A practical approach to evidence-based dentistry: V

How to appraise and use an article about diagnosis

Romina Brignardello-Petersen, DDS, MSc; Alonso Carrasco-Labra, DDS, MSc, PhD(c); Michael Glick, DMD; Gordon H. Guyatt, MD, MSc; Amir Azarpazhooh, DDS, MSc, PhD, FRCD(C)

FIFTH IN A SERIES

In the previous articles in this series, we introduced the process of evidence-based dentistry (EBD),[1] how to search for evidence to inform clinical practice,[2] and how to use articles about therapy or prevention[3] and harm.[4] In this article, we will explain how to use an article to inform clinical decisions regarding questions of diagnosis. We will introduce and describe the basic concepts needed to understand diagnostic test studies, and we will explain how to use the concepts to appraise such studies critically. In subsequent articles we will describe how to use other types of study designs.

xtra Supplemental material is available online.

ABSTRACT

Background and Overview. Questions regarding diagnosis are common in dental practice. Studies in which investigators apply a diagnostic test and a reference standard to all patients and compare their results represent the best type of design to answer these questions. The critical appraisal of these studies includes an assessment of the risk of bias, results, and applicability of the study. The authors provide the concepts and guidelines that dentists can apply to most effectively use articles regarding diagnosis to guide their clinical practice.

Practical Implications. Dentists who wish to inform their clinical decisions regarding questions related to diagnostic test properties can use these guidelines to decide what type of studies to search, define the specific question of interest to search efficiently for these studies, and critically appraise studies addressing diagnosis.

Key Words. Evidence based-dentistry; diagnosis; diagnostic test studies; critical appraisal.
JADA 2015:146(3):184-191

http://dx.doi.org/10.1016/j.adaj.2015.01.011

Dr. **Brignardello-Petersen** is a lecturer, Evidence-Based Dentistry Unit, Faculty of Dentistry, University of Chile, Santiago, Chile, and a doctoral student, Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada.

Dr. **Carrasco-Labra** is a lecturer, Evidence-Based Dentistry Unit, Faculty of Dentistry, University of Chile, Santiago, Chile, and a doctoral student, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada.

Dr. **Glick** is the dean, School of Dental Medicine, University at Buffalo, The State University of New York, Buffalo, NY. He also is the editor of The Journal of the American Dental Association.

Dr. **Guyatt** is a distinguished professor, Department of Clinical Epidemiology and Biostatistics, and a joint member, Department of Medicine, McMaster University, Hamilton, Ontario, Canada.

Dr. **Azarpazhooh** is an assistant professor, Dental Public Health and Endodontics, Faculty of Dentistry; and an assistant professor, Clinical Epidemiology and Health Care Research, Institute of Health Policy, Management and Evaluation, Faculty of Medicine, University of Toronto, Ontario, Canada. He also is the head, Endodontics, Mount Sinai Hospital, Toronto, Ontario, Canada. Address correspondence to Dr. Azarpazhooh at Biological and Diagnostic Sciences, 124 Edward St., Room 515C, Toronto, Ontario, Canada M5G 1G6, e-mail amir.azarpazhooh@dentistry.utoronto.ca.

BOX 1

Clinical scenario.

During a meeting of the clinicians in your practice, a colleague raises the idea of acquiring a new laser device to help with the diagnosis of caries. This clinician explains that a sales representative described how this device was excellent for detecting noncavitated occlusal carious lesions, and your colleague wants to know your opinion. Because you believe that evidence is necessary to inform this important decision for your practice, you tell your colleague that you need to conduct a literature search and a critical appraisal to inform your opinion.

CLINICAL QUESTIONS OF DIAGNOSIS

Dentists face diagnosis questions every day. With most patients, dentists need to use diagnostic tests before they can establish a course of action to follow. In the context of everyday practice, a diagnostic test can refer to any test performed in a laboratory or any information obtained from a medical history or clinical examination that is used to confirm or rule out a specific diagnosis.[5] Dental radiographs are a common diagnostic test used by clinicians in many dental specialties, but other procedures such as performing a vitality test and measuring probing depths can be considered diagnostic tests as well.

