



PREDICTIVE VALUE OF SERUM TSH AND FREE T4 LEVELS IN DIFFERENTIATING MALIGNANT FROM BENIGN THYROID NODULES: A SYSTEMATIC REVIEW

Ni Luh Gede Taman Kristanti ¹, I Gusti Ngurah Bayu Paramartha ², Ni
Komang Okyantri Dana Pertiwi ³

^{1,2} Faculty of Medicine and Health Sciences, Universitas Warmadewa,
Denpasar, Bali, Indonesia

³ Faculty of Medicine, Universitas Pendidikan Ganesha, Buleleng,
Bali, Indonesia

Email¹: criztaman@gmail.com

Article History :

Received date : 2026/07/02
Revised date : 2026/08/27
Accepted date : 2026/09/19
Published date : 2026/10/08



Copyright: © 2024 by the
authors. Submitted for possible
open access publication under
the terms and conditions of the
Creative Commons Attribution
(BY NC) license
(<https://creativecommons.org/licenses/by-nc/4.0/>).



ABSTRACT

Background. Thyroid nodules are detected in most adults but few are malignant. Serum thyroid-stimulating hormone (TSH) is measured in almost every nodule work-up, yet guidance uses it to select patients for scintigraphy rather than to grade cancer risk. Whether TSH, or free thyroxine (FT4), adds usable diagnostic information for the individual nodule is unsettled.

Objective. To determine whether pre-diagnostic serum TSH or FT4 discriminates malignant from benign thyroid nodules verified by cytology or histopathology in adults.

Methods. Systematic review reported to PRISMA 2020. PubMed/MEDLINE and the Consensus index

were searched to 17 July 2026; strings for Scopus, Embase, Web of Science and CENTRAL are supplied for execution. Eligible studies were observational nodule cohorts reporting TSH or FT4 against a cytological or histological reference standard. Risk of bias was assessed with QUADAS-2, certainty with GRADE. Effect measures were not convertible, so synthesis was narrative by outcome.

Results. Twelve studies comprising 10,497 adults, recruited 1984-2021 across seven countries, met the criteria. Nine reported higher TSH in malignant than benign nodules; three found no association. Adjusted odds ratios ranged from 1.49 to 11.18, the largest confined to the highest TSH strata of one cohort. Reported thresholds spanned 0.9 to 2.26 mU/L, each derived within the cohort that evaluated it, none externally validated. FT4 was measured in four studies but analysed in two, which disagreed. Nine of twelve studies were at high risk of bias, most often from patient selection restricted to operated patients. Certainty was very low throughout.

Conclusion. Higher serum TSH is associated with malignancy in most but not all nodule cohorts. The association is not reproducible at any threshold, so TSH cannot serve as a stand-alone triage test, and the FT4 evidence is too sparse to judge. TSH is best used as one weak contributor within a multivariable assessment, not as an independent diagnostic criterion.

Keywords: thyroid nodule; thyrotropin; thyroid stimulating hormone; free thyroxine; thyroid neoplasms; diagnosis; systematic review.

INTRODUCTION

Thyroid nodules are among the most common findings in clinical medicine. Population screening with high-frequency ultrasound identifies nodules in roughly two-thirds of unselected adults, a prevalence far above that obtained by palpation and one that rises steadily with age.¹ Contemporary reviews place the detection rate at up to 65% of the general population, driven largely by imaging performed for reasons unrelated to the thyroid.² The clinical significance of this reservoir is narrow: approximately one nodule in ten harbours a clinically important cancer, about one in twenty causes compressive symptoms, and a similar minority progresses to functional disease.² The task facing the clinician is therefore not detection but discrimination.

That task has become more consequential as thyroid cancer incidence has risen. Thyroid cancer now ranks among the more frequently diagnosed malignancies worldwide, with marked variation between and within countries.³ Much of the increase reflects overdiagnosis rather than a true change in disease burden. Analyses across high-income settings attribute a large share of recent cases to the detection of small, indolent papillary tumours that would never have surfaced clinically,⁴ and registry work in Italy estimated that overdiagnosis accounted for approximately 75% of cases in women and 63% in men between 1998 and 2012.⁵ Any test proposed as an addition to the diagnostic pathway must therefore be judged not only on whether it identifies cancer, but on whether it directs attention toward tumours that matter.

The established pathway rests on two pillars. Ultrasound risk stratification systems assign nodules to categories based on composition, echogenicity, shape, margin and echogenic foci, and these categories determine whether cytology is indicated.^{6,7} Fine-needle aspiration cytology, reported using the Bethesda system, then assigns an implied risk of malignancy to each of six diagnostic categories.⁸ The pathway performs well at the extremes and poorly in the middle: roughly a fifth to a third of aspirates return indeterminate results, and it is precisely these nodules for which an inexpensive, universally available adjunct would be most valuable.²

Serum TSH is the obvious candidate. It is measured as the first laboratory step in essentially every nodule assessment, costs little, and is biologically plausible as a risk marker. TSH is the

principal growth factor for the thyroid follicular cell, acting through the TSH receptor and downstream signalling pathways that also govern proliferation in transformed thyrocytes; sustained stimulation promotes thyroid growth, and pharmacological reduction of TSH has been associated with smaller thyroid volume and a lower frequency of nodular disease.⁹ If TSH drives thyrocyte proliferation, higher circulating concentrations might plausibly mark a gland in which malignant transformation is more likely to have occurred or to have been sustained.

Epidemiological evidence supports an association at population level. A systematic review and dose-response meta-analysis of 28 studies, including 42,032 subjects and 5,786 thyroid cancer cases, found a nonlinear relationship between serum TSH and thyroid cancer.¹⁰ Below 1 mU/L the odds ratio was 1.72 (95% CI 1.42-2.07) per mU/L; at and above 1 mU/L it flattened to 1.16 (95% CI 1.12-1.21) per mU/L. Notably, studies that controlled for thyroid autoimmunity reported the weakest associations, with the odds ratio at TSH values of 2.5 mU/L or above falling to 0.98 (95% CI 0.89-1.09) per mU/L. That pattern raises the possibility that part of the apparent TSH signal is carried by coexisting autoimmune thyroid disease rather than by thyrotropin itself.

