Toward a more ethical clinical Trial

J.B. KADANE and N. SEDRANSK
Carnegie-Mellon University
and
State University of New York at Albany

SUMMARY

Current methods of conducting clinical trials require the patient to agree to have his treatment assigned randomly, where his individual characteristics are taken into account only to balance the treatment groups. A Bayesian alternative involves eliciting the prior opinions of the group of clinicians who designed the study. Each patient is then guaranteed that the treatment he will receive is the best for him either in the opinion of at least one individual clinician or as a consensus of several, given the patient's characteristics and all the information available from the trial when the assignment is made.

Keywords: ETHICS, DECISION THEORY, UTILITY, ELICITATION OF OPINION, ACCEPTABLE TREATMENT, RADOMIZATION.

1. INTRODUCTION

In this paper, we explore the notion that every patient in a clinical trial should be assigned a treatment responsibly believed to afford therapeutic advantage to him.

The clinical trials considered here involve several different treatments administered to patients who arrive sequentially. In general, a patient must receive treatment shortly after arrival. Patients may be heterogeneous with respect to features or attributes recognizable prior to the determination of treatment and likely to affect prognosis. A well-defined measure of therapeutic efficacy is assumed.

For example, consider a study of patients admitted to a hospital's trauma unit where two new methods of preventing sepsis are being evaluated in comparison with surveillance. Intensive care is given to every patient in the

unit. Patients arrive following massive trauma and require immediate attention. Recognizable characteristics include type of trauma (accident or post-surgical complication; head injury or not; long-bone fracture or not) and type of patient (age; probably general physical condition immediately prior to trauma). Proportion of hospital time free from sepsis is taken to be the measure of efficacy.

Commonly in this type of clinical trial, a patient receives a treatment drawn at random from the set of treatments under study. The probabilities of selection for the treatments are fixed throughout the clinical trial and usually are equal. The randomness may be unconstrained, or an overall study design may be based on random selection of one of the possible permutations of a sequence of treatment assignments including a fixed number of assignments to each treatment. Further structure for the design may be imposed by generating separate sequences of assignments for different types of patients.

The most frequently cited motivation for these randomized designs is removal of the treatment selection from the control of the attending physician. This reduces his ability to manipulate the treatment assignments, and hence decreases the possibility of confounding effects of treatment and prognostic factors in the observed results. The use of constrained randomized designs is promoted in order to increase the efficiency of the study (to reduce the variance or mean squared-error of treatment effect comparisons). Also, hypothesis tests about treatment effects can be based on the permutation distribution induced.

Ethics for such a trial have been justified by arguments like the one given by Gilbert, McPeek and Mosteller (1977):

> "Let us consider the question of whether a present patient should give up something for future patients. We, or our insurance carriers, pay the monetary cost of our care. What we do not pay for is the contribution to the medical system by past patients. These patients, through their suffering and participation in studies, have contributed through their illness and treatments to the present state of evidence for all patients. Such contributions cannot be purchased by money but can be repaid in part by making, when appropriate, a contribution to the same system. One good way is through participation in well-designed clinical trials when the patient falls into the limbo of medical knowledge... Thus the patient has an interest not only in the trial he or she has the opportunity to engage in, but also a stake in a whole system that produces improved results that may well offer benefits in the future, if the patient survives the present difficulty. Thus, the social system will likely offer benefits through the larger system even when a particular component of the system may fail to pay off directly for a patient, his family, friends, or some other social group he belongs to".

Need for such an argument arises when some patients are asked to accept a less efficacious therapy under study in order to treat later patients more

knowledgeably. If he enters the trial shortly after it is begun, when differences among the effects of the several treatments may be imperceptible or unknown, the patient's sacrifice, if any, may be slight. However, if he enters the trial after a substancial amount of data is available, treatments which appeared equally likely to prove efficacious at the outset may no longer be equally desirable. In this case the patient's exchange of expected therapeutic benefit for expected information may be markedly to his detriment.

This approach asks the patient to accept whatever therapeutic disadvantage may come his way in the name of scientific progress. Such an emphasis on the greater social good relative to the legitimate interests of the patient is less than satisfactory. In this paper we seek an alternative which incorporates new information as it is generated by the trial to protect patients from inadvisable treatments.

