ORIGINAL ARTICLES

PHILIPPINE JOURNAL OF OTOLARYNGOLOGY-HEAD AND NECK SURGERY VOL. 40 NO. 2 JULY – DECEMBER 2025

Predictive Values of TIRADS and Bethesda Scoring Systems for Thyroid Malignancy at the Quezon City General Hospital: A Review of Records

Jericho T. Agnes, MD
Emmanuel Tadeus S. Cruz, MD

Department of Otorhinolaryngology
Head and Neck Surgery
Quezon City General Hospital

ABSTRACT

Objective: To determine the predictive values of the TIRADS and Bethesda Scoring Systems in diagnosing thyroid malignancy at the Quezon City General Hospital.

Methods:

Design: Retrospective Review of Records

Setting: Tertiary Government Training Hospital

Participants: Records of patients aged 18 years old and above who were admitted and underwent thyroidectomy under the Department of Otorhinolaryngology-Head and Neck Surgery from 2018 to 2023

Results: A total of 47 patient records were included, 16 had thyroid malignancy while 31 had benign histopathology results. The Bethesda system showed 60% sensitivity, 94% specificity, 82% positive predictive value and 83% accuracy while the TIRADS system showed 53% sensitivity, 50% specificity, 33% positive predictive value,and 51% accuracy. Using Chi-Square tests, the Bethesda system had a strong association (p value < .0001), while the TIRADS system had no significant association (p value 1.000) with thyroid malignancy [odds ratios 24.2 and 1.00] respectively.

Conclusion: Our study showed that the Bethesda classification had good specificity, positive predictive value, accuracy and fair sensitivity while the TIRADS system had poor predictive value.

Keywords: *thyroid cancer; thyroid malignancy; fine needle aspiration biopsy; fine needle aspiration cytology; TIRADS; Bethesda; thyroid nodule; thyroidectomy*

Correspondence: Dr. Jericho T. Agnes
Department of Otorhinolaryngology
Head and Neck Surgery
Quezon City General Hospital
Seminary Rd., Proj. 8, Bahay Toro, Quezon City 1106
Philippines
Phone: +63 2 8863 0800 local 407
Email: orlhns.qcgh@gmail.com

The authors declared that this represents original material that is not being considered for publication or has not been published or accepted for publication elsewhere in full or in part, in print or electronic media; that the requirements for authorship have been met by all the authors, and that each author believes that the manuscript represents honest work.

Disclosures: The authors signed a disclosure that there are no financial or other (including personal) relationships, intellectual passion, political or religious beliefs, and institutional affiliations that might lead to a conflict of interest.

Presented at the 31st Quezon City General Hospital Interdepartmental Analytical Research Paper Contest (3rd Place), Feliciano "SB" Belmonte Hall, QCGH, November 29, 2023; and the Philippine Society of Otolaryngology – Head and Neck Surgery Analytical Research Contest via Zoom, October 23, 2024.

Data Availability and Sharing Statement: Datasets are not publicly available because participants did not give written consent for their data to be shared.

Philipp J Otolaryngol Head Neck Surg 2025; 40 (2): 26-29

Patients with thyroid nodules are commonly encountered in the outpatient clinic. The advent of Thyroid Imaging Reporting and Data Systems (TIRADS) and Bethesda scoring systems dealing with thyroid ultrasound and fine needle aspiration biopsy of thyroid nodules respectively somehow simplify diagnosis and help guide clinicians with appropriate management. These diagnostic tools reduce the number of unnecessary thyroid surgeries for patients with benign nodules and identify patients with thyroid cancer requiring surgery.[1] The TIRADS system was developed to provide standards for describing thyroid nodules and avoid variation in reporting.[1] Microcalcifications, hyperechogenicity, wider than tall nodules, absent halo, nodules more than 2 centimeters and a solid composition have an associated risk of thyroid malignancy, with higher risks associated with TIRADS grades 4 and 5.[2-4] Fine needle aspiration cytology also offers accurate diagnosis for thyroid nodules, with the Bethesda System for reporting thyroid cytopathology able to communicate and categorize these nodules.[5] A large study of histologic and clinical correlations by Yang *et al.*, shows that fine needle aspiration cytology has a sensitivity and specificity of 94% and 98.5%, respectively, with Bethesda grades V and VI showing significant association with malignancy.[6]

