Impact of the ultrasonography assessment method on the malignancy risk and diagnostic performance of five risk stratification systems in thyroid nodules

Go Eun Yang (yangke00@hanmail.net)
Kangwon National University School of Medicine https://orcid.org/0000-0002-8689-8127
Dong Gyu Na
Gangneung Asan Hospital https://orcid.org/0000-0001-6422-1652

Research Article

Keywords: thyroid nodule, ultrasonography, retrospective, prospective, risk stratification system

Posted Date: March 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-319662/v1

Version of Record: A version of this preprint was published at Endocrine on September 17th, 2021. See the published version at https://doi.org/10.1007/s12020-021-02795-x.

Abstract

Purpose

Ultrasonographic (US) assessment methods may affect the estimated malignancy risk of thyroid nodules. This study aimed to investigate the impact of retrospective and prospective US assessments on the estimated malignancy risk of US features, classified categories, and diagnostic performance of five risk stratification systems (RSSs) in thyroid nodules.

Methods

A total of 3685 consecutive thyroid nodules (≥ 1 cm) with final diagnoses (retrospective dataset, n = 2180; prospective dataset, n = 1505) were included in this study. We compared the estimated malignancy risk of US features, classified categories, and diagnostic performances of the five common RSSs between retrospective (static US images without cine clips) and prospective datasets of real-time US assessment.

Results

There was no significant difference in the prevalence and histological type of malignant tumours between the two datasets ($p \geq 0.216$). The malignancy risk of solid composition and nonparallel orientation was higher and that of microcalcification was lower in the prospective dataset than in the retrospective dataset ($p < 0.001$, $p = 0.018$, $p = 0.007$, respectively). The retrospective US assessment overestimated the malignancy risk of intermediate-or high-risk nodules according to the RSSs. Prospective US assessment showed lower specificities and higher unnecessary biopsy rates by all RSSs compared to the retrospective US assessment ($p \leq 0.006$, $p \leq 0.045$, respectively).

Conclusions

The overestimated malignancy risk of microcalcification by retrospective US assessment mainly affected the estimated risk of classified categories by RSSs. The retrospective US assessment overestimated the specificities and underestimated the unnecessary biopsy rates by all RSSs.

Introduction

Ultrasonography (US) is an established primary diagnostic tool that has been widely used to predict malignancy risk of thyroid nodules [1]. Many international societies have proposed US risk stratification systems (RSSs) for thyroid nodules [2–7], which have been validated by multiple studies based on either retrospective assessment with or without cine clips [8–12] or prospective US assessment of nodules [13–15]. Although a prospective study using US evaluation of a nodule during the real-time scanning of thyroid gland is ideal for the accurate assessment of an entire nodule, it is not easy to carry out such

research in real clinical practice. Therefore, numerous previous studies assessed US features of thyroid nodules retrospectively using stored static US images.

Careful US assessment will enable accurate prediction of malignancy risk of nodules and provide a reliable basis for the development of US RSS for thyroid nodules. However, to the best of our knowledge, few studies have investigated the impact of retrospective and prospective US assessments on the estimated malignancy risk of US features and diagnostic performance of RSSs in thyroid nodules. A comparison of results between studies according to the method of US assessment (retrospective or prospective) for thyroid nodules may not provide accurate information on the differences resulting from the US assessment method because of uncontrolled confounding factors such as differences in study populations, interpreters, and applied definitions of US lexicons.

The purpose of this study was to investigate the impact of retrospective and prospective US assessment on the estimated malignancy risk of US features. Furthermore, we aimed to evaluate the impact of assessment method on classified categories of nodules by the RSSs and diagnostic performance of biopsy criteria for malignancy by the widely used five US RSSs, including the American Thyroid Association (ATA) system, American Association of Clinical Endocrinologists/American College of Endocrinology/Associazione Medici Endocrinologi (AACE/ACE/AME) system, Korean Thyroid Imaging Reporting and Data System (K-TIRADS), American College of Radiology (ACR) TI-RADS, European (EU)-TIRADS [3–7].

Materials And Methods

This study was approved by the Institutional Review Board, and informed consent was waived because of the retrospective nature of study.

Study population

A total of 3905 consecutive patients underwent US-guided fine needle aspiration (FNA) or core needle biopsy (CNB) for 4832 thyroid nodules ≥ 1 cm between January 2011 and December 2019. Among these, 998 nodules without final diagnoses confirmed by surgical or definite biopsy results (benign or malignant) and 8 nodules with US images of suboptimal quality were excluded. Among the 3826 nodules with final diagnoses by surgical or biopsy results, 35 with isolated macrocalcifications (entirely calcified nodules) [16] and 16 simple cysts were excluded because it was not possible to assess the nodule echogenicity. US images of thyroid nodules obtained between March 2017 and December 2019 before biopsy were prospectively evaluated by two radiologists, and 90 nodules that were prospectively assessed by a radiologist who did not review the US images of the retrospective dataset (January 2011 to February 2017) were excluded from the study population. Therefore, the remaining 2975 patients with 3685 nodules were included in the final cohort (2407 women and 568 men; median age, 56 years; interquartile range [IQR], 47–64 years) (Fig. 1). The data set of this study was obtained from the cohort database at a single institution, and patients and nodules evaluated here were subsets of those

discussed in a previous publication [17]. Final diagnoses of nodules were determined by definite FNA or CNB results (benign or malignant) and surgical histologic diagnoses.

US examination and image analysis

All US examinations were performed using a 5-to 12-MHz linear probe and a real-time US system (IU22 or EPIQ7, Philips Healthcare). All US images of thyroid nodules were obtained by three radiologists between January 2011 and February 2017 before biopsy based on the recommendations of the Korean Society of Thyroid Radiology [5, 18], and the static US images without cine clips were stored in the picture archiving and communication system. The stored static US images were retrospectively reviewed by an experienced radiologist (N.D.G) with 22 years' experience in performing thyroid US and intervention, who had no previous knowledge on the FNA results or final diagnoses (retrospective dataset). The US features of thyroid nodules were prospectively assessed before biopsy and recorded in daily clinical practice by the same radiologist (N.D.G) between March 2017 and December 2019 (prospective dataset). The US features of nodules were strictly assessed using the same US lexicon definitions of the RSSs in both retrospective and prospective datasets [17].

