DOI: https://doi.org/10.4038/jdp.v17i2.7753 *Journal of Diagnostic Pathology 2022;17(2):1-8.*

Research paper

Proposal for a scoring system to differentiate thyroid malignancies from benign lesions using selected demographic and sonographic data and the Bethesda system for reporting thyroid cytopathology

S. A. Gunaratne[1], M. Godakande[1], W. Sathkorala[2], K.W. Dammika[1], C. J. Wijesinghe[3]

[1] *Department of Pathology, Base Hospital Avissawella, Sri Lanka*

[2]*Department of Radiology, Base Hospital Avissawella, Sri Lanka*

[3]*Department of Community Medicine, Faculty of Medicine, University of Ruhuna, Sri Lanka*

Submitted on 21.03.2020. Accepted for publication on 27.06.2023.

Abstract

Background and objective: Differentiating benign from malignant thyroid nodules by using cytological findings alone can be misleading. Concurrent use of a reliable risk prediction tool can facilitate clinical decision-making and reduce the burden on patients and the health care system by minimizing unnecessary surgeries.

Methods: A cross-sectional study was conducted at Base Hospital, Avissawella, using data from patients with a confirmed histological diagnosis of thyroid nodules, who had ultrasonography and fine needle aspiration cytology (FNAC) prior to surgery. A scoring system was developed to predict the risk of thyroid nodules being malignant by incorporating age, sex, FNAC test results, and ultrasound findings. The benign or malignant status of the nodules determined using the scoring system was then compared with the histological diagnosis at various threshold settings. Receiver operating characteristics (ROC) analysis was used to determine an optimal cut-off threshold for the developed score analyses.

Results: A total of 138 cases were investigated. The scoring system was appropriate for clinical use as indicated by an area under ROC curve (AUC) of 0.752 (95% confidence interval = 0.622-0.882; $p<0.001$). A cut-off score of 8.5 had a sensitivity of 80% and a specificity of 54.5% for the diagnosis of malignancy. The proposed scoring system had a high negative predictive value at low score levels and a high proportion of cases were included in this category. At high score levels, it had a greater positive predictive value to differentiate benign from malignant nodules.

Conclusion: The proposed scoring system can be used as an adjunct to differentiate benign lesions from thyroid malignancy since the AUC was 0.752. We suggest a cut-off score of 8.5 as an acceptable trade-off between a sensitivity of 80% and a specificity of 54.5%. At scores of 7.5 and below, the nodules can be considered benign and safely followed up.

Keywords: thyroid malignancy, risk prediction score, ROC, Bethesda grading for cytology, ultrasonography

Corresponding author: Dr Sandhini Gunaratne
Department of Pathology
Lady Ridgeway Hospital, Sri Lanka
sandigunaratne@gmail.com

Introduction

The prevalence of thyroid cancer is increasing in Sri Lanka with a global prevalence of around 1% (1). The increase may be attributable to the use of sensitive diagnostic procedures as well as a genuine increase in the incidence. According to the 2019 National Cancer Incidence and Mortality Data for Sri Lanka, thyroid cancer is the second most common cancer diagnosed in women, with 2022 reported cases (2).

Predicting thyroid malignancy using cytological features alone is challenging to the pathologist since 10-20% of thyroid cytology specimens have an inconclusive diagnosis (1). It is important to confirm the true nature of the lesion before surgery to prevent unnecessary surgical procedures (1,2,3). If the diagnosis is inconclusive, the patient might have to undergo an initial thyroid lobectomy followed by total thyroidectomy as a second surgery if histology confirms the lesion to be malignant. If there is a possibility to reduce the rate of inconclusive diagnoses and confidently differentiate malignant thyroid nodules from benign lesions preoperatively, the appropriate treatment modality, either to be managed conservatively or by total thyroidectomy, can be planned as a single-step procedure. This approach minimizes the burden on clinicians and reduces unnecessary surgical interventions, costs and hospital stays for patients.

The current Bethesda scoring system for thyroid cytology diagnosis has several drawbacks (3). The Bethesda III category is a grey area in which patients are followed up instead of receiving a definitive surgical intervention.

Several studies have attempted to predict the malignant behaviour of thyroid nodules by using age, sex, FNAC test results, thyrotrophin levels, thyroglobulin antibody levels, hot/cold status of the nodule, and ultrasound parameters in combination (4-9). However, no definite consensus has been reached. Furthermore, to our knowledge, there are no published reports on a similar study done in the Sri Lankan setting to assess the feasibility of differentiating thyroid malignancies from benign lesions using a combination of these parameters.

