are successful doctors may then be prepared to invest more time in them. Further consideration should be given to new ways that general practitioners can deliver and follow up lifestyle interventions, including implementing the process of change model and undertaking motivational interviewing. The role of practice nurses, health visitors, and specialist clinics may be important in enhancing a general practitioner's initial intervention.

I am grateful to Dr Peter Anderson for commenting on a draft of this article.

1 Blot W. Alcohol and cancer. *Cancer Res* 1992;**52** (suppl):2119-23S.
2 Cancer Research Campaign. *Factsheet 14: oral cancer.* London: CRC, 1993.
3 Secretary of State for Health. *Health of the Nation: a strategy for health in England.* London: HMSO, 1992. (Cm 1986.)
4 Anderson P. Reducing alcohol consumption. In: Fowler G, Anderson P, Gray M, eds. *Prevention in general practice.* Oxford: Oxford University Press, 1993:136-47.
5 Anderson P. Effectiveness of general practice interventions for patients with harmful alcohol consumption. *Br J Gen Pract* 1993;**43**:386-9.

Selected additional references

Anderson P. Alcohol as a key area. *BMJ* 1991;**303**:766-9.
Anderson P. The interaction between research and policy—alcohol and general practice in the United Kingdom. *Addiction* 1993;**88**:1215-75.
Anderson P. Population and high-risk approaches both necessary to reduce population harm. *Addiction* 1993;**88**:170-4.
Anderson P, Cremona A, Paton A, Turner C, Wallace P. The risk of alcohol. *Addiction* 1993;**88**:1493-508.
Anderson P, Wallace P, Jones H. *Alcohol problems: practical guides for general practice.* Oxford: Oxford University Press, 1988.
Andserson P, Scott E. The effect of general practitioners advice to heavy drinking men. *Br J Addict* 1992;**87**:891-900.
Babor TF, Grant M. *WHO Collaborating Investigators Project on identification and management of alcohol related problems. Combined analyses of outcome data: the cross national generalizability of brief interventions. Report on phase II: a randomised clinical trial of brief interventions in primary care.* Copenhagen: WHO, 1992.
Bien TH, Miller WR, Tonigan JS. Brief interventions for alcohol problems: a review. *Addiction* 1993;**88**:315-36.
Doll R. The prevention of cancer: opportunities and challenges. In: Heller T, Davey B, Bailey L, eds. *Reducing the risk of cancers.* London: Hodder and Stoughton, 1989:14-25.
Effective Health Care. *Brief interventions and alcohol use.* Nuffield Institute for Health, University of Leeds; Centre for Health Economics, University of York; Research Unit, Royal College of Physicians, 1993.
Health Education Authority. *Cut down on your drinking pack.* London: HEA, 1991.
International Agency for Research on Cancer. *IARC monographs on the evaluation of carcinogenic risks to humans: alcohol drinking.* Lyons: IARC, 1988.
Medical Council on Alcoholism. *Hazardous drinking: a handbook for general practitioners.* London: MCA, 1987.
Office of Population Censuses and Surveys. *General household survey 1992.* London: HMSO, 1994.
Prochaska J, DiClemente C. Toward a comprehensive model of change. In: Miller WR, Heather N, eds. *Treating addictive behaviors: processes of change.* New York: Plenum, 1986:3-27.
Rollnick S, Heather N, Gold R, Hall W. Development of a short "readiness to change" questionnaire for use in brief, opportunistic interventions among excessive drinkers. *Br J Addiction* 1992;**87**:743-54.
Royal College of General Practitioners. *Alcohol—a balanced view.* London: RCGP, 1986.
Royal College of General Practitioners. *Alcohol and drugs.* London: RCGP, 1991.
Royal College of Physicians. *The medical consequences of alcohol abuse: a great and growing evil.* London: Tavistock, 1987.
Reid ALA, Webb GR, Hennrikus D, Fahey PP, Sanson-Fisher RW. Detection of patients with high alcohol intake by general practitioners. *BMJ* 1986;**293**: 735-7.
Rose G, Day S. The population mean predicts the number of deviant individuals. *BMJ* 1990;**301**:1031-4.
Saunders JB, Aasland OG, Babor TF, De La Fuente JR, Grant M. Development of the alcohol use disorders identification test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption—II. *Addiction* 1993;**88**:791-804.
Scott E, Anderson P. Randomised controlled trial of general practitioners intervention in women with excessive alcohol consumption. *Drug and Alcohol Review* 1990;**10**:313-2.
Wallace P, Cutler S, Haines A. Randomised controlled trial of general practitioners intervention in patients with excessive alcohol consumption. *BMJ* 1988;**297**:663-8.
Wallace P, Haines A. The use of a questionnaire in general practice to increase the recognition of patients with excessive alcohol consumption. *BMJ* 1985;**290**: 1949-53.

