CrossMark

META-ANALYSIS

Diagnostic accuracy of ultrasound-guided fine needle aspiration biopsy for thyroid malignancy: systematic review and meta-analysis

Naykky Singh Ospina[1,2] · Juan P. Brito[1,2] · Spyridoula Maraka[1,2] ·
Ana E. Espinosa de Ycaza[1] · Rene Rodriguez-Gutierrez[2] · Michael R. Gionfriddo[2,3] ·
Ana Castaneda-Guarderas[2,4] · Khalid Benkhadra[2,5,6] · Alaa Al Nofal[7] ·
Patricia Erwin[8] · John C. Morris[1] · M. Regina Castro[1] · Victor M. Montori[1,2]

Abstract

Purpose   To systematically appraise and summarize the available evidence about the diagnostic accuracy of ultrasound-guided fine needle aspiration biopsy (USFNA) for thyroid malignancy, and to explore the integration of these estimates with the probability of thyroid malignancy before USFNA.

Methods   A comprehensive search of multiple databases from each database's inception to August 2014 was performed. Eligible studies included those that evaluated patients with thyroid nodules who underwent USFNA and subsequent evaluation by histopathology or long-term follow-up.

Results   We identified 32 studies at moderate risk of bias evaluating the USFNA diagnostic characteristics for the diagnosis of thyroid malignancy. Results were imprecise and inconsistent across trials. The pooled likelihood ratio (LR) of thyroid malignancy for a benign USFNA result was 0.09 (95 % CI 0.06, 0.14; $I^2 = 33$ %), whereas the pooled LR for a malignant result was 197 (95 % CI, 68, 569; $I^2 = 77$ %). In the case of a suspicious for follicular neoplasm result, the pooled LR for malignancy was 0.6 (95 % CI, 0.4, 1.0; $I^2 = 84$ %) and 8.3 (95 % CI, 3.6, 19.2; $I^2 = 89$) for a result of suspicious for malignancy.

Conclusion   The available evidence regarding the diagnostic accuracy of USFNA warrants only limited confidence due to risk of bias, imprecision, and inconsistency. However, some USFNA results (benign, malignant) are likely very helpful, by significantly changing the pre-test probability of thyroid cancer.

✉ Victor M. Montori
  montori.victor@mayo.edu

1   Division of Endocrinology, Diabetes, Metabolism and Nutrition, Department of Internal Medicine, Mayo Clinic, Rochester, MN, USA

2   Knowledge and Evaluation Research Unit, Division of Diabetes, Metabolism and Nutrition, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA

3   Mayo Graduate School, Mayo Clinic, Rochester, MN, USA

4   Department of Emergency Medicine, Mayo Clinic, Rochester, MN, USA

5   Evidence-Based Practice Research Program, Mayo Clinic, Rochester, MN, USA

6   Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, MN, USA

7   Division of Pediatric Endocrinology, Mayo Clinic, Rochester, MN, USA

8   Mayo Medical Library, Mayo Clinic, Rochester, MN, USA

Introduction

About 70 % of adults have thyroid nodules on neck ultrasound [1, 2]. Most of these nodules are never discovered; however, the advent of imaging techniques has contributed to their identification [3, 4]. Nodule features detected through thyroid ultrasound help clinicians stratify patients according to their risk of cancer before fine needle aspiration biopsy is performed [5–8]. Given that 5–20 % of patients with thyroid nodules have thyroid malignancy,

clinical guidelines recommend thyroid ultrasonography in the evaluation of thyroid nodules [5, 9, 10]. Suspicious nodules then undergo ultrasound-guided fine needle aspiration biopsy (USFNA) and cytological evaluation. In general, clinicians will continue to monitor thyroid nodules with benign cytology, while those with malignant cytology will proceed to surgery. Nodules with indeterminate cytology may receive diagnostic surgery, molecular testing, or ongoing monitoring depending on the preferences of clinicians and patients [10].

This widely accepted diagnostic strategy for thyroid nodules has two important limitations. First, estimates of the accuracy of each USFNA finding to detect malignancy were reported mostly in single-center studies and vary greatly across studies [11, 12]. Second, once obtained, cytological findings loom large and dominate the diagnostic process, all but ignoring the likelihood of thyroid cancer gleaned from the patient's history and/or from the ultrasonographic findings [5, 6]. Improvements in the estimates of diagnostic accuracy of USFNA findings and integration with estimates of the probability of thyroid malignancy should improve the diagnostic process.

The goals of this study were to systematically appraise and summarize the available evidence about the diagnostic accuracy of USFNA findings, and to explore the integration of these estimates with the probability of thyroid malignancy before USFNA.

Methods

We performed a systematic review and meta-analysis regarding the diagnostic characteristics of USFNA for thyroid malignancy following a protocol and current guidelines on the conduct and reporting of systematic reviews of diagnostic accuracy [13].

Study selection

Type of studies

We searched for reports of randomized clinical trials, complete cohort, and case-control studies that enrolled patients of any age and gender with thyroid nodules undergoing USFNA published in English. We excluded studies involving patients with history of thyroid cancer, those reporting only on a single diagnostic category and studies of fine needle aspiration without ultrasound guidance.

Type of intervention (index test)

The test of interest was USFNA with results reported using the Bethesda system, the British classification, or a 4-category scheme (benign, malignant, nondiagnostic, and indeterminate) [14, 15]. This test involves the identification of a thyroid nodule by ultrasound and ultrasound-guided insertion of a fine needle for cell aspiration and subsequent cytological interpretation.

Type of outcomes (reference test)

Our outcome of interest was the diagnostic accuracy of USFNA of thyroid nodules for thyroid malignancy. The reference standard was histopathological diagnosis from a surgical specimen. When histology was not available, i.e., when thyroid nodules with benign cytology were followed clinically rather than surgically removed, we considered long-term follow-up without the emergence of malignant features as an alternative reference standard. We described the accuracy of the USFNA findings using the likelihood ratios (LR) for each diagnostic category and for each reporting system (Bethesda, British, four categories). LRs quantify the ability of the test result to modify the pre-test probability; a LR of 1 does not change it; 0.2 or 5 produce moderate changes; 0.1 or 10 produce large changes to the pre-test probability and are often seen with rule-in or rule-out test results [16, 17]. Formally, LR is defined as the probability of a specific finding in patients with disease over the probability of the same finding in patients without disease [16, 17]. We preferred to use these measures over the more traditional sensitivity and specificity, because they offer a more useful tool to integrate diagnostic accuracy measures into clinical decision-making [16–18].

