

Symposium on ethics and clinical trials

Ethics and statistical methodology in clinical trials

Christopher R Palmer *University of Cambridge, Cambridge*

Author's abstract

Statisticians in medicine can disagree on appropriate methodology applicable to the design and analysis of clinical trials. So called Bayesians and frequentists both claim ethical superiority. This paper, by defining and then linking together various dichotomies, argues there is a place for both statistical camps. The choice between them depends on the phase of clinical trial, disease prevalence and severity, but supremely on the ethics underlying the particular trial. There is always a tension present between physicians primarily obligated to their own patients (the weight of 'individual ethics') and ethical committees responsible for the scientific merit of the trial and its long-term implications ('collective ethics'). Individual ethics, it is proposed, favour the Bayesian approach; collective ethics, the frequentist. Though in some situations the choice appears clear-cut, there remain others where both methodologies can be appropriate.

Bayesian v frequentist statistics

One does not have to delve too far into statistical literature to realise there is a dichotomy of opinion between two competing methodologies, frequentism and Bayesianism (1). The former holds the traditional position and is routinely applied in the medical journals. It forms the basis of the majority of statistical training received by the medical profession, and is by far the more readily found in textbooks and statistical software packages. Its name derives from its underlying philosophy whereby the concept of probability is viewed as a hypothetical long-run average frequency. For example, if I rolled a die repeatedly, I would expect a one to occur one-sixth of the time and hence the probability of a one is deemed 1/6.

The alternative approach, named after an eighteenth century English clergyman, Thomas Bayes, views probability differently, as a degree of belief. Thus, to the Bayesian, chance events occur

according to a numerical scale with 0 and 1 representing the extremes of impossibility and certainty respectively, with intermediate values summarising the likelihood of occurrence. It makes more sense to the Bayesian to answer the question: 'What is the probability of rain tomorrow?', because the frequentist's classical definition cannot be applied without conceiving of thousands of days just like 'tomorrow'. However, aimed at Bayesians is the criticism that some, and therefore too much, subjectivity enters their view of probability, since different people may quote different probabilities for the identical event. Indeed, formally, in the Bayesian approach to statistics, one is required to propose a 'prior distribution' before the collection of any data to quantify one's beliefs. Such data subsequently modify those beliefs, however strong or weak, into a 'posterior distribution'. Frequentists sometimes argue that, with large samples, these posterior distributions and their own probability distributions tend to coincide (in form, if not interpretation) so why bother introducing 'unscientific', subjective elements into the theory?

This distinction of methodology is especially sharp when applied to medical practice and, in particular, to clinical trials. Questions of statistics pervade all stages: the design, the conduct and the analysis as well as the interpretation, publication and impact of the results. Debates abound among statisticians concerning numerous issues. For instance, do interim analyses affect statistical significance? Is it legitimate to ask about the chance some parameter is positive? Is randomisation always necessary? Will the results be persuasive enough to sway current practice? Should accruing data be exploited to adapt allocation of treatments in mid-trial? What does a 95 per cent confidence interval, or a P-value, for that matter really mean? Which methodology is the appropriate one, Bayesianism or frequentism?

Both agree that statistical inference concerns making valid generalisations from sample to population, and happily, with large samples at least, their conclusions concur. Frequentists say their sample means and standard deviations (called 'statistics' and commonly denoted \bar{x} and s) are random outcomes of an experiment that approximate fixed,

Key words

Bayesian statistics; frequentist statistics; individual ethics; collective ethics; clinical trials.

but unknown, population characteristics (called 'parameters', generally written μ and σ). Bayesians, by contrast, say sample statistics, like \bar{x} and s , are fixed and known, while the population parameters, μ and σ , are unknown and randomly distributed. In medical applications, both camps lay claim to the ethical high ground, but surely both cannot be right. Or can they?

Individual v collective ethics

It is precisely the matter of ethics that separates medical statistics from all other branches of the subject. One cares relatively little if a field of wheat perishes during or as the direct result of an agricultural experiment. However, in clinical practice the experimental units involved are priceless, so ethical concerns must dominate the clinical trial. This is especially true when the disease being combatted is life-threatening. Statisticians, like physicians and others, must be extra vigilant if their term 'failure' is a euphemism for death!

Most will be familiar with the concepts of **individual** and **collective** ethics, as described in Pocock, (2). The former is concerned with the well-being of each patient, while the latter emphasises the common good for society. He argues, rightly, that each clinical trial requires a balance between the two. Viewed globally, the *raison d'être* of clinical trials is indeed for the common good. However, it is obviously unjustifiable to argue that collective ethics (3) are the only consideration, to the neglect of the study volunteers employed. Equally, it is mistaken to hold that individual ethics are so supreme that clinical trials themselves are unjustified, the view adopted by Burkhardt and Kienle (4). Also, among statisticians who focus on the ethical aspects of randomised clinical trials, Royall (5), in stressing the 'personal care principle', overemphasises individual ethics, as borne out by the discussants of his paper. Instead, it is preferable to require individual and collective ethics to be omnipresent, in delicate balance. In one respect, individual and collective ethics are, respectively, analogous to St James's exhortations to faith and good deeds (6). The distinction is more a labelling of a continuum than a genuine dichotomy, as one does not preclude the other and chronologically, one must precede the other. I shall go on to argue that there are circumstances where one ethic can be given slightly more weight than the other in this fragile equilibrium and that, in turn, there should follow a corresponding shift between the competing statistical methodologies.

