Accuracy of Leukocyte Count in the Diagnosis of Acute Appendicitis

From the Department of Emergency Medicine, University of California, and San Diego Medical Center, San Diego, CA.

Received for publication October 10, 1997. Revision received November 12, 1998. Accepted for publication December 30, 1998.

Address for reprints: *Brian K Snyder, MD, Department of Emergency Medicine, UCSD Medical Center, 200 West Arbor Drive, Mailcode #8676, San Diego, CA 92103; E-mail bsnyder@ucsd.edu.*

Copyright © 1999 by the American College of Emergency Physicians.

0196-0644/99/$8.00 + 0
47/1/97703

Brian K Snyder, MD
Stephen R Hayden, MD

See related article, p. 575.

[Snyder BK, Hayden SR: Accuracy of leukocyte count in the diagnosis of acute appendicitis. *Ann Emerg Med* May 1999;33:565-574.]

CLINICAL SCENARIO

A 22-year-old man presents to your emergency department with abdominal pain. He is otherwise healthy, with no history of prior abdominal surgeries. He tells you that the pain began in both lower quadrants and that over the past 24 hours the right-sided pain has become more prominent. He describes the pain as generally dull and intermittent with sharp exacerbations not precipitated by movement. He denies radiation of the pain or fever. He became nauseous after his last meal, but he has not vomited. He is currently slightly hungry but is apprehensive about eating. He reports 2 loose bowel movements. He denies dysuria, frequency, flank pain, testicular pain, or swelling.

The patient's temperature is 37°C, and the remainder of the vital signs are normal. Pertinent physical findings include a nondistended abdomen with mild mid- and right-lower quadrant tenderness on deep palpation, no rebound, and no peritoneal signs. Results of the genitourinary examination are normal. Results of the rectal examination reveal no tenderness when pressure is applied to the right vault. Results of a urine dipstick test are negative.

By history and physical examination, you believe that the patient has an intermediate probability of appendicitis that you estimate at approximately 30% to 50%. Although the history is suggestive of appendicitis, the patient's description of his pain has some atypical features. You also believe that the lack of significant tenderness after 24 hours of symptoms decreases the probability of appendicitis. You doubt

that a surgeon would operate at this point but believe the patient requires observation and a surgical consultation. You inform the patient that you are concerned that he may have appendicitis. You discuss the patient's condition with the surgeon on call. After agreeing to come to the ED to evaluate the patient, she asks you to order a CBC, chemistry panel, urinalysis, and abdominal radiographs, so that the results will be available on her arrival. After you discuss this case with the surgeon and order the requested tests, you reflect on the "knee jerk" ordering of WBC counts in patients with abdominal pain. Do the results of a WBC count really help "rule in" or "rule out" acute appendicitis? Because this habit has always gnawed at you, you choose to use this as an opportunity to review the evidence regarding this question. At the end of your shift, you conduct a search on your personal computer and review the evidence. The remainder of this Evidence-Based Emergency Medicine installment will guide you through the process of formulating the question, searching for best evidence, analyzing the evidence, and applying the evidence to future similar patients. You will also be introduced to the concept of *interval likelihood ratios* made possible when multiple cutoff values of test results are provided.

FORMULATING THE QUESTION

Asking a focused and well-articulated question is integral to performing a critical appraisal exercise. The question must be phrased to facilitate conducting an efficient literature search. A good question for a diagnostic issue is composed of 4 elements: the patient care setting, the diagnostic test you are interested in, the comparison test or criterion standard against which the test is to be measured (sometimes called the "gold standard"), and the outcome of interest. In this case, the patient care setting could be defined as ED evaluation of adults with acute right lower quadrant pain. The test in question is the total WBC count as measured by Coulter counter. Other laboratory aides for the diagnosis of appendicitis, such as band count, neutrophil count, and C-reactive protein, could be considered, as well as combinations of these tests. You are specifically interested in the accuracy of the WBC count in predicting presence or absence of appendicitis and therefore choose to avoid the kind of comprehensive search, review, and appraisal that would be entailed were all possible laboratory aides to be included.

