

Information in practice

Ruling a diagnosis in or out with “SpPIn” and “SnNOut”: a note of caution

Daniel Pewsner, Markus Battaglia, Christoph Minder, Arthur Marx, Heiner C Bucher, Matthias Egger

Dr X is back from her annual leave. Dr Y, the locum doctor, reports on the patients he saw during her absence, including a 40 year old teacher who had sprained her right ankle. Returning from a conference, she had stumbled while walking down the stairs with a heavy bag. Examination revealed a moderately swollen lateral right ankle. The patient was able to walk but was clearly in pain. Her breath smelt of alcohol.

Ruling diagnoses in and out with SpPIns and SnNOuts

Dr Y had applied the Ottawa ankle rules—decision rules designed to exclude fractures of the malleolus and the midfoot—and found no bone tenderness.¹ He had previously visited the website of a centre for evidence based medicine² and printed out a list of diagnostic tests that can rule out, or rule in, the condition in question without requiring further investigations.

The probability of disease, given a positive or negative test result (post-test probability), is usually obtained by calculating the likelihood ratio of the test result and using formulas based on Bayes's theorem (see box 1), or a nomogram,³ to convert the estimated probability of the suspected diagnosis before the test result was known (pretest probability) into a post-test probability, which takes the result into account.⁴ Likelihood ratios indicate how many times more likely a test result is to be expected in a patient with the disease compared with a person free of the disease and thus measure a test's ability to modify pretest probabilities.

David Sackett and others have argued that such calculations are unnecessary when a test is highly sensitive or highly specific.⁴⁻⁶ In this situation the likelihood ratio of a negative test will generally be very small, and the likelihood ratio of a positive test very large. A negative test will thus rule out, and a positive result rule in, disease. Two mnemonics that capture the properties of such tests have been coined: SnNOut (high sensitivity, negative, rules out) and SpPIn (high specificity, positive, rules in).⁴ This concept has become increasingly popular, with many websites for evidence based medicine listing such tests and inviting users to nominate further SpPIn and SnNOut tests. The understanding of the SnNOut principle among medical students was recently examined in a randomised trial.⁷

Summary points

Negative results from highly sensitive tests can rule a diagnosis out (sensitive, negative, out = SnNOut), and positive results from highly specific tests can rule a diagnosis in (specific, positive, in = SpPIn)

Studies quoted as showing SpPIn or SnNOut properties may be affected by spectrum bias, partial verification bias, or incorporation bias. Others may be too small to define test characteristics with sufficient precision

The power of a test to rule a diagnosis out does not depend exclusively on its sensitivity, as suggested by the SnNOut rule, but is reduced by low specificity. Similarly, the power to rule in depends on both specificity and sensitivity

The evidence from studies of a test's accuracy should be critically assessed, and post-test probabilities (with 95% confidence intervals) should be calculated when evaluating potential SnNOut or SpPIn tests

Assuming that a diagnosis can be ruled in or out with confidence, when in reality it cannot, could have serious consequences for patients

The website listed the Ottawa ankle rules as a SnNOut test,² indicating that in the teacher's case a fracture could safely be ruled out without radiography. Indeed, the patient made an uneventful and full recovery within four weeks. Alerted by the patient's alcoholic breath, Dr Y wondered whether an alcohol problem might have contributed to the accident and used the CAGE questions (see box 2) to investigate this further. According to the same website,² the CAGE instrument has SpPIn properties, ruling the diagnosis in if two or more questions are answered affirmatively. The patient confirmed that she felt she should cut down on alcohol and that she had felt bad repeatedly about her drinking. Dr X, who had known her for over 10 years, explained that the patient's alcohol intake was moder-

Division of Epidemiology and Biostatistics, Department of Social and Preventive Medicine, University of Bern, Switzerland

Daniel Pewsner
senior research fellow
Markus Battaglia
senior research fellow
Christoph Minder
professor of medical statistics

Department of General Internal Medicine, Inselspital, University of Bern, Switzerland

Arthur Marx
senior registrar

Basel Institute for Clinical Epidemiology, University Hospitals, Basel, Switzerland
Heiner C Bucher
professor of clinical epidemiology

