

73452\_Auto\_Edited.docx

**Name of Journal:** *World Journal of Methodology*

**Manuscript NO:** 73452

**Manuscript Type:** META-ANALYSIS

**Assessment of diagnostic capacity and decision making based on the American Thyroid Association 2015 ultrasound system**

Luis-Mauricio Hurtado-Lopez, Alfredo Carrillo-Muñoz, Felipe-Rafael Zaldivar-Ramirez, Erich Otto Paul Basurto-Kuba, Blanca-Estela Monroy-Lozano

**Abstract**

**BACKGROUND**

This study evaluates the American Thyroid Association's ultrasound system for the initial assessment of thyroid nodules to determine if it indeed facilitates clinical decision making.

**AIM**

To perform a systematic review and meta-analysis of the diagnostic value of the American Thyroid Association's ultrasound system for the initial assessment of thyroid nodules.

**METHODS**

In accordance with the PRISMA statement for Diagnostic Test Accuracy, we selected articles that evaluated the 2015 ATA ultrasound pattern guidelines using a diagnostic gold standard. We analyzed these cases using traditional diagnostic values, as well as the threshold approach to clinical decision-making and decision curve analysis.

**RESULTS**

We reviewed 13 articles with 8,445 thyroid nodules, which were classified according to 2015 ATA patterns. Of these, 46.62% were malignant. No cancer was found in any of the ATA benign pattern nodules. The Bayesian analysis post-test probability for cancer in each classification was: very-low suspicion, 0.85%; low, 2.6%; intermediate, 6.7%; and high, 40.9%. The net benefit (NB), expressed as avoided interventions, indicated that the highest capacity to avoid unnecessary FNABs in the patterns that we studied was 42, 31, 35 and 43 of every 100 FNAB. The NB calculation for a Pt of 11% for each of the ATA suspicion patterns studied is less than that of performing FBAB on all nodules.

## CONCLUSION

These three types of analysis have shown that only the ATA high-suspicion diagnostic pattern is clinically useful, in which case, FNAB should be performed. However, the curve decision analysis has demonstrated that using the ATA ultrasound risk patterns to decide which patients need FNAB does not provide a greater benefit than performing FNAB on all thyroid nodules. Therefore, it is likely that a better way to approach the assessment of thyroid nodules would be to perform FNAB on all non-cystic nodules, as the present analysis has shown the ATA risk patterns do not provide an adequate clinical decision-making framework.

**Key Words:** thyroid nodule; thyroid cancer; ultrasound; Bayesian analysis; systematic review; Meta-analysis

Hurtado-Lopez LM, CARRILLO-MUÑOZ A, Zaldivar-Ramirez FR, Basurto-Kuba EOP, Monroy-Lozano BE. Assessment of diagnostic capacity and decision making based on the American Thyroid Association 2015 ultrasound system. *World J Methodol* 2022; In press

**Core Tip:** There is no analysis that evaluates the real diagnostic value of the 2015 ATA thyroid nodule risk patterns and their usefulness for clinical decision making; thus, we undertook this study to quantify both values.

## INTRODUCTION

Many decisions in medicine involve trade-offs, such as weighing the balance between diagnosing patients with disease <sup>5</sup> *vs* the cost of unnecessary additional testing for those who are healthy<sup>1</sup>. The traditional biostatistical approach to evaluating tests focuses on accuracy, evaluating calibration and discrimination, as well as using metrics such as sensitivity, specificity, or area-under-the-curve (AUC). These methods have several advantages: they are mathematically simple, can be used for binary or continuous predictors, and are relatively easy to interpret. However, their clinical relevance is low, because there is no way to correctly discriminate between two or more diagnostic tests when there are differences in sensitivity or specificity among them. Furthermore, they do not take into consideration the consequences of the decisions made<sup>2,3</sup>. To address these issues, analytical methods for decision making have been developed, which explicitly take into consideration <sup>5</sup> the clinical consequences of decisions. They provide data about the clinical value of tests, including either the risks associated with an incorrect diagnosis, or the benefits of a correct diagnosis, and so can determine whether or not these tests should be used to guide decisions regarding patient care<sup>4</sup>.