Some thought is required to determine the best criterion standard for testing the performance of the WBC count. The strictest gold standard would be the postsurgical pathology report. However, such a rigorous gold standard would restrict the study population to patients who underwent surgery. Many of these patients would likely have been clinically evident and would not have required any diagnostic testing. Furthermore, patients at lower risk for appendicitis, for whom diagnostic testing would be more appropriate and useful, would have been excluded from a study using only a pathology specimen gold standard. Therefore, a gold standard that includes a pathology specimen and a clinical "rule out" of patients not taken to surgery would be more appropriate. Such a clinical "rule out" would require follow-up over an adequate period of time.

The outcome you are interested in is the ability of a WBC count to predict the presence or absence of appendicitis in your patient. You estimate that your patient has a 30% to 50% chance of having acute appendicitis. A test result that would either lower a 50% risk to below 10% or increase a 30% risk to above 80% would directly contribute to your decisionmaking regarding this patient. In the former case, you could contemplate discharging the patient to home; in the latter case, it is likely that the patient would be destined for surgery.

Some physicians would be less happy sending a patient home with up to a 1-in-10 chance of having appendicitis. Establishing a specific "action threshold" can be influenced by objective factors, such as the patient's degree of sophistication and reliability, and by subjective factors, such as a practitioner's ability to clinically distinguish between a 1-in-10, a 1-in-20, or a 1-in-40 risk of disease.

The question may then be reformulated: "Will the peripheral leukocyte count enable this adult ED patient with possible appendicitis to either be safely discharged home with follow-up or taken confidently to surgery, as judged by comparison of WBC values in such patients with pathology reports or prolonged follow-up without return of symptoms or need for surgery?" Long-winded as it may seem, this is the kind of question that both facilitates an effective search strategy and sets the stage for a rational selection of primary studies to be reviewed. (Some evidence-based medicine teachers have quipped, "If you are able to recite the question in 1 breath, it is too short.") Because the question involves the accuracy of a diagnostic test,

the criteria elaborated in a published user's guide for studies on diagnosis are pertinent.[1,2]

SEARCHING FOR AND SELECTING THE BEST EVIDENCE

The MEDLINE database includes all journals included in *Index Medicus* from 1966 to the present and currently comprises more than 4,000 biomedical journals subscribed to by the National Library of Medicine in Washington DC. It constitutes the most accessible channel to a relatively comprehensive search of the medical literature for most American practitioners. You decide to use MEDLINE as your principal resource, using the Melvyl search software (University of California, Oakland), a computer-based library system available to physicians in your hospital.

Before conducting a search, it is appropriate to decide in advance what criteria will be used to accept or reject the articles found for the review. You have already decided that you are interested only in studies involving preponderantly adult patients undergoing primary evaluation for possible appendicitis in an ED setting and in which the accuracy of WBC count is measured against either the pathology reports for patients undergoing appendectomy or follow-up extending significantly beyond the acute care phase for patients who are discharged home without surgery. To be able to answer the question formulated above, the included studies will have to either report or provide data that permit the calculation of full-performance characteristics, that is, sensitivity, specificity, or likelihood ratios, for WBC counts for the entirety of the study population.

The peripheral WBC count is an example of a diagnostic test result that can take on any of an infinite number of values within a physiologic range. Investigations of the performance of such tests may define a single "cutoff" in such a way that results below this value are considered "negative" and those above, "positive." Although frequently encountered in published studies of diagnostic test performance, such a method treats differences between values close to the "cutoff," such as a WBC count of $9.8\times10^3/\mu L$ in comparison with a WBC count of $10.3\times10^3/\mu L$ (the "cutoff" being $10\times10^3/\mu L$), the same as it treats differences between values at the extremes of the physiologic range. To avoid distortions and to maximize the applicability of the study results to individual patient care, you decide to limit your review to studies that report data on multiple WBC count cutoffs or intervals.