A complete list of references can be obtained from the author.

Statistics Notes

Diagnostic tests 1: sensitivity and specificity

Douglas G Altman, J Martin Bland

This is the third in a series of occasional notes on medical statistics.

The simplest diagnostic test is one where the results of an investigation, such as an *x* ray examination or biopsy, are used to classify patients into two groups according to the presence or absence of a symptom or sign. For example, the table shows the relation between the results of a test, a liver scan, and the correct diagnosis based on either necropsy, biopsy, or surgical inspection.[1] How good is the liver scan at diagnosis of abnormal pathology?

Relation between results of liver scan and correct diagnosis[1]

	Pathology		
Liver scan	Abnormal (+)	Normal (−)	Total
Abnormal (+)	231	32	263
Normal (−)	27	54	81
Total	258	86	344

One approach is to calculate the proportions of patients with normal and abnormal liver scans who are correctly "diagnosed" by the scan. The terms positive and negative are used to refer to the presence or absence of the condition of interest, here abnormal pathology. Thus there are 258 true positives and 86 true negatives. The proportions of these two groups that were correctly diagnosed by the scan were 231/258=0.90 and 54/86=0.63 respectively. These two proportions have confusingly similar names.

Sensitivity is the proportion of true positives that are correctly identified by the test.

Specificity is the proportion of true negatives that are correctly identified by the test.

We can thus say that, based on the sample studied, we would expect 90% of patients with abnormal pathology to have abnormal (positive) liver scans, while 63% of those with normal pathology would have normal (negative) liver scans.

The sensitivity and specificity are proportions, so confidence intervals can be calculated for them using standard methods for proportions.[2]

Sensitivity and specificity are one approach to quantifying the diagnostic ability of the test. In clinical practice, however, the test result is all that is known, so we want to know how good the test is at predicting abnormality. In other words, what proportion of patients with abnormal test results are truly abnormal? This question is addressed in a subsequent note.

Medical Statistics Laboratory, Imperial Cancer Research Fund, London WC2A 3PX
Douglas G Altman, *head*

Department of Public Health Sciences, St George's Hospital Medical School, London SW17 0RE
J Martin Bland, *reader in medical statistics*

BMJ 1994;**308**:1552

1 Drum DE, Christacapoulos JS. Hepatic scintigraphy in clinical decision making. *J Nucl Med* 1972;**13**:908-15.
2 Gardner MJ, Altman DG. Calculating confidence intervals for proportions and their differences. In: Gardner MJ, Altman DG, eds. *Statistics with confidence.* London: BMJ Publishing Group, 1989: 28-33.

Soft data, as outlined in this article, are present on the ground in plenty. We have endeavoured, using our growing skills, to do responsibly what we can. We have built a relationship of trust with our family health services authority. But trust alone will not be enough in the future, and we have to work to find methods of critically evaluating the benefits of the money spent. These evaluation techniques will need to look at the work that the medical audit advisory group has done and will need to find measures that accurately reflect the outcome of such work. It is unlikely that simply counting audits done in general practice will provide an adequate measure of the changes that have been brought about.

I thank Manchester general practitioners and their teams and the medical audit advisory group members and staff for making it possible to record so much progress.

