Interpretation of diagnostic data: an unexplored field in general practice

J.A. KNOTTNERUS, MD
General Practitioner, Maastricht, The Netherlands

SUMMARY. In interpreting diagnostic information the differences between morbidity patterns in general practice and those in hospitals must be taken into consideration. This article demonstrates the importance of prevalence for predictive values of complaints, symptoms and test results. When the general practitioner refers patients to the specialist, these values may change, and it is also possible that associations between symptoms and diagnoses are distorted by selection bias. Moreover, attention must be paid to the differences in clinical stages encountered in general practice and specialist practice. It is concluded that a large part of the 'diagnostic field' of general practice has still to be discovered and developed.

Introduction

ATTENTION has been drawn to the way in which the prevalence of diseases can influence the interpretation of diagnostic data.[1,2] General practitioners and specialists deal with different patterns of morbidity and therefore they should interpret diagnostic data in different ways.

Basic concepts

Let us consider a group of women aged between 20 and 30 years in whom the general practitioner finds a palpable abnormality in the breast. After physical examination there are two possible conclusions: suspected breast cancer (positive result); no suspicion of breast cancer (negative result). Figure 1 shows the possible outcome for 1000 women; the figures are based on published data.[3,4] Of interest are the answers to the following questions:

1. What is the chance of finding breast cancer in women with suspect lumps?
2. What is the chance of *not* finding breast cancer in women with lumps which are *not* considered suspect?

The first question concerns the predictive value of a positive test result (PV+); the second question concerns the predictive value of a negative test result (PV−). The probability of a positive result in diseased persons is the *sensitivity* of the test for the disease under study. The probability of a negative test result in persons not having the disease is the *specificity* of the test. Figure 1 shows the values for sensitivity and specificity. It also reveals the prevalence: the percentage of people in the population under study who were actually suffering from the disease. In this example the prevalence was five per cent;[2] that is five per cent of the women with palpable lumps in the breast had cancer.

