REVIEW

Thyroid nodules: a review of current guidelines, practices, and prospects

H Gharib, E Papini¹ and R Paschke²

Division of Endocrinology, Diabetes, Metabolism and Nutrition, Mayo Clinic, 200 First, Street SW, Rochester, Minnesota 55905, USA, ¹Department of Endocrine and Metabolic Diseases, Regina Apostolorum Hospital, Albano Laziale 00041, Italy and ²III. Medical Department, University of Leipzig, Ph.-Rosenthal-Street 27, D-04103 Leipzig, Germany

(Correspondence should be addressed to R Paschke; Email: pasr@medizin.uni-leipzig.de)

Abstract

In 2006, two major society-sponsored guidelines and one major consensus statement for thyroid diagnosis and management were published by: the American Association of Clinical Endocrinologists/Associazione Medici Endocrinologi (AACE/AME); the American Thyroid Association (ATA); and the European Thyroid Association (ETA). A careful review of these guidelines reveals that despite many similarities, significant differences are also present, likely reflecting differences in practice patterns, interpretation of existing data, and availability of resources in different regions. The methodology of the guidelines is similar, but a few differences in the rating scale make a rapid comparison of the strength of both evidence and recommendations difficult for the use in current clinical practice. Some recommendations are based mostly on experts' opinion. Thus, a same recommendation may be based on a different evidence; on the other hand, sometimes the same evidence may induce a different recommendation. Therefore, efforts are needed to produce a few high-quality clinical studies to close the evidence gaps in the still controversial fields of thyroid disease and to create a joint task force of the most authoritative societies in the field of thyroid disease in order to reach a common document for clinical practice recommendations.

European Journal of Endocrinology 159 493–505

Introduction

Opinion regarding the optimal approach to the diagnostic evaluation and clinical management of thyroid nodules remains diverse, changing, and controversial. Clinical guidelines, developed by a panel of experts using new scientific data, attempt to bring consistency into clinical practice. Guidelines are designed to facilitate medical decision making.

In 2006, two major society-sponsored guidelines and one major consensus statement for thyroid diagnosis and management were published by: the American Association of Clinical Endocrinologists/Associazione Medici Endocrinologi (AACE/AME); the American Thyroid Association (ATA); and the European Thyroid Association (ETA) respectively (1–3). These documents take into account the best available published evidence, incorporating the experience and expertise of the writing panel. A careful review of these guidelines and the consensus statement reveals that despite many similarities, significant differences are also present, likely reflecting differences in practice patterns, interpretation of existing data, and availability of resources in different regions. Thus, the rationale for the ETA to sponsor a symposium at its 32nd annual meeting is to address some of these issues.

Epidemiology and risks of benign and malignant thyroid diseases

Thyroid nodules are very common. Moreover, with the increasing use of sensitive imaging techniques, an increasing proportion of thyroid nodules are now incidentally detected. A prospective study comparing clinical examination and ultrasound (US) showed that 46% of the nodules (> 1 cm) detected by US did escape detection by clinical examination (4). Autopsy and prospective US studies in North America detected asymptomatic thyroid nodules in 50 and 67% respectively (5, 6). A population study in Germany, a previously iodine-deficient and currently borderline-iodine-sufficient country, detected thyroid nodules by US in 20% of the population aged 20–79 years. The prevalence increased with increasing age to 52 and 29% for women and men aged 70–74 respectively (7).

In areas not affected by nuclear fall out, the annual incidence of thyroid cancer has been reported to range between 1.2 and 2.6 cases per 100,000 in men and 2.0–3.8 cases per 100,000 in women with higher incidences in countries like Sweden, France, Japan, and USA (8). An increase in thyroid cancer incidence from 3.6 per 100,000 in 1973 to 8.7 per 100,000 in 2002

has recently been reported for the USA (9). A similar increase in thyroid cancer incidence from 1983 to 2000 was reported for France (10) and the thyroid cancer incidence in Germany in 2002 was 6.7 and 3.2 per 100,000 women and men respectively (11). However, most of these increases in thyroid cancer incidences were due to an increased detection of small papillary cancers (9, 10).

In autopsy studies, clinically silent thyroid papillary microcarcinomas (<1 cm diameter) have been reported in up to 36% depending on the number of serial sections (12). Most autopsy studies report incidences ranging from 6 to 11% (13–15). A comparison of these papillary microcarcinoma incidence rates in autopsy studies with the incidence rates for clinically apparent papillary carcinomas strongly suggests that most papillary microcarcinomas will not lead to clinically apparent thyroid carcinomas. Moreover, these data suggest that the histological evaluation of resected thyroid tissue will often detect papillary microcarcinomas with an unlikely clinical relevance. A follow-up study of papillary microcarcinomas over a 9-year period demonstrates no metastasis in patients with tumors <0.8 mm (16).

Studies on the epidemiology of thyroid nodule function are rare. The scintigraphic evaluation of 60% of the solitary nodules detected by US in a random cohort of probands aged 41–71 years living in an area with borderline iodine deficiency revealed cold nodules in 46%, isofunction nodules in 44%, and hot nodules in 6% (17). In another population study, the thyroid nodules were detected by thyroid palpation of adults aged 18–64 in only 1.9% in an iodine-sufficient area and in 5.1% in an iodine-deficient area. The scintigraphic evaluation of these nodules identified cold nodules in 87 and 84%, isofunction nodules in 0.4 and 0.6%, and hot nodules in 8 and 10% in the iodine-sufficient and -deficient areas respectively (18). Most hot nodules are easily detected by a thyrotrophin (TSH) determination. However, in iodine-deficient areas, scintigraphic evidence for thyroid autonomy has been reported in 40% of patients with euthyroid endemic goiters (19). Moreover, somatic constitutively activating TSH receptor mutations have been detected in small 131 iodine-hyperfunction areas detected by autoradiography (20). It is therefore likely that not all hot nodules that are much more frequent in iodine-deficient than iodine-replete areas (21) are detectable by the determination of TSH. However, if the hot nodule volume surpasses a volume of 16 ml, a suppressed TSH was detectable even with older RIA technology (22).

The high prevalence of thyroid nodules requires evidence-based rational strategies for their differential diagnosis, risk stratification, treatment, and follow-up. These strategies should concentrate on the risk for malignancy, hyperthyroidism, and symptoms and should be adaptable to the wide spectrum of clinical manifestations of thyroid nodules ranging from the small <1.0 cm thyroid incidentaloma to the large

symptomatic thyroid nodule with progressive growth. Moreover, these strategies should also account for the different prevalences of thyroid nodules, hot nodules, and the different subtypes of differentiated thyroid carcinomas in iodine-replete and -deficient areas as well as different healthcare systems.

However, several questionnaire studies with European, North American, and Australian endocrinologists repeatedly revealed large discrepancies in the diagnosis and management of thyroid nodules (23–26). Among other discrepancies, a less frequent application of the fine needle aspiration biopsy (FNAB), more frequent use of thyroid scintigraphy, thyroid US calcitonin (CT), and thyroid peroxidase (TPO) antibody determination for the diagnosis of thyroid nodules in Europe as compared with North America and different strategies for the treatment of thyroid nodules became apparent. Most of these questionnaires and especially those performed in Europe and North America were published in 1999 and 2000. In 2006, three major society-sponsored guidelines for thyroid diagnosis and management were published by the AACE/AME; the ATA; and the ETA (1–3). Future surveys such as the one reported below and performed during an interactive symposium at the 32nd annual meeting of the ETA in Leipzig, Germany, will therefore have to clarify if these guidelines were able to impact on the divergent management strategies for thyroid nodules that have previously been documented.