When facing diagnosis questions, clinicians need to modify the classic Population, Intervention, Comparison, Outcome (PICO) framework for stating questions. The population is the patients of interest (that is, those to whom we will apply the diagnostic test when there is suspicion of a condition or disease). The intervention is the diagnostic test (that is, the test in which we are interested in learning). The comparison is a test we use as a reference to compare the diagnostic test against—this is called the reference standard (or gold standard).[6] Finally, the outcomes are either the health consequences after using the diagnostic test or the measures that describe the performance of a diagnostic test. Both of these cases are described later in this article. Table 1 shows examples of diagnostic test questions and their PICO components.

WHAT STUDY DESIGN BEST ADDRESSES QUESTIONS ABOUT DIAGNOSIS?

Clinicians can answer diagnostic questions by conducting studies using one of the following two types of designs: randomized clinical trials and cross-sectional studies. Ideally, a study's investigators would treat a diagnostic test as an intervention. Researchers would randomize patients to receive one of two diagnostic strategies, which for the purposes of this article we will call strategy A and strategy B. Clinicians would manage patients according to the results of the test, including providing whatever interventions they think might be appropriate on the basis of test results.

Ultimately, they would measure patient-important outcomes in the group whose participants received test strategy A and the group whose participants received test strategy B.

For example, when assessing how useful laser fluorescence is for detecting early interproximal carious lesions, researchers should randomly assign patients to undergo diagnosis with either laser fluorescence or bite-wing radiographs. Then clinicians would treat patients according to the results of the test the patients received, either laser fluorescence or bite-wing radiographs. The investigators would follow up with all patients to determine, for example, how many participants in each group have carious lesions extending into the dentin and what is each participant's need for restorations (outcomes). To date, we have not been able to identify any of these study designs in the dental literature, and therefore, they will not be further covered in this article.

In studies whose investigators address the accuracy of a diagnostic test, a group of patients undergo both tests (that is, the diagnostic test and the reference standard). The reference standard is considered to be the way to know whether the disease or condition is truly present or absent. The investigators compare the results of the diagnostic test with the reference standard as a way to determine the diagnostic properties of the diagnostic test.

For example, with the aim of assessing the accuracy of thermal and electrical dental pulp tests to diagnose pulp vitality, Villa-Chavez and colleagues[7] conducted the cold, hot, or electrical pulpal tests (3 different diagnostic tests) in 110 patients. They used as the reference standard the direct observation of the pulp after opening the pulp chamber, and then they estimated the sensitivity, specificity, predictive values, and accuracy of each of the diagnostic tests by comparing the results with those of the reference standard.

Ideally, clinicians would have available the results of systematic reviews of primary studies addressing test properties. We found that few such systematic reviews have been published in the dental literature. In the absence of reviews, clinicians look to the results of the best single primary diagnostic studies to inform their practice. In this article, we address such studies; in subsequent articles, we will describe how to use systematic reviews.

ABBREVIATION KEY. CT: Computed tomography. **EBD:** Evidence-based dentistry. **ICDAS:** International Caries Detection and Assessment System. **LFE:** Laser fluorescence examination. **PICO:** Population, Intervention, Comparison, Outcome. **RE:** Radiographic examination. **VE:** Visual examination.

BOX 2

The study you found.

During your search, you found a primary study[8] that seems to answer the question at hand. The investigators of the study[8] addressed whether using Diagnodent, a laser fluorescence device, was accurate for diagnosing noncavitated occlusal carious lesions extending to the dentin. You read the abstract of this study in which the researchers compared the diagnoses they obtained by using the laser fluorescence device with the diagnoses they obtained by doing an enameloplasty and by observing the carious lesions directly. The authors claimed that "…the laser device had an acceptable performance, this equipment should be used as an adjunct method to visual inspection to avoid false positive results." To find out whether the methods and results support the authors' conclusion, you retrieve the article and begin a critical appraisal.

CRITICALLY APPRAISING A STUDY ASSESSING THE PROPERTIES OF A DIAGNOSTIC TEST TO INFORM CLINICAL DECISIONS

The process of using an article from the dental literature involves 3 steps: an assessment of the risk of bias, an assessment of the results themselves, and an assessment of the applicability of the results.[9]

1. How serious is the risk of bias? The extent to which a study's results are likely to be correct for the sample of patients enrolled depends on how well the study was designed and conducted.[10] Factors to consider in judging the risk of bias of diagnostic test studies include whether any of the patient's conditions presented a diagnostic dilemma, whether the reference standard was appropriate and independent from the diagnostic test, whether the investigators independently interpreted the results of both tests and did not know the results of the other investigators, and whether all patients underwent both the diagnostic test and the reference standard irrespective of the results of the diagnostic test.[11] Table 2[8,12-14] lists questions that address the risk of bias associated with diagnostic tests used in studies and presents examples from the dental literature.

