How many ROC curves fit into one general practitioner?

The paradox between medical decision making and daily general practice

G.J. DINANT
J. DE MAESENEER
A. DERESE

Dinant GJ, De Maeseneer J. Derese A. How many ROC curves fit into one general practitioner? The paradox between medical decision making and daily general practice. Huisarts Wet 1993; 36(Suppl): 58-61.

Abstract Regarding the analysis of the daily general practitioners' activities, an increasing use of terminology from the field of clinical epidemiology and medical decision making is observed. Unfortunately, many general practitioners are not familiar with the terms prevalence, sensitivity and specifity of diagnostic tests, and ROC-curves, and this may deter them from reading articles on the subject. However, since the daily general practitioners' activities are fundamented on and improved by clinical epidemiology and medical decision making, these fields deserve a further introduction in general practitice. The importance of this is proved by the results from registration research and the implementation of results from diagnostic studies in the Standards of the Dutch College of General Practitioners.

University of Limburg, Department of General Practice, PO Box 616, 6200 MD Maastricht, The Netherlands: G.J. Dinant, MD, PhD.
University of Ghent, Department of General Practice and Primary Health Care, Belgium: Prof. J. De Maeseneer, MD, PhD; Prof. A. Derese, MD, PhD.
Correspondence: G.J. Dinant

Introduction

In daily practice, general practitioners constantly make decisions in their approach to problems presented to them. In terms of decision-making, this is described as converting pretest odds into post-test odds via anamnesis and physical and further examination (blood tests, X-rays, etcetera). The process, of which general practitioners are often unaware, is based on various factors, including contextual information. Regarding the analysis of the process, an increasing use of terminology from the field of clinical epidemiology and medical decision-making is observed. Many general practitioners are not familiar with the terminology, however, and this may deter them from reading articles on the subject. In this paper, the following questions will be answered:

- To what extent do general practitioners avail themselves of clinical epidemiology and medical decision-making in daily practice?
- To what extent does clinical-epidemiological and medical decision-making analysis of the performance of general practitioners contribute to better patient care?

Through registration research, an understanding may be gained of the performance 'in vivo' of general practitioners. Available data indicate, besides a clearly identifiable interdoctor variation, an important degree of conformity in the recorded actions of general practitioners.[1] This conformity in itself presents an argument for the proposition that general practitioners may follow recognizable decision-making patterns. Studies carried out with simulated patients have even shown that these patterns are followed more distinctly 'in vivo' than general practitioners indicate 'in vitro'.[2]

Generating hypotheses

The diagnostic strategy that is most widely used in general practice is the hypothesis-deductive strategy: on the basis of a limited amount of data on the patient, the general practitioner rather quickly sets up a short list of possible diagnoses.[3] The list may be shortened on the basis of the patient's history and the results of a physical and possibly a further examination. The following factors play a role in drawing up the list: epidemiological data, information on the patient, encyclopedic information, doctor-bound factors, signals from the patient, the seriousness of the condition, and a possible necessity to intervene quickly.[4] It is not always possible to represent the degree in which these factors contribute to the pretest odds (estimated by the general practitioner on a certain condition) in a mathematical formula. This is partially due to the fact the subject has not been studied (yet). Still, general practitioners make active use of the fact that women have a ten times greater chance of suffering from a thyroid-gland condition than men, that a patient with hypercholesterolemia has a greater risk of developing cardiovascular disease than a patient without this risk factor, and that rectal bleeding in a patient with a family history of coloncarcinoma implies an increase in the pretest odds on developing the disease. The examples show that clinical epidemiology contributes significantly to the medical decision to examine the thyroid-gland functions, reducing the cholesterol level (with medication), and requesting an endoscopic examination of the intestine respectively.

