Information in practice

Communicating accuracy of tests to general practitioners: a controlled study

Johann Steurer, Joachim E Fischer, Lucas M Bachmann, Michael Koller, Gerben ter Riet

Horten-Zentrum für praxisorientierte Forschung und Wissenstransfer, Universitätsspital Zürich, Bolleystrasse 40, Postfach Nord, CH-8091 Zurich, Switzerland
Johann Steurer *director*
Joachim E Fischer *senior research fellow*
Lucas M Bachmann *research fellow*
Michael Koller *research fellow*
Gerben ter Riet *senior research fellow*

Correspondence to: J Steurer johann.steurer@dim.usz.ch

BMJ 2002;324:824–6

Abstract

Objective To assess the extent to which different forms of summarising diagnostic test information influence general practitioners' ability to estimate disease probabilities.
Design Controlled questionnaire study.
Setting Three Swiss conferences in continuous medical education.
Participants 263 general practitioners.
Intervention Questionnaire with multiple choice questions about terms of test accuracy and a clinical vignette with the results of a diagnostic test described in three different ways (test result only, test result plus test sensitivity and specificity, test result plus the positive likelihood ratio presented in plain language).
Main outcome measures Doctors' knowledge and application of terms of test accuracy and estimation of disease probability in the clinical vignette.
Results The correct definitions for sensitivity and predictive value were chosen by 76% and 61% of the doctors respectively, but only 22% chose the correct answer for the post-test probability of a positive screening test. In the clinical vignette doctors given the test result only overestimated its diagnostic value (median attributed likelihood ratio (aLR) = 9.0, against 2.54 reported in the literature). Providing the scan's sensitivity and specificity reduced the overestimation (median aLR = 6.0) but to a lesser extent than simple wording of the likelihood ratio (median aLR = 3.0).
Conclusion Most general practitioners recognised the correct definitions for sensitivity and positive predictive value but did not apply them correctly. Conveying test accuracy information in simple, non-technical language improved their ability to estimate disease probabilities accurately.

Introduction

General practitioners are expected to be proficient in integrating diagnostic information from history taking, physical examination, and other diagnostic procedures. Effective therapeutic action rests on the correct interpretation of such data.

Usually, the accuracy of tests is reported in terms of their sensitivity, specificity, and predictive values. Where the prevalence of disease is low most doctors grossly overestimate the probability of disease in patients with a positive result from a screening test.[1] They seem to confuse the sensitivity of the test with its positive predictive value.[2 3] Less is known about doctors' understanding of test accuracy data in settings with a higher prevalence of disease. We therefore presented a structured questionnaire with a vignette of a clinical problem to general practitioners. Our primary aim was to assess the extent to which different forms of presenting test accuracy information affected the doctors' estimates of the probability of disease.

The study questionnaire appears on bmj.com

Participants and methods

Participants

We recruited general practitioners attending three conferences on continuing medical education in Switzerland. On average, the participating doctors had more than 10 years of professional experience. Although general practitioners do not formally act as gatekeepers in Switzerland, they are usually the first healthcare providers to be contacted when new medical problems arise.

Questionnaire

The questionnaire, which was developed and piloted in a different group of 45 doctors, consisted of two parts (see bmj.com). The first part consisted of multiple choice questions that asked for the definition of the terms "sensitivity" and "positive predictive value" (from a choice of four possibilities) and for the probability of disease when a screening test with a sensitivity and specificity of 95% returns a positive result in a population with a disease prevalence of 1% (from the choices < 25%, about 50%, nearly 100%, and "Don't know").

The second part evaluated the participants' ability to apply these terms to a clinical vignette. Firstly, they were asked to estimate the probability of endometrial cancer in a 65 year old woman with abnormal uterine bleeding (for simplicity, the prevalence of endometrial cancer in all women with abnormal uterine bleeding was given as 10%). Secondly, participants were asked to estimate the disease probability given the result of a transvaginal ultrasound scan. The test result was provided in three different versions: "Transvaginal ultrasound showed a pathological result compatible with cancer"; "Transvaginal ultrasound showed a pathological result compatible with cancer. The sensitivity of this test is 80%, its specificity is 60%"; or "Transvaginal ultrasound showed a pathological result compatible with cancer. A positive result is obtained twice as frequently in women with an endometrial cancer than in women without this disease." The third version was intended to present the positive likelihood ratio of 2 in non-technical language.

Data collection
Participants received a questionnaire presenting the test result in one of three versions, the allocation being concealed. The questionnaires were handed out before a lecture on evidence based medicine, and the participants were given 10 minutes to complete them. If any of the participants attended more than one of the conferences, we included only their first questionnaire in the analysis.

Data analysis
For the three multiple choice questions, we calculated the proportions of doctors (plus 95% confidence intervals) who chose the correct answer.

