The ultrasound risk stratification systems for thyroid nodule have been evaluated against papillary carcinoma. A meta-analysis

Pierpaolo Trimboli[1,2] · Marco Castellana[3] · Arnoldo Piccardo[4] · Francesco Romanelli[5] · Giorgio Grani[6] · Luca Giovanella[1,7] · Cosimo Durante[6]

Abstract

Thyroid imaging reporting and data systems (TIRADS) are used to stratify the malignancy risk of thyroid nodule by ultrasound (US) examination. We conducted a meta-analysis to evaluate the pooled cancer prevalence and the relative prevalence of papillary, medullary, follicular thyroid cancer (PTC, MTC, and FTC) and other malignancies among nodules included in studies evaluating their performance. Four databases were searched until February 2020. Original articles with at least 1000 nodules, evaluating the performance of at least one TIRADS among AACE/ACE/AME, ACR-TIRADS, ATA, EU-TIRADS, or K-TIRADS, and reporting data on the histological diagnosis of malignant lesions were included. The number of malignant nodules, PTC, FTC, MTC and other malignancies in each study was extracted. For statistical pooling of data, a random-effects model was used. Nine studies were included, evaluating 19,494 thyroid nodules. The overall prevalence of malignancy was 34% (95%CI 21 to 49). Among 6162 histologically proven malignancies, the prevalence of PTC, FTC, MTC and other malignancies was 95%, 2%, 1%, and 1%, respectively. A high heterogeneity was found for all the outcomes. A limited number of studies generally conducted using a retrospective design was found, with possible selection bias. Acknowledging this limitation, TIRADSs should be regarded as accurate tools to diagnose PTC only. Proposed patterns and/or cut-offs should be revised and other strategies considered to improve their performance in the assessment of FTC, MTC and other malignancies.

Keywords Thyroid · Ultrasound · TIRADS · Papillary · Follicular · Medullary

---

Pierpaolo Trimboli and Marco Castellana contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11154-020-09592-3) contains supplementary material, which is available to authorized users.

✉ Pierpaolo Trimboli
pierpaolo.trimboli@eoc.ch

Marco Castellana
mcastellana01@yahoo.it

Arnoldo Piccardo
arnoldo.piccardo@galliera.it

Francesco Romanelli
francesco.romanelli@uniroma1.it

Giorgio Grani
giorgio.grani@uniroma1.it

Luca Giovanella
luca.giovanella@eoc.ch

Cosimo Durante
cosimo.durante@uniroma1.it

1   Clinic for Nuclear Medicine and Competence Center for Thyroid Diseases, Imaging Institute of Southern Switzerland, Ente Ospedaliero Cantonale, Via Ospedale 12, 6500 Bellinzona, Svizzera, Switzerland

2   Biomedical Sciences, Università della Svizzera Italiana (USI), Lugano, Switzerland

3   National Institute of Gastroenterology "Saverio de Bellis", Research Hospital, Castellana Grotte, Bari, Italy

4   Department of Nuclear Medicine, Galliera Hospital, Genoa, Italy

5   Department of Experimental Medicine, "Sapienza" University of Rome, Rome, Italy

6   Department of Translational and Precision Medicine, Sapienza University of Rome, Rome, Italy

7   Clinic for Nuclear Medicine, University Hospital and University of Zurich, Zurich, Switzerland

1 Introduction

Ultrasound (US) is recognized as the most valuable imaging modality for the assessment of malignancy risk of thyroid nodules. As largely proven during the last two decades, specific nodule's US characteristics, such as hypoechogenicity, taller-than-wide shape, irregular margins, microcalcifications, and extrathyroidal extension, should be considered as suspicious and recommend fine-needle aspiration (FNA) [1]. However, their use as single parameters suffer from low/ suboptimal sensitivity and moderate/high inter- and intra-observer variability in both recognition and reporting [2]. Therefore, US risk stratification systems for thyroid nodule (RSSs, often referred to as Thyroid Imaging Reporting And Data Systems "TIRADSs") [3–14] have been developed with the aim of: 1) establishing a standard lexicon of nodule description; 2) defining the suspicious characteristics; 3) putting the nodule into a risk category; and 4) identifying those nodules in which FNA is indicated also considering the size. Since the introduction of these RSSs in clinical practice, several studies aimed to compare their performance. However, one meta-analysis on this topic showed that, regardless of the high emphasis on RSSs, only sparse data were available in the literature, limiting the number of head-to-head comparisons that could be performed [15]. In addition, some methodological limitations are present in these studies. Ideally, we should validate these systems in a cohort as truthful as possible. Such a study should: 1) contain nodules randomized to undergo FNA or not; 2) include a histologic diagnosis to confirm or not the cytological assessment (it is recognized that FNA suffers from false positives and false negatives [9, 12]); and 3) be conducted by differently experienced US operators (radiologists and endocrinologists). Unfortunately, the published data on this topic are almost all retrospective, neither from this study design nor from this patients' setting, and the indication to surgery was frequently based on FNA. In particular, the latter issue represents a major selection bias because FNA accurately diagnoses most papillary carcinomas (PTC) while follicular carcinoma (FTC) is invariably put in the indeterminate FNA category [16, 17] and medullary carcinoma (MTC) is misdiagnosed on FNA in up to 50% of cases [18]. In addition, the cancer prevalence in these studies varied largely and this influenced the results, since it is well known that the performance of a diagnostic test depends on the event/ disease frequency [15, 19].

