Debate

Problems in controlled trials – a critical response

D W Vere *Department of Pharmacology and Therapeutics, London Hospital Medical College*

The paper by Drs Burkhardt and Kienle is helpful in that it takes to be the aim of all good investigators of medical treatment that benefit to individual patients matters most. Would that everyone shared their enthusiasm in that direction. Where it seems less than helpful is in their polychromatic condemnation of clinical trials on the grounds that some have been badly done, or proved difficult (which would be reasons to condemn almost every kind of valuable human effort), and their even more serious-seeming failure to grasp the central logical problem of all applied biostatistics, including trials. So, rather than attempting to respond to their paper piecemeal, I will first try to set out the logical problem. To make this utterly clear, it will be expressed as a set of individual statements, each of which can be taken alone, analysed and attacked so as to test each logical link in a chain which I believe leads to a very different conclusion from that of Drs Burkhardt and Kienle.

- 1) The conquest of disease depends upon acquiring new knowledge.
- 2) Gaining new knowledge is very costly, not only in money but in terms of the suffering and sacrifice needed to acquire it.
- 3) In the biological world, variation between individuals is wide and obvious.
- 4) Man is within the biological world.
- 5) There is no scientific way to make sure predictions of the outcomes of interventions upon an individual. It is only possible to see the results after the intervention; prior predictions are always highly imprecise.
- 6) When multiple measurements are made within, or between, individuals, or both within and between them, it becomes possible to study their distribution. They are then statistics.
- 7) Statistics can be used to achieve two things:
 - (i) to have predictive value for a group of subjects;
 - (ii) to refine decisions about groups (through significance tests, confidence limits, likelihoods, etc) which investigators would have wished to make in any case,

but which would lack the precision conferred by statistical analysis. In other terms, the statistical test merely refines the *same* decision that an investigator would have wished to make, even without its assistance.

8) When properly made, the statistical method affords the best, most precise, and most conservative way for the investigator to reach decisions. Until that decision is reached, or rather becomes attainable, any impression that the investigator, or anyone else, may have about the best decision is speculative. It is a 'hunch'. Even if it proves later to have been correct, that confers little more respectability upon it than would adhere to a gambler's throw which landed him upon the correct answer.

It seems to me that each and every criticism levelled at clinical trial method by Burkhardt and Kienle misses the logical conclusion of these steps, namely that if one has a method of optimising decisions then every decision which anticipates that optimal point, for whatever reason, must lead to an inferior conclusion. In simple terms, if I decide what I need to guarantee a certain desirable level of knowledge and then stop short of that point, I will never know what I had decided that I needed to know to help *any* patient in that group. And, since there are no prior ways to refine decisions for the individual, but only for the group, there is nothing which can ever substitute for group methods.

Now none of this, of course, denies that many people have made bad clinical trials, that many trials go on overlong and so hurt individuals, that there is no way to anticipate definitely the end points of a trial, so that one may be surprised by what is disclosed by it, and so on and on. But no such argument can ever undo the logic of 'the good trial', any more than it would make sense to argue that, since so many marriages fail or dissolve or hurt people that marriage should be abolished, or that there is something intrinsically wrong with the concept of marriage.

Having nailed those colours to the mast let us now look at Burkhardt and Kienle's paper in detail, showing how the argument applies at each main point. Their first serious objection is that there are few trials which fail to show more or less serious shortcomings. It would be easy to dismiss such an argument as being

Key words

Controlled trials; medical ethics.

true of life in general, and to miss its deeper significance. It is *possible* to see the shortcomings of trials because they are scientific exercises, because their mechanisms are published, are open to scrutiny, are open to repetition or to falsification. It is *not* possible so easily to discern the faults of private opinions about therapies. Are we then to conclude that private opinion is less faulty than the findings of clinical trials? What nonsense that would be; trials were introduced because personal opinion was so notoriously fragile, biased and unreliable. Hence Peto's remark about the difficulties of making trials, though a fair one, is not the principal reason for the shortcomings which are apparent in many trials. The deeper point is that, unlike other methods, the shortcomings are visible because they *can* be seen.

