

2009

Diagnosis in general practice: Using probabilistic reasoning

Jenny Doust

Bond University, jenny_doust@bond.edu.au

Follow this and additional works at: http://epublications.bond.edu.au/hsm_pubs

 Part of the [Analytical, Diagnostic and Therapeutic Techniques and Equipment Commons](#)

Recommended Citation

Jenny Doust. (2009) "Diagnosis in general practice: Using probabilistic reasoning" *BMJ*, 339 (7729), 1080-1082.

http://epublications.bond.edu.au/hsm_pubs/181

DIAGNOSIS IN GENERAL PRACTICE

Using probabilistic reasoning

Jenny Doust

Having a sense of the accuracy of diagnostic tests will help general practitioners to interpret and use the tests appropriately and, as in the example of chest pain, avoid unnecessary testing

PRACTICE, p 1083

Faculty of Health Sciences and Medicine, Bond University, Gold Coast, QLD 4029, Australia

jdoust@bond.edu.au

Cite this as: *BMJ* 2009;339:b3823
doi: 10.1136/bmj.b3823

ABSTRACT

Diagnostic tests—whether clinical signs, imaging, or laboratory tests—are imperfect: there is always a possibility that test results are inaccurate and our diagnosis is wrong. However, we need to make decisions about whether to treat or not to treat patients, and so we need to feel confident that our diagnosis is above a certain threshold before we decide to treat a patient and below a certain threshold if we decide to withhold treatment. The threshold depends on the disease and the potential harms and benefits of treating or not treating patients. Unless we have clear strategies to cope with the uncertainties of testing, false positive results mislead us to treat some patients unnecessarily and false negative results lead us to fail to treat some patients adequately or in time.

What is probabilistic reasoning?

Probabilistic reasoning is used when we consider the diagnostic accuracy of tests in our clinical decisions. It is also called Bayesian reasoning, being based on Bayes' theorem, in which the probability of a hypothesis is modified by further data. As primary care doctors, we use tests every day to decide whether our patients have a particular disease, but we often ignore the uncertainty inherent in the test results. Only rarely can we define how well a test rules in or rules out a disease. Does this matter?

An example of probability revision

We can combine how likely it is that a patient has a disease before having the test (the pretest probability) with the accuracy of the diagnostic test (the sensitivity and specificity) to calculate the probability that a patient has a disease after having the test (the post-test probability). As an example, consider the scenario of a 35 year old woman who presents with symptoms of dysuria to her

Table 1 | Results of dipstick testing in 1000 women presenting to a general practitioner with dysuria

	Urinary tract infection present	Urinary tract infection absent	Total
Either nitrites or leucocyte esterase positive	True positives (n=495)	False positives (n=180)	675
Both tests negative	False negatives (n=55)	True negatives (n=270)	325
Total	550*	450	1000

*Pretest probability=55%.

general practitioner. The chance that such a woman has a urinary tract infection is approximately 55%.¹ The sensitivity for a urinary tract infection if a dipstick test is positive for either nitrites or leucocyte esterase is 90%, as measured against culture of a midstream specimen.¹ This is the proportion of patients with the disease who test positive. The specificity if both nitrites and leucocyte esterase are negative is approximately 60%. This is the proportion of patients without the disease who have a negative test result. Using this information, we can calculate the probability that the woman has a urinary tract infection after a urine dipstick result (table 1). This requires some mathematical manipulation and can be complex to follow, but is key to understanding the appropriate use and interpretation of diagnostic tests.

Of 1000 women, 550 will have a urinary tract infection and 450 women will not (based on the prevalence of disease or pretest probability). Of the 550 women with disease, there are 495 true positives (550×90% sensitivity), and of the 450 women without disease, there are 270 true negatives (450×60% specificity); completing the table, the number of false negatives is 550–495=55 and the number of false positives is 450–270=180.