MRC Health Services Research Collaboration, Department of Social Medicine, University of Bristol, Bristol
Matthias Egger
professor of epidemiology and public health medicine

Correspondence to: M Egger, Department of Social and Preventive Medicine, University of Bern, Finkenhubelweg 11, CH-3012 Berne, Switzerland
egger@ispm.unibe.ch

BMJ 2004;329:209-13

Box 1: Definitions of concepts and terms

Sensitivity—The proportion of people with the disease who are correctly identified by a positive test result ("true positive rate")

Specificity—The proportion of people free of the disease who are correctly identified by a negative test result ("true negative rate")

SnNOut—Mnemonic to indicate that a negative test result (N) of a highly sensitive test (Sn) rules out the diagnosis (Out)

SpPIn—Mnemonic to indicate that a positive test result (P) of a highly specific test (Sp) rules in the diagnosis (In)

Likelihood ratios—Measure of a test result's ability to modify pretest probabilities. Likelihood ratios indicate how many times more likely a test result is in a patient with the disease compared with a person free of the disease.

Likelihood ratio of a positive test result (LR+)—The ratio of the true positive rate to the false positive rate: sensitivity/(1 – specificity)

Likelihood ratio of a negative test result (LR–)—The ratio of the false negative to the true negative rate: (1 – sensitivity)/specificity

Pretest probability (prevalence)—The probability that an individual has the target disorder before the test is carried out

Post-test probability—The probability that an individual with a specific test result has the target condition (post-test odds/[1+post-test odds]) or

$$\text{Post-test probability} = \frac{(\text{pretest probability}/[1 - \text{pretest probability}]) \times \text{LR}}{(\text{pretest probability}/[1 - \text{pretest probability}]) \times \text{LR} + 1}$$

Pretest odds—The odds that an individual has the target disease before the test is carried out (pretest probability/[1 – pretest probability])

Post-test odds—The odds that a patient has the target disease after being tested (pretest odds × LR)

Positive predictive value (PPV)—The proportion of individuals with positive test results who have the target condition. This equals the post-test probability given a positive test result

Negative predictive value (NPV)—The proportion of individuals with negative test results who do not have the target condition. This equals one minus the post-test probability given a negative test result

ate and well controlled, and that she was socially well integrated but very health conscious and somewhat anxious. A few days later, Dr X got a telephone call from the patient, who was clearly upset about the locum doctor's suggestion that she had an alcohol problem. A further consultation was required to clarify the situation and restore trust.

Critical appraisal of test evaluation studies

In this article, we examine examples of test evaluation studies that websites and a textbook of evidence based medicine⁴ have cited as showing that the tests had SpPIn or SnNOut properties. The studies were chosen to illustrate methodological issues. We assessed the quality of studies as described elsewhere,⁸ extracted the two-by-two table from the original publication, and

Box 2: CAGE questionnaire for detecting alcohol misuse

1. Have you ever felt you should **C**ut down on your drinking?
2. Have people **A**nnoyed you by criticising your drinking?
3. Have you ever felt bad or **G**uilty about your drinking?
4. Have you ever had a drink first thing in the morning to steady your nerves or to get rid of a hangover (**E**ye opener)?

calculated likelihood ratios and post-test probabilities with exact binomial 95% confidence intervals based on the pretest probabilities observed in the studies. Finally, we examined whether the post-test probability of the condition in question in the population studied was compatible with the notion of safely ruling the condition in or out, and considered the transferability of study results to other settings and populations. Tables 1 and 2 summarise the study characteristics and results from our critical appraisal.

Random error and bias

A diagnostic study may be too small to define test performance with sufficient precision. For example, a website² and the textbook⁴ interpreted a study of ankle swelling in patients with suspected ascites¹⁰ as demonstrating SnNOut properties. The absence of a history of ankle swelling is thus assumed to rule out ascites.^{2,4} However, the study was based on only 15 patients with ascites and confidence intervals were wide, with the lower 95% confidence interval of the sensitivity including 68%. This means that absence of ankle swelling is still compatible with a 15.8% probability of ascites, which clearly is unacceptably high (table 1).