The present study was conceived to verify whether the performance of RSSs has been adequately investigated in all thyroid malignancies. Here we systematically searched studies classifying thyroid nodules according to five commonly used US RSSs and reporting the histological diagnosis of malignant lesions. Also, we performed a meta-analysis of available data to evaluate: 1) the pooled cancer prevalence; and 2) the relative prevalence of PTC, FTC, MTC and other malignances.

2 Methods

The systematic review was performed in accordance with the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) (Supplementary Table 1) [20].

2.1 Search strategy

A six-step search strategy was planned. Firstly, sentinel studies were searched in PubMed. Secondly, keywords and MeSH terms were identified in PubMed. Thirdly, in order to test the strategy, the terms "AACE/ACE/AME", "ACR TIRADS", "EU-TIRADS", "K-TIRADS" and "ATA" were searched in PubMed. Fourthly, PubMed, CENTRAL, Scopus and Web of Science were searched. Fifthly, studies meeting all the following criteria were included: 1) at least 1000 nodules should be assessed; 2) nodules should be classified according to at least one US RSS among American Association of Clinical Endocrinologist/American College of Endocrinology/ Associazione Medici Endocrinologi (AACE/ACE/AME) [10], American College of Radiology (ACR-TIRADS) [11], 2015 American Thyroid Association (ATA) [12], European Thyroid Association (EU-TIRADS) [13], and Korean Society of Thyroid Radiology and Korean Society of Radiology (K-TIRADS) [14]; 3) data on the performance of at least one of the above US RSS should be reported (e.g. the prevalence of malignancy in each US RSS class or indication to FNA according to US RSS); 4) the diagnosis of malignant nodules had not to be based on cytology only; 5) data on the overall prevalence of malignancy and the relative prevalence of PTC, FTC, MTC and other malignancies among all malignancies should be reported. Studies were excluded if focusing on pediatric patients or on specific subgroups of thyroid nodules (e.g. indeterminate, only solid or predominantly solid). Finally, references of included studies were screened for additional papers. The last search was performed on February 1st, 2020. Articles in all languages were accepted and with no restriction to the year they were published. Two investigators (MC, PT) independently and in duplicate searched papers, screened titles and abstracts of the retrieved articles, reviewed the full-texts and selected articles for their inclusion.

2.2 Data extraction

The following information was extracted independently and in duplicate by two investigators (MC, PT) in a piloted form: 1) general information on the study (author, year of publication, country, study type, number of patients, number of nodules, selection criteria of included nodules); 2) reference standard for the diagnosis of malignancy; 3) number of malignant nodules; 4) number of PTC, FTC, MTC and other malignancies. The main paper and supplementary data were searched;

if data was missing, authors were contacted via email. Data were cross-checked and any discrepancy was discussed.

2.3 Study quality assessment

The risk of bias of included studies was assessed independently by two reviewers (MC, PT). The National Heart, Lung, and Blood Institute Quality Assessment Tool was used, and the following aspects evaluated: study question; eligibility criteria; sample size calculation; description and delivering of intervention; definition of outcome measures; duration of follow-up; blinding; loss to follow-up; statistical methods. Each domain was assigned absence, unclear or possible risk of bias or as not applicable [21].

2.4 Data analysis

The characteristics of included studies were summarized. Then, separate analyses were performed according to the following steps. First, a proportion meta-analysis was carried to obtain the pooled prevalence of malignancy among all included nodules. A sub-group analysis was performed for studies including histologic series only or both histologic and cytologic series. Second, a proportion meta-analysis was carried to obtain the pooled prevalence of PTC, FTC, MTC and other malignancies among malignancies diagnosed at histology. Heterogeneity between studies was assessed by using I^2, with 50% or higher values regarded as high heterogeneity. The Egger's test was

carried out to evaluate the possible presence of significant publication bias; the trim-and-fill method was used for estimating its effect. For statistical pooling of data, a random-effects model was used. All analyses were performed on a per lesion basis and carried out using StatsDirect statistical software (StatsDirect Ltd.; Altrincham, UK) and Prometa3.0 (Internovi). A $p < 0.05$ was regarded as significant.

3 Results

A total of 1298 papers were found, of which 193 on PubMed, 56 on CENTRAL, 155 on Scopus, and 894 on Web of Science. One additional paper was retrieved from a personal database [22]. After removal of 292 duplicates, 1007 articles were analyzed for title and abstract; 879 records were excluded (guidelines, review, meta-analysis, inclusion of specific subgroups of nodules, pediatric patients, not within the field of the review). The remaining 128 papers were retrieved in full-text and 9 studies were finally included in the systematic review (Fig. 1) [22–30]. No additional study was retrieved from references of included studies.