Ideally, the results of a diagnostic test should help physicians make a clear decision, meaning that, upon testing, we would either move from an epidemiological probability that a disease is present (testing threshold) to a lesser probability, and subsequent ruling out of the disease; or, on the contrary, the results could increase the probability to levels above the test-treatment threshold, and hence, point directly to treatment. However, sometimes the change in probability is higher than the testing threshold, but lower than the test-treatment threshold, in which case, the initial diagnostic test does not provide enough certainty to support decision-making regarding treatment, and additional diagnostic testing would therefore be required. This analytical process of diagnostic testing is known as “the threshold approach to clinical decision making”<sup>5,6</sup> and it provides a clear, objective, and rational method to determine whether additional diagnostic testing is needed, or not. **Figure 1**

2

The sensitivity and specificity of a test cannot be used alone to estimate the probability of disease in a patient, but the two parameters when combined into one measure, called the likelihood ratio, may be used in conjunction with disease prevalence to estimate an individual patient's probability of having disease. This probability can then be transformed into a post-test probability through Bayesian analysis, and this post-test probability, when applied to the threshold approach to clinical decision making, can then show us the true utility of a diagnostic test<sup>7-9</sup>.

Another method for clinical decision-making is decision curve analysis<sup>2,10,11</sup>. This method calculates a clinical “net benefit” for a diagnostic test *vs* treating all or no patients, across a range of threshold probabilities, defined as “the minimum probability of disease at which further intervention would be warranted”. The net benefit, unlike accuracy metrics such as discrimination and calibration, incorporates the consequences of the decisions that were made based on the results of a diagnostic test. Therefore, if you look at the net benefit of a range of reasonable threshold probabilities (Pt) for any given intervention and, if your test has a high net benefit across the whole range, you can say that your test can help you to make an adequate decision regarding that intervention. It is clear, then, that if we use analytical methods for clinical decision-making, in addition to traditional diagnostic testing, we will have a better understanding and clinical use of the diagnostic test results.

In the context of thyroid nodule assessment, the role of ultrasound has historically been very important. It was initially used only to identify the thyroid nodule and guide fine needle aspiration biopsy (FNAB), and later, it was further developed to identify nodule characteristics that would help differentiate between benign and malignant lesions, such as internal calcifications, hypoechoic, increased central blood flow, infiltrative margins, taller than wider, absence of halo, solid nodule, and nodule size<sup>12,13</sup>. However, various meta-analyses have shown that none of these characteristics alone can differentiate with certainty between benign and malignant lesions<sup>14-16</sup>.

1

The 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer<sup>17</sup> reengineered these images,

and, taking advantage of the practical aspects of ultrasound (non-invasive and readily accessible), described five sonographic patterns to establish the risk of malignancy: 1. Benign ultrasound (US) features consist of purely cystic nodules, with no solid component, estimated risk of malignancy <1%. 2. Very-low suspicion US features consist of <sup>1</sup> spongiform or partially cystic nodules without any of the sonographic features described in low, intermediate, or high suspicion patterns, estimated <sup>4</sup> risk of malignancy <3%. 3. Low suspicion US features consist of <sup>4</sup> isoechoic or hyperechoic solid nodule, or partially cystic nodule with eccentric solid areas, without microcalcification, irregular margin or extra thyroidal extension (ETE), or taller than wide shape, estimated risk of malignancy of 5-10%. 4. Intermediate suspicion US features consist of <sup>4</sup> hypoechoic solid nodule with smooth margins without microcalcifications, ETE, or taller than wide shape, estimated risk of malignancy of 10-20%. 5. High suspicion US features consist of <sup>3</sup> solid hypoechoic nodule or solid hypoechoic component of a partially cystic nodule with one or more of the following features: irregular margins (infiltrative, microlobulated), microcalcifications, taller than wide shape, rim calcifications with small extrusive soft tissue component, evidence of ETE, estimated risk of malignancy between 70-90%.

Based on these ultrasound descriptions and the size of the thyroid nodule, the ATA's intention is to standardize diagnostic behavior, from performing FNAB to simply keeping the thyroid nodule under observation, and the ATA patterns provide a clear and simple guideline to follow.

However, in the development of these ultrasound patterns, the origins of the percentages of malignancy suspicion assigned to each pattern is not clear, nor is the diagnostic accuracy of each pattern.