Studies with methodological flaws tend to overestimate the accuracy of diagnostic tests.²¹ Bias can be introduced when tests are evaluated in patients known to have the disease and in people known to be free of it—so called diagnostic case-control studies. In this situation patients with borderline or mild expressions of the disease, and conditions mimicking the disease are excluded, which can lead to exaggeration of both sensitivity and specificity.²¹ This is called spectrum bias because the spectrum of study patients will not be representative of patients seen in practice. For example, the textbook considered auscultatory percussion in the diagnosis of pleural effusion as a SpPIn test.⁴ This assessment was based on a study that compared patients who were selected because of the presence or absence of radiological signs of effusion.¹⁶ The impressive results (100% specificity and 96% sensitivity (table 2)) may therefore not be reliable. The textbook and a website²² also claim that the presence of retinal vein pulsation in ophthalmoscopy excludes increased intracranial pressure (a SnNOut test). This is based on a study that compared patients known to have increased pressure with people not suspected to have increased intracranial pressure.¹¹

Partial verification bias may be introduced when the reference test or tests are not applied consistently to confirm negative results of the index test. Some patients are either excluded or considered true negatives. This may lead to overestimation of sensitivity and underestimation of specificity or to overestimation of sensitivity and specificity.²¹ The textbook considered the CAGE questionnaire for diagnosing alcohol misuse (box 2) to be a SpPIn test.⁴ This is based on a study that subjected only a fraction of CAGE-negative persons to further testing (liver enzymes, medical record review, and physician interviews (table 2)),¹⁷ thus possibly introducing bias.

Similarly, incorporation bias may be present if the test under evaluation is also part of the reference test.²¹ This will lead to overestimation of test accuracy because experimental and reference tests are no longer independent. For example, a website listed

Table 1 Characteristics of test evaluation studies interpreted as demonstrating SnNOut properties (high sensitivity and negative result rules out)

Test (diagnosis)	Reference test	Setting and patients	Results (tp/fp/fn/tn)	Sensitivity (95% CI)	Specificity (95% CI)	Likelihood ratio (95% CI)	Pretest probability	Post-test probability (95% CI)	Comment
Ottawa ankle rules (ankle or midfoot fractures) ⁹	Radiography	Consecutive adults in emergency departments	70/210/1/135	99% (92% to 100%)	39% (34% to 45%)	0.04 (0.01 to 0.3)	17.1%	0.7% (0.02% to 4.0%)	Post-test probability of ≤4% seems acceptable*
History of ankle swelling (ascites) ¹⁰	Ultrasound scan	Men admitted to a general medicine ward	14/16/1/32	93% (68% to 100%)	67% (52% to 80%)	0.10 (0.01 to 0.7)	23.8%	3.0% (0.1% to 15.8%)	Study compatible with post-test probability of 15.8%, but even the point estimate of 3% may be considered high in this situation
Loss of spontaneous retinal vein pulsation (increased intracranial pressure) ¹¹	Clinical evidence, including lumbar pressure in 9 patients	Patients known to have increased pressure compared with individuals without suspected increased pressure	43/18/0/128	100% (92% to 100%)	88% (81% to 93%)	0 (0 to 0.20)	22.8%	0 (0 to 2.3%)	Diagnostic case-control study, vulnerable to spectrum bias. Results not reliable
NINCDS-ADRDA criteria† (Alzheimer's disease) ¹²	Histology	Patients in secondary care	28/20/2/6	93% (78% to 99%)	23% (9% to 44%)	0.3 (0.06 to 1.3)	53.6%	25% (3.2% to 65.1%)	Power to rule diagnosis out eroded by low specificity, resulting in high post-test probability
Urinary albumin: creatinine ratio >1.8 g/mol (microalbuminuria in diabetes) ¹³	Timed overnight albumin excretion	Men with diabetes screened for microalbuminuria in primary care	405/51/24/691	94% (92% to 96%)	93% (91% to 95%)	0.06 (0.04 to 0.1)	36.6%	3.4% (2.2% to 5.0%)	Post-test probability of ≤5% is acceptable in this situation

tp=true positive, fp=false positive, fn=false negative, tn=true negative.