Data sources and searches

An experienced reference librarian (P.E.), working with the study's lead investigator (N.S.O.), designed and conducted a comprehensive search strategy using controlled vocabulary and keywords for the concepts of diagnostic accuracy, USFNA, and thyroid cancer. The strategy comprised Ovid Medline In-Process and Other NonIndexed Citations, Ovid MEDLINE, Ovid EMBASE, Ovid Cochrane Central Register of Controlled Trials, Ovid Cochrane Database of Systematic Reviews, and Scopus, from each database's inception to August 2014 (Online Appendix 1). We also searched the reference list from previous systematic reviews and consulted with local thyroid disease experts seeking additional references that may have been missed by our database search strategy.

Selection of studies

Reviewers, working independently and in duplicate, reviewed all abstracts and titles for inclusion. After abstract screening and retrieval of potentially eligible studies, the

full text publications were assessed for eligibility. Disagreements were resolved by consensus.

Data extraction

Using a standardized web-based form, reviewers collected the following information, independently and in duplicate from each study: country where study was conducted, number of patients and of thyroid nodules, average age, gender, and thyroid nodule size. Additionally, we extracted technical aspects of the index test such as operator experience (years performing USFNA), method of aspiration (negative pressure vs. capillary), number of USFNA passes as well as any measurement of inter-observer variability during the cytological examination of the specimens. Finally, we extracted true positive, true negative, false negative, and false positive values to construct a diagnostic contingency (or 2-by-2) table for each cytological outcome within each of the reporting systems evaluated.

Quality assessment

Reviewers, working independently and in duplicate, analyzed the eligible articles to assess their methods to protect them from bias using the Quality Assessment of Studies of Diagnostic Accuracy included in systematic reviews (QUADAS-2) tool [19, 20]. This tool assesses the risk of bias and applicability in terms of patient selection, index test, reference standard, flow and timing [19, 20].

Author contact

We attempted to contact 47 authors of included studies by email to verify and complete data that could not be discerned from the report; only 12 responded. We excluded studies for which required information to judge eligibility or to conduct key analyses remained unavailable after thorough review of the published report and author contact.

Subgroup and sensitivity analysis

To explain possible inconsistencies across study results we planned on conducting the following subgroup analyses for diagnostic accuracy of USFNA: method of USFNA, operator experience, needle gauge, number of aspirations, cytological examination, probe frequency, reporting system for the index test, reference standard use, risk of bias and age of population.

Data synthesis and statistical analysis

Meta-analysis

We used the random-effects model to pool likelihood ratios for the diagnosis of thyroid malignancy and their respective

95 % confidence intervals. This analysis was carried out in Meta-Disc [21] and Review Manager (Revman) [22]. This model produces variance estimates that accounts for within-study variance (precision) and between-study differences in patients, methods, test performance, and reference standard performance (inconsistency). We used the I^2 statistic to assess for inconsistency across individual studies, with $I^2 > 50$ % indicating large inconsistency [23]. As a sensitivity analysis, we conducted a bivariate meta-analysis, an approach that takes into account potential threshold effects and the correlation between sensitivity and specificity [24], as implemented in Statistical Analysis System (SAS) software.

Results

Study identification

Our literature search identified 1085 abstracts, of which 32 studies were eligible (Fig. 1) [25–56]. These studies were mostly conducted in adults, included more than 79,541 thyroid nodules (number of nodules not reported in one study) with 15,641 undergoing surgical intervention, of which 27 % harbored malignancy (Table 1). Due to inconsistent reporting of long-term follow-up, the diagnostic accuracy analysis was based only on histopathology as a reference standard. Of the 32 included studies, 22
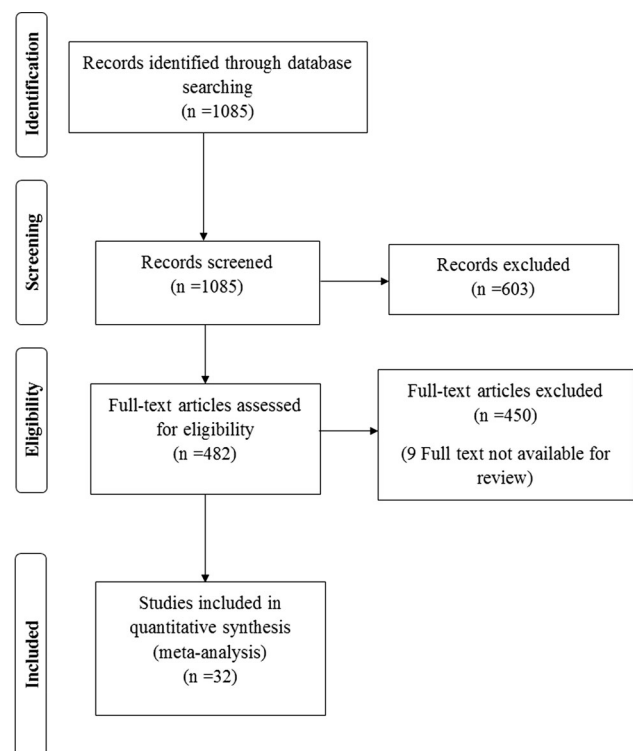
Fig. 1 Study selection