1a. Did participating patients present a diagnostic dilemma? Researchers performing diagnostic test studies should select patients who are representative of those to whom the test would be applied in clinical practice. For the results of a diagnostic test to be useful, the results have to discriminate between patients who have and patients who do not have the target condition (for example, in our scenario, the target condition was noncavitated occlusal caries lesion extending to the dentin) when there is a diagnostic dilemma. If patients clearly had the target condition, or clearly did not, there would not be a need to apply the diagnostic test, and in the setting of a study, the accuracy of the test would be overestimated. For example, if a patient had a cavitated mesio-occlusal carious lesion, the clinician would have no need to use a laser fluorescence device to confirm the presence of a carious lesion. In this case where it is so obvious that the carious lesion is present, the device will result in a correct diagnosis most (if not all) of the time. Therefore, for a diagnostic test study to provide a trustworthy assessment of the value of the test, researchers must include patients with early manifestations of the disease, in whom there is doubt regarding the diagnosis, similar to the patients whom clinicians will see in their daily practices.[11]

1b. Did investigators compare the test with an appropriate, independent reference standard? An appropriate reference standard is also a key aspect in assessing the risk of bias of diagnostic test studies. As described previously, clinicians assess the properties of the diagnostic test by comparing the results with the reference standard, which is considered to be the truth.[5,11] Clinicians should consider 2 aspects when assessing whether the reference standard is appropriate. First, the reference standard should be the test that is accepted most widely as the definitive test to establish a diagnosis.[5] For example, to diagnose oral squamous cell carcinoma, the reference standard is a biopsy and histologic

TABLE 1

Examples of diagnosis questions and the Population, Intervention, Comparison, Outcome framework.

CLINICAL QUESTION	POPULATION	DIAGNOSTIC TEST	REFERENCE STANDARD	OUTCOMES
How useful is cone-beam computed tomography at detecting the proximity of third-molar roots with the inferior alveolar nerve?	Patients undergoing third-molar extraction surgery	Cone-beam computed tomography	Direct observation of the inferior alveolar nerve when performing the surgery	True positives, true negatives, false positives, and false negatives
How useful is the oral pathologist's clinical observation for diagnosing an oral mucosal lichen planus?	Patients with oral mucosal lesions compatible with oral mucosal lichen planus	Pathologist's observation	Biopsy plus histologic confirmation	True positives, true negatives, false positives, and false negatives
How useful is the laser fluorescence device for detecting noncavitated occlusal caries lesions extending into the dentin?	Patients who might have a noncavitated occlusal caries lesion extending into the dentin	Laser fluorescence device	Direct observation of the caries lesion after performing an enameloplasty	True positives, true negatives, false positives, and false negatives

confirmation. Sometimes, however, using the reference standard is only the best available method for diagnosing a target condition instead of the ideal method. For example, to identify a carious lesion extending into the dentin, the reference standard would be to use a combination of clinical signs and radiograph images, as opposed to extract the tooth and perform a histologic confirmation. Considering this example, the clinician should judge whether the reference standard used in the study is an acceptable way to arrive at a definitive diagnosis.

Second, the reference standard should be independent from the diagnostic test. This means that the diagnostic test should not be part of the reference standard. For example, if the diagnostic test used to diagnose pulpal vital status was a cold test, and the reference standard was a combination of responses from thermal and electrical tests, the diagnostic test would not be independent from the reference standard. Including the test as part of the reference standard leads to overestimation of test accuracy.[11]

1c. Were the investigators who interpreted the test and reference standard blinded to the other results? Another factor to consider when appraising the risk of bias of a diagnostic test study is whether the investigators who interpreted the results of the diagnostic test and the reference standard were blinded to the results of the other test.[11] Many tests, such as radiographs or histologic confirmation, require interpretation by specialists. If the investigators who interpreted the diagnostic test or reference standard were aware of the results of the other test, they may have been influenced subconsciously when they interpreted the results of the diagnostic test or reference standard. Again, the result would be an overestimate of the accuracy of the diagnostic test.

