Decision logic in medical practice

The Milroy Lecture 1992

The heart has its reasons which reason does not know. Pascal

The craft of medicine

The idea of using formal logic to solve medical problems is anathema to many doctors; it is, after all, an intrusion to the mystical aura which shrouds our profession. In this Milroy lecture, I will be arguing that magic is as important in medical practice today as it ever was.

My first article, written as a medical student, was titled 'The doctor's role in terms of fundamental human needs' [1]. Here I argued that humans have a need of doctors which goes far beyond the purely physical aspects of disease—that we are all 'witch doctors'. Balint is associated, perhaps more than anyone, with the argument that clinicians should be psychotherapists and that a doctor's personality was a treatment in its own right—he referred to 'the drug-doctor' [2]. This may seem like a paradoxical introduction to a talk on formal mathematical logic in patient care. However, the difference in medical practice between our century and those that came before is that medicine is no longer virtually *all* magic. We now have at our disposal a repertoire of powerful interventions, with immense capacity for good and harm. Thus the question is not, as someone once put it to me, 'whether the whole of medicine can be reduced to a flow diagram'. It is readily conceded that most of the craft of medicine will always be an intuitive process—concerned as it is with the magic of our profession. Decision logic is concerned with the diagnosis of well-characterised diseases and in the choice between potentially dangerous and costly treatments. There are a number of levels at which formal methods may contribute to clinical decision-making. At the lowest level structured questionnaires ensure complete data collection and can incorporate checklists to act as an *aide-mémoire*. At a higher level, expert systems can improve the accuracy of diagnosis. This contribution will be concerned with a yet further extension of formal logic in medical practice—clinical decision analysis. First this method must be placed in context by considering the other formal methods.

RICHARD J. LILFORD, PhD, MRCOG, MRCP
Professor of Obstetrics and Gynaecology, and Chairman, Institute of Epidemiology and Health Services Research, University of Leeds

In collaboration with
J. D. THORNTON, MD, DTM&H, MRCOG
Senior Lecturer/Honorary Consultant of Obstetrics and Gynaecology, Institute of Epidemiology and Health Services Research, University of Leeds

Structured patient histories

It has been shown that questionnaires greatly enhance the quantity and quality of medical information in many clinical settings [3]. Furthermore, checklists of clinical actions improve the quality of medical care. This was originally demonstrated with respect to prompts provided by computer in a clinic for the treatment of hypertension [4]. It has subsequently been shown that computer prompts and paper checklists are equally effective but that both are superior to unstructured methods [5]. However, it should be said that structured history-taking systems, particularly those implemented by computer, are not always better than unstructured methods. Thus, while such systems have been successful in obtaining well-defined data sets, such as those required for antenatal care, infertility, and preparation for surgery, they have been much less successful when expanded to incorporate general outpatient clinics, such as the general medical [6] and gynaecology clinics [7]. In these settings, the structured questionnaire is called upon to emulate an essentially intuitive process—part of the craft of medicine which cannot be formalised. Dreyfus *et al* have argued convincingly in their influential book *Mind over machine* [8] that there will always be subject areas where intuition is better than an attempt at formal logic, and general outpatient histories are an example of this. Thus the attempt to take all medical histories by computer results in output where trivial symptoms are over-emphasised and where neither patient nor doctor feels that the computer has got to the nub of the problem.

Diagnostic systems

The next level of formal methods in medicine is the use of what may broadly be defined as 'expert systems'. These can be classified as: (a) those which try to emulate human reasoning (there are many variations on this theme with titles such as 'algorithmic' and 'pro-

duction rule'); (b) statistical; and (c) parallel processing (neural networks). By far the most thoroughly evaluated of these various systems are the statistical methods based on Bayes' theorem. Prominent among these Bayesian systems is a programme of automated diagnosis of abdominal pain, developed by Professor de Dombal and colleagues [9]. This is the only method which has been evaluated in a large, multicentre, randomised trial. This showed an improvement in the overall accuracy of the diagnosis of abdominal pain along with a reduction in both the number of false negative laparotomies and the incidence of delayed treatment of genuine appendicitis. Although these Bayesian systems have been applied to other areas of medical practice, such as the diagnosis of jaundice [10], the method has not been widely adopted by practising clinicians. There are two possible explanations for this. The first is that the excellent results which have been attained are specific to a small number of medical conditions and that these good results are not generalisable. The second possibility is that clinicians are not yet willing to use this technology, despite its documented benefits. The latter proposition is supported by the observation that Bayesian systems are not used widely even within the specialties (such as abdominal surgery) where their benefits have been most convincingly demonstrated. There are many examples of clear technical and scientific advances which have been translated into clinical action only after many years. A classic example is the discovery of antiseptics in the last century. In contrast, other advances are translated immediately into clinical practice, and a good example is X-rays which were in routine use within two years of their discovery. The social invention, whereby expert systems are likely to become more widely accepted, may have to await the development of comprehensive clinical information systems within hospitals and the resulting cultural change which will familiarise clinicians with direct interaction with computers.

Clinical decision analysis

Decision analysis represents a further development of formal logic in medical practice. It is the most threatening as it seeks to replace, or at least augment, clinical judgement by means of a mathematical model. It represents a fundamental, philosophical challenge to instinctive decision-making and, like all disciplines, it encompasses a language which enables its concepts to be communicated efficiently from one person to another. Decision analysis is concerned not only with the probabilities of various outcomes but also with how these are valued. It therefore contains within its framework a mathematical expression of the best treatment for a particular patient. This is called 'the treatment with the greatest expected utility', and decision analysis is therefore sometimes referred to as 'expected utility theory'.

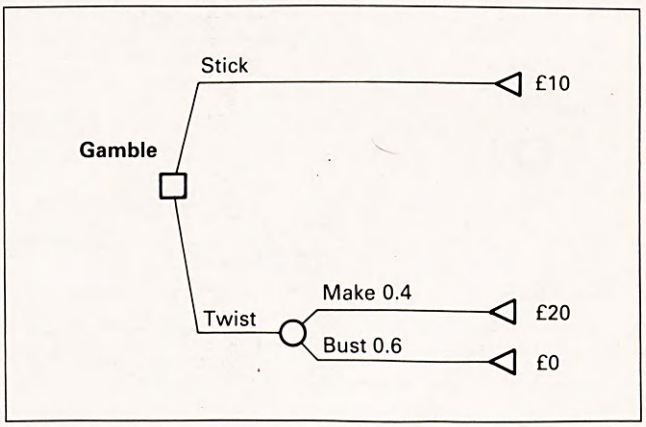
Clinical decision analysis—how to do it

Example 1

A person who has played 21 (*vingt-et-un,* pontoon, blackjack) may have wondered, when dealt, say, cards worth ten and seven, whether to 'stick' (remain with their two cards) or 'twist' (buy another card). At any particular stage of the game (eg one opponent who has not yet played) the probability of victory with a score of 17 is known to the expert. Drawing another card (twisting) may improve the odds of success but runs the risk of pushing the score over 21 (going 'bust'), with immediate loss of the game. Twisting may increase the stake and the potential winnings in some versions of the game. Let us imagine a situation where the gambler has a £10 stake which can be doubled at a 60% risk of going bust. It is obvious that the gambler should stick, but the decision diagram and arithmetic are shown in Fig. 1. The 'expected earnings' drop from £10 with the decision to stick to £8 if the gambler decides to 'twist'. Gamblers perform better if they calculate the odds and combine them correctly with the possible winnings. Decision analysis is not needed in this trivial example, but virtually all real-life situations are much more complicated and the expected earnings cannot be calculated in the head.

