Br. J. clin. Pharmac. (1982), **14**, 765–768

STATISTICAL DECISION THEORY AND CLINICAL TRIALS

C.C. SPICER
Department of Mathematical Statistics and Operational Research, University of Exeter, Exeter

Statistical decision theory was first formalized by Wald in 1950. His work arose out of a dissatisfaction with the current theory of statistical tests which took no account of the consequences that might ensue from the choice of hypothesis. This point of view arose naturally in the industrial problem of testing batches of material and deciding on criteria for rejecting or accepting them for which he developed some of the first sequential tests. In more recent years the increasing use of statistical methods in industrial and business applications has led to a much wider use of decision theory and further development of its underlying basis.

The fundamental ideas involved are similar to those used by card players and other gamblers who relate, or attempt to relate, the cost of a decision to its probable consequences. For example, the odds on improving a given poker hand on the discard are known and a player who ignores them is likely to suffer for it in monetary terms. Decision theory extends this type of argument to more subjective consequences, while maintaining its quantitative nature.

Before discussing the place of decision theory in relation to the statistical methods usually applied to clinical trials it will be useful to examine a specific example to illustrate the techniques and concepts involved.

Consider the case of a patient in a casualty department with acute abdominal pain. A firm diagnosis has not been made and it is necessary to decide whether or not to admit for further investigation. The patient may be well enough to be sent home safely or ill enough to require further treatment. There are two decisions that can be made and two types of patient. (There could be a further decision which is to admit to an observation ward but in the present purely illustrative example we shall ignore this). The situation can be represented as a 2 × 2 table as follows:

		Admission not necessary	Admission necessary
Decision	Admit	C_{11}	C_{12}
	Discharge	C_{21}	C_{22}
		P	1 − P

where the Cs represent symbolically the possible consequences of the decisions and P is the probability that the patient is ill enough to need admission. The value of P can, in principle, be determined from past records of similar patients.

The most difficult and controversial problem is to assign quantitative measures of the four Cs. If the Cs could be defined simply in terms of money gains or losses or in some other quantitative measure such as days off work then the best decision is obviously that which gives the larger expected gain (or smaller loss). In the present instance there is no obvious way in which to assign numerical values to the Cs. Many people would agree that the worst consequence would be the discharge of a sick patient. The discharge of a subject who is well is highly desirable but some mistakes would have to be accepted to avoid discharge of patients who needed admission. Admitting a healthy subject is obviously undesirable but, again, it would be inevitable if a high probability of admitting a sick patient was to be achieved.

Decision theory provides a systematic and consistent procedure for assigning numerical scores to complex and subjective judgements of the kind involved. These scores are technically known as utilities.

The method of calculating utilities depends on two basic conditions. The first is that there is a clearly defined worst outcome (or outcomes) and a clearly defined best outcome (or outcomes). In the example above, to discharge a patient who should have been admitted is obviously worst. The best is probably to be admitted when it is necessary but some people might prefer the certainty of discharge if one was not sick. The second condition (known technically as coherence) is that the outcomes can be arranged in a consistent order, so that if A is judged better than B, and B is better than C, then A is judged better than C. For three or four outcomes this is usually easy but in more complicated cases can become very difficult. The use of the utilities demands the further condition that *all* the possible outcomes and their utilities should be enumerated. If these conditions are fulfilled it can be shown that the best and worst options may be given the arbitrary values of 0 and 1 without any loss of generality or information. The intermediate cases are then given utilities between 0 and 1 by a process of idealized betting. This may sound almost frivolous, but it is not unreasonable to assume

0306–5251/82/120765–04 $01.00

that the strength of one's feelings about the utility of an object are quite well reflected by the sort of lottery one would be prepared to enter in order to obtain it. The certain loss of, say, 10 pence might be acceptable for a 200:1 chance of winning a prize in the average charitable raffle. The fact that such losses are frequently accepted is a measure of one's support for the charity in question.

The calculation of utilities is carried out by asking the decision maker to take part in the following betting game: a counter is drawn from a bag containing a proportion p of black and (1 − p) of white counters. If the value of the most desirable outcome is B and that of the worst W then drawing a black counter is rewarded by a 'prize' of p × B. A white counter wins a 'prize' of (1 − p) × W. The conventional limits as stated above being B = 1 and W = 0. The player is asked to decide what proportion p he would accept as equivalent in its consequences to the *certainty* of a given outcome and this proportion is the numerical value of the utility. In terms of our example, for C_{22} this might take the form: 'What chance would you be prepared to accept of not being admitted to hospital when you were ill to achieve certainty that you would not be admitted if you were well?' If the betting is not consistent in its ordering of the utilities it can be demonstrated that the resulting decisions are disadvantageous. If the outcomes are measured in money, lack of coherence would in fact lead to loss of money. However it is a feature of decision theory that money is not necessarily related simply to utility. Most of us would prefer the certainty of a million pounds to a 50:50 chance of 2 million pounds or nothing, but a very rich man might not.