*Confirmed in recent meta-analysis of 27 studies.¹⁴†Clinical criteria by National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and Alzheimer's Disease and Related Disorders Association (ADRDA).¹⁵**Table 2** Characteristics of test evaluation studies interpreted as demonstrating SpPIn properties (high specificity and positive result rules in)

Test (diagnosis)	Reference test	Setting and patients	Results (tp/fp/fn/tn)	Sensitivity (95% CI)	Specificity (95% CI)	Likelihood ratio (95% CI)	Pretest probability	Post-test probability (95% CI)	Comment: can positive test rule diagnosis in?
Auscultatory percussion (pleural effusion) ¹⁶	Radiography	Patients with effusion compared to patients from surgical and medical wards	113/0/5/175	96% (90% to 99%)	100% (98% to 100%)	∞ (21 to ∞)	40%	100% (97.4% to 100%)	Diagnostic case-control study, vulnerable to spectrum bias. Results not reliable
≥3 positive answers in CAGE questionnaire* (alcohol misuse or dependency) ¹⁷	Biochemical tests, medical record review, and physician interviews	Consecutive orthopaedic-medical patients. Reference test applied to sample of CAGE-negative patients only	60/1/57/400	51% (42% to 61%)	99% (98% to 100%)	206 (29 to 1468)	23%	98.4% (91.2% to 99.9%)	Results unreliable because of possible partial verification bias and wide confidence intervals
Abdominojugular reflux (congestive heart failure) ¹⁸	Clinical criteria that included reflux	Patients complaining of shortness of breath in emergency room	4/2/8/34	33% (10% to 65%)	94% (81% to 99%)	6.0 (1.3 to 29)	25%	66.7% (22.3% to 95.7%)	Results unreliable because of possible incorporation bias and wide confidence intervals. Low post-test probability
Third heart sound (heart failure) ¹⁹	Echocardiography	Consecutive patients with suspected heart failure attending general practices	10/3/31/215	24% (12% to 40%)	99% (96% to 100%)	18 (5.1 to 62)	16%	76.9% (46.2% to 95.0%)	Power to rule diagnosis in is eroded by low sensitivity, resulting in low post-test probability. Wide confidence intervals
≥2 positive answers in CAGE questionnaire* (alcohol dependence) ²⁰	ICD-10 criteria	Sample of black women admitted to trauma centre	25/12/5/287	83% (65% to 94%)	96% (93% to 98%)	21 (12 to 37)	9%	67.6% (50.2% to 82.0%)	Post-test probability too low, considering nature of diagnosis. Applicability to other settings and populations questionable

tp=true positive, fp=false positive, fn=false negative, tn=true negative.

*See box 2 for CAGE criteria.

abdominojugular reflux for the diagnosis of congestive heart failure as a SpPIn test,²² on the basis of a study that used clinicoradiographic criteria, including abdominojugular reflux, as the reference test (table 2).¹⁸

Sensitivity and specificity

The likelihood ratio associated with a negative test result does not depend on its sensitivity alone, as suggested by the SnNOut rule, but also on its specificity. For example, a website considered that the clinical criteria for the diagnosis of Alzheimer's disease

had SnNOut properties² based on a sensitivity of 93% (table 1).¹² However, despite this high sensitivity, the likelihood ratio of a negative test was a modest 0.3, because of the test's low specificity of 23% (100 – 93/23 = 0.3, see box 1). Indeed, in the population studied, the probability of Alzheimer's disease, given a negative test, was 25% (table 1). The power to rule out a diagnosis thus depends on both sensitivity and specificity.

Similarly, the ability to rule in depends not only on specificity, as suggested by the SpPIn rule, but also on

sensitivity. A study examining the presence of a third heart sound in the diagnosis of congestive heart failure (table 2)¹⁹—which a website interpreted as demonstrating SpPIn properties²³—is an example of a highly specific test (99%) that suffers from a low sensitivity (24%). The figure shows how the power to rule a disease in or out is eroded when highly specific tests are not sufficiently sensitive, or highly sensitive tests are not sufficiently specific.

