

## How to appraise a diagnostic test

Ramanitharan Manikandan, Lalgudi N. Dorairajan

Department of Urology, Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, India

### ABSTRACT

Urologists frequently encounter problems in making a clinical diagnosis whose resolution requires the use of diagnostic tests. With an ever increasing choice of investigations being available, the urologist often has to decide which diagnostic test(s) will best resolve the patient's diagnostic problem. In this article, we aim to help the urologist understand how to critically appraise studies on diagnostic tests and make a rational choice. This article presents the guiding principles in scientifically assessing studies on diagnostic tests by proposing a clinical scenario. The authors describe a standardized protocol to assess the validity of the test and its relevance to the clinical problem that can help the urologist in decision making. The three important issues to be considered when evaluating the validity of the study are to identify how the study population was chosen, how the test was performed and whether there is a comparison to the gold standard test so as to confirm or refute the diagnosis. Then, the urologist would need to know the probability of the test in providing the correct diagnosis in an individual patient in order to decide about its utility in solving the diagnostic dilemma. By performing the steps described in this article, the urologist would be able to critically appraise diagnostic studies and draw meaningful conclusions about the investigations in terms of validity, results and its applicability to the patient's problem. This would provide a scientific basis for using diagnostic tests for improving patient care.

**Key words:** Critical appraisal, diagnostic study, evidence-based medicine, likelihood ratio, user guide

### INTRODUCTION

All urologists frequently encounter difficulties in making the correct diagnosis of a patient's clinical problem that prompts them to order diagnostic tests. While many clinical problems are evaluated using standard guidelines and diagnostic algorithms, occasionally the urologist is confronted with an uncommon clinical problem and then wonders what would be the most appropriate test to resolve the diagnosis effectively and efficiently. The present day clinical practice is also dominated by an ever increasing array of diagnostic tests owing to the rapid emergence of sophisticated newer technologies. With the increasing availability of these tests, there is a pressing need for the urologist to

understand the strengths and weaknesses of these new tests to deliver an efficient patient care. Use of an inappropriate test or a lack of understanding of what the results actually mean in the clinical context results in a waste of time and money and may at times lead to an inappropriate line of management. In this article, we aim to help the urologist understand how to critically appraise a diagnostic test, how to interpret what the test actually means in the clinical context and the extrapolation of those results to the current diagnostic dilemma. This would enable the urologist to select an appropriate diagnostic test and assimilate the results to resolve diagnostic dilemmas and thereby provide an evidence-based patient-centered clinical care. We describe a clinical scenario and then in a stepwise manner demonstrate how to critically appraise a study on a diagnostic test and assess its ability to resolve the diagnostic dilemma so that a rational decision can be taken on the suitability or otherwise of ordering the test in the clinical context described.

### CLINICAL SCENARIO

In your urology outpatient department, you see a 50-year-old male patient with a history of painless hematuria. As you probe the history further, you come to know that this patient is a smoker, has a history of stable coronary heart disease and is on medications that include anticoagulants. With a preliminary diagnosis of carcinoma bladder, you evaluate the patient further. The ultrasonogram of the urinary tract

For correspondence: Dr. Lalgudi N. Dorairajan,  
Department of Urology, JIPMER, Puducherry - 605 006, India.  
E-mail: dorairajan\_in@hotmail.com

Access this article online	
Quick Response Code:	Website: www.indianjurol.com
	DOI: 10.4103/0970-1591.91444

is normal and the urinary cytology is negative for malignant cells. You explain to the patient about the need for diagnostic cystoscopy. The patient is concerned about undergoing cystoscopy in view of his cardiac disease and requests you to perform a non-invasive diagnostic test to confirm or exclude the diagnosis. You have already heard about virtual cystoscopy in the diagnosis of bladder cancer. So, you decide to perform a literature search to find out a study about the efficacy of virtual cystoscopy in detecting bladder cancer.