The frequency and malignancy risk of US features determined by the same definition used in US lexicon were compared between the retrospective and prospective datasets. The applied definitions of US lexicon for comparison of the two datasets were as follows: the solid composition was defined as a nodule with no obvious cystic component, and the nodule echogenicity was categorised as hypoechogenicity (marked or mild hypoechogenicity) and iso- or hyperechogenicity, which were defined by the predominant echogenicity in a nodule mixed echogenicity. Microcalcification was defined as a punctate echogenic focus measuring 1 mm or less, with or without posterior acoustic shadowing within the solid component of a nodule. Punctate echogenic foci at the margin or wall of cystic components and that with comet tail artefacts were not defined as microcalcifications. The nonparallel orientation (taller-than-wide shape) was defined as the ratio of the anteroposterior diameter to the transverse diameter of > 1 by visual assessment in the transverse US plane. The presence of any suspicious or high-risk US features, including marked hypoechogenicity, microcalcification (punctate echogenic foci), spiculated/microlobulated (irregular) margin, and nonparallel orientation (taller-than-wide shape), was determined only when the US features were obvious in a nodule. The malignancy risk of US patterns of nodules by combination of composition, echogenicity, and suspicious US features (microcalcification, nonparallel orientation, and spiculated/microlobulated margin) were also compared between retrospective and prospective datasets.

Classification of thyroid nodules and assessment of the diagnostic performance of risk stratification systems for thyroid malignancy

The malignancy risk of classified nodule categories according to the RSSs was compared between retrospective and prospective datasets. Isoechoic nodules with suspicious US features (microcalcification, irregular margin, taller-than-wide shape) were categorised as unclassified nodules in the ATA system. The US feature of extrathyroidal extension (ETE) was not used for the classification of thyroid nodules in this study because of a lack of standardised US criteria for ETE. The diagnostic performance of the biopsy criteria by each RSS was compared between the retrospective and prospective datasets. All nodules were dichotomised into those indicated for biopsy (test positivity) or not indicated (test negativity) according to the biopsy criteria of each RSS. Nodules classified as low-risk by the AACE/ACE/AME system, not suspicious (TR2) or benign (TR1) by the ACR TI-RADS, or benign category by the ATA system, K-TIRADS, and EU-TIRADS were considered not to be indicated for biopsy in this study because they were not routinely indicated for biopsy for diagnostic purposes according to each RSS [3–7]. A reviewer (D.G.N.), who had no previous knowledge of the FNA results or final diagnoses, determined the candidates for biopsy based on the maximal diameter, and classified the risk category of each nodule according to the RSSs.

Statistical Analyses

Continuous variables including age and nodule size were presented using the median (IQR) due to their nonparametric distribution. The Mann-Whitney U test was used to compare age and nodule size between the retrospective and prospective datasets. Categorical variables were reported as frequencies and percentages for each category. The chi-square test or Fisher exact test was used to compare the frequency and malignancy risk of US features, and malignancy risk of classified categories between the retrospective and prospective datasets. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and unnecessary biopsy rate were calculated with 95% confidence intervals. The potentially unnecessary biopsy rate for the diagnosis of thyroid malignancy was defined as the number of benign nodules among biopsy-required nodules in the total nodules. The chi-square test or Fisher's exact test was used to compare the diagnostic values and unnecessary biopsy rates by RSSs between the retrospective and prospective datasets. Statistical analyses were performed using the SPSS ver. 24 for Windows (IBM Corp., Armonk, NY, USA) software. A significant difference was defined as p-value < 0.05.

Results

Demographic data

Demographic data of the patients are summarised in Table 1. There was no difference in the sex (p = 0.919) between the retrospective and prospective datasets, and the median age of patients was slightly higher in the prospective dataset than in the retrospective dataset (57 and 55 years, respectively, p = 0.016). The median size of the nodules was slightly larger in the prospective dataset than in the retrospective dataset (19 mm and 16 mm, respectively; p < 0.001).

Of the 3685 nodules, 3156 (85.6%) were benign and 529 (14.4%) were malignant. Malignant nodules were diagnosed based on histologic diagnoses after surgery (n = 393) or malignant FNA or CNB results (n = 136). Benign nodules were diagnosed based on histologic diagnoses after surgery (n = 377), at least two benign FNA or CNB results (n = 525), and one benign FNA or CNB result (n = 2254). The 529 malignant nodules included 476 (90.0%), 30 (5.8%), 8 (1.5%) anaplastic%), 8 (1.5%), 4 (0.8%), and 3 (0.6%) medullary thyroid carcinomas. There was no significant difference in the prevalence of malignant tumours ($p = 0.216$), papillary thyroid carcinomas ($p = 0.795$), and follicular thyroid carcinomas ($p = 0.077$) between the retrospective and prospective datasets.

Comparison of frequency of US feature between retrospective and prospective datasets

Table 2 shows the frequencies of US features in the retrospective and prospective datasets of thyroid nodules. The frequency of nodules with microcalcifications, macrocalcification, spiculated/microlobulated margin, intracystic echogenic foci with comet tail artifact, and spongiform appearance were significantly higher in the prospective dataset ($p \leq 0.002$). Conversely, the frequency of nodules with solid composition was significantly higher in the retrospective dataset ($p < 0.001$), and there was no difference in nodule echogenicity and orientation between the two datasets ($p = 0.248$ and $p = 0.579$, respectively). There was no difference in the frequency of nodules with a solid hypoechoic US pattern between the retrospective and prospective datasets (27.1% and 25.0%, $p = 0.149$). Although the frequency of microcalcification was higher in the prospective dataset regardless of US pattern based on composition and echogenicity ($p \leq 0.001$), the proportion of partially cystic or iso- and hyperechoic nodules among those with microcalcifications was significantly higher in the prospective dataset than in the retrospective dataset (69.8% and 59.2%, respectively, $p < 0.001$).

Comparison of malignancy risk of US features between retrospective and prospective datasets

Among the various US features, the malignancy risk of solid composition and nonparallel orientation was significantly higher in the prospective dataset ($p < 0.001$ and $p = 0.018$, respectively), and the malignancy risk of microcalcification was significantly lower in the prospective dataset than in the retrospective dataset ($p = 0.007$) (Table 4).