In clinical practice, we need to be able to calculate the chance that a patient does or does not have a disease when they have a positive or negative test result. Given a pretest probability of 55%, if either of the dipstick tests is positive, the probability that the woman has a urinary tract infection (the positive predictive value) is the proportion of true positives to all positive test results—that is, 495 divided by 675, or 73%. You may not think this conclusive enough to determine if an infection is present, and so may decide to send a urine specimen for microbiological confirmation. Conversely, if both tests

This series aims to set out a diagnostic strategy and illustrate its application with a case. The series advisers are Kevin Barraclough, general practitioner, Painswick, and research fellow in community based medicine, University of Bristol; Paul Glasziou, professor of evidence based medicine, Department of Primary Health Care, University of Oxford; and Peter Rose, university lecturer, Department of Primary Health Care, University of Oxford

KEY POINTS

The probability that a patient has a disease is related to the diagnostic accuracy of the test and the pretest probability of disease—that is, how likely it is that the patient had the disease before the test

When the pretest probability of disease is low, such as in a general practice setting, the probability that a positive test result is a false positive is high

To avoid diagnostic errors and to be able to interpret and use diagnostic tests appropriately, general practitioners need to have a sense of the diagnostic accuracy of the tests that they use

are negative, the probability that the woman does not have a urinary tract infection (the negative predictive value) is the proportion of true negatives to all negative test results—that is, 270 divided by 325, or 83%. The probability that a woman has a urinary tract infection in this case is 17%. This may not be low enough to completely rule out infection, and further testing by microbiological culture may be necessary.

Although the positive and negative predictive values are the clinically useful measures, they are not generally reported in studies of the accuracy of diagnostic tests as predictive values vary greatly with changes in the pretest probability. To illustrate how the pretest probability affects the post-test probability, we can calculate post-test probabilities for the same test in an asymptomatic pregnant woman, using an estimate of the prevalence of asymptomatic urinary tract infection in pregnant women in Rochester, Minnesota (2.4%) (table 2).² If we assume the same sensitivity and specificity as above but a pretest probability of 2.4%, and complete the table as before, the positive predictive value is now 22 divided by 413, or 5%; 95% of all positive results are false positives. The negative predictive value is, however, now 585/587, or close to 100%.

Women with recurrent urinary tract infection have a high pretest probability, about 84% (table 3).³ Assuming the same sensitivity and specificity, but now using a pretest probability of 84%, a woman with a positive test result has a post-test probability of 756/820 or 92% (table 3). However, even if both tests are negative, the negative predictive value is 96/180 or 53%, so the probability that the woman has a urinary tract infection even with a negative test result is now 47%.

When is probabilistic reasoning used?

We use probabilistic reasoning intuitively whenever we consider the likelihood of a patient having a disease in the light of a new piece of information. In the diagnostic stages described previously, probabilistic reasoning occurs during the revision of the diagnosis (fig 1).⁴ In the accompanying article on the diagnosis of chest pain in general practice, Jelinek and Barraclough describe how different types of chest pain and the results of a

Table 2 | Results of dipstick testing in 1000 asymptomatic pregnant women

	Urinary tract infection present	Urinary tract infection absent	Total
Either nitrites or leucocyte esterase positive	22	391	413
Both tests negative	2	585	587
Total	24*	976	1000

*Pretest probability=2.4%.

Table 3 | Results of dipstick testing in 1000 women with recurrent urinary tract infection

	Urinary tract infection present	Urinary tract infection absent	Total
Either nitrites or leucocyte esterase positive	756	64	820
Both tests negative	84	96	180
Total	840*	160	1000

*Pretest probability=84%.

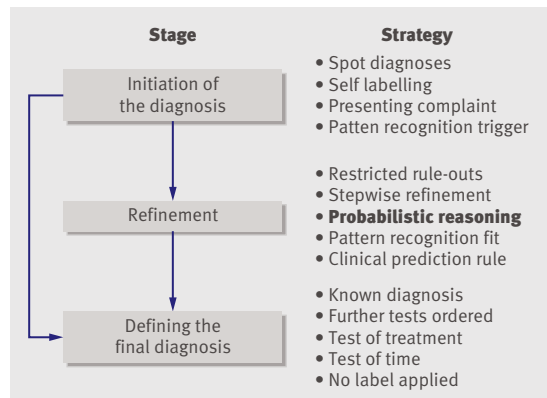


Fig 1 | Diagnostic stages and strategies

stress electrocardiogram revise the probability that a patient has coronary artery disease.⁵

Probabilistic reasoning is also used when deciding whether it is worthwhile to order further tests. For example, we might consider that the benefits outweigh the harms of treatment when the probability of a urinary tract infection is greater than about 60%. If a woman has a pretest probability of disease of 90%, even if the dipstick test is negative, her post-test probability of disease is above 60%. In this case, the dipstick test does not contribute to the decision on management and should not be ordered.

