

hormone replacement therapy have been analysed in many different studies, each concerned with different outcomes and with each outcome being valued differently.¹⁸ Furthermore, observational studies have shown that women have different risks at baseline (for example, thin women are at higher risk of fractures). Decision analysis has shown how these factors may be integrated to optimise individual treatment.¹⁹

Decision analysis is used to determine how to maximise an individual's expected utilities. By obtaining the median values of utilities from a large number of people the methodology can also be used to derive expected utilities for the community. When the costs of various options are included this is called a cost utility analysis. However, using decision analysis to make decisions for groups of patients creates some thorny ethical issues, especially when there is a conflict between maximising utility and maximising equity.

Conclusions

Decision analysis depends on probabilities and values, neither of which can be measured with certainty. These problems are not lessened when health professionals approach them intuitively; decision analysis makes these uncertainties explicit. The attempt to make complex decisions intuitively inevitably results in gross oversimplifications because it is impossible to incorporate and consider several components of a decision simultaneously. There is a large amount of empirical literature on the limitations of intuitive reasoning that is summarised by Dawes et al.²⁰

Most research findings are applied unsystematically and intuitively. If evidence based medicine is to be seen through to its logical conclusion and if both empirical evidence and human values are to be incorporated into decision making, then this duality (the explicit collection of data *v* its implicit use) must be addressed.^{21 22}

- 1 Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. *BMJ* 1995;311:1356-9.
- 2 Keeney RL, Raiffa H. *Decisions with multiple objectives*. London: Wiley, 1976.
- 3 Weinstein M, Fineberg HV. *Clinical decision analysis*. London: Saunders, 1980.
- 4 French S. *Readings in decision analysis*. London: Chapman and Hall, 1989.
- 5 Sox HC, Blatt MA, Higgins MC, Marton KI. *Medical decision making*. Boston: Butterworth Heinemann, 1988.
- 6 Second International Study of Infarct Survival (ISIS 2) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17 187 cases of suspected myocardial infarction: ISIS 2. *Lancet* 1988;ii:349-60.
- 7 Kellett J, Clarke J. Comparison of accelerated tissue plasminogen activator with streptokinase for treatment of suspected myocardial infarction. *Med Decis Making* 1995;15:297-310.
- 8 Tsevat J, Goldman L, Lamas GA. Functional status versus utilities in survivors of myocardial infarction. *Med Care* 1991;29:1153-9.
- 9 Lilford RJ, Braunholtz D. The statistical basis of public policy: a paradigm shift is overdue. *BMJ* 1996;313:603-7.
- 10 Donner A. A bayesian approach to the interpretation of sub-group results in clinical trials. *J Chronic Dis* 1982;34:429-35.
- 11 Oxman AD, Guyatt GH. A consumer's guide to sub-group analyses. *Ann Intern Med* 1992;116:78-84.
- 12 Swales J. Science in a health service. *Lancet* 1997;349:1319-21.
- 13 Lilford RJ. Trade-off between gestational age and miscarriage risk of pre-natal testing: does it vary according to genetic risk? *Lancet* 1990;336:1303-5.
- 14 Johnson N, Lilford R, Jones SE, McKenzie L, Billingsley P, Songane FF. Decision analysis to calculate optimum treatment for microinvasive cervical cancer. *Br J Cancer* 1992;65:717-22.
- 15 Lilford RJ, Jackson J. Equipoise and the ethics of randomisation. *J R Soc Med* 1995;88:552-9.
- 16 Pauker SP, Pauker SG. Prenatal diagnosis: a directive approach to genetic counselling using decision analysis. *Yale J Biol Med* 1977;50:275-89.
- 17 Sculpher M, Michaels J, McKenna M, Minor J. A cost-utility analysis of laser-assisted angioplasty for peripheral arterial occlusions. *Int J Technol Assess Health Care* 1996;12:104-25.
- 18 Johnson N, Lilford RJ, Mayers D, Johnson GC, Johnson JM. Do healthy asymptomatic post menopausal women want routine cyclical hormone replacement? A utility analysis. *J Obstet Gynecol* 1994;14:35-9.
- 19 Col NF, Eckman MH, Kara RH, Pauker SG, Goldberg RJ, Ross EM, et al. Patient-specific decisions about hormone replacement therapy in postmenopausal women. *JAMA* 1991;277:1140-7.
- 20 Dawes R, Faust D, Meehl PE. Clinical versus actuarial judgement. *Science* 1989;243:1668-74.
- 21 Dowie J. "Evidence-based," "cost-effective" and "preference-driven" medicine: decision analysis based medical decision making is the pre-requisite. *J Health Serv Res Policy* 1996;1:104-13.
- 22 Dowie J. The research-practice gap and the role of decision analysis in closing it. *Health Care Analysis* 1996;4:5-18.

