Users' guides to the surgical literature: how to use an article about a diagnostic test

Stuart Archibald, MD; Mohit Bhandari, MD, MSc; Achilleas Thoma, MD; for the Evidence-Based Surgery Working Group*

See Correction

Of the many challenges in surgical practice, none is more arduous for you, the surgeon, than the process of refining your diagnostic skill. Typically, patients are initially evaluated by the time-honoured approaches of taking a history and performing a physical examination. Further information is often needed. Diagnostic tests are selected, performed and interpreted to best discriminate between patients who will likely benefit from surgical intervention and patients who will not. In this article on evidence-based surgery, we show how to locate the best evidence in the current literature on diagnostic tests and apply this evidence to your patient.

Clinical scenario

As a busy surgeon, you are increasingly seeing patients referred to you with thyroid nodules. After you have taken the patient's history and carried out a physical examination, your usual next step, in keeping with current practice,[1] is to perform a fine-needle aspiration biopsy (FNAB) on the nodule. Your most recent case involves a 39-year-old woman who reluctantly presents to your office, armed with an article from a medical site on the Internet, which suggests that because clinical thyroid cancer is so uncommon, it is safe to ignore asymptomatic nodules. She has been aware of her nodule for 2 months and it has remained asymptomatic. She has no symptoms of airway or esophageal compression or invasion. Her voice has remained normal. There is no history of thyroid disease. She has never received previous radiotherapy to the head or neck area. The family history is negative.

Examination shows a solitary nodule (3.5 cm in diameter), which is firm but not fixed. There are no clinically positive lymph nodes. Indirect laryngoscopy yields normal findings. None of the most clinically diagnostic features of malignant disease, such as a paralyzed recurrent laryngeal nerve, fixation of the mass or metastatic nodal disease, are present. The most relevant clinical feature in this case is the firmness of the lesion, but you realize this is a subjective finding. The next most worrisome clinical feature in this case is the size of the lesion.[1] Based on your clinical examination, you believe the risk of malignant disease in this patient is low to moderate, approximately 5%. Next you perform a FNAB on the nodule. The result is not clearly "benign" or "malignant." Instead, it is reported as "uncertain," showing "cellular smear, with many follicular cells, some of which show atypia; some colloid is present." You would like to estimate the risk of malignant disease for this patient before making a recommendation to her. To use the literature for help, you carry out a MEDLINE search.

The MEDLINE search

You set out to identify an article that will give you information about the properties of FNAB as it applies to your patient, so you perform a MEDLINE search.[2] You type in the words "thyroid diseases," with the

restriction "diagnosis" and run the search initially for 1996 to 1999, sequentially adding intervals of time until the period from 1976 to 1999 is included. This yields 1027 references. For the same time period, you search under the heading "biopsy, needle" with no restrictions; not surprisingly this yields an enormous number of articles, 19 733, since all biopsy sites are included. You narrow down this number by combining these 2 searches. This reduces the number of articles to 148. You further restrict the search, limiting it to "English language" only, which yields 87 references. From these, you identify 10 papers that look promising from their abstracts because the series are large enough (you decide that at least 100 patients or diagnostic tests must be done to give adequate experience for study), the FNAB data are backed up by histologic findings, the centres from which the publications originate are in regions where the practice patterns and cytology expertise are likely similar to your own, and the raw data for further calculations are included. You retrieve 8 original papers for review, and from their references you identify further relevant articles. Consultation with local experts garners a few more papers not identified by the search strategy. From all of these, a single paper, by Hamming and associates,[3] catches your eye because it treats the clinical examination as the first diagnostic test and FNAB as the second. In this study, 530 patients with nodular thyroid disease underwent FNAB. One hundred and sixty-nine (32%) were selected for thyroidectomy based on the results of clinical examination or FNAB, or both. These patients who had surgery were retrospectively divided into 3 clinical groups with high, moderate or low suspicion of malignant neoplasms according to preoperatively recorded clinical features, without knowledge of cytologic or histologic results. These 3 clinical categories were compared to the "gold standard" (reference standard) of histologic examination. The FNAB results were categorized into malignant, uncertain or benign, and these 3 groups were also compared in a blind fashion to the same reference standard. Thus, both the clinical examination result and the cytologic diagnosis were each separately compared with the surgical pathological findings. Sufficient data were provided to allow calculation of the commonly described properties of diagnostic tests. You identify that the major weakness in this study is that 68% of the entire group of patients were test-negative; that is, they were considered so unlikely to have malignant disease based on the clinical examination supplemented by the FNAB in some cases, that the reference standard of surgery was never applied to them.