Physicians and ethics committees

In any clinical trial there are several people groups interested to varying degrees in its progress. Two of these are the doctors whose patients are being

studied and the trial's overriding ethics committee (7). Before proceeding, let me stress the following. It is abundantly clear that there exist research-minded physicians who are extremely interested in the conduct and results of the trial overall, and equally there are ethics committees supremely interested in the well-being of each patient entered into the trial. One would certainly hope both of these to be the case! However, it is also fair to say that physicians ought to have the needs of their own patients uppermost in their minds (8), while ethics committees ought to be primarily concerned with ensuring the trial as a whole is scientifically worthwhile. (For otherwise, who is responsible for doing so?) I do not wish to digress about the need for data-monitoring committees making decisions to continue or terminate trials, though that is surely one of their functions and an important area of statistical research. Instead, it suffices to note that, metaphorically speaking, a clinical trial generates two hats: one of individual beneficence that better fits physicians and another of community beneficence that better fits ethics committees.

Pragmatic v explanatory approaches

Another contrast exists according to a particular trial's purpose. Schwartz, Flamant and Lellouch (9) make a helpful distinction between the **pragmatic** and **explanatory** approaches. The first is motivated by the desire 'to reach a practical decision on the most appropriate treatment' and hence is more concerned with *identification* of treatment superiority. The second approach is motivated by the need 'to increase scientific knowledge' and so is more concerned with *estimation* of efficacy. Understandably, in a clinical trial, fewer subjects are required to identify the better (or best) treatment than to estimate their respective success rates. The statement 'A is better than B' is clearly an easier matter to establish than, say, 'A is 39 per cent successful, B is 33 per cent successful'.

Current v future patients

It is the patients who form, arguably, the most important of the above-mentioned several people groups with vested interests in any clinical trial. There are two schools of thought in operation when it comes to medical practice. One can be described as putting the needs of current patients foremost, while the other seeks to serve best the needs of future patients. These two viewpoints correspond precisely to those that should be adopted by physicians and ethics committees, at least when wearing their respective hats, as described above. Thus, the first asks: 'How should I best treat my next patient with this ailment?' while the other asks: 'What can be demonstrated to be the best therapy for all patients with this condition?' These questions are not always

compatible, and hence the conflict arises in the guise of an ethical dilemma.

I suggest that the preceding divisions are wed to one another according to these schools of thought. Thus, individual ethics tend to go along with the pragmatic approach ('What do I, as her doctor, think is best for Mrs Jones?') whereas collective ethics go with the explanatory approach ('What can be proven the best treatment for general use?'). Arguably, one could conceive of alternative pairings, but these are the more natural ones, linking together philosophies that are *subjective* (more applicable to one doctor and one patient) and *objective* (concerning all doctors and all patients).

Early v late phase trials

In order to see the practical consequences of making the above distinctions, yet more need to be introduced and considered. The first concerns the stage of research the potential new therapy is undergoing. A new treatment must pass through a number of stages, in succession, any of which could disqualify it from further clinical investigation. Formally, in practice, trials have evolved into four distinct phases. Briefly, phase I is the first time the treatment is offered to humans, who may or may not have the relevant illness, with a view to determining basic safety and dosage levels. Phase II looks for signs of efficacy among ill patients. Phase III represents full-scale testing, generally head-to-head against the current standard therapy under carefully observed conditions, while phase IV is only entered post-marketing, mostly to target rare or long-term side-effects. Encouraging results from the two earlier phase clinical trials ought to be considered as exploratory, or at best, suggestive only. Confidence in the new treatment is necessarily low. This is because with smaller sample sizes, extracted from fairly specific patient groups, its ability to generalise to the target population is still in question, more so than if it had proved successful on a larger and broader sample. At this time, one may need to exploit external sources of information, such as prior beliefs, in one's assessment of the experimental treatment. Clearly, when progress towards a new therapy is in its infancy, the school representing individual ethics and the pragmatic approach has the greater claim to attention.

By the time a treatment has reached phase III, however, the emphasis can begin to shift to collective ethics and the explanatory approach. Larger samples are called for, in keeping with growing confidence in the therapy, and if successful, results can be viewed as confirmatory or even definitive. So, in summary, the newly discovered treatment must first have emerged as the best of a number of experimental ones, but at that time, probably somewhat tentatively. Only later, after negotiating full-scale testing in phase III, can widespread acceptance be

possible. Put another way, chronologically, faith (in the treatment) must come before good deeds, as evidenced by documented improvement over existing, standard therapy.