This study aimed to develop and assess a scoring system to increase the diagnostic accuracy of differentiating thyroid malignancies from benign lesions using a combination of selected demographic and sonographic data along with the Bethesda classification.

Methodology

Study design and participants

This was a retrospective, cross-sectional study conducted using secondary data available at Base Hospital, Avissawella, Sri Lanka. The records of patients with thyroid nodules who underwent surgery from 2013 to 2016 were reviewed. Patients with a histologically confirmed diagnosis and available thyroid ultrasonography and FNAC findings, namely the Bethesda classification, were included in the study.

Sampling method and sample size

All the patients fulfilling the inclusion criteria were included with a total number of 138 subjects. This was an adequate sample size for assessing sensitivity and specificity, calculated using the formula given by Hulley and Cummings {N= 4 Zα2 P (1-P)/ d2}, with a confidence level of 95%, an expected sensitivity of 90%, a specificity of 85% and a precision of 0.2 (10).

Data collection

The data collection commenced after obtaining ethical approval from the Ethical Review Committee, Faculty of Medicine, University of Ruhuna, Sri Lanka. Secondary data available in the FNAC, ultrasound scan, and histopathology reports of the selected patients were used.

In most cases, the sonographic diagnoses were made by a single specialist. Therefore, to minimize possible variation and to increase validity, based on the specific characteristics of the nodules documented in the sonography reports, a second opinion was obtained. The

cases with doubtful ultrasonographic features were excluded from the study.

Development of the scoring system

The proposed scoring system for differentiating malignant and benign thyroid nodules was developed using five selected variables in combination: age of the patient, sex, ultrasonography impression of thyroid lesions (i.e. solitary nodule, multinodular or diffuse enlargement), final ultrasonographic diagnosis of the lesion as malignant or benign and the FNAC diagnosis. The variables were in categorical format. Based on the expert opinion of relevant specialists (radiologists/pathologists/surgeons) and existing evidence in the literature, a score was assigned to the different response categories of each variable. This score indicated the risk of a particular nodule being malignant depending on the category (Table 1).

The sonographic features of thyroid nodules: the number of nodules, the status of the lesional border (regular or irregular), the presence of microcalcification, internal vascularity, hypoechogenicity and solid/cystic nature have a predictive ability to differentiate benign and malignant nodules. For example, if there are multiple cystic nodules with regular borders, without microcalcification or internal vascularity, the nodule is more likely to be benign, and ultrasound-detected solitary nodules have a greater likelihood of being malignant. Additional weights were given to both FNAC and ultrasonography findings (variable 4 and variable 5 according to Table 1).

Considering the risk of malignancy, each of the variables 1,2 and 3, i.e. age, sex and ultrasonography impression, was assigned a score of 1. The minimum score of 1 indicates a low risk and the maximum score of 2 indicates a higher risk. The final ultrasound diagnosis was assigned a minimum score of 2 for a benign lesion and a maximum score of 6 for a malignant lesion. Similarly, the scores for FNAC diagnosis were from 2 to 10 depending on the risk of malignancy. The final score was calculated by adding up the scores of all five variables. The minimum cumulative score was 7 and the maximum score was 22.

Data analysis

The data were coded and entered into a database created using SPSS (version 20.0). The nodules were classified as likely to be benign or malignant according to the scoring system using a series of cut-off thresholds. The benign or malignant status of the nodules determined by using the scoring system was compared with the histological diagnosis, which is the gold standard. The validity parameters: sensitivity, specificity, false

Variable		The values assigned for each category				
		Score 1	**Score 2**	**Score 3**	**Score 4**	**Score 5**
1	Sex of the subject	Female (Score 1)	Male (Score 2)			
2	Age of the subject	≤ 44 years (Score 1)	>44 years (Score 2)			
3	*Ultrasonography impression of the nodule	Multinodular/ diffuse (Score 1)	Solitary nodule (Score 2)			
4	*Final ultrasonographic diagnosis	Benign Score 1x2	Suspicious for malignancy Score 2x2	Malignant Score 3x2		
5	FNAC diagnosis	Bethesda II (Score 1x2)	Bethesda III (Score 2x2)	Bethesda IV (Score 3x2)	Bethesda V (Score 4x2)	Bethesda VI Score 5x2

*Variables 4 and 5: Scores were doubled to give a greater weight

Table 1. The proposed scoring system for differentiating benign and malignant thyroid nodules.

positive/negative rates (FPR/FNR) and true positive/negative rates (TPR/TNR) at different cut-off values were calculated. ROC analysis was performed to determine an optimal cut-off threshold for the developed score (11).