## SEARCH METHODOLOGY

You formulate a focused clinical question based on the directions provided by the article “How to perform a literature search in the urological literature series” by Krupski *et al.*<sup>[1]</sup> to guide in search for the articles. “In patients at risk of bladder cancer how does virtual cystoscopy (intervention) compare with conventional cystoscopy (comparison) in confirming or ruling out the diagnosis of bladder cancer (outcome)?” This query would be best answered by a prospective study.

The best evidence that you can get is from a systematic review of several high-quality studies on this diagnostic test. You search the terms “virtual cystoscopy” and “conventional cystoscopy and bladder cancer” separately in PubMed®. This yields 82 studies and 175 studies, respectively. When the combined search of the above terms is applied restricting to English literature, we get 26 studies. You retrieve the abstracts of those studies. Out of this, you read about a study “Comparison of CT virtual cystoscopy of the contrast material-filled bladder with conventional cystoscopy in the diagnosis of bladder tumors” published in *Clinical Radiology*® in 2009.<sup>[2]</sup> The abstract reveals that patients with painless gross hematuria were evaluated initially with virtual cystoscopy using computerized tomography and then by conventional cystoscopy. The investigators conclude in a very positive note regarding the efficacy of this diagnostic test in successfully detecting bladder tumors >5 mm. This study is interesting in the sense that it is a recent study, the study topic seems to address the present clinical query and it is published in a reputed peer reviewed journal. You then download the full text of the above article.

## HOW TO CRITICALLY EVALUATE A DIAGNOSTIC STUDY

The aim of the diagnostic test is to determine whether the given test is able to accurately identify patients with and without the disease as defined by a gold standard test. There are several important issues to be considered in evaluating whether a study is likely to provide a reliable estimate of the diagnostic parameters. The “Users’ Guide to the Medical Literature” recommends evaluating the test by applying three important queries: i) Are the results of the study valid? ii) If so, what are the results of the study? and iii) Will the results help in the care of my patient?<sup>[3,4]</sup>

### *Are the results of the study valid?*

The methodology of the study is one of the most important determinants in attaching significance to its results. Standards for reporting of diagnostic accuracy (STARD) criteria provide a comprehensive guide for transparent reporting of diagnostic accuracy studies.<sup>[5,6]</sup> One has to consider how the investigators chose the subjects for the study and whether the study population would be representative of the population in whom the diagnostic test is going to be applied in routine practice. Furthermore, one has to understand how the authors applied the particular test and whether it was compared to the “reference or the gold standard”.

### *What are the results of the study?*

After establishing that the test results are valid, the next step is to determine the diagnostic test’s accuracy. This is done by calculating the likelihood ratios (LRs).

### *Will the results help me in caring for my patient?*

The next step is to analyze how to apply the test to the particular patient and to similar patients that you encounter frequently in clinical practice. Will the test provide us with the critical information necessary for decision making that was not possible with history, clinical examination and other data already available? Is the test safe, cost-effective and widely available? Is the patient really benefited by this test?

We now evaluate the study stepwise using the guidelines as outlined by Jaeschke *et al.* [Table 1].<sup>[5]</sup>

## ARE THE RESULTS OF THE STUDY VALID?

### *Primary guides*

*Was there an independent, blind comparison with a reference standard?*

The accuracy of the diagnostic tests is determined by comparison with a gold standard. An ideal reference standard should be able to differentiate the patients with and without the disease. At the same time, the test should be simple and safe to be performed on the subjects. The user should always verify that the patients have undergone an appropriate gold standard test along with the test under investigation.<sup>[7]</sup> The user must be aware that no test is absolutely perfect.

In the article of interest by Kivrak *et al.*, the authors have compared the efficacy of CT virtual cystoscopy (investigation under trial) with conventional cystoscopy (reference standard) for detection of bladder cancer in patients presenting with painless hematuria. We are all aware that conventional cystoscopy is the gold standard investigation in evaluation of patients with painless hematuria.