1d. Did investigators apply the same reference standard to all patients regardless of the results of the test under investigation? Finally, it is important that researchers apply the same reference standard to all patients, irrespective of the results obtained with the diagnostic test.[11] Researchers could overestimate the accuracy of the diagnostic test if only those patients diagnosed as "target positive" by the results of the diagnostic test undergo confirmation with the reference standard, because they would not detect patients wrongly classified as "target negative" by the results of the diagnostic test. For instance, in the example in Table 2,[8] some patients who tested negative in the diagnostic test did not undergo the reference standard, and therefore, it is possible that some of these patients had lesions but nevertheless were classified as not having lesions. Again, such misclassification will make the test look more accurate than it really is.

2. What are the results? After assessing to what extent bias may influence the results of a diagnostic test study,

---

BOX 3

Your assessment of the risk of bias of the study you identified.

With respect to the patients who had diagnostic dilemmas in the study you reviewed,[8] it seems likely that only patients whose teeth showed signs of possible carious lesions were included in the study. With respect to applying the reference standard, even though the investigators used the best method (that is, enameloplasty and direct observation of the carious lesion), they did not apply this method to all patients, but rather only to those patients for whom they highly suspected having the diagnosis. It is also not clear whether the investigators who performed the test were blinded to the results of the reference standard. Even though, owing to ethical reasons, you determined that it could have been appropriate not to apply the reference standard to all patients, this omission may have led to an overestimation of the performance of the diagnostic test (eTable,[8] available online at the end of this article, provides more details).

---

clinicians must review the results to determine how to apply the test in clinical practice. The outcomes of diagnostic test studies reflect the ability of the test to discriminate between target-positive patients and target negative patients. When thinking about test results, clinicians can consider the probability of being target positive before the test is conducted (pretest probability) and the probability of being target positive after the test is conducted (posttest probability).

The most commonly used measures of accuracy in diagnostic test studies are sensitivity, specificity, probability of having the target condition if the test result is positive (that is, the posttest probability if the test result is positive or—in an unfortunate choice of words—the positive predictive value), probability of having the target condition if the test result is negative (that is, posttest probability if the test result is negative—or, in an even worse choice of terms and meaning—the probability of not having the target condition if the test result is negative [that is, the negative predictive value]), and likelihood ratios. All of these measures are defined and explained below.

Because the diagnostic test is being compared with a reference standard, if the diagnostic test is a yes-no or dichotomous test (either positive or negative, rather than presenting a range of results), patients could be classified correctly as target positive (true positive), correctly classified as target negative (true negative), incorrectly classified as target positive (false positive), and incorrectly classified as target negative (false negative).

Assessments of the accuracy of a diagnostic test are made on the basis of this classification (Figure 1). Sensitivity is the capacity of the test to correctly classify target-positive patients, whereas specificity is the ability to correctly classify target-negative patients.[15,16] Posttest probabilities, sometimes referred to as predictive values, reflect the likelihood of patients being target positive in those studies whose investigators

TABLE 2

Examples of the critical appraisal of the validity of the results of studies about diagnosis.