Finally, because the appropriate criterion standard for the ideal study you are seeking involves clinical evaluation and observation, it is appropriate to limit the search to studies involving prospectively accumulated data. It will be important for the purpose of your question for the decision to perform appendectomy not to be influenced by knowledge of the WBC count. Given the commonplace nature of the CBC in patients evaluated in an ED setting for potentially serious disease, it is unlikely that such independence could be reliably achieved through a nonprospective study design. These selection criteria are based on the principle of looking for the "best evidence" obtainable through standard search methods practically available to clinicians. They are specific to the particular clinical question being asked. A systematic or comprehensive review on a topic, such as "laboratory aides in evaluation of acute appendicitis," would use a much broader search strategy.

Appendicitis is a common disease and assay of the peripheral WBC count is a commonly encountered test. In such a situation, it is likely that a relatively large number of studies have been published and could be located in MEDLINE. In such a situation, it is appropriate to determine whether a well-done metaanalysis has been performed. A search, including explosions and synonyms, with the Medical Subject Heading (MeSH) terms "appendicitis," "leukocyte count," and "meta-analysis" for the dates 1966 to 1998 retrieves no articles.

You change your search parameters to MeSH term "appendicitis" with "diagnosis" as a subheading, and "explode: leukocyte count." You further restrict the search to those in human subjects and to studies published after the early 1960s, because the Coulter counter, an automated cell counter, was being widely implemented around this time. Table 1 summarizes the search strategy and results.

The search retrieves 67 studies, of which 11 were prospective. Of these, 5 involved only patients who had surgery. Of the remainder, 4 reported data on only 1 WBC cutoff value. The remaining 2 studies fulfill the criteria decided on in advance and you decide to examine these in more detail.[3,4] Of these 2 studies, Miskowiak and Burcharth[4] report multiple WBC cutoff data in a histogram form that renders it difficult to accurately extrapolate absolute numbers. It is therefore

not considered further. Although other articles found in the search might provide useful information, in the interest of brevity and clarity, you choose to focus on the 1 article that meets all of your search criteria. In this situation, keep in mind that you are trying to rapidly put your hands on the best evidence that answers your clinical question. Because your time is limited, you predetermine article selection criteria to narrow the possible playing field from many articles to a handful or even 1 that best addresses and validly answers your question.

ANALYZING THE EVIDENCE

The study by Dueholm et al[3] included 204 adult patients ages 15 to 45 years. Although the authors do not explicitly state that the study subjects were enrolled from an ED setting, their protocol specifies that laboratory tests were ordered after the primary evaluation by the junior member of the surgical team confirmed appendicitis as a significant clinical possibility. The overall rule-in rate was 29%. The relatively large observation group of 104 patients also suggests a relatively unselected patient population and an appropriate spectrum of patients for a study pertinent to the question at hand.

Table 1.
Search strategy and results.

MEDLINE Search Terms		
Major subject (MeSH)		
Appendicitis		
[AND]		
Subheading Diagnosis		
[AND]		
explode Leukocyte count		
Citations found from MEDLINE	**Year**	**No. of Items**
	1994–1997	14
	1990–1993	10
	1985–1989	12
	1980–1984	9
	1976–1979	10
	1966–1974	12
	Total	67
Reason for elimination		
Retrospective		56
Surgery only		5
Incomplete data		1
Only 1 WBC cutoff		4
	Total	66
Included from MEDLINE search		1

As such, the study by Dueholm et al fulfills 1 of the 2 most important criteria determining the internal validity of a study on the accuracy of a diagnostic test.[1] Such criteria determine the degree of confidence that a clinician should have that the results reported by the investigators can properly be attributed to the intervention being studied. For example, if the study by Dueholm et al[3] had compared the performance of WBC count in distinguishing between a group of patients with perforated appendices at laparotomy and a group of healthy asymptomatic volunteers, it is likely that the WBC count would have appeared to discriminate substantially more accurately than it would in an actual clinical setting.