1 Maynard A. Ivory tower: that's why the lady is a wimp. *Health Service Journal* 1993;103 (28 Jan):21.
2 Drummond MF. *Principles of economic appraisal in health care.* Oxford: Oxford University Press, 1980:116-9.
3 Baker R. Audit and standards in new general practice. *BMJ* 1991;303:32-4.
4 Grol R, Mokkink H, Schellevis F. The effects of peer review in general practice. *J R Coll Gen Pract* 1988;38:10-3.
5 Hughes J, Humphrey C. *Medical audit in general practice.* London: King's Fund, 1990.
6 Johnson R. Summarising GP records. *Audit Trends* 1993;1 (June):756-7.
7 Foxall M. Organising large-scale multi-practice audit projects. *Audit Trends* 1993;1 (September):99-101.
8 Benett IJ, Lambert C, Hinds G, Kirton C. Emerging standards for diabetes care from a city-wide primary care audit. *Diabetic Med* (in press).
9 Humphrey C, Berrow D. Developing role of medical audit advisory groups. *Quality in Health Care* 1993;2:232-8.
10 Department of Health. *Health service developments. Working for Patients: medical audit in the family practitioner services.* London: DoH, 1990. (HC(FP)90(8).
11 House of Commons Health Committee. *Second report, 1991-92: matenity services.* London: HMSO, 1992. (Winterton Committee).
12 Department of Health. Expert Maternity Committee. *Changing Childbirth.* London: HMSO, 1993.
13 Royal College of General Practitioners. *What sort of doctor?.* London: Royal College of General Practitioners, 1985.
14 Derry J, Lawrence M, Griew K, Anderson J, Humphrey J, Pandher KS. Auditing audits: the method of Oxfordshire Medical Audit Advisory Group. *BMJ* 1991;303:1247-9.
15 Baker R. Use of psychometrics to develop a measure of patient satisfaction for general practice. In: Hopkins A, Fitzpatrick R, eds. *Measurement of patients' satisfaction with their care.* London: Royal College of Physicians, 1993.
16 General Medical Services Committee Audit Subcommittee. *The working of MAAG.* London: BMA, 1991.

(Accepted 27 April 1994)

Statistics Notes

Diagnostic tests 2: predictive values

Douglas G Altman, J Martin Bland

This is the fourth in a series of occasional notes on medical statistics.

The whole point of a diagnostic test is to use it to make a diagnosis, so we need to know the probability that the test will give the correct diagnosis. The sensitivity and specificity[1] do not give us this information. Instead we must approach the data from the direction of the test results,using predictive values.

Positive predictive value is the proportion of patients with positive test results who are correctly diagnosed.

Negative predictive value is the proportion of patients with negative test results who are correctly diagnosed.

Using the same data as in the previous note,[1] we know that 231 of 263 patients with abnormal liver scans had abnormal pathology, giving the proportion of correct diagnoses as 231/263=0·88. Similarly, among the 81 patients with normal liver scans the proportion of correct diagnoses was 54/81=0·59. These proportions are of only limited validity, however. The predictive values of a test in clinical practice depend critically on the prevalence of the abnormality in the patients being tested; this may well differ from the prevalence in a published study assessing the usefulness of the test.

In the liver scan study the prevalence of abnormality was 0·75. If the same test was used in a different clinical setting where the prevalence of abnormality was 0·25 we would have a positive predictive value of 0·45 and a negative predictive value of 0·95. The rarer the abnormality the more sure we can be that a negative test indicates no abnormality, and the less sure that a positive result really indicates an abnormality. Predictive values observed in one study do not apply universally.

The positive and negative predictive values (PPV and NPV) can be calculated for any prevalence as follows:

$$PPV=\frac{\text{sensitivity}\times\text{prevalence}}{\text{sensitivity}\times\text{prevalence}+(1-\text{specificity})\times(1-\text{prevalence})}$$

$$NPV=\frac{\text{specificity}\times(1-\text{prevalence})}{(1-\text{sensitivity})\times\text{prevalence}+\text{specificity}\times(1-\text{prevalence})}$$

If the prevalence of the disease is very low, the positive predictive value will not be close to 1 even if both the sensitivity and specificity are high. Thus in screening the general population it is inevitable that many people with positive test results will be false positives.

The prevalence can be interpreted as the probability before the test is carried out that the subject has the disease, known as the prior probability of disease. The positive and negative predictive values are the revised estimates of the same probability for those subjects who are positive and negative on the test, and are known as posterior probabilities. The difference between the prior and posterior probabilities is one way of assessing the usefulness of the test.