Based on a search of HERDIN plus, the Western Pacific Region Index Medicus (WPRIM), the Directory of Open Access Journals (DOAJ), MEDLINE (PubMed and Pubmed Central) and Google Scholar, using the search terms "Bethesda," "TIRADS," "FNAB," "thyroid cytology," "thyroid malignancy," and "thyroid ultrasound," we only found a few local studies dealing with the association of specific sonographic features of nodules, FNAB and thyroid malignancy.[7-9] A local study by Trinchera and Cruz found a significant association between sonographic findings of solid, hypoechoic nodules, irregular margins and microcalcifications with thyroid malignancy. Presence of microcalcifications proved to be a significant predictor of malignancy.[8] Another local study by Cambe *et al.* showed that using both TI-RADS and Bethesda for indeterminate thyroid nodules were good predictors for malignancy, with TI-RADS 4 and 5 having 71% and 92% risk, respectively, and Bethesda V having a 95% risk.[9]

Our present study aims to determine the predictive value of the two scoring systems in the diagnostic interpretation of thyroid nodules in our institution compared with other studies in the literature.

METHODS

With Quezon City General Hospital Institutional Ethics Review Board approval (2023.Sept 06.1552.Agnes J.TIRADS-24), this retrospective review of records retrieved and considered for inclusion medical records of patients aged 18 years old and above who were admitted and underwent thyroidectomy in the Department of Otolaryngology-Head and Neck Surgery, Quezon City General Hospital from 2018 to 2023. Excluded were records with no fine needle aspiration cytology report, no ultrasound report, records with frozen section biopsies without FNAB results, and records of patients with neck masses other than thyroid pathology.

Medical records of patients that met inclusion and exclusion criteria were assigned a numerical code, and the following were encoded in IBM SPSS Statistics version 29.0 software (IBM Corp., Armonk, NY, USA): age, sex, diagnosis, TIRADS score, Bethesda category and post-operative surgical histopathology report.

A minimum sample of 115 patient records were needed based on a target sensitivity and specificity of 0.9, concordance of 0.698, 10% margin of error and 95% confidence interval. This was computed using Buderer's formula for diagnostic accuracy.[10]

The thyroid ultrasound results were gathered based on the interpretation given by the board-certified ultrasonologist who signed the report. The ultrasound results were taken at face value and no further scrutiny was made by other radiologists. For multinodular thyroid patients with ultrasound-TIRADs reports, the dominant nodule was considered. If the record did not specify which nodule was biopsied, the nodule with the highest TIRADS score was considered.

Fine needle biopsy results were gathered from the medical records, including the Bethesda category signed by the board-certified pathologist. For repeat fine needle aspiration biopsies, the more recent biopsy was utilized in the study. The most recent FNAB results, ultrasound report and surgical histopathology report were considered for analysis.

Data Analysis

Each TIRADS score and Bethesda category were classified either True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TIRADS scores 1 to 3 and Bethesda categories 1 to 4 were considered as benign pathology. TIRADS scores 4 to 5 and Bethesda categories 5 to 6 were considered as malignant. Surgical histopathologic results were used to confirm benign or malignant pathology. Hence, if a patient had a nodule with a TIRADS score of 4, and a surgical histopathology showing malignancy, it was considered true positive (TP).

Sensitivity, specificity, negative predictive value, positive predictive value, accuracy and odds ratio were computed for the TIRADS score and Bethesda category for detecting malignancy.

IBM SPSS Statistics version 29.0 software (IBM Corp., Armonk, NY, USA) was used for statistical analysis. Two by two table analysis was

made to compute for odds ratios with 95% confidence interval. Chi square tests were used to compute for the significance of the odds ratios.

RESULTS

Out of 53 records initially considered, a total of 47 patient records were finally included in our study. Six patient records with no fine needle aspiration biopsy results were excluded. Three out of these six patient records had frozen section biopsies with no FNAB result. There were 44 (94%) females and three (6%) males. Their ages ranged from 20 to 76 years with a mean age of 42 years. Surgeries included 24 total thyroidectomies, 16 total thyroid lobectomies, and seven subtotal thyroidectomies. Thirty one patients (31;66%) had benign histopathology (23 multiple colloid adenomatous goiter, five follicular adenoma, one oncocytic adenoma, two Hashimoto's thyroiditis) while 16 (34%) had malignant tumors (papillary thyroid carcinoma).

Regarding Bethesda classification, 15 had a Bethesda category of 2, 10 had a category of four, seven had category of 1, six had a category of 5, five had a category of 3 and 4 had a category of 6. Malignancy rate was computed by dividing the number of patients with malignant tumor on post-operative histopathology by the total number of patients. Seven patients (15%) classified under category 1 had a 14% malignancy rate, 15 (32%) under category 2 had a 20% malignancy rate, five (11%) under category 3 had a 20% malignancy rate, 10 (21%) under category 4 had a 40% malignancy rate, six (13%) under category 5 had a 67% malignancy rate, and four (9%) under category 6 had a 100% malignancy rate. It is noteworthy to consider that those patients classified under categories 5 and 6 had higher malignancy rates of 67% and 100%, respectively. Two by two table analysis was used to compute the odds ratio with 95% confidence interval and Chi-Square test for the significance of the odds ratio. The Bethesda category had an odds ratio of 24.2 ($p < .0001$), showing strong association with malignant tumors.