Decision analysis is explicit and quantitative. It is explicit in that it forces the decision-maker to spell out the way decisions have been broken down into their component parts and then recombined. It is quantitative in that decision-makers are compelled to measure both key uncertainties and the values of possible outcomes. It is also prescriptive in that it aims to tell physicians what to do, not just describe what they do. There are four basic steps in a decision analysis:

Fig. 1. *Decision diagram for a gambler who may retain a £10 'stake' by 'sticking' or run a 60% risk of losing the stake versus a 40% chance of doubling it by 'twisting'.* The expected earnings of the first choice are (10 × 1) or £10 while those of the bottom are (20 × 0.4) and (0 × 0.6) or £8. If this situation were repeated a very large number of times, the gambler would lose £2



1 Identify and bound the decision problem.
2 Structure the decision problem over time.
3 Measure the uncertainties (probabilities and utilities) needed to fill in the structure.
4 Combine the uncertainties to choose a preferred course of action.

How about some more realistic examples to explain classical decision analysis?

Example 2

Consider a 'wildcatter' or oil prospector who must decide whether or not to take up an 'option' to drill at a particular site. Before deciding, a test may be performed to obtain more information about the seismic structure of the potential drilling site (in effect a diagnostic test). However, the test is expensive to perform and does not provide perfect information (it is neither completely sensitive nor specific). Wildcatters cannot afford to make many mistakes if they wish to stay in business: like gamblers, they must maximise their chances of success.

The canny wildcatter will begin by defining the problem. The choices lie between abandoning the 'option' on the basis of what is already known about it, drilling, or carrying out the seismic test. The seismic test will show whether the terrain below has no discernible structure (bad), an open structure (so-so), or a closed structure (very hopeful). The site might be dry, contain worthwhile deposits, or it might be 'soaking'. Furthermore, drilling costs may be high, medium, or low.

The prospector can now construct the decision tree (Fig. 2). This is a flow diagram in which decisions and outcomes are represented in order, with early events to the left and later events to the right. Decision points are represented by square nodes and points where outcomes occur by chance by round nodes. In Fig. 2 the left-hand decision node represents the choice between doing the seismic test and deciding without this information. The lower circle is a chance node representing the chance that the seismic test will give the various possible results.

Before the expected earnings associated with the various choices can be calculated, it is necessary to know the probabilities of the chance events, along with the conditional probabilities (eg the probability that the site will be 'wet', given an 'open' seismic test result) and the net financial returns on oil sales. All possible outcomes are included so that the probabilities at each chance node always add up to 1.

Calculation of the expected earnings is now straightforward using the mathematics shown in the first example. Numerous software packages are available for the construction of decision trees and the resulting calculations. Those planning a new career in the oil business may wish to refer to the full worked example of the prospector's problem in Raiffa's classical book, *Decision analysis* [11].

Example 3

Most high-earning professional occupations involve the responsibility for making decisions where much is at stake. Physicians will be quick to recognise similarities between many medical decisions and the problem faced by the wildcatter. Consider the management of progressive hepato-cellular jaundice which may be caused by either chronic hepatitis or cirrhosis. Steroids increase the two-year survival in the former condition

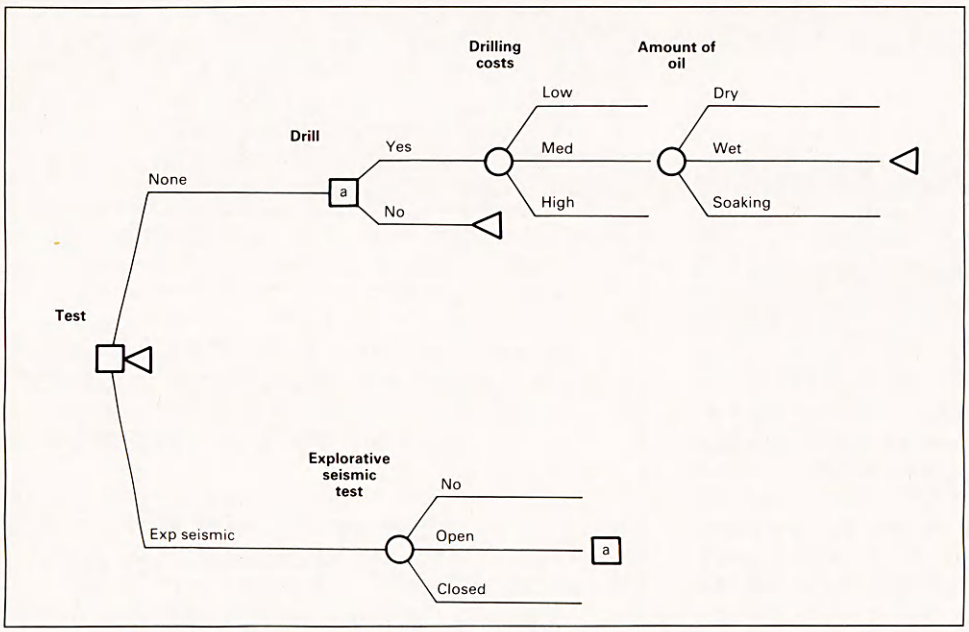


Fig. 2. *The basic decision for the oil prospector is the same as that for a doctor in many clinical situations: to obtain the result of a special test which is 'costly' and not completely accurate, or to decide to manage without this extra information*

from 67% to 85%, but they are ineffective against cirrhosis, reducing the two-year survival from 50% to 48% because they increase the risk of complications, such as gastric bleeding and thrombosis. We will assume that liver biopsy discriminates accurately between the two conditions but that it carries a mortality in these sick patients of one in 1,000. Furthermore, history and examination suggest that the prior risk of hepatitis is 20% while that of cirrhosis is 80%. It should be immediately apparent that this problem is similar to that of the wildcatter: in both cases the first decision is whether or not to do a 'costly' diagnostic test. Therefore the flow diagram (Fig. 3) is similar—in both cases the diagnostic test enables the decision-maker to determine the 'diagnosis' before the definitive choice. The chance node therefore precedes the

Fig. 3. *This example concerns the expected survival rates associated with the decision to take a biopsy versus making the decision on the basis of clinical features and blood tests.* On the basis of the latter the clinician believes that the probability of hepatitis is 0.2 (20%) and that of cirrhosis is 0.8 (80%). Furthermore, the probabilities of survival for two years are as follows:

Hepatitis with steroids	0.85	(85%)
Hepatitis without steroids	0.67	(67%)
Cirrhosis with steroids	0.48	(48%)
Cirrhosis without steroids	0.50	(50%)

If no biopsy were possible then we would wish to compare overall outcomes with and without steroids — lower branch.