One of the most important contributions of episode-oriented registration (Transition project), is the linkage of the 'reason for encounter' at the beginning of the episode with the final diagnosis. In this way, an understanding may be gained of the true pretest odds on disease given the complaint for which the patient consults the general practitioner. The time perspective provides important information as well. It was determined that 86 per cent of the new episodes 'tiredness' and 'malaise' lasted for less than 4 weeks. In the absence of positive anamnestic and clinical observations, this conclusion justifies the decision of a general practitioner to ask a patient with complaints of tiredness to come back

in 4 weeks (and not earlier).[1] With this example, the relevance of clinical-epidemiological data to the daily practice of a general practitioner is illustrated again.

The influence of patient information already known to the general practitioner in the decision-making process was investigated recently by *Hobus et al.*[5] The researchers proved that experienced general practitioners who availed themselves of foreknowledge and contextual information, drew up better diagnostic hypotheses than recently qualified doctors who did not make use of this information.

The assessment of the seriousness of the disorder (a bad feeling about the complaint or not), appears to have a significant predictive value. *Buntinx* found that in 315 patients with thoracic pain, general practitioners had in 28 per cent of the cases the feeling that the complaint was a serious problem.[6] In 31 per cent of these cases an important pathology was found, whereas the complaints of only 6 of the remaining 228 patients had erroneously been assessed as not serious. It is clear that the predictive value of the 'bad or good feeling' was very high.

Anamnesis

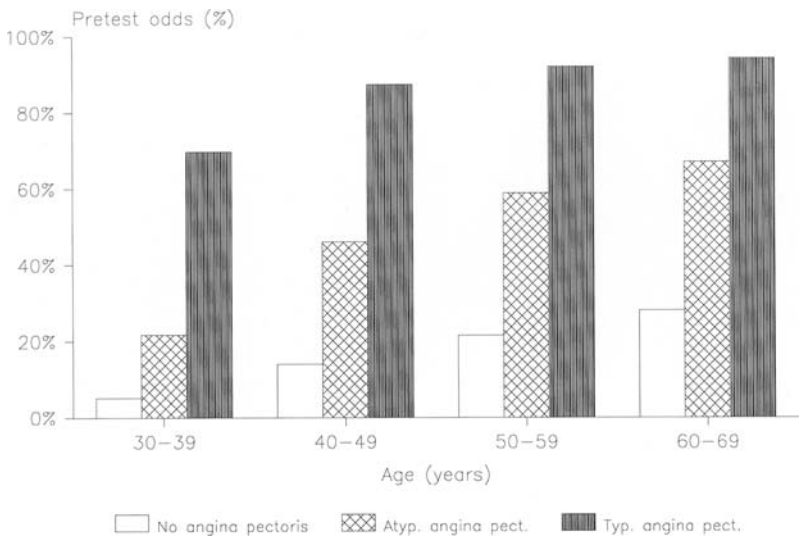
The great importance that general practitioners attach to anamnesis in diagnosing angina pectoris is based on a (perhaps unconscious) medical decision-making consideration. *Diamond & Forrester*[7] distinguish between three types of 'chest pain'; classification is done on the basis of three central questions:

- is the pain retrosternal?
- are the complaints induced by exertion?
- does the pain subside after 2-15 minutes rest and/or the use of nitrates?

If one of the questions is answered positively, the chest pain is classified as 'nonanginal chest pain'. If positive answers are given to two questions, the complaint is classified as 'atypical angina pectoris'. In case of three positive answers, the complaint is classified as 'typical angina pectoris'. *Figure 1* shows the risk of developing coronary ischaemia of men in various age groups, differentiating in one, two or three positive answers. We may read off that in a 35 year old male patient with nonanginal chest pain it is questionable whether the slim chance (5,2 per cent) of developing coronary heart disease justifies further examination. After all, was this chance not already there due to the patient's age and sex?

It is a well-known fact that in diagnosing acute appendicitis in children, the absence of the symptoms fever and vomiting has a greater predictive value in relation to not suffering from the disease than the absence of leucocytosis.[8,9] In case of an a priori chance of 10 per cent, we find a negative predictive value of 97 per cent versus 81 per cent. In other words, a child with acute abdominal pain that does not have a fever and does not vomit, is less likely to suffer from acute appendicitis than a child with a normal leucocyte count.

