Evidence-based Medicine: Answering Questions of Diagnosis

Laura Zakowski, MD, *Department of Medicine, University of Wisconsin Medical School, Madison, Wisconsin*
Christine Seibert, MD, *University of Wisconsin Medical School, Madison, Wisconsin*
Selma VanEyck, PhD, *Division of Academic Affairs, University of Wisconsin Medical School, Madison, Wisconsin*

REPRINT REQUESTS:

Laura Zakowski, MD
1300 University Avenue
326 Bradley Memorial
Madison, WI 53706-1532
Telephone: 608-263-2903
Email: zakowski@wisc.edu

GRANT SUPPORT

Department of Health and Human Services,
Health Resources and Services Administration,
#1D16HP00067-01

KEYWORDS:

Evidence-based medicine

RECEIVED:

OCTOBER 10, 2003

REVISED AND ACCEPTED:

DECEMBER 22, 2003

Clinical Medicine & Research

Volume 2, Number 1: 63-69

©2004 Clinical Medicine & Research

<http://www.mfldclin.edu/clinmedres>

ABSTRACT

Using medical evidence to effectively guide medical practice is an important skill for all physicians to learn. The purpose of this article is to understand how to ask and evaluate questions of diagnosis, and then apply this knowledge to the new diagnostic test of CT colonography to demonstrate its applicability. Sackett and colleagues¹ have developed a step-wise approach to answering questions of diagnosis:

Step 1: Define a clinical question and its four components: Patient, intervention, comparison and outcome.

Step 2: Find the evidence that will help answer the question. PubMed Clinical Queries is an efficient database to accomplish this step.

Step 3: Assess whether this evidence is valid and important. A quick review of the methods and results section will help to answer these two questions.

Step 4: Apply the evidence to the patient. This step includes: assessing whether the test can be used; determining if it will help the patient; finding whether the study patients are similar to the patient in question; determining a pretest probability; and deciding if the test will change one's management of the patient.

A relatively new diagnostic test, CT colonography, is explored as a scenario in which the steps presented by Sackett et al.¹ can be helpful. A patient who is interested in completing a CT colonography instead of a colonoscopy is the basis of the discussion. Because a CT colonography does not detect polyps of less than 10 mm accurately, many patients are not likely to prefer this test over a colonoscopy.

Evidence-based medicine is an effective strategy for finding, evaluating, and critically appraising diagnostic tests, treatment and application. This skill will help physicians interpret and explain the medical information patients read or hear about.

Evidence-based medicine (EBM) is defined as “the integration of best research evidence with clinical expertise and patient values.”¹ The use of EBM allows physicians to address questions for which answers are not readily available, assess the quality of medical information and attempt to keep pace with the ever expanding research in both basic science and clinical medicine.²

Sackett and colleagues¹ have developed a simple and straightforward approach to answering any researchable question. The four steps are:

- Step 1: Define a clinical question.
- Step 2: Find the evidence that will help answer the question.
- Step 3: Assess whether this evidence is valid and important.
- Step 4: Apply the evidence to the patient.

When one begins to use an EBM approach to find information directly applicable to a patient, the tendency is to cast a wide net to ensure finding an answer to the question. Using these steps helps narrow the focus of research to a specific piece of information, so the search and review is efficient and directly applicable to the patient.

Step 1: Define a clinical question.

In almost every patient encounter, questions arise about some portion of that patient's care. Often, these questions are easily answered through previous experience or review of easily accessed information, such as in a review article or a current textbook. Examples of more easily answered questions may include: “How is type 2 diabetes diagnosed?” or “What are the most common presenting symptoms for lymphoma?” Other questions, usually about a patient's diagnosis, prognosis, treatment or harm, may need further research for current answers. These questions are more appropriately answered through an evaluation of the current literature and application of the steps to EBM. To appropriately use the steps to EBM, one must formulate a question that is best answered by reviewing a research study. Sackett¹ considers these questions to be “foreground questions,” whereas questions best answered by other sources of information are considered “background questions.” Examples of foreground diagnostic questions include: “How effective is the *H. pylori* stool antigen compared to endoscopy in detecting *H. pylori* infection?” or “Is a magnetic resonance venogram, compared to a venogram, an accurate method to detect deep venous thrombosis?”