Survival expectancy with steroids is (probability of survival with hepatitis given steroids × probability of hepatitis) + (probability of survival with cirrhosis given steroids × probability of cirrhosis), ie

$$(0.85 \times 0.2) + (0.48 \times 0.8) = 0.554 \ (55.4\%)$$

Similarly, without steroids it is

$$(0.67 \times 0.2) + (0.5 \times 0.8) = 0.534 \ (53.4\%)$$

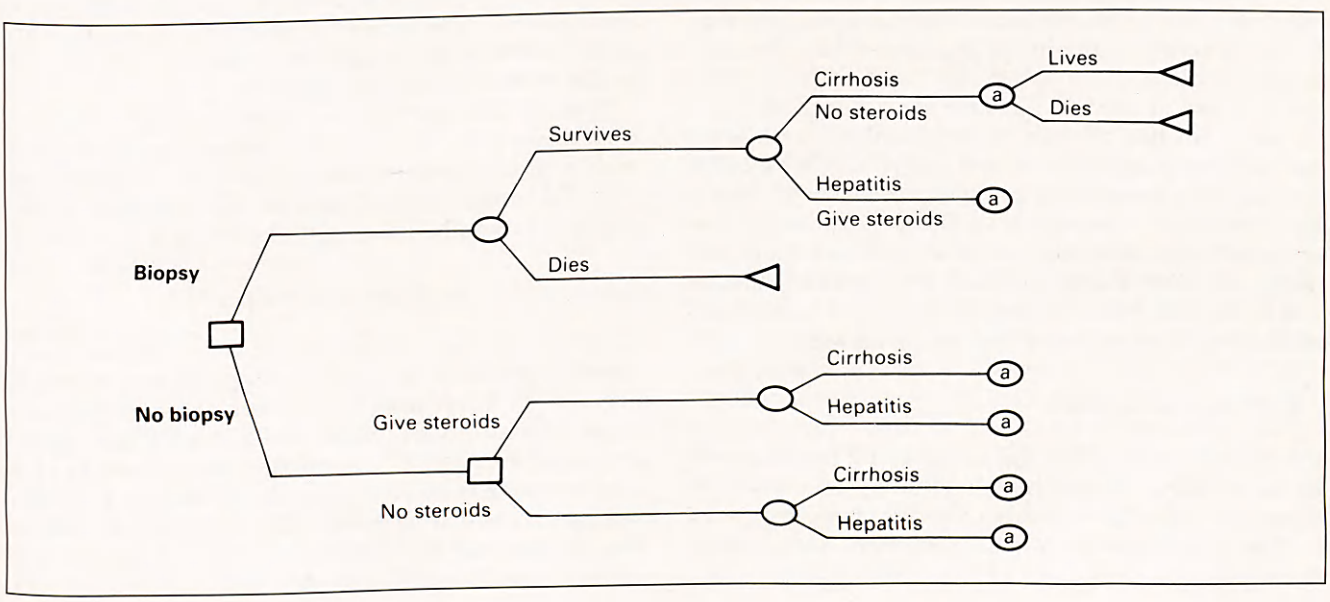
Clearly if we do not do a biopsy we must give steroids. We can therefore fold the survival figure of 0.554 back to the first branch and compare this with the greatest expected survival if we do a biopsy (top branch). We calculate first the survival expectancies at node A. If the biopsy specimen shows hepatitis we will treat with steroids, and if cirrhosis we will withhold this medicine. The overall survival when we put these two together becomes

$$(0.2 \times 0.85) + (0.8 \times 0.50) = 0.57.$$

We then go to the proximal node and 'average out' further

$$(0.999 \times 0.57) + (0.001 \times 0) = 0.5694.$$

Therefore the probability of survival with the biopsy decision exceeds that with no biopsy by 1.54%.



choice node on this branch of the decision tree. Without the diagnostic test the definitive choice comes first and the probabilities of alternative states follow. In making the calculations based on the decision tree, we could begin by working out the survival rates on the lower branch: ie if there were no opportunity to perform a biopsy would the patient be expected to fare better with or without steroids? The calculation is again based on the method used in the 'casino' example and is shown in the legend to the figure: the preferred treatment in the absence of information from the biopsy is steroids. However, when we 'fold back' to the first branch—biopsy or decide without—then the calculation shows that the chances of survival are maximised by doing the biopsy. This example is discussed in more detail by Weinstein and Fineberg [12] who give the disclaimer that the example is offered for pedagogic purposes only and not as a guide to clinicians in practice.

All three examples have one thing in common: they wish simply to maximise an outcome variable: monetary return in the first two examples and survival in the third. Many decisions in medicine, however, have multiple objectives and these may be in competition—policies designed to maximise one outcome may impair another. What does decision analysis have to say about the trade-offs inherent in these decisions? To consider this further let us again use a plausible example from clinical practice.

Example 4

Our problem here concerns a woman with very early cancer of the reproductive tract. This has been diagnosed by means of an excisional biopsy but the pathologist tells us that the probability of residual tumour is 0.02 (2%), in which case the disease can be eradicated in one-half of cases by further surgery (hysterectomy and lymph-node resection). The operation has a mortality of 5 per 1,000. We could approach this problem in the same way as the hepatologist and the wildcatter: ie simply calculate the option associated with the greatest prospect of survival. This may be appropriate for a woman who has completed her family; the decision tree and calculation are shown in Fig. 4. We find that surgery maximises the prospects of survival but a younger person who cherishes the desire to have children may wish to make a trade-off in favour of retaining her fertility. Fortunately decision analysis contains within its theoretical framework a method to evaluate and incorporate these preferences or utilities.

Measurement of utilities

Our patient must define the utility of full health, sterility, and death. The best health state is given a utility of 1 and the worst (presumably death) is given a utility of 0. The problem is to know where between 0 and 1 sterility should be placed.

The best method for measuring people's utilities is the basic reference lottery where the relative utilities of three health states are worked out together [13,14]. We obtain the utility of infertile life by asking for a choice between that and various gambles between fertile life and immediate death until a level of indifference is reached. In practice a subject would be asked to imagine two doors through one of which she had to go. Behind the left-hand door there was no risk of death but she would be rendered sterile. Behind the right-hand door she would encounter a 50% chance of intact survival but also a 50% risk of death (Fig. 5). She is likely to select the left-hand door. The risks of death through the right-hand door would then be adjusted (decreased) until a point was reached where she was unable to decide which door to select. This might occur when the risk of death through the right-hand door was 2% and of intact survival 98%. We call this the level of indifference. We can say that our patient values sterile survival at 0.98 on a scale where full health was valued 1 and death is valued 0.

There are alternative methods of measuring values, such as asking patients to mark health states on a linear scale, but, unlike the reference gamble, this method is not axiomatically correct. People avoid the extremes of the scale and, because the trade-off inherent in the technique is not obvious to the subject, values obtained this way are distorted. A better alternative makes use of natural underlying scales such as money, or years of life. People's utilities for money and years of life are rarely linear. People are usually risk-seeking or risk-averse. For example, the gambler in the first problem is likely to be risk-seeking; the utility curve for such a gambler is nonlinear, eg £50 might have twice the utility of £35. People taking out insurance policies are by definition risk-averse with nonlinear utility curves shaping the other way—the value to me of a new £25,000 car might be more than 25 times the £1,000 insurance premium. The difference between the expected losses with and without insurance is equal to the long-term administrative costs and profits of the insurance company.

Years of life expectancy is another frequently used underlying scale, but it can be shown that people tend to value the years immediately ahead more highly than those far in the future. This is another example of risk aversion, and utility scales must reflect this.