3.1 Study quality assessment

The risk of bias of the included studies is shown in Supplementary Table 2. Statement of the study question, description of the study population, participation rate, assessment of the
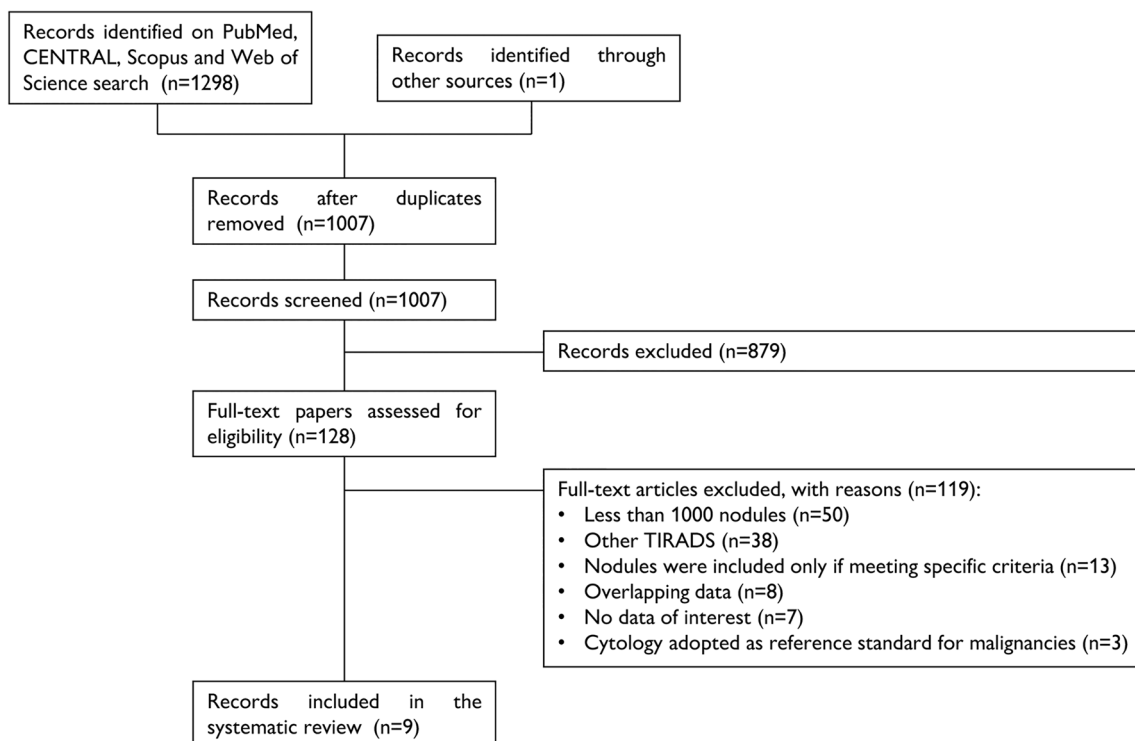

Fig. 1 Flow chart of the systematic review

exposures and outcome bias were adequate in all. Ultrasound was performed before cytology or surgery and images retrospectively reviewed but the timeframe between the two assessments was considered as adequate as cancer a chronic disease. Reviewers were generally blinded to the final diagnosis.

3.2 Qualitative analysis (systematic review)

The characteristics of the included articles are summarized in Table 1. The studies were published between 2017 and 2020 and had sample sizes ranging from 1001 to 4696 thyroid nodules. All studies were retrospective cohort and assessed the performance of at least one TIRADS among AACE/ACE/ AME, ACR-TIRADS, ATA, EU-TIRADS, K-TIRADS. Five studies were carried out in China, two in the United States of America, one in Korea, and one multicenter study in France, Switzerland and the United Kingdom. Participants were adult outpatients with US images available who had undergone either thyroid nodule FNA or surgery in six studies or surgery only in three studies [24, 26, 27]. Thyroid nodules diagnosed as non-diagnostic or indeterminate on FNA were excluded, unless a final diagnosis was met on pathology. Overall, 19,494 thyroid nodules were included in the present review, among which 6391 were malignant. Among the 6162 malignant nodules diagnosed at histology, the number of PTC, FTC, MTC and other malignancies was 5963, 97, 54 and 48, respectively.

3.3 Quantitative analysis (meta-analysis)

The overall prevalence of malignancy in all articles included in the meta-analysis was 34% (95%CI 21 to 49). When a

subgroup analysis according to the reference standard for malignancy was performed, no difference was found between studies using histology only or cytology and histology (37%; 95%CI 18 to 57 versus 31%; 95%CI 14 to 51, respectively; $p = 0.64$).