Several papers which have been published to date, both retrospective and prospective, have attempted to validate the risk patterns indicated in the ATA guidelines. These papers have shown similar findings in terms of malignancy rates in the categories of very low, low and intermediate suspicion, although not in high suspicion, which have generally been found to be a lower percentage<sup>18-40</sup>. However, these studies only

calculated the risk as a simple percentage and did not consider the diagnostic value in a clinical setting, which must be clearly established prior to decision-making.

Therefore, the objective of this paper is to determine the real diagnostic value of the ATA's classification system and to determine whether clinical decision-making based on this classification leads to an optimal management of thyroid nodules.

## **MATERIALS AND METHODS**

### **Study Design and Data Sources**

We made a systematic review of the published literature related to the American Thyroid Association ultrasound system for the initial assessment of thyroid nodules<sup>17</sup> from 2016 to date.

Data extraction was performed in accordance with the PRISMA statement for Diagnostic Test Accuracy Studies<sup>41</sup> and by searching PubMed-Medline for all articles published in the English language with the keywords: thyroid nodule, thyroid, ultrasound, US, ultrasonography, 2015 ATA. Related articles suggested by Pub-Med were also retrieved. Bibliographies of retrieved articles were searched independently and checked for additional studies.

### **Study Selection**

Our criteria for eligibility were articles that clearly reported data related to the ultrasound patterns described in the ATA 2015 Guidelines<sup>17</sup> and where the diagnosis of malignant or benign had been established either by histology reports, or two benign FNAB results.

### **Review process**

Two authors (LMHL and ACM) independently reviewed the articles and established the criteria for inclusion in the pooled data analysis, with disagreements resolved through discussion. Characteristics of the included and excluded articles are presented in **Figure 2**.

### **Methodologic Quality Assessment**



Methodologic quality was assessed using the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies 2) criterion.<sup>42</sup> Both reviewers (LMHL and ACM) scored the 7-item tool independently and disagreements were resolved by consensus (between LMHL, ACM, and FRZR) *via* a face-to-face discussion about each disagreement. **Table 1**

### **Data Analysis**

We used the Meta-DiSc, version 1.4 software (Ramon y Cajal Hospital, Madrid, Spain) in our meta-analysis<sup>43</sup>. The Mantel-Haenszel method of the random-effect model was used to calculate pooled sensitivity and specificity with corresponding 95% confidence intervals.

We analyzed these cases first using traditional diagnostic values and then the threshold approach to clinical decision making and decision curve analysis.

To perform the threshold approach to clinical decision making, it is important to understand that the indifference point for the choice between withholding therapy and performing a diagnostic test is a probability of disease designated here as the "testing" threshold (Tt). The indifference point for the choice between performing the diagnostic test and administering treatment is a probability of disease designated here as the "test-treatment" threshold (Ttrx).

Because the thresholds define these two indifference points, the physician can be guided by the calculated thresholds and estimated probability of disease in a given patient. As illustrated in the **Figure 1**, the best choices are to withhold both treatment and the test if the probability of disease is smaller than the testing threshold, to administer treatment without testing if the probability of disease is greater than the test-treatment threshold, and to perform the test only if the probability of disease falls between the two thresholds.

The threshold levels were developed as follows<sup>5</sup>: the testing threshold (Tt) consisted of the frequency of thyroid cancer, as reported in the 1015 ATA Guidelines<sup>17</sup>, estimated to be between 7% and 15%, making an average frequency of 11%.

For the test-treatment threshold (Ttrx), we used the formula described by Pauker *et al*<sup>5</sup> shown in **figure 3**. Using this formula, we estimated the Brx at 20% considering that the

survival rate of a patient with papillary thyroid cancer is 98% with an early diagnosis, and 78% when there has been distant metastases<sup>17,44</sup> Rrx at 4.2% when calculating an overall average rate of morbidity in total thyroidectomies, including permanent injury of the recurrent laryngeal nerve<sup>45-47</sup>, injury to the external branch of the upper laryngeal nerve<sup>48,49</sup>, hypoparathyroidism<sup>50-53</sup>, and hematoma<sup>54,55</sup>. Rt at zero, as performing an ultrasound does not expose the patient to any risk; Pneg/nd at 0.98, representing true negatives, calculated based on ultrasound patterns<sup>17</sup> for benign and very low suspicion; Pneg/d at 0.10 representing false negatives, as high-suspicion patterns detect 90% of cancers<sup>17</sup>. The resulting Ttrx was 67.2%.