Just as a sentence has a certain written structure, a researchable question has specific parts. Writing the question with a specific structure makes the search process easier, since terms used to formulate the question can be

used directly for the literature search. A question formulated with this specific structure is termed the “PICO” question¹ and includes the following parts:

- Patient
- Intervention
- Comparison
- Outcome

For diagnostic questions in particular, the “patient” characteristics must distinguish one particular patient from other patients. For example, if one is interested in studies of postmenopausal women with osteoporosis, the “patient” portion of the question should be written to differentiate the particular question from a study evaluating women with osteopenia or premenopausal women.

“Intervention” describes the particular diagnostic test in question, and the “comparison” is the reference standard, if this is available, or another test known to be accurate to diagnose the problem. The reference standard is the definitive test that determines whether or not the patient has the diagnosis. For example, a physical examination of the ankle helps to determine whether or not the patient has a fracture, and the ankle x-ray is the reference standard. The “outcome” is the diagnosis in question.

If one is interested in the accuracy of an increased respiratory rate to detect pneumonia in children presenting to a clinic with respiratory symptoms, an example of a PICO question is:

- P: In children with upper respiratory symptoms
- I: is measuring the respiratory rate
- C: as effective as a chest x-ray
- O: in detecting pneumonia?

Step 2: Find the evidence.

Developing the PICO question is helpful in finding the evidence. Certain key words can be used from the question to directly search the literature. A search engine that allows one to directly enter key words from a PICO question is PubMed Clinical Queries³ available at: <http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.html>. This is a subset of PubMed in which certain search terms are automatically added to the few simple words that are entered for more focused research. More specifically, for diagnostic tests, these terms are added to your search:

AND (sensitivity and specificity [MESH] OR (predictive [WORD] AND value* [WORD])).

For example, to search the PICO question stated above with PubMed Clinical Queries, one can enter the terms: pneumonia and respiratory rate and child, and the above terms are added automatically to the search terms.

Many hospitals with medical student and resident training programs have access to librarians who are most proficient at searches such as this. It can prove to be worth the time and effort to seek help from a librarian if one encounters difficulty finding an appropriate study for the question.

Step 3: Assess the evidence.

There are two parts to this section. First, one must ask whether or not the results of the study in question are likely to be valid, and second, whether or not the results of the study are important. Validity is determined by review of the methods section, and a valid study employs careful research methods that make it likely that the findings are accurate. Validity is determined by four questions, best stated by Sackett.¹

QUESTION 1: IS THERE AN INDEPENDENT BLIND COMPARISON TO A REFERENCE STANDARD?

In other words, for our study of respiratory rate, the clinician checking the patient's respiratory rates should not know the results of the chest x-rays, and radiologists reading the chest x-rays should not know respiratory rates. This is usually described in either the abstract or the methods section.

QUESTION 2: IS THE STUDY PERFORMED IN AN APPROPRIATE SPECTRUM OF PATIENTS?

If a study only tests patients who are at an advanced stage of illness or with symptoms that are easily distinguishable from other illnesses, one is unable to know whether this test can easily distinguish typical patients. In clinical practice, one typically sees patients who present with symptoms that can represent a number of illnesses, or those who are seen at early stages of disease.

QUESTION 3: IS THE REFERENCE STANDARD APPLIED REGARDLESS OF THE STUDY RESULT?

If researchers wish to determine whether the presence of an S3 (a third heart sound) will accurately determine if a patient has congestive heart failure when compared to echocardiogram findings, one would want all patients, regardless of whether they have an S3, to have a echocardiogram. If researchers decided some patients did not need an echocardiogram, the accuracy of the findings would be in question.

QUESTION 4: WAS THE STUDY VALIDATED IN A SECOND, INDEPENDENT GROUP OF PATIENTS?

This question is pertinent only to a few types of studies, where the authors develop new clinical prediction rules such as the Ottawa ankle rules.⁴ This question, therefore, is not typically asked of most diagnostic tests.

In summary, a quick assessment of a study's abstract usually provides answers to most questions of validity. If the abstract states that the diagnostic test is compared to a reference standard in a blinded fashion and all subjects

received both the diagnostic test and the reference standard, one can move on to the next step of assessing the evidence.