Among the 6162 histologically proven malignancies, four separate meta-analyses on the prevalence of PTC, FTC, MTC and other malignancies were performed and it was found a rate of 95%, 2%, 1%, and 1%, respectively (Fig. 2). A high heterogeneity and evidence of publication bias were found for all the outcomes, with the exception of the overall prevalence of malignancy; the trim-and-fill method did not change the estimates (Supplemental Table 3).
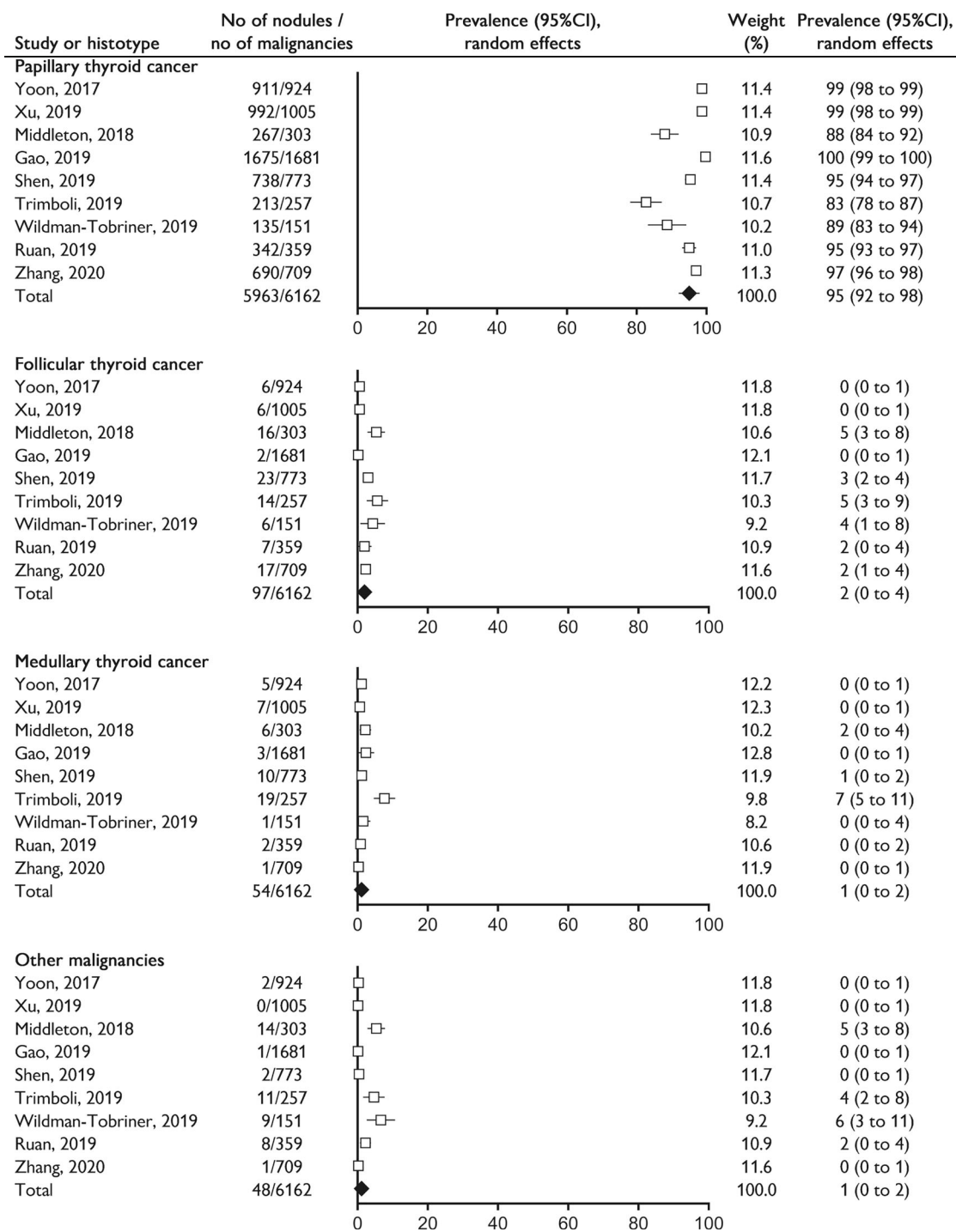
4 Discussion

Thyroid US examination represents the gold standard for the management of thyroid nodules, their risk stratification, and their indication for FNA. With the present article we raised the question of whether the RSSs have been evaluated for all types of thyroid malignancies. Accordingly, we browsed the published literature to find the largest number of original papers, with a minimum simple of one thousand nodules, which aimed to verify the accuracy of RSSs and included histologically diagnosed malignant lesions. Two main questions were addressed in our study: 1) what is the cancer rate reported in these studies? and 2) what is the relative prevalence of the histologic types of thyroid malignancy? With the introduction of the RSSs in clinical practice all thyroidologists started to select thyroid nodules for FNA or clinical follow-up according to the criteria described in these consensus or guidelines

Table 1 Characteristics of included studies

First Author, year [ref]	Country	Study design	Thyroid nodules (n)	Selection criteria of included nodules	Reference standard for malignancy
Yoon, 2017 [22]	Korea	RCS	4696	10–19 mm, benign cytology, malignant cytology or surgery	Histology or cytology
Middleton, 2018 [23]	United States of America	RCS	3422	Benign cytology, malignant cytology or surgery	Histology or cytology
Gao, 2019 [24]	China (Beijing)	RCS	2544	Surgery	Histology
Ruan, 2019 [25]	China (Guangzhou)	RCS	1001	Benign cytology, malignant cytology or surgery	Histology or cytology
Shen, 2019 [26]	China (Shanghai)	RCS	1612	>5 mm, surgery	Histology
Trimboli, 2019 [27]	France, Switzerland, United Kingdom	RCS	1058	≥5 mm, surgery	Histology
Wildman-Tobriner, 2019 [28]	United States of America	RCS	1425	Benign cytology, malignant cytology or surgery	Histology
Xu, 2019 [29]	China (Nanjing)	RCS	2465	Benign cytology or surgery	Histology
Zhang, 2020 [30]	China (Shanghai)	RCS	1271	≥5 mm, benign cytology, malignant cytology or surgery	Histology or cytology

