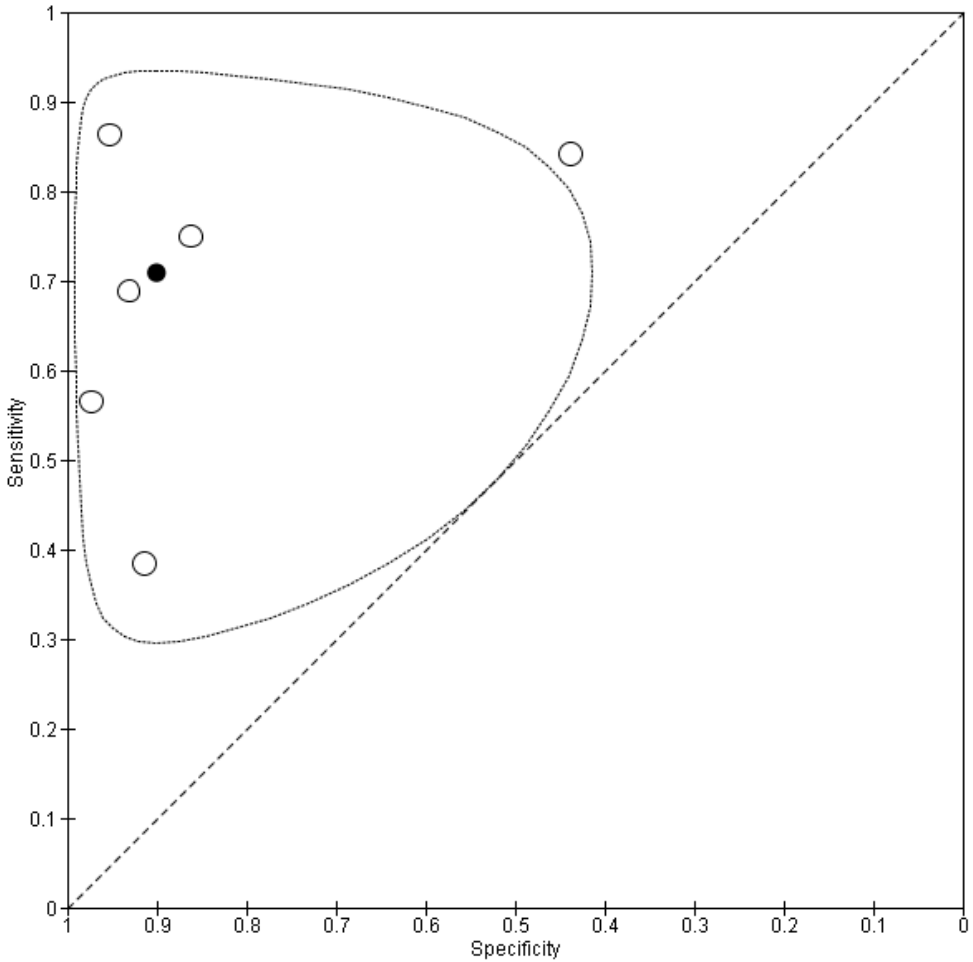


Figure 304: 4 way: malignant (negative = benign, indeterminate, suspicious)

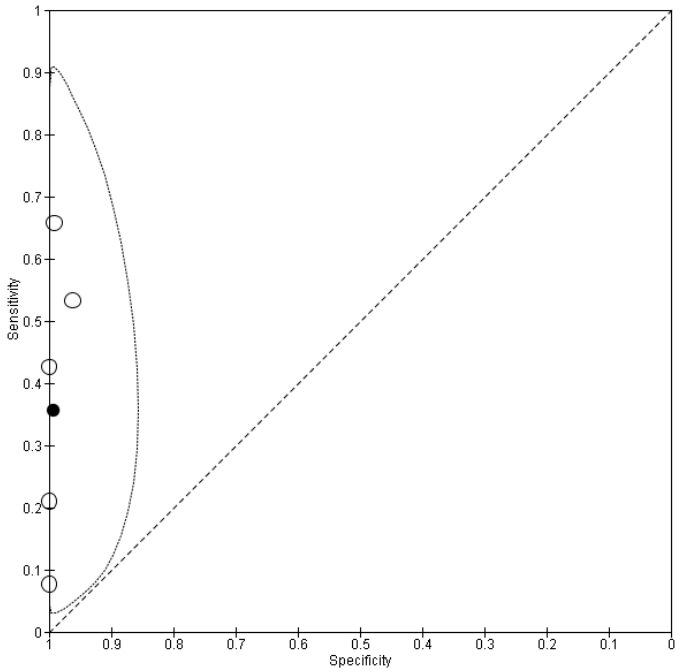


Figure 305: 5 way: malignant, suspicious, 2 grades of indeterminate (negative = benign)

No meta-analysis carried out as less than 3 studies

FNAC, no ROSE, smear, with cytospin and/or cell-block, with prior US

Figure 306: Bethesda Grade III or above

No meta-analysis carried out as less than 3 studies

Figure 307: Bethesda Grade IV or above

No meta-analysis carried out as less than 3 studies

Figure 308: Bethesda Grade V or above

No meta-analysis carried out as less than 3 studies

Figure 309: Bethesda Grade VI

No meta-analysis carried out as less than 3 studies

Figure 310: Benign or above

No meta-analysis carried out as less than 3 studies

FNAC, with ROSE, smear only, without prior US

Figure 311: Bethesda Grade III or above

No meta-analysis carried out as less than 3 studies

Figure 312: Bethesda Grade IV or above

No meta-analysis carried out as less than 3 studies

Figure 313: Bethesda Grade V or above

No meta-analysis carried out as less than 3 studies

Figure 314: Bethesda Grade VI

No meta-analysis carried out as less than 3 studies

Figure 315: 3 way: malignant and suspicious (negative = benign)