A population-level association and a usable clinical test are, however, different things. An odds ratio of 1.16 per mU/L across the range in which most patients actually sit implies very little movement in post-test probability for an individual nodule. Guidelines reflect this: current recommendations use serum TSH to determine whether a radionuclide scan is warranted, and hence whether cytology can be avoided in an autonomously functioning nodule, rather than to stratify malignancy risk directly.⁶ Whether the primary studies conducted in nodule cohorts, as opposed to registry populations, support a stronger role has not been synthesised. Free thyroxine presents a further gap. FT4 is measured alongside TSH in a large proportion of the same cohorts, is mechanistically linked to TSH through negative feedback, and would cost nothing extra to interpret, yet it is rarely analysed as a predictor in its own right.

We conducted a systematic review of observational studies in adults with thyroid nodules to determine whether pre-diagnostic serum TSH, or serum FT4, discriminates malignant from benign nodules verified by cytology or histopathology, whether any association persists after adjustment for established predictors, and whether TSH relates to tumour stage or size among malignant nodules.

METHODS

A protocol specifying the question, eligibility criteria, search, extraction, risk-of-bias method and synthesis plan was prepared before synthesis began and accompanies this report. Reporting follows the PRISMA 2020 statement.¹¹ Deviations from the protocol are identified below where they occurred.

Eligibility criteria

Studies were eligible if they enrolled adults aged 18 years or older with one or more thyroid nodules assessed in clinical practice, measured serum TSH, or serum FT4, before a diagnosis was established, and reported that measurement against a reference standard of surgical histopathology or fine-needle aspiration cytology classified by a recognised system. Patients with previously known thyroid cancer and those receiving thyroid hormone or antithyroid medication at the time of measurement were to be excluded by the primary study or accounted for in its analysis. Prospective cohort, retrospective cohort, cross-sectional and case-control designs were eligible.

Reviews, guidelines, editorials, case reports and conference abstracts without a full report were excluded, as were treatment studies of radiofrequency ablation, percutaneous ethanol injection or TSH-suppression therapy, in which the analyte is manipulated rather than observed. Studies confined to disease-defined populations, such as cohorts with acromegaly, were excluded because the relationship between thyroid growth and malignancy in those settings is confounded by other growth-factor axes. Studies that measured TSH or FT4 but did not analyse either against the malignancy outcome were excluded on the grounds of index test rather than population. Reports were restricted to English-language full papers, with no date restriction.

Information sources and search

Five databases were searched to 17 July 2026: PubMed/MEDLINE, Embase, Scopus, Web of Science and Cochrane CENTRAL. Controlled vocabulary and free-text terms for the population, the index tests and the outcome were combined with Boolean operators, and review publication types were excluded. The full strings, with database-specific settings, are reproduced in the accompanying

search-strategy document. Reference lists of retrieved reports were inspected for additional eligible studies.

Study selection

Records were exported and de-duplicated by PMID and DOI, and four further records were removed before screening as retractions or non-English reports. Titles and abstracts were screened independently against the eligibility criteria, and records surviving that stage were sought for retrieval and assessed in full. Full-text exclusions were recorded against pre-specified reason categories, shown in Table 1. Screening and full-text assessment were conducted by the human review team, with disagreements resolved by discussion.

Data extraction

One reviewer extracted, into a piloted form, the first author, year, country, setting, design, enrolment period, sample size, number and proportion of malignant nodules, reference standard, TSH assay and reporting units, TSH values by diagnostic group, any TSH thresholds and stratum-specific estimates, FT4 values where reported, the covariates entered into any adjusted model, and any reported relationship between TSH and tumour stage or size. Values not reported by a study were recorded as not reported and were never imputed or back-calculated. Extracted values require confirmation against the source reports by a second reviewer before submission.

Risk of bias

QUADAS-2 was applied to every included study, since each evaluates an index test against a reference standard.¹² Risk of bias was judged across four domains: patient selection, index test, reference standard, and flow and timing. Applicability concerns were considered for the first three. Each domain was rated Low, High or Unclear, and the overall judgment followed the domain pattern, with any High domain producing an overall High rating. Judgments recorded here were derived from the published reports available at the time of the build and are provisional pending confirmation against full texts.

Synthesis

Findings were synthesised narratively and organised by outcome: discrimination by TSH, discrimination by FT4, independence of TSH after adjustment, and the relationship between TSH and

tumour stage or size. Within each outcome the synthesis reports the number of contributing studies, the direction and range of estimates, the pattern of results in relation to study characteristics, the studies diverging from the dominant direction with likely explanations, and an overall consistency judgment. The included studies expressed results as mean or median differences between diagnostic groups, as stratum-specific odds ratios using cut-points chosen by each investigator, and as thresholds derived from receiver operating characteristic analysis within each cohort. These forms are not convertible to a shared effect measure, and the reference standards differ, so the synthesis is qualitative throughout.

Certainty of evidence

Certainty was rated per outcome using GRADE,¹³ applying the adaptation developed for questions of diagnostic test accuracy, in which accuracy is treated as a surrogate for patient-important outcomes and indirectness is assessed accordingly.¹⁴ Bodies of observational evidence started at Low certainty and were downgraded for risk of bias, inconsistency, indirectness, imprecision and publication bias, with each decision justified from the extracted data. No outcome met the criteria for upgrading.

RESULTS

Search results and study selection

The searches identified 608 records across the five databases: 128 from PubMed/MEDLINE, 174 from Embase, 151 from Scopus, 116 from Web of Science and 39 from Cochrane CENTRAL. After 219 duplicate records were merged by PMID and DOI and a further 4 were removed before screening as retractions or non-English records, 385 records were screened at title and abstract. Of these, 316 were excluded as not addressing serum TSH or FT4 as a predictor of verified thyroid malignancy in an adult nodule cohort. Sixty-nine reports were sought for retrieval, of which 3 could not be obtained, leaving 66 assessed in full. Fifty-four full-text reports were excluded, with reasons: wrong population (13), wrong intervention or exposure (10), wrong comparator (6), relevant outcomes not reported (9), ineligible study design (8), duplicate or overlapping population (5), and

insufficient data for synthesis (3). Twelve studies met all criteria and entered the narrative synthesis. The flow is summarised in Table 1.

Table 1. PRISMA 2020 flow of records through the review.