When the US patterns of nodules were compared between the two datasets (Table 3), the malignancy risk of solid hypoechoic nodules was significantly higher in the prospective dataset than in the retrospective dataset (45.7% and 38.1%, respectively, $p = 0.018$); however, the malignancy risk of partially cystic or iso- and hyperechoic nodules was not significantly different between the two datasets ($p = 0.694$). In solid hypoechoic nodules, the malignancy risk of macrocalcification and nonparallel orientation was higher in the prospective dataset than in the retrospective dataset ($p = 0.037$ and $p = 0.046$, respectively). However, the malignancy risk of microcalcification and spiculated/mirolobuated margin was not significantly different between retrospective and prospective datasets ($p = 0.148$ and $p = 0.227$, respectively). There

was no significant difference in the malignancy risk of solid hypoechoic nodules without suspicious US features between the retrospective and prospective datasets (17.0% and 15.3%, $p = 0.321$).

In partially cystic or iso-and hyperechoic nodules, the malignancy risk of microcalcification was significantly lower in the prospective dataset than in the retrospective dataset (9.2% and 15.9%, respectively, $p = 0.008$), and there was no difference in malignancy risk of other US features between the two datasets. The malignancy risk of microcalcification was significantly lower in the prospective dataset than in the retrospective dataset in the subgroup analysis of iso- and hyperechoic nodules (7.3% and 14.3%, respectively, $p = 0.008$) and in the subgroup analysis of partially cystic nodules (7.7% and 14.8%, respectively, $p = 0.017$). However, the malignancy risk of microcalcification was not significantly different between the retrospective and prospective datasets in the hypoechoic nodule subgroups (56.5% and 54.3%, respectively, $p = 0.629$) and solid nodules (49.5% and 48.6%, respectively, $p = 0.822$).

Comparison of malignancy risk of classified nodules by five risk stratification systems between retrospective and prospective datasets

Table 5 shows the malignancy risk of nodules classified by the five RSSs in the retrospective and prospective datasets. In the ATA system, there was no significant difference in the malignancy risk of classified nodules between the two datasets ($p \geq 0.105$). If unclassified nodules were classified as intermediate suspicion nodules [10, 13, 19], the malignancy risk of intermediate suspicion nodules was marginally lower in the prospective dataset than in the retrospective dataset (10.7% and 14.6%, $p = 0.050$). In the K-TIRADS, the malignancy risk of highly suspicious nodules was higher in the prospective dataset than in the retrospective dataset ($p = 0.046$), and that of intermediate suspicion nodules was marginally lower in the prospective dataset than in the retrospective dataset ($p = 0.050$). In the AACE/AME/ACE system, the malignancy risk of high-risk nodules was lower in the prospective dataset than in the retrospective dataset ($p = 0.049$). In the EU-TIRADS, the malignancy risks of high-risk and low-risk nodules were lower in the prospective dataset than in the retrospective dataset ($p = 0.038$ and $p = 0.001$, respectively). Meanwhile, the malignancy risk of intermediate-risk nodules was higher in the prospective dataset than in the retrospective dataset ($p = 0.008$). In the ACR TI-RADS, the malignancy risk of moderately suspicious nodules was lower in the prospective dataset than in the retrospective dataset ($p = 0.003$). The calculated malignancy risks of classified nodules were within or near the range of suggested malignancy risk of classified categories by each RSS in both retrospective and prospective datasets.

Comparison of diagnostic performance of biopsy criteria by five risk stratification systems between retrospective and prospective datasets

Table 6 shows the diagnostic performance of the biopsy criteria by the five RSSs in retrospective and prospective datasets. The sensitivity of biopsy criteria by the EU-TIRADS was higher in the prospective dataset than in the retrospective dataset ($p = 0.009$), and other RSSs also showed slightly higher sensitivity in the prospective dataset, but this difference was statistically insignificant ($p \geq 0.095$). Meanwhile, the specificities of all RSSs were significantly lower in the prospective dataset than in the retrospective dataset ($p \leq 0.006$). The unnecessary biopsy rates of benign nodules by all RSSs were significantly higher in the prospective dataset than in the retrospective dataset ($p \leq 0.045$), and the increased rate of unnecessary biopsy ranged from 6.3–14.2% according to the RSSs when the unclassified nodules were categorised as intermediate suspicion nodules by the ATA system. There were no significant differences in the PPV and NPV of all RSSs between the two datasets.

Discussion

Our study has demonstrated that retrospective assessment of static US images of thyroid nodules underestimated the malignancy risk of solid composition and nonparallel orientation and it overestimated the malignancy risk of microcalcification compared to the prospective US assessment. The retrospective US assessment overestimated the malignancy risk of intermediate or moderately suspicious nodules categorised by the ATA system (reclassification of unclassified nodules), K-TIRADS, and ACR TI-RADS, and overestimated the malignancy risk of high-risk nodules categorised by the AACE/AME/ACE and EU-TIRADS. Prospective US assessment had lower specificity and higher unnecessary biopsy rate by all RSSs compared to the retrospective US assessment.

The retrospective dataset without cine clips has intrinsic limitations that complete accurate assessment of US features of a nodule may not be possible because only representative static US images were obtained for clinical purposes even though the US images of nodules were collected and stored for depiction of US characteristics of nodules according to the standardised guidelines. In addition, some small US features such as minimal cystic change or microcalcification might have been overlooked by the operators, which may explain the higher frequency of solid composition and lower frequency of microcalcification in the retrospective dataset, even though the potential differences in intrinsic nodule characteristics might have existed between the two datasets in this study. If cine clips of all nodules were available, retrospective US assessment of nodules using them could minimise the limitations of the retrospective dataset compared to the prospective dataset in which US features of nodules were prospectively interpreted during real-time US imaging.

These limitations of retrospective datasets may be a major cause of the differences in the malignancy risk of solid composition and microcalcification between the retrospective and prospective datasets. Some solid nodules determined by the retrospective assessment might have included nodules with minimal cystic changes, which may have resulted in a relatively lower malignancy risk because nodules with minimal cystic changes have a lower malignancy risk compared to purely solid nodules without minimal cystic changes [20]. The lower malignancy risk of microcalcification in the prospective dataset may be explained by the higher proportion of partially cystic or iso- and hyperechoic nodules among all

nodules with microcalcifications in the prospective dataset because the malignancy risk of partially cystic or iso- and hyperechoic nodules with microcalcifications is significantly lower than that of solid hypoechoic nodules [8, 21]. Our subgroup analysis showed that the lower malignancy risk of microcalcifications in the prospective dataset was found only in partially cystic or iso- and hyperechoic nodules compared to the retrospective dataset, but it was not found in solid hypoechoic nodules. This finding suggests that prospective US assessment detected a higher number of benign partially cystic or iso- and hyperechoic nodules with microcalcifications than the retrospective US assessment did.