For any test result we can compare the probability of getting that result if the patient truly had the condition of interest with the corresponding probability if he or she were healthy. The ratio of these probabilities is called the *likelihood ratio*, calculated as sensitivity/(1−specificity).

The likelihood ratio indicates the value of the test for increasing certainty about a positive diagnosis. For the liver scan data the prevalence of abnormal pathology was 0·75, so the pre-test odds of disease were 0·75/(1−0·75)=3·0. The sensitivity was 0·895 and the specificity was 0·628. The post-test odds of disease given a positive test is 0·878/(1−0·878)=7·22, and the likelihood ratio is 0·895/(1−0·628)=2·41. The post-test odds of having the disease is the pre-test odds multiplied by the likelihood ratio.

A high likelihood ratio may show that the test is useful, but it does not necessarily follow that a positive test is a good indicator of the presence of disease.

1 Altman DG, Bland JM. Diagnostic tests 1: sensitivity and specificity. *BMJ* 1994;000:00000.

Medical Statistics Laboratory, Imperial Cancer Research Fund, London WC2A 3PX
Douglas G Altman, *head*

Department of Public Health Sciences, St George's Hospital Medical School, London SW17 0RE
J Martin Bland, *reader in medical statistics*

BMJ 1994;309:102

Correction

Diet and cancer

A printer's error occurred in this article, the fourth in the series on cancer prevention in primary care by Joan Austoker (18 June, pp 1610-4). In table I the first mention of a possibly increased risk of cancer associated with increased fat intake (in the second column) should have applied to breast cancer and not to lung cancer as published.

purport to offer a method of dealing with conflicts between the principles. But I have not found anyone who seriously argues that he or she cannot accept any of these prima facie principles or found plausible examples of concerns about health care ethics that require additional moral principles.

The four principles plus scope approach enables health care workers from totally disparate moral cultures to share a fairly basic, common moral commitment, common moral language, and common analytical framework for reflecting on problems in health care ethics. Such an approach, which is neutral between competing religious, political, cultural, and philosophical theories, can be shared by everyone regardless of their background. It is surely too important a moral prize to be rejected carelessly or ignorantly; for the sake of mere opposition; or for the fun of being a philosophical "Socratic gadfly."

1 Gillon R. *Philosophical medical ethics.* Chichester: Wiley, 1986. (From a 26 part series in *BMJ* from 1985;**290**:1117-9 to 1986;**292**:543-5.)
2 Beauchamp TL, Childress JF. *Principles of biomedical ethics.* 3rd ed. New York, Oxford: Oxford University Press, 1989.
3 Gillon R, Lloyd A, eds. *Principles of health care ethics.* Chichester: Wiley, 1994.
4 Aristotle. Nichomachean ethics. Book 5. McKeon R, ed. *The basic works of Aristotle.* New York: Random House, 1941.
5 Aristotle. Politics. Book 3, chapter 9. McKeon R, ed. *The basic works of Aristotle.* New York: Random House, 1941.
6 Klein R. On the Oregon trail: rationing health care—more politics than science. *BMJ* 1991;**302**:1-2.
7 Williams A. Economics, society and health care ethics. In: Gillon R, Lloyd A, eds. *Principles of health care ethics.* Chichester: Wiley, 1994:829-42.
8 Hunter DJ. *Rationing dilemmas in health care.* Birmingham: National Association of Health Authorities and Trusts, 1993. (NAHAT research paper No 8.)
9 Calabresi G, Bobbitt P. *Tragic choices.* New York: Norton, 1978.

(Accepted 16 March 1994)

Statistics Notes

Diagnostic tests 3: receiver operating characteristic plots

Douglas G Altman, J Martin Bland

This is the seventh in a series of occasional notes on medical statistics.

We have previously considered diagnosis based on tests that give a yes or no answer.[1,2] Many diagnostic tests, however, are quantitative, notably in clinical chemistry. The same statistical approach can be used only if we can select a cut off point to distinguish "normal" from "abnormal," which is not a trivial problem. Firstly, we can investigate to what extent the test results differ among people who do or do not have the diagnosis of interest. The receiver operating characteristic (ROC) plot is one way to do this. These plots were developed in the 1950s for evaluating radar signal detection. Only recently have they become commonly used in medicine.