Once you are in agreement with the reference standard, the next question is whether the person measuring or interpreting the diagnostic test under evaluation is blinded to the results

**Table 1: Guidelines for evaluating an article about a diagnostic test<sup>[15]</sup>****Are the results of the study valid?**

## Primary guides

- Was there an independent, blind comparison with a reference standard?
- Did the patient sample include an appropriate spectrum of patients to whom the diagnostic test will be applied in clinical practice?

## Secondary guides

- Did the results of the test being evaluated influence the decision to perform the reference standard?
- Were the methods for performing the test described in sufficient detail to permit replication?

**What were the results?**

- Are LRs for the test results presented or data necessary for their calculation provided?

**Will the results help me in caring for my patients?**

- Will the reproducibility of the test results and their 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?

Reproduced with permission from The American Medical Association; copyright© (1994) American Medical Association; all rights reserved

of the reference standard. Similarly, the person interpreting the reference standard should be blinded to the results of the diagnostic test under evaluation. Lack of blinding will lead to overestimation of sensitivity and specificity.<sup>[8]</sup> Kivrak *et al.* have explicitly reported about the blinding in their protocol, that is, the two experienced radiologists, who independently interpreted the virtual cystoscopic images, were blinded to the findings of conventional cystoscopy. Furthermore, the diagnostic test was performed prior to the reference standard, which makes it reasonable to assume that investigators were in fact blinded.

*Did the patient sample include an appropriate spectrum of patients to whom the diagnostic test will be applied in clinical practice?*

The study population should be representative of the population that is scheduled to receive the diagnostic test in clinical practice. In the scenario of bladder cancer, the test should be able to differentiate patients with and without carcinoma of the bladder. The test should be able to detect bladder lesions in high-risk patients such as older patients and smokers with painless hematuria, as well as in low-risk populations. This is critically relevant since diagnostic tests perform differently depending on the disease severity in a given patient population, which is termed as spectrum bias.<sup>[9]</sup> Kivrak *et al.* have selected the patients with painless hematuria as the study population. Since our focused clinical query involves a particular patient with similar symptomatology, this study may provide relevant information.

**Secondary guides***Did the results of the test being evaluated influence the decision to perform the reference standard?*

The properties of the diagnostic tests will be altered if their outcome influences whether patients undergo confirmation by the reference standard. This condition is known as “verification bias” or “work-up bias”.<sup>[9,10]</sup> Use of flow diagrams are encouraged in the reporting of diagnostic accuracy as they clearly explain to the reader whether there is presence or absence of verification bias.<sup>[11,12]</sup> In the virtual cystoscopy study, all the patients enrolled in the study underwent conventional cystoscopy (reference test) regardless of the outcomes of the CT virtual cystoscopy test. We are convinced that the authors decided to perform cystoscopy as the reference test eliminating verification bias.

*Were the methods for performing the test described in sufficient detail to permit replication?*

If the investigators suggest that the particular diagnostic test is useful and recommend its usage, then they must describe in detail all the issues regarding the methodology for performing the test. This should include the preparation of the patient, the drugs/chemicals to be used, the technical specification of the imaging modality if any and the analysis of the data and interpretation of its results. This is because differences in the diagnostic test protocol have been shown to produce differences in the diagnostic capability for certain tests.<sup>[13-15]</sup> Hence, the STARD criteria recommend the inclusion of the technical details of the performance of the tests.<sup>[11,12]</sup> Kivrak *et al.* have described in detail the protocol for CT cystoscopy in the “Methods” section which can be reproduced by other investigators.

By performing the above tasks for the validity assessment for CT virtual cystoscopy, we conclude that the study by Kivrak *et al.* satisfies the proposed primary and secondary criteria. Though the number of subjects included in the study is small, the diagnostic test requires further review and hence we move on to assess the results of the study.