To perform the decision curve analysis, the value for probability thresholds (Pt), defined as the minimum probability of disease at which further intervention would be warranted,<sup>2,10</sup> is a clinical judgement, and was calculated under five conditions: The first was set by simply identifying the general probability of cancer in a thyroid nodule, which is 11%, according to the ATA. The second was set to 28.1%, by taking an intermediate point between the Tt (11%) and the Ttrx (67.2%). The third condition was set to 3%, which is the probability of cancer in the very low pattern. The fourth condition was set to 7%, which is the lower range of probability of cancer according to the 2015 ATA guidelines. The fifth condition was set to 15%, which is the highest range of probability of cancer according to the 2015 ATA guidelines. We used these five Pt to calculate the net benefit of each ATA ultrasound risk pattern.

We calculated the net benefit, expressed as the number of unnecessary interventions avoided, in this case the number of FNABs avoided, using true negatives rather than true positives, and using the following formula: Net Benefit for unnecessary interventions = (true negative count/total number of patients) - (false positive count/total number of patients) x (Pt / (1- Pt)) in order to determine the number of unnecessary biopsies that had been performed, without missing any patients with cancer, in each of the ATA ultrasound patterns. We calculated these in all five possible Pt<sup>10,11</sup>.

Once we were able to determine the best Pt by calculating the highest number of unnecessary FNABs avoided, we then calculated the net benefit for this Pt using curve

decision analysis for each ultrasound pattern, using the following formula: Net benefit=(true positive count/ total number of patients) – (false-positive count / total number of patients) X (pt/1-pt). This result would then be compared with the net benefit of performing FNAB on all thyroid nodules. **Figure 4**

The resulting net benefit for each ultrasound pattern would have to be of greater value than the net benefit of performing FNAB on all thyroid nodules for it to be considered a diagnostic model that provides the correct identification of those thyroid nodules which can be safely excluded from FNAB.

We extracted the information, grouped it by author, and added all the cases together, obtaining diagnostic values for sensitivity, specificity, and positive and negative predictive values for each of the categories. We determined the Youden's J statistic using the following formula: sensitivity + (specificity -1); and the positive likelihood ratio (LR) using the following formula: sensitivity/(1-specificity), which we then calculated with Bayesian analysis using the following formula: post-test odds= pretest odds X LR, and, in the final step, we converted the post-test odds into post-test probability.

Due to the design of this study, approval by an institutional review board was not required.

## **RESULTS**

Our initial search retrieved 305 articles, and the summaries were reviewed by two authors (LMHL and ACM) to select those that met the required criteria. This resulted in 52 articles that we reviewed in their full text, before finally selecting 13 articles<sup>25,28,29-32,34-40</sup> which contained the required information in accordance with the design of the study. 9 were retrospective studies<sup>28,29,32-37,39,40</sup> and 4 were prospective studies<sup>25,30,31,38</sup> **table 2 and figure 3.**

The data from 8,445 thyroid nodules was obtained, of which 3,937 (46.62%) were malignant and 4,508 (53.38%) were benign. The average size of the tumors was 18.5 mm (5mm to 71 mm).

When grouping the nodules into risk patterns, we found that the benign pattern was reported in only 6 of the 13 articles<sup>30, 32, 34, 36, 39, 40</sup>, for a total of 62 nodules in the category, and all of these corresponded to histopathological benign nodules, therefore we decided to exclude this pattern from our analysis.

For the very-low suspicion pattern, there were a total of 848 cases. Of these, 832 were benign and 16 were malignant, meaning that in this pattern, the simple percentage of malignancy was 1.8%. The Youden's J statistic was -0.18. The diagnostic value can be found in **table 3 and figure 5**.

There were 1800 nodules in the low-suspicion pattern. Of these, 1621 were benign and 179 were malignant, meaning that in this pattern, the simple percentage of malignancy was 9.4%. The Youden's J statistic was -0.31. The diagnostic value can be found in **table 3 and figure 6**.