Legend – *RCS* retrospective cohort study

Study or histotype	No of nodules / no of malignancies	Prevalence (95%CI), random effects	Weight (%)	Prevalence (95%CI), random effects
Papillary thyroid cancer				
Yoon, 2017	911/924		11.4	99 (98 to 99)
Xu, 2019	992/1005		11.4	99 (98 to 99)
Middleton, 2018	267/303		10.9	88 (84 to 92)
Gao, 2019	1675/1681		11.6	100 (99 to 100)
Shen, 2019	738/773		11.4	95 (94 to 97)
Trimboli, 2019	213/257		10.7	83 (78 to 87)
Wildman-Tobriner, 2019	135/151		10.2	89 (83 to 94)
Ruan, 2019	342/359		11.0	95 (93 to 97)
Zhang, 2020	690/709		11.3	97 (96 to 98)
Total	5963/6162		100.0	95 (92 to 98)
Follicular thyroid cancer				
Yoon, 2017	6/924		11.8	0 (0 to 1)
Xu, 2019	6/1005		11.8	0 (0 to 1)
Middleton, 2018	16/303		10.6	5 (3 to 8)
Gao, 2019	2/1681		12.1	0 (0 to 1)
Shen, 2019	23/773		11.7	3 (2 to 4)
Trimboli, 2019	14/257		10.3	5 (3 to 9)
Wildman-Tobriner, 2019	6/151		9.2	4 (1 to 8)
Ruan, 2019	7/359		10.9	2 (0 to 4)
Zhang, 2020	17/709		11.6	2 (1 to 4)
Total	97/6162		100.0	2 (0 to 4)
Medullary thyroid cancer				
Yoon, 2017	5/924		12.2	0 (0 to 1)
Xu, 2019	7/1005		12.3	0 (0 to 1)
Middleton, 2018	6/303		10.2	2 (0 to 4)
Gao, 2019	3/1681		12.8	0 (0 to 1)
Shen, 2019	10/773		11.9	1 (0 to 2)
Trimboli, 2019	19/257		9.8	7 (5 to 11)
Wildman-Tobriner, 2019	1/151		8.2	0 (0 to 4)
Ruan, 2019	2/359		10.6	0 (0 to 2)
Zhang, 2020	1/709		11.9	0 (0 to 1)
Total	54/6162		100.0	1 (0 to 2)
Other malignancies				
Yoon, 2017	2/924		11.8	0 (0 to 1)
Xu, 2019	0/1005		11.8	0 (0 to 1)
Middleton, 2018	14/303		10.6	5 (3 to 8)
Gao, 2019	1/1681		12.1	0 (0 to 1)
Shen, 2019	2/773		11.7	0 (0 to 1)
Trimboli, 2019	11/257		10.3	4 (2 to 8)
Wildman-Tobriner, 2019	9/151		9.2	6 (3 to 11)
Ruan, 2019	8/359		10.9	2 (0 to 4)
Zhang, 2020	1/709		11.6	0 (0 to 1)
Total	48/6162		100.0	1 (0 to 2)

Fig. 2 Forest plot of the four meta-analyses of the relative prevalence of papillary, follicular, medullary thyroid cancer and other malignancies among histologically proven malignancies

[3–14]. The latter should represent a significant advancement of thyroid US culture towards a homogeneous worldwide approach to thyroid nodule [31]. However, before considering the RSSs as the gold standard to manage our patients we should have more solid proofs and be aware of what we can reasonably expect from these systems. In fact, one

thyroidologist using any RSSs might expect that these systems have been designed to identify all types of thyroid malignancies. The results of our study challenge this expectation.

According to our search strategy we found nine articles including a total of 19,494 thyroid nodules of which 6391 were malignant. The pooled cancer prevalence reported in these