Transferability and applicability

The performance of a diagnostic test often varies considerably from one setting to another, which may be due to differences in the definition of the disease, the exact nature of the test, and its calibration and the characteristics of those with and without the disease in a given setting.²¹ For example, patients attending primary care practices will generally have disease at an earlier stage than patients in secondary and tertiary care, which may reduce a test's sensitivity. Patients free of the disease in tertiary care will tend to have other conditions, which could reduce the specificity of a diagnostic test. Interpreting data on a test's accuracy thus requires defining the exact nature of the test used, the disease, and the patient population studied. For example, the website that listed the CAGE questionnaire as a SpPIn test for alcohol dependence² cited, as the evidence for this, a study that had been performed in black women admitted to a trauma centre in the United States,²⁰ which may not be applicable to other populations and settings.

Even when we assume that sensitivity and specificity do not change between settings and patient populations, test results will have different interpreta-

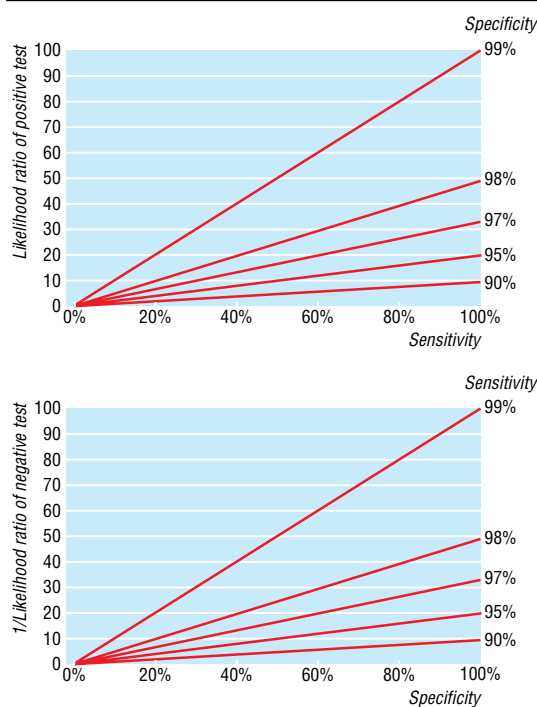
tions depending on whether a test is performed in a low risk population, such as in primary care, or high risk patients in a referral centre. For example, in the study evaluating the third heart sound in the diagnosis of heart failure,¹⁹ the pretest probability or prevalence in a general practice setting was 16%. In this situation, a positive test with a likelihood ratio of 18 will not allow the diagnosis to be ruled in with confidence: the post-test probability is only increased to 77% (table 2). If the pretest probability were 50%, however—such as in a cardiology outpatient clinic—the same positive test would produce a post-test probability of 95% (see box 1 for formula).

The interpretation of studies will be strongly influenced by the nature of the condition and the invasiveness of further investigations. For example, a study assessing urinary albumin:creatinine ratios below 1.8 g/mol for ruling out microalbuminuria in men with type 2 diabetes in primary care¹³ and the study examining the absence of a history of ankle swelling for ruling out ascites in men admitted to general internal medicine wards¹⁰ both produced post-test probabilities of about 3%. In the first case, we accepted a website's conclusion that the urinary albumin:creatinine ratio had SnNOut properties²; we thought that the post-test probability of microalbuminuria was sufficiently low with a negative test result, considering that guidelines recommend regular testing of patients with type 2 diabetes.²⁴ In the second case, however—and unlike the textbook⁴—we thought that in men with suspected ascites but no history of ankle swelling a probability of ascites of 3%, a sign often associated with serious conditions, was still too high and that sonography should be used to rule the diagnosis in or out.²⁵ As mentioned above, another problem with this study is the small sample size, which resulted in wide confidence intervals.

Conclusions

Prompted by a colleague's experience with the Ottawa ankle rule and the CAGE questionnaire, we examined diagnostic test evaluation studies, which on websites and in a textbook of evidence based medicine were interpreted as demonstrating tests' ability to conclusively rule a diagnosis in or out (SpPIn or SnNOut tests). We calculated likelihood ratios to measure the tests' power to rule the target conditions in or out and assessed the study designs for possible bias, and we have given examples of tests where these methodological issues raise questions about whether the tests have SpPIn or SnNOut properties.