**WHAT ARE THE RESULTS?**

The urologist ordering for a particular diagnostic test concerning the patient’s constellation of symptoms has an intuition regarding the probability of the disease in that patient based on clinical experience or epidemiological studies, which is referred to as the pretest probability. Consider the following scenario: The chance of a 60-year-old male with hematuria working in a leather industry to harbor bladder cancer is much higher than a 30-year-old female working as a school teacher, with the same history. This is to say that the history and physical examination remains the fundamental basis in predicting the disease probability. For some diseases like acute prostatitis and varicocele, clinical findings yield a highly accurate diagnosis. The pretest probability can be obtained from the epidemiological data,

clinical acumen, consensus view of colleagues and disease registry. The properties of diagnostic tests are usually described in terms of sensitivity and specificity or predictive values [Table 2]. LRs are alternative statistics for reporting the results of the diagnostic studies due to its powerful properties.<sup>[16]</sup> An LR is the percentage of ill people with a given test result divided by the percentage of well individuals with the same result.<sup>[17]</sup>

### Why use likelihood ratios?

Most of us are familiar using sensitivity and specificity. Then, how is LR going to help us? Is it necessary to learn to use them? The advantages of LR are the following:

- i) A number of clinical tests have continuous results. LRs help the user in better interpretation of those results<sup>[18-20]</sup>
- ii) LRs are useful over a range of disease frequencies, whereas predictive values depend on the prevalence of the disease.<sup>[7]</sup>
- iii) Dependence on sensitivity and specificity may lead to overestimation of the benefits of the tests.<sup>[21]</sup>
- iv) LR improves the clinical judgment. When tests are done in a sequence, the LR increases synergistically as the post-test odds of the first test become the pretest odds for the second.<sup>[7]</sup>

LR is calculated as shown in Table 3.

**Table 2: Terms employed in reporting diagnostic test results and their meanings**

Diagnostic terms	Meaning of the term
Case positive	An individual with the disease of interest, that is, the reference standard is positive
Case negative	An individual without the disease of interest, that is, the reference standard is negative
Test positive	An individual with a positive result for the diagnostic test under investigation
Test negative	An individual with a negative result for the diagnostic test under investigation
Prevalence	The proportion of the population with the condition of interest
True positives	Patients correctly identified by the diagnostic test as having the disease
True negatives	Patients correctly identified by the diagnostic test as not having the disease
False positives	Patients without the disease who are incorrectly labeled by the diagnostic test as having the disease
False negatives	Patients with the disease who are incorrectly labeled by the diagnostic test as not having the disease
Sensitivity	The proportion of patients with the disease who are correctly identified by the test
Specificity	The proportion of patients without the disease who are correctly identified by the test
Positive predictive value	The proportion of patients with a positive test who genuinely have the disease
Negative predictive value	The proportion of patients with a negative test who genuinely do not have the disease

### How to use likelihood ratios?

LR+ is defined as the likelihood of a positive test in individuals with the disease compared to the likelihood of a positive test in those without the disease. LR+ is calculated as sensitivity/(1 – specificity). We shall consider the example of the results of the CT virtual cystoscopy study. Kivrak *et al.* have tabulated the results of the study in Table 2 of their manuscript. The detection rates of both the polypoid and sessile lesions of bladder cancer have been calculated. Out of the 78 lesions detected by conventional cystoscopy, CT virtual cystoscopy was able to detect 71 (91%) lesions with a sensitivity of 90% and specificity of 94%. The LR+ of detecting a bladder lesion is  $90/(100 - 94) = 15$ . This means that a positive CT virtual cystoscopy is 15 times more likely to occur in patients with than without bladder cancer. Similarly LR- is stated as the likelihood of a negative test result in patients with the condition compared to the likelihood of a negative test in patients without the condition. LR- is calculated as  $(1 - \text{sensitivity})/\text{specificity}$ . In the above study, LR- of detecting a bladder lesion is  $(100 - 90)/94 = 0.10$ . LRs have not been reported by Kivrak *et al.* in their study, but they can be calculated from a 2 × 2 table or from sensitivity and specificity as described above.