Introduction

Thyroid nodules are common[4,5] and occur as solitary clinical masses in about 4% of the population, the prevalence increasing with age. Yet, cancers arising in thyroid nodules are rare, usually characterized by indolent growth, and are frequently curable with surgical treatment. The difficulty is in identifying the few patients who will benefit from thyroid surgery. Ultrasonography evaluates the morphologic characteristics of the nodule.[6] Radionuclide scanning provides indirect assessment of the malignant potential by examining activity in the nodule.[7] These tests typically identify many benign lesions as probably malignant, and if surgery is performed on the basis of the results, many unnecessary procedures are done.[6,7]

FNAB, initially developed in Europe,[8] has become widely accepted in North America. It is rapidly performed as a simple, safe, inexpensive office procedure. It has been standardized, as to technique and cytologic interpretation.[9] Typically, 4 categories are reported: benign; malignant; uncertain, indeterminate or suspicious; and inadequate, no reading or insufficient cells. Inadequate specimens account for 5% to 20% of thyroid aspirations,[10,11] with repeat aspiration reducing this by half. There remain 10% to 20% of aspirates for which the results are uncertain, usually because the cytologist could not discriminate between the benign and malignant forms of follicular and Hurthle cell nodules.[12] Your referred patient falls into this category.

In using the literature to evaluate a diagnostic test, you must now follow 3 key steps, posed here as 3 questions (Table 1):[13]
- Are the results of the study valid?
- What are the results?
- Will the results help me in caring for my patient?

Table 1
Guidelines for Evaluating Studies About a Diagnostic Test
Are the results of the study valid?
Primary guides
Is there an independent, blind comparison with a reference standard?
Does the patient sample include an appropriate spectrum of patients to whom the diagnostic test will be applied in clinical practice?
Secondary guides
Do the results of the test being evaluated influence the decision to perform the reference standard?
Are the methods for the performing the test described in sufficient detail to permit replication?
What are the results?
Are likelihood ratios of the test being evaluated or data necessary for their calculation provided?
Will the results help me in caring for my patients?
Will the reproducibility of the test result and its interpretation be satisfactory in my setting?
Are the results applicable to my patient?
Will the results change my management?
Will patients be better off as a result of the test?

Are the results of the study valid?

Primary guides

The validity of the results are only as good as the methods used to perform the study. First, you must determine if the study results can be believed by considering how the authors assembled the patients, and if the diagnostic test was compared to an appropriate reference standard.

- **Is there an independent, blind comparison with a reference standard?** The primary guide to examining the validity of a study includes the use of an independent, blind comparison with a reference standard. The accuracy of a diagnostic test is best determined by comparing it with the "truth." In the study by Hamming and associates,[3] 2 diagnostic tests were studied. The first, clinical examination at the time of presentation, covered a range of signs and symptoms that allowed each patient to be assigned retrospectively to 1 of 3 categories according to their probability of having cancer. This assignment was done without knowledge of the FNAB or pathological findings. All of the patients who were assigned to a category had the reference standard of surgical pathological examination applied to their nodules. The second test used was the FNAB. All patients who underwent thyroidectomy (and thus had surgical pathological examination) had FNAB done preoperatively. Thus, you are satisfied that for each of the 2 tests, an appropriate reference standard has been used without knowledge of the results of either test. It is safe to conclude that the histologic findings are independent of both the initial clinical examination and the preoperative FNAB.
- **Does the study sample include an appropriate spectrum of patients to which the diagnostic test is to be applied?** Another im-

portant guide to assessing validity involves the inclusion of an appropriate spectrum of patients to whom the diagnostic test will be applied in clinical practice. Any test can distinguish between severely affected and healthy patients; this tells us nothing about the clinical utility of a test. The practical value of a diagnostic test is based on its usefulness in those patients who are commonly encountered in clinical practice. In the paper by Hamming and associates the patients were accrued through a university hospital. It is possible that some preselection bias was present because of this recruitment setting, so that those more likely to have serious problems were overrepresented. For a rare condition such as cancer of the thyroid, such over-representation can be an advantage. It ensures that the test is applied to a large enough sample of "diseased" patients so that the advantages and disadvantages of the test can be determined. After all, our purpose is not to learn of the prevalence of thyroid cancer in the community but to assess diagnostic tests. A broad spectrum of diagnoses, from low to moderate to high clinical suspicion were included, with a total of 169 subjects, 39 of whom had cancer. Therefore we conclude that an appropriate patient sample was chosen.