We believe that the concept of SpPIn or SnNOut tests can help clinicians in interpreting diagnostic test results. However, identifying and promoting tests with SpPIn or SnNOut properties should be based on a careful appraisal of the evidence, including the methodological quality of the test evaluation studies, and not simply on the test's sensitivity or specificity. Likelihood ratios and typical post-test probabilities should be calculated and reported, together with measures of statistical uncertainty, as recommended in the Standards for Reporting Diagnostic Accuracy Studies (STARD).²⁶ Assessments should ideally be based on a systematic review of all available studies, which may include a meta-analysis to increase the



Ability of a test to rule in disease (measured by the likelihood ratio of a positive test) as a function of specificity and sensitivity (upper panel), and ability of a test to rule out disease (measured by the inverse of the likelihood ratio of a negative test result) as a function of sensitivity and specificity (lower panel)

precision of estimates of test accuracy. For example, a recent meta-analysis of 27 test accuracy studies of the Ottawa ankle rules confirmed that in many settings this decision aid can indeed exclude fractures and reduce the number of unnecessary radiographs.¹⁴ The fact that results may not be transferable to other populations and settings should be stressed, and the information required to judge transferability and applicability should be provided. Clearly, assuming that a diagnosis can be ruled in or ruled out with confidence, when in reality it cannot, could have serious consequences for patients.

We thank Nicola Low for helpful comments on an earlier version of this manuscript.

Contributors: DP had the idea of critically appraising test accuracy studies that were interpreted as demonstrating SpPin or SnNOut properties. CM advised on statistical issues. ME and DP wrote the first draft of the article. All authors contributed to the appraisal of the evidence from studies and to writing the final draft of the article.

Funding: The Krankenfürsorgestiftung der Gesellschaft für das Gute und Gemeinnützige (GGG), Basle, Switzerland, and the Swiss Academy of Medical Sciences supported this study.

Competing interests: None declared.

- 1 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.
- 2 Centre for Evidence-Based Medicine, University Health Network, Mount Sinai Hospital Toronto. Sensitivity & specificity (SnNouts and SpPins). www.cebm.utoronto.ca/glossary/spn.htm (accessed March 2004).
- 3 Fagan TJ. Nomogram for Bayes theorem. *N Engl J Med* 1975;293:257.
- 4 Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based medicine. How to practice and teach EBM*. New York: Churchill Livingstone, 2000.
- 5 Barry HC, Ebell MH. Test characteristics and decision rules. *Endocrinol Metab Clin North Am* 1997;26:45-65.
- 6 Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology, a basic science for clinical medicine*. Boston: Little Brown, 1992.
- 7 Schwartz A, Hupert J. Medical students' application of published evidence: randomised trial. *BMJ* 2003;326:536-8.
- 8 Horvath AR, Pewsner D. Systematic reviews in laboratory medicine: principles, processes and practical considerations. *Clin Chim Acta* 2004;342:23-39.