There were 1673 nodules in the intermediate-suspicion pattern. Of these, 1340 were benign and 333 were malignant, meaning that in this pattern, the simple percentage of malignancy was 19.9%. The Youden's J statistic was -0.22. The diagnostic value can be found in **table 3 and figure 7**.

There were 4124 nodules in the high-suspicion pattern. Of these, 715 were benign and 3404 were malignant, meaning that in this pattern, the simple percentage of malignancy was 82.5%. The Youden's J statistic was 0.71. The diagnostic value can be found in **table 3 and figure 8**.

After using Bayesian analysis to determine the post-test probability of cancer, our results were as follows: very low, 0.85%; low, 2.6%; intermediate, 6.7%; and high, 40.9%.

#### **Figure 9.**

The net benefit, expressed as the number of interventions (FNAB) avoided, with a Pt calculated for all five possibilities, can be seen in **figure 10**.

The NB calculation for a Pt of 11% was: very low, -0.01028; low, -0.002526; intermediate, 0.019821; and high, 0.393207. This means that the intermediate pattern was only able to identify 1.9 of every 100 patients with cancer, and the high pattern was only able to

identify 39 of every 100 patients with cancer. In the very low and low pattern cases, there is no obvious interpretation for negative net benefit using this type of framework. The true- and false-positive count for performing FNAB in all thyroid nodules is simply the number of patients with and without thyroid cancer, respectively. Calculating the net benefit for this strategy gives:  $(3937/8445) - (4508/8445) \times (0.11/0.89) = 0.40022$ . The net benefit for each of the ATA suspicion patterns studied is less than that of performing FNAB on all nodules.

There was heterogeneity in the initial results, which could create a bias risk, so we made an analysis of the subgroups. **Figure 10**. The homogenous group was reduced to a population of 5151 thyroid nodules: 443 from the very low risk category with 2.9% frequency of cancer, 316 cases from the low risk category with 16.1% frequency of cancer, 946 from the intermediate risk category with 20.7% frequency of cancer, and finally in the high risk category, 3446 nodules with 81.5% frequency of cancer. **Figure 11**. After using Bayesian analysis to determine the post-test probability of cancer in the subgroups analysis, our results were: very low, 0.7%; low, 1.8%; intermediate, 0.2%; and high, 37.4%. **Figure 12**.

## DISCUSSION

This article demonstrates how complicated it can be to interpret diagnostic tests and use them for clinical decision making.

It is important to recognize that when we analyze decision-making based on diagnostic tests, there is no one “perfect test” that can either rule out or diagnose a disease with 100% accuracy (in this discussion, thyroid cancer). Therefore, for every patient who undergoes a diagnostic test to help guide decision-making, there will always be three possible paths; the first is to keep the nodule under observation, the second is to do additional diagnostic testing, and the third is to proceed with treatment.

The results found in this study show an overall frequency of cancer of 46.6%. This frequency is higher the 7 to 15% reported by the ATA because the data was obtained from reference center hospitals with a high volume of thyroid cancer cases. Although

this could be considered as selection bias, it is important to emphasize that the diagnostic values which we used in this analysis, in particular sensitivity and specificity, and the decision-making based on these diagnostic tests, are not affected by the frequency of this disease.

We must also keep in mind that while indeed cancer was most frequent in the high-suspicion pattern, 15.5% of the cases were found among the very low, low and intermediate-suspicion patterns.

It is also important to note that we did not include the benign pattern, as it was only reported in 6 of the studies analyzed, and of the 62 nodules reported with this pattern, all were confirmed to be benign. Therefore, it is reasonable to assume that any thyroid nodule that is purely cystic, regardless of size, is benign, and as such, we did not consider it necessary to include them in this study.

Another potential bias is that we had to exclude the size of the thyroid nodules, as detailed information for each nodule was not available in the articles we reviewed. However, we do not consider this as selection bias, because the range of sizes reported in the articles that we reviewed were an average size of 18.5 mm (ranging from 5mm to 71 mm), which is within the typical range found in most nodules in the day-to-day medical practice. We therefore consider that this variable is not of particular importance in deciding which patients will need FNAB. Nodule size has always generated controversy in terms of how it might affect the diagnosis of malignancy<sup>56-59</sup>.