Stage	n	Notes
Records identified - PubMed/MEDLINE	128	Searched to 17 July 2026
Records identified - Embase	174	Emtree plus free text; searched to 17 July 2026
Records identified - Scopus	151	TITLE-ABS-KEY; searched to 17 July 2026
Records identified - Web of Science	116	Core Collection, Topic field; searched to 17 July 2026
Records identified - Cochrane CENTRAL	39	Trials; searched to 17 July 2026
Records identified - total	608	Sum across the five databases
Duplicate records removed	219	Merged by PMID and DOI
Records removed for other reasons	4	Retractions and non-English records flagged before screening
Records screened (title/abstract)	385	608 identified minus 219 duplicates minus 4 removed
Records excluded at title/abstract	316	Not serum TSH or FT4 as a predictor of verified malignancy in an adult nodule cohort
Reports sought for retrieval	69	385 screened minus 316 excluded
Reports not retrieved	3	Full text unobtainable
Full-text reports assessed for eligibility	66	69 sought minus 3 not retrieved
Full-text reports excluded, with reasons	54	Wrong population 13; wrong intervention or exposure 10; wrong comparator 6; relevant

	outcomes not reported 9; ineligible study design 8; duplicate or overlapping population 5; insufficient data for synthesis 3
Studies included in narrative synthesis	12 10,497 adults with thyroid nodules

Note. Flow follows PRISMA 2020. All five core databases (PubMed/MEDLINE, Embase, Scopus, Web of Science and Cochrane CENTRAL) were searched to 17 July 2026 using the strings in the accompanying search-strategy document. Independent duplicate screening by a second reviewer is outstanding and is recorded in the audit trail.

Characteristics of included studies

The characteristics of the twelve included studies are shown in Table 2. Together they contributed 10,497 adults with thyroid nodules and enrolled patients between 1984 and 2021. Seven countries were represented: the United Kingdom, the United States, Brazil, Turkey, China, Saudi Arabia and Syria. Three studies were prospective cohorts and the remainder were retrospective cohorts or cross-sectional analyses of retrospectively assembled cohorts. Sample sizes ranged from 143 to 2,672.

Two features of the included cohorts bear directly on interpretation. First, the reference standard differed. Six studies verified diagnosis by surgical histopathology alone, applying it to patients who proceeded to thyroidectomy; four used cytology, alone or combined with histology, across the whole aspirated cohort; and one verified diagnosis by histology in operated patients and by a minimum of two years of clinical follow-up, mean 9.5 years, in the remainder.¹⁵ Second, the point at which patients entered the cohort differed. Studies recruiting at the time of fine-needle aspiration captured the full spectrum of nodules presenting for assessment, while those recruiting at the time of surgery captured a spectrum already filtered by the very risk factors under study. The reported proportion of malignancy ranged from 4.1% in a cytology-based aspiration cohort¹⁸ to a majority of nodules in a surgical series assembled from a hospital data centre,²² a spread that reflects selection rather than any real difference in underlying disease frequency.

Table 2. Characteristics of the twelve included studies.

Study	Country	Design	n	Malignant, n (%)	Reference standard	FT4 reported
Boelaert 2006[15]	United Kingdom	Prospective cohort	1,183	Not separately reported for the TSH subgroup	Histology or ≥ 2 -year follow-up	No
Haymart 2008[16]	United States	Retrospective cohort	843	241 (29%)	Histology	No
Maia 2011[17]	Brazil	Retrospective cohort	143	50 (35%)	Histology	Yes, analysed
Duran 2013[18]	Turkey	Retrospective cohort	2,672	110 (4.1%)	Cytology	No
Azizi 2014[19]	United States	Prospective cohort	2,023	233 patients (248 malignant nodules)	Histology	Yes, measured only
Figuera 2015[20]	Brazil	Retrospective cohort	622	Not separately reported	Cytology and/or histology	Yes, measured only
Golbert 2017[21]	Brazil	Prospective cohort	615	47 of 160 operated (29.4%)	Anatomopathological examination of thyroidectomy tissue	No
Miao 2020[22]	China	Retrospective cohort	811	784 malignant of 1290 nodules	Pathological confirmation	Yes, analysed

Ruhaiye m 2023[23]	Saudi Arabia	Retrospective cohort	263	17.5% carcinoma after thyroidectomy	Histology	No
Alaraifi 2023[24]	Saudi Arabia	Retrospective cohort	421	Not separately reported	Final histopathology	No
Shahrokh 2023[25]	Syria	Cross- sectional (retrospective)	305	65 (21.3%)	Histology	No
Sah Unal 2023[26]	Turkey	Retrospective cohort (matched comparison)	596	42/296 (14.2%) macrocalcified ; 16/300 (5.3%) control	Cytology	No

Note. n is the number of patients contributing a pre-diagnostic TSH measurement, except in Miao 2020 where 811 patients contributed 1,290 nodules. Reference standard is the standard applied to the analysed cohort. FT4 reported distinguishes studies that analysed FT4 against the malignancy outcome from those that measured it only to establish thyroid status.

Risk of bias

Risk of bias, assessed with QUADAS-2, is summarised in Table 3. Nine of the twelve studies were judged at high overall risk of bias and three at unclear risk; none reached low risk across all four domains. Patient selection was the dominant concern, rated high in eight studies, in each case because the analysed cohort comprised patients who had already been selected for surgery. In such a design the index test contributes to the decision that determines cohort entry, and the resulting spectrum is not the spectrum in which the test would be applied. The reference standard domain was rated high in the four studies that verified diagnosis by cytology alone, since cytology is an imperfect standard whose own accuracy has been reported at 88% sensitivity and 84% specificity in one of the included

cohorts.¹⁵ The index test domain was more often unclear than high: most reports did not state whether TSH was interpreted without knowledge of the cytological or histological result, and none pre-specified the threshold used, with several deriving it from the same data in which it was then evaluated. Flow and timing was rated low in most studies, reflecting the short and well-defined interval between the blood draw and the diagnostic procedure.

Table 3. QUADAS-2 risk-of-bias assessment of the included studies.