Agreement and disagreement between the guidelines

This presentation attempted to review and compare a few similarities and differences, but clearly is not an exercise in systematic comparison of the recently published guidelines. The format consisted of a case presentation, followed by a management question with multiple choices, followed by audience response immediately obtained by an audience response system (ARS), and then a brief discussion of current recommendations by the major published guidelines vis-à-vis audience response. Although several hundred physicians were present at the symposium, only 200 ARS devices were picked up with responses that are included and discussed herein.

Case 1. A 40-year-old woman presents with a recently discovered thyroid nodule. She reports no history of radiation exposure and there is no pain, tenderness, dysphonia, or dysphagia. Examination reveals an easily palpable, firm, solitary 2.5 cm right thyroid nodule. There is no cervical adenopathy.

Question 1 – Would you obtain a thyroid US?

- a. Yes, always
- b. No, never
- c. It depends!

Answer 1 – ($n=146$; n represents number of responders)

- a. 89%
- b. 5%
- c. 6%

Discussion. The overwhelming majority here would obtain a thyroid US, which is consistent with the recommendations from all three recent guidelines (Table 1). The AACE/AME guidelines suggest a thyroid US for any patient with a palpable nodule, history of neck radiation, family history of thyroid cancer, or the presence of unexplained cervical adenopathy. The ATA guidelines suggest that thyroid US should be performed in all patients with one or more suspected thyroid nodules. The ETA guidelines state that thyroid US is mandatory when a nodule is discovered at palpation.

Single thyroid US characteristics of thyroid nodules are of limited sensitivity and specificity (27). However, results are highly operator dependent and clearly superior in clinics or centers with good experience and expertise (27). Therefore, despite the fact that US data are compelling, they should be viewed with caution. In the US an ever-increasing number of endocrinologists are attending US courses and learning to use US machines in their daily thyroid practices. Current US machines are user-friendly, sensitive, portable, and affordable. Although office use of sonography, unlike nuclear medicine practice, does not require licensure, it is predicted that in the near future physicians using US machines will be required to obtain a license.

Question 2 – Serum TSH is 0.6 mIU/l (normal 0.5–4.5). Would you now obtain a radioisotope scan?

- a. Yes
- b. No
- c. Maybe

Answer 2 – ($n=171$).

- a. 45%
- b. 41%
- c. 12%

Table 1 Comparison of recommendations for thyroid nodular diagnostic tests in different guidelines.

Society	EB	Initial test	US	Scan	CT	Rebiopsy
AACE/AME ^a	Yes	TSH+FNA	Yes	No	No	No
ATA ^b	Yes	TSH+FNA	Yes	No	?	No
ETA ^c	No	TSH+FNA	Yes	No	Yes	Yes
UK ^d	No	TSH+FNA	No	No	?	Yes

EB, evidence based; US, ultrasound; CT, calcitonin.

^aEndocrine Practice 2006 12 63–102 (1).

^bThyroid 2006 16 109–142 (2).

^cEuropean Journal of Endocrinology 2006 154 787–803 (3).

^dwww.british-thyroid-association.org (52).

Discussion. The audience was evenly divided between ordering and not ordering a thyroid scan. Recent AACE/AME guidelines have suggested that a radioisotope thyroid scan should be ordered only if TSH is below the lower limit of normal range or if the patient has a large single nodule or a multinodular goiter, and is from an iodine-deficient area. ATA guidelines recommend a radioisotope scan when TSH is low or low-normal to rule out an autonomous nodule, similar to the ETA guidelines that suggest a thyroid scan when TSH is low or undetectable.

Table 2 illustrates results of surveys of ATA members in 1996 and 2000, and ETA members in 2000. Two points emerge: in evaluating thyroid nodules, ATA members use imaging less often than their ETA colleagues, and ATA respondents use thyroid scan less and US more in 2000, compared with 1996 (28).

Question 3 – How would you perform FNA?

- a. By palpation
- b. By US guidance (US-FNA)

Answer 3 – ($n=186$).

- a. 26%
- b. 74%

Discussion. The fact that two-thirds of this audience selects FNA with US guidance seems surprising, considering that the nodule was easily palpable. However, as illustrated by a response to question 1, the vast majority of the people in this room would use an US for the initial evaluation of this patient. Accordingly, it appears logical that they would also perform an US-guided FNA rather than a palpation-directed FNA.

Several recent reports have suggested that US-FNA is more reliable than palpation-FNA (1, 27). With the use of US guidance, the sensitivity, positive predictive value, and negative predictive value of the test increase significantly. Accordingly, as the use of thyroid US by endocrinologists is becoming more widespread, one should expect that more biopsies are done with rather than without US. Clearly, the audience feels that US-FNA has greater diagnostic accuracy.

AACE/AME guidelines suggest US-FNA for the following clinical settings: any size nodule with a history of radiation, or family history of RET; any size nodule with suspicious US features; for nodules with

Table 2 Answers to the question 'which imaging test would you obtain in a patient with a solitary thyroid nodule?' in published questionnaires (24, 53).

Test	ATA, 1996 (%)	ATA, 2000 (%)	ETA, 2000
Scan	56	23	66
US	28	34	80
Scan+US	–	13	58

extra capsular growth or cervical nodes; and impalpable or small (<1 cm) nodules. The other two guidelines do not make specific recommendations for US-FNA.

Question 4 – A palpation-directed FNA was performed and smears were 'nondiagnostic' or 'unsatisfactory.' Would you proceed with:

- a. a repeat FNA by palpation
- b. an US-FNA
- c. a repeat FNA only if nodule is primarily (>75%) solid on US
- d. surgery

Answer 4 – (n=170)

- a. 3%
- b. 76%
- c. 15%
- d. 6%

Discussion. Nondiagnostic or unsatisfactory smears are caused by inadequate cellularity and should be repeated. Approximately 50% are satisfactory on reaspiration, the yield increasing if rebiopsy is done by US-FNA (27). The majority of those in this room feel that repeat biopsy should be done with the help of an US, an answer that is consistent with both AACE/AME and ATA guidelines. AACE/AME suggests that in this case, US-FNA is preferred and biopsy should be directed at the periphery of the lesion, noting that despite good technique ~5% of nodules remain nondiagnostic and should be surgically removed. ATA guidelines suggest close observation for repeatedly nondiagnostic cystic nodules, and thyroidectomy if nodule is solid.

Table 3 illustrates that FNA results are 65% benign or negative for malignancy; 5% positive or malignant; 20% nondiagnostic or unsatisfactory; and 10% suspicious or indeterminate. The probability of malignancy in each category is listed in the Table and should be <2% in the benign group in laboratories with good FNA experience (27).

Question 5 – FNA was done and result is benign. You ask this patient to return in 6–12 months for:

- a. Thyroid palpation only
- b. Thyroid palpation + US
- c. Thyroid palpation + FNA
- d. Thyroid palpation + US + FNA
- e. None of the above

Table 3 FNA results (according to (27)).