Another aspect of this argument is that clinical trials are seen, as it were, to be 'on trial', to stand in isolation before some legal inquiry. But in medicine we can and must do something which could never be done in a law court; we are faced with a choice between alternative methods of discovering new therapies. The legal process pillories one subject; we are deciding which is the least culpable. This aspect never emerges in Burkhardt and Kienle's paper; what valid alternatives do they offer to replace their incriminated clinical trials? The fact is that every individual therapy is as much an experiment, a venture into the unknown, a constraint upon nature, as is a clinical trial. The only difference is that whereas both can and should be entered upon in the interests of individuals, a trial has the added benefit of being in the interests of the group, given always that it is properly designed so as not to sacrifice individuals for group ends.

The next main problem cited is that treatment strategies rather than drugs are compared in trials. It is hard to see why this should seem to be a stumbling block, for two impressive reasons. First, it is therapies that matter to patients, not drugs. Studies of pure drug action matter to scientists, for they enable them to understand existing drugs and to improve them. But patients' needs are concerned with treatments, so this is really an advantage of clinical trials; they are at least relevant to patients' needs, or can be so designed. The second point here is that although treatment strategies rather than drugs are under test, the characteristic signs of specific drug action usually show through unambiguously, given, that is, that the drug element of the therapy has effects. For example, it is very difficult to conduct a 'blind' trial of beta adrenoceptor blocking agents because the group treated with them have slower pulses than those given most other remedies.

Burkhardt and Kienle next turn for ammunition to the undoubted heterogeneity among patients. This is undoubtedly a problem, which has been recognised rather later than it deserved. However, numerous techniques of stratification, minimisation and subgroup analysis exist to offset the problem, provided sufficient patients can be recruited to fill out the most important subgroups so that tests of generalisable

significance can be made. Though real, this is certainly not a problem so general, so insoluble, as to warrant the abandonment of trial methods as such. It is a problem which is intrinsic to the nature of natural enquiry; one could advocate the abandonment of all scientific method in biology on the same grounds.

The next objection voiced to clinical trials is that random selection of patients cannot be achieved. Of course it cannot; patients are a highly selected group. But what counts in testing remedies is not so much the random selection of patients as the random allocation of treatments, which is something which *can* be achieved. This point is so widely misunderstood that it is worth closer attention. Trials cannot eliminate accidental bias; they can eliminate systematic bias. Hence random allocation is always preferable to biased allocation. But no one should pretend that this includes random case selection, even though that matters far less than random therapy allocation.

That statistical testing often involves inductive knowledge is undeniable. That Popper has shown reasons to mistrust induction is also undeniable. That Popper is a great philosopher is also true. But the argument is not as simple as these isolated facts suggest, for the following reasons. First, the inductive step in statistical reasoning is the step which envisages a model, or a test distribution, which is believed to represent the form, or the distribution, of the data. Many would regard that step as very weak; I do myself (1). But there are statistical tests which are 'distribution-free', which do not commit the investigator to any imaginative step beyond the image in his mind of how a purely chance mechanism would behave were it generating the results he has observed. What is more, it seems to me that Popper only seems to eliminate the inductive step in logic by burying it within that inventive or imaginative leap without which he seems to agree that science cannot progress. But the conclusions drawn from that leap are, as he rightly asserts, always and necessarily open to test and further experiment. So, I cannot see the quarrel; it is induction *on its own* which has no substance, but seen as an invention whose fruits are testable by steps of observation and deduction, it is usable whatever we may call it, provided we always treat it critically (2).

And, what now of the alternatives? If induction in clinical trials is wicked, what about induction based on the experience of isolated individual patients? How can one ever reach any sort of valid generalisation from such experiences, unless it be of the crudest and most obvious kind?

