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.

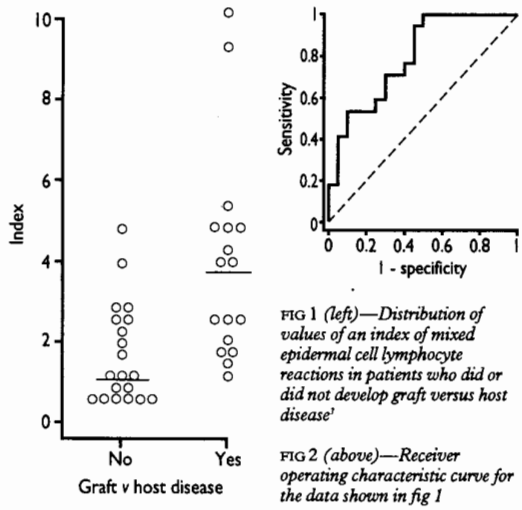

FIG 1 *(left)—Distribution of values of an index of mixed epidermal cell lymphocyte reactions in patients who did or did not develop graft versus host disease*[3]

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.