Table 1 Description of the included studies

Study	Year	Country	Special Population	Population	Total number of nodules reported in the study	Total number of nodules with surgery	% Nodules with surgery	% Nodules with malignancy on GS	%Patients with PTC	Reporting system
Bohacek	2012	USA	Surgeon performed biopsies	Adults	1000	450	45.1	28.6	85.3	Bethesda
Bongiovanni	2012	Multicenter	None	Adults	3724	1358	36.5	40.7	NR	Bethesda
Braga	2001	Brazil	Complex nodules	Adults	124	14	11.3	35.7	80	4 categories
Can	2008	Turkey	None	Adults	184	23	12.5	34.8	87.5	4 categories
Cochand-Priollet	1994	France	Cold nodules	Adults	132	132	100	15.1	50	4 categories
Danese	1998	Italy	None	Adults and children	4697	540	11.5	19.1	NR	4 categories
Das	2009	Kuwait	None	Adults and children	759	44	5.8	18.2	NR	4 categories
Giovagnoli	1998	Italy	None	Adults and children	138	138	100	19.6	77.7	4 categories
Hen Lee	2009	Korea	None	Adults	2912	528	18.1	68.2	91.4	4 categories
Kawai	2012	Japan	Accessory nodules	Unknown	365	365	100	44.4	NR	4 categories
Kim SJ	2003	Korea	Nonpalpable thyroid nodules	Adults	149	65	43.6	67.8	NR	4 categories
Kim SK	2011	Korea	None	Adults	865	204	23.5	89.6	97.8	Bethesda
Maia	2014	Brazil	None	Adults	643	242	100	45.9	87.4	Bethesda
Mikosch	2001	Austria	Only single nodules	Unknown	Not available	538	100	11.5	85.4	4 categories
Mirshemirani	2009	Iran	None	Children	32	32	100	25	87.5	4 categories
Nacamulli	2012	Italy	none	Adults	164	164	100	54.9	92.2	British
Ochoa	2011	Spain	None	Adults	475	92	19.4	25	80	4 categories
Pisani	2000	Italy	None	Adults	136	42	30.9	31	61.5	4 categories
Rago	2010	Italy	None	Adults and children	47,775	3406	7.1	37.4	NR	British
Rorive	2010	Belgium	None	Adults and children	5283	924	17.5	19.1	61	4 categories
Rosen	1993	Canada	None	Unknown	59	41	69.5	39	NR	4 categories
Sarkis	2014	Australia	None	Unknown	2076	425	24.5	10.1	83.1	Bethesda
Schoedel	2008	USA	None	Adults	180	46	25.6	45.7	95.2	Bethesda
Schueller-Weidekamm	2010	Austria	Cold nodules	Adults	21	21	100	61.9	NR	4 categories
Smadi	2008	Jordan	Clinical palpable solitary thyroid nodules	Adults	100	88	88	22.7	60	4 categories

Table 1 continued

Study	Year	Country	Special Population	Population	Total number of nodules reported in the study	Total number of nodules with surgery	% Nodules with surgery	% Nodules with malignancy on GS	%Patients with PTC	Reporting system
Solymosi	2013	Hungary	An iodine-deficient region	Adults and children	4601	4601	100	10.6	NR	4 categories
Tabain	2004	Croatia	None	Adults and children	457	457	100	20.3	NR	4 categories
Tunca	2007	Turkey	Patients with indication for surgery	Adults	31	31	100	25.8	87.5	4 categories
Unluturk	2012	Turkey	None	Adults	237	72	30.4	79.2	84.5	Bethesda
Wharry	2014	USA	Nodules >4 cm	Adults	290	290	100	22.4	NR	Bethesda
Yoon	2011	Korea	Nodules with microcalcification	Adults	26	17	65.4	82.4	100	4 categories
Young	2011	Philippines	Cystic nodules excluded	Adults	1906	251	13.2	27.5	86.9	4 categories

PTC papillary thyroid cancer, *GS* gold standard

reported the USFNA outcomes in four categories [27–34, 36, 37, 39, 41, 42, 45–48, 51–53, 55, 56], eight used the Bethesda system [25, 26, 35, 43, 44, 49, 50, 54], and 2 the British classification [34, 36].

Risk of bias

Overall, the risk of bias was moderate to high due to the study of nonconsecutive samples, lack of blinding when retrospectively assessing the results of the index and reference tests, having few patients undergo evaluation with the histopathological reference standard, and having inconsistent follow-up and reporting of those followed clinically. These features increase the risk of overestimation of the diagnostic accuracy of the index test. A summary of the risk of bias evaluation of the included studies is depicted in Fig. 2, representing the proportion of studies deemed at high, moderate, or low risk of bias by the reviewers in each of the seven QUADAS-2 domains.

Diagnostic accuracy by USFNA outcome and reporting system

The diagnostic accuracy of USFNA was analyzed according to USFNA outcome (all reporting systems included), the Bethesda reporting system, or a 4 category reporting system (Table 2). Overall, results in the categories of benign, malignant, and suspicious for malignancy were associated with LR that significantly changed the pretest probability. A benign USFNA result was associated with a LR of 0.09, 95 % CI 0.06–0.14, $I^2 = 33$ %, whereas a malignancy finding on USFNA was associated with a LR of 197, 95 % CI 68–569, $I^2 = 77$ %. These results suggest these USFNA results (benign and malignant) all but rule-out and rule-in malignancy, respectively. Consistent results were found across reporting systems (Table 2). A distribution of histology results according to USFNA category and reporting system is found on Table 3.

Due to inconsistent reporting, we were unable to perform any of the planned subgroup analyses. As a sensitivity analysis, we evaluated only studies deemed to be at lower risk of bias (5 out of 7 QUADAS-2 criteria considered low risk). These estimates were compatible with the main analysis. (Supplemental Table 1).