### What do these numbers signify?

LRs indicate by what magnitude a given diagnostic test will increase or decrease the pretest probability of the disease.<sup>[15]</sup> If the diagnostic test has an LR of 1, the post-test probability of the disease is the same as the pretest probability. In other words, the result does not discriminate illness and health. Hence, the test has caused inconvenience and cost to the patient rather than being beneficial. LR greater than 1 increases the probability that the disease is present, and when the LR is less than 1 the probability of the presence of disease decreases. How does the absolute value of the LRs influence clinical practice? Practically, they can be used roughly as follows:

- i) LR greater than 10 or less than 0.1 produces greater and often conclusive changes from pretest to post-test probability.
- ii) LR of 5-10 and 0.1-0.2 causes moderate shifts in pretest to post-test probability.
- iii) LR of 2-5 and 0.5-0.2 causes small (but sometimes significant) changes in post-test probability.
- iv) LR of 1-2 and 0.5-1 changes post-test probability to an insignificant degree.<sup>[15]</sup>

**Table 3: Method of calculation of likelihood ratios**

	Case positive	Case negative
Test positive	a	b
Test negative	c	d

Sensitivity =  $a/(a + c)$ ; Specificity =  $d/(b + d)$ ; Positive predictive value =  $a/(a + b)$ ; Negative predictive value =  $d/(c + d)$ ; Likelihood ratio for positive test result =  $[a/(a + c)]/[b/(b + d)]$ ; Likelihood ratio for negative test result =  $[c/(a + c)]/[d/(b + d)]$



Having calculated the LRs, how do we use them to connect the pretest probability to the post-test probability? One way of doing this task is by converting the pretest probability to odds, multiplying the result by the LR, and converting the consequent post-test odds into a post-test probability. This is a difficult exercise and requires the following equation to convert probabilities into odds:  $\text{odds} = \text{probability}/(1 - \text{probability})$ . With the pretest odds, the post-test odds can be calculated by multiplying the pretest odds by LR. The post-test odds can be converted back into probabilities using the formula:  $\text{odds}/(\text{odds} + 1)$ .<sup>[22]</sup> To overcome this tedious exercise, one can use the Fagan nomogram.<sup>[23]</sup> The nomogram has three columns [Figure 1]. The first column shows the pretest probability, the second column the LR and the third shows the post-test probability. A line is drawn from the pretest probability linking the LR of the result, and the point where it strikes the third column denotes the post-test probability. An interactive version of the Fagan nomogram is available at the Centre for Evidence- Based Medicine (CEBM) website (<http://www.cebm.net>).

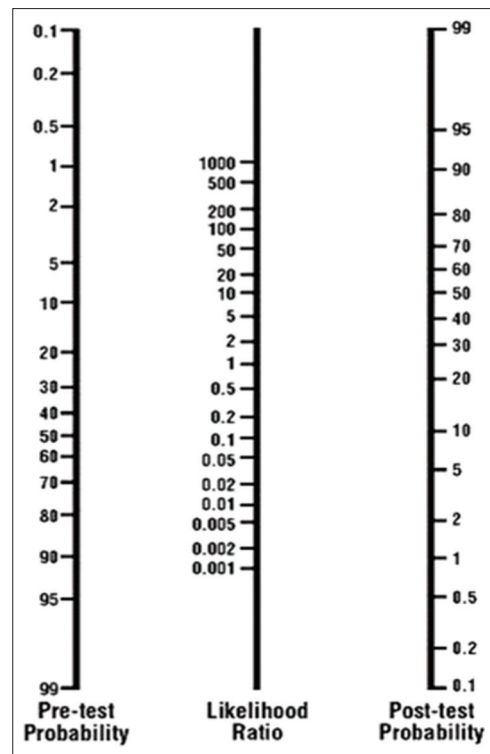
In the present clinical scenario of a 50-year-old male with a history of smoking, we can assume a pretest probability of 40% of harboring a bladder cancer based upon recently published epidemiological studies.<sup>[24,25]</sup> Using a pretest probability of 40% and a positive CT virtual cystoscopy with an LR+ of 15, the Fagan nomogram yields a post-test probability of approximately 90%. Next, we will analyze whether a negative CT virtual cystoscopy report would result in lowering the probability of bladder cancer. Similarly, with a pretest probability of 40% and an LR- of 0.1, we arrive at a post-test probability of approximately 7%, thereby implying that there is a 7% chance of missing a bladder tumor despite a negative CT virtual cystoscopy in our patient.