Results

Sample characteristics

There were 138 cases, with the majority being females (n=124;90%). The majority (89.1%; n=123) of the nodules were benign. Histological assessment confirmed 10.9% (n=15) of lesions to be malignant. Of the benign nodules, 52.8% (n=71) were colloid nodules, 23.1% (n=32) showed chronic autoimmune thyroiditis, 3.6% (n=6) were hyperplastic nodules, 8.7% (n=12) were follicular adenomas and one was a Hurthle cell adenoma. The histological subtypes of malignancy included papillary carcinoma (7.2%;n=10), follicular variant of papillary carcinoma (1.4%; n=2) and follicular carcinoma (2.1%;n=3). Most patients who had malignant thyroid nodules were females (93%; n=14).

Predictors of malignancy

The ROC curve was drawn for the data set using sensitivity and the FPR at different cut-off thresholds (Figure 1).

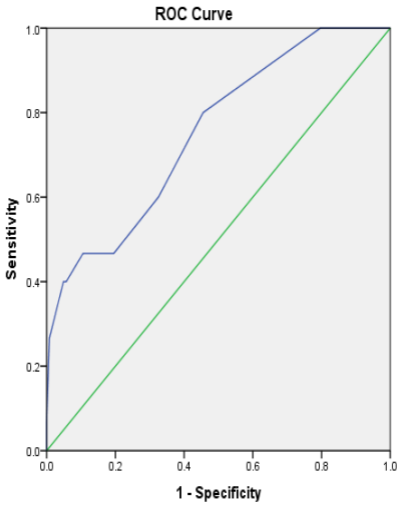

Figure 1. ROC curve for predicting thyroid malignancy using the proposed scoring system.

Cut-off score	**Sensitivity (TPR)	**1 – Specificity (FPR)	Specificity (TNR)	1 – Sensitivity (FNR)
7.0	1.000 (100%)	1.000 (100%)	0.000 (0%)	0.000 (0%)
7.5	1.000 (100%)	0.797 (79.7%)	0.203 (20.3%)	0.000(0%)
8.5	**0.800 (80%)**	**0.455 (45.5%)**	**0.545 (54.5%)**	**0.200 (20%)**
9.5	0.600 (60%)	0.325 (32.5%)	0.675 (67.5%)	0.400 (40%)
10.5	0.467 (46.7%)	0.195 (19.5%)	0.805 (80.5%)	0.533 (53.3%)
11.5	0.467 (46.7%)	0.106 (10.6%)	0.894 (89.4%)	0.533 (53.3%)
12.5	0.400 (40%)	0.057 (5.7%)	0.943 (94.3%)	0.600 (60%)
13.5	0.400 (40%)	0.049 (4.9%)	0.951 (95.1%)	0.600 (60%)
14.5	0.267 (26.7%)	0.008 (.8%)	0.992 (99.2%)	0.730 (73.3%)
15.5	0.067(6.7%)	0.000 (0%)	0.000 (100%)	0.933 (93.3%)

**FPR, false positive rate; **TPR, True positive rate); **(coordinates of the curve)
TNR, True negative rate; FNR, false negative rate (calculated values)

Table 2. Coordinates of the ROC curve for predicting thyroid malignancy using the proposed scoring system along with calculated values

The area under the curve (AUC) was 0.752 (95% confidence interval = 0.622-0.882), which was considered a fair test for predicting the benign or malignant nature of nodules (10). Based on the coordinates of the ROC curve, all patients having a cumulative score of 7.5 or less could be categorized as benign (Table 2). On the other hand, when the cut-off score was set at 15.5, all nodules with scores that were equal or greater could be categorized as malignant.

A cut-off score of 8.5 was considered an acceptable trade-off between sensitivity (TPR) at 80% and specificity (TNR) at 54.5%. The FNR was 20%. The proposed scoring system had high negative predictive value at low score levels.

The findings revealed that the proposed scoring system has a greater ability to predict benign status than malignancy as evidenced by the high negative predictive values at low scores. At high score levels, it has a greater positive predictive value for distinguishing benign from malignant nodules (Table 3).

Considering the Bethesda classification, a total of 41 cases were classified into Bethesda III and IV categories which were considered indeterminate cytology categories. In the Bethesda III category, there were no malignant cases in cut-off scores below 9, indicating that it is a reliable threshold score to exclude malignancy, and therefore, these patients could be safely followed up.

