STATISTICS IN MEDICINE
Statist. Med. **17**, 2675–2683 (1998)

ATTITUDES IN CLINICAL TRIALS

PETER ARMITAGE*

2 Reading Road, Wallingford OX10 9DP, U.K.

SUMMARY

Two important papers by Schwartz and Lellouch have drawn distinctions between explanatory and pragmatic attitudes to clinical trials, and between individual and collective ethics. The pragmatic approach accords with the widespread use of 'intention-to-treat' analyses, but recent research has started to explore models for non-compliance adjustment which seek an explanatory approach. Individual and collective ethics often co-exist, and standard approaches to randomization may provide the most satisfactory compromise. © 1998 John Wiley & Sons, Ltd.

1. INTRODUCTION

Some 30 years ago the French statisticians, Daniel Schwartz and Joseph Lellouch, published two seminal papers[1,2] on what might loosely be termed the 'philosophy' of the clinical trial. Each of these presented the clinical trialist with a choice between two approaches. The choice in Schwartz and Lellouch[1] was between two attitudes towards the purpose, and consequently the design, analysis and interpretation, of a trial. In Lellouch and Schwartz[2] the authors contrasted two approaches to the ethics of medical experimentation, of which the clinical trial is a particular form.

Before discussing these papers in more detail, I need to explain why this topic is peculiarly appropriate to a collection of papers celebrating Michael Healy's distinguished career in medical statistics. There are two interlinked reasons. First, his own work, and his reactions to that of other workers, have always been tempered by an interest in the logic and purposes of statistical investigation. This interest has been manifested not in the complexities of axiomatic probability theory, but rather in a critical attitude towards the intentions of the investigator and the ability (or inability) of standard statistical procedures to satisfy these intentions.[3,4]

Secondly, Michael Healy is among the least insular of British statisticians. A fluent linguist and admirer of Latin culture, he was especially sympathetic towards the incisive and analytic approach of Schwartz and Lellouch. Indeed, he had been a member of the Controlled Clinical Trials Committee of the Union Internationale contre le Cancer (UICC) chaired by Daniel Schwartz, and he translated both the French text published in English as Schwartz and Lellouch[1] and the book by Schwartz *et al.*[5] which presented these ideas to a larger audience.

* Correspondence to: Peter Armitage, 2 Reading Road, Wallingford OX10 9DP, U.K.

CCC 0277–6715/98/232675–09$17.50

2. EXPLANATORY AND PRAGMATIC ATTITUDES

Schwartz and Lellouch[1] draw a clear distinction between two sorts of objectives, for which a clinical trial might be proposed:

(a) *The explanatory attitude*: an attempt to understand precisely the relative effects of alternative treatments when administered under ideal conditions.
(b) *The pragmatic attitude*: a comparison of treatments made under the less rigorous conditions likely to be found in practice.

The explanatory approach may be said to simulate laboratory conditions, and the pragmatic approach to resemble the conditions of normal clinical practice. The terms *efficacy* and *effectiveness* are often used for the two types of comparison, following Cochrane.[6]

Although the distinction was first articulated clearly by Schwartz and Lellouch, it was not entirely unfamiliar to clinical investigators. Marks[7] (p. 185), in a detailed discussion of an abortive trial of low-fat diet in the prevention of heart disease during the 1960s (the Diet-Heart Study), describes the differences of attitude amongst proponents of the trial:

> 'The diet-heart researchers found themselves trying to reconcile in one study the tensions between two intellectual traditions: one, a pragmatic public health tradition aimed, in the final analysis, at changing public policy and social behavior; the other, a research tradition intended to establish a rigorous scientific basis for all therapeutic practices.'

Schwartz and Lellouch exemplify the problem by considering various aspects of the design and analysis of a clinical trial:

(i) *Definition of treatment regimens.* In the treatment of cancer, is it useful to precede radiotherapy by a 30-day period of administration of a drug? The explanatory approach might require a control group to have an initial 30-day period without active treatment, so that the radiotherapy starts at the same time as in the test group. The pragmatic approach would be to start radiotherapy immediately in the control group.
(ii) *Assessment of results.* The explanatory approach will lay emphasis on outcome measures of biological significance, such as tumour regression or biochemical changes. The pragmatic approach will emphasize practical measures such as functional capacity, relief of symptoms and duration of survival.
(iii) *Choice of subjects.* In the explanatory approach, subjects may be highly selected so that most or all can follow the strictly defined treatment regimen; departures from this regimen damage the authority of the study. In the pragmatic approach, a more heterogeneous patient population is acceptable, and treatment regimens may be more loosely defined so as to accommodate the needs of individual patients. The treatments are then best regarded as 'strategies', and patients are retained in the groups to which they were assigned (the policy now usually called 'intent(ion)-to-treat').
(iv) *Method of comparison.* The classical apparatus of significance testing is regarded as appropriate for explanatory trials. For pragmatic trials, though, a decision may be needed as to which of a number of rival treatments should be used in future. In that case, the relevant measure of operational effectiveness is the 'error of the third kind', the probability that the wrong treatment is chosen, when the true difference is more than some specified

amount. Indeed, the whole apparatus of significance tests should preferably be replaced by recourse to some form of decision theory.

Schwartz and Lellouch tend to argue that the two attitudes correspond to a clear choice between two sorts of trial. It may be more realistic to suggest that the two attitudes are likely to co-exist, and to compete for ascendency, in any one trial, for the investigators planning a trial may well wish to kill two birds with one stone ('faire d'une pierre deux coups') – to gain information about biological mechanisms, and to assess strategies of treatment. In so far as a choice has to made, it is fair to say that current practice leans very heavily in the pragmatic direction. Ethical and financial constraints are likely to prohibit any but a very small randomized trial of treatment regimens which were not plausible therapeutic strategies.

In a predominantly pragmatic trial, are we led irrevocably to the conclusions (i)–(iv) listed previously? There is little doubt about (i)–(iii); indeed, the 'intention-to-treat' approach mentioned in (iii) has almost reached the status of a 'sacred cow'. (I discuss this in more detail below.) There is more scope for debate about (iv). The concept of a clinical trial as a selection procedure has been much discussed[8] since the early paper by Colton.[9] It now seems over-simplistic; clear decisions to change therapies rarely follow as an immediate consequence of a clinical trial, particularly on the mere basis of the *sign* of a difference which is not significant. Even with pragmatic treatment choices, probabilistic assessments of some form are important, and will play their part in the interpretation of results, both by the investigators and by other medical practitioners.

The explanatory/pragmatic antithesis bears some resemblance to the contrast drawn by Healy[3] between science and technology – between the search for truth, and 'the efficient achievement of pre-defined ends'. Healy maintains persuasively that statistics is 'best considered as a technology rather than as a science'. On further reflection, though, the analogy breaks down. From Healy's standpoint, both explanatory and pragmatic trials fall into the 'technology' half of his classification. An explanatory trial, à la Schwartz–Lellouch, may provide some insight into biological mechanisms, or suggest hypotheses for further laboratory studies, but can hardly be said to provide fundamental information on, say, biochemical reactions or DNA structure. The science/technology contrast is more akin to the distinction drawn by Claude Bernard in the 19th century between laboratory science and statistics, to the detriment of the latter. We can accept the distinction, and the implied limitations of statistical investigation, without in any way depreciating the contributions of statistical investigations, and clinical trials in particular, to the technology of therapeutic medicine – as helping to show what is *useful*, rather than what is *true*.

3. INTENTION TO TREAT

The pragmatic attitude is often taken to be exemplified *par excellence* by the intention-to-treat approach to the analysis of results, whereby comparisons are made between the outcomes for the complete groups assigned to different treatment regimens, irrespective of the extent of departure from the treatment schedules laid down in the protocol. Professors Schwartz and Lellouch have pointed out (in a personal communication) that this identification would go beyond their intentions. They would support the adoption of the intention-to-treat approach only if this was authorized in the trial protocol, as defining the strategy under study. In other circumstances they would, I think, regard protocol deviations as regrettable, and as a mark of unsatisfactory research methods.

