Diagnosis and general practice

NICHOLAS SUMMERTON

SUMMARY

This discussion paper presents the case for a rigorous diagnostic research agenda within primary care. In view of unique aspects of primary care medicine and the relatively unselected nature of the populations encountered by general practitioners, it is clear that diagnostic research undertaken in other settings may be inapplicable.

Most diagnostic studies adopt designs that seek to compare items, or clusters of items, of clinical information against a gold standard. In order to enhance the feasibility and rigour of such research within primary care, suggestions are made about priorities and specific key methodological issues.

It is essential that the information generated by primary care diagnostic research should be reliable, valid, useful, and useable within general practice. The ultimate objective must be the identification of specific items, or small clusters of items, of clinical information of high discriminant ability among the groups of patients encountered in general practice.

Keywords: diagnostic research; clinical decision making; treatment versus non-treatment.

Introduction

ACCORDING to Morrell, general practitioners (GPs) are primarily diagnosticians[1] and yet, as pointed out by Howie in 1972, diagnosis remains the Achilles heel in general practice.[2] Today, over 25 years later, it seems that medical research has yet to address in any meaningful manner what symptoms and signs indicate in primary care; just how useful is a particular symptom at predicting a certain disease, which symptoms are not useful, and which symptoms will rule out disease.[3] This view has been echoed by both the *Medical Research Council Topic Review on Primary Health Care* and the report of the *National Working Group on R & D in Primary Care* (Mant Report).[4,5] In addition, the Medical Defence Union has reported that failure or delay in diagnosis consistently accounts for nearly one-third of notified complaints concerning GPs.[6] As the most frequent clinical condition associated with such diagnostic failure or delay is missed malignancy, it is salutary to note that the evidence for the significance of chronic cough as a feature of lung cancer or haematuria as a symptom of urological malignancy within primary care populations remains inadequate.[7,8]

There are a number of possible reasons for the lack of emphasis on diagnostics within primary care academic research. Diagnostic research is difficult: it requires large numbers of patients, often involves long time periods, and requires very robust methods. Such research is clinical and GP vocational training is designed on the assumption that clinical medicine is best taught by an appropriate systems specialist. The tradition is that the diagnosis and assessment of, for example, chest pain is best learnt from a cardiologist, the diagnosis and assessment of abdominal pain from a surgeon. At a macro level, health authorities, public health departments, and many evidence-based institutions and initiatives seem to perpetuate the myth that patients present with 'colorectal cancer' or 'heart failure' rather than a complex array of items of clinical information.

In this climate diagnostic research is given low priority; therefore, in developing a diagnostic research agenda within primary care settings, there is a requirement to ensure that it is both necessary and feasible. In addition, as it focuses on an extremely important aspect of clinical decision making, we need to be well aware of the potential pitfalls of such research. Above all, any discriminant information generated should be both useful and useable in day-to-day clinical general practice.

Why do we need to undertake diagnostic research in primary care settings?

Increasingly it is being recognised that there are important differences between primary care medicine and secondary/tertiary care medicine.[9] For example, there are distinct dissimilarities between the patients, pathologies, and presentations encountered by the GP in comparison with his or her specialist colleagues. Within primary care conditions will often be seen at an evolutionary stage when classical descriptions and classifications simply do not apply. According to McWhinney, many illnesses even defy categorisation as they are transient and self-limiting, or are treated early before reaching the stage of traditional diagnosis.[10]

Decisions made by GPs are also different from those made by specialists — the precise diagnostic labels are often less important than deciding on an appropriate course of action. Diagnoses may be framed in terms of dichotomous decisions: treatment versus non-treatment, referral versus non-referral, and serious versus non-serious. In a well-known general practice-based study of respiratory illness, Howie concluded that a specific diagnostic label may be merely a justification of antibiotic treatment rather than a reason for it.[2]