Figure 1 *Pretest odds for a coronary disease in men, related to age*[7]



Physical examination

General practitioners generally attach greater value to anamnesis than to results of physical examination. Given the low sensitivity and great specificity, it is questionable whether auscultating the bronchi contributes to diagnosing bronchus obstruction.[10] If a general practitioner considers this diagnosis, it is advisable to carry out a peak-flow measurement in addition to the anamnesis. This advice may also be observed in the Standard of the Dutch College of General Practitioners (NHG) for diagnosing COPD in adults.

In diagnosing influenza, general practitioners mainly focus on the complex of symptoms (fever, myalgia, headache, coughing and acute onset), besides establishing the absence of abnormalities in physical examination.[11] In addition, the presence of an influenza epidemia is also regarded as a relevant factor. This example is illustrative for the situation in which the positive predictive value of the complex of symptoms increases with a (temporary) rise in the incidence of the disease.

Further examination

Many publications have been issued on medical decision-making aspects in relation to further examination. Well-known

in this context is the two-centuries-old theorum of Bayes:

$$PV+ = \frac{se * po}{(se * po) + [(100\% - sp) * (100\% - po)]}$$

$$PV- = \frac{sp * (100\% - po)}{[(100\% - se) * po] + [sp * (100\% - po)]}$$

PV+ predictive value for a positive test result
PV– predictive value for a negative test result
se sensitivity; sp specificity; po pretest odds

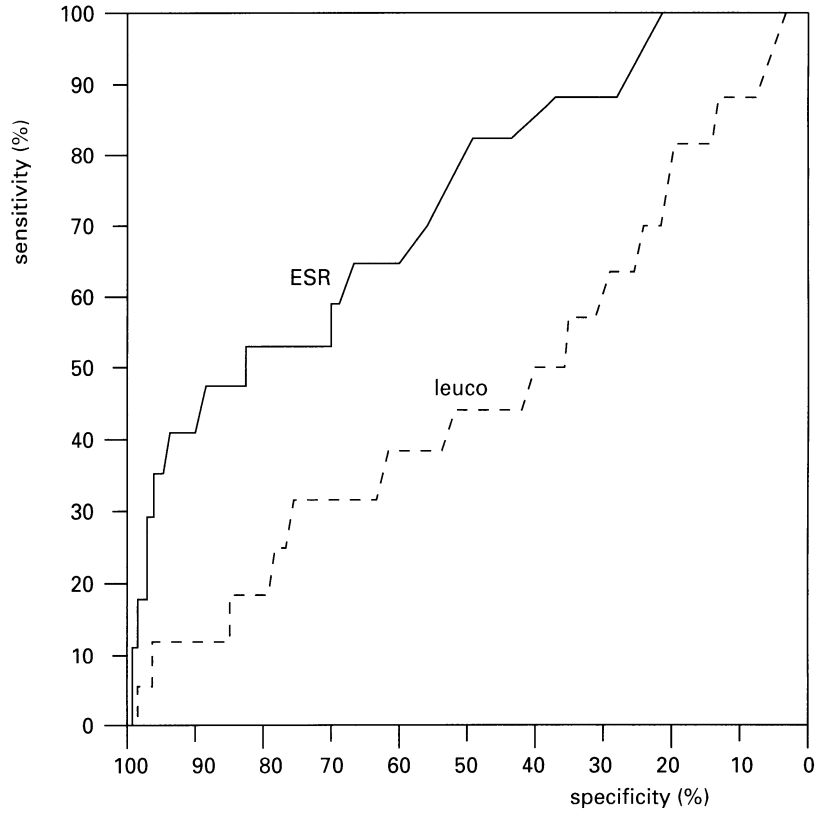
If the presence (or absence) of a certain disease is suspected in a patient, the formula of Bayes facilitates the assessment of a post-test odds given a pretest odds and the test results. The post-test odds corresponds with the predictive value of the test result, and can be calculated if the sensitivity and specificity of the test are known.