Kivrak *et al.* in their study have evaluated the sensitivity and specificity in their results but have not provided the same for conventional cystoscopy (reference standard). But these diagnostic parameters are not of great help for the urologists as they are keen to know to what extent a positive or negative test changes the patient's probability of the target disorder. There are other measures such as a positive predictive value (PPV) and a negative predictive value (NPV) of 87 and 93%, respectively, for virtual cystoscopy brought out in this study. Since PPV and NPV vary with the prevalence of the disease, they are an unreliable estimation of the test performance.<sup>[26]</sup>

## WILL THE RESULTS HELP ME IN CARING FOR MY PATIENT?

### *Will the reproducibility of the test result and its interpretation be satisfactory in my setting?*

The value of a diagnostic test depends on its ability to demonstrate the same result when reapplied to stable



**Figure 1:** The Fagan nomogram<sup>[23]</sup> [reproduced with permission from The Massachusetts Medical Society; copyright© (1975) Massachusetts Medical Society; all rights reserved]

patients. Lack of reproducibility may be due to the nature of the test itself. Considering our example of CT virtual cystoscopy, the interpretation of the CT images requires the services of an experienced radiologist. In the study of Kivrak *et al.*, two radiologists were involved in reporting the results. But the authors have not explicitly revealed whether there was any discordance between the two radiologists in interpreting any of the images and if so, whether there was any bias involved. With experience, the radiologists will be able to interpret the images consistently, and hence, this test is likely to be reproducible in your setting.

### *Are the results applicable to my patient?*

It is important to consider whether the study population is comparable to the patients that you encounter in your clinical practice. Then only, the results of the diagnostic tests reported in the article can be applied to the particular patient. In this study, the authors have chosen a study population which has a high risk of bladder cancer. This resulted in increased sensitivity in bladder cancer detection. Hence, the results of this study cannot be applied in screening of patients presenting with painless hematuria, which has been accepted by the authors themselves. However, the results can be applied to the present clinical scenario as our patient also falls in the high risk group.

### *Will the results change my management?*

After seeing whether the results of the study are applicable

to the present clinical scenario, we have to decide whether the results can change the line of management. The test parameters may vary in a population having a composition of varying disease severity. The urologist should make up his mind of what degree of probability one is willing to accept to confirm or rule out the disease in order to decide on the further course of management, be it follow up alone, further testing or the start of treatment.<sup>[22]</sup> The probability thresholds that determine initiation of treatment or further testing varies among patients and disease states and are usually a matter of clinical judgment and patient's perception and values.<sup>[22]</sup> The factors that influence these judgments include the danger of missing a disease and the consequences of its delayed treatment, the potential risks and the cost of the diagnostic test or therapy (in this case, the risks associated with cystoscopy) and the overall medical condition of an individual patient (the presence of a stable cardiovascular disease in the present clinical scenario). A useful new diagnostic test should, ideally, increase the probability of presence of disease in that patient beyond the threshold for initiating a treatment or eliminating the need for another diagnostic test (in this case, cystoscopy) or lower it to such a level that we are comfortable in assuming that the patient does not have the target condition or at least that the risk of missing the target condition (in this case, bladder cancer) is less than the risk of treatment or another diagnostic test.