Making the right choice early is important; diagnostic inefficiency and diagnostic inaccuracy may have untoward consequences. In relation to cancer, general practice diagnostic delay can have adverse effects on prognosis[11] as well as the nature of the interventions required. For example, patients with late stage testicular cancer often require more extensive, aggressive, and disabling treatment than those with early stage disease.[12] On the other hand, overdiagnosis or unnecessarily excessive testing (diagnostic inefficiency) can lead to both physical and psychological damage. Normal children who were misdiagnosed as having organic heart disease show as much deterioration in physical and social function as children who really do have damaged hearts.[13]

In view of the low prevalences of many serious or important conditions within primary care, GPs necessarily differ from specialists in the amount of risk they are willing to take when confronted with a particular symptom. On moving from hospital practice to general practice a doctor needs to adjust his or her perception of the balance between probability and pay off. The problem is that in taking any risks based on probabilities, even in low-prevalence populations, it is essential to be certain of the adequacy of the diagnostic knowledge base. In view of the severe consequences of missing some diagnoses (e.g. myocardial infarction) GPs must be aware of the validity and, in particular, the

N Summerton, MA, MFPHM, MRCGP, clinical senior lecturer in primary care medicine, Clinical Assessment Programme, University of Hull.
Submitted: 14 September 1999; Editor's response: 27 March 2000; final acceptance: 30 May 2000.

false-negative rates associated with the clinical information they are using. If the validity is dubious or the false-negative rate is too high the information may become inapplicable in relation to the amount of risk GPs are willing to accommodate.

The discriminant ability of any item of clinical information can be expressed in a number of ways — for example, predictive value, sensitivity, specificity, and likelihood ratio. The predictive value often makes the most intuitive sense to clinicians — i.e. the positive predictive value is the probability that the disease is present if the 'test' is positive, the negative predictive value is the probability that the disease is absent if the 'test' is negative. For the primary care clinician it is important to be aware that the predictive value is affected by the prevalence: as the prevalence falls, the number of false positives tends to increase resulting in a lowering of the positive predictive value.[14] The effect of prevalence can also be readily understood in relation to Bayes Theorem — i.e. posterior probability = likelihood ratio x prior probability. This emphasises that the interpretation of any new information should depend on the prior probability (prevalence) of disease. Thus, in primary care, the posterior probability of disease will be lower than in a selected population even if the same discriminant functions with identical 'weights of evidence' can be applied.[15] In 1990, Sox and colleagues were able to demonstrate that the prevalence of coronary artery disease was much lower in two primary care populations than in a referred arteriography population, even when patients with similar logistic scores, and thus similar clinical histories, were compared.[16]

The sensitivity is the probability of a positive result with a test if the disease is present; specificity is the probability of a negative result with the test if the disease is absent. Traditionally, sensitivities and specificities have been applied to laboratory testing but the functions can clearly be used for any type of clinical information. Sackett has even coined the terms 'SpPin' and 'SnNout' to make the terms more meaningful for clinicians.[17] The rule of SnNout states that if a test has sufficiently high sensitivity, a negative result rules out the target disorder (i.e. very few false negatives). The associated rule of SpPin states that if a test has sufficiently high specificity, a positive result rules in the disorder (i.e. very few false positives).

It is often assumed that the values for the sensitivity and specificity are fixed. This is not the case: the values for sensitivity and specificity must themselves be derived from a study undertaken within a defined study population in which the test is compared against some gold standard approximating to the truth. Unfortunately, the majority of such studies have been conducted outside primary care settings and GPs need to know the nature of this defined study population if they are going to give meaning to the sensitivity and specificity values. Many such studies recruit cohorts of either healthy, self-selected volunteers, such as medical students, or sicker populations from outpatient clinics. This is important as a test may only be positive in extensive/advanced disease or in sicker patients but negative in localised disease, or in situations where there are significant co-morbidities. Within primary care, conditions will often be seen at an evolutionary stage when sensitivities and specificities of clinical information are changing; by the time the patient reaches the ward or the clinic the discriminant functions may have become more fixed. The process of referral may itself alter the discriminant value of the clinical information — some being used up during the course of the referral process.[18]