Discussion

Summary of the evidence

The body of evidence about the diagnostic accuracy of USFNA is at moderate risk of bias which may overestimate diagnostic accuracy estimates. These estimates are
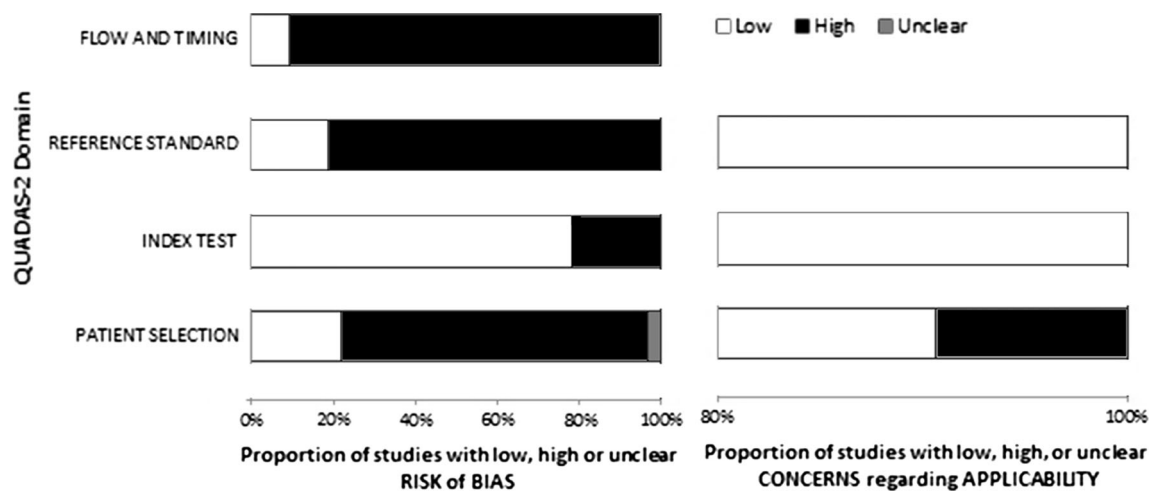
Fig. 2 Risk of Bias using the QUADAS-2 instrument

imprecise (wide confidence intervals) despite pooling and come from studies that report inconsistent estimates, inconsistency that cannot be fully explained by the use of different reporting systems. With these caveats, USFNA results that are reported as benign, suspicious for malignancy, or malignant have a large impact on the probability of thyroid malignancy after the test; other results have a smaller effect making management decisions dependent on other clinical variables (nondiagnostic and indeterminate results). These findings correlate with a study that evaluated the concordance between central and local histopathologists when interpreting USFNA results. As in our review, the benign and malignant categories were associated with the highest concordance between pathologists; with central pathologist reporting a fewer number of indeterminate diagnosis (found to be less helpful clinically) compared to the local pathologists, highlighting the importance of experience in the interpretation of USFNA [57]. In addition, even when the USFNA biopsy has been carefully performed and interpreted, the test has an intrinsic limitation when trying to differentiate follicular adenomas from carcinomas [14].

Limitations

Our review is limited by the possibility that our search strategy may have missed studies, our language restriction, and that studies had to be excluded due to incomplete reporting of data. We have tried to mitigate these limitations by engaging an expert librarian in designing the search strategy [58] and by contacting study authors although this was accomplished with only partial success.

Additionally, pooling methods such as the random effect model have been questioned in diagnostic systematic reviews. To overcome this potential limitation, we

conducted the analysis using a bivariate model. This method accounts for the aforementioned variation and analyzes sensitivity and specificity jointly [24]. Furthermore, we were not able to conduct subgroup analysis of accuracy of USFNA based on important clinical factors and help explain inconsistency in results across studies. Due to the moderate risk of bias of the included studies (given than only a minority of patients undergo histological evaluation), our results can overestimate the diagnostic accuracy of USFNA. Also, our findings do not apply to the less common thyroid malignancies poorly represented in the included studies.

Applications for practice

To better understand these results, it is best to apply these findings to clinical cases. Consider a nodule labeled as highly suspicious for thyroid cancer based on US features, that is, for every 10 such nodules we would expect to find malignancy in eight of them [10]. If the USFNA result is benign that number drops to about 3 in 10. One would hardly stop evaluation and treatment at such a high likelihood of malignancy. But one would stop for a nodule characterized on US features as having a low suspicion for malignancy, that is, for every ten such nodules at most we would expect to find malignancy in one of them [10]. In that case, a benign USFNA result all but excludes (<1 %) the possibility of malignancy. So the same test result, a benign USFNA has different implications depending on the probability of malignancy—giving the characteristics of the patient and of the nodule on US—before the USFNA is performed.

Another example would be if a highly suspicious nodule is found to have a biopsy suspicious for follicular neoplasm, where the risk of thyroid cancer will be 71 %. In this case, patients and clinicians might feel comfortable with a recommendation for diagnostic thyroidectomy. In contrast, a nodule with intermediate suspicion for

Table 2 Summary of the meta-analysis of diagnostic performance of USFNA for the diagnosis of thyroid malignancy

Category	Sensitivity	Specificity	LR	I^2 (%)
A) Diagnostic performance of USFNA based on diagnostic outcome				
Non diagnostic	0.029	0.942	0.515	33
	(0.020–0.0452)	(0.918–0.959)	(0.336–0.791)	
Benign	0.053	0.436	0.093	83
	(0.035–0.079)	(0.363–0.512)	(0.063–0.139)	
Suspicious for follicular neoplasm	0.166	0.727	0.608	84
	(0.097–0.269)	(0.605–0.823)	(0.365–1.01)	
Suspicious for malignancy	0.166	0.980	8.35	89
	(0.112–0.239)	(0.956–0.991)	(3.63–19.23)	
Malignant	0.559	0.998	196.77	77
	(0.477–0.638)	(0.991–0.999)	(67.99–569.51)	
B) Diagnostic performance of USFNA using a 4 category reporting system				
Non diagnostic	0.031	0.924	0.403	16
	(0.016–0.057)	(0.887–0.949)	(0.213–0.763)	
Benign	0.056	0.375	0.089	64
	(0.354–0.087)	(0.294–0.464)	(0.057–0.139)	
Suspicious/indeterminate	0.224	0.758	0.923	89
	(0.161–0.302)	(0.678–0.823)	(0.624–1.37)	
Malignant	0.642	0.993	98.142	81
	(0.557–0.719)	(0.980–0.998)	(33.674–286.28)	
C) Diagnostic performance of USFNA using the Bethesda reporting system				
Non diagnostic	0.024	0.952	0.504	44
	(0.012–0.048)	(0.910–0.974)	(0.216–1.17)	
Benign	0.059	0.583	0.141	82
	(0.022–0.147)	(0.468–0.690)	(0.064–0.312)	
AUS	0.068	0.830	0.405	73
	(0.033–0.134)	(0.7–0.911)	(0.228–0.719)	
Suspicious for follicular neoplasm	0.159	0.770	0.694	82
	(0.081–0.290)	(0.646–0.860)	(0.398–1.21)	
Suspicious for malignancy	0.136	0.973	5.07	43
	(0.088–0.204)	(0.943–0.987)	(2.59–9.89)	
Malignancy	0.394	0.998	239.83	0
	(0.262–0.546)	(0.994–0.999)	(73.98–777.47)	