Cytology	Results %	Probability of malignancy %
Benign (negative)	65	<1
Malignant (positive)	5	>99
Nondiagnostic (unsatisfactory)	20	<3
Suspicious (indeterminate)	10	20

Answer 5 – (n=166)

- a. 6%
- b. 59%
- c. 4%
- d. 27%
- e. 4%

Discussion. AACE/AME guidelines suggest simple follow-up for cytologically benign thyroid nodules; repeat US was not recommended. ATA guidelines suggest clinical follow-up at 6–18 months, without US monitoring for easily palpable benign nodules. Opinion on reaspiration of benign nodule remains divided. AACE/AME suggests reaspiration only for enlarging nodules, recurrent cysts, or for nodules not shrinking after thyroxine (T₄) therapy. ATA guidelines suggest either reaspiration or surgery for growing nodules.

Wiersinga has recommended repeat palpation and FNA 1 year after a benign FNA result (29). Lucas *et al.* rebiopsied 116 patients with benign FNA and found no missed malignancy, concluding that reaspiration is not necessary (30). On the other hand, Chehade *et al.* followed 235 patients with benign FNA for an average of 2.9 years, and on repeat FNA found malignancy in 1 (0.4%), concluding that rebiopsy reduces false-negative rates (31).

Question 6 – Remember that our patient is a 40-year-old woman with a single, 2.5 cm nodule; benign, colloid by FNA; solid by US; and with serum TSH 0.6 mIU/L. Would you recommend T₄ suppressive therapy?

- a. Yes
- b. Never
- c. Only in iodine-deficient areas
- d. I do not like this question

Answer 4 – (n=182)

- a. 14%
- b. 65%
- c. 12%
- d. 9%

Discussion. The practice of routine T₄ therapy for benign thyroid nodules has undergone dramatic change in the past two decades. Whereas previously most endocrinologists would have used T₄ to suppress TSH in this case, current guidelines do not endorse this practice. Therefore, it is gratifying that 65% in this audience agreed with the guidelines and chose not to use T₄ therapy.

AACE/AME guidelines state that routine T₄ therapy in patients with benign thyroid nodules is not appropriate but it may be considered in iodine deficiency. Today, 12% of this group voted to use T₄ if the patient is in an iodine-deficient area. The ATA panel did not recommend suppression therapy for benign nodules.

A recent meta-analysis of nine studies including 596 patients has showed that nodule volume decreased significantly in only less than 20% of the treated group. Moreover, T₄ suppressive therapy led to a nonsignificant improvement in the rate of response to therapy (defined as $\geq 50\%$ nodule volume reduction by US; pooled relative risk (RR) 1.83, 95% CI 0.9–3.73) (32). In summary, neither the guidelines nor the majority of this audience use T₄ to suppress benign thyroid nodules.

Question 7 – What would you do next if cytology showed ‘suspicious for malignancy-follicular neoplasm’?

- a. Obtain immunohistochemical markers
- b. Order radioisotope scan
- c. Respire nodule
- d. Surgical excision

Answer 7 – (n = 195)

- a. 6%
- b. 12%
- c. 2%
- d. 80%

Discussion. Management of a nodule with indeterminate cytology still generates controversy. The cancer risk among these specimens ranges from 15 to 75%, $\sim 15\%$ for follicular neoplasms. Immunohistochemical markers have neither regularly nor reliably separated benign from malignant lesions (27). Repeat biopsy is not helpful and can even lead to confusion, because if reaspiration is benign, the clinician has to reconcile between a benign and a suspicious result. AACE/AME guidelines consider surgical excision as the best management; repeat biopsy or large-needle biopsy is not recommended. ATA guidelines discourage the use of molecular markers and prefer a radioiodine thyroid scan, to rule out nodule hyperfunction, when cytology is suspicious. ETA guidelines find immunocytochemistry neither sensitive nor specific, believing surgical treatment is the best approach.

Question 8 – Final cytology is ‘follicular neoplasm’ and you recommend surgery. Which of the following do you choose?

- a. Lobectomy and postop histological review
- b. Near-total thyroidectomy
- c. Lobectomy and intraoperative frozen section exam
- d. My surgeon is smart; I’ll leave it up to him/her!

Answer 8 – (n = 188)

- a. 24%
- b. 34%
- c. 36%
- d. 6%

Discussion. AACE/AME guidelines recommend surgical treatment but do not specify the extent of surgery.

Guidelines by the ATA suggest thyroid lobectomy for an isolated, indeterminate solitary nodule, whereas ETA recommends lobectomy for a solitary nodule and a near-total thyroidectomy for an MNG when cytology is suspicious. Moreover, the ETA does not endorse frozen section because of the high frequency of false-negative results.

Several recent reports have suggested that in experienced hands intraoperative frozen section can accurately separate benign from malignant follicular or Hurthle cell neoplasms. For example, Paphavasit and colleagues report that intraoperative frozen section was correct in 78% of patients, with sensitivity, specificity, positive predictive value, negative predictive value, and accuracy of 78, 99, 90, 98, and 98% respectively (33).

It is, therefore, not surprising that this audience seems evenly divided between the surgical options noted above. While there is no majority of opinion here, these differences likely represent availability of surgical and pathology expertise available to each participant in his/her clinic or location of practice.

Case 2. A 40-year-old woman has a palpable, solitary thyroid nodule; serum TSH is 0.6 mIU/l; an US shows a solid $3.0 \times 2.5 \times 2.0$ cm right lobe mass, left lobe is normal, and no cervical adenopathy identified.

Question 9 – Would you measure baseline serum CT level?

- a. Yes
- b. No
- c. Only if it is free!

Answer 9 – (n = 125)

- a. 49%
- b. 43%
- c. 8%

Discussion. In this presentation, this was the most controversial question that generated a lengthy and heated discussion. Serum CT is a useful marker for C-cell disease and correlates well with tumor burden. Medullary thyroid carcinoma (MTC) accounts for only 5% of thyroid malignancies; however, recent reports have shown that prevalence of MTC ranges from 0.4 to 1.4% in unselected patients with nodular thyroid disease. Data from nonrandomized, prospective studies mostly from European countries suggest that routine CT measurement can detect early and unsuspected MTC (3). Early diagnosis and prompt thyroidectomy result in decreased morbidity and increased survival. Yet, there seems to be no consensus on this issue, and outside Europe, the enthusiasm for ordering routine CT has not been high except for a recent publication suggesting cost effectiveness (34). Figure 1 illustrates how often clinical members of the ATA, the Australian endocrinologists, the ETA, and the Latin American Thyroid Society (LATS) would order CT in patients with nodular thyroids (28, 35).

Figure 1 Answers with yes to the question 'Would you measure serum calcitonin for the work up of a patient with nodular thyroid disease?' in published questionnaires (*Journal of Clinical Endocrinology and Metabolism*, 2005; *British Journal of Surgery*, 2003).

AACE/AME guidelines do not endorse routine CT measurement recommending the test only if FNA is suspicious for MTC or family history is positive for RET. The ATA panel could not recommend either for or against routine CT measurement. And the ETA recommends CT measurement in the initial diagnostic evaluation of thyroid nodules. Recent expert opinion has underlined this practice difference on opposite sides of the Atlantic: Borget *et al.* writing from France, state that 'based on their assumption, plasma CT determination in the assessment of thyroid nodule pt would appear to be highly favorable compared with a number of other accepted health interventions' (36). On the other hand, Podak and Burman, writing from the US, observe that 'the issue of CT testing in pt with thyroid disease remains controversial. It does not seem that the use of basal CT levels in the routine screening of pt with nodular thyroid disease is warranted without the ability to use gastrin stimulation as a confirmatory test' (37).