Our study showed that the malignancy risk of intermediate suspicion or moderately suspicious nodules by the ATA system with categorizing unclassified nodules into intermediate suspicion category, K-TIRADS, and ACR TI-RADS, and the malignancy risk of high-risk category nodules by the AACE/AME/ACE system and EU-TIRADS were significantly lower in the prospective dataset than in the retrospective dataset. This finding may be mainly explained by the lower malignancy risk of partially cystic or iso- and hyperechoic nodules with microcalcifications in the prospective dataset compared to the retrospective dataset because other suspicious US features of non-parallel orientation and spiculated/microlobulated margins did not decrease the malignancy risk of nodules in the prospective dataset. Furthermore, there was no difference in malignancy risk in solid hypoechoic nodules without suspicious US features between the two datasets. Although the point-based ACR TI-RADS is not directly correlated with US patterns of nodules, most moderately suspicious nodules classified by the ACR TI-RADS correspond to the intermediate suspicion nodules classified by the ATA system with categorizing unclassified nodules into intermediate suspicion category and K-TIRADS [22]. The lower malignancy risk of these category nodules by the RSSs may also explain the lower specificity of biopsy criteria and higher unnecessary biopsy rate of the five RSSs in the prospective dataset than in the retrospective dataset. Therefore, the diagnostic performance of the biopsy criteria by the five RSSs estimated from the retrospective dataset may overestimate the specificity and underestimate the unnecessary biopsy rate.

The estimated malignancy risk of US features of thyroid nodules in cohort studies may be influenced by many factors, including the characteristics and disease spectrum of the study population, prevalence and histological type of malignant tumours, applied definitions of US lexicon, interobserver variability of interpreters, US assessment method (retrospective or prospective), and US machines. In the present study, we believe that these confounding factors were minimised because there was no significant difference in the prevalence and histological type of malignant tumours between the retrospective and prospective datasets, and the same interpreter assessed US features using the same definitions of US lexicon and the same high-resolution US machines in one institution. Although the median nodule size was larger in the prospective dataset, the difference in nodule size should not have affected the main result of this study suggesting that the estimated malignancy risk of microcalcification was lower in the prospective dataset than in the retrospective dataset because increased nodule size was not significantly associated with malignancy risk in overall nodules [23, 24] and may be associated with a higher malignancy risk in nodules with intermediate or low suspicion US patterns [25].

Our study has several limitations. First, we could not evaluate the retrospective US assessment of nodules using cine clips because they were not routinely obtained in clinical practice. Second, there may have been a selection bias because we have excluded some patients without a final diagnosis and included only thyroid nodules in which US-guided biopsy was performed. Third, the final diagnosis was based on the results of biopsy and histology of surgical specimens, which has an inherent risk of rare false-negative or false-positive results. Fourth, only one experienced radiologist interpreted the US features of the thyroid nodules at one institution. Further multicentre studies are required to verify the reproducibility of our results.

In conclusion, the retrospective assessment of static US images of thyroid nodules overestimated the malignancy risk of microcalcification in partially cystic or iso-and hyperechoic nodules compared to prospective US assessment. The difference in the estimated malignancy risk of microcalcification mainly affected the estimated risk of classified categories by the five common RSSs. Furthermore, retrospective US assessment overestimated specificities and it underestimated unnecessary biopsy rates by all RSSs compared to the prospective US assessment. A prospective US assessment will be necessary to determine the accurate diagnostic performance of RSSs to estimate risk of malignancy in thyroid nodules.

Abbreviations

AACE/ACE/AME
American Association of Clinical Endocrinologists/American College of Endocrinology/Associazione Medici Endocrinologi
ACR
American College of Radiology
ATA
American Thyroid Association
EU-TIRADS
European Thyroid Imaging Reporting and and Data system
K-TIRADS
Korean Thyroid Imaging Reporting and and Data system
US
ultrasonography,

Declarations

Acknowledgements

This research was supported by the Medical Research Promotion Program through the Gangneung Asan Hospital, funded by the Asan Foundation (2020IC001).

Disclosures and declarations

Funding

This research was supported by the Medical Research Promotion Program through the Gangneung Asan Hospital, funded by the Asan Foundation (2020IC001).

Conflict of Interest

The authors have no relevant financial or non-financial interests to disclose.

Ethics approval

This retrospective study was approved by our institutional review board.

Consent to participate/ Consent to publish

The requirement for patient informed consent was waived.

Authors' contributions

Conceptualization: Dong Gyu Na; Methodology: Go Eun Yang, Dong Gyu Na; Formal analysis and investigation: Go Eun Yang, Dong Gyu Na; Writing - original draft preparation: Go Eun Yang; Writing - review and editing: Dong Gyu Na; Funding acquisition: Dong Gyu Na.

References

1. E.J. Ha, H.K. Lim, J.H. Yoon, J.H. Baek, K.H. Do, M. Choi et al., Primary imaging test and appropriate biopsy methods for thyroid nodules: guidelines by Korean Society of Radiology and Nat ional Evidence-Based Healthcare Collaborating Agency. Korean J. Radiol. **19**(4), 623–631 (2018). https://doi.org/10.3348/kjr.2018.19.4.623
2. P. Perros, K. Boelaert, S. Colley, C. Evans, R.M. Evans, G. Gerrard Ba et al., Guidelines for the management of thyroid cancer. Clin. Endocrinol. **81**, 1–122 (2014). https://doi.org/10.1111/cen.12515
3. B.R. Haugen, E.K. Alexander, K.C. Bible, G.M. Doherty, S.J. Mandel, Y.E. Nikiforov et al., 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer. Thyroid **26**(1), 1–133 (2016). https://doi.org/10.1089/thy.2015.0020
4. H. Gharib, E. Papini, J.R. Garber, D.S. Duick, R.M. Harrell, L. Hegedus et al., American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid

Nodules-2016 Update Appendix. Endocrine practice **22**, 622–623 (2016). https://doi.org/10.4158/EP161208.GL