Cut-off score	Negative predictive value	Positive predictive value	True positives	True negatives	False positives	False negatives
07	100%	12%	14 (10%)	25 (18%)	99 (72%)	0 (0%)
08	95%	17%	12 (8%)	67 (49%)	56 (41%)	03 (2%)
09	93%	18%	09 (7%)	83 (60%)	40 (29%)	06 (4%)
10	92%	22%	07 (5%)	99 (72%)	24 (17%)	08 (6%)
11	93%	35%	07 (5%)	110 (80%)	13 (9%)	08 (6%)
12	92%	50%	06 (4%)	115 (84%)	07 (5%)	10 (7%)
13	92%	50%	06 (4%)	117 (85%)	06 (4%)	09 (7%)
14	91%	80%	04 (3%)	122 (88%)	01 (1%)	11 (8%)
15	89%	100%	01 (1%)	123 (89%)	0 (0%)	14 (10%)

Table 3. The distribution of samples according to diagnostic categories at different cut-off values of the proposed

Cut-off score	Bethesda III			Bethesda IV		
	Benign n (%)	Malignant n (%)	Total n (%)	Benign n (%)	Malignant n (%)	Total n (%)
9.00	2 (100%)	0 (0%)	2 (100%)	0 (0%)	0 (0%)	0 (0%)
10.00	11 (84.6%)	2 (15.3%)	13 (100%)	0 (0%)	0 (0%)	0 (0%)
11.00	7 (100%)	0 (0%)	7 (100%)	1 (100%)	0 (0%)	1 (100%)
12.00	4 (100%)	0 (0%)	4 (100%)	2 (66%)	1 (33%)	3 (100%)
13.00	0 (0%)	0 (0%)	0 (0%)	1 (100%)	0 (0%)	1 (100%)
14.00	1 (100%)	0 (0%)	1 (100%)	4 (66%)	2 (33%)	6 (100%)
15.00	0 (0%)	0 (0%)	0 (0%)	1 (33%)	2 (66%)	3 (100%)
Total	**25**	**2**	**27**	**9**	**5**	**14**

Table 4. Classification of the benign and malignant status of nodules according to different cut-off levels in relation to Bethesda categories III and 1V

Discussion

Unnecessary thyroid surgical interventions can be minimized to a considerable extent if a reliable measure is available to differentiate thyroid malignancies from benign lesions, particularly when the FNAC findings alone are inconclusive. In the present study, we developed a composite score using the age of the subject, sex, ultrasonography impression of the thyroid lesions (solitary nodule, multinodular or diffuse enlargement), final ultrasonographic diagnosis (malignant or benign) and the Bethesda category of the lesion as variables to predict whether a thyroid nodule is malignant or benign.

The subsequent ROC analysis, which compared these scores with the histology as the gold standard diagnosis, revealed that the proposed scoring system can be adopted for clinical use with considerable certainty as indicated by an AUC of 0.752 (95% CI=0.622-0.882; p<0.001). According to our findings, all nodules having a score of 7.5 or less can be considered benign since the sensitivity is 100%. They can be followed up conservatively avoiding unnecessary surgical interventions. This is of great importance to patients who have contraindications for surgery.

In a study conducted by Witczak et al. (4), the TSH values, male gender and significant ultrasound characteristics were used in a scoring system to screen for thyroid malignancies, and the ROC analysis revealed an AUC of 0.77, which was comparable with our findings (AUC=0.752). In both the present study and Witczak's study, there was a comparable level of sensitivity at low scores (80% and 86.9%, respectively) and high specificity at higher scores (99% and 94.8%, respectively).

A study done in Sri Lanka by Pathirana et al. (7) using a score for screening thyroid nodules based on ultrasonography findings showed that AUC obtained by ROC analysis was 0.822, indicating that a combination of ultrasonographic criteria increases the accuracy in predicting the benign nature of thyroid nodules.

Boelaert et al. (8) used a risk stratification algorithm derived from multivariate analysis as a predictive formula for malignancy in thyroid nodules. This study used age, sex, type of thyroid enlargement (solitary/multinodular /diffuse) and serum TSH levels as predictor variables. This study revealed that if the serum TSH level was in the upper third of the normal range there was a significant association with malignancy [(OR=2.91 (95% CI=1.49-5.71)]. In the present study, TSH level was not considered as a variable in risk prediction because laboratories use different methods to measure TSH levels (8), which may affect the test values.

By using a regression model, Maia et al. (5) found a strong risk of malignancy in patients aged >39 years having ultrasonically suspicious thyroid nodules with microcalcification, central flow, irregular borders and hypoechogenicity, with a predictive accuracy of 81.7% (6).