Today, the audience response seems to acknowledge this ongoing controversy in thyroid practice, with 49% using, whereas 43% not using, routine CT measurement. It is surprising, however, that as many as 43% of this group does not order CT measurement, considering that the overwhelming majority of them today are likely practicing in Europe. We would submit that despite the recommendation from the ETA for routine CT determination, many members still consider this issue unresolved.

Conclusions

Guidelines provide useful information and recommendations for practice, and impact positively on patient care. Additionally, guidelines should be considered as suggestions but not a rigid formula for practice. In the opinion of the authors, thyroid practice is influenced by published scientific data, personal preferences, patient priorities, and availability of resources. The recent thyroid guidelines underscore regional practice differences, but they also illustrate many similarities. It will be desirable to evaluate the impact of guidelines on

practice patterns by a follow-up survey of clinical members of the AACE, AME, and ETA.

One aim of published guidelines is to reach a consensus. We feel consensus in health-related issues is useful and provides a framework for improved practice. However, even on the issue of consensus, there is no consensus! For example, Dr Terry Davies has recently observed that 'I am always suspicious of the word 'consensus.' To me this word means that not everyone agrees... We have to rely on the art of practical medicine, not arbitrary guidelines which can lead to problems for the individual patient'.

Recommendations based on evidence and lack of evidence

Despite many strong similarities, a few relevant differences are present in the recently issued clinical practice guidelines on thyroid nodule management. These controversies probably reflect not only the possible absence of high-quality evidence but also the frequently observed lack of consistency of results and the presence of some differences among the authors in the interpretation of the existing data. This is probably due to the variable availability of technical resources and professional skills and to the changing prevalence of thyroid disease in different regions of the world.

We will therefore review the methodology followed by the AACE/AME and ATA guidelines and by the ETA consensus statement to assess their evidence-based medicine (EBM) quality on the basis of a validated framework, and to evaluate the strength of the evidence at the basis of the, sometimes controversial, recommendations for clinical practice.

The methodology of the recently issued guidelines on thyroid nodule is similar and comparable?

The AACE/AME guideline follows the AACE *ad hoc* task force system (38). Four levels of evidence are recognized: good quality randomized controlled trials and

meta-analysis (level 1 evidence), limited quality randomized controlled trials and well-conducted prospective cohort studies (level 2 evidence), observational studies or conflicting evidence (level 3 evidence), and expert consensus and opinions (level 4 evidence). Recommendations are graded 'A' (homogeneous evidence from multiple randomized controlled trials or multiple cohort studies), 'B' (evidence from at least one randomized controlled trial, cohort or case-control study, or meta-analysis), 'C' (evidence based on clinical experience, descriptive studies, or expert consensus/opinion), or 'D' (not rated evidence). Risk-to-benefit ratio is also taken into account. See Table 4 for details.

The ATA guideline grades the recommendations according to the US Preventive Services Task Force system (39). Rating: 'A – Strongly recommends' recommendations are based on consistent evidence from well-designed and well-conducted studies; 'B – Recommends' means that the evidence is sufficient to recommend the intervention but is of limited value regarding number, quality, or consistence of individual studies; 'C – Recommends' refers to expert opinions. 'D – Recommends against'; 'E – Recommends against'; and 'F – Strongly recommends against' are based on the

same evidence as C, B, and A recommendations respectively, but are negative ('do not do'); 'I – Recommends neither for nor against' recommendation is released when evidence is insufficient to recommend for or against providing the intervention/service, due to lacking or conflicting data. The impact on relevant health outcomes and harm-to-benefits ratio is considered. See Tables 5 and 6 for details.

No official method for grading the level of evidence and the strength of recommendations has been followed by the ETA panel for the European Consensus. The ETA approached the National Endocrine Societies in Europe requesting the appointment of two thyroid cancer experts for each country. The experts have discussed some issues about thyroid nodules in a meeting held in Athens on 24 May, 2005. The experts were advised to base their statements on clinical and scientific evidence whenever available in the current literature. After a state-of-the-art session and a subgroup's work on the selected issues, the results were reported to the whole group of experts. During the following months, the coordinators drafted a text that was circulated among the members of the task force and the discussion took place via electronic mail up to the delivery of the final

Table 4 Classification of the level of evidence and the recommendation grade according to the American Association of Clinical Endocrinologists (AACE) *ad hoc* task force system for the AACE/Associazione Medici Endocrinologi (AME) guideline (38).

Level of evidence	Recommendation grade	Description
Various strength-of-evidence scales reported in the medical literature		
1		Well-controlled, generalizable, randomized trial Adequately powered Well-controlled multicenter trial Large meta-analysis with quality ratings
2		All-or-none evidence Randomized controlled trial-limited body of data Well-conducted prospective cohort study Well-conducted meta-analysis of cohort studies
3		Methodologically flawed randomized clinical trials Observational studies Case series or case reports
4		Conflicting evidence with weight of evidence supporting the recommendation Expert consensus Expert opinion based on experience 'Theory-driven conclusions' 'Unproved claims'
	A	Homogeneous evidence from multiple well-designed randomized controlled trials with sufficient statistical power Homogeneous evidence from multiple well-designed cohort controlled trials with sufficient statistical power ≤ 1 conclusive level 1 publications demonstrating benefit $\gg$ risk
	B	Evidence from at least one large well-designed clinical trial, cohort or case-controlled analytic study, or meta analysis No conclusive level 1 publication ≥ 1 conclusive level 2 publications demonstrating benefit $\gg$ risk
	C	Evidence based on clinical experience, descriptive studies, or expert consensus opinion No conclusive levels 1 or 2 publications ≥ 1 conclusive level 3 publications demonstrating benefit $\gg$ risk
	D	No conclusive risk at all and no conclusive benefit demonstrated by evidence Not rated No conclusive levels 1, 2, or 3 publication demonstrating benefit $\gg$ risk Conclusive levels 1, 2, or 3 publications demonstrating risk $\gg$ benefit

From the American Association of Clinical Endocrinologists *Ad Hoc* Task Force for Standardized Production of Clinical Practice Guidelines.

Table 5 Classification of the strength of recommendation based on the American Thyroid Association (ATA) US Preventive Services Task Force (USPSTF) system for the ATA guideline.

Rating	Definition
Strength of panelists' recommendations based on available evidence	
A	Strongly recommends. The recommendation is based on good evidence that the service or intervention can improve important health outcomes. Evidence includes consistent results from well-designed, well-conducted studies in representative populations that directly assess effects on health outcomes
B	Recommends. The recommendation is based on fair evidence that the service or intervention can improve important health outcomes. The evidence is sufficient to determine effects on health outcomes, but the strength of the evidence is limited by the number, quality, or consistency of the individual studies; generalizability to routine practice; or indirect nature of the evidence on health outcomes
C	Recommends. The recommendation is based on expert opinion
D	Recommends against. The recommendation is based on expert opinion
E	Recommends against. The recommendation is based on fair evidence that the service or intervention does not improve important health outcomes or that harms outweigh benefits
F	Strongly recommends against. The recommendation is based on good evidence that the service or intervention does not improve important health outcomes or that harms outweigh benefits
I	Recommends neither for nor against. The panel concludes that the evidence is insufficient to recommend for or against providing the service or intervention because evidence is lacking that the service or intervention improves important health outcomes, the evidence is of poor quality, or the evidence is conflicting. As a result, the balance of benefits and harms cannot be determined

Source: Adapted from the US Preventive Services Task Force, Agency for Healthcare Research and Quality.

text. According to the proper characteristics of a consensus statement paper, the rating of evidence and the grade of recommendations are not provided.