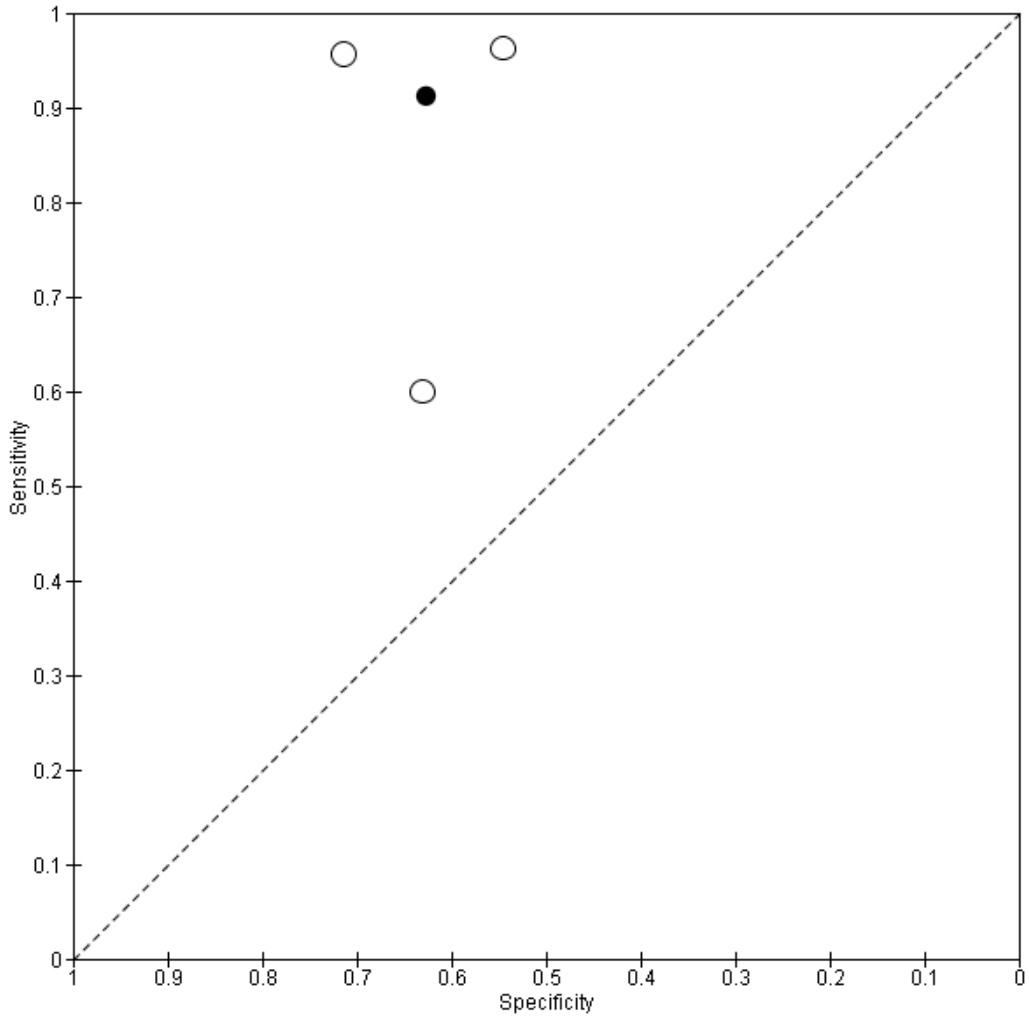


Figure 316: 3 way: malignant (negative = benign and suspicious)

No meta-analysis carried out as less than 3 studies

Figure 317: 4 way: malignant, suspicious, indeterminate (negative = benign)

No meta-analysis carried out as less than 3 studies

Figure 318: 4 way: malignant, suspicious (negative = benign, indeterminate)

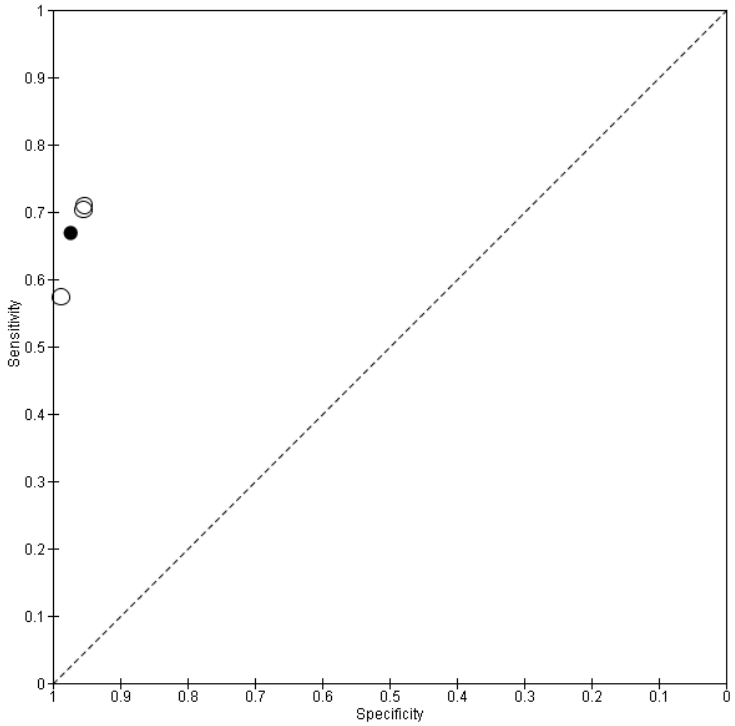


Figure 319: 4 way: malignant (negative = benign, indeterminate, suspicious)

No meta-analysis carried out as less than 3 studies

FNAC, with ROSE, smear only, with prior US

Figure 320: intermediate or malignant

No meta-analysis carried out as less than 3 studies

FNAC, with ROSE, smear, with cytospin and/or cell-block, without prior US

Figure 321: 3 way: suspicious or malignant (negative = benign)