In almost any clinical trial, though, departures from protocol are likely to occur to some extent. A patient may experience unwanted side-effects; deterioration in the patient's condition may lead the physician to prescribe alternative treatments; or the patient may decline to co-operate for any of a number of reasons. The trial statistician, therefore, is inevitably faced with the problem of how to take them into account in the analysis. In a trial conceived of as essentially pragmatic in nature, the investigators are likely to lean towards an intention-to-treat approach. The arguments usually adduced for this approach are:

1. The benefits of randomization are maintained; differences in outcome between the groups cannot be ascribed to systematic differences in the pre-treatment characteristics of the patients.
2. The comparison is essentially one of different strategies of treatment, defined by ideal schedules but with the recognition that, in the trial just as in routine clinical practice, departures from these schedules will occasionally occur. In this sense, the trial may be said to simulate routine practice.
3. Any attempt to measure relative *efficacy*, by comparing groups of patients with 100 per cent compliance to protocol, is deeply suspect because of selection biases. Patients failing to comply are likely to differ in number, and in pre-treatment characteristics, between different regimens, and comparisons between the reduced groups of compliers do not enjoy the benefits conferred by randomization.

Possible counter-arguments to points 2 and 3 are:

2*. The extent of non-compliance in a trial is likely to be different from that in routine practice, so the simulation of the latter by the former is of dubious value.
3*. Although a *per protocol* analysis comparing compliers is not fully randomized, and is therefore subject to selection bias, it might be better to accept such bias (particularly if it could be shown to be small), in order to achieve a measure of relative efficacy; in other words, to attempt an explanatory approach.

Underlying 3* is the often-expressed view that an intention-to-treat analysis is likely to underestimate 'true' differences in effect, because the non-compliers will dilute whatever effect might have been revealed by the compliers. It is sometimes asserted that physicians involved as investigators in a clinical trial find the intention-to-treat approach unnatural, presumably because they take an explanatory rather than a pragmatic view of the purpose of the trial.

Several recent research papers have presented possible methods for obtaining reliable measures of relative efficacy from trials in which non-compliance occurs, in spite of the selection biases mentioned above. One very simple model (stated in Sommer and Zeger[10] to have been first proposed by Tarwotjo *et al.*[11]) relates to a trial of an active treatment (say, a drug) against a control (possibly a placebo), with a binary outcome (success or failure). It is supposed that the success proportions are known separately for compliers and non-compliers in the drug group, but only for the total control group. We are, of course, unable to identify the patients in the control group who would have been compliers if they had received the drug.

It is, however, easy to estimate the success probabilities for drug compliers, separately for those assigned to the drug and for those forming part of the control group, and hence to estimate the treatment effect for drug compliers. As we should expect, this estimate differs from that obtained from an intention-to-treat analysis of the two complete treatment groups,

and also from an inappropriate comparison of the complete control group with the compliers in the drug group.

One rather crucial assumption in this model should be noted. The success probability for the drug non-compliers is assumed to be the same whichever treatment they are assigned to. This may be reasonable for a 'hard' endpoint such as survival, but may be less so for 'softer' endpoints such as improvement in symptoms. The experiences of being assigned to a control group and receiving an inert treatment, or being withdrawn from an active treatment regimen, may be quite different in their subjective effects and lead to different outcome measures.

A generalization of this model refers to a comparison of two active treatments, A and B, with three categories of patients: non-compliers who will receive A whatever their assignment; those who will receive B whatever their assignment; and compliers who will follow the assigned treatment. Again, the endpoint is binary. A simple solution is available, but, as before, the model involves crucial assumptions. The non-compliers who receive A are assumed to have the same outcome whether they are assigned to A or whether they default from a B assignment; and similarly with those receiving B. Moreover, the basic assumption that non-compliers with the assigned treatment merely transfer to the other treatment is unduly simplistic. Protocol departures can take many forms. In some instances a patient may take the assigned treatment, but in non-prescribed form; in others, the assigned treatment may be dropped and replaced by a regimen different from either A or B.

Another defect of these particular models is the specification of compliance as a binary characteristic; a patient is supposed either to comply or not. In practice, some examples of non-compliance are more serious than others. A patient may depart from the protocol in some relatively minor way, clearly of less importance than a wholesale abandonment of a prescribed medicament. In many chronic disease trials, patients are required to have regular doses of a prescribed drug, and the proportion of doses missed may vary considerably from one patient to another.

Many authors have presented models in which compliance is measured as a proportion of the assigned dose actually taken. Efron and Feldman,[12] in a paper followed by an instructive discussion, consider a trial of drug versus placebo, each group providing a regression of the outcome variable on the compliance measure. The distribution of compliance differs between groups, and the authors assume that patients would be ranked for compliance in the same order, in whichever group they were placed. This assumption enables a comparison to be made between the efficacy of the two groups. Goetghebeur and Lapp[13] consider a similar trial in which compliance is not measured for individual patients in the control group, additional information now being provided by inclusion in the model of a suitable set of baseline covariates.