2. MODELING THE CIRCUMSTANCES OF A CLINICAL TRIAL

A clinical trial may be proposed in order to reduce controversy about the relative merits of the therapies to be studied and/or to gain information about the efficacy of one or more of the therapies in the absence of any strong prior opinions. In a study to evaluate several therapies, there may be agreement among responsible scientists, physicians in this case, about some of the treatments, disagreements about some and lack of firm opinion about others.

In general, it is reasonable to assume that a set of prevailing opinions within the scientific medical community is represented by various physicians involved in the clinical trial. Establishing and defending criteria for selecting the "prevailing opinions" to use is outside the purview of this paper. It is usefull to express each of these opinions about the efficacies of the therapies studied as a probability distribution for the efficacy measure conditional upon the prognostic factors considered important by one or more of the physicians involved. Once "prevailing opinions" are expressed as distributions, the acquisition of data during the conduct of the study permits updating in the usual fashion. Thus at each point during the study, all opinions are "current" an important divergence from a classical, fixed design for a randomized trial.

Two distinct sets of utilities are involved in the conduct of a clinical trial. One, obviously, is defined in terms of accomplishing the study objectives, *i.e.*, reaching a consensus about preferred treatments and/or acquiring information about treatments. The role of this set of utilities is akin to that of efficiency measures or power function requirements for tests of hypotheses in conventional (non-Bayesian) designs for clinical trials. The second set of utilities is the set of patients' utilities. For each patient, this utility function represents his own valuation of therapeutic results; there is no apparent counterpart in conventional designs for clinical trials.

Much of the difficulty in reconciling ethics with efficiency in the design of clinical trials seems to arise from ignoring the patients' utilities, or from confusing the two sets of utilities with each other, or arguing as Gilbert, *et al.* do that it is reasonable to assume that the sets are the same.

The two sets of utilities govern our experimental design in different ways: the former as an objective function to be maximized, the latter as a constraint on the solution set.

In the formulations to be discussed, each patient's set of utilities is used to define an acceptable set of possible treatments for this particular patient. Then the selection of treatment from this set is made with respect to the overall scientific objectives of the study.

3. IDENTIFYING ACCEPTABLE TREATMENTS

An acceptable treatment for a patient is considered to be one which in some sense maximizes the patient's interests, where these are expressed as a utility function; an unacceptable treatment is one which in no sense maximizes the patient's expected utility. The patient may himself express a utility function or a general form for the utility function may be supplied for him. In either case, the patient himself is not assumed to have a prior opinion, although he has at his disposal the collection of current "prevailing opinions". Thus the expected utility for the patient reflects his own utility function and the expert opinion he consults. Following the custom of "seeking a second medical opinion", there is a set of expected utilities for a particular treatment corresponding to the set of (updated) prevailing opinions.

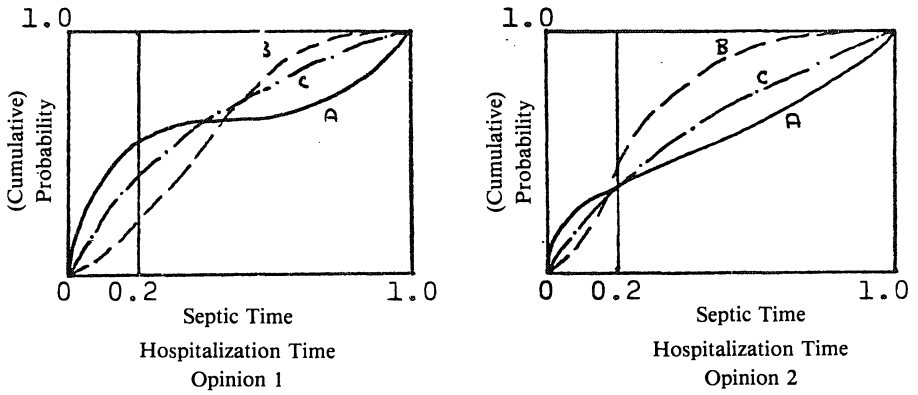
For a study of T treatments where P prevailing opinions are available, a particular patient will have a T x P array of expected utilities $[u_{tp}]$, where u_{tp} is the expected utility of each treatment t according to the updated opinion of expert p. Define a treatment t as *acceptable* if there is some set of convex weights for prevailing opinions $\{w_p\}$ satisfying $w_p \geq 0$ for all p and $\Sigma\, w_p = 1$, such that

$$\sum_p u_{tp}\ w_p \geq \sum_p u_{t',p} w_p \tag{1}$$

for all other treatments, t'. Acceptable treatments include the treatment most-favored by each expert, and possibly other, generally well-favored treatments. *We propose that each patient be guaranteed an acceptable treatment.*