Study	Design	Tool	D1 Patient selection	D2 Index test	D3 Reference standard	D4 Flow and timing	Overall
Boelaert 2006[15]	Prospective cohort	QUADAS- 2	Unclear	Unclear	Low	Low	Unclear
Haymart 2008[16]	Retrospective cohort	QUADAS- 2	High	Unclear	Low	Low	High
Maia 2011[17]	Retrospective cohort	QUADAS- 2	High	Unclear	Low	Low	High
Duran 2013[18]	Retrospective cohort	QUADAS- 2	Low	Unclear	High	Low	High
Azizi 2014[19]	Prospective cohort	QUADAS- 2	Unclear	Unclear	Low	Low	Unclear
Fighera 2015[20]	Retrospective cohort	QUADAS- 2	Low	Unclear	High	Low	High
Golbert 2017[21]	Prospective cohort	QUADAS- 2	Low	High	Low	Low	High
Miao 2020[22]	Retrospective cohort	QUADAS- 2	High	Unclear	Low	Unclear	High

Ruhaiyem 2023[23]	Retrospective cohort	QUADAS- 2	High	Unclear	High	Unclear	High
Alaraifi 2023[24]	Retrospective cohort	QUADAS- 2	High	Unclear	Low	Low	High
Shahrokh 2023[25]	Cross- sectional (retrospective)	QUADAS- 2	High	Unclear	Low	Low	High
Sah Unal 2023[26]	Retrospective cohort (matched comparison)	QUADAS- 2	High	Unclear	High	Low	High

Note. Risk of bias assessed with QUADAS-2. D1 = patient selection, D2 = index test, D3 = reference standard, D4 = flow and timing. Overall follows the QUADAS-2 algorithm: any domain rated High produces an overall High rating; an overall rating of Unclear indicates no High domain but at least one Unclear domain. Domain-level signalling questions, judgments and rationales are given in the RoB_Detail sheet of the accompanying workbook. Judgments are provisional pending confirmation against full texts.

Serum TSH and the discrimination of malignant from benign nodules

All twelve studies reported serum TSH against the malignancy outcome. Nine found higher TSH in patients with malignant nodules and three found no association, so the dominant direction is positive but not uniform. The magnitude of the association, where it could be quantified, varied by almost an order of magnitude. The largest estimates came from the single cohort that stratified TSH most finely: adjusted odds ratios rose from 2.72 (95% CI 1.02-7.27) at TSH 1.0-1.7 mU/L, to 3.88 (95% CI 1.48-10.19) at 1.8-5.5 mU/L, to 11.18 at values above 5.5 mU/L, each compared with TSH below 0.4 mU/L.¹⁵ At the other end, the two studies reporting a single adjusted odds ratio across the whole range found much smaller effects: 1.49 (95% CI 1.09-2.03) for TSH at or above 1 microIU/mL,¹⁹ and 1.54 for higher TSH in a euthyroid surgical cohort.²⁴ This gradient is not a contradiction. It is what a nonlinear, monotonic relationship looks like when different investigators

choose different contrasts, and it means that the reported odds ratio in any single study is largely a function of how extreme its reference and comparison strata were.

The absolute risks reported in the largest surgical series make the same point in more clinically legible terms. Malignancy was found in 16% (9 of 55) of patients with TSH below 0.06 mIU/L and in 52% (15 of 29) of those with TSH at or above 5.00 mIU/L.¹⁶ Between those extremes the separation collapsed: 25% (85 of 347) at TSH 0.40-1.39 mIU/L against 35% (109 of 308) at TSH 1.40-4.99 mIU/L. The comparison retained statistical significance ($P=0.002$) on 655 patients, but a ten-percentage-point difference in malignancy risk across the interval in which the great majority of patients fall is not a difference that redirects management.

Three studies reported no association, and their characteristics are informative rather than anomalous. A surgical cohort of 143 patients found no difference in TSH between benign and malignant groups and retained only age, border irregularity, microcalcifications and nodule size in its multivariate model.¹⁷ A primary-care aspiration cohort of 263 patients found no association between TSH and cytology, with a mean TSH of 2.3 mU/L and a standard deviation of 5.9 that indicates a markedly skewed distribution.²³ A hospital data-centre analysis of 1,290 nodules found no significant effect of TSH at all.²² The first two are the smallest studies in the set and the third had the most atypical case mix, with malignant nodules outnumbering benign. The null findings therefore cluster where power was lowest or where selection was most extreme, which weakens rather than refutes the positive direction seen elsewhere, but also shows that the association is not robust to how a cohort is assembled.

The thresholds proposed by individual studies did not converge. Reported inflection points included TSH above 0.9 mU/L,¹⁵ 1 microIU/mL,¹⁹ 1.64 mU/L,²⁰ and 2.26 microU/mL from receiver operating characteristic analysis,²¹ and one study reported a stepwise increase across Bethesda categories without nominating a cut-point.²⁴ These values span more than a twofold range and, importantly, each was derived within the cohort in which it was then evaluated. No included study validated a threshold proposed by another. This is the central practical finding of the review: the direction of the association replicates, but the operating point does not, and a test without a transferable threshold cannot be applied prospectively to a new patient.

Consistency for this outcome is therefore judged moderate for direction and low for magnitude. Nine of twelve studies agree on the sign of the effect, but the estimates cannot be reconciled to a single number, and the discrepancies track study design features rather than random error.

Serum free thyroxine

The evidence for FT4 is much thinner than the review question anticipated. FT4 was measured in four of the twelve studies but analysed against the malignancy outcome in only two, and those two disagree. A surgical cohort of 143 patients reported no difference in FT4 between benign and malignant nodules.¹⁷ A hospital data-centre analysis of 1,290 nodules reported that FT4, together with thyroglobulin and free triiodothyronine, had a significant effect on malignancy ($P=0.003$), in the same model in which TSH did not.²² The remaining two studies recorded FT4 as part of the thyroid function panel used to establish euthyroid status but did not carry it into any analysis of the outcome.^{19,20}

The one study reporting discrimination for a thyroid-function model in quantitative terms found sensitivity of 80.6% and specificity of 55.8%, against sensitivity of 89.0% and specificity of 89.9% for ultrasound risk stratification in the same nodules.²² A specificity near 56% would classify almost half of benign nodules as suspicious, which in a population where roughly 90% of nodules are benign would generate a large excess of unnecessary aspirations. That single comparison is the only direct head-to-head estimate available, and it comes from the cohort with the most atypical case mix in the review.

No conclusion about FT4 can be drawn from two discordant studies, one of which did not isolate FT4 from co-entered analytes. Consistency is low and the evidence base is insufficient rather than merely uncertain. This finding is worth stating plainly because the absence of analysis, rather than an absence of measurement, is what limits inference: FT4 values exist in these datasets and could be analysed without new data collection.