The study by Dueholm et al also appears, on the whole, to rate satisfactorily on other important internal validity tests. The make and manufacturer of the automated cell counter used for calculation of the WBC counts are identified. The criterion standard used by Dueholm et al was either findings at laparotomy (100 patients) or observation for return of symptoms or need for surgery (104 patients). The length of the observation period to which the latter group of patients was subjected is not made clear by the authors and was apparently limited to the period of the initial hospital stay. The average length of stay is also not reported. It is therefore possible that some of these patients may have "ruled in" at surgery at some point subsequent to their initial hospital discharge.

Although the effect of such an eventuality on the observed results cannot entirely be predicted, it seems unlikely that it would have served to substantially *improve* the observed accuracy of the WBC count. Dueholm et al report that patients judged to have appendicitis had significantly higher WBC counts (median $12.6\times10^3/\mu L$) than those judged to be negative (median $8.9\times10^3/\mu L$). Redefining a substantial number of patients in the observation group as positive for appendicitis would therefore be likely to *decrease* the difference in WBC count between the 2 groups, and hence the ability of the test to discriminate between those patients with and those patients without appendicitis. Should you ultimately find that the WBC count, as judged by the Dueholm study, fulfills the criteria for a clinically useful test identified at the outset of this review, you may need to reevaluate the potential impact of this weakness of the study on that conclusion. Should the results indicate that the WBC count falls short by our criteria, it is unlikely that correction of this potential bias, if present, would alter the assessment.

Three patients without appendicitis had surgically correctable diseases. They were classified as "negative" by

Dueholm et al for the purposes of data analysis. Although this decision could be challenged from the standpoint of an emergency physician's perspective, the small number makes a significant effect on the observed results unlikely. The percentage of patients undergoing surgery in the study by Dueholm et al that were "ruled out" for appendicitis was rather high, 38% or 41%, depending on how the 3 patients with operable lesions other than acute appendicitis are classified. Much lower negative laparotomy rates for appendicitis have been reported[5] and reflect the goals and expectations of many hospital general surgery departments. The disproportionately large number of appendicitis-negative patients undergoing surgery, the definitive criterion assessment for the disease in question, serves to strengthen the internal validity of the study by Dueholm et al.

It is reassuring that Dueholm et al report that the results of the initial WBC count and other laboratory tests remained unavailable to the physicians responsible for deciding whether to take patients to surgery, not only during the initial phase of evaluation but throughout the hospital stay. You may probably safely assume that the pathologists who examined the tissue samples from removed appendices were similarly blinded. It is not explicitly stated that results of postadmission laboratory tests were similarly unavailable or to what extent they were performed. Nonetheless, it seems that the results of the initial WBC count could not have influenced whether a particular patient went to surgery or was simply observed. This is very important in a situation in which a "clinical equivalent" to the conventional criterion standard is used. Dueholm et al[3] observe that some patients in the nonoperated group may have had undiagnosed appendicitis that resolved without surgery. Were this the case and had these patients had an unblinded elevation of the WBC count, they may well have gone to surgery and have been categorized as "positives" by laparotomy rather than "negatives" by observation. This in turn would have served to inflate the accuracy of the WBC count above the more clinically appropriate value. After reviewing the internal validity of the study by Dueholm et al, you decide that, on a 4-point categorical scale of "weak, fair, good, excellent," it deserves a "good" rating and that it is appropriate to review the results.

The principal results of the study by Dueholm et al that pertain to WBC count are summarized in Table 2. The authors provide data on sensitivity, specificity, and predictive values for multiple WBC cutoffs from 7.0 to $19.0 \times 10^3/\mu L$. Unfortunately, sensitivity and specificity do not allow a clinician to directly apply diagnostic test results to patient care.[2] Predictive values calculated from research data only pertain to the prevalence of disease that existed in the research population itself. Although they can be recalculated for any estimated disease prevalence using the sensitivity and specificity of the test in question, the process is time-consuming and not conducive to practical clinical application. The sensitivity and specificity of a test represent the likelihood of a particular test result in patients with and without the disease. They do not themselves provide the likelihood that a

Table 2.