Secondary guides

Having met the primary guidelines for study validity, you can be confident that the study in question likely represents an unbiased estimate of the real accuracy of clinical examination and of FNAB; that is, it does not distort the truth. You can further reduce your chances of being misled by asking a few more key questions.

- **Do the results of the test being evaluated influence the decision to perform the reference standard test?** The properties of a diag-

nostic test can be distorted if its results influence the use of the reference standard. This "verification bias" occurs when patients suspected of having malignant disease as a result of the clinical examination or FNAB are more likely to undergo surgical resection for pathological examination than those patients who had negative test results. As you examine the methods section of the current article you realize that "verification bias" was indeed a problem in the study. Those patients with "negative" test results did not undergo surgery, therefore no reference standard was applied to this group. As a result, the true-negative and false-negative rates for the tests are inaccurate. Those patients with positive test results (including those with malignant or uncertain findings on FNAB) underwent further reference standard examination with surgical resection and pathological examination. Thus the true-positive and false-positive rates can be determined. The problem of verification bias is unavoidable now that FNAB has become standard practice in evaluating thyroid nodules. It would be impossible now to conduct a study in which the FNAB played no role in influencing the decision to operate. Verification bias will give a false elevation to the likelihood ratios.

- **Are the methods for performing the test described in sufficient detail to permit replication?** If the authors recommend a diagnostic test, they should provide sufficient detail on how to perform it. This description should cover all the issues from patient preparation, technique, analysis and interpretation guides to the diagnostic test. In the paper by Hamming and associates, the Methods section includes the relevant information and references for the 2 diagnostic tests, the clinical examination and FNAB.

What are the results?

Since your article has met the primary guides for study validity, you are reasonably confident that the study's results will be believable. A guide to interpreting results follows. Accuracy of a diagnostic test refers to the percentage of correct diagnoses made by the test (true-positive results plus true-negative results) out of the total number of tests performed. This characteristic, which is the most often quoted, does not give enough information about the test to truly evaluate its performance, because no test is 100% accurate. Errors can occur in 2 ways. A false-negative diagnosis in our scenario will provide a false sense of security concerning a malignant lesion, thus delaying treatment. For example, if the FNAB specimen was reported as "benign" but the patient actually had cancer, no operation would be recommended. On the other hand, a false-positive result will lead to an unnecessary operation. More informative measures of a diagnostic test are the likelihood ratio (LR) of a positive test, the LR of a negative test, sensitivity, specificity, positive predictive value and negative predictive value. The Appendix shows how these are defined and calculated using a two-by-two table.

LRs possess important properties: they answer a clinically important question, can be calculated for each stratum of the test result and are unaffected by changes in the prevalence of the disease in the population under study.[14] Thus, the next question is: **Are likelihood ratios for the test results presented or is the data necessary for their calculation included**?

The starting point for any diagnostic process is the pre-test chance, expressed as a percentage, of the outcome in question. For instance, when all you know on referral is that a particular patient has a thyroid nodule, then the pre-test (or preclinical examination) chance that the nodule is malignant is equal to the chance of malignant disease in all patients with thyroid nodules referred to you. Once you have carried out the diagnostic test of clinical examination, then a new "post-test" chance that the nodule is malignant can be stated. This post-test chance becomes the pre-test chance of the next diagnostic test (in this case FNAB), if such is necessary, which allows you to refine the chance of malignant diagnosis as test information becomes available from this and successive tests. Sometimes, the post-test chance from the first test will be high enough to make the diagnosis secure, and no further tests will be necessary. In the scenario we present, this would have been the case if the clinical examination had identified some of the conclusively diagnostic signs, such as a fixed mass with a paralyzed recurrent laryngeal nerve. Each diagnostic test, whether used alone or in a sequence, alters the pre-test chance of the diagnosis to a new post-test chance of the disease in question. The direction and magnitude of this change from pre-test to post-test chance of the diagnosis are determined by the properties of the LR test.