Estimate, 95 % confidence interval

To calculate the diagnostic performance of each USFNA category of the included studies, we constructed a two by two table, that included the test results as positive (cases with the diagnostic category of interest) and the test result as negative (cases with any result except the diagnostic category of interest). For the reference standard, a positive result were those found to have cancer on histology and a negative result all cases without cancer on histology

The sensitivity was then calculated as TP/(TP + FN); specificity as TN/(TN + FP)

USFNA ultrasound-guided fine needle aspiration biopsy, *AUS* atypia of undetermined significance, *LR* likelihood ratio

malignancy based on US features and the same USFNA result (suspicious for follicular neoplasm) will have a 10 % risk of malignancy. In this case, further diagnostic testing (e.g., molecular testing) might be helpful before diagnostic thyroidectomy is recommended.

Clinicians make intuitive and often implicit judgments as they receive reports from US and USFNA about the likelihood of malignancy. These may be open to error that an explicit approach may reduce. Similarly, an explicit approach may also be helpful in discussing cases with trainees who have yet to form the kinds of intuitions that guide judgments in more senior colleagues. Several approaches are available to bring this so-called Bayesian reasoning, to the clinic [18]. To facilitate a more explicit

Table 3 Distribution of histology results by USFNA category and reporting system

	Benign	Malignant	Total	% Malignant
USFNA diagnostic outcome				
Nondiagnostic	814	175	989	17.7
Benign	6874	192	7066	2.7
Suspicious for follicular neoplasm	1479	622	2101	29.6
Suspicious for malignancy	89	644	733	87.9
Malignancy	61	2127	2188	97.2
			13,077	
USFNA 4 category reporting system				
Nondiagnostic	557	58	615	9.4
Benign	4899	106	5005	2.1
Suspicious/indeterminate	1749	433	2182	19.8
Malignant	58	1124	1182	95.1
			8984	
USFNA Bethesda reporting system				
Nondiagnostic	94	28	122	23.0
Benign	679	59	738	8.0
AUS	293	86	379	22.7
Suspicious for follicular neoplasm	748	326	1074	30.4
Suspicious for malignancy	75	236	311	75.9
Malignancy	3	460	463	99.4
			3087	

use of the estimates reported here, we have developed a calculator to determine the risk of malignancy for thyroid nodules after a USFNA (freely available at http://www.thyroidcarisk.mayo.edu/). In this tool, the user enters a best guess of the probability of thyroid malignancy based on clinical judgment (informed by prevalence, clinical, and ultrasound features and their experience) or, alternatively, uses the ATA ultrasonography risk stratification system for thyroid nodules [10]. Then the user selects an USFNA finding (e.g., malignant, benign). The resulting post-test probability is represented in a 100-person pictogram [59]. This tool brings visual clarity to the process, allows for what-if scenarios, and offers the opportunity to involve patients into the decision-making process [60].

Applications for research

Our study identifies large knowledge gaps in the diagnostic process of a common clinical problem, the management of thyroid nodules. To the subjective estimates of a probability of malignancy from clinical and ultrasound features, we add imprecise and perhaps biased estimates of the LR for USFNA results based on the best available evidence. At the extremes of these ranges, the resulting probability of malignancy after USFNA may suggest different 'next steps' emphasizing the need for additional research to improve the validity and precision of our diagnostic

estimates. The Institute of Medicine has recently called for a greater investment in diagnostic research to improve its contribution to the value and safety of clinical care [61].

Opportunities for research are rich. For example, tools that integrate clinical and ultrasound features to determine a more precise pre-test probability are required. Moreover, given that an important limitation of the included studies is that these estimates are based only on the patients that underwent surgery, [62] diagnostic studies with ideally low risk of bias that provide consistent and complete follow-up for patients should be conducted. Finally, this framework provides clinicians with the opportunity to involve patients into the decision-making process, especially when deciding the next step of management based on action thresholds taking into consideration their values and preferences. However, formal studies assessing the impact of the use of this tool on patient important outcomes and on engaging them in shared decision-making would be needed.

Conclusion

The available evidence regarding the diagnostic accuracy of USFNA warrants only limited confidence due to risk of bias, imprecision, and inconsistency. However, some USFNA results (benign, malignant) are likely very helpful, by significantly changing the pre-test probability of thyroid cancer.

References

1. S. Ezzat, D.A. Sarti, D.R. Cain, G.D. Braunstein, Thyroid incidentalomas. Prevalence by palpation and ultrasonography. Arch. Intern. Med. **154**(16), 1838–1840 (1994)

2. J.J. Cronan, Thyroid nodules: is it time to turn off the US machines? Radiology **247**(3), 602–604 (2008). doi:10.1148/radiol.2473072233

3. J.P. Brito, J.C. Morris, V.M. Montori, Thyroid cancer: zealous imaging has increased detection and treatment of low risk tumours. BMJ **347**, f4706 (2013). doi:10.1136/bmj.f4706

4. J.A. Sosa, J.W. Hanna, K.A. Robinson, R.B. Lanman, Increases in thyroid nodule fine-needle aspirations, operations, and diagnoses of thyroid cancer in the United States. Surgery **154**(6), 1420–1426 (2014). doi:10.1016/j.surg.2013.07.006

5. J.P. Brito, I.D. Hay, J.C. Morris, Low risk papillary thyroid cancer. Bmj **348**, g3045 (2014). doi:10.1136/bmj.g3045