Adding utilities to the decision analysis

Utilities must be combined with the probabilities to select a preferred course of action, ie that with the greatest expected utility. We start by estimating the utility of each chance node which is calculated as the weighted average of the utilities of its possible outcomes, where the weights are the probabilities of each outcome. The mathematics is shown in the legend to Fig. 4, where we assume that our young patient has a utility for infertile life of 0.98. This implies that she

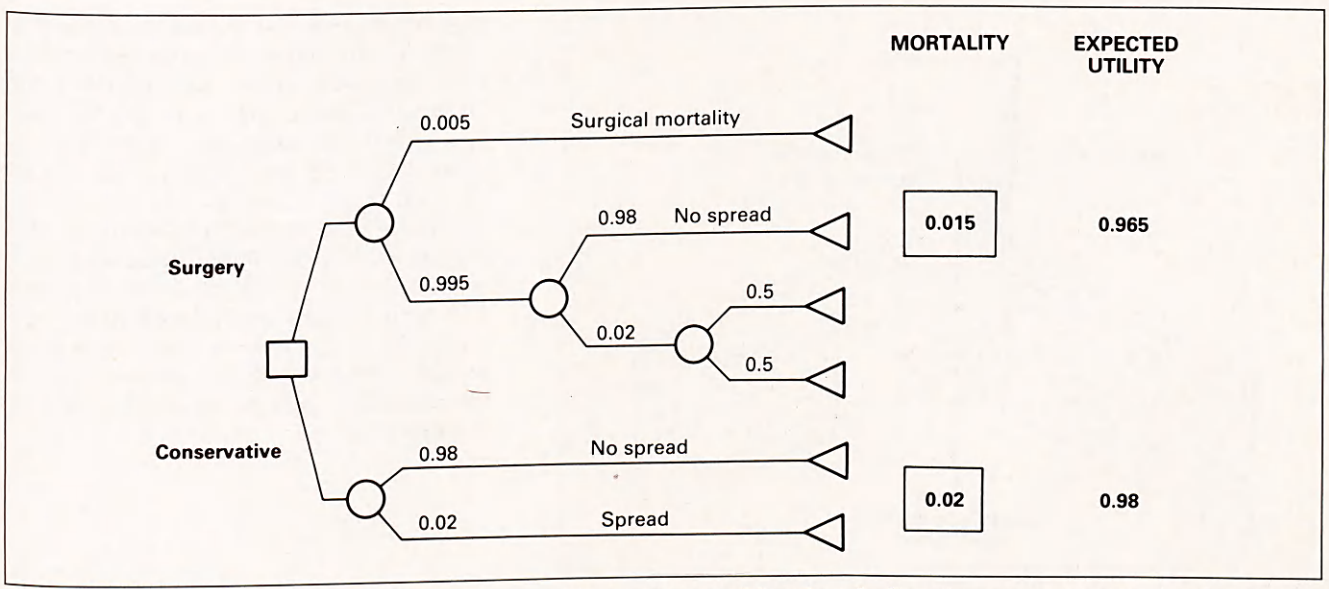


Fig. 4. *Decision analysis for the decision to carry out hysterectomy and lymph-node dissection versus conservative therapy in a woman with cancer of the reproductive tract.* We believe that the risk that the tumour has spread beyond the original excisional biopsy (PS) is 2% (0.02). If it has done so, the probability of cure (PC) is 50% (0.50). The operative mortality (OM) is 0.5% (0.005). If we simply wish to maximise the chance of survival, then we could compare the mortality of conservative management (0.02) with that of further surgery $0.005 + (0.995 \times 0.5 \times 0.02) = 0.015$. Clearly the latter maximises the chance of survival. However, if we assume that all forms of death are equally undesirable (UD = 0) and that the patient would run a 2% risk of death to avoid infertility (UI = 0.98), then the preferred decision changes. The utility for fertile life (UFL) is 1.0. The formula for the expected utility of conservative management is now:

$$(UD \times PS) + [UFL \times (1 - PS)]$$
$$= (0 \times 0.02) + (1 \times 0.98) = 0.98$$

For surgical management the expected utility is:

$$(UD \times OM) + (1 - OM)\{[(1 - PS) \times UI]$$
$$+ PS(PC \times UI)[(1 - PC) \times UD]\}$$
$$= 0.005 \times 0 + 0.995\{(0.98 \times 0.98)$$
$$+ 0.02\ [(0.5 \times 0.98) + (0.5 \times 0.0)]\}$$
$$= 0.9654$$

This is a lower figure than the expected utility for conservative surgery which now becomes the preferred option. We could determine the chance of spread where the expected utilities would be equal—a threshold analysis.

would run a 2% risk of death to retain her fertility. Under this assumption, surgery is no longer the preferred option (as it was when our sole objective was to maximise chance of survival).

Sensitivity analysis

We have now shown how the problem is structured, probabilities are selected, values measured, and the course of maximum expected utility identified. The final part of a full decision analysis should include a sensitivity analysis, because conclusions are dependent on the probabilities and utilities used, and in real life we are never certain what these are. In a sensitivity analysis, each of the key probabilities and values is varied in turn within the range of reasonable uncertainty to test the robustness of the conclusion. Fig. 6 shows a one-way sensitivity analysis in which our wildcatter has examined the effect of a range of drilling costs on the expected value of the decision to drill for oil: a profit is expected over a wide range of plausible estimates. However, when we did sensitivity analysis for the cancer example above, we found that the decision was very sensitive to the utility of sterility. When we extended our model to take into account possible preferences for mode of death (whether by cancer in the long term or complications of surgery in the short term) we found that previous conclusions were unaltered throughout the plausible range: the decision is not sensitive to this factor [15].

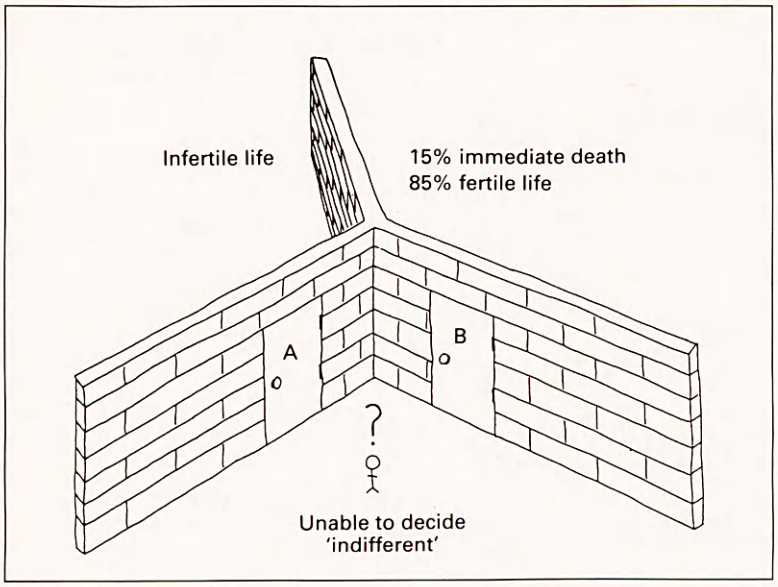


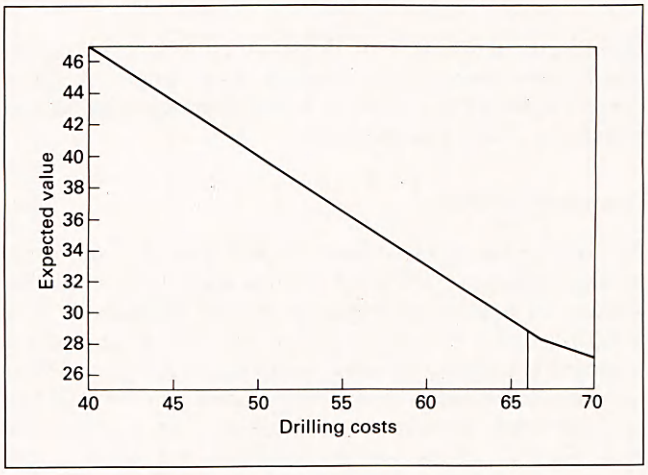
Fig. 5. *Diagram of the lottery method to determine an individual's trade-off between three outcomes: in this case sterility from treatment, intact life, and death.* The intermediate outcome (sterility) is placed behind the left-hand door and the extreme outcomes (intact life/death) behind the right-hand door. The risk of the intermediate outcome is fixed at 100% but the relative risks of intact life versus death (right-hand door) are varied from the extreme ends of a scale to the point where the subject cannot decide which door to take—her point of indifference.