Independence of TSH after adjustment

Six studies fitted multivariable models that could establish whether TSH is independent of established predictors. TSH was retained in four. In a cohort of 1,183 patients it remained significant after adjustment for sex, age and goitre type, with the adjusted odds ratios described above.¹⁵ In 2,023

prospectively enrolled patients it was retained after adjustment for age and sex, at an odds ratio of 1.49 (95% CI 1.09-2.03), alongside antithyroglobulin antibody positivity at an odds ratio of 2.24 (95% CI 1.57-3.19).¹⁹ In 615 patients undergoing aspiration, elevated TSH was retained together with family history of thyroid cancer, microcalcifications and solitary nodule.²¹ In a surgical cohort of 843 it was retained on multivariable analysis ($P=0.007$).¹⁶

TSH was not retained in two models, and the pattern of what displaced it is consistent. In the 143-patient surgical series the model retained age above 39 years, border irregularity, microcalcifications and nodule size above 2 cm, reaching 81.7% accuracy without TSH.¹⁷ In the matched macrocalcification study, younger age, higher TI-RADS category and nodule location survived multivariable adjustment while TSH did not.²⁶ In both, sonographic features displaced the biochemical marker. The comparison in the antibody-adjusted cohort points the same way from a different angle: antithyroglobulin antibody carried a larger adjusted odds ratio than TSH in the same model,¹⁹ which is congruent with the meta-analytic observation that studies controlling for autoimmunity report the weakest TSH associations.¹⁰

One report requires a note of caution. The macrocalcification study states that TSH levels of malignant nodules were higher than those of benign nodules, and elsewhere that TSH level was not associated with malignancy in univariate analysis.²⁶ These statements are difficult to reconcile from the published report, and the extraction has been flagged for verification against the full text. The study is retained because its multivariable finding, that TSH was not an independent predictor, is unambiguous and is the value used in this synthesis.

TSH and tumour stage or size

Three studies examined TSH in relation to tumour extent among malignant nodules, and all three found a positive relationship. Mean TSH was 4.9 (SD 1.5) mIU/L in patients with stage III or IV disease against 2.1 (SD 0.2) mIU/L in stage I or II.¹⁶ TSH increased stepwise from benign nodular disease through papillary microcarcinoma to larger cancers ($P=0.001$), and higher TSH was associated with lymph node metastases ($P=0.008$).²⁵ TSH correlated with tumour size in a third cohort ($P=0.005$).²⁰ Consistency for this outcome is high in direction across three studies, though the total

number of malignant cases contributing is modest and no study adjusted these comparisons for the duration of nodule surveillance, which plausibly confounds both TSH exposure and tumour size.

This secondary observation is the more coherent signal in the review, and it points away from a diagnostic use. A marker that rises with tumour extent among cancers already diagnosed behaves like a correlate of disease burden or of prolonged thyrotropic stimulation, not like a test that separates cancer from non-cancer at first presentation. The qualitative summary of all four outcomes appears in Table 4 and the certainty assessment in Table 5.

Table 4. Qualitative evidence summary by study and outcome.

Study	Outcome	Effect direction	Effect magnitude	Certainty	Notes
Boelaert 2006[15]	TSH: malignant vs benign	up	AOR 2.72-11.18 across ascending TSH strata	Very low	Unclear risk of bias; United Kingdom
Haymart 2008[16]	TSH: malignant vs benign	up	16% to 52% absolute malignancy risk across TSH range	Very low	High risk of bias; United States
Maia 2011[17]	TSH: malignant vs benign	null	No significant between-group difference reported	Very low	High risk of bias; Brazil
Duran 2013[18]	TSH: malignant vs benign	up	Higher TSH in malignant cytology (P=0.017); effect size not reported	Very low	High risk of bias; Turkey
Azizi 2014[19]	TSH: malignant vs benign	up	OR 1.49 (95% CI 1.09-2.03) for TSH ≥ 1 microIU/mL	Very low	Unclear risk of bias; United States

Fighera 2015[20]	TSH: malignant vs benign	up	Significant risk increase above TSH 1.64 mU/L (P<0.001)	Very low	High risk of bias; Brazil
Golbert 2017[21]	TSH: malignant vs benign	up	Median TSH 2.25 vs 1.50 microU/mL; ~3-fold risk above 2.26 microU/mL	Very low	High risk of bias; Brazil
Miao 2020[22]	TSH: malignant vs benign	null (TSH); up (FT4)	TSH P>0.05; FT4 P=0.003; thyroid- function model sensitivity 80.6%, specificity 55.8%	Very low	High risk of bias; China
Ruhaiyem 2023[23]	TSH: malignant vs benign	null	No significant association reported	Very low	High risk of bias; Saudi Arabia
Alaraifi 2023[24]	TSH: malignant vs benign	up	TSH 1.94 vs 1.62; OR 1.54 (P=0.038)	Very low	High risk of bias; Saudi Arabia
Shahrokh 2023[25]	TSH: malignant vs benign	up	Stepwise TSH increase across benign, microcarcinoma and larger cancer (P=0.001)	Very low	High risk of bias; Syria
Sah Unal 2023[26]	TSH: malignant vs benign	mixed	Higher TSH reported in malignant nodules; not retained on multivariate analysis	Very low	High risk of bias; Turkey

Maia 2011[17]	FT4: malignant vs benign	null	No significant between- group difference	Very low	High risk of bias; n=143
Miao 2020[22]	FT4: malignant vs benign	up	P=0.003, co-entered with thyroglobulin and FT3; thyroid-function model sensitivity 80.6%, specificity 55.8%	Very low	High risk of bias; FT4 not isolated from co-entered analytes

Note. Effect direction is reported relative to malignancy: up = higher analyte in malignant nodules; null = no significant difference; mixed = internally discordant report. Values are as extracted from the published reports and are not converted to a common metric, since the reported forms (mean or median difference, stratum-specific odds ratio, ROC-derived threshold) are not interconvertible. Certainty is the GRADE rating for the outcome as a body of evidence, not for the individual study.

Table 5. GRADE summary of findings.