How does probabilistic reasoning go wrong?

We make errors by believing false positive and false negative test results and by ordering inappropriate diagnostic tests. To avoid these errors, we need to have a sense of both the pretest probability of disease and the diagnostic accuracy of test results. We do not need to be able to do these calculations exactly. Ultimately we need to decide whether it is worthwhile to treat a patient and whether it is worthwhile to order a diagnostic test.

When the prevalence or pretest probability of disease is low, the probability that a positive test result is a false positive becomes quite high. This is often the case in general practice, and it is always the case in screening tests. For example, only about 1 in 7 women with a positive screening mammogram will have breast cancer, and only 1 in 88 patients with a positive fecal occult blood test will have colorectal cancer⁶—most patients with a positive test result will be false positives.

One of the specific skills of a general practitioner is to understand the pretest probabilities of disease in his or her clinical setting and to interpret the test results and order appropriate diagnostic tests. The difference in pretest probabilities between primary care and secondary care is one reason why clinicians find it difficult to move between these settings.

We also need to have better evidence about the diagnostic accuracy of tests, and particularly in the clinical setting in which they are going to be used.

How can we improve?

In clinical practice, we need to be aware of false positive and false negative test results in our clinical decision

making. The information in the three case scenarios can be shown on a graph that plots the post-test probabilities for positive and negative results against each possible pretest probability from 0 to 100%, using the sensitivity and specificity of the test (fig 2). The woman in the first case scenario moves from the pretest probability of 55% (table 1) to a post-test probability of 73% if the dipstick test is positive and to a post-test probability of 17% if the dipstick test is negative. An asymptomatic pregnant woman moves from a pretest probability of 2.4% (table 2) to a post-test probability of 5% if the dipstick test is positive and to a post-test probability of about 0% if the dipstick test is negative. A woman with recurrent urinary tract infection moves from a pretest probability of 84% (table 3) to a post-test probability of 92% if the test is positive and to a post-test probability of 47% if the test is negative.

Using this graph, which is based on the sensitivity and specificity of the test, means we do not have to recalculate a 2x2 table for each case and shows graphically how well a test rules in or rules out the diagnosis. The further the curve above the diagonal is away from the diagonal, the greater the ability of the test to rule the disease in, and the further the curve below the diagonal is away from the diagonal the greater the ability of the test to rule the disease out. A useful mnemonic is SpPin (when a specific test is positive, it rules the diagnosis in) and Snnout (when a sensitive test is negative, it rules the diagnosis out).⁷

Likelihood ratios

Another method for describing the diagnostic accuracy of tests is likelihood ratios. The positive likelihood ratio is the probability of a positive test result in patients with the disease divided by the probability of a positive test result in patients without the disease, or sensitivity/(1-specificity).⁸ In the example above, the positive likelihood ratio is 90%/40%, or 2.25. A test is moderately good at ruling in disease if the positive likelihood ratio is >2 and very good at ruling in disease if the positive likelihood ratio is >10, so the urine dipstick test is moderately good at ruling in the diagnosis.

Conversely, the negative likelihood ratio is the probability of a negative test result in patients with the disease divided by the probability of a negative test result in patients without the disease, or (1-sensitivity)/specificity. The negative likelihood ratio in the example above is 10%/60%, or 0.17. A test is moderately good at ruling out disease if the negative likelihood ratio is <0.5 and is very good at ruling out disease if the negative likelihood ratio is <0.1, so this test is moderately good at ruling out the diagnosis.