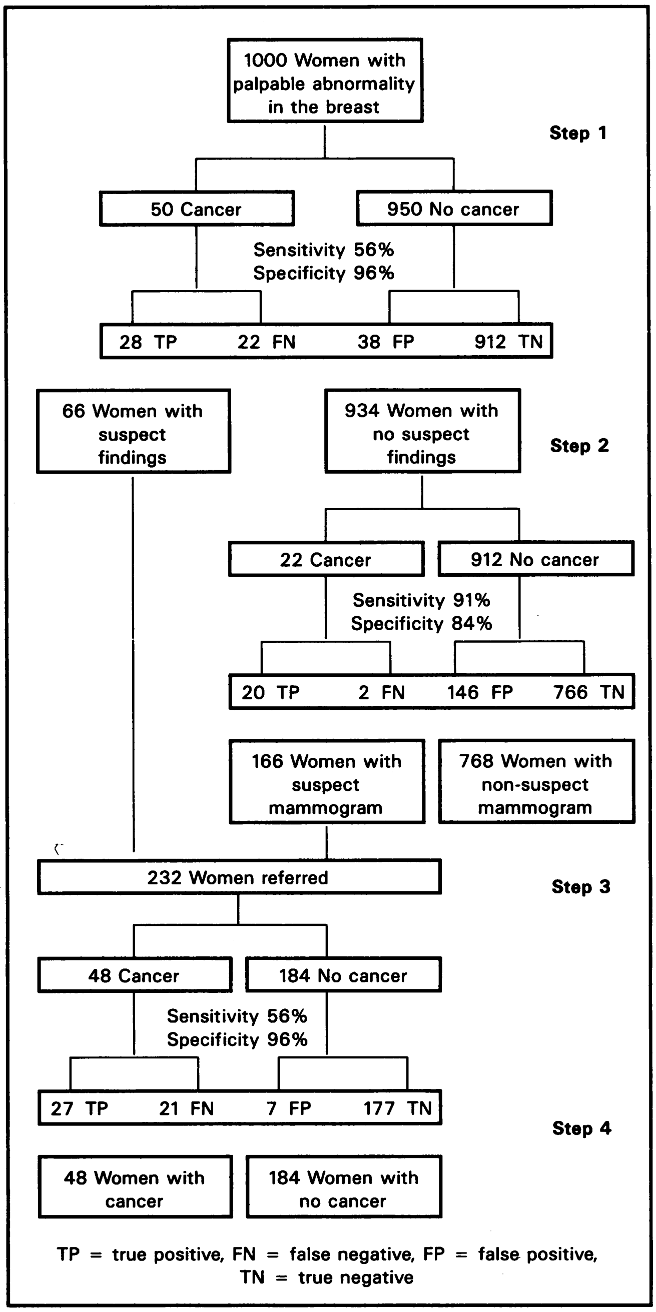


Figure 1. *The possible results of diagnostic procedure in young women with palpable abnormality in the breast. Step 1. The general practitioner carries out a physical examination and refers in the event of suspect findings. Step 2. The general practitioner carries out a mammography when no suspect abnormalities have been found by physical examination; refers if mammogram suspect. There are two false negatives, which have no further examination. Step 3. At referral surgeon carries out a physical examination. Prevalence = 48/232 = 21%. Step 4. Biopsy reveals the true situation.*

Prevalence of the disease and predictive values of tests

The policy of the general practitioner may be to refer those women having a suspect lump directly to a surgeon for biopsy. Other women having a palpable abnormality are sent for

© *Journal of the Royal College of General Practitioners,* 1985, 35, 270-274.

mammography; if this is positive, these women are also referred for biopsy. Assuming that mammography has a sensitivity of 91 per cent and a specificity of 84 per cent[4-7] and that biopsy eventually reveals the true situation, the prevalence of breast cancer can be calculated as 21 per cent after referral (Figure 1).

If the surgeon carries out palpation before biopsy and is no better or worse than the general practitioner in diagnostic skill, then this palpation has a positive predictive value (PV+) of 79 per cent and a negative predictive value (PV−) of 89 per cent. The surgeon will find relatively less false positives than the general practitioner, which could, wrongly, create the impression that the surgeon is more skilled at palpation than the general practitioner. The surgeon may make a false-negative diagnosis more often than the general practitioner but this is less conspicuous because the surgeon often performs a biopsy as well, and will often follow up the women who did not have a biopsy. The general practitioner may make relatively few false-negative diagnoses — patients with cancer who had a negative palpation and a negative mammogram. These patients will be reassured for a time but later these cases may be regarded as 'missed' carcinomas or as examples of doctor delay.

It is clear that the general practitioner could only have avoided all false-negative diagnoses by referring all women having a palpable abnormality for biopsy. In the example given, this would have meant that 768 extra biopsies would have been carried out in order to discover two more cases of cancer.

When choosing between diagnostic strategies, the prevalence of the disease is important. Knowledge of prevalence allows an appropriate choice of strategy in order to avoid missing carcinomas or carrying out too many biopsies on healthy women.

In general it should be noted that:

1. As prevalence increases, the predictive value of every positive test result increases, and as a consequence there will be fewer false positives (less needless biopsies) and more false negatives (more missed cases).
2. As prevalence decreases, the reverse applies and the predictive value of every positive test result decreases, and so there are more false positives (more needless biopsies) but fewer false negatives (less missed cases).

The positive association between the predictive value of a positive test result and the prevalence of the disease under study is not too difficult to understand. It is harder to understand the reverse — that the predictive value of a negative test result decreases as the prevalence increases. Another example may help to clarify this.

If a woman aged 25 years, without symptoms, is given a cervical smear and no abnormal cells are apparent, then the doctor will assume that this is not a case of cervical cancer. However, for a 40-year-old woman who complains of intermenstrual blood loss and bleeding during intercourse, the doctor will be less confident of the negative test result. The doctor may consider it to be a false-negative result and would not exclude the possibility of cancer. The predictive value of the normal smear test is considered to be low. The doctor is thus unconsciously applying the principles summarized above: he interprets the test result in the knowledge that cervical cancer seldom occurs in women of 25 years of age without symptoms (prevalence in such a population is very low) and in the knowledge that the disease occurs more often in women of 40 years of age who have the complaints described above (prevalence in such a population is higher).

The prevalence of serious diseases requiring treatment is higher in the population of patients visiting specialists than in the population of patients seen only by the general practitioner owing to preselection by the general practitioner. This means that the predictive value of an abnormal finding from an examination carried out by a specialist is higher than from the same examination carried out by a general practitioner. However, for serious diseases the predictive value of a normal finding is higher in general practice. The opposite is true for diseases seen frequently by the general practitioner but less often by the specialist, for example minor ailments and exanthemas in children.