A synopsis of the strength of recommendations and of the types of evidence that support the suggested interventions in the ATA and AACE/AME guidelines is provided in Table 6.

Two relevant thyroid cancer guidelines, which also discuss a few issues about clinical management of thyroid nodules, have been published in 2007: the National Comprehensive Cancer Network Guidelines (40) and the British Thyroid Association and Royal College of Physicians Guidelines (41). These documents are mainly focused on the management of thyroid cancer and have not been systematically evaluated for the ETA symposium.

EBM quality of guidelines

The overall EBM quality of the guidelines has been assessed by the Appraisal of Guidelines for Research & Evaluation (AGREE) instrument. This framework is the product of an international collaboration, endorsed by

the European Community and developed at St George's Hospital Medical School, London. The last release is of September 2001 (42).

The AGREE instrument is a complex of 23 key items organized in six domains. Each domain is intended to capture a separate dimension of guideline quality. The domains evaluate scope and purpose (three items), stakeholder involvement (four items), rigour of development (seven items), clarity and presentation (four items), applicability (three items), and editorial independence (two items). Each item ranks 1–4, generating a maximum score of 4×23=92 and a minimum score of 1×23=23.

The three documents considered here have been independently evaluated by four AME EBM experts. The mean AGREE scores have been the following: AME/AACE, 63/92; ATA, 61/92; ETA, 52/92.

The AACE/AME, ATA, and ETA documents had good ratings, with a similar AGREE score. Although the score may be useful for comparing guidelines and will inform the decision as to whether or not use or recommend a guideline, it is not possible to set a threshold for the score to mark a 'good or bad' guideline. The lower score

Table 6 Synopsis of the strength of recommendations and of the types of evidence that support the suggested interventions in the American Thyroid Association (ATA) and American Association of Clinical Endocrinologists/Associazione Medici Endocrinologi (AACE/AME) guidelines.

Recommendation grade	Type of evidence supporting the recommendation	Action
Methodology – synopsis		
A	Randomized trials	Strongly recommended
B	Not randomized trials	Recommended
C	Expert opinion	Recommended low grade
D (AACE/AME) or I (ATA)	No evidence available	Recommended neither for nor against

ATA guideline reports also ratings D, E, and F. These have the same strength of ratings C, B, and A respectively, but the recommendations are negative ('against').

of the ETA statement is due to the absence of links between the reported evidence and the recommendations for clinical practice. Moreover, recommendations are not reported as definite sentences.

A synopsis of the specific comments of the reviewers is presented in [Tables 6 and 7](#).

Levels of evidence and grading of recommendations

In order to assess the quality of the links between the strength of evidence and the grade of recommendations reported by the various guidelines, we examined the EBM basis of the recommendations concerning a few clinical questions.

Question 1: When should one use US and FNA in thyroid nodule evaluation?

Answer. Basically US examination is indicated by all guidelines (GL) if a nodule or goiter is palpable and by AACE/AME GLs even in patients without palpable lesions, who are at high risk of thyroid cancer, a practice also reported by the Society of Radiologists (42). FNA biopsy (either US or palpation guided) is recognized as the diagnostic procedure of choice for the assessment of clinically or incidentally discovered thyroid nodules.

Comments. Concordance between GLs is very high for both the clinical recommendations and the grade of their strength, but the evidence available for recommending US examination is correctly reported only as fair (ATA, grade B; AACE/AME, grade C). Indeed, although US is generally appreciated as a diagnostic procedure that induces a powerful effect on thyroid outcomes, the quality of EBM evidence in favor of clinical use of US thyroid scan may be rated just as fair due to the absence of level 1 and 2 clinical evidence.

The examination of the linked references confirms these remarks. For the thyroid US issue, ATA reports three observational studies (level of evidence: 3 according to AACE scale, [Table 4](#)), ETA one retrospective observational study (level of evidence: 3) and one consensus (level of evidence: 4) and AACE/AME reports five prospective observational studies (level of evidence: 3) and six reviews (level of evidence: 4). For FNA biopsy, ATA reports three observational studies (level of evidence: 3), ETA three observational studies (level of evidence: 3) and one review (level of evidence: 4), and AACE/AME six observational studies (level of evidence: 3) and seven reviews with a pooled analysis (level of evidence: 3–4).

It is noteworthy to observe that the evidence reported by ATA, ETA, and AACE/AME GLs on thyroid US and FNA biopsy lacks consistency. We found only as shared by the three GLs:

- one reference in ATA and ETA (Marqusee *et al.* *Annals of Internal Medicine* 2000: retrospective observational study (43)),
- two references in ATA and AACE/AME (Tan *et al.* *Archives of Internal Medicine*, 1995: retrospective observational study (44); Hagag *et al.* *Thyroid* 1998: prospective observational study (45)), and
- no reference in the ETA Consensus and the AACE/AME GL.

Question 2: Which laboratory evaluations should be performed for thyroid nodules?

Answer. In the absence of a specific clinical suspect, the initial thyroid laboratory evaluation requires the determination of TSH only according to all the guidelines. The ATA panel cannot recommend either for or against the routine measurement of serum CT,

Table 7 Results of the guideline evaluation by the Appraisal of Guidelines for Research & Evaluation (AGREE) instrument.

AGREE score		Pros	Cons
Guidelines quality – overall evaluation			
AME/AACE	63/92	Well-described link between levels of evidence and strength of recommendations Key recommendations clearly indicated in specific tables	Some paragraphs seem of a text book more than a guideline A few recent procedures with a low level of evidence are addressed
ATA	61/92	Comprehensible review of thyroid nodule management The link between level of evidence and strength of recommendations is indicated Mainly recommendations are clearly indicated Algorithm provided Addressed to DTC as well	Disclosure policy: most authors have ‘relationships with commercial companies that could potentially affect the information presented’ The strength of evidence is not rated
ETA (consensus statement)	52/92	Synthetic overview of the main issues regarding thyroid nodule and DTC Summarizes what European endocrinologists are advised to do for the diagnosis and treatment of DTC	The links between evidence and recommendation are lacking Recommendations are not reported as definite sentences

while AACE/AME recommends that serum CT should be measured if FNA or family history suggests MTC. On the other hand, ETA consensus recommends the routine use of serum CT measurement in the initial diagnostic evaluation of thyroid nodules.

Comments. There is a complete concordance among the panels of the three societies about the measurement of serum TSH concentration as the single-most useful laboratory test in the initial evaluation of thyroid nodules because of the high sensitivity of the TSH assay in detecting even subtle thyroid dysfunction.

On the contrary, there is a near-complete discordance about CT testing. ATA claims that there remain unresolved issues of sensitivity, specificity, assay performance, and cost effectiveness about serum CT measurement and adds that most studies rely on pentagastrin stimulation testing, a drug no longer available in the United States. ETA recommends routine use of CT testing as several – mostly European – prospective studies of unselected thyroid patients have demonstrated that routine measurement of serum CT allows the detection of unsuspected medullary thyroid carcinoma. According to these reports, serum CT has a better sensitivity for medullary cancer than thyroid FNA biopsy and using routine CT screening should improve the outcome of clinically unapparent medullary cancer. Finally, AACE/AME does not endorse routine CT measurement because testing of serum CT in all patients with unselected thyroid nodules does not seem to be cost-effective but recommends the test if FNA biopsy is suspicious or positive for malignancy or if family history is positive.