Sheiner and Rubin[14] present a spirited attack on 'intention-to-treat' analyses, advocating the wider use of model-based analyses of efficacy. It is, though, possible to offer an unreserved welcome to research on model-based methods, while cautioning against too ready an assumption that the need for 'intention-to-treat' has disappeared. Compliance is a complex and multifaceted phenomenon, not necessarily summarized by a single variable. The reasons for compliance are equally complex, and include the possibility (not commonly discussed) that it may depend on, rather than influence, a patient's health status. Models capturing the salient features of compliance are likely to vary in character from trial to trial, so that their use will be time-consuming and their nature exploratory. Research is

needed on methods of analysis for more complex situations. Ideally, in any one trial, several different models might well be used, to present a sensitivity analysis which would carry more conviction than the results of applying a single model the basis for which might be easily criticized.

A recent issue of *Statistics in Medicine*[15] contains a number of useful papers on the analysis of non-compliance in clinical trials.

4. INDIVIDUAL OR COLLECTIVE ETHICS

Lellouch and Schwartz[2] draw an antithesis between two approaches to the ethics of clinical trials:

(a) *individual ethics*, in which the interests of the individual patient are paramount;
(b) *collective ethics*, stressing the benefits to society, or more narrowly to the population of patients who might be exposed to whichever treatment is chosen.

They distinguish starkly between these two attitudes, for each of which they outline an approach to the design and analysis of trials. For simplicity they consider in detail only trials to compare two treatments, with a binary outcome measure.

They see collective ethics as giving rise to a selection procedure of the type referred to earlier. Following the models studied by Colton[9] and others, the population of eligible patients (the 'horizon') has size N, and the task is to determine the optimal size, $2n$, of a randomized trial, and perhaps the number of patients allocated to each treatment and/or the method of assignment. The authors adopt a Bayesian approach, to allow for different degrees of prior knowledge, and obtain some theoretical and computational results for certain cases. They consider the effect of a type of sequential design for the trial, but express some reserve about its merits, on the grounds that just before a sequential trial is stopped there will be strong evidence in favour of one treatment, and the last few patients given the other treatment will be treated inhumanely:

> 'L'optique "éthique collective" est ici évidente: les derniers malades, du moins la moitié d'entre eux, servent véritablement à la communauté.'

Lellouch and Schwartz discuss briefly the generalization to the 'two-armed bandit problem' (enquiring in passing whether the French equivalent should be 'bandit à deux bras' or 'bandit à deux armes'). Here,[16] the assignment is made individually for each patient, but still on a basis which optimizes results for the population rather than the individual patient.

Their prescription for individual ethics is to assign to any patient that treatment which is judged to be the better. This judgement is based quite simply on the posterior means of the success probabilities for the two treatments, taking into account the prior distributions and the observed results so far in the trial. Some simulations suggest that, even judged on the 'collective ethics' criterion of the expected number of patients correctly treated, this rule does not perform too badly. It does, though, have the tendency to yield very unbalanced numbers on the two treatments, and so may provide rather poor statistical information about the magnitude of the treatment effect.

This solution for individual ethics departs widely, perhaps dangerously, from the ideal of a randomized trial, and perhaps should not be regarded as a controlled trial at all. Lellouch and Schwartz recognize the problem, and somewhat reluctantly come down in favour of the

'collective ethics' solution:

> 'Aussi, si avec quelques regrets on abandonne l'éthique individuelle pour l'éthique collective, ce n'est pas parce qu'elle conduit à un bénéfice global moindre, mais parce qu'elle semble devoir conduire à des essais non correctement menés.'