Observed sensitivity, specificity, and false-positive rates for various WBC counts from data reported by Dueholm et al.[3]*

WBC Count ($\times 10^3/\mu L$)	Sensitivity % (95% CI)	Specificity % (95% CI)	False-Positive Rate % (95% CI)
>4.2	100 (NA)	0 (NA)	100 (NA)
>7.0	98 (91–100)	21 (15–29)	79 (71–85)
>9.0	83 (71–92)	50 (42–59)	50 (42–59)
>11.0	76 (63–86)	74 (62–84)	26 (16–38)
>13.0	39 (27–53)	87 (73–98)	13 (2–27)
>15.0	29 (18–42)	93 (78–100)	7 (0–22)
>17.0	15 (7–27)	98 (80–100)	2 (0–20)
>19.0	8 (3–19)	100 (85–100)	0 (0–15)

The false-positive rate equals 1–Specificity and represents the percentage of patients without appendicitis with the WBC count value in question. The percentage of patients with and without appendicitis by the criteria used in the study having WBC counts within any given interval may be directly calculated from this table. This allows the likelihood ratios for WBC counts within each interval to be calculated, as tabulated in Table 3. For example, 2% of patients with appendicitis and 21% of patients without appendicitis had WBC counts between 4.2 and $7.0 \times 10^3/\mu L$. This yields an LR of 2÷21=.1 for WBC counts within this interval.

patient with such a test result has the disease. This latter information is the very thing that clinicians receiving test results on patients undergoing evaluation need to know. It is made possible by the likelihood ratio (LR) for a particular test result.[1,6,7]

The LR for a test result in which only 2 outcomes are possible, "positive" and "negative," can be easily derived from the reported sensitivity and specificity of the test. However, when this approach is applied to continuous diagnostic test variables, such as WBC count, the resulting LRs are distorted in a fashion that exaggerates the diagnostic accuracy of the test.[8] This is because the very high and very low LRs corresponding to extreme test values are "averaged in" to the LRs otherwise attributed to nonextreme test results. An alternative approach, recommended by the Evidence-Based Medicine Working Group and others,[1,6,7,9-11] is to calculate LRs for multiple discrete intervals within the range of clinically likely test results. This approach is usually possible when the data on multiple cutoff values of the test results are provided.

Interval LR values for the WBC count derived from the data reported by Dueholm et al[3] are presented in Table 3. Scanning this table, it is immediately apparent that there is a trend for higher WBC counts to result in higher LRs, that is, a greater probability that patients with appendicitis will be assigned to results in this range than patients without the disease. Within this trend, LRs for different WBC intervals appear to move in different directions. This is likely a reflection of variation related to statistical sampling, the numbers of patients within each WBC

Table 3.
*Likelihood ratios for WBC counts within defined intervals calculated from data reported by Dueholm et al.[3]**

WBC Count ($\times 10^3/\mu L$)	Appendicitis + (% of 59 patients)	Appendicitis – (% of 145 patients)	Likelihood Ratio (95% CI)
4–7	2	21	.10 (0–.39)
7–9	15	29	.52 (0–1.57)
9–11	7	24	.29 (0–.62)
11–13	37	13	2.8 (1.2–4.4)
13–15	10	6	1.7 (0–3.6)
15–17	14	5	2.8 (0–6.0)
17–19	7	2	3.5 (0–10)
19–22	8	0	∞ (NA)

*For each WBC count interval, the LR is the ratio of the percentage of patients found to have appendicitis by laparotomy with WBC counts in that range to the percentage of patients ruled out by laparotomy or extended follow-up with similar WBC counts.

interval being relatively small, and the confidence intervals around the parameters relatively wide.

APPLYING THE EVIDENCE

At this point, it has been determined that the 1 study fulfilling our original criteria for studies likely to answer the clinical question we have asked is of relatively high internal validity and that the study report allows us to derive a data set in the proper form needed for clinical application. It remains for us to determine whether these results derived from a relatively unbiased study are actually usefully applicable to clinical practice. This last remaining task is commonly referred to as assessing the external validity of the study results, that is, their applicability outside of the research setting in which they were derived.