The predictive value of a positive result from examination by a specialist is relatively high because of preselection by the general practitioner. Without such preselection there would be much unnecessary investigation and treatment. As examinations or tests by a general practitioner have a high negative predictive value, specialist assessment can achieve a high positive predictive value.

When the general practitioner — by means of history taking and physical examinations — makes a selection of those patients more likely to have certain diseases, he increases the predictive value of abnormal findings from specific examination (for example, electrocardiography) in his own practice. Table 1 shows the relationships between sensitivity, specificity, prevalence and predictive value summarized in the Formula of Bayes. Note that this formula can be directly concluded from the 2 x 2 multiplication table and that it does not supply any more information than this table. Figure 2 represents the relationships graphically for physical examination and mammography.

Table 1. The formula of Bayes and its derivation.

Test result	Patient has disease	Patient does not have disease
Positive	TP = sensitivity x prevalence	FP = (100% − specificity) x (100% − prevalence)
Negative	FN = (100% − sensitivity) x prevalence	TN = specificity x (100% − prevalence)

Derivation of formula

1. The predictive value of a positive test result $= \dfrac{TP}{TP + FP}$

$$= \frac{\text{sensitivity} \times \text{prevalence}}{(\text{sensitivity} \times \text{prevalence}) + (100\% - \text{specificity}) \times (100\% - \text{prevalence})}$$

2. The predictive value of a negative test result $= \dfrac{TN}{TN + FN}$

$$= \frac{\text{specificity} \times (100\% - \text{prevalence})}{\text{specificity} \times (100\% - \text{prevalence}) + (100\% - \text{sensitivity}) \times \text{prevalence}}$$

TP = true positive, FP = false positive, FN = false negative, TN = true negative.

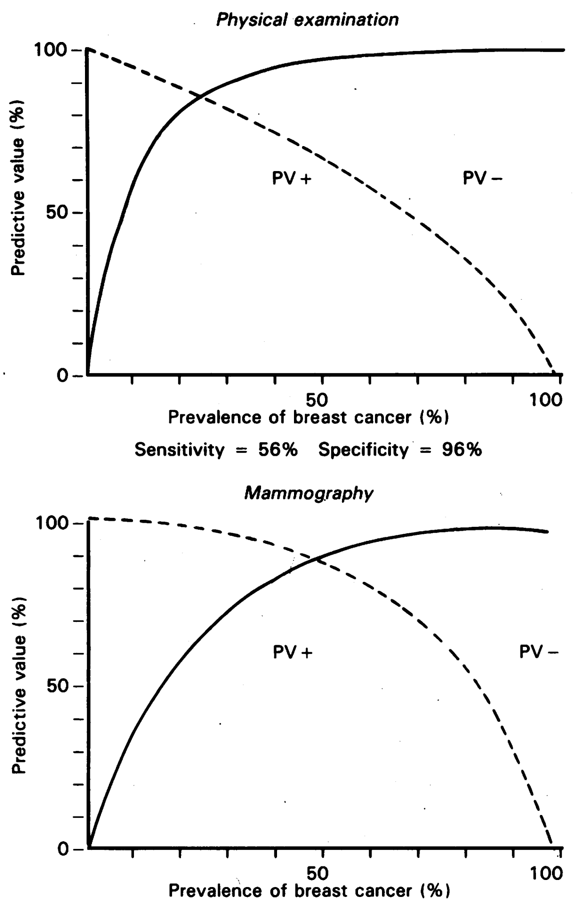


Figure 2. *Relationships between sensitivity and specificity and predictive values for physical examination and mammography of palpable abnormalities in the breast. PV+ = predictive value of suspect findings. PV− = predictive value of non-suspect findings.*

Meaning of complaints and symptoms

It is not only test results that have a predictive value but also symptoms and complaints. In this context they are a form of 'test'. In the first example the prevalence of breast cancer in a population of women with palpable abnormalities was discussed. The predictive value of a palpable abnormality for breast cancer can also be considered. This value will also be five per cent, providing a second test has not been used to differentiate between suspect and non-suspect abnormalities. The implications of this for the diagnostic interpretation of complaints and symptoms will be discussed by means of examples.