Regarding TIRADS scores, 17 had a TIRADS score of 4, 11 had a score of 3, 8 had a score of 5, 7 had a score of 2 and 4 had a score of 1. Malignancy rate was computed by dividing the number of patients with malignant tumors on post-operative histopathology by the total number of patients. Four patients (9%) had a score of 1 with malignancy rate of 0, 7 (15%) had a score of 2 with a 57% malignancy rate, 11 (23%) had a score of 3 with a 27% malignancy rate, 17 (36%) had a score of 4 with a 29% malignancy rate, and eight (17%) had a score of 5 with a 50% malignancy rate. It is noteworthy to consider that those patients with scores of 4 and 5 had relatively low malignancy rates of 29% and 50%, respectively. Two by two table analysis was used to compute the odds ratio with 95% confidence interval and Chi-Square test for the significance of the odds ratio. The TIRADS system was shown to have an odds ratio of 1.0 ($p = 1.00$) showing no significant association with malignant tumors.

The Bethesda system had 60% sensitivity, 94% specificity, 82% positive predictive value, 83% negative predictive value and 83% accuracy while the TIRADS system had 53% sensitivity, 50% specificity, 33% positive predictive value, 70% negative predictive value and 51% accuracy.

DISCUSSION

Our study showed the Bethesda classification had a good predictive value in diagnosing thyroid malignancy, congruent with the findings of George *et al.* who reported that the Bethesda classification had a sensitivity of 78.72%, specificity of 100% and an accuracy of 80% in predicting malignancy in 144 patients.[11] Another study by Abdelkader *et al.*, showed that the Bethesda system had a sensitivity of 81.8%, specificity 98% and accuracy of 95% in predicting thyroid malignancy in 100 patients.[12]

Although the Bethesda classification had a good predictive value in our study, six cases diagnosed as benign nodules under Bethesda 2 and three turned out to be malignant post-operatively or in other words, were classified as false negatives in our study, which may be reflected by the low sensitivity and high specificity rates. Low sensitivity means that it may not detect thyroid malignancy accurately while high specificity means it can rule out thyroid malignancy reliably. The low sensitivity of the Bethesda system may be attributed to incorrect or inadequate fine needle aspiration technique, errors in specimen preparation, and variability in interpretation of the cytology by different cytopathologists. A thorough review on fine needle aspiration biopsy techniques, preparation, and interpretation are recommended to improve the procedure.

The results of our study showed that the TIRADS score was a poor predictor of thyroid malignancy, with a sensitivity of 53%, specificity of 50% and accuracy rate of 51%. This is relatively lower compared to the study of George *et al.* which showed a sensitivity of 72.3%, specificity 66.7% and accuracy of 75%[11] and another study by Abdelkader *et al.*, where TIRADS had a sensitivity of 76.9%, specificity of 91.3% and accuracy of 75.4 in predicting malignancy in 100 patients.[12] Our study showed TIRADS 4 and 5 categories had malignancy rates of 29% and 50% respectively which were lower in comparison to the study of George *et al.* showing a 97% malignancy rate for TIRADS 5 nodules in 105 patients.[11] In another study by Moifo *et al.*, malignancy rates were 58% and 100% for TIRADS 4 and 5, respectively.[13]

In our study, the TIRADS system had low or fair predictive value, sensitivity, specificity and accuracy which means that the ultrasound scoring system was not able to rule in or rule out thyroid malignancy. This can be attributed to interobserver variability in the interpretation of thyroid nodules by different radiologists, having some ultrasound results come from other institutions, and taking reports at face value with no further review by an independent radiologist.

Our study has several other limitations. First, we were not able to attain the required sample population. However, despite having only 47 samples, the computed achieved power was still acceptable at 82%. Second, our study only covered five years and was limited to our department, and some of the data lacked ultrasound and FNAB results. Third, the FNAB and ultrasound reports and their respective Bethesda and TIRADS scores were interpreted by different pathologists and radiologists respectively which may have affected the grading or rating given in the interpretation. Inherent selection bias in retrospective studies is inevitable and interobserver variability was not accounted for. A study involving more participants is recommended to obtain more accurate results that represent the population under study. The FNAB and ultrasonography results should be reviewed by a panel of pathologists and radiologists respectively and interobserver variability should be computed in future studies.

In conclusion, our study showed that the Bethesda classification had good specificity, positive predictive value, accuracy, and fair sensitivity, while the TIRADS system had poor predictive value.