How can we use the LR? Basically, the LR tells you how much the pre-test chance of a specific diagnosis increases or decreases. For instance, when the LR = 1.0, there is no change in the pre-test to post-test chance of the diagnosis. When the LR is greater than 1, the post-test chance of the diagnosis has been increased by the test from the pre-test chance. A rough guide to the interpretation of LRs is as follows: an LR greater than 10 or an LR less than 0.1 generates large and often conclusive changes in the post-test chance of a diagnosis. LRs between 2 and 5 or between 0.5 and 0.2 generate small, but sometimes important, changes in the chance of a diagnosis, and an LR between 1 and 2 or between 0.5 and 1.0 alters the post-test chance of the diagnosis only to a small degree.[14]

Table 2 has been constructed from the results of the clinical examination diagnostic test, using Hamming's data. The LR for "high" clinical suspicion is 8, which means that such patients have a moderately large shift from pre-test to post-test chance of the disease. The LR for "moderate" clinical suspicion is 0.5, meaning that such patients are half as likely to have the disease after the clinical examination as they were before the examination! The "moderate" and the "low" groups cannot be distinguished from each other by clinical examination only. We cannot conclude that these patients do not have malignant disease, because in Hamming's paper, 9 of 64 (14%) patients in this category did have cancer. The clinical examination alone is therefore a poor test to determine which patient has a malignant thyroid nodule and which patient does not. A further diagnostic test is needed, in this case the FNAB. By a similar process, the properties of the FNAB diagnostic test can be calculated from the paper (Table 3). There were 38 patients with proven malignant disease and 126 patients in whom malignant disease was ruled

Table 2			
Clinical Examination of a Thyroid Nodule			
	No. (%) of patients		
Clinical chance of malignant disease	Cancer present (positive)	Cancer absent (negative)	Total
High	22 (56)	9 (6.9)	31
Moderate	9 (23)	55 (42.3)	64
Low	8 (20)	66 (50.8)	74
Total	39 (100)	130 (100)	169
Likelihood ratios for high clinical suspicion = 8.1 (22/39)/(9/130): for moderate clinical suspicion = 0.5 (9/39)/(55/130): for low clinical suspicion = 0.4 (8/39)/(66/130).			

out. For each patient, the FNAB specimen was categorized as indicating high, uncertain or low chance of malignancy. Twenty-nine of the 38 patients (76%) with malignant disease had a positive FNAB result. Two of 126 patients (1.6%) without malignant disease had a positive FNAB result. The LR for these 2 is 48. In other words, a positive FNAB result is 48 times more likely to occur in patients with malignant disease than in those without malignant disease, which is conclusive. Using similar thinking, we can calculate the LR for an uncertain FNAB result as 5.7, and for a negative FNAB result as 1.1.

Once the LRs are determined, how do we use them to move from pre-test to post-test chances of malignant disease? A nomogram for converting the pre-test chance of a diagnosis to the post-test chance has been published by Fagan[15] (Fig. 1). The post-test chance is obtained by anchoring a ruler at the pre-test chance of the diagnosis, crossing the LR for the given diagnostic test and reading the post-test chance of the diagnosis from the right column.

We do not know the original chance that a thyroid nodule is malignant. In Hamming's paper, 39 cancers were discovered from a total sample of 530 patients, a 7.4% chance. Such a chance of cancer is much higher than the prevalence of clinically important thyroid cancers in the general population.[16,17] Although clinically recognized thyroid nodules occur in 2% to 7% of the population,[11] clinically recognized thyroid cancer occurs in only 1:25 000 people per year,[15] or 0.004%. Instead of using Hamming's chance of malignant disease, let us arbitrarily use 1% as a "guesstimate" of the chance of malignant disease in patients referred to your surgical practice. Your clinical examination has categorized the patient as having a high risk of malignant disease, according to Hamming's categories, based on the firmness of the lesion. The LR for this category is 8. From the nomogram we see that the post-test chance of malignant disease is 5%. Using this as the new pre-test chance of malignant disease for the FNAB result obtained, we see from the nomogram that the post-FNAB test chance for cancer using the LR of an uncertain FNAB of 5.7 (Table 3) is about 22%.