In recent years there has been an increasing vogue for the use of the likelihood ratio, a ratio between the sensitivity and the specificity. This is useful as the size of the likelihood ratio gives an indication of the 'weight of evidence' provided by the test and the characteristics of the test can be described in a more clinically meaningful series of strata. Calculating a ratio also overcomes the prevalence problem but, as in the case of individual sensitivity and specificity values, likelihood ratios cannot be easily transposed from one population to another.

Thus, in view of the unique nature of primary care medicine and the relatively unselected nature of the populations encountered by GPs, it is clear that primary care diagnostic research needs to be undertaken in the setting and among the patients in which the results will be applied.

Is it feasible to undertake diagnostic research in primary care settings?

As is evident from published diagnostic work undertaken in Holland and Norway, it is possible to estimate discriminant functions within primary care populations.[19–22] However, such studies require the involvement of large numbers of patients and practitioners. Significant amounts of time, funding, collaboration, and organisation are required to complete the work.

Within the United Kingdom, the existence of the *MRC General Practice Research Framework* and the advent of primary care research networks have made such diagnostic research more feasible in British general practice settings. However, in view of the costs of such studies it is clearly necessary to develop a prioritisation framework. Wilson and Jungner's modified screening criteria provide useful guidance[23] (Box 1).

The diagnosis of most of the more common cancers, many cardiological conditions, and certain rheumatological conditions, e.g. rheumatoid arthritis or carpal tunnel syndrome, clearly fit into such a framework.

In general the studies designed to derive diagnostic prediction rules are cross-sectional and analytical and seek to compare items (or clusters of items) of clinical information against a diagnostic benchmark (or gold standard). Where follow-up is used this is to develop or enhance the validity of the gold standard (e.g. 'delayed-type' cross-sectional studies where the gold standard diagnosis is arrived at after a pre-determined period of follow-up[24]), or to also develop prognostic indicators. Using a variety of mathematical techniques it is then possible to decide what features in a patient's initial presentation may have the most significance and provide the greatest 'weight of evidence' for diagnosis in a primary care setting. In order to allow exploration of the relationships of the items of clinical information with each other and with the outcome, multivariate statistical techniques such as logistic regression, discriminant function analysis, and recursive partitioning may be required.[25]

What are the major problems of diagnostic research in primary care settings?

1. Deciding on the clinical information

The first problem in developing primary care diagnostic research is to decide what clinical information ought to be examined. Traditionally there is a tendency to focus on the medical history, the clinical examination, and investigations. However, there may be other features of a patient's presentation that have predictive value, e.g. the pattern and characteristics of attendance in the three years prior to consultation, or previous alcohol/smoking history and recent changes in consumption. Over 30 years ago, Pereira-Gray noted the importance of behaviour change in indicating a likelihood of malignancy; one specific example often cited is a recent decision to stop smoking.[26] In a large cohort study, Nylenna discovered that the patient's fear of cancer was an important predictor of malignancy.[27] In 1988 Oleson examined the pattern of attendance in patients for three years prior to

The disease	• Can the disease be clearly defined? • Is the condition an important problem in terms of prevalence or seriousness? • Is there a recognised early or evolutionary stage? • Does the condition present a diagnostic problem (i.e. inaccuracy or inefficiency) and would it be useful to have better diagnostic tools? • Could the diagnostic process for the disease be improved by obtaining information in a less risky fashion or in a manner more acceptable to patients?
The clinical information being considered for use as a diagnostic tool	• Is it useable within primary care? (i.e. simple and quick to extract from patients) • Is it reliable? • Is it safe and acceptable to extract from patients? • Is there a valid and reliable gold standard against which the clinical information can be compared in order to assess its validity?
The impact of better diagnosis	• Is there an effective treatment for the target condition and could improved diagnosis lead to either better treatment delivery or shorter treatment delay? • Would intervention at an earlier stage in the disease result in lowered mortality (taking into account lead time bias)? • Would better diagnosis result in lowered morbidity both from the disease and from the diagnostic process?