6. P. Campanella, F. Ianni, C.A. Rota, S.M. Corsello, A. Pontecorvi, Quantification of cancer risk of each clinical and ultrasonographic suspicious feature of thyroid nodules: a systematic review and meta-analysis. Eur. J. Endocrinol. **170**(5), R203–211 (2014). doi:10.1530/EJE-13-0995

7. R. Smith-Bindman, P. Lebda, V.A. Feldstein, D. Sellami, R.B. Goldstein, N. Brasic, C. Jin, J. Kornak, Risk of thyroid cancer based on thyroid ultrasound imaging characteristics: results of a population-based study. JAMA Intern. Med. **173**(19), 1788–1796 (2013). doi:10.1001/jamainternmed.2013.9245

8. A. Al Nofal, M.R. Gionfriddo, A. Javed, Q. Haydour, J.P. Prokop, S.T. Pittock, M.H. Murad, Accuracy of thyroid nodule sonography for the detection of thyroid cancer in children: systematic review and meta-analysis. Clin. Endocrinol. (Oxf) (2015). doi:10.1111/cen.12786

9. J.P. Brito, A.J. Yarur, L.J. Prokop, B. McIver, M.H. Murad, V.M. Montori, Prevalence of thyroid cancer in multinodular goiter versus single nodule: a systematic review and meta-analysis. Thyroid **23**(4), 449–455 (2013). doi:10.1089/thy.2012.0156

10. B.R.M. Haugen, E.K. Alexander, K.C. Bible, G. Doherty, S.J. Mandel, Y.E. Nikiforov, F. Pacini, G. Randolph, A. Sawka, M. Schlumberger, K.G. Schuff, S.I. Sherman, J.A. Sosa, D. Steward, R.M.M. Tuttle, L. Wartofsky, 2015 American Thyroid Association Management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer. Thyroid (2015). doi:10.1089/thy.2015.0020

11. M. Bongiovanni, A. Spitale, W.C. Faquin, L. Mazzucchelli, Z.W. Baloch, The Bethesda system for reporting thyroid cytopathology: a meta-analysis. Acta. Cytol. **56**(4), 333–339 (2012). doi:10.1159/000339959

12. C.C. Wang, L. Friedman, G.C. Kennedy, H. Wang, E. Kebebew, D.L. Steward, M.A. Zeiger, W.H. Westra, Y. Wang, E. Khanafshar, G. Fellegara, J. Rosai, V. Livolsi, R.B. Lanman, A large multicenter correlation study of thyroid nodule cytopathology and histopathology. Thyroid **21**(3), 243–251 (2011). doi:10.1089/thy.2010.0243

13. D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J. Clin. Epidemiol. **62**(10), 1006–1012 (2009). doi:10.1016/j.jclinepi.2009.06.005

14. E.S. Cibas, S.Z. Ali, The Bethesda system for reporting thyroid cytopathology. Thyroid **19**(11), 1159–1165 (2009). doi:10.1089/thy.2009.0274

15. P. Perros, K. Boelaert, S. Colley, C. Evans, R.M. Evans, G. Gerrard Ba, J. Gilbert, B. Harrison, S.J. Johnson, T.E. Giles, L. Moss, V. Lewington, K. Newbold, J. Taylor, R.V. Thakker, J. Watkinson, G.R. Williams, Guidelines for the management of thyroid cancer. Clin. Endocrinol. (Oxf) **81**(Suppl 1), 1–122 (2014). doi:10.1111/cen.12515

16. S. McGee, Simplifying likelihood ratios. J. Gen. Intern. Med. **17**(8), 646–649 (2002)

17. T.S. Furukawa, H. Bucher, T. Agoritsas, G. Guyatt, in Diagnostic Tests, ed. by G.D. Guyatt, M. Meade, D. Cook. User's Guides to the Medical Literature: A munal for Evidence Based Clinical Practive, vol. 3 (McGraw hill, 2014), p. 345–357

18. M.B. Elamin, M.H. Murad, R. Mullan, D. Erickson, K. Harris, S. Nadeem, R. Ennis, P.J. Erwin, V.M. Montori, Accuracy of diagnostic tests for Cushing's syndrome: a systematic review and metaanalyses. J. Clin. Endocrinol. Metab. **93**(5), 1553–1562 (2008). doi:10.1210/jc.2008-0139

19. S. Schueler, G.M. Schuetz, M. Dewey, The revised QUADAS-2 tool. Ann. Intern. Med. **156**(4), 323–324 (2012). doi:10.7326/0003-4819-156-4-201202210-00018

20. P.F. Whiting, A.W. Rutjes, M.E. Westwood, S. Mallett, J.J. Deeks, J.B. Reitsma, M.M. Leeflang, J.A. Sterne, P.M. Bossuyt, QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann. Intern. Med. **155**(8), 529–536 (2011). doi:10.7326/0003-4819-155-8-201110180-00009

21. J. Zamora, V. Abraira, A. Muriel, K. Khan, A. Coomarasamy, Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med. Res. Methodol. **6**, 31 (2006). doi:10.1186/1471-2288-6-31

22. Review Manager (RevMan) [Computer program]. in Copenhagen: the Nordic Cochrane Centre, The Cochrane Collaboration, (2014)

23. J.P. Higgins, S.G. Thompson, J.J. Deeks, D.G. Altman, Measuring inconsistency in meta-analyses. Bmj **327**(7414), 557–560 (2003). doi:10.1136/bmj.327.7414.557

24. A.H. Zwinderman, P.M. Bossuyt, We should not pool diagnostic likelihood ratios in systematic reviews. Stat. Med. **27**(5), 687–697 (2008). doi:10.1002/sim.2992

25. L. Bohacek, M. Milas, J. Mitchell, A. Siperstein, E. Berber, Diagnostic accuracy of surgeon-performed ultrasound-guided fine-needle aspiration of thyroid nodules. Ann. Surg. Oncol. **19**(1), 45–51 (2012). doi:10.1245/s10434-011-1807-z