The examination of the references linked to the issue of serum CT testing shows that ATA GLs report four observational papers (EBM level: 3); AACE/AME two reviews (level of evidence: 4), one guideline (level of evidence: 4), and three observational papers (level of evidence: 3); and ETA four observational papers (level of evidence: 3). We can observe how the evidence reported by ATA, ETA, and AACE/AME GLs on CT testing is definitely consistent. We found as shared by the three GLs:

- two references in ATA and ETA (Niccoli, *Journal of Clinical Endocrinology and Metabolism* 1997 (46); Elisei, *Journal of Clinical Endocrinology and Metabolism* 2004: observational studies (47)),
- two references in ATA and AACE/AME (Niccoli, *Journal of Clinical Endocrinology and Metabolism* 1997 (46); Elisei, *Journal of Clinical Endocrinology and Metabolism* 2004: observational studies (47)), and
- two references in ETA and AACE/AME (Niccoli, *Journal of Clinical Endocrinology and Metabolism* 1997 (46); Elisei, *Journal of Clinical Endocrinology and Metabolism* 2004: observational studies (47)).

Thus, although the references reported in the American and the European documents are rather the same, their recommendations are strongly discordant. This

controversy may be explained by the low level of the available evidence, but probably socioeconomic factors and personal opinions of the panelists play a major role in this enduring controversy. A large scale, prospective, controlled study with an adequate follow-up is needed to provide a final high-quality evidence about the diagnostic efficacy and cost-effectiveness of routine serum CT determination.

Question 3: When should a radioisotope thyroid scan be performed?

Answer. Basically radioisotope thyroid scan is indicated by the GLs if serum TSH is suppressed and in nodules with a 'follicular neoplasm' cytological report.

Comments. According to ATA GLs, in the presence of a low or low-normal serum TSH concentration, a radioiodine scan should be performed directly compared with the US images to determine functionality of each nodule larger than 1–1.5 cm. AACE/AME and ETA extend the indication to radioisotope scan, suggesting to perform thyroid scintigraphy for a multinodular goiter in iodine-deficient areas even if the TSH level is still in the normal range, in order to identify the presence of an autonomous nodule.

Concordance between the guidelines is very high for the suggested actions even if the clinical evidence available for using radioisotope scan is just fair (ATA GL, grade B; AACE/AME GL, grades B and C).

The references linked to the issue of thyroid scintigraphy are scarce. Their examination shows that ATA GLs quote no reference, ETA mentions one reference (Pacini, *European Journal of Nuclear Medicine and Molecular Imaging* 2004; 31: 1443–1449: review (48)), and AACE/AME reports seven papers (one review and six observational studies). Actually, the quality of EBM evidence about the use of radioisotope scan is quite low due to the absence of level 1 and 2 evidence and recommendations are at present based mostly on experts opinion and largely accepted thyroid practice.

Question 4. What is the role of medical therapy for benign thyroid nodules?

Answer. L-T₄ suppressive treatment for thyroid nodules is not indicated (ATA) or should be restricted to a minority of selected patients, preferably in iodine-deficient regions (AACE/AME). This issue is not addressed by the ETA consensus statement.

Comments. There is a partial discordance between the guidelines. ATA GL strongly recommends against the use of L-T₄, while AACE/AME GL recommends that the use of L-T₄ therapy should be avoided in most cases but may be considered in the following conditions: patients from geographical areas with iodine deficiency, young patients with small thyroid nodules, and nodular goiters with no evidence of functional autonomy.

Thus, even if the evidence from multiple randomized control trials and three meta-analyses suggest that thyroid hormone in doses that suppress the serum TSH

to subnormal levels may result in a decrease in nodule size in regions of the world with borderline low iodine intake, ATA strongly recommends against (grade F). On the other hand, AACE/AME considers $L-T_4$ at nearly suppressive doses as a possible treatment (grade C) in selected cases.

It's noteworthy that the evaluation of the references used to support the grade of recommendations in the two GLs demonstrates that in this controversial case the quality of the available clinical evidence is high.

ATA GLs strongly recommend against $L-T_4$ on the basis of two well-conducted randomized clinical trials (RCT) and two meta-analyses with a high level of evidence.

AACE/AME GLs fairly recommend to consider $L-T_4$ for a minority of cases on the basis of three observational studies (level of evidence: 3), eight randomized controlled trials, and two meta-analyses (level of evidence: 1–2).

Evidence reported by ATA and AACE/AME GLs has three references in common (Zelmanovitz *et al.*, *Journal of Clinical Endocrinology and Metabolism* 1998, RCT+meta-analysis (49); Wemeau *et al.*, *Journal of Clinical Endocrinology and Metabolism* 2002, RCT (50); Castro *et al.*, *Journal of Clinical Endocrinology and Metabolism* 2002, meta-analysis (51)). Nine references are quoted only in AACE/AME guidelines. So, although both the references and the strength of clinical evidence in the two GLs are quite similar ('strong'), their recommendations are at least in part discordant.

How do the European endocrinologists evaluate and implement the guidelines?

A partial response to this question may be obtained by evaluating the answers of the ETA audience to a few clinical questions.

'Is the methodology of guidelines similar, and are their recommendations comparable?'

The answer was: Yes, 23%; No, 42%; I do not know, 35%.

Hence, only a minority of the participating endocrinologists were aware that the methodology of the GLs is basically similar but that the differences in the rating scales may make difficult a rapid evaluation of the strength of evidence and recommendations for their use in clinical practice.

The question 'which of these recommendations is based on good quality evidence?' was addressed to several clinical problems.

The above-mentioned recommendations 'US is indicated if a nodule or goiter are palpable' and 'FNA is the diagnostic procedure of choice for thyroid nodules' had the following score when participants were asked about their strength of evidence: High, 73%; Low, 18%; I do not know, 7%.

On the other hand, the recommendation ' T_4 treatment for thyroid nodules is not indicated OR it should be restricted to a minority of selected patients' had the following score: High, 57%; Low, 38%; I do not know, 3%.

Because US is generally appreciated as an important procedure for the diagnosis of thyroid nodules, the quality of EBM evidence in favor of the use of US thyroid scan was supposed by part of the audience to be elevated even if it can be rated only as low due to the absence of controlled studies. On the other hand, the still controversial use of $L-T_4$ suppressive therapy deceptively induced part of the audience to suppose that the available evidence is low, while the quality of several RCTs is elevated.

The strength of evidence in endocrinology is sometimes weak even if clinical experience clearly demonstrates the efficacy of diagnostic procedures or treatments. Hence, some GLs recommendations and algorithms are presently based more on the expert's opinion or the practical evidence of a clinical usefulness and cost-effectiveness than on the results of high-quality trials.

Conclusions

The overall EBM quality of the recently issued thyroid nodule GLs is good. The AGREE score of the ETA consensus statement is obviously slightly lower than the rating of the GLs documents.

The methodology of the GLs is similar, but a few differences in the rating scale make difficult a rapid comparison of the strength of both evidence and recommendations for use in current clinical practice.

The strength of evidence in endocrinology is frequently weak; hence some recommendations are based mostly on experts' opinion.

Thus, a same recommendation may be based on different evidence; on the other hand, sometimes the same evidence may induce a different recommendation.