5. J.H. Shin, J.H. Baek, J. Chung, E.J. Ha, J. Kim, Y.H. Lee et al., Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of Thyroid Radiology consensus statement and recommendations. Korean J. Radiol. **17**(3), 370–395 (2016). https://doi.org/10.3348/kjr.2016.17.3.370
6. F.N. Tessler, W.D. Middleton, E.G. Grant, J.K. Hoang, L.L. Berland, S.A. Teefey et al., ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee. Journal of the American college of radiology **14**, 587–595 (2017). https://doi.org/10.1016/j.jacr.2017.01.046
7. G. Russ, S.J. Bonnema, M.F. Erdogan, C. Durante, R. Ngu, L. Leenhardt, European Thyroid Association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. European thyroid journal **6**, 225–237 (2017). https://doi.org/10.1159/000478927
8. D.G. Na, J.H. Baek, J.Y. Sung, J.-H. Kim, J.K. Kim, Y.J. Choi, H. Seo, Thyroid imaging reporting and data system risk stratification of thyroid nodules: categorization based on solidity and echogenicity. Thyroid **26**(4), 562–572 (2016). https://doi.org/10.1089/thy.2015.0460
9. W.D. Middleton, S.A. Teefey, C.C. Reading, J.E. Langer, M.D. Beland, M.M. Szabunio et al., Comparison of performance characteristics of american college of radiology TI-RADS, Korean Society of thyroid radiology TIRADS, and American Thyroid Association guidelines. Am. J. Roentgenol. **210**, 1148–1154 (2018). https://doi.org/10.2214/AJR.17.18822
10. E.J. Ha, D.G. Na, J.H. Baek, J.Y. Sung, J.h. Kim, S.Y. Kang, US fine-needle aspiration biopsy for thyroid malignancy: diagnostic performance of seven society guidelines applied to 2000 thyroid nodules. Radiology **287**(3), 893–900 (2018). https://doi.org/10.1148/radiol.2018171074
11. S.J. Yoon, D.G. Na, H.Y. Gwon, W. Paik, W.J. Kim, J.S. Song, M.S. Shim, Similarities and differences between thyroid imaging reporting and data systems. Am. J. Roentgenol. **213**, W76–W84 (2019). https://doi.org/10.2214/AJR.18.20510
12. P. Trimboli, R. Ngu, B. Royer, L. Giovanella, C. Bigorgne, R. Simo et al., A multicentre validation study for the EU-TIRADS using histological diagnosis as a gold standard. Clin. Endocrinol. **91**, 340–347 (2019). https://doi.org/10.1111/cen.13997
13. E.J. Ha, W.-J. Moon, D.G. Na, Y.H. Lee, N. Choi, S.J. Kim, J.K. Kim, A multicenter prospective validation study for the Korean thyroid imaging reporting and data system in patients with thyroid nodules. Korean J. Radiol. **17**(5), 811–821 (2016). https://doi.org/10.3348/kjr.2016.17.5.811
14. A. Persichetti, E. Di Stasio, R. Guglielmi, G. Bizzarri, S. Taccogna, I. Misischi et al., Predictive value of malignancy of thyroid nodule ultrasound classification systems: a prospective study. The Journal of Clinical Endocrinology & Metabolism **103**(4), 1359–1368 (2018). https://doi.org/10.1210/jc.2017-01708
15. G. Grani, L. Lamartina, V. Ascoli, D. Bosco, M. Biffoni, L. Giacomelli et al., Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: toward the “right” TIRADS. The

Journal of Clinical Endocrinology & Metabolism **104**, 95–102 (2019). https://doi.org/10.1210/jc.2018-01674

16. H.Y. Gwon, D.G. Na, B.-J. Noh, W. Paik, S.J. Yoon, S.J. Choi, D.R. Shin, Thyroid nodules with isolated macrocalcifications: malignancy risk of isolated macrocalcifications and postoperative risk stratification of malignant tumors manifesting as isolated macrocalcifications. Korean J. Radiol. **21**(5), 605–613 (2020). https://doi.org/10.3348/kjr.2019.0523
17. D.G. Na, W. Paik, J. Cha, H.Y. Gwon, S.Y. Kim, R.-E. Yoo, Diagnostic performance of modified Korean Thyroid Imaging Reporting and Data System for thyroid malignancy according to nodule size: comparison with five society guidelines, Ultrasonography (2020). https://doi.org/10.14366/usg.20148
18. W.J. Moon, J.H. Baek, S.L. Jung, D.W. Kim, E.K. Kim, J.Y. Kim et al., Ultrasonography and the ultrasound-based management of thyroid nodules: consensus statement and recommendations. Korean J. Radiol. **12**(1), 1–14 (2011). https://doi.org/10.3348/kjr.2011.12.1.1
19. J.H. Yoon, H.S. Lee, E.-K. Kim, H.J. Moon, J.Y. Kwak, Malignancy risk stratification of thyroid nodules: comparison between the thyroid imaging reporting and data system and the 2014 American Thyroid Association management guidelines. Radiology **278**(3), 917–924 (2016). https://doi.org/10.1148/radiol.2015150056
20. D.G. Na, J. Kim, D.S. Kim, S.J. Kim, Thyroid nodules with minimal cystic changes have a low risk of malignancy. Ultrasonography **35**(2), 153–158 (2016). https://doi.org/10.14366/usg.15070
21. Y.M. Sohn, D.G. Na, W. Paik, H.Y. Gwon, B.J. Noh, Malignancy risk of thyroid nodules with nonshadowing echogenic foci. Ultrasonography **40**, 115–125 (2021). https://doi.org/10.14366/usg.20012
22. Y. Yim, D.G. Na, E.J. Ha, J.H. Baek, J.Y. Sung, J. Kim, W.-J. Moon, Concordance of three international guidelines for thyroid nodules classified by ultrasonography and diagnostic performance of biopsy criteria. Korean J. Radiol. **21**(1), 108–116 (2020). https://doi.org/10.3348/kjr.2019.0215
23. C.R. McHenry, E.S. Huh, R.N. Machekano, Is nodule size an independent predictor of thyroid malignancy? Surgery **144**(6), 1062–1069 (2008). https://doi.org/10.1016/j.surg.2008.07.021
24. A. Cavallo, D.N. Johnson, M.G. White, S. Siddiqui, T. Antic, M. Mathew et al., Thyroid nodule size at ultrasound as a predictor of malignancy and final pathologic size. Thyroid **27**(5), 641–650 (2017). https://doi.org/10.1089/thy.2016.0336
25. M.J. Hong, D.G. Na, J.H. Baek, J.Y. Sung, J.-H. Kim, Impact of nodule size on malignancy risk differs according to the ultrasonography pattern of thyroid nodules. Korean J. Radiol. **19**(3), 534–541 (2018). https://doi.org/10.3348/kjr.2018.19.3.534