articles was 34%, with heterogeneity. Moreover, among all malignant nodules, PTC represented the 95%. Both findings are of high clinical relevance. First, all RSSs were conceived for selecting thyroid nodules for FNA. Then, when comparing their performance, only summary operating measures assumed to be independent of the disease prevalence should be used (e.g. diagnostic odds ratio) [15]. On the other hand, biased result can be obtained if a comparison is based only on other parameters (e.g. sensitivity, specificity) [32]. Second, from a clinical point of view, this histologic prevalence deserves more thorough discussion. What is particularly striking is that the percentage of FTC and MTC seems much lower than expected in such selected populations [33]. This finding can be due to the challenges faced by clinicians when making a diagnosis of FTC or MTC. FTC has often an unsuspicious echo-structural presentation and cannot reliably be diagnosed on cytology, as stated [17]. Therefore, cytologically indeterminate thyroid nodules without suspicious sonographic patterns warrant particularly careful follow-up strategies. In addition, FTC rate is heavily influenced by the epidemiological curves, which are consistently showing a decrease in the frequency of new cases over the last three decades [34, 35]. Similarly, MTC has an heterogenous US presentation and is difficult to detect on FNA [17, 36]. The routinely use of serum calcitonin, the most sensitive tool for MTC diagnosis, can possibly improve diagnosis, but it is still a debated matter [4, 7, 10, 12]. Finally, the rate of the other thyroid malignancies (i.e., lymphoma, metastases from other organs) is expected low and these lesions may have heterogeneous US presentation too [37]. It has also to be taken into account that PTC is the most frequently diagnosed thyroid cancer, it can be easily detected in the clinical practice due to its typical US features and the high performance of the cytological assessment. Therefore, when reviewing a histological series of thyroid nodules, a large number of PTC is widely expected. All these clinical issues may have affected the relative rates of the different types of thyroid malignancies included in the studies evaluating the accuracy of RSSs.

Beyond all the above considerations, it is indisputable based on our data that RSSs' performance has been tested almost exclusively in PTC patients, thus supporting the view that the clinical validity of these systems cannot be unconditionally extended to other forms of malignancy [31]. Therefore, while a sonographic-centered diagnostic work-up can effectively identify PTCs, RSSs cannot be advocated to reliable diagnose those cancers burdened by the greatest risk of mortality, i.e., FTCs [16] and MTCs [38]. Patients diagnosed with a large FTC have a higher risk of developing distant macro-metastases, for which radioactive iodine therapy may be ineffective [39]. MTCs are expected to spread-out early to loco-regional lymph-nodes and to distant sites, even if diagnosed when small in size [40]. All cases falling in these clinical scenarios are invariably not curable, require lifelong treatments often affecting the quality of life, and

have a lower life expectancy [41]. An important diagnostic effort should be fielded to allow clinicians not to miss these diagnoses. This implies an effort to develop US RSSs able to intercept FTC and MTC cases. On the other hand, a great effort is underway in validating molecular tests to improve cytological diagnosis [42–45]. In the meantime, international guidelines have sped up this process by promoting the potential of molecular tests in clinical practice [12, 41].

This review has several limitations. The first limitation relates to the design of included studies: a retrospective review of nodules that have been submitted to FNA or surgery was performed in most of them, and this introduced a significant selection bias. We selected only those studies in which at least 1000 nodules were included, and this is a second limitation. However, the prevalence of FTC, MTC and other cancers is expected to be low compared to PTC, then only studies with an adequate sample size could be deemed sufficiently powered to reliably determine their frequency. Lastly, despite being classified as PTC, specific subtypes have been associated with a worse prognosis [12]. Further studies are needed to assess the representativeness of these subtypes in studies assessing the performance of US RSSs.

In conclusion, almost all histologically proven cancers found in the studies evaluating the accuracy of RSSs are PTCs. On one hand, this suggests that US classifications are an accurate tool to diagnose PTC. Their reliability in detecting FTC, MTC and other malignancies should still be improved, by either modifying patterns and cut-offs for FNA or integrating US with other technologies. From another perspective, our results raise the question of whether during our clinical practice we are on the hunt of PTCs while we are neglecting the most aggressive thyroid cancers. We advise for further studies investigating the latter issue.

Compliance with ethical standards

Ethical approval These systematic review and meta-analysis were in accordance with the principles of the Declaration of Helsinki. Analyses were performed on data extracted from published papers.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Durante C, Grani G, Lamartina L, Filetti S, Mandel SJ, Cooper DS. The diagnosis and management of thyroid nodules: a review. JAMA. 2018;319(9):914–24. https://doi.org/10.1001/jama.2018.0898.

2. Grani G, Lamartina L, Cantisani V, Maranghi M, Lucia P, Durante C. Interobserver agreement of various thyroid imaging reporting and data systems. Endocr Connect. 2018;7(1):1–7. https://doi.org/10.1530/EC-17-0336.

3. Frates MC, Benson CB, Charboneau JW, Cibas ES, Clark OH, Coleman BG, et al. Management of thyroid nodules detected at US: Society of Radiologists in Ultrasound consensus conference statement. Radiology. 2005;237(3):794–800. https://doi.org/10.1148/radiol.2373050220.

4. American Thyroid Association (ATA) Guidelines Taskforce on Thyroid Nodules and Differentiated Thyroid Cancer, Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. Thyroid. 2009;19(11):1167–214. https://doi.org/10.1089/thy.2009.0110.

5. Horvath E, Majlis S, Rossi R, Franco C, Niedmann JP, Castro A, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. J Clin Endocrinol Metab. 2009;94(5):1748–51. https://doi.org/10.1210/jc.2008-1724.

6. Park JY, Lee HJ, Jang HW, Kim HK, Yi JH, Lee W, et al. A proposal for a thyroid imaging reporting and data system for ultrasound features of thyroid carcinoma. Thyroid. 2009;19(11):1257–64. https://doi.org/10.1089/thy.2008.0021.