QUESTION	EXAMPLES	EXPLANATION
Did participating patients present a diagnostic dilemma?	"…135 patients (161 impacted teeth), …, who underwent additional examination by cone-beam CT* because of panoramic features suggesting a close relationship of the tooth root to the mandibular canal were included…"[12] "Each surface had to meet one of the three listed criteria to be included in the study: macroscopically intact occlusal fissure that exhibited absolutely no signs of caries; occlusal fissure with a discolored, brown or black area at the clinical examination without cavitation; grey discoloration from the underlying dentin without enamel breakdown."[8]	In both examples, the authors indicated that patients included in the study had characteristics that presented a diagnostic dilemma, such as features suggesting proximity of the third-molar root to the alveolar canal, or occlusal appearances representing a spectrum of patients, ranging from those who seemed healthy to those who seemed diseased. Therefore, in both of these examples, there is a low risk of bias on the basis of this criterion.
Did investigators compare the results of a test with an appropriate, independent reference standard?	"Panoramic and cone-beam CT features were correlated with the intraoperative findings, that is, the presence or absence of the inferior alveolar neurovascular bundle exposure at the time of extraction."[12] "All teeth included in the study underwent endodontic treatment, and the presence or absence of bleeding pulp in the pulp chamber on access was used as a true positive or true negative."[13]	In the first example, the reference standard was the direct observation of the inferior alveolar neurovascular bundle during surgery, which most clinicians would agree is the best method to diagnose proximity of the third-molar roots with the alveolar canal. In the second example, the clinician must judge how appropriate it is that only bleeding was considered as a sign of pulp vitality and whether this may have affected the results of the study. In both cases, the reference standard was independent from the diagnostic test, which was the cone-beam CT scan and thermal tests, respectively.
Were those investigators who interpreted the test and reference standard blinded to the other results?	"LFE† scores were analyzed using the manufacturer's cutoff points, taking into account the absence of a histological examination and the in vivo nature of the study. VE‡ and RE§ data were analyzed using the ICDAS II¶ method (Table 3), and modified criteria were validated in vivo by the method of …"[14]	When the authors described the way in which the tests were interpreted, they did not mention whether this was done by different clinicians or by the same clinician in a blinded fashion. Therefore, clinicians should consider the potential for bias owing to this factor.
Did investigators perform the same reference standard with all patients regardless of the results of the test under investigation?	"The validation method for diagnosis (gold standard) was determined by fissure eradication or enameloplasty using an invasive fissure sealing kit … However, not all fissures could be validated as this is an invasive method. Thus, for ethical reasons, opening of the cavities occurred only in cases when both examiners agreed to the presence of dentin caries."[8]	The authors mentioned not applying the reference standard to some patients, which is not appropriate (that is, it seems likely that the authors assumed that those patients were healthy). There is a risk that the investigators misclassified those patients for whom a substandard reference was used, which could have biased the results. Because this misclassificiation was done owing to ethical reasons, the clinician should judge how likely it is that bias could have occurred and what the magnitude of this bias could have been.

* CT: Computed tomography.
† LFE: Laser fluorescence examination.
‡ VE: Visual examination.
§ RE: Radiographic examination.
¶ ICDAS: International Caries Detection and Assessment System.

used positive and negative tests in the particular sample studied,[16]

Posttest probabilities when test results are positive or negative are what clinicians want to know. Unfortunately, the posttest probabilities reported in a particular study will only be accurate for clinicians' use if their patients have the same pretest probability as the population studied, which likely will be true for only a few patients.

The posttest probability of having the target condition if the test is positive is sometimes referred to as the positive predictive value. Unfortunately, instead of reporting the posttest probability of having the target condition if the test is negative (that is, the intuitive way to think of the situation), the investigators of studies often report the posttest probability of not having the target condition if the test is negative; this is referred to as the negative predictive value.[5,15] Figure 2 provides an

example of how these outcomes are calculated. Because all of these outcomes are proportions or probabilities, they can range from 0 to 1 (or from 0% to 100%), and thus, values closer to 1 reflect greater accuracy, or a greater posttest probability, of a patient's having the target condition.

The likelihood ratio (LR) is an outcome that helps clinicians moving from a pretest probability to a posttest probability of having the target condition. This number can range from zero to infinity. An LR of 1 indicates that the posttest probability of having the target condition is the same as the pretest probability. In other words, when the LR is 1, the test result has provided no useful information. Large LRs imply large increases in posttest probabilities relative to the pretest probabilities, whereas LRs close to zero imply large decreases in posttest probabilities relative to pretest probabilities.[11,17] As a rule of thumb, LRs higher than 5.0 or lower than 0.2 mean that the diagnostic test is useful for arriving at a diagnosis.[11] Unfortunately, investigators do not commonly report the LR in diagnostic test studies in dentistry; however, it can be calculated easily using the sensitivity and specificity values reported frequently in these studies (Table 3[11,17]). After LRs have been calculated, they can be used to move from specific pretest to posttest probabilities by using mathematical formulas, or more easily, graphical tools such as the Fagan nomogram.[17]

Figure 1. Outcomes of a diagnostic test study based on the relationship between the results from the diagnostic test and the reference standard.

Figure 2. Calculation of diagnostic test studies outcomes.

In summary, when appraising the results of a diagnostic test study, clinicians should look at sensitivity, specificity, and predictive values, which ideally should be close to 1. Calculating the LR for positive and negative results also can be valuable, although clinicians need to keep in mind the fact that the pretest probabilities differ

BOX 4

Your assessment of the results of the study you identified.