The automated WBC count is readily available, inexpensive, and objectively reproducible. Results of WBC counts ordered in an ED setting tend to be available within 30 to 60 minutes of venipuncture. Because patients who are being evaluated for a significant possibility of appendicitis tend to have intravenous access placement during the observation period, no additional discomfort or inconvenience to the patient need be attributed to the ordering of a WBC count. The population studied by Dueholm et al was, furthermore, very similar to your 22-year-old male patient. The median age of the study subjects was between 23 and 26 years, and the median time from onset of symptoms was 24 hours. Although only 35% of the study population was male, 33 women with clinically identifiable gynecologic diagnoses, as judged by senior surgeons, were excluded. The principal issue pertaining to the applicability of the results of the Dueholm study to your patient population is, therefore, whether the results of the WBC count, judged by the performance characteristics observed by Dueholm et al, will change your patient's management.

On the basis of your clinical experience, you estimated the likelihood of appendicitis in your patient to be between 30% and 50% before ordering the WBC count. Although there is no way to estimate your individual patient's risk exactly, you are confident that, out of 100 patients clinically indistinguishable from the patient in question, 30% to 50% would rule in for appendicitis at surgery. You also believed, before conducting the literature search, that the results of the WBC count, or of any other test, would have to either raise the likelihood above 80% or lower it to 10% or below to be clinically useful.

For you to be confident of the impact of a test result on your decisionmaking, the result would, therefore, have to either raise a pretest probability of 30% to more than 80% or lower a 50% risk to below 10%. A standard nomogram[2,6,7] allows you to determine that the LR of a WBC count result would have to be either 8 or greater or .1 or lower to accomplish this.

Scanning the LRs for WBC counts in different ranges tabulated in Table 3 makes it clear that no values of the WBC count between 7.0 and $19.0\times10^3/\mu L$ result in LRs satisfying these threshold criteria. A WBC count greater than $19.0\times10^3/\mu L$ would apparently justify a decision to operate. However, the sensitivity of 8% for a WBC count greater than $19.0\times10^3/\mu L$ (Table 2) tells us that only 5 patients, comprising less than 3% of the total study population, had results in this range. It is furthermore likely that some, if not all, of these patients were clinically identifiable as being at extremely high risk for acute appendicitis and clinically appropriate for surgery irrespective of the WBC count result.

On the low end of the WBC count range, a WBC count between 4.0 and $7.0\times10^3/\mu L$ would apparently just meet the threshold for allowing consideration of outpatient management. Thirty-two patients, representing 16% of the study population, had WBC counts within this range. All but 1 of these were judged negative for acute appendicitis. The LR of .1 represents a kind of weighted average of the LRs that would correspond to the exact test results in this group.[8] It is likely, therefore, that the WBC count would have to be as low as 5.0 or $6.0\times10^3/\mu L$ for an LR of .1 to apply.

Another way of scanning a data set such as that presented in Table 3 is to ask, "eliminating the extremes, what would my clinical estimates of an individual patient's probability of having appendicitis have to be for the best observed LR values to affect decisionmaking?" The highest interval LR between WBC counts of 7.0 and $19.0\times10^3/\mu L$ was 3.5; the lowest was .29. Using the nomogram alluded to previously, it is easily determined that for a test result with an LR of 3.5 to result in a decision to operate, the patient would have to be judged to have a pretest probability of greater than 50% ("more likely than not to have appendicitis"). Similarly, for an LR of .29 to result in a decision to send the patient home from the ED, the estimate of the likelihood of appendicitis in a particular patient would have to be 30% or lower. The range of likelihoods originally defined through qualitative evaluation of the patient presented at the outset of this review, considering appendicitis as an important possibility, therefore precludes an important impact of WBC counts between 7.0 and $19.0\times10^3/\mu L$ on clinical decisionmaking.

Clinical assessments of individual patients are qualitative judgments. Quantifying them in the terms demanded by the use of LRs may seem to run counter to physicians' clinical instincts. Each clinician must decide for himself or herself what level of suspicion constitutes an appropriate threshold for changing the clinical approach, either in the direction of decisive interventions or of outpatient management. Indeed, these thresholds may vary from 1 patient to the next.