No meta-analysis carried out as less than 3 studies

Figure 322: 3 way: malignant (negative = suspicious or benign)

No meta-analysis carried out as less than 3 studies

Figure 323: 4 way: malignant, suspicious, indeterminate (negative = benign)

No meta-analysis carried out as less than 3 studies

Figure 324: 4 way: malignant, suspicious (negative = benign, indeterminate)

No meta-analysis carried out as less than 3 studies

Figure 325: 5 way: malignant, suspicious, 2 grades of indeterminate (negative = benign)

No meta-analysis carried out as less than 3 studies

Figure 326: 5 way: malignant, suspicious (negative = 2 grades of indeterminate, benign)

No meta-analysis carried out as less than 3 studies

Figure 327: 5 way: malignant, suspicious (negative = suspicious, lower grade of indeterminate, benign)

No meta-analysis carried out as less than 3 studies

Figure 328: 5 way: malignant (negative = suspicious, 2 grades of indeterminate, benign)

No meta-analysis carried out as less than 3 studies

FNAC, with ROSE, smear, with cytospin and/or cell-block, with prior US

Figure 329: indeterminate follicular, indeterminate Hurtle, Suspicious for malignancy, or positive

No meta-analysis carried out as less than 3 studies

Figure 330: Suspicious for malignancy, or indeterminate follicular or positive

No meta-analysis carried out as less than 3 studies

Figure 331: Suspicious for malignancy, or positive

No meta-analysis carried out as less than 3 studies

Figure 332: Positive for malignancy

No meta-analysis carried out as less than 3 studies

Core biopsy, without prior US

Figure 333: carcinoma or neoplasm (versus benign)

No meta-analysis carried out as less than 3 studies

Figure 334: carcinoma (versus benign/indeterminate)

No meta-analysis carried out as less than 3 studies

Figure 335: CB grades V and VI

No meta-analysis carried out as less than 3 studies

Figure 336: CB grades III, V and VI

No meta-analysis carried out as less than 3 studies

Figure 337: positive (versus negative) with CEUS guidance

No meta-analysis carried out as less than 3 studies

Figure 338: positive (versus negative) with US guidance

No meta-analysis carried out as less than 3 studies

Core biopsy, with prior US

Figure 339: indeterminate, follicular neoplasm, suspicious for malignancy, or malignant

No meta-analysis carried out as less than 3 studies

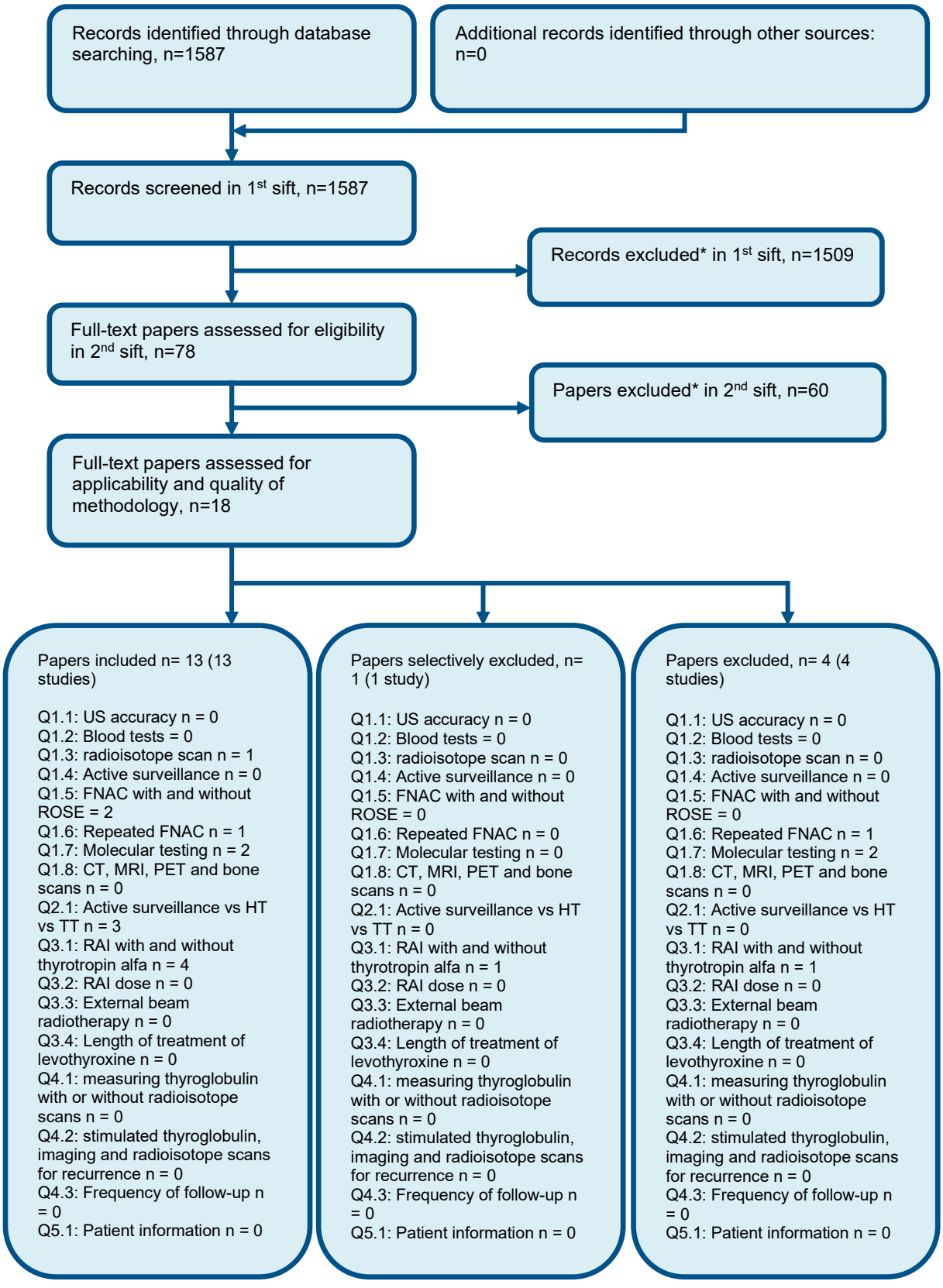
Figure 340: follicular neoplasm, suspicious for malignancy, or malignant