The final limitation in our study is that most of the ultrasound images, and consequently their ATA ultrasound classifications, were performed retrospectively. However, this is unlikely to be significant, as the cases come from high-volume thyroid disease reference centers, and were interpreted by highly qualified ultrasound experts<sup>60,61</sup>.

Our results show that, if we were to assume that the simple percentage for the presence of cancer in each ultrasound pattern was a reliable diagnostic tool, they would coincide perfectly with the ranges published by the 2015 ATA guidelines<sup>17</sup>, in that the very low pattern had a 1.8% malignancy rate and the ATA reports <3%; in the low pattern, a

9.4% malignancy rate and the ATA reports 5 to 10%; in the intermediate pattern, a 19.9% malignancy rate and the ATA reports 10 to 20%; and finally, in the high pattern, an 82.5% malignancy rate and the ATA reports between 70 and 90%.

However, when we analyze the results of the traditional biostatistical approach to evaluating tests and Youden's J statistic, together with the threshold approach to clinical decision-making and curve decision analysis, we can see that these ultrasound patterns are no longer as clear, or as practical.

When we analyze a traditional diagnostic value, we can see that the very low, low and intermediate patterns have regular to good specificity, but very low sensitivity, therefore they could, in theory, be used to correctly identify and rule out patients who do not have the disease, with very few false-positive results. In effect, this is the reason the ATA has grouped thyroid nodules into patterns; to guide decisions regarding whether to do further FNAB testing or not. However, our study shows that relying strictly on these patterns would have resulted in false positives: for very low, 18.1%; low, 36%; and intermediate, 29.7%, meaning that there would have been patients with cancer that went undetected using the ultrasound patterns, and, because the Youden's J statistic for these three patterns was below 0, they would have been identified as "without diagnostic value".

On the other hand, the high pattern gave a more accurate diagnostic value with sensitivity of 86.6%, specificity of 84.1%, and a Youden's J statistic of 0.7, and so, this pattern is better able to discriminate between malignant and benign cases, and therefore, can be used reliably.

When analyzing decisions based on the threshold approach to clinical decision making, we were able to determine that the testing threshold was 11% and the test-treatment threshold was 67.2%. Once we determined the values of the ultrasound patterns using Bayesian analysis, we concluded that the very low pattern had a cancer risk of 0.3%, low of 1.5%, and intermediate of 3.4%; meaning that these three patterns would be below the testing threshold and, as a logical consequence of this method of analysis, thyroid nodules classified within these three patterns could simply be left under

observation, since by definition, cancer had been ruled out, rendering further testing unnecessary. However, decisions using this type of analysis would have resulted in 15.5% of the cancers in this study going undiagnosed, troubling if we keep in mind that intermediate-risk nodules are mostly isoechoic nodules, which could be follicular carcinoma, a potentially high-risk thyroid cancer with a poor prognosis. If physicians do not perform FNAB on Bethesda IV intermediate-risk nodules, this group of potentially high-risk cancers can be missed.

The high-risk pattern, on the other hand, analyzed in the same way, raises the probability of thyroid nodule cancer to 40.3%, which is between the testing threshold (11%) and the test-treatment threshold (67.2%), therefore indicating the need to do additional FNAB testing.

With regard to heterogeneity, we made an analysis by subgroups, and it was interesting to note that the frequency of malignancy was quite similar to the initial analysis. Even more importantly, after using Bayesian analysis with the threshold approach to clinical decision making, the results were still the same, where, in the groups of very low, low, and intermediate risk, cancer was ruled out, and the high-risk group fell in the area where additional testing would be necessary.

When analyzing decision-making based on the net benefit expressed as the number of unnecessary interventions avoided, we made our calculations using several different Pt scenarios, with a Pt of 11% being the one that most frequently matched with the test threshold defined in the analysis discussed above. We made this analysis, because the classification of ultrasound patterns was intended to determine who should undergo FNAB and who could be kept under observation, in order to reduce the number of unnecessary FNABs. When evaluating the capacity to avoid unnecessary FNABs, we found that the highest capacity for avoidance was: 42 of 100 FNABs in the very low pattern with a Pt of 11%, 31 of 100 FNAB in the low pattern with a Pt of 11%, 35 of 100 FNAB in the intermediate pattern with a Pt of 11%, and 43 of 100 FNABs in the high-risk pattern with a Pt of 11%. This means that attempting to avoid unnecessary FNABs



is inadvisable, since over half of them were indeed necessary, and a cancer diagnosis would have been missed if FNAB had not been performed.