The next step asks whether or not the results of a study are important. In other words, can this test distinguish patients with and without the diagnosis? No test is perfect; therefore, one searches for tests with the best record for correctly identifying the diagnosis. The ability of the test to correctly identify the diagnosis is best expressed by the likelihood ratio, a derivation of sensitivity and specificity. A review of the results section of a research study helps determine the answer to this question.

A brief review of sensitivity and specificity will provide some background. The use of a 2x2 table as illustrated in table 1 can help to organize the findings of the study in a standardized manner. Sensitivity is the proportion of patients with the disease who show a positive test result. From the 2x2 table, sensitivity is the result of the calculation:

$$a/a+c.$$

If a test has a high sensitivity, a negative test will rule it out. This can easily be remember by the term "**SnNout**,"¹ highly **S**ensitive test, **N**egative test rules it **Out**. This is easy to see from the 2x2 table. If the test is 99% sensitive, almost all of the patients with the diagnosis must be in box "a", and almost no one is in box "c" (negative test result with diagnosis). Therefore, all patients with a negative test result do *not* have the diagnosis and are in box "d," "**SnNout**."

Specificity is the proportion of patients without a disease who have a negative test result, and it can be calculated from table 1:

$$d/b+d.$$

If a test has a high specificity, a positive test will rule it in. This is easily remembered by the term "**SpPin**,"¹ highly **S**pecific test, **P**ositive test rules it **In**. This is easy to see from the 2x2 table. If the test is 99% specific, almost all of the patients who do not have the diagnosis must be in box "d", and almost no one is in box "b" (positive test result without diagnosis). Therefore, almost all patients with a positive test result *do* have the diagnosis and are in box "a", therefore "**SpPin**."

Before ordering the test, the clinician should consider how likely it is that the patient has the diagnosis? This is called the *pretest probability*. After performing the test, the

Table 1. 2 X 2 Table For Determining Sensitivity And Specificity.

		Diagnosis/Disorder	
		(+)	(-)
Test Result	(+)	a	b
	(-)	c	d

likelihood that the patient has the diagnosis is referred to as *posttest probability*. A perfect test will take any pretest probability (0-100%) and reveal, after the test is completed, that the patient definitively has the diagnosis (a 100% posttest probability). However, this is unrealistic for most tests. A likelihood ratio, which is derived from the sensitivity and specificity, easily converts any pretest probability into a posttest probability. The likelihood ratio is just as the words indicate: a ratio of likelihoods.

$$\text{Likelihood ratio} = \text{Likelihood of test result with disease} / \text{Likelihood of test result without disease}$$

Thus, if your test result is positive, the likelihood ratio is:

$$\frac{\text{Likelihood of a positive test result with disease}}{\text{Likelihood of a positive test result without disease}}$$

This is the same as:

$$\frac{\text{Sensitivity (positive test result with disease)}}{1 - \text{Specificity (positive test result without disease)}}$$

Table 2. Likelihood ratio scale.

The likelihood ratio can be used on the likelihood ratio scale in table 2. Drawing a line from the predetermined pretest probability on the left side of the diagram through the likelihood ratio determines posttest probability. For example, an excellent test with a sensitivity of 99% and a specificity of 98%, yields a likelihood ratio for a positive test at $.99 / 1 - .98 = 49.5$. Using almost any pretest probability on the likelihood ratio scale, one can see how helpful this excellent test is for confirming the diagnosis. For example, if one believed that a patient with unilateral lower extremity edema and no other findings has about a 20% chance of having a DVT (deep venous thrombosis), and a test with a likelihood ratio of 49.5 for a positive result is performed, one will find that this is indeed a very likely diagnosis with a posttest probability of more than 90% when this test is positive.