In summary, accepting the results of the study by Dueholm et al at face value, you conclude that, from the standpoint of an "accessible best-evidence" approach, the total WBC count could contribute significantly to clinical decisionmaking for patients with WBC counts greater than 19.0 or less than $7.0\times10^3/\mu L$. In a population similar to that studied by Dueholm et al, this would constitute about 18% of all patients evaluated for acute appendicitis. Restated, there is a less than 1-in-5 chance that the WBC count will appropriately affect clinical decisionmaking in patients with possible, but not clinically unequivocal, acute appendicitis.

Dueholm et al[3] reported results for 3 cutoff values of manually calculated neutrophil counts, for single cutoff values of manual band count and C-reactive protein, and for different single cutoff combinations of WBC count, neutrophil, band, and C-reactive protein. Although some of these combinations suggest utility beyond that of the simple WBC count as a single test, rigorous evaluation of such combinations would require an analysis beyond the context of that possible from the data reported by Dueholm et al. To evaluate the clinical utility of such options would require an entirely different and much more broad-based search strategy than that used for this review.

After completing this exercise, you discuss the process and conclusions of your evidence-based medicine exercise with your emergency medicine colleagues. As a group, you decide to approach the hospital surgeons to discuss streamlining diagnostic testing in patients with suspected appendicitis presenting to the ED. Thus, evidence-based medicine can serve as both an impetus and a bridge to constructively solve traditionally thorny problems that cross departmental boundaries.

We acknowledge the assistance of Michael K Parides, PhD, Division of Biostatistics, Joseph L Mailman School of Public Health of Columbia University, in calculating the confidence intervals around the likelihood ratios for different interval values of the WBC count reported in Table 3 and elsewhere.

REFERENCES

1. 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? *JAMA* 1994;271:389-391.

2. 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? *JAMA* 1994;271:703-707.

3. Dueholm S, Bagi P, Bud M: Laboratory aid in the diagnosis of acute appendicitis: A blinded prospective trial concerning diagnostic value of leukocyte count, neutrophil differential count, and C-reactive protein. *Dis Col Rectum* 1989;32:855-859.

4. Miskowiak J, Burcharth F: The white cell count in acute appendicitis: A prospective blind study. *Dan Med Bull* 1982;29:210-211.

5. Lau WY, Ho YC, Chu KW, et al: Leucocyte count and neutrophil percentage in appendectomy for suspected appendicitis. *Aust N Z J Surg* 1989;59:395-398.

6. Sackett DL, Richardson WS, Rosenberg W, et al: *Evidence-Based Medicine: How to Practice and Teach EBM*. New York: Churchill Livingstone, 1997:118-128.

7. Gallagher EJ: Clinical utility of likelihood ratios. *Ann Emerg Med* 1998;31:391-397.

8. Tandberg D, Deely JJ, O'Malley AJ: Generalized likelihood ratios for quantitative diagnostic test scores. *Am J Emerg Med* 1997;15:694-699.

9. Guyatt GH, Patterson C, Ali M, et al: Diagnosis of iron deficiency anemia in the elderly. *Am J Med* 1990;88:205-209.

10. Guyatt GH, Oxman A, Ali M, et al: Laboratory diagnosis of iron deficiency anemia: An overview. *J Gen Intern Med* 1992;7:145-153.

11. Carey M, Brennan J, Ciconelli R, et al: Likelihood ratios when the results of diagnostic tests are continuous variables [letter]. *Am J Emerg Med* 1998;16:722-723.

Critically Appraised Topic (CAT): The accuracy of leukocyte count in the diagnosis of acute appendicitis

Title

Will the peripheral leukocyte count enable this adult emergency department patient with possible appendicitis to either be safely discharged home with follow-up or taken confidently to surgery, as judged by comparison of WBC values in such patients with pathology reports or prolonged follow-up without return of symptoms or need for surgery?