Acute appendicitis and pain in the lower abdomen

From morbidity data it can be concluded that acute appendicitis presents between four and five times a year on average in a practice of 2500 patients.[8-10] Clinical investigation shows that approximately 75 per cent of patients with acute appendicitis have a pain in the right lower abdomen.[11] Therefore, the sensitivity of this symptom is 75 per cent with regard to appendicitis. It can also be estimated, on the basis of general morbidity recording, that in a practice of 2500 patients approximately 40 persons per year report to their doctor with a pain in the right lower abdomen. The positive predictive value of the complaint can be calculated — three out of four persons with appendicitis (75 per cent) have a pain in the right lower abdomen, but a total of 40 persons report this complaint. Therefore, the positive predictive value of right lower abdomen pain for appendicitis is 3/40 = 7.5 per cent. Therefore, by itself, pain in the right lower abdomen seems to be of moderate value in predicting the presence of appendicitis.

Lump in the neck

Another example illustrates the way in which the different populations visiting specialists and general practitioners may result in the two types of doctor arriving at a different predictive value for the same symptom.

A female patient attends the doctor with a lump in the neck that has been present for a few weeks. It is sensitive to pressure, is smooth and has a cross-section of 0.5 cm. This clinical picture has a different meaning to a general practitioner and an oncologist. On the basis of exactly the same findings, the general practitioner predicts an inflamed lymphatic gland and adopts a waiting attitude whereas the oncologist wants to exclude primary malignancy or metastasis by means of further investigation. When teaching, it would be expected that these two types of doctors would differ widely in their recommendations.

Haemoglobin and 'anaemic complaints'

Research in the general population[12] and in general practice[13] has not shown a clear relationship between classical 'anaemic complaints' and a low level of haemoglobin. General practitioners, however, learn at medical school that anaemia causes certain complaints. Therefore it is more likely that they measure the haemoglobin levels of patients with these complaints than of patients with complaints not supposedly related to anaemia. The probability of finding a low haemoglobin level by accident is higher in persons with 'anaemic complaints' than in persons with other complaints. If the general practitioner refers people with low haemoglobin levels, for which there is no explanation, to a specialist for further examination, the latter will see comparatively more patients with a low haemoglobin level who have 'anaemic complaints'. The sensitivity of 'anaemic complaints' to low haemoglobin levels can then be relatively high, and even if there is no relationship between them, the specialist will be confirmed in his view that these factors are related. This type of selection bias in the association of complaints and described diseases may be widespread.

Cut-off points between 'normal' and 'abnormal' results

Where the line is drawn between a normal and an abnormal result is seldom objective. In general it is the result of an arbitrary decision based on a consensus of opinions. Examples are: the lower limit of haemoglobin levels, the upper limit of blood pressure, liver function tests or blood sugar levels, the level of the S–T segment in an exertion electrocardiogram and the acceptable number of mitoses in cytological examination. Sensitivity and specificity depend on the choice of cut-off point. This is shown in Figure 3a, which shows haematocrit readings for Swedish women with and without iron deficiency; Garby could make this distinction from the response to the administration of iron of a sample of women taken from the general population.[14] It can be seen that if the haematocrit cut-off point is raised (that is, made less stringent) more iron-deficiency anaemia patients will be found. At the same time, however, there will be more false-positive results — sensitivity increases, but specificity decreases. If the cut-off point is lowered, the opposite will occur. In other words, the more stringent the cut-off point chosen for a certain population, the more the predictive value of an abnormal result increases. In the case of a less stringent cut-off point, the predictive value of an abnormal result decreases.

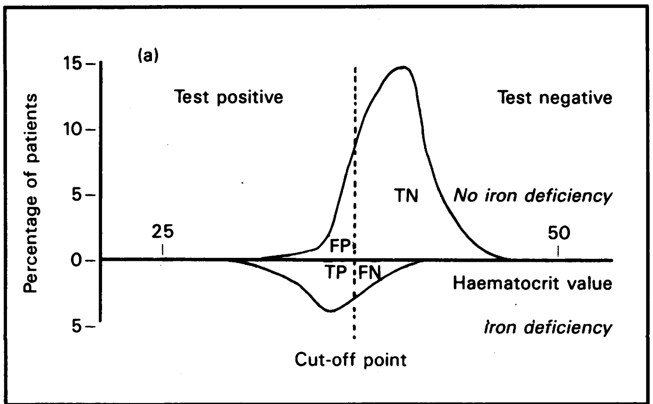
What is the optimal cut-off point for a test? This question cannot easily be answered. The answer depends on the subjective values attached to false-positive and false-negative results. If the consequences of a false-negative result are considered to be more serious than those of a false-positive result, emphasis will be placed on a test with high sensitivity. If, on the other hand, the consequences of a false-positive result are considered to be the more serious, a more specific test is needed. At present, the consequences of a false-negative result are considered to be more serious than those of a false-positive result, and general practitioners tend to choose less stringent cut-off points. These decisions lead to a large number of false positives which could possibly result in iatrogenic illness.