If the test result is negative, then the ratio of likelihoods has essentially the same definition but with a *negative* test as part of the equation:

$$\frac{\text{Likelihood of a negative test result with disease}}{\text{Likelihood of negative test result without disease}}$$

This is the same as:

$$\frac{1 - \text{Sensitivity (negative test result with disease)}}{\text{Specificity (negative test result without disease)}}$$

As with a positive test result, one can again use a likelihood ratio of a negative test result to change a pretest probability into a posttest probability. If one again uses an excellent test with a sensitivity of 99% and a specificity of 98%, the likelihood ratio for a negative test is $1 - .99 / .98 = 0.01$. Using almost any pretest probability on the likelihood ratio scale, it is easy to see how helpful this excellent test is for ruling out the diagnosis. For example, if the patient with right lower quadrant pain and fever has about a 50% chance of having appendicitis and a test is available where the likelihood ratio is 0.01 for a negative result, one finds that this is indeed an unlikely diagnosis with a posttest probability of less than 0.2% when the result is negative.

In comparison, if one uses a test with poor sensitivity and specificity, one will find that this test is not helpful in determining the diagnosis. If one uses a test with a sensitivity and specificity near 50%, which indicates that test has just about an equal chance of either truly or falsely identifying a diagnosis, then the likelihood ratios with this test will be equally unhelpful. If the calculation for the likelihood ratio of either a positive or negative test is near one, and when using the likelihood ratio scale, one sees that the pretest and the posttest probability are almost the same. Table 3 shows that the further a likelihood ratio is from 1.0,

Table 3. Usefulness of likelihood ratios.

Usefulness	Likelihood ratio (+)	Likelihood ratio (-)
Conclusive	>10	<0.1
Moderately helpful	5-10	0.1-0.2
Possibly helpful	2-5	0.5-0.2
Not helpful	1-2	0.5-1

the more helpful it is in determining the diagnosis. Since this is derived from the sensitivity and specificity, the higher the percentages, the better the likelihood ratios are and the more helpful the test result will be.

In summary, Step 3 moves through the methods section with a few simple questions about validity. Then, the results are computed. If helpful or conclusive likelihood ratios are calculated from the data, the results are deemed important.

Step 4: Apply the evidence to your patient.

This step is divided into 5 parts or questions, and directly relates back to the patient for whom you are using EBM skills. The first two questions pertain to whether or not the test should be used, and whether or not the test will help the patient. For example, if the test under review is unavailable, unaffordable or not covered by the patient's insurance, it is not applicable. It is also important to know if it is the best test for the patient. If there is a better test, or one that is easier for the patient to complete with nearly the same likelihood ratios, it is unlikely that one would use this particular test. In addition, the patient's preferences need to be considered. One must determine, are there are reasons why the patient does or does not want to complete the test?

The third question asks if the study patients similar to the one in question. Studies are often conducted at referral centers or through specialty clinics, and this may be quite different from patients in a given practice. When one reads the methods of the study, it is important to determine whether or not the patient in question would qualify for the study. If the patient does qualify, this may indicate that the population is similar to the patient in question. If the patient does not qualify, one should ascertain whether or not this patient would be excluded from the study. If so, this should lead one to look further for another study.

The fourth question is whether or not a pretest probability can be generated. This is sometimes the most difficult part for students or others with limited clinical experience to answer. There are a number of ways, other than guessing, that may be easily accessible. For instance the answer to the question may be in a textbook or in a review article. For example, if the accuracy of the physical examination is being evaluated to determine whether or not the patient has an aortic aneurysm, one wants to know the probability that he has one. A text can be consulted to determine the percentage of patients who have risk factors for vascular

disease (such as history of hypertension, smoking, etc.) and have an aneurysm; that percentage can be used for a pretest probability. Asking colleagues with more experience or experts in the specialty in question is also helpful. A useful web site from the Mount Sinai School of Medicine provides a few pretest probabilities for common inpatient disorders, and it can be accessed at: <http://www.mssm.edu/medicine/general-medicine/ebm/>.⁵ One can also refer to the study under evaluation. The prevalence of the disease in the study in question is a valid pretest probability if the patient is similar to the research subjects, as discussed in the previous paragraph. Referring back to the question about aortic aneurysms, the prevalence of the disease in the study population is:

Patients in the study with an aneurysm / All patients in the study. Thus, from the 2x2 table (see table 1): $a + c / a + b + c + d$

The final question relates to likelihood ratios. More specifically, one must determine whether or not the results of the study change the management. If the likelihood ratios for either a positive or negative test result (likelihood ratio positive and likelihood ratio negative) are in the conclusive or helpful range as seen in table 3, the diagnostic test will help the clinician decide to either treat the patient or discard the diagnosis. However, as illustrated in table 4, before ordering the test one needs to consider how this will affect the patient's management. One must decide if the test will result in the need for more testing or will it give a definitive answer. The latter is more likely with conclusive or helpful likelihood ratios that will move one from the "test then treat based on the results" range to either "don't test or treat" (in other words, discard the diagnosis), or "don't test, treat" range (where it is believed the patient has the diagnosis).