Box 1. Suggested prioritisation criteria for clinical diagnostic research within primary care populations.

the diagnosis of cervical carcinoma and remarked that significantly more patients than controls had no doctor contact.[28] These factors, termed 'quantifiable features of health-seeking and health-modifying behaviour', may be of particular value in diagnostic decision making within primary care settings.

Diagnostic research in secondary care often ignores the temporal factor. In hospital settings patients attend the clinic on a specific date and, based on this assessment supplemented with appropriate investigations, are provided with a 'diagnostic label'. The emphasis is on making a clear and prompt diagnosis in a patient who has already been sifted by primary care. In contrast, general practice patients often present early in the course of an illness when typical symptoms and signs are absent and, as the condition evolves, GPs may rely more on the assimilation of information gained over a period of time ('dynamic evidence', e.g. the addition of new features, the persistence or changes in the characteristics of previous problems) rather than the traditional static information obtained at one point. Significant symptoms (or clusters of symptoms) developing per unit time are often useful diagnostic tools in the real world of general practice.

In deciding on the extent of clinical information about which to enquire, guidance can be sought from theoretical constructs such as the health belief model,[29] previous research or research tools (e.g. use of validated descriptive instruments as discriminant tools),[30] or from qualitative and quantitative research among both patients and GPs. Salander *et al* recently published a descriptive study of the pathways from symptoms to medical care in a consecutive sample of 28 patients with malignant gliomas.[31] This study emphasised the importance of the clinical information provided by the spouse and whether the spouse escorted the patient to the GP. Other work has indicated a requirement to consider the 'information within the information' ('symptoms within symptoms'). Bland labels need to be given more precise meanings and operationalised; for example, the term 'palpitations', as used both by patients and doctors, is too imprecise, ambiguous, and vague.[32] There is an important requirement for precision in the definition of any potential predictive variable.

In view of the low prevalences of many serious and important conditions within primary care it seems likely that clusters of information may have a greater role than individual items. However, regression shrinkage techniques will need to be applied in order to ensure that the potential interactions and dependency between the variables are appropriately modelled.[33] Muris *et al* have demonstrated how simple, individual items of clinical information can be built into 'clusters' with powerful predictive value in the diagnosis of intra-abdominal malignancy within primary care populations (Table 1).[34]

When assessing the usefulness of any item (or cluster) of clinical information it is always necessary to consider its reliability as well as its validity; it is not only necessary to hit the bull's-eye, but to hit it consistently. Some clinical information exhibits significant interobserver variability (e.g. in relation to routine chest examination[35]) and Laupacis and colleagues have suggested that this is the most important aspect of reliability to assess as the discriminant information, once published, will be used by a variety of different clinicians.[25] The kappa statistic provides a helpful measure of the extent of reliability by assessing the agreement between observers or observations.[36] A kappa value of 0 indicates no agreement beyond that expected by chance, and a kappa of 1.0 reflects perfect agreement; values of less than 0.6 are taken as indicating unreliability.

Among patients, there may be specific unreliability difficulties; for example, the knowledge of the cancer family history may vary significantly between members of the same family.[37,38] Other questions (e.g. history of pregnancies or terminations) may become unreliable among certain groups as they may be greeted by socially acceptable responses.[39] Crowne and Marlowe have developed a useful tool to assess the extent of this 'socially acceptable bias'.[40]

2. Biases within diagnostic studies

Cross-sectional studies that seek to examine the discriminant value of key items or clusters of items of clinical information can be quite problematic. Sackett has identified a large collection of biases that need to be addressed in designing analytical studies.[41] A number of these biases are worthy of specific recognition in relation to diagnostic research.[42] Misclassification and selection bias are particular problems in undertaking such research within primary care settings.