26. M. Bongiovanni, S. Crippa, Z. Baloch, S. Piana, A. Spitale, F. Pagni, L. Mazzucchelli, C. Di Bella, W. Faquin, Comparison of

5-tiered and 6-tiered diagnostic systems for the reporting of thyroid cytopathology: a multi-institutional study. Cancer Cytopathol. **120**(2), 117–125 (2012). doi:10.1002/cncy.20195

27. M. Braga, T.C. Cavalcanti, L.M. Collaco, H. Graf, Efficacy of ultrasound-guided fine-needle aspiration biopsy in the diagnosis of complex thyroid nodules. J. Clin. Endocrinol. Metab. **86**(9), 4089–4091 (2001). doi:10.1210/jcem.86.9.7824

28. A.S. Can, K. Peker, Comparison of palpation-versus ultrasound-guided fine-needle aspiration biopsies in the evaluation of thyroid nodules. BMC Res. Notes. **1**, 12 (2008). doi:10.1186/1756-0500-1-12

29. B. Cochand-Priollet, P.J. Guillausseau, S. Chagnon, C. Hoang, C. Guillausseau-Scholer, P. Chanson, H. Dahan, A. Warnet, P.T. Tran Ba Huy, P. Valleur, The diagnostic value of fine-needle aspiration biopsy under ultrasonography in nonfunctional thyroid nodules: a prospective study comparing cytologic and histologic findings. Am. J. Med. **97**(2), 152–157 (1994)

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

31. D.K. Das, C.S. Pant, K.L. Chachra, A.K. Mandal, R. Kumar, S.K. Ganguly, S. Bhambhani, R.P. Tripathy, Solitary nodular goiter: role of ultrasonography in fine-needle aspiration cytology diagnosis. Med. Princ. Pract. **9**, 205–213 (2009)

32. M.R. Giovagnoli, T. Pisani, A. Drusco, L. Scardella, A. Antonaci, A. Vecchione, Fine needle aspiration biopsy in the preoperative management of patients with thyroid nodules. Anticancer Res. **18**(5B), 3741–3745 (1998)

33. T. Kawai, E. Nishihara, T. Kudo, H. Ota, S. Morita, K. Kobayashi, M. Ito, S. Kubota, N. Amino, A. Miyauchi, Histopathological diagnoses of "accessory" thyroid nodules diagnosed as benign by fine-needle aspiration cytology and ultrasonography. Thyroid **22**(3), 299–303 (2012). doi:10.1089/thy.2011.0272

34. S.J. Kim, E.K. Kim, C.S. Park, W.Y. Chung, K.K. Oh, H.S. Yoo, Ultrasound-guided fine-needle aspiration biopsy in nonpalpable thyroid nodules: is it useful in infracentimetric nodules? Yonsei Med. J. **44**(4), 635–640 (2003)

35. S.K. Kim, T.S. Hwang, Y.B. Yoo, H.S. Han, D.L. Kim, K.H. Song, S.D. Lim, W.S. Kim, N.S. Paik, Surgical results of thyroid nodules according to a management guideline based on the BRAF(V600E) mutation status. J. Clin. Endocrinol. Metab. **96**(3), 658–664 (2011). doi:10.1210/jc.2010-1082

36. P. Mikosch, U. Wartner, E. Kresnik, H.J. Gallowitsch, M. Heinisch, H.P. Dinges, P. Lind, Results of preoperative ultrasound guided fine needle aspiration biopsy of solitary thyroid nodules as compared with the histology. A retrospective analysis of 538 patients. Nuklearmedizin **40**(5), 148–154 (2001)

37. A. Mirshemirani, F. Roshanzamir, A.K. Tabari, J. Ghorobi, S. Salehpoor, F.A. Gorji, Thyroid nodules in childhood: a single institute experience. Iran J. Pediatr. **20**(1), 91–96 (2010)

38. D. Nacamulli, L. Nico, S. Barollo, L. Zambonin, G. Pennelli, M.E. Girelli, E. Casal Ide, M.R. Pelizzo, F. Vianello, I. Negro, S. Watutantrige-Fernando, F. Mantero, M. Rugge, C. Mian, Comparison of the diagnostic accuracy of combined elastosonography and BRAF analysis vs cytology and ultrasonography for thyroid nodule suspected of malignancy. Clin. Endocrinol. (Oxf) **77**(4), 608–614 (2012). doi:10.1111/j.1365-2265.2012.04427.x

39. T. Pisani, M. Bononi, C. Nagar, M. Angelini, M. Bezzi, A. Vecchione, Fine needle aspiration and core needle biopsy techniques in the diagnosis of nodular thyroid pathologies. Anticancer Res. **20**(5C), 3843–3847 (2000)

40. T. Rago, E. Fiore, M. Scutari, F. Santini, G. Di Coscio, R. Romani, P. Piaggi, C. Ugolini, F. Basolo, P. Miccoli, A. Pinchera, P. Vitti, Male sex, single nodularity, and young age are associated with the risk of finding a papillary thyroid cancer on fine-needle aspiration cytology in a large series of patients with nodular thyroid disease. Eur. J. Endocrinol. **162**(4), 763–770 (2010). doi:10.1530/EJE-09-0895

41. S. Rorive, N. D'Haene, C. Fossion, I. Delpierre, N. Abarguia, F. Avni, C. Decaestecker, I. Salmon, Ultrasound-guided fine-needle aspiration of thyroid nodules: stratification of malignancy risk using follicular proliferation grading, clinical and ultrasonographic features. Eur. J. Endocrinol. **162**(6), 1107–1115 (2010). doi:10.1530/EJE-09-1103

42. I.B. Rosen, A. Azadian, P.G. Walfish, S. Salem, E. Lansdown, Y.C. Bedard, Ultrasound-guided fine-needle aspiration biopsy in the management of thyroid disease. Am. J. Surg. **166**(4), 346–349 (1993)