Tables

Table 1. Demographic patient data of retrospective and prospective datasets

Parameter	Retrospective dataset	Prospective dataset	*p*
No. of patients	1749	1226	
No. of women	1414 (80.8)	993(81.0)	0.919
Age (years, median [IQR])	55 (47-64)	57 (48-65)	0.016
No. of nodules	2180	1505	
Maximal nodule size (mm, median [IQR])	16 (12-24)	19 (14-29)	<0.001
No. of malignant tumors	300 (13.8)	229 (15.2)	0.216
Papillary thyroid carcinomas	279 (12.8)	197 (13.1)	0.795
Follicular thyroid carcinomas	13 (0.6)	17 (1.1)	0.077

Unless otherwise indicated, data are numbers with percentages in parentheses for categorical variables and ranges in parentheses for continuous variables
IQR interquartile range

Table 2. Frequency of US features in retrospective and prospective datasets

US features	Retrospective Dataset (n = 2180)		Prospective Dataset (n = 1505)		*p*
	No. of nodules	Frequency (%)	No. of nodules	Frequency (%)	
Composition					
Solid	1335	61.2	682	45.3	<0.001
Partially cystic	845	38.8	823	54.7	
Echogenicity					
Hypoechoic	732	33.6	533	35.4	0.248
Iso- or hyperechoic	1448	66.4	972	64.6	
Calcification					
Microcalcification	478	21.9	547	36.3	<0.001
Macrocalcification	270	12.4	240	15.9	0.002
Rim calcification	40	1.8	42	2.8	0.053
Orientation (shape)					
Nonparallel orientation (taller-than-wide)	182	8.3	118	7.8	0.579
Parallel orientation	1998	91.7	1387	92.2	
Margin					
Spiculated/microlobulated (irregular)	103	4.7	108	7.2	0.002
Smooth or ill-defined	2077	95.3	1397	92.8	
Intracystic echogenic foci with comet tail artifact	15	0.7	57	3.8	<0.001
Spongiform	13	0.6	27	1.8	0.001

US ultrasonography

Table 3. Comparison of malignancy risk of US features between retrospective and prospective datasets

US features	Retrospective Dataset (n = 2180)			Prospective Dataset (n = 1505)			*p*	Malignancy risk difference (%)[a]
	Benign (n = 1880)	Malignancy (n = 300)	Malignancy risk (%)	Benign (n =1276)	Malignancy (n = 229)	Malignancy risk (%)		
Composition								
Solid	1074	261	19.6 (17.4, 21.7)	486	196	28.7 (25.3, 32.1)	< 0.001	9.1
Partially cystic	806	39	4.6 (3.2, 6.0)	790	33	4.0 (2.7, 5.4)	0.543	
Echogenicity								
Hypoechoic	492	240	32.8 (29.4, 36.2)	346	187	35.1 (31.0, 39.1)	0.394	
Isoechoic or hyperechoic	1388	60.0	4.1 (3.1, 5.2)	930	42	4.3 (3.0, 5.6)	0.831	
Calcification								
Microcalcification	305	173	36.2 (31.9, 40.5)	392	155	28.3 (24.6, 32.1)	0.007	-7.9
Macrocalcification	175	95	35.2 (29.5, 40.9)	152	88	36.7 (30.6, 42.8)	0.728	
Rim calcification	34	6.0	15.0 (3.9, 26.1)	34	8	19.0 (7.2, 30.9)	0.626	
Orientation								
Nonparallel orientation (taller-than-wide)	101	81	44.5 (37.3, 51.7)	49	69	58.5 (49.6, 67.4)	0.018	14
Parallel orientation	1779	219.0	11.0 (9.6, 12.3)	1227	160	11.5 (9.9, 13.2)	0.602	
Margin								
Spiculated or microlobulated	25	78	75.7 (67.4, 84.0)	20	88	81.5 (74.2, 88.8)	0.308	
Smooth or ill-defined	1855	222.0	10.7 (9.4, 12.0)	1256	141	10.1 (8.5, 11.7)	0.574	
Intracystic echogenic foci with comet tail artifact	15	0	0.0	55	2	3.5 (1.3, 8.3)	0.462	
Spongiform	13	0	0.0	27	0	0.0	N/A	

Data in parenthesis are 95% confidence intervals
US ultrasonography
[a]Malignancy risk in prospective dataset minus malignancy risk in retrospective dataset

Table 4. Comparison of malignancy risk of US features according to composition and echogenicity between retrospective and prospective datasets

US features	Retrospective Dataset (n = 2180)			Prospective Dataset (n = 1505)			*p*	Malignancy risk difference (%)[a]
	Benign (n = 1880)	Malignancy (n = 300)	Malignancy risk (%)	Benign (n =1276)	Malignancy (n = 229)	Malignancy risk (%)		
Solid and hypoechoic	366	225	38.1 (34.2, 42.0)	204	172	45.7 (40.7, 50.8)	0.018	7.6
Microcalcification	67	128	65.6 (59.0, 72.3)	45	120	72.7 (65.9, 79.5)	0.148	
Macrocalcification	46	81	63.8 (55.4, 72.1)	22	73	76.8 (68.4, 85.3)	0.037	13
Rim calcification	20	4	16.7 (1.8, 31.6)	11	2	15.4 (-4.2, 35.0)	0.920	
Nonparallel orientation (taller than wide)	33	68	67.3 (58.2, 76.5)	13	56	81.2 (71.9, 90.4)	0.046	13.9
Spiculated/microlobulated	20	75	78.9 (70.7, 87.1)	13	78	85.7 (78.5, 92.9)	0.227	
Partially cystic or iso- and hyperechoic	1514	75	4.7 (3.7, 5.8)	1072	57	5.0 (3.8, 6.3)	0.694	
Microcalcification	238	45	15.9 (11.6, 20.2)	347	35	9.2 (6.3, 12.1)	0.008	-6.7
Macrocalcification	129	14	9.8 (4.9, 14.7)	130	15	10.3 (5.4, 15.3)	0.876	
Rim calcification	14	2	12.5 (-3.7, 28.7)	23	6	20.7 (5.9, 35.4)	0.492	
Nonparallel orientation (taller than wide)	68	13	16.0 (8.1, 24.0)	36	13	26.5 (14.2, 38.9)	0.304	
Spiculated/microlobulated	5	3	37.5 (4.0, 71.0)	7	10	58.8 (35.4, 82.2)	0.319	