No meta-analysis carried out as less than 3 studies

Figure 341: suspicious for malignancy, or malignant

No meta-analysis carried out as less than 3 studies

Appendix G – Economic evidence study selection



* Non-relevant population, intervention, comparison, design or setting; non-English language

Appendix H – Economic evidence tables

Study	Breeze 2014[54]			
Study details	**Population & interventions**	**Costs**	**Other outcomes**	**Cost effectiveness**
Economic analysis: Cost-effectiveness analysis **Study design:** Cross-sectional diagnostic study **Approach to analysis:** FNAC results for patients prior to a trial of biomedical scientist rapid onsite evaluation were compared prospectively with the results from four such clinics in which rapid onsite evaluation by a biomedical scientist was performed. **Perspective:** UK NHS **Time horizon:** NR **Discounting:** Costs: NR Outcomes: NR	**Population:** Adults with suspected thyroid cancer who underwent ultrasound guided FNAC with and without rapid onsite evaluation by a biomedical scientist **Cohort settings:** Median age: NR Male: NR N: 138 **Intervention 1:** FNA cytology without rapid onsite evaluation (ROSE) **Intervention 2:** FNA cytology with rapid onsite evaluation by a biomedical scientist (ROSE)	**Total costs (mean per patient):** Intervention 1: £182.95 Intervention 2: £235 Incremental (2−1): £52.05 (95% CI: NR; p=NR) **Currency & cost year:** 2012 UK pounds **Cost components incorporated:** Ultrasound-guided FNAC, repeated FNAC, biomedical scientist evaluation	**Primary outcomes:** Adequate samples (not requiring repeated FNAC): Intervention 1: 72% Intervention 2: 86% Incremental (2−1): 14% (95% CI: NR; P = 0.448) **Secondary outcomes:** Duration of visit (mean per patient): Intervention 1: 13 mins Intervention 2: 19 mins Incremental (2−1): 6 mins (95% CI: NR; p=NR) Number of patients receiving a FNAC in a day in an average clinic: Intervention 1: 13 people Intervention 2: 10 people Incremental (2−1): -3 people (95% CI: NR; p=NR)	FNAC with ROSE costs £378 more for each additional satisfactory sample (different than non-diagnostic Thy1) **Analysis of uncertainty:** NR

Data sources				
Health outcomes: Adequacy rates were determined by retrospective review of the written pathology reports for the 20 consecutive clinics preceding the trial, and by review of the final pathology reports for each case taken after implementation of rapid onsite evaluation. The result used for statistical purposes was the final pathology result of all an individual patient's slides taken including any in-clinic re-aspiration samples. The adequacy rate of FNA samples and accuracy of histological diagnosis were determined before and after the introduction of rapid onsite evaluation by a biomedical scientist. The diagnosis determined by FNA cytology was also compared with the eventual diagnosis in those patients in whom surgery was undertaken and therefore histology was available. The accuracy of FNA cytology was determined using those samples from which a diagnosis could be made (not just those deemed adequate) and which subsequently went on to have a tissue sample taken. For non-thyroid aspirates as there are no generally accepted criteria for cellular adequacy the criteria for cell adequacy were those used by the reporting pathologist, based on the subjective assessment of all the submitted slides taken from the final diagnostic cytology report. **Quality-of-life weights:** NA **Cost sources:** Cost of ultrasound-guided FNA cytology was obtained from Borget 2008. The cost of in-clinic rapid onsite evaluation by biomedical scientists was obtained from Poller 2013. The effect on timing of introducing a biomedical scientist was assessed using a time-in-motion analysis in a representative sample of 10 out of the total of 20 clinics. However, the cost of additional time for ultrasound or radiology attendance was not included.				
Comments				
Source of funding: NR **Limitations:** Small sample size in the ROSE arm. Clinical outcomes were not reported. Time horizon or duration over which clinic visits took place was not reported. FNAC costs were based on a French source. The estimation of the additional cost for ROSE is not adequately explained and likely overestimates the cost per hour of a cytopathologist in the UK. Cost and consequences of surgery or further testing if the second FNAC is inadequate (e.g. diagnostic thyroid lobectomy) were not included, potentially underestimating the impact of improved sampling associated with rapid onsite evaluation by biomedical scientist. Resource use was obtained from single centre study of unclear generalizability to wider UK context. Sensitivity analyses were not reported. Potential conflicts of interests were not declared. Funding source was not reported. **Other:** None				
Overall applicability:[a] Partially applicable **Overall quality:**[b] Potentially serious limitations				

Abbreviations: 95% CI= 95% confidence interval; CC= cost–comparison; da= deterministic analysis; FNAC = fine needle aspiration cytology; ICER= incremental cost-effectiveness ratio; NA = not applicable; NR= not reported; pa= probabilistic analysis; QALYs= quality-adjusted life years; ROSE= Rapid on-site evaluation.
(a) Directly applicable / Partially applicable / Not applicable
(b) Minor limitations / Potentially serious limitations / Very serious limitations

Study	**Feletti 2021[105]**			
Study details	**Population & interventions**	**Costs**	**Other outcomes**	**Cost effectiveness**
Economic analysis: Cost-effectiveness analysis	**Population:** people with suspected thyroid cancer who underwent ultrasound guided FNAC with and without the	**Total costs (mean per patient):** Intervention 1: £99 Intervention 2: £114	**Thy1 samples** Intervention 1: 7.9% Intervention 2: 2.9% Incremental (2−1): - 5%	FNAC with ROSE costs £300 more for each additional satisfactory sample (different than non-diagnostic Thy1) **Analysis of uncertainty:**