The articles in this series are adapted from *Getting Research Findings into Practice*, edited by Andrew Haines and Anna Donald, and published by BMJ Books.

Statistics Notes

Generalisation and extrapolation

Douglas G Altman, J Martin Bland

All medical research is carried out on selected individuals, although the selection criteria are not always clear. The usefulness of research lies primarily in the generalisation of the findings rather than in the information gained about those particular individuals. We study the patients in a trial not to find out anything about them but to predict what might happen to future patients given these treatments.

A recent randomised trial showed no benefit of fine needle aspiration over expectant management in women with simple ovarian cysts.¹ The clinical question is whether the results can be deemed to apply to a given patient. For most conditions it is widely accepted that a finding like this validly predicts the effect of treatment in other hospitals and in other countries. It would not, however, be safe to make predictions about patients with another condition,

such as a breast lump. In between these extremes lie some cases where generalisability is less clear.

For example, when trials showed the benefits of β blockers after myocardial infarction the studies had been carried out on middle aged men. Could the findings reasonably be extrapolated to women, or to older men? It is probably rare that treatment effectiveness truly varies by sex, and claims of this kind often arise from faulty subgroup analysis.² Age too rarely seems to affect the benefit of a treatment, but clinical characteristics certainly do. Treatments that work in mild disease may not be equally effective in patients with severe disease, or vice versa. Likewise the mode of delivery—for example, oral versus subcutaneous—or dose may affect treatment benefit. Clinical variation is likely to affect the size of benefit of a treatment, not whether any benefit exists.

ICRF Medical Statistics Group, Centre for Statistics in Medicine, Institute of Health Sciences, Oxford OX3 7LF

Douglas G Altman, head

Department of Public Health Sciences, St George's Hospital Medical School, London SW17 0RE

J Martin Bland, professor of medical statistics

Correspondence to: Mr Altman.

BMJ 1998;317:409-10

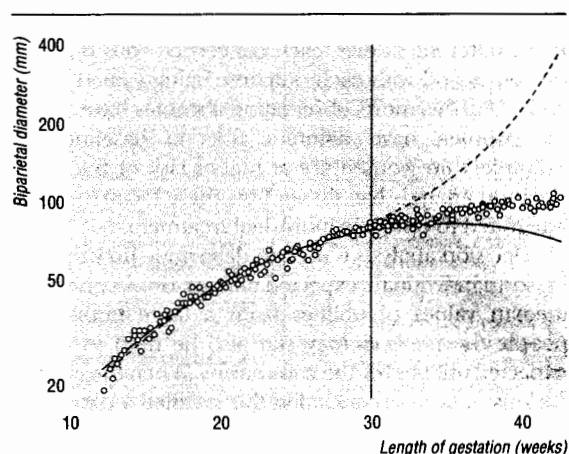
The extent to which it is wise or safe to generalise must be judged in individual circumstances, and there may not be a consensus. Arguably many studies (especially randomised controlled trials) use over-restrictive inclusion criteria, so that the degree of safe generalisability is reduced.³ Even geographical generalisation may sometimes be unwarranted. For example, BCG vaccination against tuberculosis is much less effective in India than in Europe, probably because of greater exposure in India.⁴ For the clinician treating a patient the question can be expressed as: "Is my patient so different from those in the trial that its results cannot help me make my treatment decision?"⁵

In a clinical trial we are interested in the difference in effectiveness between two treatments. There is no need to generalise the success rate of a particular treatment. In some other types of research, such as surveys to establish prevalence and prognostic or diagnostic studies, we may be trying to estimate a single population value rather than the difference between two of them. Here generalisation may be less safe. For example, the prevalence of many diseases varies across social and geographical groups. Results may not even hold up across time. For example, changes in case mix over time can affect the properties of a diagnostic test.⁶

Many studies use regression analysis to derive a model for predicting an outcome from one or more explanatory variables. The model, represented by an equation, is strictly valid only within the range of the observed data on the explanatory variable(s). When a measurement is included in the regression model it is possible to make predictions for patients outside the range of the original data (perhaps inadvertently). This numerical form of generalisation is called extrapolation. It can be seriously misleading.