In the present scenario, will a positive result on CT virtual cystoscopy change the management? The pretest probability of the patient having a bladder tumor is sufficiently high to justify cystoscopy and further management of the patient. In this study of CT virtual cystoscopy, LR+ is so high that it will increase the pretest probability of presence of bladder tumor from 40% to a post-test probability of 90%. Thus, a positive result from CT virtual cystoscopy, while dramatically increasing the pretest probability of the disease due to the high LR+, will only further justify rather than eliminate the need to perform cystoscopy.

Will a negative result on CT virtual cystoscopy eliminate the need for cystoscopy? Considering that even nonmuscle invasive bladder cancers can progress to invasive or metastatic bladder cancers in 10-20% of cases<sup>[27]</sup> and considering the low procedure-related mortality rate of conventional cystoscopy even when performed under general anesthesia, we may assume that any diagnostic test that misses more than 2-3% of bladder tumors would not eliminate the need for cystoscopy. We see that although the LR- value of CT virtual cystoscopy is so low that it changes the pretest probability from 40% to a post-test probability of 7%, this value is still quite high and above our threshold and, therefore, does not eliminate the need for cystoscopy.

### ***Will the patient be better off as a result of the test?***

The usefulness of the diagnostic test is accepted only if it adds such additional information beyond what is available at that

point in time that is going to change the management plan and thereby prove beneficial for the patient. As compared to a more accurate test, a less accurate diagnostic test may still be of great value if the risks involved in performing the tests are substantially less and missing the target disease that results in a delayed diagnosis and treatment is not particularly detrimental to the overall well-being of the patient. However, the benefits of an accurate diagnostic test will not be questioned, even if it is invasive, if the risk of the target disorder going undiagnosed, and therefore missing a potential opportunity for cure, is higher than the risk of undergoing the test as in the case of conventional cystoscopy in the present scenario.

### ***Resolution of the clinical scenario***

After critically appraising the original research article on the efficacy of CT virtual cystoscopy in detecting bladder lesions, you need to decide whether you can apply the diagnostic test in your patient. You conclude that the study was well conducted with acceptable results. But you decide that even when the test results are negative, the probability of missing a tumor in your patient is significantly high, and as the consequences of missing the diagnosis are significant, cystoscopy would still be warranted. As the diagnostic test (CT virtual cystoscopy) is not going to change the management plan, irrespective of whether the result is positive or negative, in the given case you conclude that it is not beneficial to the patient. You explain your conclusions to the patient who agrees to undergo cystoscopy.

You also understand from your appraisal that the decision to adopt a new diagnostic test should ideally be based not on one but on several well-designed studies on such patient populations as are encountered in your practice. You realize that the usefulness of a test will vary depending not only on its accuracy but also on the nature of the disease it seeks to identify and its prognosis, the risks associated with its management, the ease of availability of the test and its costs.

## **CONCLUSIONS**

In this article, the authors have described a systematic approach in critically analyzing the results of the diagnostic tests. By addressing the key issues of validity, interpretation of results, applicability and diagnostic parameters, one has to decide whether the diagnostic test will help in better patient care. This is an important aspect of application of evidence-based medicine that every urologist must learn

## **REFERENCES**

1. Krupski TL, Dahm P, Fesperman SF, Schardt CM. User's guide to the urological literature: How to perform a literature search. *J Urol* 2008;179:1264-70.
2. Kivrak AS, Kiresi D, Emlik D, Odev K, Kilinc M. Comparison of CT virtual cystoscopy of the contrast material-filled bladder with conventional cystoscopy in the diagnosis of bladder tumours. *Clin Radiol* 2009;64:30-7.