Study design: Decision tree model based on retrospective accuracy analysis **Approach to analysis:** US-guided FNACs of thyroid nodules conducted in a single centre were retrospectively compared with some randomly adopting cytopathologist assistance (including ROSE). A decision tree model was developed alongside to estimate cost-effectiveness **Perspective:** Italian NHS **Time horizon:** 1 year **Discounting:** Costs: NR Outcomes: NR	assistance of a cytopathology **Cohort settings:** Median age: 58 Male: 25.7% N: 4589 **Intervention 1:** US-guided FNAC without cytopathologist assistance **Intervention 2:** US-guided FNAC with cytopathologist assistance	Incremental (2−1): £15 (95% CI: NR; p=NR) **Currency & cost year:** 2020 Euros (presented here as 2020 UK pounds[b]) **Cost components incorporated:** Ultrasound-guided FNAC, repeated FNAC, cyto-assistance assessment	(95% CI: NR; P > 0.001)	No analysis of uncertainty was conducted
Data sources				
Health outcomes: Adequacy rates were determined by retrospective review of FNACs conducted in a single centre with some randomly receiving cytopathology assistance. FNACs conducted to refine a diagnosis of thyroiditis and FNACs performed on anatomic structures other than thyroids (e.g. parathyroid or lymph-nodes) were excluded. **Quality-of-life weights:** NA **Cost sources:** The cost of a FNAC without assistance was calculated with the assistance of the institution's quality control department splitting the cost of the laboratory analysis and radiological component. The cost of adding a cytopathologist was separately calculated estimating 20 minutes needed for the execution of FNAC.				
Comments				

Source of funding: No funding was obtained for this research **Limitations:** No analysis of uncertainty was conducted. Cytology assistance in this analysis is not limited to on-site evaluation (ROSE) but includes the presence of the cytopathologist during the entire procedure, who helps the radiologist choosing the best site of the nodule to perform the biopsy and assists the procedure in other ways. Thus, benefits estimated in this analysis may be larger than the results of other analyses based on ROSE only. Baseline inadequate rates come from a single Italian centre with an excellent performance. This may underestimate the cost-effectiveness of ROSE and cytopathology assistance as these are known to be particularly cost-effective when introduced to centres with poor performance. Relative treatment effects expressed as the reduction of FNAC receiving a non-diagnostic cytology THY1 were estimated from a single centre and it is unclear whether they can be generalised to other centres. Cost and consequences of surgery or further testing if the second FNAC is inadequate (e.g. diagnostic thyroid lobectomy) were not included, potentially underestimating the impact of improved sampling associated with rapid onsite evaluation by biomedical scientist. Resource use and unit costs were obtained from a single Italian centre of unclear generalisability to UK context. **Other:** None

Overall applicability:[b] Partially applicable **Overall quality:**[c] Potentially serious limitations

Abbreviations: 95% CI= 95% confidence interval; CC= cost–comparison; da= deterministic analysis; FNAC = fine needle aspiration cytology; ICER= incremental cost-effectiveness ratio; NA = not applicable; NR= not reported; pa= probabilistic analysis; QALYs= quality-adjusted life years; ROSE= Rapid on-site evaluation.

(a) Converted using 2020/2021 purchasing power parities[268]

(b) Directly applicable / Partially applicable / Not applicable

(c) Minor limitations / Potentially serious limitations / Very serious limitations

Appendix I – Excluded studies

I.1 Clinical studies

Table 30: Studies excluded from the clinical review

Reference	Reason for exclusion
Aftab, 2005 #1090[5]	Cannot be sourced
Ahari, 2020 #1095[10]	No diagnostic accuracy data provided
Ahn, 2010 #1097[12]	Looked at the diagnostic accuracy of US
Ahn, 2021 #1096[11]	Not all participants had histopathological gold standard (some had cytological gold standard)
Akerman, 1985 #1098[13]	Data insufficient for diagnostic accuracy calculation
Akhavan, 2016 #1099[14]	No details of FNAC type
Akhtar, 2007 #1100[15]	No details of FNAC type
Alalawi, 2019 #1101[20]	No details of FNAC type
Al-Chalabi, 2019 #1102[16]	No diagnostic accuracy data relating to FNAC
Al-Dbahri, 2001 #1103[17]	No details of FNAC type
Alhashem, 2021[21]	Type of FNAC not reported
Alshaikh, 2018 #1105[22]	Type of FNAC not reported for all particpants
Anderson, 2014 #1110[26]	Not a diagnostic accuracy study
Archondakis, 2009 #1111[27]	No details of FNAC type
Arena, 2014 #1112[28]	Restricted to people at THY4 and 5
Aysan, 2017 #1115[31]	Not all CNB categories given opportunity for surgery - therefore the diagnostic accuracy analysis only performed with appropriate GS for people of thy3 and above. This will skew accuracy of the categories given surgery.
Bahar, 2003 #1116[33]	No diagnostic accuracy data provided
Bajaj, 2006[34]	Serious inconsistencies between tabular results and text
Balas, 1985 #1118[35]	Statistics paper; no diagnostic accuracy analysis
Bapat, 1992 #1119[36]	No details on FNAC type
Basharat, 2011 #1120[37]	No details of FNAC type
Baskin, 1987 #1122[39]	Not all participants had histopathological gold standard
Beecham, 1988 #1123[40]	Not all participants had histopathological gold standard
Bernante, 1998 #1126[43]	Did not evaluate diagnostic accuracy of FNA
Bhartiya, 2016 #1127[44]	Data not reported clearly enough to permit extraction of raw data
Bhatki, 2008 #1128[45]	No definition of gold standard
Bhatti, 2010 #1129[46]	No details of FNAC type
Bisi, 1992 #1131[48]	Non-systematic review of literature
Blumenfeld, 1999 #1132[49]	Not relevant to protocol question
Bozbiyik, 2017 #1135[52]	No details of FNAC type
Breeze, 2014 #74[54]	Insufficient data to calculate sensitivity and specificity
Burch, 1996 #1139[56]	No details of FNAC type