Application

The above considerations fit well with the Bayesian and frequentist approaches to statistics. Which methodology ought to prevail depends fundamentally on which type of ethic is considered the more relevant. This, of course, depends on the particular clinical trial. Two further factors beyond the phase of the trial are of utmost importance: the prevalence/incidence of the disease in question and its severity. As already suggested, the later the phase, the greater can be the emphasis on collective rather than individual ethics. Similarly, the more common the disease and the further it is from being life-threatening, the more weight can be put upon collective ethics. Where this is the case, as in most phase III (and later) trials whose purpose is to advance knowledge for the treatment of a fairly common and non-lethal disease, frequentist methodology seems preferable. But, in most early phase trials or in those phase III trials concerning a comparatively rare, lethal disease, Bayesian methodology, complete with its subjective shortcomings, seems the more appropriate.

Table 1 summarises the state of affairs, including a necessarily fuzzy boundary where neither individual nor collective ethics should obviously predominate. An imaginary, diagonal line discriminates between predicated choice of statistical methodologies: individual ethics favours the Bayesian approach to design and analysis, whereas collective ethics favours the frequentist approach.

Discussion

It is perhaps fitting that Bayesian methodology is associated with individual ethics and physicians' viewpoints, for it is generally agreed that most non-statisticians tend to think like Bayesians when it comes to interpreting probability. For example, the natural interpretation given to a statement such as 'a 95 per cent confidence interval for some difference is from -3.1 to 7.5 ', is that there is a 95 per cent chance the true difference lies between those values. The strict frequentist definition, however, disallows such an interpretation and actually requires some mental gymnastics to conceive of a sequence of hypothetical trials and sets of data beyond the one observed. (It says that in repeated sampling, we can expect 95 out of every 100 such confidence intervals to contain the true value of the parameter.)

Besides labelling the asterisks, one matter Table 1 raises is the precise definition of disease type. It begs the questions of just what constitutes 'rare' or 'severe', say, as well as who is to judge such things,

Table 1
Suggested shift between individual ('Ind') and collective ('Coll') ethics according to phase of trial, prevalence and severity of disease

| Phase | I | | II | | III | | IV | |
|------------------|------|--------|------|--------|------|--------|------|--------|
| | Rare | Common | Rare | Common | Rare | Common | Rare | Common |
| Mild | Ind | * | * | Coll | Coll | Coll | Coll | Coll |
| Moderate | Ind | Ind | * | * | * | Coll | Coll | Coll |
| Severe | Ind | Ind | Ind | * | * | * | Coll | Coll |
| Life-threatening | Ind | Ind | Ind | Ind | Ind | * | * | Coll |

Asterisks denote where claims to supremacy seem equivocal. The further away a cell in the table is from the asterisks, the stronger the case for the type of ethic recommended.

although these are not, I suggest, statistical concerns alone. Clearly, too, severity and prevalence are not classified as simple, categorical variables. However, Table 1 does serve as a starting point for further bio-statistical/ethical debate and at least illustrates the trend of where the choice between methodologies seems more clear-cut and those situations where there remains room for ambivalence.

Acknowledgements

This work was conducted partially while the author belonged to the University of Reading's Department of Applied Statistics. It was stimulated by the 1991 joint Royal Statistical Society and London School of Economics conference on *Methodological and Ethical Issues in Clinical Trials* (10) and has benefited from the comments of referees.

Christopher R Palmer, MA, MS, PhD, is a medical statistician in Cambridge University's Department of Community Medicine, based at the Institute of Public Health, Robinson Way, Cambridge CB2 2SR.

References and notes

(1) In the medical context, some recent remarks on this dichotomy can be found in: Herson J. Bayesian analysis of cancer clinical trials: an introduction to four papers. *Statistics in medicine* 1992; 11: 1-4.

(2) Pocock S J. *Clinical trials*. Wiley: New York, 1983: 104-105.

(3) A note on terminology. Here, 'collective ethics' refers to the needs of the population of *patients*, in contrast to 'collective equipoise'. This refers to the population of *physicians* and describes a necessary condition even to begin a randomised controlled trial. Further details are in: Johnson N, Lilford R J, Brazier W. At what level of collective equipoise does a clinical trial become ethical? *Journal of medical ethics* 1991; 17: 30-34.

(4) Burkhardt R, Kienle G. Ethical problems of controlled clinical trials, with critical response by Vere D W. *Journal of medical ethics* 1983; 9: 80-89.

(5) Royall R M. Ethics and statistics in randomised clinical trials, with discussion. *Statistical science* 1991; 6: 52-88.

(6) See James 2: 14-26, *Holy Bible*, for some theology.

(7) Throughout this paper, 'ethics committee' refers to whichever body is responsible for granting (ethical) permission for the clinical trial. For early phase trials such a committee may be an institutional review board, while for later phases it may be called a data-monitoring committee.

(8) Succinctly put, 'Doctors have a duty to avoid harm to their patients and to serve their best interest', quoting from the Institute of Medical Ethics Working Party on the Ethical Implications of AIDS. Boyd K. AIDS, ethics and clinical trials. *British medical journal* 1992; 305: 699-701.

(9) Schwartz D, Flamant R, Lellouch J. *Clinical trials*. London: Academic Press, 1980 (Translator: Healy M J R).

(10) Proceedings of this conference are due to be published in *Statistics in medicine* 1993; 12.