This result means that approximately 1 out of every 5 patients who have thyroidectomy in this scenario will benefit by having their cancer treated. However, in discussing surgery with the patient, we must keep in mind that all thyroidectomy patients will be exposed to the risks of the procedure.

Will the results help me in caring for my patient?

Having assessed the validity of the article and performed the necessary simple calculations to determine the results, we can ask ourselves whether these results can help us in caring for our patient. The value of a diagnostic test often depends on its reproducibility when applied to patients. If a test requires much interpretation (e.g., electrocardiogram, and pathological and cytologic specimens) or uses laboratory assays (staining, biochemical assays), variation in test results can occur. In our case, the LRs from other studies show that for the second test you used, namely FNAB, the LR of a positive result is consistently conclusive, with LRs
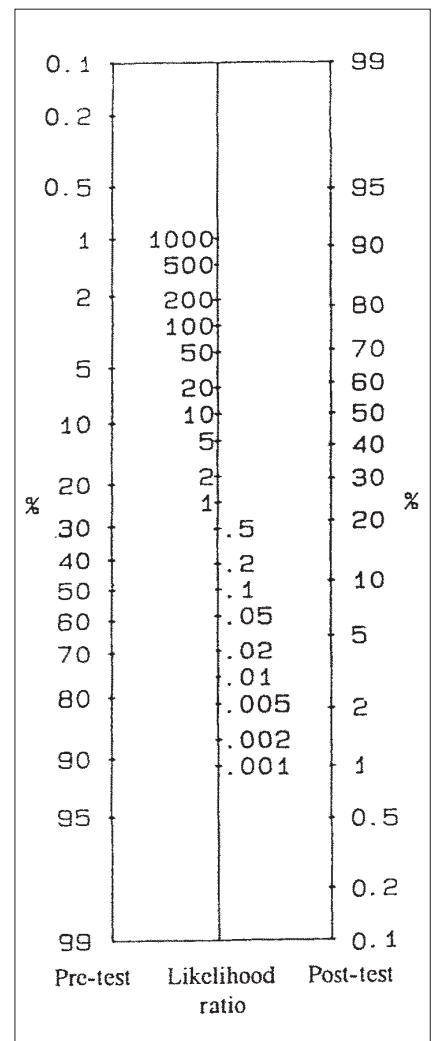

FIG. 1. Nomogram for interpreting diagnostic test results. Adapted from Fagan TJ. A nomogram for applying likelihood ratios [letter]. *N Engl J Med* 1975; 293: 257. Copyright © 1975 Massachusetts Medical Society. Adapted with permission, 2000. All rights reserved.

Table 3			
Fine-Needle Aspiration Biopsy (FNAB) of a Solitary Thyroid Nodule: Likelihood Ratios for Malignant Disease			
FNAB result	Cancer present (positive)	Cancer absent (negative)	Total
Positive	29	2	31
Uncertain	6	35	41
Negative	3	89	92
Total	38	126	164*

*5 of 169 cases were excluded from this tabulation because the specimens were considered inadequate.
Likelihood ratio of a cancer when FNAB is read as positive = 48.1 (29/38)/(2/126); 94% of these are cancer.
Likelihood ratio of cancer when FNAB is read as uncertain = 5.7 (6/38)/(35/126); 15% of these are cancer.
Likelihood ratio of cancer when FNAB is read as negative = 1.1 (3/38)/(89/126); 3% of these are cancer.

varying from 9.7 to 260; the LR of an uncertain result varies from 1.1 to 7.6 and the LR of a negative result varies from 0.07 to 1.1.[3,18–20]

Before making any treatment decisions, you must have a threshold for recommending surgery. Above this threshold, you will recommend surgery, and below it you will not. This threshold is set by balancing the risks of a particular treatment with the benefits of such treatment. Some of the risks will be inherent to a particular treatment. Others will be determined by patient factors such as comorbid conditions and age, and by surgeon factors such as training and experience. The risks of this particular treatment, thyroidectomy, are low, and adverse events are almost never life-threatening. Furthermore, thyroid cancer almost always requires surgery. Therefore, you have a relatively low threshold for recommending thyroidectomy. For patients with a thyroid nodule, as in most clinical situations, you as a surgeon will have already set a threshold. Let us say it is set at 20%;[21] that is, if the post-test chance of cancer is 20% or more, you will recommend operation, whereas if the post-test chance of cancer is less than 20%, you will not recommend surgery. Given this information, the nomogram can be used in reverse. Set your ruler at 20% on the right column and use the upper and lower reported LRs for the uncertain result for FNAB in the literature, noted as 7.6 and 1.1 . This gives the lower and upper limits for the pre-test chance that the patient has cancer as about 3% and 20%. Once this interval is known, you then ask yourself whether the patient, in your estimation, falls within this interval for the pre-test chance of cancer; that is, do you believe that this patient, after the clinical examination, had a 3% to 20% chance of having cancer? As you will recall from the initial account above, you estimated that this patient had about a 5% chance of having cancer after the clinical examination only. This helps to confirm that