Reference	Reason for exclusion
Buzdar, 2016[57]	Type of FNAC not reported
Caleo, 2016 #1140[58]	Not all CNB categories given opportunity for surgery - therefore the diagnostic accuracy analysis only performed with appropriate GS for people of thy3B and above. This will skew accuracy of the categories given surgery.
Camargo, 2007 #1141[59]	Evaluated a combined US and FNAC score
Can, 2009 #77[60]	Cost effectiveness paper
Cappelli, 2009 #1144[62]	Opinion piece
Caraci, 2002 #1145[63]	No details of FNAC type
Carpi, 1994 #1146[64]	unavailable for loan
Cavallo, 2017 #1147[65]	No details of FNAC type
Chakravarthy, 2018 #1148[66]	Not all participants had histopathological gold standard
Chen, 1998 #1150[68]	No details of FNAC type
Choi, 2014 #1152[71]	Not all participants had histopathological gold standard
Chowdhury, 2008 #1154[73]	No details of FNAC type
Christ, 1979 #1155[74]	Unavailable for loan
Chu, 1979 #1156[75]	Unavailable for loan
Ciatti, 1983 #1157[76]	Unable to source
Ciobanu, 2006 #1158[77]	No diagnostic accuracy analysis
Clary, 2005 #1159[78]	FNAC ratings limited to follicular lesions and follicular neoplasms
Colacchio, 1980 #1160[79]	Not all participants had histopathological gold standard
Cristo, 2016 #1162[81]	Excluded from accuracy analysis those with unsatisfactory, indeterminate (class III) and class IV lesions
Crowe, 2011 #1163[83]	Gold standard unclear - not reported that all had histopathology
Daskalakis, 2008 #1165[86]	Theoretical paper involving design of a multi-classifier system
Davidov, 2010 #1166[87]	No details of FNAC type
Davoudi, 1997 #1168[89]	No details of FNAC type
Dellal, 2021 #1171[92]	No details of FNAC type
Deshpande, 1997 #1172[93]	Restricted to FNAC grading of follicular neoplasms
Di Benedetto, 2013 #1173[94]	Not all participants had histopathological gold standard
Duek, 2002 #1174[95]	No details of FNAC type
Dumitriu, 1984 #1175[96]	Not all participants had histopathological gold standard
El Hag, 2003 #1178[99]	Gold standard differentiated neoplasms from benign, not malignant from benign
Erdogan, 1998 #1179[100]	No diagnostic accuracy analysis
Ersoz, 2016 #1180[101]	No UK source
Essex-Sorlie, 2000 #1181[102]	No details of FNAC type
F, 2011 #1182[103]	No details of FNAC type
Fadda, 1998 #1183[104]	Restricted to FNAC grading of follicular lesions

Reference	Reason for exclusion
Ferraz de Oliveira, 2019 #1185[107]	Unclear if histopathology used as GS for all patients
Flanagan, 2006 #1186[109]	Repeat FNAC in people with initially benign cytological results
Fon, 1996 #1187[110]	No details of FNAC type
Frable, 1979 #1191[114]	Not all participants had histopathological gold standard (some had long term clinical observation)
Frable, 1980 #1188[112]	Not all participants had histopathological gold standard (some had long term clinical observation)
Frable, 1982 #1189[111]	No useful data pertaining to thyroid nodules
Frable, 1986 #1190[113]	Unclear if histopathology used as GS for all patients
Franklyn, 1987 #1194[117]	Likely that clinical follow up used as GS for most patients
Franklyn, 1993 #1193[116]	Unclear if all participants had histopathological gold standard
Friedman, 1979 #1195[118]	Likely that clinical follow up used as GS for most patients
Frost, 1998 #1196[119]	Not all participants had histopathological gold standard (some had cytological gold standard)
Fulciniti, 2001 #1197[120]	Restricted to FNAC grading of follicular lesions
Furlan, 2005 #86[121]	Raw data not available in the paper
Galimberti, 1997 #1199[122]	No details of FNA; all patients had malignancy
Garg, 2015 #1202[125]	No details of FNAC type
Garg, 2018 #762[124]	Patients with bethesda score of benign not given histopathological gold standard (conservatively followed up)
Gibb, 1995 #1205[128]	Unavailable for loan
Godinho-Matos, 1992 #1206[129]	Tabular data conflated FNAC and clinical data; gold standard did not evaluate malignancy (neoplasms not malignancy)
Goldfarb, 1982 #1207[130]	Review article
Goulart, 2021 #1209[132]	Bethesda I,III and IV nodules excluded so does not represent population
Granados-Garcia, 2010 #1211[134]	In Spanish
Greenblatt, 2006 #1212[135]	No details of FNAC type
Guadagni, 1988 #1213[136]	No details of FNAC type
Gunes, 2015 #1214[137]	No details of FNAC type
Gupta, 2016 #1216[139]	No details of FNAC type
H, 2019 #1217[140]	Not all participants had histopathological gold standard (some had 1 year clinical follow up)
Ha, 2018 #1218[141]	Diagnostic accuracy of US (GS not wholly surgical histopathology)
Ha, 2021[142]	Combined FNAC and CNB biopsies in same analysis, without subgrouping
Haas, 1993 #1219[143]	Histopathology not used as GS for all patients
Haider, 2011 #1221[145]	Restricted to analysis of inadequate smears
Hajmanoochehri, 2015	Gold standard differentiated neoplasms and non-neoplasms, not malignancy versus non-malignancy
Hamaker, 1983 #1223[146]	Histopathology not used as GS for all patients
Hamburger, 1985 #1225[148]	No details of FNAC type