The choice of an optimal cut-off point also depends on the prevalence of the disease under study.[15] This is shown in Figure 3b. In this instance the population represents a larger number of women suffering from iron-deficiency anaemia than in Figure 3a. Therefore, more false-negative results are to be expected with the same cut-off point. In order to reduce the number of false negatives, a less stringent cut-off point must be chosen. For this reason, it cannot be assumed that the same cut-off point is appropriate for specialists and general practitioners. Figure 3a shows that it is also possible to interpret test results without using single cut-off points. For every haematocrit reading, the probability of iron deficiency is the quotient of the height of the curve for iron deficiency over the sum of the heights of both curves for that haematocrit value. In the example shown this probability for the haematocrit readings 31 (and lower), 33, 36, 39 and 41 (and higher) is 100, 80, 59, 9 and 0 per cent, respectively. Even these probabilities, however, are not independent of prevalence.

The severity of the clinical picture

One of the most important differences in the morbidity patterns observed by the general practitioner and the specialist is the extent to which the clinical picture of any disease has developed. In general, symptoms will be less well-differentiated when observed by the general practitioner than in the later phase after referral to the specialist. In general practice sick persons are not as easily distinguished from healthy persons and the discrimination of test methods is much lower than in the later phase after referral to the specialist. Sensitivity — and if choosing the right cut-off point, specificity also — is easier for the specialist to achieve.

One can see that in a population with more serious cases of iron-deficiency anaemia, the distribution of haematocrit values for iron deficiency moves to the left (Figure 3c), that is, to lower haematocrit levels. The discrimination of the haematocrit reading in detecting iron deficiency increases. It could be said that the discrimination of a temperature recording is nil in the first phase of acute appendicitis, and that this is also true of liver function tests in the early stages of moderate hepatitis. Similarly, with breast cancer — the specialist examines larger tumours than the general practitioner, and these tumours are more easily palpable. In summary — in general practice sick people need to be distinguished from healthy people and in hospital healthy people need to be distinguished from sick people. That the validity of tests and the interpretation of symptoms should be directly related to the populations consulting and to the degree of clinical differentiation of the disease at the time of reporting has only recently received attention in the literature.[16-18] This has important implications for general practice, as only a few interpretations of this kind have been verified in general practice.



Cut-off point	Sensitivity (%)	Specificity (%)
39	94	63
38	83	79
37	68	90
36	47	97

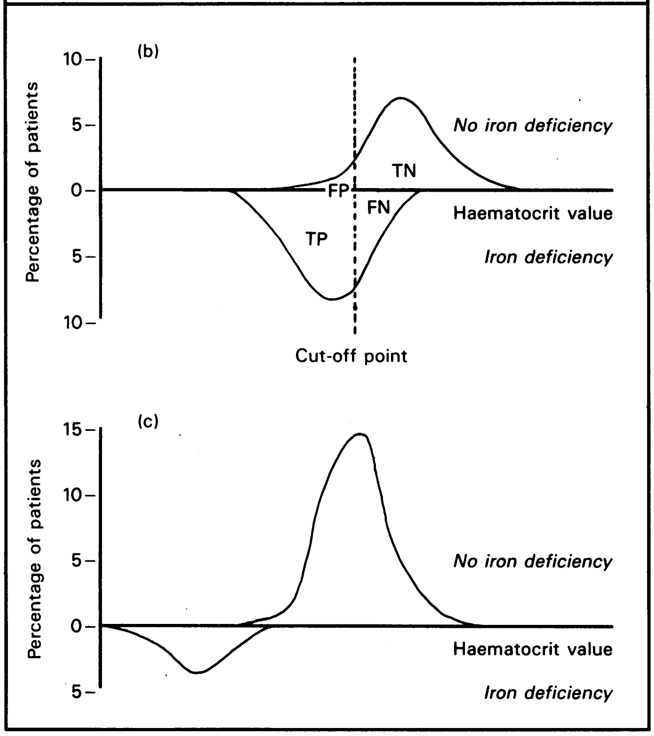


Figure 3. *The relationship between cut-off point and sensitivity and specificity for haematocrit readings. (a) Change in sensitivity and specificity when altering the cut-off point. (b) In case of higher prevalence there are more false-negative and less false-positive results at the same cut-off point. (c) If the iron deficiency is more serious and the distribution among healthy people remains unchanged the discrimination increases. TP = true positive, TN = true negative, FP = false positive, FN = false negative.*