43. L.M. Sarkis, O. Norlen, A. Aniss, N. Watson, L.W. Delbridge, S.B. Sidhu, M.S. Sywak, A.J. Gill, The Australian experience with the Bethesda classification system for thyroid fine needle aspiration biopsies. Pathology **46**(7), 592–595 (2014). doi:10.1097/PAT.0000000000000168

44. K.E. Schoedel, M.E. Tublin, K. Pealer, N.P. Ohori, Ultrasound-guided biopsy of the thyroid: a comparison of technique with respect to diagnostic accuracy. Diagn. Cytopathol. **36**(11), 787–789 (2008). doi:10.1002/dc.20896

45. C. Schueller-Weidekamm, G. Schueller, K. Kaserer, C. Scheuba, H. Ringl, M. Weber, C. Czerny, A.M. Herneth, Diagnostic value of sonography, ultrasound-guided fine-needle aspiration cytology, and diffusion-weighted MRI in the characterization of cold thyroid nodules. Eur. J. Radiol. **73**(3), 538–544 (2010). doi:10.1016/j.ejrad.2008.12.013

46. N. Sebastian-Ochoa, J.C. Fernandez-Garcia, I. Mancha Doblas, A. Sebastian-Ochoa, D. Fernandez Garcia, M.V. Ortega Jimenez, E. Gallego Dominguez, F. Tinahones Madueno, [Clinical experience in a high-resolution thyroid nodule clinic]. Endocrinol. Nutr. **58**(8), 409–415 (2011). doi:10.1016/j.endonu.2011.06.007

47. T. Solymosi, G. Lukacs Toth, D. Nagy, I. Gal, Twenty years of experience with the preoperative diagnosis of medullary cancer in a moderately iodine-deficient region. Int. J. Endocrinol. **2013**, 571606 (2013). doi:10.1155/2013/571606

48. F. Tunca, Y. Giles, A. Salmaslioglu, A. Poyanli, D. Yilmazbayhan, T. Terzioglu, S. Tezelman, The preoperative exclusion of thyroid carcinoma in multinodular goiter: Dynamic contrast-enhanced magnetic resonance imaging versus ultrasonography-guided fine-needle aspiration biopsy. Surgery **142**(6), 992–1002 (2007). doi:10.1016/j.surg.2007.09.027

49. U. Unluturk, M.F. Erdogan, O. Demir, S. Gullu, N. Baskal, Ultrasound elastography is not superior to grayscale ultrasound in predicting malignancy in thyroid nodules. Thyroid **22**(10), 1031–1038 (2012). doi:10.1089/thy.2011.0502

50. L.I. Wharry, K.L. McCoy, M.T. Stang, M.J. Armstrong, S.O. LeBeau, M.E. Tublin, B. Sholosh, A. Silbermann, N.P. Ohori, Y.E. Nikiforov, S.P. Hodak, S.E. Carty, L. Yip, Thyroid nodules > 4 cm: can ultrasound and cytology reliably exclude cancer? World J. Surg. **38**(3), 614–621 (2014)

51. J.H. Yoon, E.K. Kim, E.J. Son, H.J. Moon, J.Y. Kwak, Diffuse microcalcifications only of the thyroid gland seen on ultrasound: clinical implication and diagnostic approach. Ann. Surg. Oncol. **18**(10), 2899–2906 (2011). doi:10.1245/s10434-011-1717-0

52. J.K. Young, C.G. Lumapas, R. Mirasol, Sonographically guided fine-needle aspiration biopsy of thyroid nodules: Correlation between cytologic and histopathologic findings. Phillippine J. Intern. Med. **49**(1), 8–14 (2011)

53. Y.H. Lee, B.H. Kim, S.I. Suh, H.S. Seo, B.K. Seo, K.R. Cho, J.H. Lee, N.H. Kim, J.A. Seo, Comparison of cytological results obtained by repeated US-guided fine-needle aspiration biopsies of thyroid nodules: focus on the rate of malignancy and diagnostic concordance. Diagn. Cytopathol. **37**(7), 492–497 (2009). doi:10.1002/dc.21043

54. F.F. Maia, P.S. Matos, E.J. Pavin, D.E. Zantut-Wittmann, Thyroid imaging reporting and data system score combined with Bethesda system for malignancy risk stratification in thyroid nodules with indeterminate results on cytology. Clin. Endocrinol. (Oxf) **82**(3), 439–444 (2015). doi:10.1111/cen.12525

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

56. A.A. Smadi, K. Ajarmeh, F. Wreikat, Fine-needle aspiration of thyroid nodules has high sensitivity and specificity. Rawal. Med. J. **33**(2), 221–224 (2008)

57. E.S. Cibas, Z.W. Baloch, G. Fellegara, V.A. LiVolsi, S.S. Raab, J. Rosai, J. Diggans, L. Friedman, G.C. Kennedy, R.T. Kloos, R.B. Lanman, S.J. Mandel, N. Sindy, D.L. Steward, M.A. Zeiger, B.R. Haugen, E.K. Alexander, A prospective assessment defining the limitations of thyroid nodule pathologic evaluation. Ann. Intern. Med. **159**(5), 325–332 (2013). doi:10.7326/0003-4819-159-5-201309030-00006

58. M. Rethlefsen, A. Farrell, L. Osterhaus Trzasko, T. Brigham, Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. J. Clin. Epidemiol. **68**(6), 617–626 (2015)

59. M. Price, R. Cameron, P. Butow, Communicating risk information: the influence of graphical display format on quantitative information perception-Accuracy, comprehension and preferences. Patient Educ. Couns. **69**(1–3), 121–128 (2007). doi:10.1016/j.pec.2007.08.006

60. D.R. Gordon Guyatt, M.O. Meade, D.J. Cook (ed.,) Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 3rd edn. (2015)

61. National Academies of Sciences, in E.A.M.: Improving diagnosis in health care, (The national academies press, Washington, DC, 2015)

62. Y.Y. Tee, A.J. Lowe, C.A. Brand, R.T. Judson, Fine-needle aspiration may miss a third of all malignancy in palpable thyroid nodules: a comprehensive literature review. Ann. Surg. **246**(5), 714–720 (2007). doi:10.1097/SLA.0b013e3180f61adc