You find that the likelihood ratio (LR) of a positive test is 3.68, which is a somewhat large value and indicates that the test produces moderate shifts from pretest to posttest probabilities. For instance, in a patient with a 40% pretest probability, if the test is positive, the posttest probability is 71%. In addition, the LR negative is 0.09, a small LR that indicates that a negative result with the laser fluorescence test will be helpful for ruling out the presence of caries lesions (eTable,[8] available online at the end of this article, provides more details). For instance, in the same patients with a 40% pretest probability, a negative result of the test would result in a posttest probability of 5.7%. Therefore, you conclude that the laser fluorescence device is accurate for confirming or ruling out the presence of caries lesions.

TABLE 3

Calculation and interpretation of likelihood ratios.*

	Calculation	Interpretation
Likelihood Ratio for Positive Test (LR+)	$\frac{\text{Sensitivity}}{1-\text{Specificity}} = \frac{0.86}{1-0.89} = 7.81$	When the result of the diagnostic test is positive, the probability of having the disease is high. The exact probability of having the disease depends on the pretest probability and the specific LR+ result.[†]
Likelihood Ratio for Negative Test (LR−)	$\frac{1-\text{Sensitivity}}{\text{Specificity}} = \frac{1-0.86}{0.89} = 0.16$	When the result of the diagnostic test is negative, the probability of having the disease is low. The exact probability of having the disease depends on the pretest probability and the specific LR− result.[†]

* Described on the basis of numbers from Figure 2.
† The pretest probability is specific to individual patients and is estimated by the physician on the basis of the patient's medical history and clinical examination results. The shift from pretest to posttest probabilities can be estimated easily using an LR nomogram.[11,17]

among patients and individual patients in a clinical practice may not have the same pretest probability as that of the patients in the study.

3. How can I apply the results to patient care? Finally, it is necessary to assess to what extent the results of a study are applicable to a particular context. The following factors should be considered when evaluating the applicability of results published in articles about diagnosis.

3a. Will the reproducibility of the test results and its interpretation be satisfactory in my clinical setting? First, it is necessary to judge whether the diagnostic test had adequate reproducibility (that is, whether the study yielded the same results when reapplied to the same patient in the study). The investigators of diagnostic test studies often report interrater reproducibility, intrarater reproducibility, or both. These values range from 0 to 1, and values closer to 1 represent better reproducibility. For example, Jablonski-Momeni and colleagues[18] reported that the "intra-class correlation coefficient (ICC) was 0.89" when using a fluorescence-based camera for detecting occlusal carious lesions. This is a high value that makes us confident that, in a study setting, the results of the diagnostic test would be highly reproducible.

Clinicians, however, need to consider whether the test will be equally reproducible in their own clinical setting. Limitations of reproducibility include lack of optimal training or experience in applying the test or interpreting its results, or limitations in the maintenance or calibration of the necessary equipment.

3b. Are the study results applicable to the patients in my practice? If the patients in the study had different characteristics than the patients to whom the results would be applied, the performance of the test may change. A test could have different properties when evaluated in patients with a different mix of disease severity or comorbidities that could confuse the diagnosis.[11] For example, if investigators assessed a diagnostic test for detecting interproximal carious lesions in a population with enamel defects, a clinician can expect that the test would perform differently when used in patients without these defects. Therefore, the clinician should look at the selection criteria and characteristics of the patients included in the study and judge whether they are similar to the patients in their practice.

3c. Will the test results change my management strategy? A third aspect clinicians have to consider when judging the extent to which a test will be useful in their practice is the frequency with which results with LRs far from 1.0 occur. For instance, consider a positive test that has an LR of 1.5 (extremely uninformative) and a negative test that has an LR of 0.1 (leading to large decreases in posttest versus pretest probability). This will be a useful test if the negative result occurs frequently, but not if it is rarely seen.

3d. Will patients be better off as a result of the test? This final point requires clinicians to consider the consequences of applying a diagnostic test.[11] When the consequences of not diagnosing a disease are severe, it is likely that the clinician would want to apply the diagnostic test. The consequences of false-positive and false-negative results should be considered as well. Other considerations include possible adverse effects from having the test if it is invasive as well as taking on the financial cost of the test and addressing issues of convenience and burden. These judgments require clinical expertise.

BOX 5

Your assessment of the applicability of the study you identified.