Reference	Reason for exclusion
Hamburger, 1988 #1224[147]	No diagnostic accuracy analysis
Harach, 1989 #1228[151]	unavailable for loan
Hawkins, 2021[154]	No diagnostic accuracy analysis
Hirokawa, 2020 #1232[156]	No non-malignant participants in sample so specificity not measured
Hoffman, 1986 #1233[157]	Non-systematic-review paper
Hong, 2020[158]	No diagnostic accuracy analysis
Hurtado-López, 2004 #1578[162]	Data not reported clearly enough to permit extraction of raw data
Irish, 1992 #1239[164]	No details of FNAC type
Irkorucu, 2007 #1240[165]	No details of FNAC type
Jing, 2012 #1244[169]	re-analysis of group of aspirates previously interpreted as AUS/FLUS - likely to be a narrow band of applicability
Kakudo, 2015 #1245[170]	Indeterminate nodules only evaluated
Karadeniz, 2019 #1246[171]	No details of FNAC type
Karstrup, 2001 #1247[172]	GS differentiated neoplasms and non-neoplasms, not malignancy versus non-malignancy
Katagiri, 1994 #1248[173]	No details of FNAC type
Kawai, 2012 #1249[174]	No details of FNAC type
Kendall, 1989 #1251[176]	No diagnostic accuracy analysis
Khan, 1996 #1254[179]	No diagnostic accuracy analysis relevant to FNAC
Khan, 2004 #1252[177]	Cases restricted to people with FNAC grades of follicular neoplasms, Hurthle cell neoplasms and follicular carcinomas
Khan, 2013 #1253[178]	No UK source
Kikuchi, 2003 #1255[180]	No details of FNAC type
Kim, 2003 #1259[186]	Not all participants had histopathological gold standard (some had cytological gold standard)
Kim, 2008 #1256[181]	Only patients with suggestive malignant cytology or clinically suspicious of malignancy among the indeterminate category were referred to surgery for GS
Kim, 2014 #1258[184]	No details of FNAC type
Kim, 2021[183]	All benign on FNAC
Kim, 2022[185]	differentiated subtypes of follicular variant papillary thyroid carcinoma
Kini, 1980 #1261[189]	Vast majority in study were malignant or indeterminate on cytology (no benign)
Kizilkaya, 2014 #1263[190]	No details of FNAC type
Kline, 1973 #1264[191]	Not specific to thyroid cancer
Knezevic-Usaj, 2012 #1265[192]	Not in English
Kollur, 2003 #1268[195]	unavailable for loan
Krishnappa, 2013 #1270[197]	Gold standard differentiated neoplasms from benign, not malignant from benign
Kulstad, 2016 #1271[198]	No details of FNAC type
Lee, 2002 #1275[203]	raw data not clear enough to allow extraction of data
Lee, 2013 #1274[202]	Not all participants had histopathological gold standard

Reference	Reason for exclusion
Lewis, 2009 #1277[205]	Review paper
Linhares, 2021[209]	Type of FNAC not reported
Liu, 2021[212]	Restricted to patients with elevated serum calcitonin
Lo Gerfo, 1982 #1282[213]	Nonbenign on FNAC so not representative
Lobo, 2011 #1283[214]	Restricted to Thy 3a to Thy 5 only
Lodewijk, 2016 #1284[215]	No details of FNAC type
Lopez, 1997 #1285[216]	Not all participants had histopathological gold standard (some had 4 year follow up)
Lyu, 2019 #1078[218]	Nodules at Bethesda I,III and IV excluded from analysis
Makes, 2007 #1288[219]	No details of FNAC type
Malberger, 1985 #1289[220]	Unclear reporting of results
Manchanda, 2018 #1291[222]	Cannot be sourced
Mandal, 2011 #1293[223]	Cannot be sourced
Martinek, 2004 #1295[225]	No details of FNAC type
Mary Lilly, 2019 #1297[227]	Cannot be sourced
Masatsugu, 2005 #1298[228]	No details of FNAC type
Mathur, 2005 #1300[230]	Sample were restricted to people with cytology suggesting goitre or histology suggesting goitre
Maxwell, 1996 #1301[231]	No details of FNAC type
McCoy, 2007 #1302[232]	No details of FNAC type
McHenry, 1999 #1304[234]	Restricted to indeterminate findings on cytology
McIvor, 1993 #1305[235]	Restricted to Hurthle cell neoplasia on cytology/histology
Meng, 2019 #1308[238]	Special population with Hashimoto's thyroiditis
Miller, 1981 #1313[243]	No diagnostic accuracy analysis that specifically and clearly used histopathological findings as the GS
Miller, 1985 #1314[244]	Unclear description of gold standard
Miller, 1986 #1315[245]	Case control study where the gold standard was papillary cancer vs no cancer, as opposed to any thyroid malignancy vs no cancer.
Mo, 2017 #1316[246]	Not all participants had histopathological gold standard (some had 1 year clinical follow up)
Montironi, 1989 #1317[247]	Only discriminated between follicular adenoma and follicular carcinoma, not the wider issue of thyroid malignancy vs no malignancy
Montironi, 1990 #1319[249]	Sufficient quantitative data not provided for data extraction
Montironi, 1992 #1318[248]	Unable to access
Mora-Guzman, 2018 #1320[250]	No details of FNAC type
Morgan, 2003 #1321[251]	No details of FNAC type
Muratli 2014, #1323[253]	No details on FNAC type
Na, 2012 #1324[254]	Patients previously had non-diagnostic FNAC readings so atypical population
Na, 2015 #1325[255]	Patients previously had atypia/follicular lesion of undetermined significance FNAC readings so atypical population