Limitations of decision analysis

Probability estimates

Ideally each probability estimate used in making a decision should be derived from the literature; the example of early reproductive cancer was based on the problem of occult cervical tumours and in this case we were fortunate to find numerous relevant articles in the literature. The accuracy and precision of the estimates will depend on the study design. Large randomised controlled studies should be used with safeguards against post-randomisation bias. Such studies are not always available, even when they are, judgement is required to extrapolate results from one time and place to another. It is frequently necessary to use semi-objective probability estimates in which probabilities obtained from the literature have to be modified according to local circumstances or changes in practice. In the example of occult cervical cancer, we adjusted the published estimate of operation mortality downwards to take into account improvements in surgical technique, anaesthetics and intensive care since the studies were undertaken. Such adjustments must be arbitrary but they cannot be avoided. One of the most commonly heard and least justified criticisms of decision analysis is that the need to make revisions of probability estimates somehow invalidates the technique. These adjustments invalidate decision analysis no more than they do conventional intuitive decision-making. The latter is also based on probabilities which are no more accurate for not being made explicit. Indeed, the process of making probabilities explicit is a reason to use rather than abandon decision analysis, since this exercise exposes the source of disagreement about treatment policy. The process of decision analysis, because it is transparent, encourages a comprehensive review of the literature: a great improvement on informal probability estimates based on incomplete data and subject to 'availability bias'—the systematic tendency of the human mind to overestimate the likelihood of events which have occurred recently or which stand out from the ordinary. This is one of the many 'faulty heuristics' to which the human mind is prone [16,17]. The traditional review article has come in for much justified criticism and should be replaced by structured overviews in which literature review is *systematic* and where the results of similar studies are combined by the technique of quantitative meta-analy-

Fig. 6. *A typical sensitivity analysis showing in this case how the expected value of the decision to drill for oil varies with drilling costs to the point—beyond the vertical line—where it is less than the expected value of not drilling.*



sis. Decision and sensitivity analysis impose a further tier of rigour and relevance by showing what the results could mean for clinical practice in a logical, transparent and numerical way.

Often the probability of an event is based not on one but on many items of information, ie on many 'tests'. When this is the case we use Bayes' theorem to calculate the probability of an event by combining prior probability (prevalence) and the result of all these tests. This is superior to intuitive methods since it avoids 'anchoring'—the systematic tendency of the human mind to give undue weight to the information which is collected first. The probabilities of different diagnoses produced by Bayes' theorem are eminently suitable for inclusion on decision trees and hence for decision analysis.

There are, of course, times when objective probabilities cannot be derived from the literature at all and subjective numerical probabilities must be used. These can be obtained by deriving a consensus from a panel of experts in a way which minimises the biases mentioned above (the Delphi method) [12].

Utility measurements

We saw in the cancer example that the decision between conservative management and radical surgery was sensitive to the value placed by the patient on preserving her fertility. We also showed that if these utilities were to be used in decision analysis, then the trade-off (in this case risk of death versus preservation of fertility) must be explicit in the method by which the utilities were obtained. Nevertheless, even if such a method is used, no one can pretend that these utility estimates are themselves not subject to bias. For example, the way that a lottery is framed (eg risk of death or chance of survival) can influence the results. Nevertheless, it is possible to test whether the subject has understood the choices by measuring the internal consistency of a series of responses in which the answers to one gamble should be predictable on the basis of the responses to two earlier gambles [18,19].

Another feature of utility measurement is that decision analysis assumes that utilities remain proportionately the same as risk changes; for example, a patient who regards a Down's birth as twice as bad as the accidental miscarriage of a normal fetus would be expected to regard a 1-in-200 risk of Down's birth as equivalent to a 1-in-100 risk of the loss of a normal fetus. This principle is called the 'constant proportion risk attitude'. Unfortunately there is evidence that human reasoning does not strictly adhere to this principle [20]. We would argue that lack of a constant proportion risk attitude is irrational and a failure of human mental decision processes which can be overcome by decision analysis. Nevertheless, it is impossible to deny the sheer practical difficulty of obtaining utility functions from bewildered patients on a busy clinical service—a point to which we shall return.

Representing life's infinite complexity

Decisions in real life have almost infinite complexity. In the early cancer example, for instance, we took the importance of infertility into account and in the full analysis (not shown here) even incorporated relative dislike of different modes of death, but we could have gone further to enumerate the described complications of surgery, their probabilities and utilities. We could also have considered other forms of treatment, such as radiotherapy. However, the resulting tree would have been very complicated—a messy bush! What we did was to use our judgement: we selected a question which is the subject of real disagreement and which seemed most important. Decision analysis requires explicit articulation of a thought process and some people think that this biases the result. Such people argue that the intuitive decision-maker is able to incorporate the full myriad of interconnected considerations. The trouble with this claim is that all the formal psychological studies of human reasoning show that, far from incorporating these subtleties, the mind makes major simplifying assumptions and is prone to numerous biases [16,20]. There is no intellectual justification for hiding behind a mysterious mental mechanism which has been shown to fail in careful experiments. Equally, the results of a decision analysis should not be accepted uncritically; the reader must determine whether all these important treatments, probabilities, and outcomes have been incorporated, and whether the range of probability and utility measurement in a sensitivity analysis are reasonable. Decision analysis, for all its internal rigour, must be based on at least one assumption—which decision options and outcomes to include.

Decision analysis and medical practice

Care of individual patients

Formal decision analysis, in its full rigour, is seldom used in everyday practice even by ardent advocates of the technique. The need to employ a series of lotteries for all possible outcomes, let alone incorporate 'coherence' measures, along with the incomprehension with which many patients might regard any such attempt, are powerful limiting factors. Prenatal diagnosis/clinical genetics are possible exceptions [21]; they are more overtly value-led than many other areas of medicine, and the patients are often young and articulate. They themselves are not in the 'sick role'. Such subjects appreciate the clarity with which issues are highlighted, and for them counselling is more effective [21]. Decision analysis exposes the reasons for any disagreement between parents about the need for prenatal testing; different estimates of risk or differing values can then be explored in more detail. Nevertheless, even in this circumscribed area of medical practice, use of decision analysis in its full rigour to aid the

management of *individual* patients remains the exception rather than the rule. What then is the use of the technique if it can seldom be used in real time? The value of decision analysis derives from the observation that much of medical practice is determined not in the consulting room but in the wider arenas of medical and public debate, and it is here that the technique is most useful.

Decision analysis and treatment policy

Medical controversies are based on disputes over the best treatment for groups of patients with similar features, eg the management of women with occult cancers or young people with chronic progressive liver failure. Decision analysis encourages decision-makers to structure the decision correctly by means of a decision tree, to search for the relevant probabilities, and to acknowledge the importance of those values to which the decision is sensitive. Any debate can then be focused on the specific features of the decomposed problem, and mistakes will not occur from incorrectly analysing the correct data. The tradition in all civilisations is to improve decisions through debate; if not, why discuss medical ethics, argue over controversial treatments, or debate serious issues in Parliament? Decision analysis provides the logical framework for these debates—for 'testing' one decision against another. If two experts disagree, then instead of resorting to slogans ('People die of cancer' . . . 'How would you like to have no leg?' . . . 'Women value their fertility'), the precise source of disagreement can be pinpointed. Often this will be because of a disagreement over the probabilities such as the probability of death from radical hysterectomy or liver biopsy. Even more often, different values will be found to be the source of disagreement. For example, we found that women selecting home birth had a lower utility for avoiding fetal death than those selecting hospital confinement [22].