Conclusion

It is concluded that the general practitioner and specialist must learn to deal with widely different patterns of predictive values of complaints, symptoms and test results. Since teaching in medical school is to a great extent based on hospital experience, the specialist has generally learned the working methods relevant to his situation from the beginning of his professional education. The general practitioner, however, has to build up a great deal of relevant knowledge later on.

Although we should learn from one another, it may be harmful to transplant experience, knowledge and standards from one situation to another. For general practice it is necessary to understand the unexplored field of diagnostics. In doing so, attention should be given to the predictive values of complaints, symptoms and test results, to the extent to which test results are accepted as normal, to the finding and improving of diagnostic instruments of particular importance to general practice, and to the correction of self-confirming clinical distortions of the interrelation between symptoms and diseases.

References

1. McWhinney IR. Problem solving and decision making. *An introduction to family medicine.* Oxford University Press, 1981.
2. Velden HGM van der. Diagnose of prognose; de betekenis van de epidemiologie voor het handelen van de huisarts. *Huisarts en Wetenschap* 1983; **26:** 125-128.
3. Rombach JJ. *Breast cancer screening, results and implications for diagnostic decision making.* Proefschrift: Stafleu, Alphen aan de Rijn, 1980.
4. Griner PF, Mayewski RJ, Mushlin AI, Greenland P. Selection and interpretation of diagnostic tests and procedures. *Ann Intern Med* 1981; **94:** 533-596.
5. Clarijs WWJ, Van Deelen-Willems M, Hendriks HHCL, *et al.* Het occulte mammacarcinoom. *Ned Tijdschr Geneeskd* 1977; **123:** 526-529.
6. Dronkers DJ. De betrouwbaarheid van mammografie. *Ned Tijdschr Geneeskd* 1977; **121:** 955-957.
7. Molenaar JC, Herder-Kroon B den. De practische waarde van de mammografie voor de chirurg. *Ned Tijdschr Geneeskd* 1970; **114:** 1168-1690.
8. Oliemans AP. *Morbiditeit in de huisartsenpraktijk.* Leiden: Stenfert-Kroese, 1969.
9. Lamberts H. De morbiditeitsanalyse van de groepspraktijk ommoord, een nieuwe ordening van ziete-en probleemgedrag in de huisartsgeneekunde 11. *Huisarts en Wetenschap* 1975; **18:** 7-39.
10. Lamberts H. Incidentie en prevalentie van gezondheidsproblemen in de huisartsenpraktijk. *Huisarts en Wetenschap* 1982; **25:** 401-414.
11. Dombal FT de. Acute abdominal pain, an OMGE survey. *Scand J Gastroenterol* 1979; **56:** Suppl 14, 29-43.
12. Elwood PC. Symptoms and circulating haemoglobin level. *J Chronic Dis* 1969; **22:** 615-628.
13. Watt CM. The estimation of haemoglobin and the diagnosis of anaemia in general practice. *Practitioner* 1972; **209:** 218-222.
14. Garby L. The normal haemoglobin level. *Br J Haematol* 1970; **19:** 429-434.
15. Weinstein MC, Fineberg HV. *Clinical decision analysis.* Philadelphia: WB Saunders, 1980.
16. Ranschoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *New Engl J Med* 1978; **299:** 926-930.
17. Anonymous. Help with diagnostic decisions. *Ann Intern Med* 1982; **97:** 444-445.
18. Rumke Chr L. Kanttekeningen over de gevoeligheid, de specificiteit en de voorspellende waarde van diagnostische tests. *Ned Tijdschr Geneeskd* 1983; **127:** 556-561.

Acknowledgement

In preparing and translating the manuscript the author is grateful for the assistance of Miss S.A.A. Breevoort.

Address for correspondence

Dr J.A. Knottnerus, Capacity Group of Epidemiology/Health Care Research, PO Box 616, 6200 MD Maastricht, The Netherlands.