7. Gharib H, Papini E, Paschke R, Duick DS, Valcavi R, Hegedüs L, et al. American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association Medical guidelines for clinical practice for the diagnosis and management of thyroid nodules: executive summary of recommendations. Endocr Pract. 2010;16(3):468–75. https://doi.org/10.4158/EP.16.3.468.

8. Kwak JY, Han KH, Yoon JH, Moon HJ, Son EJ, Park SH, et al. Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. Radiology. 2011;260(3):892–9. https://doi.org/10.1148/radiol.11110206.

9. Perros P, Boelaert K, Colley S, Evans C, Evans RM, Gerrard Ba G, et al. Guidelines for the management of thyroid cancer. Clin Endocrinol. 2014;81(Suppl 1):1–122. https://doi.org/10.1111/cen.12515.

10. Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedüs L, 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. Endocr Pract. 2016;22(5):622–39. https://doi.org/10.4158/EP161208.GL.

11. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS Committee. J Am Coll Radiol. 2017;14(5):587–95. https://doi.org/10.1016/j.jacr.2017.01.046.

12. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, 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. 2016;26(1):1–133. https://doi.org/10.1089/thy.2015.0020.

13. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European thyroid association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. Eur Thyroid J. 2017;6(5):225–37. https://doi.org/10.1159/000478927.

14. Shin JH, Baek JH, Chung J, Ha EJ, Kim JH, Lee YH, et al. Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of thyroid radiology consensus statement and Recommendations. Korean J Radiol. 2016;17(3):370–95. https://doi.org/10.3348/kjr.2016.17.3.370.

15. Castellana M, Castellana C, Treglia G, Giorgino F, Giovanella L, Russ G, et al. Performance of five ultrasound risk stratification systems in selecting thyroid nodules for FNA. J Clin Endocrinol Metab. 2020;105(5):dgz170. https://doi.org/10.1210/clinem/dgz170.

16. Grani G, Lamartina L, Durante C, Filetti S, Cooper DS. Follicular thyroid cancer and Hürthle cell carcinoma: challenges in diagnosis, treatment, and clinical management. Lancet Diabetes Endocrinol. 2018;6(6):500–14. https://doi.org/10.1016/S2213-8587(17)30325-X.

17. Castellana M, Piccardo A, Virili C, Scappaticcio L, Grani G, Durante C, et al. Can ultrasound systems for risk stratification of thyroid nodules identify follicular carcinoma? Cancer Cytopathol. 2020;128(4):250–9. https://doi.org/10.1002/cncy.22235.

18. Trimboli P, Treglia G, Guidobaldi L, Romanelli F, Nigri G, Valabrega S, et al. Detection rate of FNA cytology in medullary thyroid carcinoma: a meta-analysis. Clin Endocrinol. 2015;82(2):280–5. https://doi.org/10.1111/cen.12563.

19. Eusebi P. Diagnostic accuracy measures. Cerebrovasc Dis. 2013;36(4):267–72. https://doi.org/10.1159/000353863.

20. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. JAMA. 2000;283(15):2008–12. https://doi.org/10.1001/jama.283.15.2008.

21. National Heart, Lung, and Blood Institute. Study Quality Assessment Tools. https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools. Accessed 1 Mar 2020.

22. Yoon JH, Han K, Kim EK, Moon HJ, Kwak JY. Diagnosis and management of small thyroid nodules: a comparative study with six guidelines for thyroid nodules. Radiology. 2017;283(2):560–9. https://doi.org/10.1148/radiol.2016160641.

23. Middleton WD, Teefey SA, Reading CC, Langer JE, Beland MD, Szabunio MM, et al. Comparison of performance characteristics of

American College of Radiology TI-RADS, Korean Society of Thyroid Radiology TIRADS, and American Thyroid Association guidelines. AJR Am J Roentgenol. 2018;210(5):1148–54. https://doi.org/10.2214/AJR.17.18822.

24. Gao L, Xi X, Jiang Y, Yang X, Wang Y, Zhu S, et al. Comparison among TIRADS (ACR TI-RADS and KWAK- TI-RADS) and 2015 ATA guidelines in the diagnostic efficiency of thyroid nodules. Endocrine. 2019;64(1):90–6. https://doi.org/10.1007/s12020-019-01843-x.

25. Ruan JL, Yang HY, Liu RB, Liang M, Han P, Xu XL, et al. Fine needle aspiration biopsy indications for thyroid nodules: compare a point-based risk stratification system with a pattern-based risk stratification system. Eur Radiol. 2019;29(9):4871–8. https://doi.org/10.1007/s00330-018-5992-z.

26. Shen Y, Liu M, He J, Wu S, Chen M, Wan Y, et al. Comparison of different risk-stratification systems for the diagnosis of benign and malignant thyroid nodules. Front Oncol. 2019;9:378. https://doi.org/10.3389/fonc.2019.00378.

27. Trimboli P, Ngu R, Royer B, Giovanella L, Bigorgne C, Simo R, et al. A multicentre validation study for the EU-TIRADS using histological diagnosis as a gold standard. Clin Endocrinol. 2019;91(2):340–7. https://doi.org/10.1111/cen.13997.