REFERENCES

1. Grant EG, Tessler FN, Hoang JK, Langer JE, Beland MD, Berland LL, et al. Thyroid ultrasound reporting lexicon: white paper of the ACR Thyroid Imaging, Reporting and Data System (TIRADS) Committee. *J Am Coll Radiol.* 2015 Dec;12(12 Pt A):1272-9. DOI:10.1016/j.jacr.2015.07.011; PubMed PMID: 26419308.
2. Singaporewalla RM, Hwee J, Lang TU, Desai V. Clinico-pathological correlation of thyroid nodule ultrasound and cytology using the TIRADS and Bethesda Classifications. *World J Surg.* 2017 Jul;41(7):1807-1811. DOI:10.1007/s00268-017-3919-5; PubMed PMID: 28251273.
3. Periakaruppan G, Seshadri KG, Vignesh Krishna GM, Mandava R, Sai VPM, Rajendiran S. Correlation between ultrasound-based TIRADS and Bethesda system for reporting thyroid-cytopathology: 2-year experience at a tertiary care center in India. *Indian J Endocrinol Metab.* 2018 Sep-Oct;22(5):651-655. DOI: 10.4103/ijem.IJEM_27_18; PubMed PMID: 30294576; PubMed Central PMCID: PMC6166562.
4. Alshaikh R, Almaghribi K, Alshammari DM, Mohamad H, Ebrahim W, Alshammari SM, et al. Correlation between ultrasound and cytological findings of patients with suspicious thyroid nodules: the King Hamad University Hospital experience. *Cureus.* 2022 Mar 5;14(3):e22877. DOI: 10.7759/cureus.22877; PubMed PMID: 35399395; PubMed Central PMCID: PMC8980678.
5. Cibas ES, Ali SZ; NCI Thyroid FNA state of the science conference. The Bethesda system for reporting thyroid cytopathology. *Am J Clin Pathol.* 2009 Nov;132(5):658-65. DOI: 10.1309/AJCPPHLWMI3JV4LA; PubMed PMID: 19846805.
6. Yang J, Schnadig V, Logrono R, Wasserman PG. Fine-needle aspiration of thyroid nodules: a study of 4703 patients with histologic and clinical correlations. *Cancer.* 2007 Oct 25;111(5):306-15. DOI: 10.1002/cncr.22955; PubMed PMID: 17680588.
7. Sio MC, Uy J, Soriano R. Calcifications in thyroid ultrasonography and thyroid carcinoma. *Philipp J Otolaryngol Head Neck Surg.* 2014 Jul-Dec;29(2):15-18. DOI: https://doi.org/10.32412/pjohns.v29i2.413.
8. Trinchera C, Cruz ET. Association of ultrasonographic findings and thyroid malignancy: A cross-sectional study. *Philipp J Otolaryngol Head Neck Surg.* 2019 Jul-Dec;34(2):24-8. DOI:10.32412/pjohns.v34i2.115.
9. Cambe SM, Arañas JA, Manzala JL, Manlapaz KT. Stratifying indeterminate cytology thyroid nodules by combining thyroid imaging reporting and data systems (TI-RADS) and the Bethesda system for reporting thyroid cytopathology (TBSRTC). *Philipp J Otolaryngol Head Neck Surg.* 2023 Nov 20; 38(2):42. DOI: https://doi.org/10.32412/pjohns.v38i2.1703.
10. Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med.* 1996 Sep;3(9):895-900. DOI: 10.1111/j.1553-2712.1996.tb03538.x; PubMed PMID: 8870764.
11. George NA, Suresh S, Jiji V, Renu S, Thomas S, Janardhan D, et al. Correlation of TIRADS and Bethesda scoring systems with final histopathology of thyroid nodules – an institutional experience. *Indian J Otolaryngol Head Neck Surg.* DOI:10.1007/s12070-021-02380-8; PubMed PMID: 36742706; PubMed Central PMCID: PMC9895457.
12. Abdelkader AM, Zidan AM, Younis MT, Dawa SK. Preoperative evaluation of thyroid nodules: a prospective study comparing the accuracy of ultrasound (TI-RADS) versus the FNAC Bethesda system in relation to the final postoperative histo-pathological diagnosis. *Ann Pathol Lab Med.* 2018 Oct;5(10):A801-A809. DOI: https://doi.org/10.21276/apalm.2110.
13. Moifo B, Takoeta EO, Tambe J, Blanc F, Fotsin JG. Reliability of thyroid imaging reporting and data system (TIRADS) classification in differentiating benign from malignant thyroid nodules. *Open J Radiol.* 2013 Aug;3(3):103-107. DOI:10.4236/ojrad.2013.33016.