We assume that high values are more likely among those dubbed "abnormal." Figure 1 shows the values of an index of mixed epidermal cell lymphocyte reactions in bone marrow transplant recipients who did or did not develop graft versus host disease.[3] The usefulness of the test for predicting graft versus host disease will clearly relate to the degree of non- overlap between the two distributions.

A receiver operating characteristic plot is obtained by calculating the sensitivity and specificity of every observed data value and plotting sensitivity against 1—specificity, as in Figure 2. A test that perfectly discriminates between the two groups would yield a "curve" that coincided with the left and top sides of the plot. A test that is completely useless would give a straight line from the bottom left corner to the top right corner. In practice there is virtually always some overlap of the values in the two groups, so the curve will lie somewhere between these extremes.

A global assessment of the performance of the test (sometimes called diagnostic accuracy[4]) is given by the area under the receiver operating characteristic curve. This area is equal to the probability that a random person with the disease has a higher value of the measurement than a random person without the disease. (This probability is a half for an uninformative test—equivalent to tossing a coin.)

No test will be clinically useful if it cannot discriminate,[4] so a global assessment of discriminatory power is an important step. Having determined that a test does provide good discrimination the choice can be made of the best cut off point for clinical use. This requires the choice of a particular point, and is thus a local assessment. The simple approach of minimising "errors" (equivalent to maximising the sum of the sensitivity and specificity) is not necessarily best. Consideration needs to be given to the costs (not just financial) of false negative and false positive diagnoses and to the prevalence of the disease in the subjects being tested.[4] For example, when screening the general population for cancer the cut off point would be chosen to ensure that most cases were detected (high sensitivity) at the cost of many false positives (low specificity), who could then be eliminated by a further test.

A receiver operating characteristic plot is particularly useful when comparing two or more measures. A test with a curve that lies wholly above the curve of another will be clearly better. Methods for comparing the areas under two curves for both paired and unpaired data are reviewed by Zweig and Campbell,[4] who give a full assessment of this method.

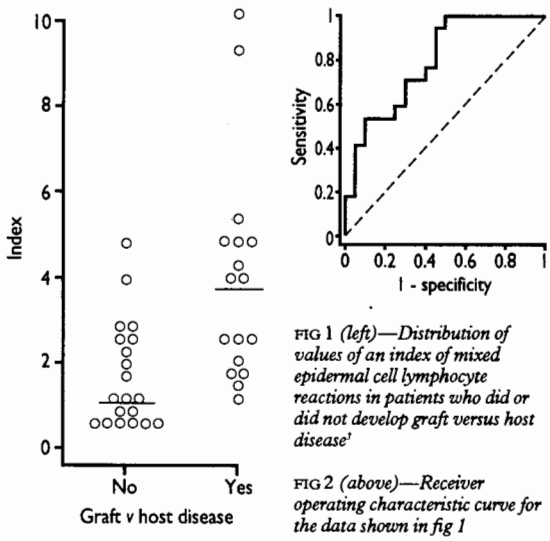

FIG 1 *(left)—Distribution of values of an index of mixed epidermal cell lymphocyte reactions in patients who did or did not develop graft versus host disease*[3]

FIG 2 *(above)—Receiver operating characteristic curve for the data shown in fig 1*

Medical Statistics Laboratory, Imperial Cancer Research Fund, London WC2A 3PX
Douglas G Altman, *head*

Department of Public Health Sciences, St George's Hospital Medical School, London SW17 1RE
J Martin Bland, *reader in medical statistics*

BMJ 1994;**309**:188

1 Altman DG, Bland M. Diagnostic tests 1: sensitivity and specificity. *BMJ* 1994;**308**:1552.
2 Altman DG, Bland M. Diagnostic tests 2: predictive values. *BMJ* 1994;**309**: 102.
3 Bagot M, Mary J-Y, Heslan M, *et al.* The mixed epidermal cell lymphocyte—reaction is the most predictive factor of acute graft- versus-host disease in bone marrow graft recipients. *Br J Haematol* 1988;**70**:403-9.
4 Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;**39**: 561-77.