All that glistens is not gold

Johannes L.H. Evers
Deputy Editor
E-mail: jlh.evers@mumc.nl

André C. Van Steirteghem
Editor-in-Chief
E-mail: vansteirteghema@humanreproduction.co.uk

As a general gynaecologist you see a 28 year old patient who presents with increasingly painful menstrual periods since she stopped oral contraceptives one year ago in order to conceive. Since intercourse is now also becoming painful, she decided to consult you. After taking her history and doing a pelvic examination you consider that she might have endometriosis and propose a diagnostic laparoscopy. Operation room capacity is limited however and long waiting lists exist. Are there easier ways to make the diagnosis?

Not until recently. In this issue of the journal, two groups, one from Australia (Al-Jefout et al., 2009) and one from Belgium (Bokor et al., 2009) report on a new diagnostic test with a reliability ‘close to the accuracy of laparoscopic assessment by experienced gynecological laparoscopists’ (Al-Jefout et al., 2009): the demonstration of nerve fibers in the functional layer of the endometrium. Both studies were relatively small for reliable evaluation of a new diagnostic procedure \( n \approx 99 \) and \( n = 40 \), respectively. In the biggest, Al-Jefout’s study, 91% of women in whom they found nerve fibers indeed had endometriosis diagnosed at subsequent (blinded) laparoscopy, as compared with only 4% of those in whom they failed to find nerve fibers. This seems too good to be true. Or is it?

Sensitivity (Sn) and specificity (Sp) are diagnostic test characteristics. They are independent of the prevalence of disease in the population investigated. Positive and negative predictive values (PPV and NPV) are also test characteristics, but they are dependent on the prevalence of disease. The ‘normal’ prevalence of endometriosis in women in the reproductive age group is 10% (Rogers et al., 2009). It may rise to 30–50% in women with subfertility and/or pain. The prevalence in the two studies in this issue was 65% (Al-Jefout et al., 2009) and 50% (Bokor et al., 2009), respectively, relatively high prevalences therefore. This will not affect the test’s Sn and Sp; the PPV and NPV will be affected however. If we were to recalculate the composition of Al-Jefout’s study group to reflect a prevalence of endometriosis of 10%, obviously the Sn would remain 98% and the Sp 83%, but the PPV would decrease from 91 to 39% and the NPV would increase from 97% to over 99%. In other words, the test would still succeed in ruling out the diagnosis of endometriosis and prevent unnecessary laparoscopies in over 99% of patients, but in those that we would do a diagnostic laparoscopy based on the findings of the endometrial nerve fiber test we now would not find endometriosis in the majority, 61%.

This illustrates that diagnostic tests may do extremely well in tertiary care centers (where new tests are usually developed) that see a high prevalence of patients with a given disease, but that at the same time, with the same test characteristics, they may disappoint the general gynaecologist who sees patients with the same disease, but much less frequently, in his daily outpatient clinic.

A way around this problem may be to use likelihood ratios (LR) to characterize a new diagnostic test. A positive likelihood ratio (LR+) gives the likelihood of an abnormal test in a person with a given disease as compared with a person without that disease. Values between 5 and 10 generate moderate shifts from pre- to post-test probability of disease; values above 10 generate large and often conclusive shifts from pre- to post-test probability (Guyatt and Rennie, 2002). The LR+ of the new endometrial nerve fiber test is 5.76 and will therefore only generate a modest change in probability of finding endometriosis at laparoscopy. A negative likelihood ratio (LR–) gives the likelihood of a normal test in a patient with the disease compared with a patient without the disease. Values between 0.2 and 0.1 generate modest and values below 0.1 large shifts from pre- to post-test probability of disease. The new test has a LR– of 0.02 and as such (and if confirmed in larger patient groups) appears to be a clinically useful test to rule out endometriosis and postpone a diagnostic laparoscopy, at least for diagnosing endometriosis.

This brings us to the second point that demands attention: what really is it that this test detects? Endometrial implants wax and wane in many patients, do we really wish to identify such an ephemeral condition? And if so, what are we going to do? Are we prepared to perform a laparoscopy in all positive patients to remove the (visible) implants of endometrium in their pelvic lining? And what about the dozens (or perhaps even hundreds) of still occult lesions? And for that matter, even if we were to succeed in removing all visible and invisible lesions, is there evidence that early diagnosis really leads to improved future fertility or a better quality of life?

The new test at first sight looks very promising, and that is the reason why we decided to publish it. We also have to realize, however, that a diagnostic test, no matter how promising, may do more harm than good, e.g. by subjecting patients to unnecessary or...
even potentially harmful procedures. Do we really understand the target disorder well enough to warrant the degree of effort and expenditure? The obvious next step will be to study the clinical value of the endometrial nerve fiber test prospectively in large and more representative groups of patients. Until a rational treatment of endometriosis becomes available, however, the risk exists that even this effort may prove futile.

References