5. COMMENTARY

There can be little doubt about the continued validity and importance of the ethical distinction drawn by Lellouch and Schwartz. Lindley,[17] in a book review, writes

> 'The trial is conducted for social reasons, namely to acquire knowledge about the drugs for the future benefit of society. Yet the trial is conducted on individuals whose utilities are different from those of society. It is considered undesirable to give an individual a supposedly inferior treatment, thereby losing personal utility, in order more effectively to gain scientific information and to increase social utility. It is this contrast between social and personal utility that dominates the conduct of trials … '

One's main reservation about the Lellouch–Schwartz scenario is that the contrast between the two attitudes is too highly polarized. The authors themselves seem to sense this, remarking 'Notre impression – actuelle – est qu'il n'y a pas beaucoup de différence, du point de vue des pourcentages de succès entre les deux éthiques, d'autant plus que ces deux attitudes ne peuvant être jamais totalement pures: … '. The contrast with which they are concerned is perhaps relevant to a wider range of medical practice than merely the clinical trial. The late Professor Geoffrey Rose pointed out to me that it is implicit in the acceptance of surgical operations by trainee surgeons. It would be to the benefit of an individual patient not to be the target for the first operation conducted by a trainee; yet the needs of society require that inexperienced surgeons should gain the experience they initially lack. If one accepts that the interests of both the individual patient and society need to be safeguarded, it seems to follow that the specific models proposed by Lellouch and Schwartz for their two approaches are both unsatisfactory. The decision-theoretic horizon model (apart from any uncertainties about its specific formulation or about the values of its parameters) runs the risk of occasional assignment of a patient to a treatment which is almost certainly inferior, thus ignoring individual ethics. The model for adaptive allocation on the basis of posterior probabilities leads to inefficient and unreliable experiments, thus ignoring collective ethics. What seems to be needed is an approach which recognizes and respects both ends of the spectrum.

Strangely, and paradoxically, the approach to clinical trials most widely adopted in practice appears to a large extent to fulfil this need. On the one hand, the modern clinical trial is, at its best, an efficiently organized controlled experiment, carefully planned and executed, and capable of interpretation by valid statistical analysis. On the other hand the interests of individual patients are safeguarded by the use of informed consent, the provision for patient withdrawal if necessary, and the almost ubiquitous data and safety monitoring procedures.

To what extent are the demands of individual ethics transgressed by randomization? The 'classical' view of the ethics of randomization was perhaps adumbrated most strongly by Bradford Hill in a number of publications collected in reference 18. A trial was justified if, and only if, the investigators and physicians in charge of patients were genuinely uncertain which treatment was preferable. In these circumstances they would be prepared to administer any of the

rival treatments to any patient, and it was clearly better to do so in a way (i.e. by randomization) that permitted valid inferences about treatment effects.

Probably most current trial investigators accept this approach. There are, though, some problems to be resolved:

(i) Is a physician ever in a state of complete 'equipoise', with no preference one way or the other? If such a state is possible, will it not be shattered after the results are obtained from the first patient in the trial?
(ii) If uncertainty can continue as the trial progresses, what are its limits? There is a continuous spectrum between an absolutely blank mind and an uneasy feeling that A is probably better than B.
(iii) As evidence accrues in a trial in favour of one treatment, an initially agnostic attitude will gradually weaken. When, and how, should a decision be taken to abandon randomization?

The dilemmas indicated by (i) and (ii) may be illusory. A physician's judgement as to the relative merits of two or more treatments is often highly multivariate. There may be many measures of treatment efficacy, long- and short-term; side-effects of various kinds need to be considered, and many of these may be long-term; the degrees of acceptability by the patient and ease of administration by the doctor may not be immediately apparent; and so on. In a trial in which one or more treatment is relatively untried, many of these issues will be resolved only by extensive observation. The physician may well, therefore, remain for a considerable time uncertain which treatment is, on balance, to be preferred.

Frequently, though, there will be one major efficacy endpoint, and accumulating evidence that one treatment is more effective than another will, in due course, give rise to dilemma (iii). Investigators in the same trial may well move away from the region of uncertainty at different rates, varying in their prior judgements, the weights attached to different criteria, psychological characteristics, and so on. Some of this variation may become submerged by the taking of collective decisions, or by delegating data monitoring to an independent committee whose responsibility it is to make recommendations about stopping or continuing the trial. Ultimately, though, assessments about the limits of uncertainty must be personal matters, and a particular member of a collaborating team may need to take unilateral action in deciding to stop randomization.