In conclusion, using the five steps to answer questions of diagnosis ensures that the question is defined and focused enough to aide in finding a study. A review of the methods and results section enables one to use the likelihood ratio scale to determine whether this test is accurate enough to help when findings are applied to the patient.

The following review of the technology of CT colonography, a new form of colon cancer screening and diagnosis will provide an example for EBM steps.

Table 4. Thresholds for testing and treating.

Don't test or treat	Test, then treat based on results	Don't test, TREAT!
0%	?%	?%
PROBABILITY		
		100%

EBM EXAMPLE

A female, 60 years of age, with a history of polyp removal via colonoscopy 3 years ago is due for a repeat examination. She is interested in obtaining a CT scan for repeat evaluation instead of a colonoscopy, since she heard that the CT requires less sedation, and she had some problems with nausea and vomiting after the sedation for her last colonoscopy. Your PICO question for this case is:

- Patient: In patients needing a surveillance colonoscopy
- Intervention: is a CT scan
- Comparison: equivalent to traditional colonoscopy
- Outcome: for detecting colonic polyps?

CT colonography, and a related test, virtual colonoscopy, are new technologies using CT scan instead of the colonoscope for both screening and surveillance of the colon after the diagnosis and treatment of colon cancer.⁶ Virtual colonoscopy utilizes a software program that converts multiple two dimensional images of the colon into a three dimensional path that the radiologist follows through the colon. CT colonography also uses multiple two-dimensional images but does not convert this to three dimensional images. With both studies structures within the abdomen are visualized.

Both methods require fastidious preparation of the colon, since stool can look like polyps, and fluid can obscure polyps. At the beginning of the CT colonography procedure, air is inserted to distend the entire colon, and because of cramping a patient often experiences, a smooth muscle relaxant may be given. For a more complete detection of polyps, both supine and prone images are usually obtained to move any stool or liquid. Intravenous contrast can be given, and the patient is required to hold his or her breath for about 30 seconds.⁶

One of the advantages of CT colonography when compared to traditional colonoscopy includes the utility in patients in whom a traditional colonoscopy cannot be completed. It is also helpful in patients who have a previous history of colon cancer, since other structures such as the liver can be evaluated at the same time. Furthermore, as this case suggests, there is less of a need for sedation compared to traditional colonoscopy.

Other than the need for a perfect preparation, other disadvantages of CT colonography include its decreased accuracy with smaller polyps and flat lesions, its tendency to result in more tests (if a lesion were found on the kidney, for example) and its relative increased cost. A recent cost-effective analysis⁷ indicated that screening by CT colonography costs \$24,586 per life-year saved, compared with \$20,930 for traditional colonoscopy. Even if CT colonography were 100% sensitive and specific, colonoscopy screening remains more cost effective, since positive CT findings will result in referring the patient for colonoscopy, therefore increasing the cost. Thus, CT colonography must be either 54% cheaper, to adjust for the additional testing required of a positive result, or be associated with 15% to 20% improved

compliance rates, to improve the number of patients screened (and life-years saved), for at least equivalent cost-effectiveness compared to traditional colonoscopy.

For the second step, the PICO question search terms for this example are: computed tomography and colonoscopy, and surveillance. When using PubMed Clinical Queries, thirteen articles are listed³, and with a quick assessment of the title or abstract, the articles that are not directly pertinent to our case can be eliminated. For example, studies that fail to include patients with a diagnosis of polyps, or studies that do not compare CT colonography to colonoscopy, or review articles are not pertinent. After eliminating these articles, four possible articles are left for review. One should select a recent article that, upon quick review, gives clear answers to the validity and importance questions. In this example, an article by Miao and colleagues⁶ qualifies.