- 9 Auleley GR, Kerboul L, Durieux P, Cosquer M, Courpied JP, Ravaud P. Validation of the Ottawa ankle rules in France: a study in the surgical emergency department of a teaching hospital. *Ann Emerg Med* 1998;32:14-8.
- 10 Simel DL, Halvorsen RA Jr, Feussner JR. Quantitating bedside diagnosis: clinical evaluation of ascites. *J Gen Intern Med* 1988;3:423-8.
- 11 Levin BE. The clinical significance of spontaneous pulsations of the retinal vein. *Arch Neurol* 1978;35:37-40.
- 12 Varma AR, Snowden JS, Lloyd JJ, Talbot PR, Mann DM, Neary D. Evaluation of the NINCDS-ADRDA criteria in the differentiation of Alzheimer's disease and frontotemporal dementia. *J Neurol Neurosurg Psychiatry* 1999;66:184-8.
- 13 Bakker AJ. Detection of microalbuminuria. Receiver operating characteristic curve analysis favors albumin-to-creatinine ratio over albumin concentration. *Diabetes Care* 1999;22:307-13.
- 14 Bachmann LM, Kolb E, Koller MT, Steurer J, ter Riet G. Accuracy of Ottawa ankle rules to exclude fractures of the ankle and mid-foot: systematic review. *BMJ* 2003;326:417.
- 15 McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;34:939-44.
- 16 Guarino JR, Guarino JC. Auscultatory percussion: a simple method to detect pleural effusion. *J Gen Intern Med* 1994;9:71-4.
- 17 Bush B, Shaw S, Delbanco TL, Aronson MD. Screening for alcohol abuse using the CAGE questionnaire. *Am J Med* 1987;82:231-5.
- 18 Marantz PR, Kaplan MC, Alderman MH. Clinical diagnosis of congestive heart failure in patients with acute dyspnea. *Chest* 1990;97:776-81.
- 19 Davie AP, Francis CM, Caruana L, Sutherland GR, McMurray JJ. Assessing diagnosis in heart failure: which features are any use? *QJM* 1997;90:335-9.
- 20 Cherpitel CJ, Clark WB. Ethnic differences in performance of screening instruments for identifying harmful drinking and alcohol dependence in the emergency room. *Alcohol Clin Exp Res* 1995;19:628-34.
- 21 Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004;140:189-202.
- 22 Centre for Evidence Based Medicine, University of Oxford. SpPins and SnNouts. www.cebm.net/sppins_snnouts.asp (accessed January 2003).
- 23 Centro per la formazione permanente e l'aggiornamento del personale del servizio sanitario (CEFPAS), Caltanissetta, Italy. http://www.cefpas.it/ebm/diagnosis/dia_appraising.htm (accessed March 2004).
- 24 National Electronic Library for Health. Full-text guideline collection. Clinical guidelines for type 2 diabetes: renal disease—prevention and early management. www.nelh.nhs.uk/guidelinesdb/html/fulltext-summary/RcgpRenal.html (accessed March 2004).
- 25 Williams JW Jr, Simel DL. The rational clinical examination. Does this patient have ascites? How to divine fluid in the abdomen. *JAMA* 1992;267:2645-8.
- 26 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;326:41-4.

(Accepted 20 April 2004)

A sequence of errors

The weekly 500 mile aeroplane journey to the South African hospital shortened the working day for the visiting radiologist, so a car was always ready to take him straight from the airstrip to the waiting patients. If an examination under general anaesthetic was required, an anaesthetist would start to prepare the patient as soon as the aeroplane's wheels touched the tarmac, so that the radiologist wasted no time at the hospital. The senior registrars from Cape Town, who took it in turns to provide the service, enjoyed the visit; it was good for radiological experience and even better for the ego. My memory of one of my visits remains particularly vivid.

The chauffeur who met me at the airport incorrectly but deferentially addressed me as "Professor" as I climbed into the large limousine. When I arrived at the hospital a patient was already anaesthetised, waiting for me to carry out a carotid angiogram to help diagnose a mass in the neck. I quickly washed my hands and located the carotid pulse; no scanner was available, and direct carotid puncture, occasionally under general anaesthesia, was used. I inserted the angiography needle and obtained a jet of arterial blood, so I connected the needle to a syringe and injected the contrast material, and the radiographs were exposed. However, when these were developed, none of the films showed the carotid circulation. Instead, they all just showed contrast material opacifying both vertebral arteries: I had obviously injected the wrong artery and, in error, performed a vertebral angiogram.

The vertebral arteries run up on each side of the neck deep to the carotids, passing from their origin in the subclavian arteries to the head, where they join to form the basilar artery. Erroneously puncturing a vertebral artery at carotid angiography was unusual. Opacifying the vertebral artery on the opposite side of the neck was totally unexpected. It meant that contrast material had passed up the vertebral artery on the injected side and down the one on the other side.

This retrograde flow indicated that there was an occlusion or stenosis at the origin of the contra lateral subclavian artery, and I was therefore able to diagnose instantly a subclavian steal syndrome on the opposite side of the neck to the injection. Closer clinical examination of the neck revealed that the mass on the ipsilateral side of the neck was due to obvious tuberculous glands, and a carotid angiogram had not been indicated in the first place. In the twinkling of an eye, I had attempted an inappropriate examination, performed an unintended procedure, and diagnosed an unsuspected condition.

Unfortunately I didn't learn of the effect of the diagnosis on the patient's outcome, but when, on the return journey to the airstrip, the chauffeur again incorrectly called me professor the sequence of errors seemed complete.

Brian Witcombe *consultant radiologist, Gloucestershire Royal Hospital, Gloucester*