Although general practitioners do not immediately grab their calculator in interpreting laboratory results, sensitivity and specificity do provide them with important information. In the first instance, general practitioners are mainly interested in the following question: 'With what certainty can pathology be excluded?' In answering this question, a high-sensitivity test may be useful: a negative test result implies a clearly increased chance of the absence of pathology. In the second phase, a goal-oriented high-specificity examination may be performed to prove the presence of a certain disease. A practical example is diagnosing a thyroid-gland disorder. The sensitivity of the sensitive TSH-test is a 100 per cent and the specificity 93 per cent, which means that a normal TSH-test can be used for excluding thyroid-gland suffering.[12] On the other hand, an abnormal TSH justifies the determination of the free T4 in order to be able to prove the suspected thyroid-gland suffering. This example illustrates how medical decision-making and cost/benefit go hand in hand: needless determination of the free T4 and needless follow-up diagnostics of a false positive or a false negative free T4 are prevented.

Knowledge of test characteristics has stopped general practitioners from requesting both creatinine and urea for the purpose of evaluating renal function.[13] Medical decision-making analysis of the value of doing preoperative laboratory research in nonsymptomatic patients, has provided a reason for deciding not to do so in patients under 40 years of age.[14]

Sensitivity and specificity are essential in comparing the discriminating ability of various tests. Drawing these comparisons is generally performed by means of Receiver Operating Characteristic (ROC) curves. An ROC-curve gives a graphical presentation of the relation between the sensitivity, specificity and reference values of a test. ROC-curves were used, for example, for proving that there is no point in using the leucocyte count as an alternative for ESR in daily practice (see *figure 2*),[15] independent of the selected reference-values. Figure 2 shows that the sensitivity and specificity of the leucocyte count are in all cases lower than those of ESR. In itself this is not a spectacular observation; every general practitioner would have sensed as such. In practice, however, many patients are referred to the laboratory for a leucocyte count, or, even worse, for a complete blood picture in an attempt of the general practitioner to compensate for the, in his eyes, insufficient discriminating ability of ESR. The consequences of this approach are clear: burden on the patient, laboratory costs, and, last but not least, delegation of laboratory activities to the hospital (once again).

Figure 2 *ROC curves illustrating the relation between sensitivity, specificity and reference values of the erythrocyte sedimentation rate (ESR) and leucocyte count (leuco), with respect to 'pathology' (inflammatory diseases and malignancies), using 'no patholgy' as a control group, in patients indicated for the determination of ESR*[15]



ROC-curves also turn out to be useful in calculating reference values, at least in case of tests with more than two possible results (e.g. tests for determining glucose, cholesterol or haemoglobine levels). Specifying cut-off points is relevant to daily practice, as, for instance, the Standards of the Dutch College of General Practitioners (NHG) on diabetes mellitus and hypercholesterolemia show. Medicinal and nonmedicinal advices are based on the use of reference intervals. The Standards are founded, in a nearly unrecognizable way, on Bayes' theorum and, fortunately, not on the results of tasting urine samples and holding up blood samples against the light that were common in Bayes' younger years.

In regard to the above-mentioned ESR-study, it is advised to use a larger reference-value range in women than in men, without correcting for the age of the patient.[16] This advice, which is relevant to daily practice, was not exclusively drawn up on the basis of ROC-curves. Complicated statistical analyses, such as logistic regression, were applied as well. This illustrates the paradox between the complicatedness of the statistical analyses, which many regard as difficult to follow, and the simplicity of the resultant advice.

Epilogue

The daily general practitioners' activities form a complex unity. Case history, clinical findings, and definitions of disease are considered against the light of therapeutical alternatives, opportunities for further examination, financial and organisational limitations, and opinions of colleagues and patients. The above-described examples show the legitimacy of a clinical-epidemiological and medical decision-making approach to anamnesis and physical and further examination, and to generating hypotheses. Adequately fitting medical decision-making insights in daily patient care is not at right angles with a context-directed integral approach. A better assessment of objective interventions creates more room for the most individual expression of the most individual emotion, and moreover contributes to affordable care!