As an example recall the sepsis-prevention study in the trauma unit. For

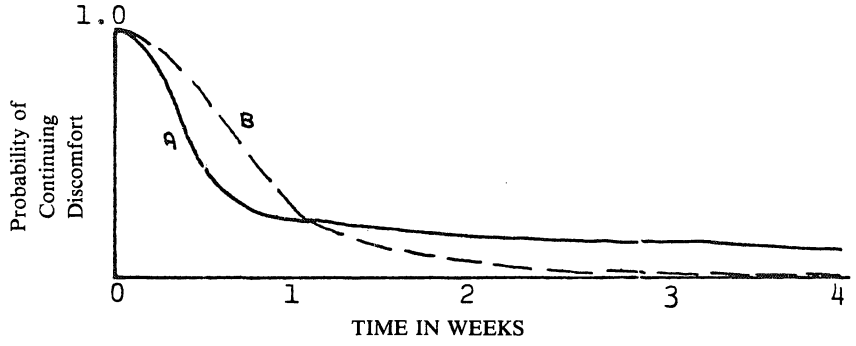
the three treatments, A, B, C, two hypothetical prior opinions of participating physicians are shown below for the proportion of hospitalization time likely to be spent in a septic state.



In this case the patient's utility function may be supplied for him, in view of his inability due to ignorance as well as to incapacitation, to express his own interests. Thus medical knowledge is imposed, in this case by recognition that total septic time in excess of 20 percent of hospitalization time is decisive for a patient's recovery. Hence expected utility is defined here to be the probability that less than 20 percent of hospitalization time is spent a septic state.

Using the prior opinions depicted, all treatments are acceptable, since the three vectors of weights $w_1 = (1,0)$, $w_2 = (0,1)$, $w_3 = (1/2, 1/2)$ satisfy inequality (1) for the three treatments, A, B, C, respectively.

However, even for identical patients the acceptable set of treatments may differ, due to differences in utility function. As an example consider a clinical trial of "soft" contact lenses. In this case, the objective may be the minimization of adjustment time (duration of accelerated eye-fatigue and increased eye-strain). A single expert opinion about the lenght of adjustment time might have the form depicted below.



Thus the expected utilities for two patients who measure utility differently, that is immediate adjustment (0.5 week of discomfort) and eventual adjustment (2 weeks of discomfort), will be maximized by lenses of types A and B, respectively. A third patient whose utility is defined in terms of one week of discomfort will be indifferent between the two types of contact lens. Consequently, the acceptable sets of treatments (lens types) differ for the three patients.

Finally, note that the acceptable set of treatments is defined for each patient, as the patient arrives. Thus the opinions used in the calculation of expected utilities are the original prior distributions (representing "prevailing prestudy opinions") *updated* by all accrued data. Hence, the definition of acceptable treatment is current for each patient at the time his treatment must be determined.

4. DESIGNING WITHIN THE CONSTRAINTS

Restricting treatment selection to the set of acceptable treatments does not, in general, completely specify the design for a clinical trial. Under this restriction the ethical considerations are satisfied for any design, therefore other criteria can be used to determine the design.

For example, a group of physicians committed to the idea of randomized clinical trials could use a random process to choose among acceptable treatments for a patient. In this case, definition of the proper permutation distribution and proper consequent analysis would be complicated greatly. Furthermore, it is logically inconsistent to discard philosophy and to ignore the experts' opinions in this aspect of the design.

Consider, therefore, a criterion based on the overall scientific study objectives, *i.e.,* reaching a consensus and/or acquiring information about the treatments. The relevant set of utilities is defined for the experts in terms of the information to be gained following treatment of the patient. Thus the treatment is selected from the acceptable set to maximize progress of the study, expressed as a function of the experts expected utilities.

A fully optimal sequential design would take into account the history of the clinical trial, including patient characteristics, assignments and results. It would require specification of a probability distribution characteristics of future patients. It would also require specification of a probability distribution for future patients' utility functions in order to consider the acceptable sets of treatments for the future patients. In face of such complexities, we restrict attention to myopic designs, treating each patient as if he were the last one to be studied.

