In practice, laboratory tests are often ordered in a highly uncritical manner. They are comparatively cheap (for example, in comparison to imaging procedures), but highly sensitive and specific. This implies that if many different laboratory parameters are measured, this will supply clinically relevant information on the disease fast, with little effort and relatively cheaply. This is even taken to be the case if the tested parameters have little or nothing to do with the patient's symptoms. This includes routine profiles (which may be very extensive), as well as screening for diseases such as cancer which should be diagnosed before clinical symptoms develop and infectious diseases, such as borreliosis, which develop in phases.

This overlooks the fact that the reliability of test results depends on a clear indication. Although this aspect is frequently mentioned in public discussions of the value of screening (1), it is also important in daily medical practice. This does not of course apply to recommended screening tests (such as neonatal screening), as these issues are explicitly considered in the recommendations.

The present article sketches the underlying relationships in a largely non-mathematical form and explains the consequences for ordering diagnostic tests in daily medical practice. This problem is related to statistics, an area in which intuitive ideas are often misleading. The underlying problem is displayed in the following multiple choice question:

A laboratory test (for example, for borreliosis) has a diagnostic specificity of 98%. How probable is it that a patient who gives a positive test result does in fact have this disease?

a) You have to know the sensitivity too to be able to answer this question.
b) 98%
c) (1-specificity) × 100 (%) = 2%
d) None of these answers is correct.

Readers who can answer this question correctly can stop here. (The solution is at the end of the article). This article can be very helpful for practical medical work, as the underlying problem appears repeatedly in many different variations.

Confronted with this problem, most people attempt to solve it with the help of specificity alone. The specificity states the proportion of healthy subjects for whom a negative test result is (correctly) given. Conversely, 1-specificity gives the proportion of healthy subjects for
Sensitivity, specificity and positive predictive value

Sensitivity and specificity are statistical parameters which are well known for most tests, as they are easy to determine in principle. This is done by testing a defined number of samples from patients who are either known to be healthy or are known to be suffering from the disease in question. Each patient sample can give either a positive or a negative result. The results can be presented in a 2 × 2 matrix (4-field table). Instead of defined samples, a field study can be performed. A reference procedure must then be used retrospectively to establish whether the tested material comes from a healthy or an ill person. Table 1 gives the result of such a field experiment.

The sensitivity gives the proportion of ill persons positively recognized by the test. It can be seen immediately from the 4-field table that the sensitivity must be a/(a+c), where (a+c) is the number of ill persons in the test cohort, of whom (a) patients give a positive test result. Thus the sensitivity of the test in this example is 398/(398+22) = 0.9476, or roughly 94.8%.

The specificity gives the proportion of healthy persons in this test giving a negative test result, corresponding to d/(b+d), where (b+d) is the total number of healthy test persons in the test cohort, of whom (d) persons give a negative test result. Thus the specificity of the test in this example is 1012/(1012+12) = 0.9883, or roughly 98.8%.

How probable is it that a person with a positive test result is in fact ill? This is what mainly interests the responsible physician in a clinical situation. This probability is called the positive predictive value (PPV). It can be seen in the 2x2 matrix that the number of persons testing positive is (a+b) = 410. Of these, 398 (a) are in fact ill. The probability that a person with a positive test result is also ill in our example (the PPV) is then a/(a+b) = 398/410 = 0.9707 or about 97.1%.

This is somewhat different from the specificity. When determining the specificity, the right column (healthy) must be evaluated, but when determining the PPV, the upper line (positive test result) must be used. In the first case lines are evaluated, in the second columns.

Now the difference in this example is not very great (specificity 98.8%, in comparison with PPV 97.1%). The reason for this is that the proportion of ill persons (prevalence of the disease) was very high in this example, corresponding to (a+c)/(a+b+c+d)—about 29.1%. The number of healthy persons in the field experiment could be ten times higher, or 96.1%, corresponding to the prevalence of 3.9% (Table 2). This is often a realistic assumption. This gives rise to the following values:

<table>
<thead>
<tr>
<th>Test result</th>
<th>“True Value”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ill</td>
</tr>
<tr>
<td>positive</td>
<td>a (398)</td>
</tr>
<tr>
<td>negative</td>
<td>c (22)</td>
</tr>
</tbody>
</table>

The sensitivity is then still 94.8% and the specificity is also unchanged: d/(b+d) = 1012/(1012 + 12) = 0.9883 or about 98.8%.

If we set the prevalence even lower, for example, to 0.41%—corresponding to a further 10-fold increase in the number of healthy test persons—the PPV would drop to 24.9%. In this case, a positive test result means that there is a probability of more than 75% that the test person does not have the disease.

Thus, a positive result can come from either an ill or a healthy person. In the latter case, it is a false positive or non-specific result. When the prevalence of the disease is lower, there are fewer ill persons in the test cohort, more healthy persons are tested and the probability increases that a positive test result is false.