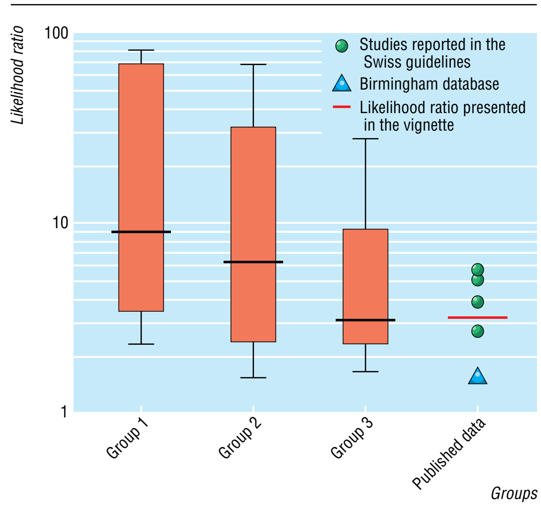
For the second part of the questionnaire, we derived the implicitly attributed likelihood ratios (aLR) by comparing the given probability of disease (10%) with the participants' estimate of probability after being given information on patient's age and result of ultrasound scan. We used the equation aLR = post-test odds/pretest odds, where odds = probability/(1 – probability). Likewise, we calculated the likelihood ratio attributed to the positive ultrasound result (probability estimate based on age and test information compared with probability estimate based on age alone). To avoid needless missing values, we converted eight post-test probability estimates of 100% to 99.999%. We made an overall comparison between the three versions of the test information using the Kruskal-Wallis test using SAS statistical software (version 8.1, SAS, Cary, NC, USA). We tested other differences using the Mann-Whitney rank sum test.

To obtain an empirical likelihood ratio of endometrial cancer in a 65 year old woman with abnormal uterine bleeding, we used data from 248 consecutive outpatients presenting with abnormal uterine bleeding at the Birmingham Women's Hospital rapid access ambulatory diagnostic clinic (RAAD) between November 1996 and December 1997.[4] This database contains information on patients' age and uses endometrial biopsy as the definitive test for cancer. In this database women aged 60-70 with abnormal bleeding are 3.1 times more likely to have endometrial cancer than younger women. The sensitivities and specificities for transvaginal ultrasound that we provided approximated to the median values given in the literature,[5 6] with rounding to simplify calculation.

Results

Of the 263 eligible general practitioners, between 251 and 261 answered the three multiple choice questions. Of those answering the question, 76% (95% confidence interval 70% to 81%) chose the correct definition of "sensitivity," and 61% (54% to 67%) chose the correct definition of "positive predictive value." However, only 22% (17% to 27%) chose the correct option of "< 25%" for the probability of disease in the example of a positive result from a screening test (sensitivity and specificity 95%, disease prevalence 1%), while 56% (49% to 62%) selected a probability of "close to 100%."

In the clinical vignette, providing the information that the woman was aged 65 led 48% of participants to



Probability of presence of disease attributed by general practitioners to a positive result from transvaginal ultrasound scanning in women with abnormal uterine bleeding. Group 1 received no information on test accuracy, group 2 was provided with test sensitivity (80%) and specificity (60%), group 3 was presented with the positive likelihood ratio of 2 in non-technical language. (Box and whisker plots show medians, 25th and 75th centiles, and ranges)

change their estimates of the probability of disease. The figure shows the effect of presenting the results of the ultrasound scan in three different ways, with the three groups producing significantly different attributed likelihood ratios (P = 0.0013). The 92 participants who were not given any information on the test's accuracy seemed to grossly overestimate the probability of endometrial cancer (median attributed likelihood ratio (aLR) 9; interquartile range 3.25-68.5; P = 0.0006 com-

Terms used to describe the accuracy of a diagnostic test

Sensitivity—The number of people with a positive result both on the test under study and on the reference test divided by the number of people with a positive result on the reference test[1] (also called the true positive rate)

Specificity—The number of people with a negative result both on the test under study and on the reference test divided by the number of people with a negative result on the reference test (also called the true negative rate)

Positive likelihood ratio for a dichotomous test—The percentage of patients who have a positive test result among those with the target disease divided by the percentage of patients who have a positive test result among those without the target disease[2 3]

Likelihood ratio for a positive test result (general definition)
The percentage of patients who have test result *t* among those with the target disease divided by the percentage of patients who have test result *t* among those without the target disease

- Note that the numerator corresponds to the sensitivity and that the denominator corresponds to (1 – specificity)
- Likelihood ratio of positive test result = [a/(a + c + e)]/[b/(b + d + f)]
- Likelihood ratio of intermediate test result = [c/(a + c + e)]/[d/(b + d + f)]
- Likelihood ratio of positive test result = [e/(a + c + e)]/[f/(b + d + f)]

	Reference test positive	Reference test negative
Test positive	a	b
Test intermediate	c	d
Test negative	e	f
Total	a+c+e	b+d+e

What is already known on this topic

Many doctors confuse the sensitivity of clinical tests and their positive predictive value

Doctors tend to overestimate information derived from such tests and underestimate information from a patient's clinical history

Most primary research on diagnostic accuracy is reported using sensitivity and specificity or likelihood ratios

What this study adds

In a cohort of experienced Swiss general practitioners most were unable to interpret correctly numerical information on the diagnostic accuracy of a screening test

When presented with a positive result alone they grossly overestimated its value

Adding information on the test's sensitivity and specificity moderated these overestimates, and expressing the same numerical information as a positive likelihood ratio in simple, non-technical language brought the estimates still closer to their true values

pared with the other two groups combined). The 92 doctors provided with the sensitivity and specificity of the scan had lower estimates of the likelihood of disease (aLR = 6; 2.3-22.1; P = 0.019 compared with group 1). The 79 doctors given the test accuracy in plain language had an attributed likelihood ratio still closer to the literature based ones (aLR = 3.0; 2.25-9; P = 0.228 compared with group 2).