There are two distinct aspects of defining the treatment selection procedure. First, a reasonable utility function must be determined for each expert.

Second, and conceptually more difficult, individual utilities must be agregated to form a group decision.

For a single expert, Raiffa and Schlaifer (1961) propose choosing treatment t to maximize

$$\int_X dx \max_d \int_\Theta d\theta \, V(d,\theta,t,x)\, p(\theta \mid x,t)\, p(x \mid t) \tag{2}$$

where $\theta \in \Theta$ is the parameter, $x \in X$ is the outcome of the experiment, V is the expert's utility function, $d \in D$ is the decision reached by the clinical trial, in Lindley's (1971) notation. Here D is the set of possible recommendations of treatments at the conclusion of the clinical trial. Note that D may include decisions of equivalence among a subset of preferred treatments, as well as selection of a single recommended treatment for each patient. Good (1956), Lindley (1956) and Lindley (1971) suggest maximizing expected information over possible experiments, in this case possible treatment selections. Bernardo (1979) shows that maximizing information can be treated as a special case of maximizing expected utility.

In general, expected utilities for experts will differ because of initially differing opinions, whether or not the experts' utility functions have a common form.

Suppose that treatment t belongs to the acceptable set for the current patient. Denote by V_{tp} the expected utility of treatment for the expert holding prevailing opinion p. For the first patient, without loss of generality, $0 \leq V_{tp} \leq 1$ for all t and p, since $\{V_{tp}\}$ are unique only up to a positive linear transformation (see Savage, 1954) and hence can be standardized with $\max_t V_{tp} = 1$ and $\min_t V_{tp} = 0$ for each p. This standardization may be repeated with each subsequent patient.

Alternatively the expected utilities for the first patient to enter the trial can be standardized in the foregoing manner. Then the standardization coefficients for each expert can be used throughout the remainder of the study. In this case the range restriction on V_{tp} will not necessarily hold for patients after the first.

Selection of treatment can then be made to maximize a suitable function of $\{V_{tp}\}$. For example, one treatment selection procedure is given by

Choose t to maximize $\min_p V_{tp}$.

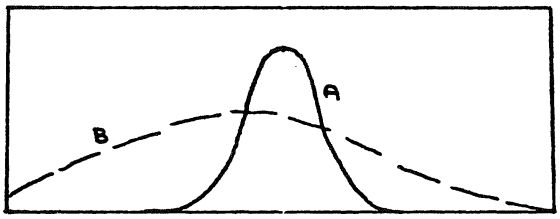
Clearly, a variety of other measures of aggregate utility are possible, as well.

5. CRITIQUE

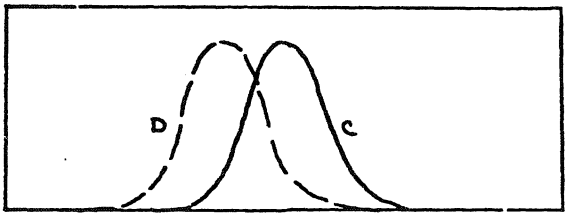
For clinical trials conducted in an atmosphere of conflicting views the formulation given thus far seems workable. However, studies undertaken in

the absence of prior information are vulnerable to premature discontinuation or to unwarranted degeneration of the design. In the definition of the set of acceptable treatments, this difficulty arises from the limitations of determining expected utility as a single number for each treatment. The result is an (unnecessary) narrowing of the definition of acceptable treatment, with resultant loss of flexibility in the overall study design.

Consider two possible situations, one in which there is very little information about one of the two treatments, the other in which there is ample information about both. For a particular expert, the prior opinions for these two cases might be depicted by the densities shown below.



I: PRIOR DENSITIES
TREATMENTS *A* AND *B*



II: PRIOR DENSITIES
TREATMENTS *C* AND *D*

There are smooth monotone utility functions for which the expected utilities calculated for case I and II are the same, with treatments *A* and *C* considered acceptable. In Case II, the rejection of treatment *D* may be considered desirable; whereas in Case I it would be desirable to consider both treatments *A* and *B* acceptable.

It is useful that different patients may express different utility functions resulting, for example, in Treatment *A* being acceptable for some patients, Treatment *B* being acceptable for others, as in Case I. However, one objective of a clinical trial design is that the study be viable without dependence upon a broad distribution of patient utility functions. For the case where diagnosis is

inconclusive, Lindley (1975) describes circumstances permitting valid inference, despite the identical utility functions for all patients. This case is considered further by Good (1978).