Much primary care diagnostic research focusing on malignancy adopts a delayed-type cross-sectional design in which items of clinical information obtained at one point are compared against a gold standard derived by following up patients for a pre-defined period. For example, during 1989, 933 patients with new onset non-acute abdominal complaints within Limburg, Holland were recruited. All patients underwent a physical examination by their own GP and were asked to complete a questionnaire on symptoms/psychological factors and were then subjected to some simple investigations. The patients' records were subsequently

Table 1. Clinical information clusters and the diagnosis of intra-abdominal malignancy.

	Probability for neoplastic disease (%)
Age > 65, male	3
Age > 65, male, non-specific abdominal pain	18
Age > 65, male, non-specific abdominal pain and weight loss	50
Age > 65, male, non-specific abdominal pain, weight loss, and ESR > 20mm/h	75

reviewed after one year to assess the nature of the final diagnostic gold standard.[34] Information was then provided on which symptoms and signs were most helpful in predicting which of the patients with non-acute abdominal symptoms at one point subsequently turned out to have serious disease.

An ideal study of diagnosis in primary care using 'pathology at follow-up' as the gold standard would adopt a study design that closely reflected the natural history of patient follow-up in primary care. Three to four weekly re-assessments (if possible combined with symptom diaries) of patients would not only provide more useful 'dynamic' clinical information but would also enable the study to assess more reliably the association between the clinical information and the eventual diagnosis. In the absence of a clear link between the symptoms and the outcome, the extent of misclassification bias is very dependent on the length of follow-up. If the observation period is too short then conditions may be missed; too long and false linkages may be construed between symptoms and outcomes. Detection bias is a related problem resulting from, for example, preferential referral of patients with abdominal pain and rectal bleeding as opposed to abdominal pain alone. Thus, an early cancer may be more likely to be detected in a patient with abdominal pain and rectal bleeding than in a patient solely with abdominal pain. Clearly, the length of follow-up also needs to be designed to address this issue.

Selection bias is a major problem in diagnostic studies that rely on recruitment of patients by GPs. Despite taking great care, GPs may selectively re-interpret inclusion and exclusion criteria. For example, GPs might only recruit patients whom they would have referred onwards anyway, e.g. patients with moderate symptoms of prostatism may only be included in a study of such symptoms if they have also been noted to have an enlarged prostate (selection by test outcome). In other situations there may be a tendency to enlist patients with intermediate or greater probabilities of illness (selection by clinical spectrum).[43] Patients with significant co-morbidities may also be excluded from many diagnostic studies and the range of differential diagnoses in the study population may consequently not reflect the true situation in primary care.

One approach to the problem of selection is to consent a random proportion of patients to be re-interviewed by phone in order to validate the eligibility criteria being applied by the recruiting GPs. Another way is to contact all patients directly: this was the method adopted by Stoffers *et al* when they sought to assess the diagnostic value of signs and symptoms associated with peripheral arterial occlusive disease seen in general practice.[44] A list was obtained of all the patients aged 40–75 years registered with 18 GPs in Limburg, Holland. These 26 620 subjects then received a simple postal questionnaire enquiring about 'leg complaints on walking' and those reporting in the affirmative were invited to attend for assessment. However, in adopting such an approach care must be taken to enquire about whether such patients do consult, or would consider consulting their GP about such symptoms, as we clearly need to understand both the significance of community symptoms as well as identifying those patients that may present to the GP. Clearly some symptoms, such as haemoptysis, have a greater 'iatrotrophic stimulus'[45] than, perhaps, a change in bowel habit. Such findings could therefore also have important implications for education within the community.

In all diagnostic studies undertaken within primary care, an appraisal ought to be made of the extent of selection bias by checking against other sources of information (e.g. GP records, hospital records, registries) to identify potential cases not recruited into the study. If possible, key demographic characteristics should be compared between patients selected and those not selected for inclusion in a diagnostic study.