The very act of structuring decisions properly may give valuable insight. For example, until decision analysis was conducted, the entire literature on breech delivery compared outcome following Caesarean section with that following vaginal delivery. The decision tree showed that trial of vaginal delivery should be compared with planned Caesarean section [23], since many cases with the worst fetal and maternal outcome will follow hurried intrapartum Caesarean section. The act of incorporating patients' values has an even greater effect on our ways of thinking. In the field of prenatal diagnosis the facile analysis that amniocentesis should be recommended when the risk of Down's syndrome is greater than the risk of miscarriage from the procedure is wrong, since a decision theorist can immediately see that this implies an equal value for Down's syndrome and fetal loss—a situation which applies to only a few patients [19,21]. Decision theory emphasises those components of the decision which are value-led. A chest physician recently declared, on the basis of the lung function tests obtained from a patient with kyphoscoliosis, that he did not 'think the patient should contemplate pregnancy'. When pressed, he estimated the risk of death from pregnancy at 0.5%. Clearly, he had no business to attempt to proscribe pregnancy since this was a value decision which should be based on, but not determined by, the probability estimate. These are examples of some simple insights which decision analysis provides, but the technique is most powerful when complex medical controversies are examined. The example which we used (microscopic occult cancer) was inspired by a debate about the histological diagnosis of so-called 'micro-invasive cancer'. The analysis shows how treatment can be individualised according to a range of histological and other criteria, and that any attempt to treat patients according to fixed cut-off criteria for histological diagnosis is doomed to failure. Decision analysis has been used to shed light on some important medical controversies. Some examples are listed in Table 1.

Decision analysis and research

Decision and sensitivity analyses show which factors influence expected utility most strongly, and therefore point the way to future research. In addition to showing us which research questions are most pressing, decision analysis can be used to determine the sample sizes for trials [34,35]. Any trial should be sufficiently large to detect the size of difference which would influence clinical practice. This can be determined by

Table 1. Some published examples of decision analysis

Pauker and Pauker [21]	Amniocentesis for prenatal diagnosis
Heckerling and Verp [24]	Amniocentesis versus chorion sampling
Bingham and Lilford [23]	Management of the term breech
Hillner and Smith [25]	Chemotherapy for breast cancer
Verhoef *et al* [26]	Mastectomy versus lumpectomy in breast cancer
Klein and Pauker [27]	Anticoagulation for DVT in pregnancy
Elstein *et al* [28]	Oestrogen replacement in the menopause
Speroff *et al* [29]	Elective oöphorectomy at hysterectomy
Feldman and Freiman [30]	Elective Caesarean section
Weinstein *et al* [31]	Coronary artery bypass surgery
Tompkins *et al* [32]	Antibiotic therapy for sore throats
Neutra [33]	Management of appendicitis

decision theory. Clinicians are unlikely to place patients in trials if the trial has no chance of producing a result which would influence clinical practice. Delta is the term used to describe the magnitude of the effect in a trial, and ideally delta should be of sufficient magnitude to change clinical practice. Let us imagine two treatments, A and B. Treatment B has greater side-effects than treatment A but may be more effective. Treatment B could be radical mastectomy for breast cancer, elective Caesarean section for the premature breech delivery, or a new antihypertensive for use in pregnancy. Treatment A might be local excision of the breast tumour, the intention to deliver the breech vaginally, or 'tried and tested' pregnancy antihypertensives. In each case, value trade-offs among clinicians or (preferably) other subjects could be used to measure the size of delta which would make the cost of treatment B worthwhile. Let us imagine that 60% of women treated by 'lumpectomy' for T1 N0 M0 breast cancer will die of their disease. If the mean utility for survival without a breast is known, then it is possible to calculate the minimum improvement in survival which would justify mastectomy in preference to lumpectomy. As in the previous example, a utility for life after mastectomy of 0.98* (life post-lumpectomy = 1 and death = 0) corresponds to a willingness to run a 2% risk of earlier death to avoid this disfigurement. However, if the mean utility was 0.9, then we would know that half of the women would run a 10% risk of death from their cancer, above the baseline, to avoid the more radical operation. (It would, of course, be necessary for these women to understand that they were trading death many years in the future against immediate mastectomy.) Delta can be determined for these utilities as shown in Fig. 7. It will be noted that delta is not large; a change in mortality of 4% must be demonstrated to satisfy half of all women if the utility of life after a mastectomy is 0.9, whereas much smaller treatment effects are clinically important if the utility is 0.98. It is obviously important to bear these considerations in mind when designing or interpreting clinical trials. A trial must not be larger than necessary to detect the clinically meaningful effect, but equally it should be large enough to detect effects which most women would consider worthwhile. At the very least it should be sufficiently powerful in combination (by meta-analysis) with other trials that might be under way or which are thought likely to take place.

An important corollary of the use of decision theory in the design of trials is that the alpha (type 1—false positive) and beta (type 2—false negative) errors should be equal [36]. This is in contrast to the usual teaching that a false positive result is more serious than a false negative if one of the treatments is more 'costly', eg if it has worse side-effects. Under the scheme proposed here this greater cost is discounted in the size of delta. Once this is done it is no longer advantageous to plan for different type 1 and type 2 errors, since a false negative trial result is no longer preferable to a false positive result. In addition, there is no intuitive way to relate the relative sizes of the type 1 and 2 errors to clinical practice, whereas a reader can immediately relate delta to the effect which would be required to justify the use of new or more radical treatment†.

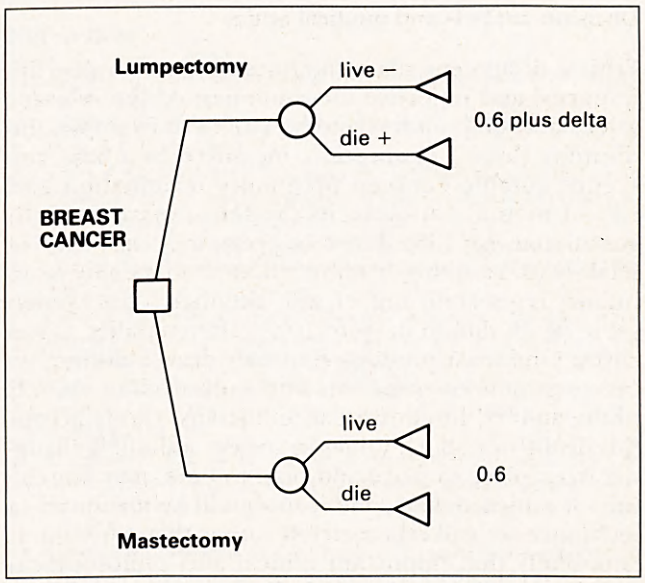


Fig. 7. *Decision analysis for a trial of surgical treatments for early cancer of the breast.* A trial is planned of mastectomy versus 'lumpectomy'. Most 'experts' think that the treatments are equivalent in terms of survival or mastectomy is slightly better. Therefore a trial needs to measure an improvement of delta with mastectomy, from the 60% base-line risk of dying from cancer following lumpectomy. If a small delta (eg 2% improvement in mortality) would justify the larger operation, then the trial will need to be very large.

*A utility of 0.98 seems about right to us—as husbands we would go for a higher figure still. We are interested in getting direct estimates from women.