indeed you should recommend surgery for your patient. However, if you set your threshold for recommending surgery much higher, say at 50% due to premorbid conditions, then you would not recommend surgery now.

If a study reports a test as being highly reproducible, 2 possibilities are likely. Either the test is quite simple and easy to apply to patients or the investigators in the study are highly skilled in applying this diagnostic test to the study patients. If the latter is true, the diagnostic test may not be useful in a setting in which nonskilled interpretation of the test is likely. Hamming and associates[3] refer to a standard classification by the WHO,[22] but there is no information regarding the reliability (or inter- and intra-observer variation) of the FNAB among the investigators in this study.

Another important issue to consider is the similarity of your patient to those included in the study. The properties of a diagnostic test can change with different degrees of disease severity. For instance, LRs tend to increase when patients with the target disorder all have severe disease, and tend to diminish toward the value of 1 when patients with the target disorder have mild disease.[23] In general, however, if you practise in a similar setting to that presented in the study and your patient meets the study eligibility criteria, you can be confident in applying the results of the study to your patient. The patients in the current study by Hamming and associates spanned a wide spectrum of disease. However, they were accrued in a university hospital setting in the Netherlands, which may affect the generalizability of the results to your practice.

The value of the clinical examination and the FNAB in your patient has been that you now have a quantitative estimate of the patient's probability of cancer. Of course, there are nuances on clinical examination that are not discussed in any published series, which may make

the published series less generalizeable to your particular patient. Certainly, all diagnostic tests have to be considered in the overall clinical context of the patient.

Finally, you can ask yourself if your patient will be better off having had the test. In other words, does this test add to your current knowledge of the patient's condition. A test becomes more valuable when it has acceptable risks, if the target disorder left untreated has major consequences, and if the disorder can be readily treated if diagnosed. These conditions are met in the case of the thyroid nodule.

Conclusions

Application of the guidelines presented in this article will allow surgeons to critically assess studies about a diagnostic test. Surgeons are increasingly overwhelmed with a growing body of literature describing new and innovative diagnostic tests. Using the approach outlined here, you the surgeon can determine from the literature the validity of a study, the results from a study and the applicability of these study results to your patient in order to optimize patient care.

References

1. Tuttle RM, Lemar H, Burch HB. Clinical features associated with an increased risk of thyroid malignancy in patients with follicular neoplasia by

fine-needle aspiration. *Thyroid* 1998; 8(5):377-83.

2. Oxman AD, Sackett DL, Guyatt GH, The Evidence-Based Medicine Working Group. User's guides to the medical literature. I. How to get started. *JAMA* 1993;270(17):2093-5.

3. Hamming JF, Goslings BM, van Steenis GJ, van Ravenswaay Claasen H, Hermans J, van de Velde CJ. The value of fine-needle aspiration biopsy in patients with nodular thyroid disease divided into groups of suspicion of malignant neoplasms on clinical grounds. *Arch Intern Med* 1990;150 (1):113-6.

4. Mazzaferri EL. Management of a solitary thyroid nodule. *N Engl J Med* 1993;328(8):553-9.

5. Tunbridge WM, Evered DC, Hall R, Appleton D, Brewis M, Clark F, et al. The spectrum of thyroid disease in a community: the Whickham survey. *Clin Endocrinol (Oxf)* 1977;7(6): 481-93.

6. Ashcraft MW, Van Herle AJ. Management of thyroid nodules. I: history and physical examination, blood tests, X-ray tests, and ultrasonography [review]. *Head Neck Surg* 1981;3(3):216-27.