To take an extreme example, a linear relation was found between ear size and age in men aged 30 to 93, with ear length (in mm) estimated as $55.9 + 0.22 \times \text{age}$ in years.⁷ The value of 55.9 corresponds to an age of zero. A baby with ears 5.6 cm long would look like Dumbo.

Extrapolating may be especially dangerous when a curved relation is found. Figure 1 shows fetal biparietal diameter (on a log scale) in relation to gestational age. Also shown are quadratic and cubic models fitted to the log biparietal diameter measurements from only those fetuses less than 30 weeks' gestation. Both curves fit the data well up to 30 weeks, but both give highly misleading predictions thereafter. The quadratic model shows a



Fetal biparietal diameter (on log scale) in relation to gestational age⁸ with quadratic (solid line) and cubic (broken line) regression models fitted to data from only those fetuses less than 30 weeks' gestation (n=119)

spurious maximum at around 34 weeks, while the cubic curve takes us again into elephantine regions.

When we have two explanatory variables it will not usually be apparent (unless we examine a scatter diagram) when a patient has a combination of characteristics which do not fall within the span of the original data set. With more than two variables, such as in many prognostic models, it is not possible to be sure that the original data included any patients with the combination of values of a new patient. Nevertheless, it is reasonable to use such models to make predictions for patients whose important characteristics are within the range in the original data.

Clearly patient characteristics, including the criteria for sample selection, need to be fully reported in medical papers. Yet such basic information is not always provided.

- 1 Zanetta G, Lissoni A, Torri V, Dalle Valle C, Trio D, Rangoni G, Mangioni C. Role of puncture and aspiration in expectant management of simple ovarian cysts: a randomised study. *BMJ* 1996;31:1110-3.
- 2 Matthews JNS, Altman DG. Interaction 1: Heterogeneity of effects. *BMJ* 1996;313:486.
- 3 Ellenberg JH. Selection bias in observational and experimental studies. *Statistics in Medicine* 1994;13:557-67.
- 4 Anonymous. BCG: bad news from India. *Lancet* 1980;i:73-4.
- 5 Sackett DL, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based medicine. How to practice and teach EBM*. London: Churchill-Livingstone, 1997:167.
- 6 Begg CB. Biases in the assessment of diagnostic tests. *Statistics in Medicine* 1987;6:411-23.
- 7 Heathcote JA. Why do old men have big ears? *BMJ* 1995;311:1668.
- 8 Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 2. Head measurements. *Br J Obstet Gynaecol* 1994;101:35-43.

A memorable meeting

A view from the man in the seat opposite

The train journey, the pile of papers, and the *BMJ*. The man opposite catches my eye.

"I see you're a doctor."

Self-aiming at closure: "Yes, I've a lot of work to get through."

"I'm on my way back from London. I was in Harley Street being interviewed as a volunteer for a drug trial of antidepressants. I had to use some 'kidology'."

"Yes?"

"I'm not depressed but they pay better in London than they do in Scotland or Manchester but I don't think I succeeded. The doctor advised me to use cognitive therapy. They said they would pay my expenses by cheque but eventually agreed to give me cash."

"Oh?"

"Yes, the trials are advertised, the best pay about £100 a day to volunteers. For a 20 day trial that's £2000. The worst trial was when I had to be woken up every hour to do mental tests, but usually it's like being on a health farm."

"What about making sure you don't come to harm?"

"Oh they have a committee of vicars and lawyers to decide it's all right, and it's nice to see your regular friends."

My train drew into the station; I was no longer irritated at the interruption.

Robert Boyd, *principal, St George's Hospital Medical School*