Reference	Reason for exclusion
Ng, 1999 #1331[262]	Only discriminated between Hurthle cell adenoma and Hurthle cell carcinoma, not the wider issue of thyroid malignancy vs no malignancy
Nirmal, 2017 #1332[264]	Cannot be sourced
Norton, 1981 #1333[265]	Gold standard did not differentiate between adenoma and carcinoma
Pan, 2018 #1337[270]	Not all participants had histopathological gold standard (some had US follow up)
Pasha, 2021[271]	Type of FNAC not reported
Patel, 2014 #1338[272]	Gold standard differentiated neoplasms from benign, not malignant from benign
Pavithra, 2014 #1339[274]	No UK source
Postma, 2009 #1344[281]	No UK source
Raab, 1995 #1346[283]	Not all had histopathological gold standard
Rangaswamy, 2013 #1351[288]	Population only included malignant cases
Renshaw, 2001 #1353[290]	Not all participants had histopathological gold standard (some had cytological follow up)
Renshaw, 2002 #1354[291]	No diagnostic accuracy analysis
Renshaw, 2007 #1356[293]	Not all participants had histopathological gold standard (some had cytological follow up)
Renshaw, 2018 #1355[292]	Does not provide diagnostic accuracy data (no false positive rates)
Reyaz, 2020 #1357[294]	Not possible to extract accuracy data because data unclearly reported
Rosen, 1986 #1360[297]	Inadequate diagnostic accuracy data to allow extraction
Sabel, 1997 #1365[302]	Insufficient data to enable extraction (data for all FNAC categories not provided)
Sahin, 2006 #1366[303]	No details of FNAC type
Sangalli, 2001 #1367 [304]	All cases were lymphomas
Sarda, 1997 #1368[305]	No details of FNAC type
Sarkis, 2014 #1369[306]	No details of FNAC type
Schnurer, 1978 #1371[308]	No details of FNAC type
Seifman, 2011 #1376[313]	No details of FNAC type
Sengul, 2020[314]	Unclearly reported in terms of gold standard and the threshold of index test accuracy
Sharma, 2016 #1380[318]	No details on FNAC type
Sharma, 2017 #1381[319]	No details of FNAC type
Sharma, 2019[320]	Type of FNAC not reported
Sheahan, 2004 #1382[321]	General paper on neck masses
Shirzad, 2003 #1383[323]	No details of FNAC type
Shrestha, 2012 #1384[324]	No details of FNAC type
Sidawy, 1997 #1385[325]	Unclear reporting of results made it difficult to extract accuracy data
Silver, 1984 #1386[326]	No details of FNAC type
Silverman, 1986 #1388[328]	No details of FNAC type
Smadi, 2008 #1391[331]	No details of FNAC type
Soreide, 1979 #1393[333]	No diagnostic accuracy analysis

Reference	Reason for exclusion
Stanek-Widera, 2016 #1395[335]	Patients restricted to Bethesda category V in primary test
Stanek-Widera, 2016 #1396[336]	Patients restricted to Bethesda category IV in primary test
Stavric, 1980 #1397[337]	Not all participants had histopathological gold standard (some had 6 month - 3.5 year clinical follow up)
Suh, 2017 #1398[338]	Not a diagnostic accuracy analysis
Sulejmanovic, 2019 #1400[340]	All in study had thyroid cancer
Suwatthanarak, 2021[341]	Type of FNAC not reported
Taki, 1997 #1405[346]	Unclear data
Talpur, 2007 #1407[348]	No details of FNAC type
Tan, 2010 #943[349]	No details of FNAC type
Tao, 2021[350]	Type of FNAC not reported
Tee, 2007 #1409[351]	Literature review
Tele, 2020[352]	Type of FNAC not reported
Thomas, 1999 #1413[356]	Not relevant to diagnostic accuracy of FNAC in thyroid cancer
Thomsen, 1973 #1414[357]	insufficient data for inclusion (no data on TP and TN)
Tilak, 2002 #1415[358]	Covered head and neck region - no specific analysis for thyroid gland
Tomimori, 1999 #1416[359]	evaluated a combination of US and FNA
Werga, 2000 #1423[366]	Review - useful info on FNAC techniques
Williams, 2013 #1424[367]	No details of FNAC type
Wong, 1993 #1426[370]	insufficient data for inclusion (no data on TP and TN)
Wong, 2012 #1425[369]	Literature review
Wood, 2005 #1427[371]	Restricted to cellular follicular lesions
Wu, 2016 #1430[374]	No details of FNAC type
Wu, 2017 #1431[375]	restricted to nodules with indeterminate elastography
Wu, 2021 #1429[373]	Did not consider all classes of Bethesda in diagnostic accuracy evaluation
Xavier-Junior, 2020[376]	No diagnostic accuracy analysis; restricted to cystic nodules
Yagmur, 2018 #1434[379]	No details of FNAC type
Yassa, 2007 #1435[380]	Patients referred for surgery because of abnormal FNAC - therefore not possible to analyse accuracy in benign categories of FNAC, and exclusion of these groups will heavily skew accuracy in the remaining groups
Yildirim, 2021[382]	Type of FNAC not reported
Yilmaz, 2020[383]	Type of FNAC not reported
Ylagan, 2004 #1437[384]	Not possible to extract diagnostic accuracy data from the data provided
Yokozawa, 1995 #1439[386]	Surgery only offered to those with strong suspicion on FNA
Yoo, 2013 #1440[387]	No details of FNAC type
Zaidan, 2010 #1441[388]	No UK source
Zhang, 2012 #1446[393]	Unclear reporting of results making extraction of data impossible

Reference	Reason for exclusion
Zhong, 2015 #1447[394]	Not all participants had histopathological gold standard (some had 1 year clinical follow up)
Zosin, 2013 #1448[395]	Population with Hashimoto's thyroiditis
Zoulias, 2011 #1449[396]	No UK source

I.2 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.