In summary, this means that the positive predictive value (PPV) not only depends on the sensitivity and specificity, but also on the prevalence of the disease in the test cohort. The lower the prevalence, the lower the PPV is.

Analogous arguments apply to the negative predictive value (NPV), the probability that a person with a negative test result is indeed not ill. It can be seen from Table 1 that the NPV = d/(c+d) = 1012/(1012 + 22) = 97.7%. In contrast to the PPV, the NPV decreases with increasing prevalence. If the number of ill persons in the test cohort is increased by a factor of 100, corresponding to the prevalence of 97.6%, the following values can be calculated (Table 3):

<table>
<thead>
<tr>
<th>Test result</th>
<th>“True Value”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ill</td>
</tr>
<tr>
<td>positive</td>
<td>a (39800)</td>
</tr>
<tr>
<td>negative</td>
<td>c (2200)</td>
</tr>
</tbody>
</table>

The sensitivity, a/(a+c) = 39800/(39800 + 2200) = 94.8%, and specificity, d/(d+c) = 1012/(1012 + 12) = 98.8% are unchanged. However, the NPV is now only d/(c+d) = 1012/(2200+1200) = 31.5%, i.e. only 31.5% of patients with a negative test result are in fact healthy. Thus, a negative test result can come from either healthy or ill persons. In the latter case, it is a false negative result. With increasing prevalence, more and more ill persons are tested. Corresponding to this, the probability that a negative test result is a false negative also increases. These mathematical relationships can also be presented as calculations of probability (Box 2).
This corresponds to a probability of about 1.25% that a positive test result will be 2025 genuine positive findings. The probability of 1-specificity = 2% of the negative patients, but there will be 100 genuine positive findings. The probability of 25 genuine positive results is actually obtained for 100 000 tests in the population. It will be neglected that the prevalence of the disease and should not be underestimated. The arguments apply equally well to laboratory tests or to other investigations.

Conclusion
Sensitivity and specificity are test-specific properties which the physician cannot actively influence. This assumes that the test is properly performed and evaluated, taking into account the steps before and after the analysis. On the other hand, the reliability of a positive test result—the positive predictive value—is critically dependent on the prevalence of the disease in the test cohort and this is something the physician can influence. As a matter of principle, tests should only be ordered when they are indicated, as it is only then that the test result can be clinically evaluated. Results from non-indicated orders are clinically useless without a well-founded database on the prevalence of the disease and should therefore not be ordered. This is unrelated to economic or ethical considerations.

Although the borreliosis test was used as an example, this applies to all laboratory tests. The arguments apply equally well to laboratory tests or to other investigations.

Bayes’ theorem
Sensitivity, specificity, and positive predictive value can also be expressed as conditional probabilities. Let p(B/A) be the probability that the event B occurs under the condition A. For example, the conditional probability p(B) (test result positive/test person ill) is exactly the same as sensitivity as defined in Box 1 and is given by a/(a+c). However, the inverse probability p(A/B) (test person ill/test result positive), the positive predictive value, is usually of more interest. In some sense, effect and cause are swapped. The probability is usually known that the cause (the disease) leads to a positive test result.

We are often interested in another aspect, namely how a positive test result can lead to the conclusion of the cause (disease). The correct mathematical relationship is given by Bayes’ theorem, as originally presented by Thomas Bayes: p(A/B) = p(B/A) × p(A)/p(B), or the transformation of this, as p(B) is not known directly: p(A/B) = p(B/A) × p(A) : [p(B/Ac) × p(Ac) + p(B/A) × p(A)], where Ac is the complement of A, i.e., A is not present.
including X-ray, endoscopic, sonographic, electrocardio-
graphic or clinical procedures. If the test or investigation
is not indicated, this reduces its positive predictive value
and increases the number of false positive test results.

The correct answer to the initial multiple choice
question was—d.

Conflict of interest statement
The authors declare that no conflict of interest exists according to the
guidelines of the International Committee of Medical Journal Editors.

Manuscript received on 6 February 2007, revised version accepted on
19 October 2007.

Translated from the original German by Rodney A. Yeates, M.A., Ph.D.

REFERENCES

2. O’Connell S, Granström M, Gray JS, Stanek G: Epidemiology of European
3. Talaska T: Borrellose-Epidemiologie. Brandenburgisches Ärzteblatt
www.laekb.de/15/15Beitraege/95021TH0211.pdf
4. Mehnert WH, Krause G: Surveillance of lyme borreliosis in Germany,
Fingerle V: Improvement of Lyme borreliosis serodiagnosis by a newly
developed recombinant immunoglobulin G (IgG) and IgM line
immunoblot assay and addition of VlsE and DbpA homologues. J

Corresponding author
Prof. Dr. med. habil., Dr. rer. nat. Wilfried Bautsch
Institut für Mikrobiologie,
Immunologie und Krankenhaushygiene
Städtisches Klinikum Braunschweig gGmbH
Celler Str. 38
38814 Braunschweig, Germany
w.bautsch@klinikumbraunschweig.de