The paper of Burkhardt and Kienle next moves to its central critique, that trials can never lead to the choice of the best treatment for the individual. Though true in an absolute and isolated sense, in practical contexts this contention (to my mind) degenerates into nonsense for reasons which I have tried to explain at the beginning. This is so for two important reasons. First, it seems to me to be a matter of ineluctable logic that, if a group of colleagues agree that they would become convinced of a

treatment's superiority for the group after, and only after, a prefixed level of statistical significance had been exceeded, then until that moment is reached the best treatment for every individual is *either* of the therapies under test, chosen at random, for no one then knows which is the better. Were that uncertainty not present, it would surely be unethical to make a trial. Once the uncertainty has been removed it is unethical to make one. And, notice that this has little to do with what those investigators may *believe* up to that point, what hunches they may have, or what their friends may say. Until the facts have been demonstrated, whatever they may believe, they do not *know*. Their level of foundation for their beliefs must be raised by evidence to the point where no reasonable person is likely to reject it. Until that point, the best therapy for the individual is the random choice. The second reason for rejecting this criticism can be seen by taking the inverse of one of Burkhardt and Kienle's points. As they suggest, trials often disclose surprising end points, which are unable to be anticipated, or at least may not be anticipated. But what a good thing this is, for there is no better way to help individual patients than to expose the differences between their responses to therapy by serendipity, if not by design! The point is that whenever nature is exposed to view in orderly ways, new facts of value are likely to emerge. It is argued that trial data hold 'no privileged position' amongst other medical data; true, if those data are collected in a scientific way. But most clinical data are not; they are observational and unique. Though 'true', they are of limited value because they do not afford a basis for comparison, generalisation or prediction beyond the simplest level; that is why scientific medicine evolved.

The authors then turn to ethics, with the same logical approach which characterised their handling of the scientific aspects. Leaving aside the blemish at the end of their first paragraph (modern manufacturers do not, in general, begin clinical trials with short supplies of trial remedies), their chief contention is the unethical quality of withholding remedies from a control group of subjects. That drugs have, at times, been wrongly withheld from control subjects is beyond doubt. Whether this is done in the name of a drug firm, a regulatory authority, or a group of investigators it is still wrong. But it *need* not happen, and is in no sense a necessary entailment of clinical trials. In properly conducted trials the comparator for a new therapy must always be the best available alternative. Placebos should only be used if there is reason to think there is no more effective remedy (and they, of course, are not devoid of effect), or if the patient's complaints are minor, not dangerous to life or limb, and the patient knows that placebo usage will be part of the trial to which he or she freely assents. In other words, there surely is a proper place for trials where placebos are given, or active therapy is withheld or washout periods are included. As with all else in life, from garden forks to guns, there is also an improper or unethical use of

such means. To stigmatise all trials on the grounds that some have been unethical is also improper. Indeed, since the clinical trial is often the only reliable means to gain new knowledge it is unethical to withhold it. The ethics of withholding trials is a central problem to which further reference will soon be made.

But the next argument advanced is that withdrawals of individuals may provide a means to evade the imagined ethical faults of trials; even though such withdrawals, if they amount to a substantial fraction of trial subjects, effectively destroy the trial. This, to my mind, is that same central logical error in a new guise. If there is objective evidence that an individual may be suffering by inclusion in a trial, then let him indeed be withdrawn. But if the reason for withdrawing him is the doctor's *subjective* belief that one remedy is proving superior to the other, before that superiority has attained the pre-agreed statistical limits, then he is simply confused. He has prejudged the very question that he sought to ask, having agreed beforehand how he would accept an answer. He is shadow boxing with the truth, and that is indeed unethical. How can it ever be right to accept standards of proof, set out on the costly road to attain them, and then draw back whilst still upon that road, having made it *seem* that justice has been done? If those prior judgments are shown by the proceeding trial to have been faulty, then indeed it would be right to discontinue the whole trial. Even that in itself would be invaluable; human error, which is inescapable in a position of partial knowledge, would have occurred. But it would have occurred in an observational situation which makes it possible to profit from it; new knowledge will have been gained, and set to good account. I have had experience of just this situation; in a set of trials of therapies for the nephrotic syndrome, made many years ago, the answers were totally surprising. The reason, when pursued, turned out to be faults in the hospital drug administration system. These were studied and threw a large amount of quite unanticipated light on treatments in general (3). There simply is no substitute for collecting new knowledge in orderly ways.

