Clinicians may have poor sensitivity in determining whether a given diabetic patient with a foot ulcer has osteomyelitis. Thus many turn to diagnostic tests. The literature was reviewed and data on the sensitivity and specificity of five commonly used diagnostic tests were compiled. Using Bayes’ theorem, the authors’ analysis suggests that pretest probability may be more important in the decision-making process than individual test characteristics. Also, a positive probe-to-bone test is as predictive of osteomyelitis as the other four tests. A negative magnetic resonance imaging test most likely rules out osteomyelitis. Interpretation of any test result is greatly influenced by the pretest probability of disease. Future work needs to focus on aiding the clinician in determining the patient’s probability of disease prior to testing in order to optimize patient care. (J Am Podiatr Med Assoc 88(7): 337-343, 1998)
ease, probability of a test result conditional upon the patient having disease, and probability of a test result conditional upon the patient not having disease.8

The aim of this article is to use Bayes’ theorem to suggest that pretest probability may be more important in the decision-making process than individual test characteristics and that clinicians would benefit from this information prior to the use and interpretation of diagnostic tests.

Materials and Methods

A search of the MEDLINE database (1990–1996) and bibliographies of selected articles revealed 17 articles reporting sensitivity and specificity for five diagnostic tests (ie, probing to bone,10 radiographs,1, 11-18 three-phase technetium bone scan [99mTc],1, 11-16, 19, 20 combination $^{99m}$Tc with indium-labeled leukocyte scan [99mTc/$^{111}$In],13, 14, 16, 19-23 and magnetic resonance imaging [MRI]2, 15, 17, 18, 24, 25) used for diagnosing osteomyelitis in the presence of diabetic foot ulcers. This framework was selected because it has been reported that 94% of diabetic pedal osteomyelitis is contiguous with open ulceration.26 Medical Subject Headings (MeSH) used included the following: osteomyelitis, radiography, ROC (receiver operator characteristic) curve, diabetes mellitus, diagnosis, diagnostic imaging, and diagnostic error.

For each test, the authors extracted estimates of sensitivity and specificity. For tests reported more than once, summary estimates were calculated by weighting for sample size.8 Summary sensitivity and specificity data are reported in Table 1. The distribution of these data is shown in Figures 2 and 3. Intermediate pretest probabilities (0.25–0.50) may represent the range in which clinicians will most commonly use diagnostic tests. Therefore, these pretest probabilities are reported in the results.

Bayes’ theorem8, 9 was used to calculate probabilities of disease given positive and negative test results. Prior probabilities included the following: 0.05, 0.25, 0.50, 0.75, and 0.95. A glossary of terms for the application of Bayes’ theorem appears at the end of this article. Bayes’ theorem can be represented by the following equations:

Probability of disease after a positive test =

\[
\frac{(x \text{ (sensitivity)})}{(x \text{ (sensitivity)}) + (1-x \text{ (1-specificity)})}
\]

Probability of disease after a negative test =

\[
\frac{(x \text{ (1-specificity)})}{(x \text{ (1-specificity)}) + (1-x \text{ (specificity)})}
\]

where \(x\) = pretest probability

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe to bone</td>
<td>76</td>
<td>66</td>
<td>85</td>
</tr>
<tr>
<td>Radiographs</td>
<td>363</td>
<td>54</td>
<td>80</td>
</tr>
<tr>
<td>$^{99m}$Tc</td>
<td>333</td>
<td>91</td>
<td>46</td>
</tr>
<tr>
<td>$^{99m}$Tc/$^{111}$In</td>
<td>250</td>
<td>88</td>
<td>82</td>
</tr>
<tr>
<td>MRI</td>
<td>183</td>
<td>92</td>
<td>84</td>
</tr>
</tbody>
</table>