Decisions about stopping may be helped by statistical calculations of various sorts: Bayesian computations to combine prior judgements with observed information; stopping rules based on boundaries with known frequentist properties; subgroup analyses; and so on. However, these should not be automatic decision rules. Termination of a trial must depend on multiple factors, many of which may not be foreseeable at the outset of a trial and are unlikely to be comprehensively covered by any mathematical model. Ultimate responsibility rests with the investigators, for the termination as well as start of a clinical trial.

The viewpoints expressed in this article will not be accepted by all those who participate in trials or who study the theoretical aspects of trials. I do not know whether they are in any way in line with the views of Michael Healy, Daniel Schwartz and Joseph Lellouch. What is, I think, indisputable, is that the improvement of clinical practice relies greatly on the conduct of clinical trials, and that the promulgation of trials is stimulated and enhanced by lively discussion about their logical foundations. In this discussion, the contributions of these three authors have been important and influential, and I am glad to be able to pay tribute to them.

ACKNOWLEDGEMENTS

I am grateful to Daniel Schwartz and Joseph Lellouch for comments on the first draft of this paper.

REFERENCES

1. Schwartz, D. and Lellouch, J. 'Explanatory and pragmatic attitudes in therapeutic trials', *Journal of Chronic Diseases*, **20**, 637–648 (1967).
2. Lellouch, J. and Schwartz, D. 'L'essai thérapeutique: éthique individuelle ou éthique collective?', *Revue de l'Institut Internationale de Statistique*, **39**, 127–136 (1971).
3. Healy, M. J. R. 'Is statistics a science?', *Journal of the Royal Statistical Society*, *Series A*, **141**, 385–393 (1978).
4. Healy, M. J. R. 'Does medical statistics exist?', *Bulletin in Applied Statistics*, **6**, 137–182 (1979).
5. Schwartz, D., Flamant, R. and Lellouch, J. *L'essai Thérapeutique chez l'Homme*, Editions Médicales Flammarion, Paris, 1969. Translated by M.J.R. Healy as *Clinical Trials*, Academic Press, London, 1980.
6. Cochrane, A. L. *Effectiveness and Efficiency*: *Random Reflections on Health Services*, Nuffield Provincial Hospitals Trust, London, 1972.
7. Marks, H. M. *The Progress of Experiment*: *Science and Therapeutic Reform in the United States*, 1900–1990, Cambridge University Press, Cambridge, 1997.
8. Armitage, P. 'The search for optimality in clinical trials', *International Statistical Review*, **53**, 15–24 (1985).
9. Colton, T. 'A model for selecting one of two treatments', *Bulletin de l'Institut International de Statistique*, **39** (3), 185–200 (1962). (Also in *Journal of the American Statistical Association*, **58**, 388–400 (1963)).
10. Sommer, A. and Zeger, S. L. 'On estimating efficacy from clinical trials', *Statistics in Medicine*, **10**, 45–52 (1991).
11. Tarwotjo, I., Sommer, A., West, K. P., Djunaedi, E., Loedin, A. A., Mele, L. and Hawkins, B. 'Influence of participation on mortality in a randomized trial of vitamin A prophylaxis', *American Journal of Clinical Nutrition*, **45**, 1466–1471 (1987).
12. Efron, B. and Feldman, D. 'Compliance as an explanatory variable in clinical trials' with discussion, *Journal of the American Statistical Association*, **86**, 9–26 (1991).
13. Goetghebeur, E. and Lapp, K. 'The effect of treatment compliance in a placebo–controlled trial: regression with unpaired data', *Applied Statistics*, **46,** 351–364 (1997).
14. Sheiner, L. B. and Rubin, D. B. 'Intention-to-treat analysis and the goals of clinical trials', *Clinical Pharmacology and Therapeutics*, **57**, 6–15 (1995).
15. Goetghebeur, E. and Van Houwelingen, H. (eds). 'Proceedings of Symposium on Analyzing Non-compliance in Clinical Trials', *Statistics in Medicine*, **17**, 247–389 (1998).
16. Berry, D. A. and Fristedt, B. *Bandit Problems*: *Sequential Allocation of Experiments*, Chapman and Hall, London, 1985.
17. Lindley, D. V. 'Review of *Bayesian Methods and Ethics in Clinical Trial Design*, J. B. Kadane (ed),' *Statistician*, **46**, 454–455 (1997).
18. Hill, A. B. *Statistical Methods in Clinical and Preventive Medicine*, Livingstone, Edinburgh, 1962.