Usually, it is relatively easy to answer questions of validity, and this article is no exception. The radiologists and the gastroenterologists were unaware of the results of the other study, and all patients received both tests. The spectrum of patients studied included patients who were in need of surveillance colonoscopies for neoplasia and patients presenting with colorectal symptoms. Patients were excluded if they were less than 55 years of age, had a history of inflammatory bowel disease, or were asymptomatic and had a family history of colorectal cancer.

Our evaluation of the results of the study is represented in table 5. Two hundred and one patients were studied. Of the 76 patients with polyps or invasive carcinoma discovered on colonoscopy, CT scan correctly identified 30 patients. Of the 125 patients with no findings on colonoscopy, CT scan was negative in 107 patients. The likelihood ratio for the numbers can easily be calculated. For a positive CT scan, the likelihood ratio positive is 2.74 and for a negative CT scan, the likelihood ratio negative is 0.71.

Finally, this information must be applied to the patient. This test is not available at every institution and it is under experimental use in others, therefore, a review of its availability is necessary to answer the question of whether one can use the test. Although the patient in our case is concerned about sedation, she may be even more concerned that it is not nearly as accurate at detecting polyps as colonoscopy.

A pretest probability can be generated from the study by determining prevalence. The authors state that 36% of the patients had either polyps or invasive carcinoma. A review

Table 5. 2x2 Table for CT diagnosis of colonic polyps.

		Colonic polyps	
		(+)	(-)
CT Scan	(+)	30	18
	(-)	46	107

of a medical textbook⁸ reveals that this matches fairly well with the known percent of patients who have polyps (30 to 40%), especially after the age of 60 years.

Whether or not the findings of the study will change the management of the patient is an important question and is determined from the likelihood ratios. The values for both the likelihood ratio positive and negative are not in the conclusive or helpful range; therefore, CT colonography will not clearly help to determine whether our patient has polyps or not.

In summary, this review of CT colonography illustrates how to ask a question, search the literature, and then assess and apply the evidence. This effective and practical procedure assists the patient in deciding whether or not to proceed with a test that is not as definitive as a colonoscopy.

The authors developed a web site that further describes the steps to EBM, www.fammed.wisc.edu/pds/three/EBM.⁹ The authors of the text referenced in this article¹, have also developed a web site that closely follows the text, www.cebm.utoronto.ca/.¹⁰ EBM is an effective strategy for finding, evaluating, and critically appraising diagnostic tests and treatment. Its concepts help address clinical questions using Internet databases, evaluate the strengths and weaknesses of study designs, and interpret evidence for the patient. It is a powerful and invaluable tool for life-long learning, a basic tenet of medical education.

REFERENCES

1. Sackett D, Straus S, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine. Churchill, Livingstone: 2000, 2nd Edition.
2. Welch HG, Lurie JD. Teaching evidence-based medicine: caveats and challenges. *Acad Med* 2000;75:235-240.
3. PubMed Clinical Queries Web Site. National Library of Medicine. Available at: <http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.html>. Accessed September 27, 2003.
4. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A Study to develop clinical decision rules for the use of radiography in acute ankle injuries. *Ann Emerg Med* 1992;21:384-390.
5. Division of General Medicine Evidence-Based Medicine Web Site. Mount Sinai School of Medicine. Available at: <http://www.mssm.edu/medicine/general-medicine/ebm/>. Accessed October 1, 2003.
6. Miao YM, et al. A prospective single centre study comparing computed tomography pneumocolon against colonoscopy in the detection of colorectal neoplasms. *Gut* 2000;47:832-837.
7. Sonnenberg A, Delco F, Baurefeind P. Is virtual colonoscopy a cost-effective option to screen for colorectal cancer? *Am J Gastroenterol* 1999;94:2268-2274.
8. Powell D. Approach to the patient with gastrointestinal disease. In: Goldman L, Bennett JC, eds. *Cecil Textbook of Medicine*. 21st ed. W.B. Saunders Company; 2000.
9. PDS (Patient, Doctor and Society) 3: Evidence Based Medicine Curriculum. University of Wisconsin Medical School. Available at: www.fammed.wisc.edu/pds/three/EBM. Accessed September 27, 2003.
10. Centre for Evidence-Based Medicine. University Health Network-Mount Sinai Hospital. Available at: <http://www.cebm.utoronto.ca/>. Accessed October 1, 2003.