Several alternatives deserve investigation. Note that the difficulty arises when the treatment with smaller expected utility also has the more diffuse prior distribution. This suggests that Treatments A and B may both be considered acceptable because their expected utilities differ by less than some $\epsilon > 0$.

Since a resolution of this problem acts as a governance on the study, adequate solution is essential to the viability of the method.

6. IMPLEMENTING A CLINICAL TRIAL

Many of the essentials for carrying out a clinical trial are available now; others require only moderate efforts to be developed. Representations of "prevailing opinions" must be elicited from physicians holding these views. When then measure of efficacy can be assumed to have a normal distribution or a lognormal distribution, a member of the conjugate prior family can be elicited using the methods of Kadane *et al.* (1978) in the univariate case or of Dawid *et al.* (1979) in the multivariate case.

Updating of prior distributions can be done automatically as data is acquired; and for the conjugate family this can be accomplished quite easily. The major technical difficulty in this regard is the incorporation of censored observations, particularly when the lognormal model is used. Seeking adequate approximations may provide the most effective solution to this problem.

Substantial commitment of programming effort will be required to develop and implement efficient algorithms for the definition of the set of acceptable treatments and for the treatment selection procedure.

Finally, careful selection for a pilot effort should include the following favorable circumstances: primary objective of resolving sharp conflict of opinion (case where the formulation seems to have least vulverability), modest rate of accrual of patients, and single prominent measure of efficacy with clear definition.

7. CONCLUSION

In reply to John Tukey (1977),

> "Many of us are convinced, by what seems to me to be very strong evidence, that the only sources of reliable evidence about the usefulness of almost any sort of therapy or surgical intervention is that obtained from well-planned and carefully conducted randomized, and, where possible, double-blind experiments [see review papers of Byar *et al.* (1977) and Peto *et al.* (1977)]. Dare we prevent outselves

from obtaining reliable evidence?",

the only word we question is "randomized".

8. ACKOWLEDGEMENTS

This research was supported in part by the National Science Foundation under grant SOC-77-27575. The authors are indebted to Morris H. DeGroot and James M. Dickey for helpful discussions before the meeting in Valencia.

REFERENCES

BERNARDO, J.M. (1979). Expected information as expected utility. *Ann. Math. Stat.* **7**, 686-690.

BYAR, D.P., SIMON, R.M., FRIEDEWALD, W.T., SCHLESSELMAN, J.J., DEMETS, D.L., ELLENBERG, J.N., GAIL, M.H. and WARE, J.H. (1976). Randomized clinical trials: Perspectives on some recent ideas. *N. Engl. J. Med.* **295**, 74.

DAWID, A.P., DICKEY, J. and KADANE, J.B. (1979). Matrix t and Multivariate t Assessment. *Tech. Report,* Carnegie-Mellon University

GILBERT, J.P., MCPEEK,B. and MOSTELLER, F. (1977). Statistics and Ethics in Surgery and Anesthesia. *Science* **198**, 684-689.

GOOD, I.J. (1956). Some terminology and notation in information theory. *Proc. IEEE Part C* **103**, 200-204.

— (1978). Ethical Treatments. *J. Statist. Comput. Simul* **7**, 292-295.

KADANE, J.B., DICKEY, J., WINKLER, R., SMITH, W. and PETERS, S. (1978). Interactive Elicitation for the Normal Linear Model. *Tech. Report,* Carnegie-Mellon University.

LINDLEY, D.V. (1956). On a measure of the information provided by an experiment. *Ann. Math. Stat.* **27**, 986-1005.

— (1971). *Bayesian Statistics; A Review.* Philadelphia: SIAM.

— (1975). The effect of ethical design considerations on statistical analysis. *Applied Statist.* **24**, 218-228.

PETO, R., PIKE, M., ARMITAGE, P., BRESLOW, N.E., COX, D.R., HOWARD, S.V., MANTEL N., MACPHERSON, K., PETO, J. and SMITH, P.G. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. *B.J. Cancer* **34**, 585; **35**, 1.

RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory.* Boston: Harvard Business School.

SAVAGE, L.J. (1954). *The Foundations of Statistics.* New York: Wiley.

TUKEY, J.W. (1977). Some Thoughts on Clinical Trials especially Problems of Multiplicity. *Science* **198**, 679-684.