Data in parentheses are 95% confidence intervals

[a]Malignancy risk in prospective dataset minus malignancy risk in retrospective dataset

Table 5. Comparison of malignancy risk of classified nodules between retrospective and prospective datasets by five risk stratification systems

US risk stratification system	Suggested malignancy risk (%)	Retrospective Dataset (n = 2180)			Prospective Dataset (n = 1505)			*p*	Malignancy risk difference (%)[a]
		Benign (n = 1880)	Malignancy (n = 300)	Malignancy risk (%)	Benign (n =1276)	Malignancy (n = 229)	Malignancy risk (%)		
ATA									
High suspicion	> 70-90	149	184	55.3 (49.9, 60.6)	134	159	54.3 (48.6, 60.0)	0.804	
Intermediate suspicion	10-20	258	53	17.0 (12.9, 21.2)	140	26	15.7 (10.1, 21.2)	0.700	
Low suspicion	5-10	657	17	2.5 (1.3, 3.7)	254	5	1.9 (0.3, 3.6)	0.594	
Very low suspicion	< 3	561	11	1.9 (0.8, 3.0)	444	12	2.6 (1.2, 4.1)	0.445	
Benign	< 1	-	-	-	-	-	-	-	
Unclassified	-	255	35	12.1 (8.3, 15.8)	304	27	8.2 (5.2, 11.1)	0.105	
K-TIRADS									
High suspicion	> 60	105	174	62.4 (56.7, 68.1)	60	147	71.0 (64.8, 77.2)	0.046	8.6
Intermediate suspicion	15-50	561	98	14.9 (12.2, 17.6)	520	65	11.1 (8.6, 13.7)	0.050	-3.8
Low suspicion	3-15	1190	28	2.3 (1.5, 3.1)	642	15	2.3 (1.1, 3.4)	0.983	
Benign	< 3	24	0	0	54	2	3.6 (-1.3, 8.4)	0.348	
AACE/AME/ACE									
High-risk	50-90	409	225	35.5 (31.8, 39.2)	433	188	30.3 (26.7, 33.9)	0.049	-5.2
Intermediate-risk	5–15	1446	75	4.9 (3.8, 6.0)	788	39	4.7 (3.3, 6.2)	0.817	
Low-risk	1	25	0	0.0	55	2	3.5 (-1.3, 8.3)	0.343	
EU-TIRADS									
High risk	26-87	416	227	35.3 (31.6, 39.0)	444	189	29.9 (26.3, 33.4)	0.038	-5.4
Intermediate risk	6-27	684	35	4.9 (3.3, 6.4)	347	34	8.9 (6.1, 11.8)	0.008	4
Low risk	2-4	774	38	4.7 (3.2, 6.1)	483	6	1.2 (0.3, 2.2)	0.001	-3.5
Benign	0	6	0	0.0	2	0	0.0	N/A	
ACR									
Highly suspicious	> 20	151	177	54 (48.6, 59.4)	117	164	58.4 (52.6, 64.1)	0.276	
Moderately suspicious	5-20	569	99	14.8 (12.1, 17.5)	479	48	9.1 (6.7, 11.6)	0.003	-5.7
Mildly suspicious	5	636	15	2.3 (1.2, 3.5)	249	7	2.7 (0.7, 4.7)	0.705	
Not suspicious	< 2	509	9	1.7 (0.6, 2.9)	404	10	2.4 (0.9, 3.9)	0.467	
Benign	< 2	15	0	0.0	27	0	0.0	N/A	

Data in parentheses are 95% confidence intervals

[a]Malignancy risk in prospective dataset minus malignancy risk in retrospective dataset

AACE/ACE/AME American Association of Clinical Endocrinologists/American College of Endocrinology/Associazione Medici Endocrinologi Medical Guidelines, *ACR TI-RADS* American College of Radiology Thyroid Imaging Reporting and Data System, *ATA* American Thyroid Association Management Guideline, *EU-TIRADS* European Thyroid Imaging Reporting and Data System, *K-TIRADS* Korean Thyroid Imaging Reporting and Data System

Table 6. Comparison of diagnostic performance of biopsy criteria by five risk stratification systems between retrospective and prospective datasets