7. Ashcraft MW, Van Herle AJ. Management of thyroid nodules. II: scanning techniques, thyroid suppressive therapy, and fine needle aspiration [review]. *Head Neck Surg* 1981;3(4): 297-322.

8. Soderström N. Puncture of goiters for aspiration biopsy. *Acta Med Scand* 1952;144:237-44.

9. Guidelines of the Papanicolaou Society of Cytopathology for the Examination of Fine-Needle Aspiration Specimens from Thyroid Nodules. The Papanicolaou Society of Cytopathology Task Force on Standards of Practice. *Mod Pathol* 1996;9(6):710-5.

10. Goellner JR, Gharib H, Grant CS, Johnson DA. Fine needle aspiration cytology of the thyroid, 1980 to 1986. *Acta Cytol* 1987;31(5):587-90.

11. Mazzaferri EL, de los Santos ET, Rofagha-Keyhani S. Solitary thyroid nodule: diagnosis and management *Med Clin North Am* 1988;72(5): 1177-211.

12. Gharib H, Goellner JR, Zinsmeister AR, Grant CS, Van Heerden JA. Fine-needle aspiration biopsy of the thyroid. The problem of suspicious cytologic findings. *Ann Intern Med* 1984;101(1):25-8.

13. Jaeschke R, Guyatt G, Sackett DL, The Evidence-Based Medicine Working Group. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A: Are the results of the study valid? *JAMA* 1994;271(5):389-91.

14. Jaeschke R, Guyatt G, Sackett DL, The Evidence-Based Medicine Working Group. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B: What are the results and will they help me in caring for my patients? *JAMA* 1994;271(9):703-7.

15. Fagan TJ. A nomogram for applying likelihood ratios [letter]. *N Engl J Med* 1975;293:257.

16. National Cancer Institute, Division of Cancer Prevention and Control. *1986 annual cancer statistics review.* Bethesda (MD): National Institutes of Health; 1986. NIH pub no. 86-2789.

17. *Third National Cancer Survey: Incidence Data.* Monograph 41, Bethesda (MD): National Cancer Institute; 1975. p. 107. DHEW pub no. (NIH) 75-787.

18. Leonard N, Melcher DH. To operate or not to operate? The value of fine needle aspiration cytology in the assessment of thyroid swellings. *J Clin Pathol* 1997;50:941-3.

19. Grant CS, Hay ID, Gough IR, McCarthy PM, Goellner JR. Long-term follow-up of patients with benign thyroid fine-needle aspiration cytologic diagnoses. *Surgery* 1989;106(6):980-6.

20. Van Herle AJ. The thyroid nodule. *Ann Intern Med* 1982;96:221-32.

21. Berry SM, editor. *Mont Reid surgical handbook.* 4th ed. Chicago: Mosby-Year Book Publishers; 1997. p. 332-6.

22. Löwhagen T, Willems JS, Lundell G, Sundblad R, Granberg PO. Aspiration biopsy cytology and diagnosis of thyroid cancer. *World J Surg* 1981; 5:61-73.

23. Moons KG, van Es GA, Deckers JW, Habbema JD, Grobbee DE. Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: a clinical example. *Epidemiology* 1997; 8(1):12-7.

Appendix

Properties of a Diagnostic Test

Property	Reference standard is positive	Reference standard is negative	Total
Test is positive	a	b	a + b
Test is negative	c	d	c + d
Total	a + c	b + d	a + b + c + d

Terms and Calculations

a = true-positive results, b = false-positive results, c = false-negative results, d = true-negative results.

Sensitivity = a /(a + c) = true-positive rate = "positivity in disease." Question answered: What proportion of patients with disease is correctly identified by the test?

Specificity = d/(b + d) = true-negative rate = "negativity in health." Question answered: What proportion of patients without the disease is correctly identified by the test?

Positive predictive value = a/(a + b). Question answered: What proportion of positive tests is correct?

Negative predictive value = d/(c + d). Question answered: What proportion of negative tests is correct?

Likelihood ratio of a positive test = probability of a positive test in diseased people/probability of a positive test in non-diseased people = true-positive rate/false-positive rate = [a/(a+c)]/[b/(b+d)].

Likelihood ratio of a negative test = probability of a negative test in diseased people /probability of a negative test in non-diseased people = false-negative rate/true-negative rate = [/(a+c)]/[d/(b+d)].