28. Wildman-Tobriner B, Buda M, Hoang JK, Middleton WD, Thayer D, Short RG, et al. Using artificial intelligence to revise ACR TI-RADS risk stratification of thyroid nodules: diagnostic accuracy and utility. Radiology. 2019;292(1):112–9. https://doi.org/10.1148/radiol.2019182128.

29. Xu T, Wu Y, Wu RX, Zhang YZ, Gu JY, Ye XH, et al. Validation and comparison of three newly-released thyroid imaging reporting and data systems for cancer risk determination. Endocrine. 2019;64(2):299–307. https://doi.org/10.1007/s12020-018-1817-8.

30. Zhang WB, Xu HX, Zhang YF, Guo LH, Xu SH, Zhao CK, et al. Comparisons of ACR TI-RADS, ATA guidelines, Kwak TI-RADS, and KTA/KSThR guidelines in malignancy risk stratification of thyroid nodules. Clin Hemorheol Microcirc. 2020;75:219–32. https://doi.org/10.3233/CH-190778.

31. Trimboli P, Durante C. Ultrasound risk stratification systems for thyroid nodule: between lights and shadows, we are moving towards a new era. Endocrine. 2020;69:1–4. https://doi.org/10.1007/s12020-020-02196-6.

32. Kim PH, Suh CH, Baek JH, Chung SR, Choi YJ, Lee JH. Diagnostic performance of four ultrasound risk stratification systems: a systematic review and meta-analysis. Thyroid. 2020;30:1159–68. https://doi.org/10.1089/thy.2019.0812.

33. Lloyd RV, Osamura RY, Klöppel G, Rosai J. WHO classification of TUMOURS of endocrine organs. 4th ed. Lyon: IARC; 2017.

34. Trimboli P, Ulisse S, Graziano FM, Marzullo A, Ruggieri M, Calvanese A, et al. Trend in thyroid carcinoma size, age at diagnosis, and histology in a retrospective study of 500 cases diagnosed over 20 years. Thyroid. 2006;16:1151–5. https://doi.org/10.1089/thy.2006.16.1151.

35. Livolsi VA, Asa SL. The demise of follicular carcinoma of the thyroid gland. Thyroid. 1994;4:233–6. https://doi.org/10.1089/thy.1994.4.233.

36. Trimboli P, Giovanella L, Valabrega S, Andrioli M, Baldelli R, Cremonini N, et al. Ultrasound features of medullary thyroid carcinoma correlate with cancer aggressiveness: a retrospective multi-center study. J Exp Clin Cancer Res. 2014;33:87. https://doi.org/10.1186/s13046-014-0087-4.

37. Falcone R, Ramundo V, Lamartina L, Ascoli V, Bosco D, Di Gioia C, et al. Sonographic presentation of metastases to the thyroid gland: a case series. J Endocr Soc. 2018;2(8):855–9. https://doi.org/10.1210/js.2018-00124.

38. Cabanillas ME, McFadden DG, Durante C. Thyroid cancer. Lancet. 2016;388(10061):2783–95. https://doi.org/10.1016/S0140-6736(16)30172-6.

39. Durante C, Haddy N, Baudin E, Leboulleux S, Hartl D, Travagli JP, et al. Long-term outcome of 444 patients with distant metastases from papillary and follicular thyroid carcinoma: benefits and limits of radioiodine therapy. J Clin Endocrinol Metab. 2006;91(8):2892–9. https://doi.org/10.1210/jc.2005-2838.

40. Machens A, Dralle H. Biomarker-based risk stratification for previously untreated medullary thyroid cancer. J Clin Endocrinol Metab. 2010;95(6):2655–63. https://doi.org/10.1210/jc.2009-2368.

41. Filetti S, Durante C, Hartl D, Leboulleux S, Locati LD, Newbold K, et al. Thyroid cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up†. Ann Oncol. 2019;30(12):1856–83. https://doi.org/10.1093/annonc/mdz400.

42. Steward DL, Carty SE, Sippel RS, Yang SP, Sosa JA, Sipos JA, et al. Performance of a multigene genomic classifier in thyroid nodules with indeterminate cytology: a prospective blinded multicenter study. JAMA Oncol. 2019;5(2):204–12. https://doi.org/10.1001/jamaoncol.2018.4616.

43. San Martin VT, Lawrence L, Bena J, Madhun NZ, Berber E, Elsheikh TM, et al. Real-world comparison of afirma GEC and GSC for the assessment of cytologically indeterminate thyroid nodules. J Clin Endocrinol Metab. 2020;105(3):dgz099. https://doi.org/10.1210/clinem/dgz099.

44. Banizs AB, Silverman JF. The utility of combined mutation analysis and microRNA classification in reclassifying cancer risk of cytologically indeterminate thyroid nodules. Diagn Cytopathol. 2019;47(4):268–74. https://doi.org/10.1002/dc.24087.

45. Sponziello M, Brunelli C, Verrienti A, Grani G, Pecce V, Abballe L, et al. Performance of a dual-component molecular assay in cytologically indeterminate thyroid nodules. Endocrine. 2020;68:458–65. https://doi.org/10.1007/s12020-020-02271-y.