When assessing the applicability of the results, you note that even though the test would be easy to use and its results would be reproducible in your clinical practice, one applicability issue is that the researchers only included teeth from young patients (that is, 10 to 13 years old), which is not the age group you usually see in your practice. As you continue to think about whether to use the test, you remember its good performance, which may be satisfactory enough to change your clinical practice. However, you also remember that in this study the conventional methods for detecting noncavitated dentin carious lesions (for example, bite-wing radiographs) performed better. You conclude that your patients would not benefit from the additional use of the laser fluorescence device for identifying the presence of caries lesions (eTable,[8] available online at the end of this article, provides more details).

CONCLUSION

The investigators of the best studies of diagnostic test accuracy will enroll a population in whom there is genuine uncertainty about the diagnosis, and these investigators will undertake a blinded comparison between the test and a reference standard to patients. The critical appraisal of diagnostic test studies focuses on aspects of risk of bias, results, and applicability. Clinicians should apply these guidelines using their best judgment. ■

BOX 6

What you say to your colleague.

You explain to your colleague that the performance of the laser fluorescence device for detecting noncavitated carious lesions seems to be good, but that the numbers reported by the authors of the article you found[8] probably overestimate its performance. Also, you explain to your colleague the applicability issues that you could face, which make you think that you should be cautious and continue to investigate the device before making a final decision regarding acquiring it, especially considering that the diagnostic methods you usually use in your practice seem to perform better.

SUPPLEMENTAL DATA

Supplemental data related to this article can be found at http://dx.doi.org/10.1016/j.adaj.2015.01.011.

1. Brignardello-Petersen R, Carrasco-Labra A, Sash P, Azarpazhooh A. A practitioner's guide to developing critical appraisal skills: what is the difference between clinical and statistical significance? *JADA.* 2013;144(7):780-786.

2. Brignardello-Petersen R, Carrasco-Labra A, Booth HA, et al. A practical approach to evidence-based dentistry: how to search for evidence to inform clinical decisions. *JADA.* 2014;145(12):1262-1267.

3. Brignardello-Petersen R, Carrasco-Labra A, Glick M, Guyatt GH, Azarpazhooh A. A practical approach to evidence-based dentistry: III: how to appraise and use an article about therapy. *JADA.* 2015;146(1):42-49.e1.

4. Brignardello-Petersen R, Carrasco-Labra A, Glick M, Guyatt GH, Azarpazhooh A. A practical approach to evidence-based dentistry: IV: how to use an article about harm. *JADA.* 2015;146(2):94-101.e1.

5. Fletcher R, Fletcher S. Diagnosis. In: Fletcher R, Fletcher S, eds. *Clinical Epidemiology: The Essentials.* 4th ed. Baltimore, MD: Lippincott Williams & Wilkins; 2005:35-58.

6. Guyatt GH, Meade MO, Richardson S, Jaeschke R. What is the question? In: Guyatt GH, Rennie D, Meade MO, Cook DJ, eds. *Users' Guide to the Medical Literature: A Manual for Evidence-Based Clinical Practice.* 2nd ed. New York, NY: McGraw-Hill; 2008:17-28.

7. Villa-Chavez CE, Patino-Marin N, Loyola-Rodriguez JP, et al. Predictive values of thermal and electrical dental pulp tests: a clinical study. *J Endod.* 2013;39(8):965-969.

8. Costa AM, Paula LM, Bezerra AC. Use of Diagnodent for diagnosis of non-cavitated occlusal dentin caries. *J Appl Oral Sci.* 2008;16(1):18-23.

9. Guyatt GH, Straus S, Meade MO, et al. Therapy. In: Guyatt GH, Rennie D, Meade MO, Cook DJ, eds. *Users' Guide to the Medical Literature: A Manual for Evidence-Based Clinical Practice.* 2nd ed. New York, NY: McGraw-Hill; 2008:67-86.

10. Fletcher R, Fletcher S. Introduction. In: Fletcher R, Fletcher S, eds. *Clinical Epidemiology: The Essentials.* 4th ed. Baltimore, MD: Lippincott Williams & Wilkins; 2005:1-16.

11. Furukawa TA, Strauss S, Bucher HC, Guyatt GH. Diagnostic tests. In: Guyatt GH, Rennie D, Meade MO, Cook D, eds. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice.* 2nd ed. New York, NY: McGraw Hill; 2008:419-438.