\(a\) Except for probe to bone, which was based on one study.
Results

Of the five diagnostic tests studied, radiographs exhibited the least desirable test characteristics, with a mean sensitivity of 54% and a mean specificity of 80%. Magnetic resonance imaging exhibited the most desirable test characteristics, with a mean sensitivity of 92% and a mean specificity of 84% (Figs. 2 and 3). A comparison of radiographs and MRI illustrates the impact of prior probability. At a pretest probability of 0.25, a positive MRI suggests a post-test probability of osteomyelitis of 0.66. At a pretest probability of 0.50, a positive radiograph suggests a post-test probability of osteomyelitis of 0.73. At a pretest probability of 0.25, a negative radiograph suggests a post-test probability of 0.16. At a pretest probability of 0.50, a negative MRI suggests a post-test probability of 0.09 (Figs. 4 and 5).
Figure 4 suggests that given a positive test, probing to bone performs approximately as well as the other tests. Figure 5 suggests that MRI most likely rules out osteomyelitis.

Figures 2 and 3 display the variation in the individual studies’ reports of sensitivity and specificity for each of the five diagnostic tests. Figure 3 displays the variation of radiographs and MRI. Figure 4 displays the variation of radiographs, $^{99m}$Tc/$^{111}$In, and $^{99m}$Tc.

In the 17 studies described, 71% (ranging from 29% to 100%) of patients had the diagnosis made using bone biopsy as the “gold standard,” and 35% (6/17) of the studies made the diagnosis based solely on bone biopsy.

Discussion

The authors’ analysis clearly demonstrates that the pretest probability may be a more important factor in the post-test probability than the individual diagnostic test characteristics (eg, sensitivity and specificity). The example of MRI compared with radiographs, at differing pretest probabilities, makes this argument quite clear. Even when summary estimates of test characteristics were used, wide variation existed in sensitivity (54% to 92%) and specificity (46% to 84%) for each of the five diagnostic tests. Given this variation in individual test performance, pretest probability still was a more important predictor of post-test probability. For example, probing to bone (sensitivity of 66%) performed as well as MRI (sensitivity of 92%) at all pretest probabilities given a positive test.

There are limitations to this analysis. Individual study limitations have been well described and are beyond the scope of this article. Probing-to-bone test characteristics were based on a single study. With the wide ranges of specificity and sensitivity demonstrated with other tests, it would seem likely that variation is also possible with probing to bone. Also, the authors have mentioned elsewhere the potential of spectrum bias. This is the difference in disease rate between the testing population and the clinically relevant population of diabetic patients with infected foot ulcers. Lastly, probing to bone was determined prior to debridement and after removal of eschar. Even though two examiners were involved, whether or not standard operational definitions existed for debridement and eschar remains uncertain.

Individual institutions may vary in their laboratories’ processes of care. For instance, the results of $^{111}$In scanning studies may not be generalizable for institutions that require transportation of blood products for labeling. The authors identified no studies that addressed the in vitro survival of leukocytes over time. At each subsequent stage of the $^{111}$In scanning process (eg, drawing, labeling, and reinjection), there is loss of leukocyte number and activity.

Bone biopsy is often cited as the “gold standard” for diagnosing osteomyelitis. However, there was great variation in the use and yield of bone biopsies in the 17 studies reviewed. The percentage of patients evaluated with bone biopsy as the first-choice technique ranged from 29% to 100%. In all of the studies, biopsies were performed on 312 patients: 215 were positive and 97 were negative. In two similarly sized studies, the first reported 9 positive and 12 negative, and the second reported 23 positive and none negative. Part of the variation in yield may be explained by the clinician’s inaccuracy in determining the presence of osteomyelitis. This referral bias combined with variations in surgical technique could further drive the differences in yield. Underutilization may also explain the variation in use. For example, a decision-analytic model concluded that at prior probabilities of 0.08 to 0.5, the optimal strategy would be to start with bone biopsy. This conclusion gives empirical support to the strategy supported by others, who “initially consider bone biopsy and culture, then assess whether an imaging modality may obviate the need for surgery or guide the surgical plan.” These recommendations are quite clear; however, are they the current standard of care?

One could also argue that weighting by sample size could further compound the cumulative limitations of the studies in the authors’ analysis, particularly if there were more bias in the larger studies. Conversely, weighting by sample size could result in more accurate information about sensitivity and specificity by randomly distributing error in both directions.