Likelihood ratios can be used to convert pretest to post-test probabilities using the formula:

Post-test odds of disease = pretest odds of disease x likelihood ratio
 where the odds of disease are (probability of disease/1-probability of disease).

Likelihood ratios can be used to describe more than two test outcomes. For example, we can calculate the

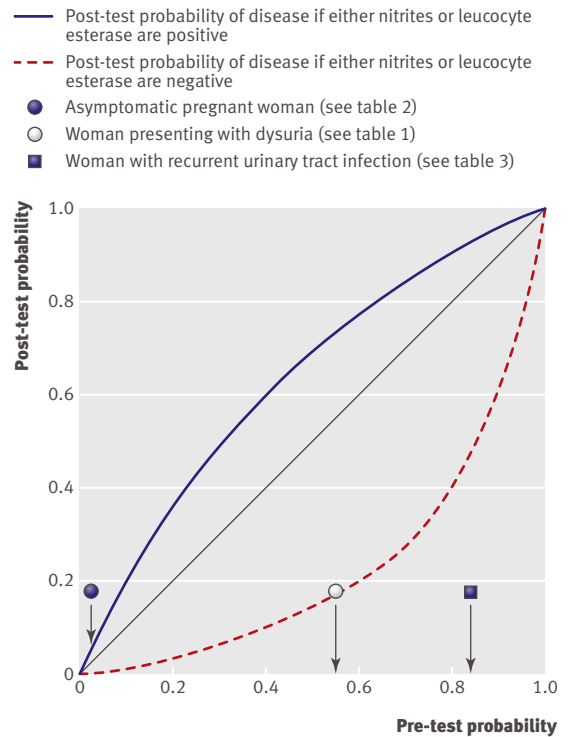


Fig 2 | Pretest post-test graph of urine dipstick results for detecting urinary tract infection

likelihood ratio for both urine dipstick tests being positive, for each test being positive and the other negative, and for both tests being negative. The combination of likelihood ratios is the basis of clinical prediction rules, as described earlier in this series.⁹

We need to be able to recognise the potential for diagnostic test results to be wrong, particularly the probability of false positive test results in low prevalence settings and false negative test results when the pretest probability of disease is high. We also need to have better evidence about the diagnostic accuracy of tests, and particularly in the clinical setting in which they are going to be used. The publication in the Cochrane Library of systematic reviews of the accuracy of diagnostic tests will allow more of this information to be available to general practitioners.

Contributors: JD is sole author and guarantor.

Competing interests: None declared.

Provenance and peer review: Commissioned; externally peer reviewed.

- 1 Deville WL, Yzermans JC, van Duijn NP, Bezemer PD, van der Windt DA, Bouter LM. The urine dipstick test useful to rule out infections. A meta-analysis of the accuracy. *BMC Urol* 2004;4:4.
- 2 Bachman JW, Heise RH, Naessens JM, Timmerman MG. A study of various tests to detect asymptomatic urinary tract infections in an obstetric population. *JAMA* 1993;270:1971-4.
- 3 Gupta K, Hooton TM, Roberts PL, Stamm WE. Patient-initiated treatment of uncomplicated recurrent urinary tract infections in young women. *Ann Intern Med* 2001;135:9-16.
- 4 Heneghan C, Glasziou P, Thompson M, Rose P, Balla J, Lasserson D, Scott C, et al. Diagnostic strategies used in primary care. *BMJ* 2009;338:b946.
- 5 Jelinek M, Barraclough K. Chest pain. *BMJ* 2009;339:b4117.
- 6 Trevena L. Informing patients about screening. *Australian Doctor* 2006 June 2:25-32. www.health.usyd.edu.au/shdg/docs/AD_HTT_025_032_JUN02_06.pdf
- 7 Straus SE, Richardson WS, Glasziou P, Haynes RB. *Evidence-based medicine: how to practice and teach EBM*. 3rd ed. Edinburgh; New York: Elsevier/Churchill Livingstone, 2005: 77.
- 8 Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ* 2004;329:168-9.
- 9 Gavin Falk G, Fahey T. Clinical prediction rules. *BMJ* 2009;339:b2899.