3. The gold standard

In all cross-sectional diagnostic analytical studies, the method involves the comparison of a battery of clinical information against a gold standard. There are a variety of types of gold standards, e.g. invasive or costly techniques involving, for example, biopsy, radiology, or electrophysiology; simple observation over a pre-defined period; or the use of an independent expert panel. The ultimate objective is for the gold standard to approximate the truth as closely as possible. Misclassification has already been discussed in relation to follow-up type gold standards. However, it is important to appreciate that misclassification can even occur with invasive tests, e.g. by biopsying the wrong area or misinterpreting an electophysiological measurement.

Gold standards derived by observation for a pre-defined follow-up period are particularly applicable to primary care diagnostic research. However, there may be pathological and clinical information dissimilarities in the presentation of patients who develop definitive disease after a shorter period of observation. Thus, there may also be a need for time-specific stratification of the gold standard categorisation within delayed-type cross-sectional studies.

Either through losses to follow-up in delayed-type cross-sectional studies, or resource constraints in the case of other gold standards, only a proportion of individuals may be subjected to the gold standard. To avoid verification or work-up bias it is essential that every individual (or a representative random selection of sufficient power) from whom clinical information is extracted are also submitted for assessment by the gold standard. If this does not occur there tends to be an overestimate of the strength of the association.[46] It is also well recognised that knowledge of clinical or other factors can, in certain circumstances, influence a diagnostic test result. Thus, if the gold standard involves an interpretable component then it is important that it is assessed independently and, ideally, blind in relation to the clinical information already obtained.[42]

In all diagnostic research, the underlying assumption is that there really is an identifiable truth; a diagnostic category that fits the clinical information and that such a diagnosis can be verified by an independent gold standard method. This may not be the case and, in the absence of an adequate gold standard, statistical methods are being developed in order to estimate the accuracy of new tests.[47,48] There may even need to be a re-assessment of the nature of diagnostic gold standards in primary care research with,

perhaps, more of an emphasis on the diagnosis of 'wellness' by using a battery of gold standards to exclude serious or important illness (reverse gold standard), or even gold standards related to a course of action rather than a clinical diagnosis.

In deciding on the significance of a possible link between the clinical information and a gold standard, it is not necessary to demonstrate causality but rather a strong association.[49] However, in primary care where we may rely more on surrogate makers of potential disease we need to be clear about what information has been considered within a study and what has not. To further support a diagnostic role for an item (or cluster) of clinical information, the evidence generated should be of sufficient strength (as assessed by the magnitude of the numerical association), if possible exhibit a dose-response relationship, and, most importantly, there should be consistency with other evidence.[50] Thus, diagnostic studies undertaken in one setting need to be repeated elsewhere by different practitioners in different populations. In addition, pragmatic randomised controlled trials will be required in order to assess the practical application and impact of the diagnostic information generated by the analytical studies. We eventually need to be able to focus downstream from the test to assess its influence on subsequent patient management and the eventual course of the disease.

Conclusion

Within this discussion paper I have sought to argue the case for developing a rigorous diagnostic research agenda within primary care. It is essential to ensure that the information generated is valid, reliable, useful, and useable within general practice. The ultimate objective must be the identification of specific items or small clusters of items of clinical information of high discriminant ability that can be applied with confidence within general practice when required during the course of a busy surgery. The research community needs to assist GPs with appropriate decision making in the settings where they practice and among the types of patients they encounter.