The authors’ analysis contradicts the current focus of the literature. Most studies have focused on the later stages of the decision process, such as diagnosis and treatment. The one study that addressed the prevalence of osteomyelitis in this context found it to be 68%. The implications of better knowledge of pretest probability are elucidated in Figure 6. Note how better knowledge of the pretest probability improves the utility of all the “downstream” stages in the decision process, without directly altering either tests or treatments. It may now be assumed that as certainty of disease increases and the post-test probability increases, more practitioners will reach their individual treatment threshold and initiate treatment. This may translate into better care of those with disease currently not clinically identified. Conversely, this explicit knowledge may decrease a patient’s pretest probability, sparing him or her unnecessary testing and treatment offered by practitioners with a particularly low treatment threshold. Therefore, added
clinician certainty may help the patient and the medical system avoid excessive loss of financial resources, function, and well-being.

Pretest probability and prevalence are occasionally used interchangeably. However, there are limitations for using prevalence in the diagnosis of osteomyelitis. Traditionally, prevalence rates are determined through cross-sectional studies. These represent "snapshots of the health experience of the population at a specified time." A problem may arise when applying a population-based rate to an individual clinical patient. Diabetic foot ulcers present at various

Figure 6. Decision-making process for diagnosis and treatment of osteomyelitis in patients with diabetic foot ulcers.
stages, with various clinical findings. Another limitation of prevalence is that it gives limited information on the prognosis and natural history.33

Perhaps a more useful method would be the formulation of a clinical prediction rule. This alternative involves determining pretest probability based on clinical findings made through systematic clinical observations.34 Others have defined a clinical prediction rule as a “decision-making tool for clinicians that included three or more variables obtained from history, physical examination, and simple diagnostic tests.”35

Commonly, this results in a probability determined through a multivariate regression model that uses risk-factor data derived from a prospective cohort design.

Methodological standards have been proposed for the formation of clinical prediction rules.34, 35 One study34 suggested that: 1) the outcome be defined, clinically important, and blindly assessed; 2) predictive variables be identified, defined, and blindly assessed; 3) patient age and sex be stated; 4) study site(s) be described; 5) test misclassification be reported; 6) the effects of clinical use be prospectively measured; and 7) mathematical techniques be described.34 Another study35 suggested four additional standards: 1) describing the results of the clinical prediction rule; 2) prospective validation replacing test misclassification rate; 3) reproducibility; and 4) sensibility.

A clinical prediction rule for diagnosing osteomyelitis in the presence of foot ulceration may have special considerations. First, a review of the literature would define potential predictive variables. Many predictor variables have already been described: probing to bone,10 periulcer inflammation,1 and ulcer size.1 Second, the rule should be sensible. It must make clinical sense, be easy to use, and suggest a course of action.8 If history can be a proxy for physician adoption and use of the rule, an action-oriented and simple approach may be effective. Evidence for this may be the risk-factor stratification36 and treatment interval algorithm37 used by many who treat diabetic feet. Third, a large multi-institutional prospective design would probably be the only way to study this rare outcome. These noninclusive considerations may explain why this work has not yet been done.

Conclusion

The authors’ analysis suggests that interpretation of any test result is greatly influenced by the pretest probability of disease. Future work needs to focus on aiding the clinician in determining the patient’s probability of disease prior to testing in order to optimize patient care. Probing to bone performs well against more expensive and often less accessible tests in diagnosing diabetic pedal osteomyelitis. A negative MRI test most likely rules out osteomyelitis.

Glossary

Prevalence: The proportion of the population that has a disease at a point in time.32

Incidence: The rate of occurrence of new disease during a period of time.32

Predictive value: The proportion of study patients with a test result that is calculated from a defined population of patients; dependent on prevalence.8

Prior probability (pretest probability): The probability of disease prior to testing; occasionally used interchangeably with prevalence.

Posterior probability (post-test probability): The probability of disease given new information calculated using Bayes’ theorem; dependent on the prior probability of disease.8

Sensitivity (true-positive rate): The likelihood that a diseased patient has a positive test.8

Specificity (true-negative rate): The likelihood that a nondiseased patient has a normal test result.8


References


342 Journal of the American Podiatric Medical Association