The imagery of a house that stands or falls with the underground foundations it was built on also applies to the not always visible role that clinical epidemiology and medical decision-making play in a general practitioner's daily practice. A courageous house owner who inspects the catacombs of his property via the crawl space may be compared with a researcher who follows the sometimes twisting lines of Bayes. The imagery can be carried even further. Like the house owner uses a torch underground to see more, the researcher avails himself of ROC-curves to shed light on his observations.

References

1 Lamberts H, Brouwer H, Mohrs J. Reason for encounter-, episode- and process-oriented standard output from the Transition Project. Amsterdam: Department of General Practice, University of Amsterdam, 1991.

2 Rethans JJ, Van Boven CPA. Simulated patients in general practice: a different look at the consultation. Br Med J 1987; 294: 809-12.

3 Sacket DL, Haynes RB, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. Boston: Little, Brown and Company, 1989.

4 De Maeseneer J. Diagnostische strategieën in de huisartsgeneeskunde. Belgisch Tijdschr Geneeskd 1991; 47: 1075-83 [Summary in English].

5 Hobus PPM. De context van de klacht als diagnosticum. Huisarts Wet 1988; 31: 261-3 [Synopsis in English].

6 Buntinx F. Pijn op de borst. Leuven: Academisch Centrum Huisartsgeneeskunde, 1988.

7 Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary artery disease. N Eng J Med 1979; 300: 1350-4.

8 O'Shea JS, et al. Diagnosing appendicitis in children with acute abdominal pain. Pediatr Emerg Care 1988; 3: 172-6.

9 Marchand A, Van Leute F, Galen F. The assessment of laboratory tests in the diagnosis of acute appendicitis. Am J Clin Pathol 1983; 80: 369-74.

10 Muris JWM. Auscultatie van de longen in de huisartspraktijk. Huisarts Wet 1990; 33: 258-62 [Abstract in English].

11 Fleming DM, Ayres JG. Diagnosis and patterns of incidence of influenza, influenza-like illness and the common cold in general practice. J R Coll Gen Pract 1988; 38: 159-62.

12 Tal A. Screening for thyroid disease [Letter]. Ann Intern Med 1990; 113: 896.

13 Van Rijn HJM, Hené RJ, Struyvenberg A. Het mes in de ureumbepaling. Ned Tijdschr Geneeskd 1990; 134: 1177-9.

14 Narr BJ, Hansen TR, Warner MA. Preoperative laboratory screening in healthy Mayo patients: cost-effective elimination of tests and unchanged outcomes. Mayo Clinic Proceedings 1991; 66: 155-9.

15 Dinant GJ, Knottnerus JA, Van Wersch JWJ. Leucocyte count as an alternative to ESR in general practice? Scand J Prim Health Care 1991; 9: 281-4.

16 Dinant GJ, Knottnerus JA, Van Wersch JWJ. Discrimating ability of the erythrocyte sedimentation rate: a prospective study in general practice. Br J Gen Pract 1991; 41: 365-70. ■

Samenvatting

In toenemende mate wordt bij de analyse van het handelen van huisartsen in de dagelijkse praktijk gebruik gemaakt van kennis uit de klinische epidemiologie en medische besliskunde. Begrippen als prevalentie, sensitiviteit en specificiteit van diagnostische tests en ROC-curves zijn echter veelal onbekend en mede daardoor onbemind bij huisartsen. Toch verdienen zij verdere introductie, omdat de klinische epidemiologie en medische besliskunde het fundament zijn van het dagelijkse handelen van de huisarts en bovendien dit handelen verbeteren. Deze stelling wordt gestaafd door gegevens uit het Transitieproject en de implementatie van resultaten uit wetenschappelijk onderzoek in onder andere NHG-Standaarden.