A limitation of the current study was the unavailability of Thyroid Imaging and Reporting and Data Systems Score (TIRADS) in most reports. Therefore, in practice, the best approach would be to have multidisciplinary team meetings (MDT) on cases that fall into a grey area to decide on further management.

The indeterminate thyroid FNAC findings categorized under Bethesda III and IV are challenging for patient management, and in these categories, the proposed scoring system can be of great additional benefit in deciding the management. The findings of the present study were based on a limited number of Bethesda III/IV cases (n=41) and need to be tried out in a much larger sample.

The strength of our study design was that the scoring was based on generally available patient data. The other studies on similar scoring systems to differentiate thyroid malignancies had incorporated the thyroid hormone levels, tumour markers such as galectin-3 in cytology samples and BRAF mutation which is characteristic of papillary thyroid carcinoma (5). However, these additional tests are not available in the local setting.

Conclusion

The proposed scoring system has an AUC of 0.752, which is considered a fair test and can be used as an adjunct for differentiating benign thyroid lesions from thyroid malignancies. We suggest that a cut-off score of 8.5 is an acceptable trade-off between a sensitivity (TPR) of 80% and a specificity (TNR) of 54.5 %. At scores of 7.5 and below, the nodules can be considered benign, and patients can be safely followed up.

The proposed scoring system in this study is a simple and practical scoring system that can be easily applied to diagnoses and facilitate decision-making in thyroid nodules with inconclusive diagnoses, thereby reducing unnecessary surgeries. As this study was carried out in a small sample, further studies should be carried out in larger samples involving other centres to ensure the reproducibility of results.

References

1.Lioyd RV, Osamura RY, Kloppel G, Rosai J. WHO classification of Tumours. Pathology and Genetics Tumours of endocrine Organs :De Lellis, Williams ED (eds): Thyroid and parathyroid tumours.2004;51-93. 4th edition.

2.National cancer Incidence and Mortality Data Sri Lanka 2019: Ministry of health and indigenous medicine http://nccp.health.gov.lk (accessed 05 December 2022)

3.Ali SZ, Cibas ES (eds) : The Bethesda System for Reporting Thyroid Cytopathology. 1st edition. New York: Springer. 2010.

4.Witczak J, Taylor P, Chai J, Amphlett B, Soukias Jean-Marc, Das Get al Predicting malignancy in thyroid nodules: feasibility of predictive model integrating clinical,

biochemical, and ultrasound characteristics. Thyroid Research 2016; 9:4. https://doi.org/10.1186/s13044-016-0033-y

5. Maia FR, Matos PS, Silva BP, Pallone A T, Pavin EJ, Vassallo J Role of ultrasound, clinical and scintigraphic parameters to predict malignancy in thyroid nodule .Head & Neck Oncology 2011; (3):17. https://doi.org/10.1186/1758-3284-3-17

6. Maia FRM, Zantut-Wittmann DE, Thyroid nodule management: Clinical, ultrasound and cytopathological parameters for predicting malignancy. Clinics 2012; 67(8): 945-954. https://doi.org/10.6061/clinics/2012(08)15

7. Pathirana AA, Bandara KGMW, Faleel MA, Kuruppumullage SD, Solangarachchi N, Rupesinghe R et al A Sonographic scoring system to assess the risk of thyroid malignancy .Ceylon Medical Journal. 2016;(61):32-34 https://cmj.sljol.info/articles/10.4038/cmj.v61i1.8268

8. Boelaert K, Horacek J , Holder RL, Watkinson J C, Sheppard M C, Franklyn J A Serum Thyrotropin Concentration as a Novel Predictor of Malignancy in Thyroid Nodules Investigated by Fine Needle Aspiration. The Journal of Clinical Endocrinology & Metabolism.2006;91 (11): 4295- 4301 https://doi.org/10.1210/jc.2006-0527

9. Lin-Zheng H, ,Zeng T, Pu L, Pan S, Xia W, Chen L Thyroid Hormones, Autoantibodies, Ultrasonography, and clinical Parameters for Predicting Thyroid Cancer. International Journal of Endocrinology 2016; 2016:8215834 http://dx.doi.org/10.1155/2016/8215834.

10. Hulley, SB, Cummings, SR, Browner, WS, Grady, DG, Newman, TB Designing Clinical Research, 2nd Edition; Philadelphia: Lippincott Williams and Wilkins.2001

11.ROC curve analysis. MedCalc easy to use statistical software . Available at https://www.medcalc.org/manual/roc-curves.php (accessed on20.02.2020).