Reviewed by

Brian K Snyder, MD, and Stephen R Hayden, MD

Date

September 1, 1998

Expiration date

September 1, 2000

Clinical bottom line

If the threshold for probable surgery is taken as a likelihood of appendicitis of 80% or greater, and the threshold for discharge from the ED as below 10%, the total WBC count could contribute significantly to clinical decisionmaking for patients with WBC counts greater than 19.0 or less than 7.0 $\times 10^3/\mu L$. This would constitute about 18% of a population comprised of patients with possible, but not clinically unequivocal, acute appendicitis.

Search strategy

MEDLINE 1966–1998 via Melvyl: using MeSH term “appendicitis” AND Subheading “diagnosis,” AND “explode: leukocyte count,” limited to human studies.

Citation

Dueholm S, Bagi P, Bud M: Laboratory aid in the diagnosis of acute appendicitis: A blinded prospective trial concerning diagnostic value of leukocyte count, neutrophil differential count, and C-reactive protein. *Dis Col Rectum* 1989;32:855-859.

Summary of study

Population 204 adult patients, ages 15–45 years, 35% male presenting with signs/symptoms suggestive of acute appendicitis. One hundred patients went to laparotomy, with 59 having appendicitis for prevalence of the disease in the population of 29%. Thirty-three women with clinically identifiable gynecologic disease excluded. Median duration of symptoms 24 hours.
Intervention WBC count by automated counter together with other cell counts and C-reactive protein.
Outcomes Presence or absence of appendicitis judged by pathology reports or observation for return of symptoms or need for surgery during hospital stay.

Critical review form for diagnostic tests

Dueholm S, Bagi P, Bud M: Laboratory aid in the diagnosis of acute appendicitis: A blinded prospectivetrial concerning diagnostic value of leukocyte count, neutrophil differential count, and C-reactive protein. *Dis Col Rectum* 1989;32:855-859.

Guide	Comments
Are the results valid?	
Was there an independent "blind" comparison with a reference standard?	Yes: The reference standard was histologic evaluation of the appendix or in-hospital observation until symptoms resolved or another disease process was identified. The result of the laboratory CBC was unlikely to have influenced the pathologist's report.
Did the patient sample include an appropriate spectrum of patients to whom the test will be applied in clinical practice?	Yes: All patients between 15 and 45 years of age admitted with right lower quadrant abdominal pain, including all patients who were operated on and those who were observed for suspected appendicitis were included.
Did the result of the test being evaluated influence the decision to perform the reference standard?	No: The WBC count was not available to the surgeons during the entire hospital stay.
Were the methods of the test described in sufficient detail to permit replication?	The WBC counts were determined with a Technicon Hemalog cell counter.
Overall validity rating (weak–fair–good–excellent)=Good	
What are the results?	
What were the likelihood ratios? (calculated from data in Dueholm et al[3])	LR for WBC 4–7*=.10 (0–.39)† LR for WBC 7–9=.52 (0–1.57) LR for WBC 9–11=.29 (0–.62) LR for WBC 11–13=2.8 (1.2–4.4) LR for WBC 13–15=1.7 (0–3.6) LR for WBC 15–17=2.8 (0–6.0) LR for WBC 17–19=3.5 (0–10) LR for WBC 19–22=∞ (NA)
Will the results improve my patient care?	
Will the reproducibility of the test result and its interpretation be satisfactory in my setting?	Yes: Automated WBC count is reproducible across most hospital laboratory settings.
Are the results applicable to my patients?	Yes: The study population observed by Dueholm was very similar in important respects to the average, otherwise healthy, adult patient evaluated in an ED setting for possible appendicitis.
Will the results change my management?	Probably not. The LRs for WBC counts other than extreme values, using the interval cutoff method, are not in the range that fundamentally alters clinical assessment in an intermediate risk population. Fewer than one 1 of 5 patients in the study by Dueholm et al had extreme WBC count values.
Will patients be better off as a result of the test?	Probably not, because the LRs are so poor. Cost, risk, and patient discomfort are minimal.

*Unit of measure=$\times 10^3/\mu L$.
†95% CI indicated in parentheses.