Outcome	Studies (n)	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Certainty
TSH discriminating malignant from benign nodules	12 (10,497)	Observational	Serious (downgraded 1): 9/12 high risk, patient selection dominant	Serious (downgraded 1): 3/12 null; magnitude 1.49-11.18 not reconcilable	Not serious: population and testing match the review question	Not serious: large total sample; individual CIs mostly exclude the null	Suspected: positive hypothesis long dominant; small null studies scarce	Very low

FT4 discriminating malignant from benign nodules	2 (954)	Observational	Serious (downgraded 1): both studies high risk	Very serious (downgraded 2): the two studies reach opposite conclusions	Serious (downgraded 1): FT4 co-entered with thyroglobulin and FT3 in the positive study	Serious: only 2 of 12 studies analysed FT4 despite 4 measuring it	Suspected: FT4 measure d but unreported in 2 further cohorts	Very low
TSH independent of established predictors after adjustment	6 (5,403)	Observational	Serious (downgraded 1): high or unclear risk throughout	Serious (downgraded 1): retained in 4/6; covariate sets differ across models	Not serious	Not serious	Suspected	Very low
TSH in relation to tumour stage or size among	3 (1,770)	Observational	Serious (downgraded 1): high or unclear	Not serious: direction consistent across all	Serious (downgraded 1): a prognostic correlate,	Serious: few contributing malignant cases;	Suspected	Very low

malignant	risk in all	three	not	a	no
nodules	three	studies	diagnosti	adjustm	
			c	ent	for
			outcome	surveilla	
				nce	
				duration	

Note. Certainty rated with GRADE as adapted for diagnostic test accuracy. Bodies of observational evidence start at Low and are downgraded from there; no outcome met the criteria for upgrading. Studies (n) gives the number of studies and the total participants contributing to that outcome.

DISCUSSION

Across twelve observational studies and 10,497 adults, higher pre-diagnostic serum TSH was associated with thyroid malignancy in nine cohorts and showed no association in three. The direction of the association replicates reasonably well; its magnitude does not. Adjusted odds ratios ranged from 1.49 to 11.18, and this range is largely explained by how widely each study spaced its reference and comparison strata rather than by disagreement about the underlying relationship. The thresholds proposed for clinical use spanned 0.9 to 2.26 mU/L, every one was derived in the cohort in which it was assessed, and none has been validated externally. TSH remained an independent predictor in four of six multivariable models, and in the two where it did not, sonographic features displaced it. Evidence on FT4 is too sparse to support any conclusion: it was measured in four studies, analysed in two, and those two disagree. Certainty is very low for every outcome.

Comparison with prior evidence

These findings are consistent with, and help to interpret, the dose-response meta-analysis of 28 studies, 42,032 subjects and 5,786 cancer cases that established the population-level association.¹⁰ That analysis found the relationship to be nonlinear, with an odds ratio of 1.72 (95% CI 1.42-2.07) per mU/L below 1 mU/L flattening to 1.16 (95% CI 1.12-1.21) per mU/L above it. The present review shows what that shape implies at the bedside. Because most patients presenting with a nodule sit in

the flat portion of the curve, the incremental information available from their TSH value is small, and this is exactly what the included studies report: a ten-percentage-point separation in malignancy risk between TSH 0.40-1.39 and 1.40-4.99 mIU/L,¹⁶ against a 36-point separation only when the extreme tails are contrasted. The prior meta-analysis answered whether TSH is associated with thyroid cancer. This review answers a different question, whether that association discriminates, and the answer is that it does so weakly and inconsistently.

The earlier synthesis also noted that studies controlling for thyroid autoimmunity reported the lowest odds ratios, with the estimate at TSH values of 2.5 mU/L or above falling to 0.98 (95% CI 0.89-1.09) per mU/L.¹⁰ The one included study here that adjusted for thyroid antibodies found antithyroglobulin antibody positivity to carry a larger adjusted odds ratio than TSH itself, while thyroid peroxidase antibody was not associated.¹⁹ Autoimmune thyroid disease raises TSH and has been linked, with continuing disagreement, to thyroid cancer risk.²⁷ Some studies of aspirated cohorts have found no excess malignancy in patients with Hashimoto thyroiditis.²⁸ The plausible reading is that thyroid autoimmunity accounts for part of the TSH signal, and that adjustment for it is one of the design choices separating the positive from the null studies here.

Possible mechanisms

Two mechanisms could produce the observed association, and they carry different clinical implications. The first is causal: TSH acts on the TSH receptor to drive follicular cell proliferation, and this signalling is preserved in differentiated thyroid carcinoma, so sustained thyrotropic stimulation could promote the growth of transformed clones.⁹ The stepwise rise in TSH from benign disease through microcarcinoma to larger tumours,²⁵ the association with stage,¹⁶ and the correlation with tumour size²⁰ all fit a growth-promoting role. Related observations, including the smaller thyroid volume and lower nodule frequency reported in patients treated with metformin,⁹ and the associations between obesity, insulin resistance and thyroid proliferation,²⁹ are congruent with a proliferative axis in which TSH is one input among several.

The second mechanism is confounding by thyroid architecture. Malignant nodules in these cohorts were consistently smaller and occurred in smaller glands: nodules above 4 cm were more common among benign than malignant lesions (43.1% against 21.1%), and larger nodule size was

protective (OR 0.760, $P=0.004$).²⁴ Thyroid volume was lower in patients with malignant cytology.¹⁸ A gland with less functioning tissue produces less thyroxine and therefore, through negative feedback, runs a higher TSH. On this account TSH is elevated not because it causes cancer but because malignant nodules arise in glands with less compensatory reserve, and because coexisting thyroiditis both raises TSH and destroys functioning tissue. Both mechanisms predict the association that is observed; only the first would predict any benefit from intervening on TSH. The available data cannot separate them, and the review offers this as an interpretation to be tested rather than a conclusion.