References

1. Morrell DC. *Diagnosis in general practice. Art or science?* London: Nuffield Provincial Hospitals Trust, 1993.
2. Howie JGR. Diagnosis — the Achilles heel? *J R Coll Gen Pract* 1972; **22:** 310-315.
3. Owen P. Clinical practice and medical research: bridging the divide between the two cultures. *Br J Gen Pract* 1995; **45:** 557-560.
4. Medical Research Council. *MRC topic review: primary health care.* London: MRC, 1997.
5. NHS Executive. *R & D in primary care. National Working Group Report*. London: Department of Health, 1997.
6. Green S, Price J. Complaints. *Pulse,* 4 April 1998; 63: 67.
7. Liedekerken BMJ, Hoogendam A, Buntix F, *et al*. Prolonged cough and lung cancer: the need for more general practice research to inform clinical decision making. *Br J Gen Pract* 1997; **47:** 505.
8. Bruntix F, Wauters H. The diagnostic value of macroscopic haematuria in diagnosing urological cancers: a meta analysis. *Fam Pract* 1997; **14:** 63-68.
9. Summerton N. The medicine of primary care. *Br J Gen Pract* 1999; **49:** 604-605.
10. McWhinney IR. *A textbook of family medicine.* Oxford: Oxford University Press, 1997.
11. Richards MA, Westcombe AM, Love SB, *et al*. Influence of delay on survival in breast cancer: a systematic review. *Lancet* 1999; **353:** 1119-1126.
12. Neal AJ, Hoskin PJ. *Clinical oncology.* Oxford: Oxford University Press, 1997.
13. Bergman AB, Stamm SJ. The morbidity of cardiac non-disease in schoolchildren. *N Engl J Med* 1967; **276:** 1008-1013.
14. Vecchio TJ. Predictive value of a single diagnostic test in unselected populations. *N Engl J Med* 1966; **274:** 1171-1173.
15. Deeks JJ. Using evaluations of diagnostic tests: understanding their limitations and making the most of available evidence. *Ann Oncol* 1999; **10:** 761-768.
16. Sox HC, Hickman DH, Marton KI, *et al*. Using the patient's history to estimate the probability of coronary artery disease: a comparison of primary care and referral practices. *Am J Med* 1990; **89:** 7-14.
17. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine.* Boston, MA: Little, Brown and Company, 1991.
18. Knottnerus JA, Leffers P. The influence of referral patterns on the characteristics of diagnostic tests. *J Clin Epidemiol* 1992; **45:** 1143-1154.
19. Melbye H, Straume B, Aaseb U, Dale K. The diagnosis of adult pneumonia in general practice. Relative importance of typical symptoms and abnormal chest signs evaluated against a radiographic reference standard. *Scand J Prim Health Care* 1992; **10:** 226-233.
20. Lindboek M, Hjortdahl P, Johnsen ULH. Use of symptoms, signs, and blood tests to diagnose acute sinus infections in primary care: comparison with computed tomography. *Fam Med* 1996; **28:** 183-188.
21. Govaert TME, Dinant GJ, Aretz K, Knotnerus JA. The predictive value of influenza symptomatology in elderly people. *Fam Pract* 1998; **15:** 16-22.
22. Fijten GH, Starmans R, Muris JWM, *et al*. Predictive value of signs and symptoms for colorectal cancer in patients with rectal bleeding in general practice. *Fam Pract* 1995; **12:** 279-286.
23. Wilson J, Jungner G. *Principles and practice of screening for disease, World Health Organisation Public Health Paper 34*. Geneva: WHO, 1968.
24. Knottnerus JA. Diagnostic prediction rules: principles, requirements, and pitfalls. *Prim Care* 1995; **22:** 341-362.
25. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methological standards. *JAMA* 1997; **277:** 488-494.
26. Pereira Gray DJ. The role of general practitioners in the early detection of malignant disease. *Trans Hunt Soc* 1966; **25:**135-179.
27. Nylenna M. Diagnosing cancer in general practice; from suspicion to certainty. *BMJ* 1986; **293:** 314-317.
28. Olesen F. Pattern of attendance at general practice in the years before the diagnosis of cervical cancer: a case control study. *Scand J Prim Health Care* 1988; **6:** 199-203.
29. Rosenstock I. Why people use health services. *Millbank Q* 1966; **44:** 54-127.
30. Streiner DL, Norman GR. *Health measurement scales.* Oxford: Oxford University Press, 1995.
31. Salander, P, Bergenheim AT, Hamberg K, Henriksson R. Pathways from symptoms to medical care: a descriptive study of symptom development and obstacles to early diagnosis in brain tumour patients. *Fam Pract* 1999; **16:** 143-148.
32. Barsky AJ, Cleary PD, Barnett MC, *et al*. The accuracy of symptom reporting in patients complaining of palpitations. *Am J Med* 1994; **97:** 214-221.
33. Spiegelhalter DJ. Statistical methodology for evaluating gastrointestinal symptoms. *Clin Gastroenterol* 1985; **14:** 489-515.
34. Muris JWM, Starmans R, Fijten GH, *et al*. Non-acute abdominal complaints in general practice: diagnostic value of signs and symptoms. *Br J Gen Pract* 1995; **45:** 313-316.
35. Spiteri MA, Cook DG, Clarke SW. Reliability of eliciting physical signs in examination of the chest. *Lancet* 1988; **I:** 873-875.
36. Fleiss JL. *Statistical methods for rates and proportions.* New York, NY: John Wiley & Sons, 1981.
37. Love R, Evans AM, Josten DM. The accuracy of patient reports of a family history of cancer. *J Chron Dis* 1985; **38:** 289-293.
38. Theis B, Boyd N, Lockwood G, Tritchler D. Accuracy of family cancer history in breast cancer patients. *Eur J Cancer Prev* 1994; **3:** 321-327.
39. Bluestein D, Levin JS. Symptom reporting in wanted and unwanted pregnancies. *Fam Med* 1991; **23:** 271-274.
40. Crowne DP, Marlowe D. A new scale of social desirability independent of psychopathology. *J Consult Psych* 1969; **24:** 349-354.
41. Sackett DL. Bias in analytic research. *J Chron Dis* 1979; **32:** 51-63.
42. Begg CB. Biases in the assessment of diagnostic tests. *Stat Med* 1987; **6:** 411-423.
43. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficiency of diagnostic tests. *N Engl J Med* 1978; **299:** 926-930.
44. Stoffers HEJH, Kester ADM, Kaiser V, *et al*. Diagnostic value of signs and symptoms associated with peripheral arterial occlusive disease seen in general practice: a multivariable approach. *Med Decis Making* 1997; **17:** 61-70.