When calculating NB with a Pt of 11%, the very low and low patterns had a net benefit of less than zero, so there was no obvious interpretation for negative net benefit using this type of framework. In the case of intermediate patterns, NB could only detect slightly fewer than 2 out of every 100 patients with cancer, and in high-risk patterns, it could only detect 39 out of every 100 patients with cancer.

Further still, when comparing the net benefit of each ultrasound risk pattern studied *vs* that of systematically performing FNAB on all thyroid nodules, we found that none were higher than that of performing FNAB on all thyroid nodules. This clearly demonstrates that using these ultrasound categories to guide the decision regarding who should undergo FNAB is inferior to performing FNAB on all thyroid nodules.

It is also important to understand that clinical decisions should not be made based on simple percentages. Instead, it is preferable to apply a rigorous study based on clinical decision models, which, although they may seem complicated to calculate, they are, in reality, the only way to make an accurate professional diagnosis for the thyroid nodule patient. It is also clear that the clinically low aggressivity of malignant thyroid cancer has allowed for a margin of error since it can be detected in a formerly undiagnosed patient at a later date. However, this diagnostic behavior would be unprofessional, as any patient who consults a physician for a thyroid nodule expects an accurate diagnosis.

From a practical standpoint, the results of this study indicate to the physician that, when evaluating thyroid nodules with ultrasound, only the high-risk and benign categories are clinically useful, indicating FNAB for high-risk cases and observation for the benign pattern. However, if the ultrasound shows a pattern of very low, low, or intermediate risk, the physician should recommend FNAB, as opposed to the current recommendations of observation only, as there is a risk of cancer of up to 15% that could go undiagnosed.

## **CONCLUSION**

It is clear from our three types of analysis, that the only ATA diagnostic pattern that is clinically useful is the high-suspicion pattern, in which case, without a doubt, FNAB should be performed. However, and even more importantly, curve decision analysis has demonstrated that using these ultrasound risk patterns to decide which patients need FNAB does not provide a greater benefit than performing FNAB on all thyroid nodule patients. Therefore, we conclude that a better way to approach the assessment of thyroid nodules would be to perform FNAB on all non-cystic nodules, as the present study has shown that the ATA risk patterns do not provide an adequate clinical decision-making framework.

# 5%

SIMILARITY INDEX

### PRIMARY SOURCES

- |          |   |               |
|----------|---|---------------|
| <b>1</b> | <a href="http://www.liebertpub.com">www.liebertpub.com</a><br><small>Internet</small>   | 51 words — 1% |
| <hr/>    |   |               |
| <b>2</b> | <a href="http://onlinelibrary.wiley.com">onlinelibrary.wiley.com</a><br><small>Internet</small>   | 49 words — 1% |
| <hr/>    |   |               |
| <b>3</b> | <a href="#">"Practical Management of Thyroid Cancer", Springer Science and Business Media LLC, 2018</a><br><small>Crossref</small>  | 43 words — 1% |
| <hr/>    |   |               |
| <b>4</b> | <a href="#">Raquel Andrade de Siqueira, Ana Paula dos Santos Rodrigues, Lucas Massao Miamae, Eduardo Kiyoshi Tomimori et al. "Thyroid Nodules in Severely Obese Patients: Frequency and Risk of Malignancy on Ultrasonography", Endocrine Research, 2019</a><br><small>Crossref</small> | 43 words — 1% |
| <hr/>    |   |               |
| <b>5</b> | <a href="http://Bmcmedinformdecismak.Biomedcentral.Com">Bmcmedinformdecismak.Biomedcentral.Com</a><br><small>Internet</small>   | 40 words — 1% |
| <hr/>    |   |               |
| <b>6</b> | <a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a><br><small>Internet</small>   | 29 words — 1% |