Heterogeneity

The differences between studies are not random and are worth naming precisely, because they identify what a better study would have to do. The strongest determinant is who entered the cohort. Studies recruiting at surgery analyse a population already filtered by the risk factors under investigation, which compresses the spectrum and can move the estimate in either direction; studies recruiting at aspiration retain the spectrum in which the test would actually be used but verify most diagnoses by cytology, an imperfect standard. The second determinant is the reference standard itself: cytology-verified cohorts reported malignancy proportions as low as 4.1%,¹⁸ histology-verified surgical cohorts reported roughly 29%,^{16,21} and one data-centre series reported a majority.²² These are not different diseases but different denominators. The third is adjustment for autoimmunity, discussed above. The fourth is analytic: assay platform mattered enough that one study reported different median TSH values for the same patients on chemiluminescent and electrochemiluminescent immunoassay (2.25 against 2.33 microU/mL in malignant nodules, 1.50 against 1.27 in benign),²¹ which is a caution against transporting a numerical threshold between laboratories. Iodine status and the background prevalence of autoimmune disease differ across the seven countries represented and plausibly contribute further, though no included study reported iodine status.

Limitations

The limitations of this review are systematic rather than incidental. Restriction to English-language full reports risks language bias, which is of particular concern for a topic where iodine status varies geographically and much relevant work is published regionally. Extracted values are drawn from published reports and require confirmation against source documents. Beyond the review

process, the evidence base itself is limited: no included study was at low risk of bias on all QUADAS-2 domains, none pre-specified a TSH threshold, none validated a threshold proposed elsewhere, and none reported the incremental discrimination of TSH added to a model already containing ultrasound features, which is the only comparison that would establish clinical value. Publication bias is a realistic concern, since a positive TSH-cancer association has been the prevailing hypothesis for two decades and small null studies are less likely to have been written up or accepted.

Certainty of evidence

Certainty is very low for all four outcomes. For TSH discrimination, the body of evidence began at Low as observational data and was downgraded one level for risk of bias, since nine of twelve studies were at high risk with patient selection the dominant concern, and one further level for inconsistency, since three of twelve studies found no association and the magnitude of effect could not be reconciled across the remainder. For FT4 discrimination, downgrading was applied for risk of bias and for very serious inconsistency, two studies reporting opposite conclusions, with imprecision contributing further. For independence after adjustment, downgrading reflected risk of bias and inconsistent covariate sets across models. For the relationship with tumour extent, direction was consistent across three studies but risk of bias and imprecision, arising from the small number of contributing malignant cases, produced the same rating. Suspected publication bias contributed to every assessment. These ratings are consistent with the summary of findings in Table 5.

Implications for practice

The evidence does not support the use of serum TSH as a stand-alone triage test for thyroid nodules, and it does not support any specific threshold. The reasons are concrete: the association is weak across the range in which most patients sit, the proposed operating points span more than a twofold range and none has been externally validated, the value depends on the assay platform, and sonographic features displace TSH when both are offered to a multivariable model. Current guidance, which uses TSH to determine whether scintigraphy is warranted rather than to grade malignancy risk,⁶ remains consistent with the evidence assembled here. TSH is best understood as one weak contributor to an overall clinical impression, alongside age, sex, nodule number, sonographic pattern and thyroid antibody status, and it should not displace or defer cytology in a nodule that meets sonographic criteria

for aspiration.⁷³⁰ Where TSH is suppressed, its established role in identifying an autonomously functioning nodule, and thereby sparing an aspiration, remains the one use with a secure basis.

Implications for research

The most efficient next step requires no new patients. FT4 has been measured in cohorts that have already been assembled and published; analysing it against the malignancy outcome, and reporting it whether or not it is significant, would resolve a gap that currently rests on two discordant studies. The more demanding requirement is a prospective cohort recruiting at the point of aspiration rather than surgery, with a TSH threshold pre-specified from prior work rather than derived in the same data, verification of all nodules against a consistent reference standard, adjustment for thyroid antibodies and thyroid volume, reporting of iodine status and assay platform, and, critically, a comparison of discrimination for a model with and without TSH added to established sonographic features. Until such a study exists, the incremental value of TSH over what ultrasound already provides remains unmeasured. Given the scale of thyroid cancer overdiagnosis,⁴⁵³¹³² any proposal to widen the diagnostic net on the basis of a weak biochemical marker should be held to that standard.

CONCLUSION

Higher pre-diagnostic serum TSH is associated with thyroid malignancy in most, though not all, adult nodule cohorts, but the association is weak across the range in which most patients present, no proposed threshold has been reproduced outside the cohort that generated it, and the evidence for free thyroxine is too sparse to judge; on very low certainty evidence, TSH should be treated as one weak contributor within a multivariable clinical assessment rather than as an independent diagnostic criterion.

REFERENCES

1. Guth S, Theune U, Aberle J, Galach A, Bamberger CM. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur J Clin Invest.* 2009;39(8):699-706. PMID: 19601965; doi: 10.1111/j.1365-2362.2009.02162.x.

2. 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. PMID: 29509871; doi: 10.1001/jama.2018.0898.
3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-49. PMID: 33538338; doi: 10.3322/caac.21660.
4. Vaccarella S, Franceschi S, Bray F, Wild CP, Plummer M, Dal Maso L. Worldwide thyroid-cancer epidemic? The increasing impact of overdiagnosis. *N Engl J Med*. 2016;375(7):614-7. PMID: 27532827; doi: 10.1056/NEJMp1604412.
5. Dal Maso L, Panato C, Franceschi S, Serraino D, Buzzoni C, Busco S, et al. The impact of overdiagnosis on thyroid cancer epidemic in Italy, 1998-2012. *Eur J Cancer*. 2018;94:6-15. PMID: 29502036; doi: 10.1016/j.ejca.2018.01.083.
6. 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. *Thyroid*. 2016;26(1):1-133. PMID: 26462967; doi: 10.1089/thy.2015.0020.
7. 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. PMID: 28372962; doi: 10.1016/j.jacr.2017.01.046.
8. Ali SZ, Baloch ZW, Cochand-Priollet B, Schmitt FC, Vielh P, VanderLaan PA. The 2023 Bethesda System for Reporting Thyroid Cytopathology. *Thyroid*. 2023;33(9):1039-44. PMID: 37427847; doi: 10.1089/thy.2023.0141.
9. Meng X, Xu S, Chen G, Derwahl M, Liu C. Metformin and thyroid disease. *J Endocrinol*. 2017;233(1):R43-51. PMID: 28196954; doi: 10.1530/JOE-16-0450.
10. McLeod DSA, Watters KF, Carpenter AD, Ladenson PW, Cooper DS, Ding EL. Thyrotropin and thyroid cancer diagnosis: a systematic review and dose-response meta-analysis. *J Clin Endocrinol Metab*. 2012;97(8):2682-92. PMID: 22622023; doi: 10.1210/jc.2012-1083.

11. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. PMID: 33782057; doi: 10.1136/bmj.n71.
12. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-36. PMID: 22007046; doi: 10.7326/0003-4819-155-8-201110180-00009.
13. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011;64(4):401-6. PMID: 21208779; doi: 10.1016/j.jclinepi.2010.07.015.
14. Brozek JL, Akl EA, Jaeschke R, Lang DM, Bossuyt P, Glasziou P, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines: part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. *Allergy*. 2009;64(8):1109-16. PMID: 19489757; doi: 10.1111/j.1398-9995.2009.02083.x.
15. Boelaert K, Horacek J, Holder RL, Watkinson JC, Sheppard MC, Franklyn JA. Serum thyrotropin concentration as a novel predictor of malignancy in thyroid nodules investigated by fine-needle aspiration. *J Clin Endocrinol Metab*. 2006;91(11):4295-301. PMID: 16868053; doi: 10.1210/jc.2006-0527.
16. Haymart MR, Repplinger DJ, Levenson GE, Elson DF, Sippel RS, Jaume JC, et al. Higher serum thyroid stimulating hormone level in thyroid nodule patients is associated with greater risks of differentiated thyroid cancer and advanced tumor stage. *J Clin Endocrinol Metab*. 2008;93(3):809-14. PMID: 18160464; doi: 10.1210/jc.2007-2215.
17. Maia FFR, Matos PS, Silva BP, Pallone AT, Pavin EJ, Vassallo J, et al. Role of ultrasound, clinical and scintigraphic parameters to predict malignancy in thyroid nodule. *Head Neck Oncol*. 2011;3:17. PMID: 21426548; doi: 10.1186/1758-3284-3-17.
18. Duran AO, Anil C, Gursoy A, Nar A, Altundag O, Inanc M, et al. The relationship between thyroid volume and malignant thyroid disease. *Med Oncol*. 2013;31(1):814. PMID: 24338169; doi: 10.1007/s12032-013-0814-2.

19. Azizi G, Keller JM, Lewis M, Piper K, Puett D, Rivenbark KM, et al. Association of Hashimoto's thyroiditis with thyroid cancer. *Endocr Relat Cancer*. 2014;21(6):845-52. PMID: 25217233; doi: 10.1530/ERC-14-0258.
20. Fighera TM, Perez CLS, Faris N, Scarabotto PC, da Silva TT, Cavalcanti TCS, et al. TSH levels are associated with increased risk of thyroid carcinoma in patients with nodular disease. *Endokrynol Pol*. 2015;66(6):480-5. PMID: 26662646; doi: 10.5603/EP.a2015.0059.
21. Golbert L, de Cristo AP, Faccin CS, Farenzena M, Folgieri H, Graudenz MS, et al. Serum TSH levels as a predictor of malignancy in thyroid nodules: a prospective study. *PLoS One*. 2017;12(11):e0188123. PMID: 29145466; doi: 10.1371/journal.pone.0188123.
22. Miao S, Jing M, Sheng R, Cui D, Lu S, Zhang X, et al. The analysis of differential diagnosis of benign and malignant thyroid nodules based on ultrasound reports. *Gland Surg*. 2020;9(3):653-60. PMID: 32775255; doi: 10.21037/gs.2020.04.03.
23. Ruhaiyem ME, Al Khenizan A, Hussain A. The yield of thyroid fine-needle aspiration in a primary care setting in Riyadh, Saudi Arabia. *J Family Med Prim Care*. 2023;11(12):7602-6. PMID: 36994071; doi: 10.4103/jfmpc.jfmpc_2132_21.
24. Alaraifi AK, Alessa M, Hijazi LO, Alayed AM, Alsalem AA. TSH level as a risk factor of thyroid malignancy for nodules in euthyroid patients. *Acta Otorhinolaryngol Ital*. 2023;43(3):183-8. PMID: 37204842; doi: 10.14639/0392-100X-N2288.
25. Shahrokh M, Alsultan M, Kabalan Y. The relationship between papillary thyroid carcinoma and preoperative TSH level: a cross-sectional study from Syria. *Medicine (Baltimore)*. 2023;102(28):e34283. PMID: 37443517; doi: 10.1097/MD.00000000000034283.
26. Sah Unal FT, Gokcay Canpolat A, Elhan AH, Sevim S, Sak SD, Emral R, et al. Cancer rates and characteristics of thyroid nodules with macrocalcification. *Endocrine*. 2023;84(3):1021-9. PMID: 38147262; doi: 10.1007/s12020-023-03650-x.
27. Penta L, Cofini M, Lanciotti L, Leonardi A, Principi N, Esposito S. Hashimoto's disease and thyroid cancer in children: are they associated? *Front Endocrinol (Lausanne)*. 2018;9:565. PMID: 30356680; doi: 10.3389/fendo.2018.00565.

28. Grani G, Calvanese A, Carbotta G, D'Alessandri M, Nesca A, Bianchini M, et al. Thyroid autoimmunity and risk of malignancy in thyroid nodules submitted to fine-needle aspiration cytology. *Head Neck*. 2015;37(2):260-4. PMID: 24375752; doi: 10.1002/hed.23587.
29. Demetriou E, Fokou M, Frangos S, Papageorgis P, Economides PA, Economides A. Thyroid nodules and obesity. *Life (Basel)*. 2023;13(6):1292. PMID: 37374075; doi: 10.3390/life13061292.
30. Giovanella L, Campenni A, Tuncel M, Petranovic Ovcariček P. Integrated diagnostics of thyroid nodules. *Cancers (Basel)*. 2024;16(2):311. PMID: 38254799; doi: 10.3390/cancers16020311.
31. Acosta GJ, Singh Ospina N, Brito JP. Epidemiologic changes in thyroid disease. *Curr Opin Endocrinol Diabetes Obes*. 2024;31(5):184-90. PMID: 39087407; doi: 10.1097/MED.0000000000000877.
32. Li M, Zheng R, Dal Maso L, Zhang S, Wei W, Vaccarella S. Mapping overdiagnosis of thyroid cancer in China. *Lancet Diabetes Endocrinol*. 2021;9(6):330-2. PMID: 33891886; doi: 10.1016/S2213-8587(21)00083-8.