†The Bayesian approach to randomised trials is very much in keeping with the decision analysis approach. Decision analysts are attracted to the idea of starting a trial with a 'prior' distribution of expectations of the effects of competing treatments since they are similar to semi-objective probability estimates in decision analysis. Similarly, the presentation of the results as a frequency distribution of the 'posterior' odds of the two treatments takes into account the consideration that alpha and beta errors should be equal. Lastly, standard (frequentist) statistics do not tell clinicians the probability that the difference is as big as measured. In a trial where we had no prior knowledge at all (ie where we were equipoised around completely uninformative prior probabilities), a two-tailed p value of 0.05 implies a 50% chance that the difference between two treatments is as big as or bigger than measured, 47.5% that it was smaller but in the same direction, and 2.5% the direction of the effect was wrong. However, as soon as we have some prior belief, even about the shape of the likely distribution, the above probabilities change.

Decision analysis and medical ethics

Ethical disagreements sometimes hinge on value differences, and objective measurement of the relevant values may help understanding [37]. For example, the disutility ('cost') of abortion, measured by a basic reference gamble between pregnancy termination and risk of mental handicap in the fetus, increases with gestational age [38]. Later in pregnancy the 'cost' of fetal death resulting from inaction after viability or in infancy is generally higher still, although a few women value death due to inaction, even after viability, as less costly. One must presume that they draw a distinction between acts of omission and commission. Moral philosophers, in contrast to clinicians, rarely accept this distinction. Such values in society, although changing from place to place and time to time, may be relevant to ethical debate, and if so should be measured (a technique we call ethometrics) rather than guessed. It is unlikely that important ethical and philosophical questions will ever be resolved by measuring differences in people's values; after all, we seek to influence public attitudes by ethical debate, not the other way round. However, at any one time public perceptions cannot be ignored if society as a whole is to have a stake in medical ethics. Furthermore, the philosophy may lead us to conclude that an issue is subject to gradualism, eg that abortion becomes gradually less acceptable as the fetus gets older. If this is so, then philosophers might also agree that the point on the scale where an action should be judged unlawful should be determined by the values of society. Decision analysis may also give valuable insight into the ethics of trials. Bayesian trialists talk about 'prior' expectations before starting a trial. If the expected difference is not zero, then we are not equipoised (agnostic). If we start from the premise that equipoise is ethically important, then a trial might still be ethical, even if we expect one treatment to be more successful than another: first, access to one of the treatments may be restricted, on the basis of inadequate resource, to patients in a trial; second, if one of the treatments has higher perceived 'cost' (eg side-effects), the trial is ethical provided our starting expectation is that the more costly treatment has an advantage sufficient to make up for (but not exceed) this cost, ie if our 'prior' odds equal delta in the breast cancer example.

Distribution of scarce resources

Readers interested in health economics will have recognised the similarity between the expected utility model which we have presented and cost-utility analysis [39,40]; indeed, the logical framework and mathematics are identical. Thus decision analysis can be extended from how to care for an individual patient to decisions affecting communities. At its simplest level the sensitivity analysis approach can be used to calculate the expected benefits of extending a screening programme, for example, for fetal anomaly [41,42] or cancer [43]. Full cost-utility studies are based on decision trees where the expected values achievable by competing resource allocation decisions are compared [44–46]. The difficult part is to reduce years of life gained or suffering averted to a single scale. One approach is based on the Quality Adjusted Life Year (QALY). This requires, first, that the possible health states are modelled over time—the Markov process [47]; then the number of years spent in each state are adjusted for quality.

Thus we do not simply calculate the number of years gained with different treatments but also take into account the quality of those years.

As with utilities, the quality adjustment in QALYs can be derived in three ways: using rating scales for various hypothetical health states [48,49], the lottery method, and the time trade-off method.

Rating scales. These have been used to derive valuation matrices for hypothetical health states varying in two dimensions: first, a disability rating varying in seven steps from 'no disability' to 'unconscious' and, second, a distress rating varying in four steps from 'no distress' to 'severe distress' [48]. Real health states are then placed on the valuation matrix and the relevant utility of a year in that state is read off.

Lottery method. Patients are asked to imagine that they face a choice between a lifetime of impaired health and a gamble between a lifetime of full health and death. The probability (p) of full health at which they would be unable to decide defines the value of the impaired health state. If the answer is near 1 the health state is rated near to full health, while if it is near 0 the health state is rated little better than death.

Time trade-off method. Patients are asked what is the smallest fraction (f) of a year of life at full health they would exchange for a full year in the relevant health state. The value of f defines the utility of the health state.

The Markov process or Markov chain. For this method a limited set of health states and permitted transitions between these health states is defined. In our example a patient in the 'post cone biopsy' state may in any given year move into the state 'post cone biopsy one child', the state 'radical operation no child', and the state 'recurrence no child'. From the state 'radical operation no child' she can move into 'post operation no child', 'recurrence no child', or 'death', and so on. The probabilities of each move would be derived from the literature. In a simple Markov chain transition probabilities do not vary from year to year. In biology they usually do, with, for example, risk of recurrence being higher in the earlier years after diagnosis. Such variations can be accommodated in modern computer analyses. The utilities of each of these states would be

calculated by one of the above methods, but preferably by lotteries.

Typical figures for a year in the following health states might be:

post cone biopsy no child	0.95
operation no child	0.95
post cone biopsy 1 child	0.99
post cone biopsy 2 children	1.0
post operation no child	0.95
post operation 1 child	0.99
post operation 2 children	1.0
symptomatic residual disease	0.9
dead	0

The model is used by placing a hypothetical cohort of patients in a certain state at the beginning of the period of analysis (eg placing 10,000 patients in the post cone biopsy state) and following their course year by year. Besides the relevant transition rates (eg residual disease to death), age-specific mortality from other causes is fed into the model so that after a sufficiently long time all patients will have died. To allow for the lower utility of years far in the future a discount can be applied. The rate of this discount can also be calculated individually by lotteries.

The programme is run with various policies, eg immediate surgery, surgery after the first child, surgery after the second child, or no surgery at all. The total and average numbers of QALYs obtained with each policy are compared and that producing the most is chosen.

Conclusion

The language and methodology of decision analysis and, more specifically, of expected utility theory can change how we think. The discovery that there is a specific mathematical function (expected utility) which measures the benefits of a course of action cannot be ignored by those who wish to analyse their treatment choices. Since it emphasises the way that values and probabilities underlie decisions, it is threatening to those who like to work with certainties. Many doctors reject it, believing that they have made good decisions for many years without it. We hope that the examples in this lecture have shown that this may not be the case. Decision analysis has been widely used in business for years, and has entered the mainstream of medical thinking in North America and, more recently, the Antipodes. It is incorporated in the medical curriculum in centres as far apart as Hamilton, Ontario and Dunedin in the South Island of New Zealand. We believe that doctors in Europe may come to love it or hate it, but cannot ignore it.

Address for correspondence: Professor R. J. Lilford, Institute of Epidemiology and Health Services Research, University of Leeds, Leeds LS2 9LN.