US risk stratification system	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)	Unnecessary biopsy rate[a] (%)
ATA[b]					
Retrospective dataset	83.7 (251/300) [79.5, 87.8]	42.7 (803/1880) [40.5, 44.9]	18.9 (251/1328) [16.8, 21.0]	94.2 (803/852) [92.7, 95.8]	49.4 (1077/2180) [47.3, 51.5]
Prospective dataset	86.9 (199/229) [82.5, 91.3]	37.8 (482/1276) [35.1, 40.4]	20.0 (199/993) [17.6, 22.5]	94.1 (482/512) [92.1, 96.2]	52.8 (794/1505) [50.2, 55.3]
P-value	0.301	0.006	0.492	0.934	0.045
K-TIRADS					
Retrospective dataset	95.7 (287/300) [93.4, 98.0]	23.9 (449/1880) [22.0, 25.8]	16.7 (287/1718) [14.9, 18.5]	97.2 (449/462) [95.7, 98.7]	65.6 (1431/2180) [63.6, 67.6]
Prospective dataset	98.3 (225/229) [96.6, 100.0]	10.4 (133/1276) [8.7, 12.1]	16.4 (225/1368) [14.5, 18.4]	97.1 (133/137) [94.3, 99.9]	75.9 (1143/1505) [73.8, 78.1]
P-value	0.095	< 0.001	0.848	0.948	< 0.001
AACE/AME/ACE					
Retrospective dataset	84.3 (253/300) [80.2, 88.4]	48.2 (906/1880) [45.9, 50.5]	20.6 (253/1227) [18.4, 22.9]	95.1 (906/953) [93.7, 96.4]	44.7 (974/2180) [42.6, 46.8]
Prospective dataset	89.1 (204/229) [85.0, 93.1]	32.1 (410/1276) [29.6, 34.7]	19.1 (204/1070) [16.7, 21.4]	94.3 (410/435) [92.1, 96.4]	57.5 (866/1505) [55.0, 60.0]
P-value	0.114	< 0.001	0.352	0.525	< 0.001
EU-TIRADS					
Retrospective dataset	87.0 (261/300) [83.2, 90.8]	39.6 (745/1880) [37.4, 41.8]	18.7 (261/1396) [16.7, 20.7]	95.0 (745/784) [93.5, 96.5]	52.1 (1135/2180) [50.0, 54.2]
Prospective dataset	93.9 (215/229) [90.8, 97.0]	21.9 (279/1276) [19.6, 24.1]	17.7 (215/1212) [15.6, 19.9]	95.2 (279/293) [92.8, 97.7]	66.2 (997/1505) [63.9, 68.6]
P-value	0.009	< 0.001	0.528	0.895	< 0.001
ACR					
Retrospective dataset	78.0 (234/300) [73.3, 82.7]	68.1 (1281/1880) [66.0, 70.2]	28.1 (234/833) [25.0, 31.1]	95.1 (1281/1347) [93.9, 96.3]	27.5 (599/2180) [25.6, 29.4]
Prospective dataset	83.4 (191/229) [78.6, 88.2]	60.2 (768/1276) [57.5, 62.9]	27.3 (191/699) [24.0, 30.6]	95.3 (768/806) [93.8, 96.7]	33.8 (508/1505) [31.4, 36.1]
P-value	0.121	< 0.001	0.739	0.846	< 0.001

Data in parentheses are the raw data used to calculate the percentages, and data in square brackets are 95% confidence intervals

[a]Unnecessary biopsy rate = number of benign nodules among biopsy-required nodules / total nodules.

[b]Unclassified nodules were categorized as nodules not indicated for biopsy. If unclassified nodules were classified as intermediate suspicion nodules, the unnecessary biopsy rate was significantly higher in the prospective dataset compared to the retrospective dataset (73.0% and 61.1%, $p < 0.001$)

All acronyms of US risk stratification systems are as given in Table 4.

Figures

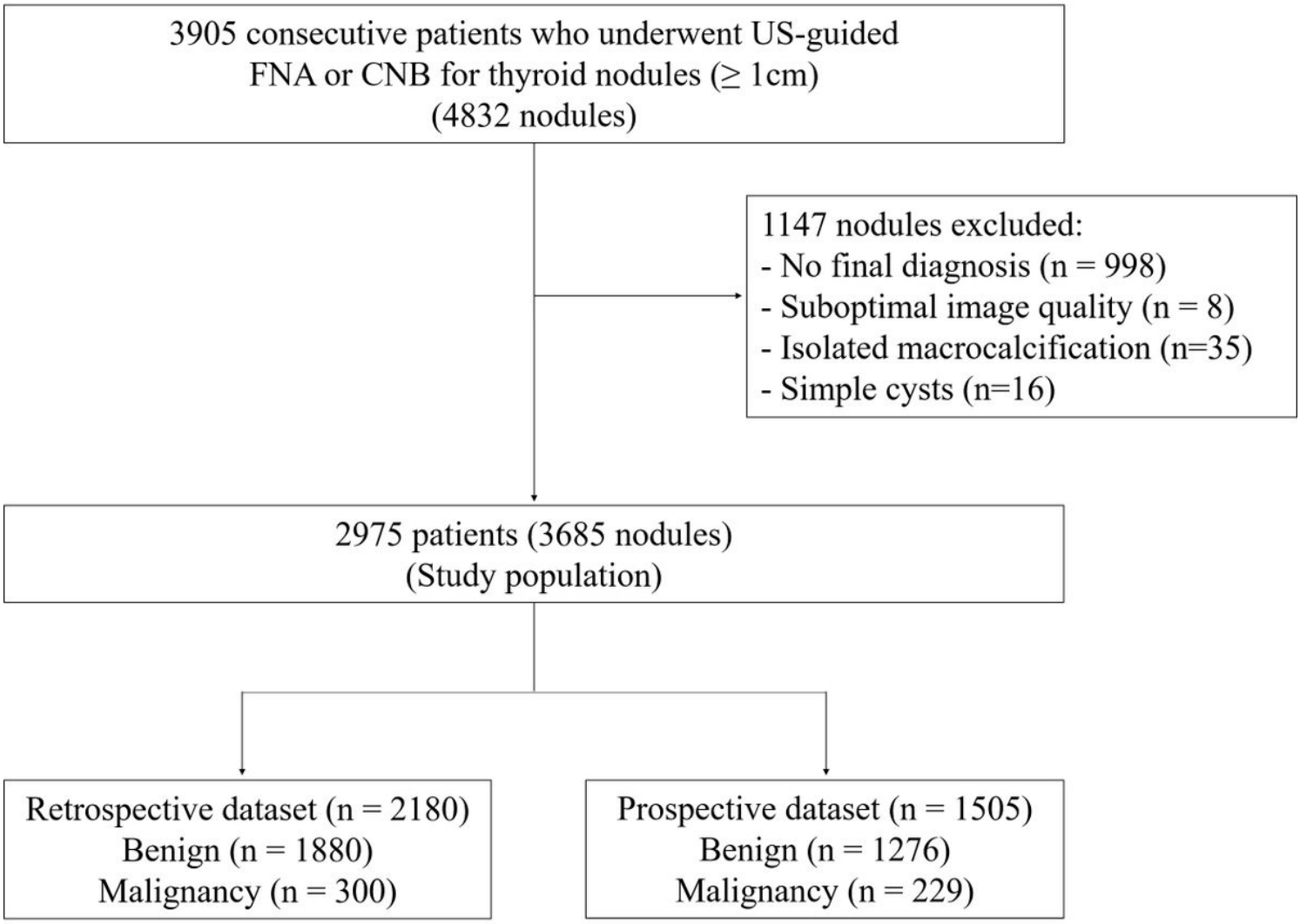

Figure 1

Flowchart of study participants. FNA fine-needle aspiration, CNB core needle biopsy.