Future perspectives

Therefore, efforts are needed:

- to produce a few high-quality clinical studies to close the evidence gaps in the still controversial fields of thyroid disease like:
 - the natural history of thyroid nodules,
 - evaluation of new therapeutic options for benign thyroid nodules except follow-up,
 - relevance and ways of CT screening, and
 - data on quality of life and costs;
- to agree on similar/comparable assay and US technologies;
- to agree on the characterization of phenotypes (solitary, multiple, solid, or cystic);
- to agree on when a volume change is significant;
- to evaluate inter- and intra-observer variations for US characteristics of thyroid nodules;
- to agree on standardized survey cases that take different disease epidemiologies into account;

- to agree on criteria for the adequacy/quality and reporting for FNABs;
- to use a similar grading and rating scale for the guidelines in order to reach a better understanding of the actual strength of their recommendations;
- to improve the correct clinical implementation of the guidelines with the addition of a practical appendix and a guide for primary care physicians and for countries with other disease epidemiologies or less economic and technical resources;
- to also be aware that guidelines are not a substitute for sound clinical judgement and the art of medicine but only a way to summarize current clinical evidence that needs to be applied to the individual patient; and
- to create a joint task force of the most authoritative societies in the field of thyroid disease in order to reach a common document for clinical practice recommendations.

Some of these goals will be approached by a joint revision of the current AACE/AME guideline for the diagnosis and management of thyroid nodules by a joint AACE/AME/ETA committee until 2009.

Declaration of interest

The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.

Funding

This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

Acknowledgements

We wish to thank Michele Zini, MD, and Rinaldo Guglielmi, MD, for their relevant contribution to the appraisal of the methodology followed by the guidelines, to the assessment of their EBM quality, and to the evaluation of the strength of evidence as the basis of their recommendations for clinical practice. According to a symposium at the 32nd Annual Meeting of the European Thyroid Association, Leipzig, Germany, September 2, 2007.

References

- 1 American Association of Clinical Endocrinologists and Associazione Medici. Endocrinologi Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules. *Endocrine Practice* 2006 **12** 63–102.
- 2 Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel SJ, Mazzaferri EL, McIver B, Sherman SI & Tuttle RM. Management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* 2006 **16** 109–142.
- 3 Pacini F, Schlumberger M, Dralle H, Elisei R, Smit JW & Wiersinga W. European Consensus for the management of patients with differentiated thyroid carcinoma of the follicular epithelium. *European Journal of Endocrinology* 2006 **154** 787–803.
- 4 Brander A, Viikinkoski P, Tuuhea J, Voutilainen L & Kivisaari L. Clinical versus ultrasound examination of the thyroid gland in common clinical practice. *Journal of Clinical Ultrasound* 1992 **20** 37–42.
- 5 Ezzat S, Sarti DA, Cain DR & Braunstein GD. Thyroid incidentalomas. Prevalence by palpation and ultrasonography. *Archives of Internal Medicine* 1994 **154** 1838–1840.
- 6 Mortensen JD, Woolner LB & Bennett WA. Gross and microscopic findings in clinically normal thyroid glands. *Journal of Clinical Endocrinology and Metabolism* 1955 **15** 1270–1280.
- 7 Volzke H, Ludemann J, Robinson DM, Spieker KW, Schwahn C, Kramer A, John U & Meng W. The prevalence of undiagnosed thyroid disorders in a previously iodine-deficient area. *Thyroid* 2003 **13** 803–810.
- 8 Nagataki S & Nystrom E. Epidemiology and primary prevention of thyroid cancer. *Thyroid* 2002 **12** 889–896.
- 9 Davies L & Welch HG. Increasing incidence of thyroid cancer in the United States, 1973–2002. *Journal of the American Medical Association* 2006 **295** 2164–2167.
- 10 Colonna M, Guizard AV, Schwartz C, Velten M, Raverdy N, Molinie F, Delafosse P, Franc B & Grosclaude P. A time trend analysis of papillary and follicular cancers as a function of tumour size: a study of data from six cancer registries in France (1983–2000). *European Journal of Cancer* 2007 **43** 891–900.
- 11 Krebs in Deutschland, Häufigkeiten und Trends. 2006. Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V.(GEKID) in Zusammenarbeit mit dem Robert Koch Institut (RKI).
- 12 Harach HR, Franssila KO & Wasenius VM. Occult papillary carcinoma of the thyroid. A 'Normal' finding in Finland. A systematic autopsy study. *Cancer* 1985 **56** 531–538.
- 13 Brauer VF, Hentschel B & Paschke R. Euthyroid thyroid nodules, aims, results and perspectives concerning drug therapy. *Deutsche Medizinische Wochenschrift* 2003 **128** 2381–2387.
- 14 Lang W, Borrusch H & Bauer L. Occult carcinomas of the thyroid. Evaluation of 1,020 sequential autopsies. *American Journal of Clinical Pathology* 1988 **90** 72–76.
- 15 Martinez-Tello FJ, Martinez-Cabrera R, Fernandez-Martin J, Lasso-Oria C & Ballestin-Carcavilla C. Occult carcinoma of the thyroid. A systematic autopsy study from Spain of two series performed with two different methods. *Cancer* 1993 **71** 4022–4029.
- 16 Roti E, Rossi R, Trasforini G, Bertelli F, Ambrosio MR, Busutti L, Pearce EN, Braverman LE & Degli Uberti EC. Clinical and histological characteristics of papillary thyroid microcarcinoma: results of a retrospective study in 243 patients. *Journal of Clinical Endocrinology and Metabolism* 2006 **91** 2171–2178.
- 17 Knudsen N, Perrild H, Christiansen E, Rasmussen S, Dige-Petersen H & Jorgensen T. Thyroid structure and size and two-year follow-up of solitary cold thyroid nodules in an unselected population with borderline iodine deficiency. *European Journal of Endocrinology* 2000 **142** 224–230.
- 18 Belfiore A, La Rosa GL, Padova G, Sava L, Ippolito O & Vigneri R. The frequency of cold thyroid nodules and thyroid malignancies in patients from an iodine-deficient area. *Cancer* 1987 **60** 3096–3102.
- 19 Bahre M, Hilgers R, Lindemann C & Emrich D. Thyroid autonomy: sensitive detection *in vivo* and estimation of its functional relevance using quantified high-resolution scintigraphy. *Acta Endocrinologica* 1988 **117** 145–153.
- 20 Krohn K, Wohlgemuth S, Gerber H & Paschke R. Hot microscopic areas of iodine-deficient euthyroid goitres contain constitutively activating TSH receptor mutations. *Journal of Pathology* 2000 **192** 37–42.
- 21 Laurberg P, Pedersen KM, Vestergaard H & Sigurdsson G. High incidence of multinodular toxic goitre in the elderly population in a low iodine intake area vs. high incidence of Graves' disease in the young in a high iodine intake area: comparative surveys of thyrotoxicosis epidemiology in East-Jutland Denmark and Iceland. *Journal of Internal Medicine* 1991 **229** 415–420.
- 22 Emrich D, Erlenmaier U, Pohl M & Luig H. Determination of the autonomously functioning volume of the thyroid. *European Journal of Nuclear Medicine* 1993 **20** 410–414.
- 23 Bennedbaek FN, Perrild H & Hegedus L. Diagnosis and treatment of the solitary thyroid nodule. Results of a European Survey. *Clinical Endocrinology* 1999 **50** 357–363.