References

1 Lilford RJ. The doctor's role in terms of fundamental human needs. *Leech* 1972;**41**(1–2):7–8.
2 Balint M. *The doctor, his patient and the illness* (2nd edn). London, Pitman Paperbacks, 1968.
3 Lilford RJ, Bingham P, Fawdry R, Setchell M, Chard T. The development of on-line history-taking systems in antenatal care. *Meth Inform Med* 1983;**26**(5):474–87.
4 McDonald CJ. Protocol-based reminders — the quality of care and the non-perfectibility of man. *N Engl J Med* 1976;**195**: 1351–5.
5 Lilford RJ, Kelly M, Baines A, *et al.* Improved medical care through the use of protocols: a randomised trial based on the first antenatal visit. *Br Med J* (in press).
6 Quaak MJ, Westerman RF, van Bemmel JH. Comparisons between written and computerised histories. *Br Med J* 1987;**295**:184–95.
7 Bingham P, Lilford RJ, Chard T. Strengths and weaknesses of direct patient interviewing by a microcomputer system in specialist gynaecological practice. *Eur J Obstet Gynaecol Reprod Biol* 1984;**18**:43–56.
8 Dreyfus HL, Dreyfus SE, Athenafion T. *Mind over machine* (the power of human intuition and expertise in the area of the computer) Oxford: Blackwell Scientific Publications, 1986.
9 Adams ID, Chan M, Clifford PC, *et al.* Computer-aided diagnosis of abdominal pain: a multicentre study. *Br Med J* 1986;**293**:800–4.
10 Knill-Jones RP, Stern RB, Grimes DH, Maxwell JD, Thompson RPH, Williams R. Use of a sequential Bayesian model in diagnosis of jaundice by computer. *Br Med J* 1973;**1**:530–5.
11 Raiffa H. *Decision analysis: introductory lectures on choices under uncertainty.* New York: Random House, 1968.
12 Weinstein MC, Fineberg HV, eds. *Clinical decision analysis.* Philadelphia: WB Saunders, 1980.
13 Keeny RL, Raiffa H. *Decisions with multiple objectives: preferences and value trade-offs.* New York: J Wiley, 1976.
14 Lindley DV. The role of utility in decision making. *J R Coll Physicians Lond* 1975;**9**:225–30.
15 Johnson N. Computer-aided clinical management using the treatment of microscopic cancer of the ectocervix as an example. In *Baillière's clinics in obstetrics and gynaecology,* ed Lilford RJ. London: Baillière Tindall, 1990.
16 Tversky A, Kahnemann D. Judgement under uncertainty: heuristics and biases. *Science* 1974:**185**:1124–31.
17 Pauker SG, Kassirer JP. Decision analysis. *N Engl J Med* 1987;**316**:250–8.
18 Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, Ciampi A, Till JE, Boyd NF. The measurement of patients' values in medicine. *Med Decision Making* 1982;**2**:449–62.
19 Thornton JG, Lilford RJ. Prenatal diagnosis of Down's syndrome: a method for measuring the consistency of women's decisions. *Med Decision Making* 1982;**10**:288–93.
20 Elstein AS, Bordage G. Psychology of clinical reasoning. In *Professional judgement: a reader in clinical decision-making.* Dowie J, Elstein A, eds. Cambridge: Cambridge University Press, 1988.
21 Pauker SP, Pauker SG. Prenatal diagnosis: a directive approach to genetic counselling using decision analysis. *Yale J Biol Med* 1977;**50**:275–89.
22 Thornton JG, Lilford RJ. The Caesarean section decision: patients' choices are not determined by their immediate emotional reactions. *J Obstet Gynaecol* 1989;**9**:283–8.
23 Bingham P, Lilford RJ. Management of the selected term breech presentation: assessment of the risks of selected vaginal delivery versus Caesarean section for all cases. *Obstet Gynecol* 1987;**69**:965–78.
24 Heckerling PS, Verp MS. Amniocentesis or chorionic villus sampling for prenatal genetic testing: a decision analysis. *J Clin Epidemiol* 1991;**44**:657–70.
25 Hillner BE, Smith TJ. Efficacy and cost-effectiveness of adjuvant chemotherapy in node negative breast cancer. *N Engl J Med* 1991;**324**:160–8.

26 Verhoef LCG, Stalpers LJA, Verbeek ALM, Wobbes T, van Daal WAJ. Breast-conserving treatment or mastectomy in early breast cancer: a clinical decision analysis with special reference to the risk of local recurrence. *Eur J Cancer* 1991;**27**:1132–7.

27 Klein K, Pauker SG. Recurrent deep venous thrombosis in pregnancy: analysis of the risks and benefits of anticoagulation. *Med Decision Making* 1981;**1**:181–202.

28 Elstein AS, Holzman GB, Ravitch MM, *et al.* Comparison of physicians' decisions regarding estrogen replacement therapy for menopausal women and decisions derived from a decision analytic model. *Am J Med* 1986;**80**:246–58.

29 Speroff T, Dawson NV, Speroff L, Haber RJ. A risk-benefit analysis of elective bilateral oophorectomy: effect of changes in compliance with oestrogen. *Am J Obstet Gynecol* 1991;**164**:165–74.

30 Feldman GB, Freiman JA. Prophylactic Caesarean section at term? *N Engl J Med* 1985;**312**:1264–7.

31 Weinstein MC, Pliskin JS, Stason WB. Coronary artery bypass surgery; decision and policy analysis. In *Costs, benefits, and risks of surgery,* Bunker JP, Barnes BA, Mosteller F, eds. Oxford: Oxford University Press, 1977.

32 Tompkins RK, Burnes DC, Cable WE. An analysis of the cost-effectiveness of pharyngitis management and acute rheumatic fever prevention. *Ann Intern Med* 1977;**86**:481–92.

33 Neutra R. Indications for the surgical treatment of suspected acute appendicitis. A cost-effectiveness approach. In *Costs, benefits and risks of surgery,* Bunker JP, Barnes BA, Mosteller F, eds. Oxford, Oxford University Press, 1977.

34 Lilford RJ. Decision theory to estimate delta. *Lancet* 1990;**i**:483–4.

35 Thompson MS. Decision analytic determination of study size: the case of electronic fetal monitoring. *Med Decision Making* 1981;**1**:165–79.

36 Lilford RJ, Johnson N. The alpha and beta errors in randomised trials. *N Engl J Med* 1990;**332**:780.

37 Dear PRF. Attitudes to the risk of adverse neonatal outcome. *Contemp Rev Obstet Gynaecol* 1989;**1**:207–13.

38 Thornton JG, Bryce FC, Bhabra K. Withholding intensive care from premature babies. *Lancet* 1988;**ii**:332–3.

39 Maynard A. Logic in medicine: an economic perspective. *Logic in Medicine* 1988;73–88.

40 Drummond M, Stoddart G, Torrance G. *Methods for the economic evaluation of health care programmes.* Oxford: Oxford Medical Publications, 1987.

41 Thornton JG, Lilford RJ, Newcombe RG. Tables for estimation of individual risks of fetal neural tube and ventral wall defects, incorporating prior probability, maternal serum α-fetoprotein levels, and ultrasonographic examination results. *Am J Obstet Gynecol* 1991;**164**:154–60.

42 Tyrrell S, Howel D, Bark M, Allibone E, Lilford RJ. Should maternal alphafetoprotein estimation be carried out in centres where ultrasound screening is routine? *Am J Obstet Gynecol* 1988;**158**:1092–9.

43 Neuhauser D, Lewicki AM. What do we gain from the sixth stool guaiac? *N Engl J Med* 1975;**293**:226–8.

44 Hillner BE, Smith TJ. Efficacy and cost-effectiveness of adjuvant chemotherapy in women with node-negative breast cancer. *N Engl J Med* 1991;**324**:160–8.

45 Weinstein MC, Schiff I. Cost-effectiveness of hormone replacement therapy in the menopause. *Obstet Gynecol Survey* 1983;**38**:445–54.

46 Kawachi I, Malcolm LA. The cost-effectiveness of treating mild-to-moderate hypertension: a reappraisal. *J Hypertension* 1991;**9**:199–208.

47 Beck JR, Pauker SG. The Markov process in medical prognosis. *Med Decision Making* 1983;**3**:419–58.

48 Kind P, Rosser R, Williams A. Valuation of quality of life: some psychometric evidence. In *The value of life and safety,* Jones-Lee MW, ed. Amsterdam: Elsevier, 1982.

49 Torrance GW. Utility approach to measuring health-related quality of life. *J Chronic Dis* 1987;**40**:593–600.