12. Tantanapornkul W, Okouchi K, Fujiwara Y, et al. A comparative study of cone-beam computed tomography and conventional panoramic radiography in assessing the topographic relationship between the mandibular canal and impacted third molars. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 2007;103(2):253-259.

13. Jespersen JJ, Hellstein J, Williamson A, Johnson WT, Qian F. Evaluation of dental pulp sensibility tests in a clinical setting. *J Endod.* 2014;40(3):351-354.

14. Sinanoglu A, Ozturk E, Ozel E. Diagnosis of occlusal caries using laser fluorescence versus conventional methods in permanent posterior teeth: a clinical study. *Photomed Laser Surg.* 2014;32(3):130-137.

15. Grimes DA, Schulz KF. Uses and abuses of screening tests. *Lancet.* 2002;359(9309):881-884.

16. Fields MM, Chevlen E. Screening for disease: making evidence-based choices. *Clin J Oncol Nurs.* 2006;10(1):73-76.

17. Fagan TJ. Letter: nomogram for Bayes theorem. *N Engl J Med.* 1975;293(5):257.

18. Jablonski-Momeni A, Heinzel-Gutenbrunner M, Klein SM. In vivo performance of the VistaProof fluorescence-based camera for detection of occlusal lesions. *Clin Oral Investig.* 2014;18(7):1757-1762.

eTABLE

Critical appraisal of an article about diagnosis*

GUIDE	EXPLANATION
Are the Results Valid?	
Did participating patients present a diagnostic dilemma?	Some of them. Teeth included in the study could be at any stage of the disease, ranging from healthy to decayed appearance. There was a subgroup of healthy teeth that were excluded from the sample after the teeth had been enrolled. In addition, when looking more closely at the proportion of teeth from each category included, the results showed that only teeth with signs of being decayed were included for the assessment of the diagnostic test properties. This indicates that the accuracy of the diagnostic test probably was overestimated.
Did investigators compare the results of the test with an appropriate, independent reference standard?	Yes, the reference standard was fissure eradication or enameloplasty, a method that most clinicians would agree is the best available to diagnose noncavitated carious lesions.
Were the investigators interpreting the test and reference standard blinded to the other results?	Probably not. It is not clear from reading the description of the methods whether the clinicians who applied the different diagnostic tests were the same clinicians who applied all the tests. If the clinicians were the same, then they were not blinded to the results of the other tests, as the clinicians may have recognized the patients they were examining.
Did investigators perform the same reference standard with all patients regardless of the results of the test under investigation?	No, owing to ethical reasons, only patients classified as sick underwent confirmation with the reference standard. This could have resulted in an overestimation of the accuracy of the diagnostic test.
What Are the Results?	
What LRs† were associated with the range of possible test results?	The positive LR (that is, LR+) value was 3.68. This means that a positive test result would lead to moderate to low shifts in pretest to posttest probabilities of having noncavitated occlusal carious lesions. The negative LR (that is, LR-) value was not reported by the investigators of this article, but it can be calculated using the sensitivity and specificity values. Because the sensitivity is 0.93 and the specificity is 0.75, the LR- value is 0.09. This means that, when the test is negative, the change from pretest to posttest probability of being healthy would be important, and the disease could be ruled out with more confidence.
How Can I Apply the Results to My Patients' Care?	
Will the reproducibility of the test result and its interpretation be satisfactory in my setting?	Interrater and intrarater reproducibility values ranged from 0.730 to 0.747, which represents substantial agreement according to the criteria followed by the authors. The laser device assessed in this study seems to be easy to use and its results easy to interpret; therefore, it is likely that the study results could be applied to many settings. However, the clinician should assess whether there would be any limitation in his or her practice.
Are the results applicable to the patients in my practice?	The study investigators included permanent molars and premolars with and without apparent carious lesions from 26 patients aged 10 to 13 years. Clinicians should consider whether their patients might have any different features that could alter the performance of the test.
Will the results change my management strategy?	Probably not. The LR shows that the diagnostic test alone would not be enough to confirm or rule out the diagnosis of noncavitated carious lesions.
Will patients be better off as a result of this test?	Probably not. Even though the consequence of not diagnosing a noncavitated carious lesion could be severe, potentially leading to pulp necrosis, the test results were not accurate enough. Conventional methods performed better than the diagnostic test. Therefore, patients are unlikely to benefit from the additional use of this diagnostic test.

* Source: Costa and colleagues.[8]
† LR: Likelihood ratio.