45. Feinstein AR. *Clinical judgement.* Baltimore, MD: Williams and Wilkins, 1967.
46. Knottnerus JA. The effects of disease verification and referral on the relationship between symptoms and diseases. *Med Decis Making* 1987; **7:** 139-148.
47. Hui SL, Zhou XH. Evaluation of diagnostic tests without gold standards. *Stat Methods Med Res* 1998; **7:** 354-370.
48. De Bock GH, Houwing-Duistermaat JJ, Springer MP, *et al*. Sensitivity and specificity of diagnostic tests in acute maxillary sinusitis determined by maximum likelihood in the absence of an external standard. *J Clin Epidemiol* 1994; **47:** 1343-1352.
49. Elwood M. *Critical appraisal of epidemiological studies and clinical trials*. Oxford: Oxford University Press, 1998.
50. Schlesselman JJ. 'Proof' of cause and effect in epidemiologic studies: criteria for judgement. *Prev Med* 1987; **16:** 195-210.

Acknowledgements

In developing the ideas within this paper I should like to express my gratitude to the following for helpful advice and information: David Mant, Andre Knottnerus, Jean Muris, Ian McWhinney, David Jewell, and the anonymous referee. I am also very grateful to the Eric Gambrill Memorial Fund for providing the support to enable me to visit colleagues in Maastricht.

Address for correspondence

Dr Nicholas Summerton, The Surgery, Manlake Avenue, Winterton, Scunthorpe DN15 9TA. E-mail: N.Summerton@medschool.hull.ac.uk