The process described requires much more justification than is possible in this article and anyone interested will finds books by Lindley (1971) and Raiffa (1974) give a full and relatively non-mathematical discussion of the issues involved, and it is fair to suggest that before condemning the procedure out-of-hand one should consider the alternatives. In many cases these consist of an intuitive judgement based mainly on a jumble of ill-defined and unmeasurable notions. It seems to me that there is much to be said for supplementing this with an attempt to frame the problem in the more precisely defined terms of decision theory. A considerable advantage of the latter is that they restate the problem as a set of relatively simple comparisons and compel the user to define his problem much more clearly. It must be said that the use of decision theory for really complicated problems is in its infancy. The basic theory seems to be fairly well established but there is a great shortage of practical techniques of the kind available in applied statistics, and a rather noticeable shortage of actual practical examples in the text-books as compared with those in books on statistical method.

The example used above for illustration is, of course, greatly over-simplified and unrealistic. It is worth remarking that about 40% of cases of acute abdominal pain do not require admission and that the decision to admit or discharge them is made thousands of times a week, often on rather nebulous grounds. On the other hand, the processes of clinical examination and diagnosis do not alter its basic structure. All that clinical examination can do is to alter the values of p and (1 − p) until one or the other is near zero. So long as uncertainty remains, the basic utilities are unaltered until a probable diagnosis or diagnoses are made. At this point the choice is more specific: the utility of being admitted for a perforated peptic ulcer would probably differ from that of acute cholecystitis or appendicitis and the possibility of wrong admission might be much more (or less) acceptable.

Clearly the reduction of a complex group of subjective factors such as pain, mental distress, or family complications, to a single numerical value may present great difficulties.

The exact way in which further information can be used to improve the choice of decision uses a fundamental statistical theorem named after the Reverend Thomas Bayes who discovered it in about 1760. Its use in the problem of testing statistical hypotheses was for many years in disrepute but its correct application is quite rigorous given the accepted axioms of the theory of probability.

In the present example the decisions are based on two complementary probabilities which can be determined empirically in the population under study. The process of clinical examination produces evidence in the form of signs and symptoms whose frequencies differ in the two diagnostic groups. Take, for example, the presence or absence of abdominal rigidity. In those requiring admission the frequency of this sign is about 80%: in those not requiring admission about 20%.

These probabilities are written symbolically as $P(S|C)$ which is the probability that the sign or symptom is present in patients belonging to category C. Bayes's theorem gives the relationship between this and the converse probability $P(C|S)$, that the patient belongs to category C when S is present. Namely

$$P(C|S) = \frac{P(S|C).\ P(C)}{P(S)}$$

where $P(S)$ is the overal frequency of the symptom S in the population under study.

The probability $P(C)$ is the frequency of the category C and is known as the prior or *a priori* probability of C. In the example above there are only two categories: A, requiring admission and D, fit for discharge, hence $P(D) = 1 - P(A)$, and

$$P(S) = P(A)P(S/A) + (1 - P(A))P(S/D).$$

Using the frequencies given above

$$P(S) = 0.6 \times 0.8 + 0.4 \times 0.2 = 0.56$$

$$P(A|S) = 0.6 \times 0.8/0.56 = 0.86$$

The ratio

$$\frac{P(A|S)}{P(D|S)} = \frac{P(S|A).\ P(A)}{P(S|D).\ P(D)}$$

expresses the fundamental relationship between the state of information before and after eliciting the symptom. If its incidence in the two groups is identical, i.e. if $P(S|A) = P(S|D)$ then the ratio of the prior probabilities is unchanged. In the present example the values are approximately: $P(S|A) = 0.8$, $P(S|D) = 0.2$ and the priors are modified by a factor of 4 if abdominal rigidity is present, by a factor of 1/4 if it is not.

In the example all the probabilities can be clearly defined as observable frequencies but most proponents of decision theory regard this as too narrow an approach and maintain that probabilities are, in general, measures of a subjective 'degree of belief' not necessarily related to any set of observable frequencies. Numerical values can be assigned to such probabilities by a betting process identical with that described above for utilities, resting on the same condition of coherence (consistency) with limits which can be given the values 1 and 0 without loss of generality. From this point of view utilities are in fact a type of probability. As in the case of utilities it is not possible to give here a detailed justification of the procedure and the reader is again referred to the literature. It should be clear, however, that the concepts are essentially individual and subjective, though it may be possible in many practical situations to work out a consensus value when several people are involved.

The methods described above can be applied to decisions based on clinical trials at roughly two levels: (a) For policy making on a large scale, e.g. by a drug company or a government department; and (b) By individuals or small groups of doctors to decide on actions to be taken in the management of individual patients.