- 24 Bennedbaek FN & Hegedus L. Management of the solitary thyroid nodule: results of a North American Survey. *Journal of Clinical Endocrinology and Metabolism* 2000 **85** 2493–2498.
- 25 Fuhrer D, Mugge C & Paschke R. Questionnaire on management of nodular thyroid disease (Annual Meeting of the Thyroid Section of the German Society of Endocrinology 2003). *Experimental and Clinical Endocrinology and Diabetes* 2005 **113** 152–159.
- 26 Paschke R, Reinert C, Fuhrer D, Schmid KW, Dralle H & Brabant G. Recommendations and unanswered questions in the diagnosis and treatment of thyroid nodules. Opinion of the Thyroid Section of the German Society for Endocrinology. *Deutsche Medizinische Wochenschrift* 2005 **130** 1831–1836.
- 27 Gharib H & Papini E. Thyroid nodules: clinical importance, assessment, and treatment. *Endocrinology and Metabolism Clinics of North America* 2007 **36** 707–735 (vi).
- 28 Diehl LA, Garcia V, Bonnema SJ, Hegedus L, Albino CC & Graf H. Management of the nontoxic multinodular goiter in Latin America: comparison with North America and Europe, an Electronic Survey. *Journal of Clinical Endocrinology and Metabolism* 2005 **90** 117–123.
- 29 Wiersinga WM. Is repeated fine-needle aspiration cytology indicated in (benign) thyroid nodules? *European Journal of Endocrinology* 1995 **132** 661–662.
- 30 Lucas A, Llatjos M, Salinas I, Reverter J, Pizarro E & Sanmarti A. Fine-needle aspiration cytology of benign nodular thyroid disease. Value of re-aspiration. *European Journal of Endocrinology* 1995 **132** 677–680.
- 31 Chehade JM, Silverberg AB, Kim J, Case C & Mooradian AD. Role of repeated fine-needle aspiration of thyroid nodules with benign cytologic features. *Endocrine Practice* 2001 **7** 237–243.
- 32 Richter B, Neises G & Clar C. Pharmacotherapy for thyroid nodules. A systematic review and meta-analysis. *Endocrinology and Metabolism Clinics of North America* 2002 **31** 699–722.
- 33 Paphavasit A, Thompson GB, Hay ID, Grant CS, van Heerden JA, Ilstrup DM, Schleck C & Goellner JR. Follicular and Hurthle cell thyroid neoplasms. Is frozen-section evaluation worthwhile? *Archives of Surgery* 1997 **132** 674–678.
- 34 Cheung K, Roman SA, Wang TS, Walker HD & Sosa JA. Calcitonin measurement in the evaluation of thyroid nodules in the United States: a cost-effectiveness and decision analysis. *Journal of Clinical Endocrinology and Metabolism* 2008 **93** 2173–2180.
- 35 Bhagat MC, Dhaliwal SS, Bonnema SJ, Hegedus L & Walsh JP. Differences between endocrine surgeons and endocrinologists in the management of non-toxic multinodular goitre. *British Journal of Surgery* 2003 **90** 1103–1112.
- 36 Borget I, De Pouvourville G & Schlumberger M. Editorial: calcitonin determination in patients with nodular thyroid disease. *Journal of Clinical Endocrinology and Metabolism* 2007 **92** 425–427.
- 37 Hodak SP & Burman KD. The Calcitonin Conundrum – is it time for routine measurement of serum calcitonin in patients with thyroid nodules? *Journal of Clinical Endocrinology and Metabolism* 2004 **89** 511–514.
- 38 American Association of Clinical Endocrinologists. Protocol for Standardized Production of Clinical Practice Guidelines. *Endocrine Practice* 2004 **10** 353–361.
- 39 US Preventive Services Task Force Ratings: strength of recommendations and quality of evidence. Guide to clinical preventive services, Third edition: periodic updates, 2000–2003. Agency for Healthcare Research and Quality, Rockville, MD. <http://www.ahrq.gov/clinic/3rduspstf/ratings.htm>.
- 40 National Comprehensive Cancer Network. Thyroid Cancer. http://www.nccn.org/professionals/physician_gls/PDF/thyroid. Accessed December 1, 2007.
- 41 British Thyroid Association & Royal College of Physicians. Guidelines for management of thyroid cancer. <http://www.btf-thyroid.org/Accessed December 1, 2007>.
- 42 AGREE collaboration. Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument. <http://www.agreecollaboration.org>. Accessed November 24, 2007.
- 43 Marqusee E, Benson CB, Frates MC, Doubilet PM, Larsen PR, Cibas ES & Mandel SJ. Usefulness of ultrasonography in the management of nodular thyroid disease. *Annals of Internal Medicine* 2000 **133** 696–700.
- 44 Tan GH, Gharib H & Reading CC. Solitary thyroid nodule. Comparison between palpation and ultrasonography. *Archives of Internal Medicine* 1995 **155** 2418–2423.
- 45 Hagag P, Strauss S & Weiss M. Role of ultrasound-guided fine-needle aspiration biopsy in evaluation of nonpalpable thyroid nodules. *Thyroid* 1998 **8** 989–995.
- 46 Niccoli P, Wion-Barbot N, Caron P, Henry JF, de Micco C, Saint Andre JP, Bigorgne JC, Modigliani E & Conte-Devolx B. Interest of routine measurement of serum calcitonin: study in a large series of thyroidectomized patients. The French Medullary Study Group. *Journal of Clinical Endocrinology and Metabolism* 1997 **82** 338–341.
- 47 Elisei R, Bottici V, Luchetti F, Di Coscio G, Romei C, Grasso L, Miccoli P, Iacconi P, Basolo F, Pinchera A & Pacini F. Impact of routine measurement of serum calcitonin on the diagnosis and outcome of medullary thyroid cancer: experience in 10,864 patients with nodular thyroid disorders. *Journal of Clinical Endocrinology and Metabolism* 2004 **89** 163–168.
- 48 Pacini F, Burroni L, Ciuli C, Di Cairano G & Guarino E. Management of thyroid nodules: a clinicopathological, evidence-based approach. *European Journal of Nuclear Medicine and Molecular Imaging* 2004 **31** 1443–1449.
- 49 Zelmanovitz F, Genro S & Gross JL. Suppressive therapy with levothyroxine for solitary thyroid nodules: a double-blind controlled clinical study and cumulative meta-analyses. *Journal of Clinical Endocrinology and Metabolism* 1998 **83** 3881–3885.
- 50 Wemeau JL, Caron P, Schvartz C, Schlienger JL, Orgiazzi J, Cousty C & Vlaeminck-Guillem V. Effects of thyroid-stimulating hormone suppression with levothyroxine in reducing the volume of solitary thyroid nodules and improving extranodular nonpalpable changes: a randomized, double-blind, placebo-controlled trial by the French Thyroid Research Group. *Journal of Clinical Endocrinology and Metabolism* 2002 **87** 4928–4934.
- 51 Castro MR, Caraballo PJ & Morris JC. Effectiveness of thyroid hormone suppressive therapy in benign solitary thyroid nodules: a meta-analysis. *Journal of Clinical Endocrinology and Metabolism* 2002 **87** 4154–4159.
- 52 www.british-thyroid-association.org.
- 53 Solomon BL, Wartofsky L & Burman KD. Current trends in the management of well differentiated papillary thyroid carcinoma. *Journal of Clinical Endocrinology and Metabolism* 1996 **81** 333–339.

Received 21 July 2008

Accepted 24 July 2008