3. Guyatt G, Rennie D. Users' Guide to the Medical Literature. 4<sup>th</sup> ed. Chicago: AMA Press; 2002.
4. Bajammal S, Bhandari M, Dahm P. What every urologist should know about surgical trials Part I: Are the results valid? *Indian J Urol* 2008;24:281-8.
5. Jaeschke R, Guyatt GH, Sackett DL. 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? The Evidence-Based Medicine Working Group. *JAMA* 1994;271:703-7.
6. Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. *CMAJ* 1986;134:587-91.
7. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. 2<sup>nd</sup> ed. Boston, Mass: Little Brown and Co; 1991. p. 53-7.
8. Goodacre S. Critical appraisal for emergency medicine: 5. Evaluation of a diagnostic test. *Emerg Med J* 2009;26:53-6.
9. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978;299:926-30.
10. Choi BC. Sensitivity and specificity of a single test in the presence of work-up bias. *J Clin Epidemiol* 1992;45:581-6.
11. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, *et al*. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med* 2003;138:40-4.
12. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, *et al*. The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. *Ann Intern Med* 2003;138: W1-12.
13. Detrano R, Gianrossi R, Froelicher V. The diagnostic accuracy of the exercise electrocardiogram: A meta-analysis of 22 years of research. *Prog Cardiovasc Dis* 1989;32:173-206.
14. Philbrick JT, Horwitz RI, Feinstein AR. Methodologic problems of exercise testing for coronary artery disease: Groups, analysis and bias. *Am J Cardiol* 1980;46:807.
15. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1994;271:389-91.
16. Sackett DL, Straus S, Richardson WS, Rosenberg W, Haynes RB. Evidence based medicine. How to practice and teach EBM. 2nd ed. Edinburgh: Churchill Livingstone; 2000. p. 67-93.
17. Grimes DA, Schultz KF. Refining clinical diagnosis with likelihood ratios. *Lancet* 2005;365:1500-5.
18. Choi BC. Slopes of a receiver operating characteristic curve and likelihood ratios for a diagnostic test. *Am J Epidemiol* 1998;148:1127-32.
19. Giard RW, Hermans J. Interpretation of diagnostic cytology with likelihood ratios. *Arch Pathol Lab Med* 1990;114:852-4.
20. Lee WC. Selecting diagnostic tests for ruling out or ruling in disease: The use of the Kullback-Leibler distance. *Int J Epidemiol* 1999;28:521-5.
21. Khan KS, Khan SF, Nwosu CR, Arnott N, Chien PF. Misleading authors' inferences in obstetric diagnostic test literature. *Am J Obstet Gynecol* 1999;181:112-5.
22. Scales CD, Dahm P, Sultan S, Scherer DC, Devereaux PJ. How to use an article about a diagnostic test. *J Urol* 2008;180:469-76.
23. Fagan TJ. Letter to the editor: A nomogram for applying likelihood ratios. *N Engl J Med* 1975;293:257.
24. Baris D, Karagas MR, Verrill C, Johnson A, Andrew AS, Marsit CJ, *et al*. A case-control study of smoking and bladder cancer risk: Emergent patterns over time. *J Natl Cancer Inst* 2009;101:1553-61.
25. Khadra MH, Pickard RS, Charlton M, Powell PH, Neal DE. A prospective analysis of 1,930 patients with hematuria to evaluate current diagnostic practice. *J Urol* 2000;163:524-7.
26. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. *BMJ* 1994;309:102.
27. Messing EM. Urothelial tumors of the bladder. In: Wein AJ, editor. *Campbell-Walsh Urology*, 9<sup>th</sup> ed. Philadelphia: WB Saunders Co; 2007. p. 2407-46.

**How to cite this article:** Manikandan R, Dorairajan LN. How to appraise a diagnostic test. *Indian J Urol* 2011;27:513-9.

**Source of Support:** Nil, **Conflict of Interest:** None declared.