Discussion

In this study we evaluated general practitioners' knowledge of terms commonly used to describe a test's accuracy. Although most identified the correct definitions of sensitivity and positive predictive value, only 22% correctly estimated the (low) probability of disease after a positive test result when told of the disease prevalence in the population and the test's sensitivity and specificity. In the clinical vignette the participants underestimated the diagnostic value of the patient's age. Those who were not provided with data on test accuracy grossly overestimated the diagnostic accuracy of a positive transvaginal ultrasound result compared with data from a recently published systematic review[6] and with data provided in the Swiss guidelines on the management of women with postmenopausal bleeding.[5] Presenting test accuracy as the positive likelihood ratio expressed in plain language seemed to be more effective for eliciting correct estimates of disease probability than presenting it as sensitivity and specificity.

Our findings might overestimate the average general practitioner's performance with our questionnaire because e selected doctors attending educational sessions on evidence based medicine. Their responses might have been affected by their prior knowledge of measures of test accuracy and their presentation.

Implications of results

In clinical practice not all wrong estimations of disease probability are of equal importance. Two numerically different estimates may not be clinically different if they lead to the same clinical decision. However, it is difficult to be specific about these thresholds for action as they may depend on many subjective factors.

Despite a long tradition of reporting test accuracy in terms of sensitivity and specificity, only a minority of our participants could correctly apply this information. This difficulty in performing the required calculations probably explains their underuse in general practice.[7] Rather than blaming doctors for this lack of aptitude, authors of diagnostic test data should reconsider the way they communicate their research data. We showed that presentation of a positive likelihood ratio in simple, non-technical wording improved the participants' ability to estimate accurately the probability of disease. Other ways to present diagnostic data—such as disease probability estimates,[8] prediction rules,[9] or decision trees[10 11]—should be explored.

Our study raises the question to what extent overestimation of the diagnostic value of screening procedure contributes to the steadily increasing use of laboratory and imaging tests.[12] One reason for underestimating the diagnostic value of information from a patient's history may be the lack of well designed studies tackling this issue.[13]

This study gives no insight into what conclusions general practitioners would draw from a positive ultrasound result in real practice. However, if other considerations do not correct for the observed overestimation of the accuracy of the test, there might be adverse consequences for doctor-patient communication and further action.

We thank Wim Verstappen (Maastricht University) for sharing his questionnaire and K S Khan and J K Gupta for providing the outpatient database.

Contributors: JS, GtR, and LMB initiated the study. GtR and JEF contributed to study design and data analysis. All authors helped to interpret data and write the paper. JS and GtR are guarantors for the study.

Funding: None declared.

Competing interests: None declared.

1 Eddy DM. Probabilistic reasoning in clinical medicine: problems and opportunities. In: Kahnemann D, Slovic P, Tversky A, eds. *Judgement under uncertainty: heuristics and biases.* Cambridge: Cambridge University Press, 1982:249-67.
2 Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. Communicating statistical information. *Science* 2000;290:2261.
3 Kahnemann D, Tversky A. *Choices, values and frames.* Cambridge: Cambridge University Press, Russel Sage Foundation, 2000.
4 Clark TJ, Khan KS, Bakour SH, Gupta JK. Evaluation of outpatient hysteroscopy and ultrasonography in the diagnosis of endometrial disease. *Obstet Gynecol* (in press).
5 Bronz L, Dreher E, Almendral A, Studer A, Haller U. Abklärung von postmenopausalen Blutungen (PMBP). *Schweizerische Aerztezeitung* 2000;81:1635-46.
6 Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;323:157-62.
7 Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med* 1998;104:374-80.
8 Miettinen OS, Henschke CI, Yankelevitz DF. Evaluation of diagnostic imaging tests: diagnostic probability estimation. *J Clin Epidemiol* 1998;51:1293-8.
9 Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. *Ann Emerg Med* 1992;21:384-90.
10 Buntinx F, Truyen J, Embrechts P, Moreels G, Peeters R. Evaluating patients with chest pain using classification and regression trees. *Fam Pract* 1992;9:149-53.
11 Lieu SA, Quesenberry CP, Sorel ME, Mendoza GR, Leong AB. Computer-based models to identify high risk children with asthma. *Am J Respir Crit Care Med* 1998;157:1173-80.
12 Van Walraven C, Naylor D. Do we know what inappropriate laboratory utilization is? A systematic review of laboratory clinical audits. *JAMA* 1998;280:550-8.
13 McAlister FA, Straus SE, Sackett DL. Why we need large, simple studies of the clinical examination: the problem and a proposed solution. *Lancet* 1999;354:1721-4.

(Accepted 22 January 2002)