The first of these categories is frequently met in the literature of decision theory. Interesting examples will be found in a recent number of the Journal of the Operational Research Society (**33**, Number 5) which illustrate the process of actually using the methods to determine utilities and probabilities. These examples show that utilities and subjective probabilities can be obtained in quite complex cases such as the siting of nuclear power stations or sites for the disposal of nuclear waste, in which a number of competing interest groups are involved. Simpler situations are correspondingly easier to deal with but it is, not surprisingly, often difficult to convince the non-specialist of the validity of the procedures. However, it seems that the act of setting out the options and their consequences explicitly can often be very useful in clarifying a problem. At the least this may be a valuable supplement to traditional procedures.

The results of clinical trials form an important but not exclusive part of some medical policy decisions. The current controversy about the use of whooping cough vaccine whose introduction rested on two clinical trials might be the subject of an interesting study in this connection. The utilities of the general public here seem to be different from those of the health administrators. Other examples are influenza vaccination and screening for breast cancer. There is good reason from clinical trials to believe that the former might be highly cost effective in terms of days off work but it is not very well accepted by individuals. On the other hand, screening for breast cancer has been shown by a well designed trial to reduce mortality in women over 50 and there is much public pressure for its general introduction. But from the point of view of the health administrators this may represent an enormous diversion of scarce resources. Problems such as these could be profitably studied from the viewpoint of decision theory.

In industry the theory might find application in deciding whether a trial justified research and development effort on a new compound.

The use of decision theory by clinicians in the case of individual patients seems to be quite practical given that the clinician *and* the patient are prepared to go through the process of estimating their utilities and personal probabilities. In many (perhaps most) instances, fairly simple decisions are involved such as whether to adopt a new treatment which is more efficacious and no less productive of adverse effects than the old one; so that none of the niceties of decision theory is required. Use of the theory becomes a possibility when the treatment is complicated by side effects and complications. Professor Card has shown that patients and doctors can be led to give their utilities by the betting procedures described above. More general adoption of these would, perhaps, provide a useful corrective to the crude assessment of the value of trial results in terms of survival times or other very simple measures. These considerations apply particularly to major procedures accompanied by much risk, pain and other complication such as pneumonectomy for lung cancer or endocrine ablation in late cancer of the breast.

The applications suggested would frequently need to be based on the adoption of decision theory techniques for estimating utilities. It is likely too that these will also be required for the estimation of personal probabilities. Many scientists will not be willing to do this though refusal to do so should not be made without examining more carefully the traditional procedures. Only experience can show how useful the theory is but those who try to apply it will often find that the exercise enforces a more critical and exact understanding of the way in which their decisions are arrived at.

The probabilities provided by the significance levels of the results and the power function of the test are not necessarily applicable if the decision maker feels that they should be modified in the light of other information. For example, the χ^2 test for homogeneity of a set of rates or proportions tests the hypothesis of homogeneity against the whole range of possibilities for inhomogeneity, many of which will be quite implausible and others inherently likely. Another consideration is that few experiments are actually carried out unless the scientist has reasonable cause to believe that their outcome will be either definitely positive or negative and the significance levels do not reflect this nor the fact that negative findings frequently remain unpublished.

The notions of decision theory are not fundamental to the main body of scientific thought. In scientific research an experiment need not lead to any specific decision, though it may suggest possibilities for other experiments, for modification of existing theories and improvements in methodology. Decision theory does not play any essential part in this process and in so far as a clinical trial is a scientific experiment the use of its results in a decision process may be comparatively unimportant. Decision theory is concerned with managerial types of problem, particularly those in which a decision has to be made in the light of incomplete or poorly defined and vague information of a kind not commonly acceptable in scientific research.

It should always be realized that all the possible decisions and their outcomes must be clearly set out. It is not sufficient simply to decide not to use a treatment without specifying what alternatives exist and their outcomes. At the present moment there is a growing movement towards introducing the methods of decision theory into clinical medicine (a journal and society for this have already been founded in the USA) but not enough practical experience has accumulated to define the place of formal decision theory in this field.

The ideas of subjective probability and utility can only be justified empirically but there does seem reason to think that they are sufficiently useful in clarifying the process of decision making to be worth trying. They do not, after all, demand more than an effort by the users to define more clearly the process by which their decisions are made.

Further reading

Professor D.V. Lindley's book 'Making Decisions' is a clear exposition of the basic ideas of decision theory. The textbook by Raiffa is also worth reading and gives references to the literature. A very thorough and balanced, but mathematical account is given in Cox & Hinkley's textbook, Chapter 11. A symposium printed in the Journal of the Royal College of Physicians (1975), **9**, contains a number of medical applications and expository papers. The number of the Journal of Operational Research quoted in this text is also worth consulting and gives many references.

References

COX, D.R. & HINKLEY, D.V. (1970). *Theoretical statistics*. London: Chapman and Hall.

LINDLEY, D.V. (1965). *Introduction to probability and statistics from a Bayesian viewpoint*. Part I – Probability, pp. 6–21. Cambridge University Press.

LINDLEY, D.V. (1971). *Making decisions*. New York: John Wiley Interscience.

RAIFFA, H. (1974). *Decision analysis*. Reading, Massachusetts: Addison-Wesley Publishing Co.