Burkhardt and Kienle next embark upon a long critique of the ethics of biostatistical method. Again, they seem not to have grasped the essential nature of statistical technique. It is designed, and is only properly used, to refine a decision which the investigators would have in any case wished to make, even without its aid. It must and can be the servant, not the master, of clinical practice. It need and should not be imposed in ways which satisfy scientific desiderata but damage patients' interests (4). That it has been misapplied is again not denied. That it must by its nature be misapplied I reject utterly. (The problem which troubles them so much about randomisation of life-saving treatments could have been dealt with it seems, quite simply, by stratification.)

They then repeat their former arguments for the individual versus 'the group'. At risk of seeming repetitious, it can only be restated that until what is best for

the group is known, no one can make any *generalisable* or *predictive* statement about what is best for the individual, about which remedy will have the best chance of helping any individual. And, the corollary which follows is that it is impossible, despite these authors' suggestion that it should be done, to apply 'individual ethics' to a trial prior to the point of pre-agreed decision, for all participants have agreed that until that point is reached they will not accept lesser evidence as convincing. Hence, until that moment, no one can know which therapy is most likely to help any patient, therefore it is impossible to make any rational decisions on behalf of individuals. Needless to say, a valid exception to this rule is any case where the grounds of the prior decision are shown by the trial to have been false. Even then, the result is helpful, and the trial should be stopped, not the individual withdrawn, unless, of course, an error of judgment has allowed the patient to enter the trial.

The legal problems which Burkhardt and Kienle discuss are, fortunately, largely peculiar to their own country. However, it happens that they raise the gravest ethical difficulties. If one nation decides that, in order to protect the assumed interests of its own individual subjects it will establish legal rules which make clinical trials difficult or impossible to conduct on any reasonably scientific basis, how will that nation obtain new remedies? Either it will deprive its people of novel and effective remedies, which scarcely seems to be an outcome they might have wished, or it will rely upon data obtained through what to them must seem to be the wrongful sacrifice of the interests of patients in other countries, which is also unethical. Now, I do not wish in any way to condemn a group of people for holding such views; they are indeed the views of many British people, and of people of other nationalities. But it does seem to me that the attitude just described is unethical, and it is a great pity when it becomes enacted in law. What is worse, I do not believe that German law as such *does* condemn or outlaw clinical trials, indeed it enjoins them as a normal method to validate new therapies (5). It is rather the narrow view of what constitutes a patient's interests that is the problem. Indeed, there is plentiful evidence that patients in some countries are suffering great damage as a result of such views, which lead them by uninformed choices to refuse the very remedies which could have helped them (6). That is a sad, and I think an unethical position.

The central problem here, to my mind, is that large numbers of people are unprepared to accept the cost and suffering required to gain new knowledge, especially when faced by the cruel turn of events, which, like a war, brings them to the point of risk simply because they happened to fall ill just as a new remedy became available for trial (7). So many prefer the remedy to be tested just on others. But is that an ethical preference?

Lastly, the two authors discuss the patient's 'right of self determination' which is at the centre of the dilemma in German legislation. But, pause for one

moment to consider upon what basis a patient may have the 'right' to 'determine' himself – on a basis of ignorance? How can he determine himself when even the best, the most informed, of his doctors do not know which remedy is best for him? Is it, if there is to be no clinical trial, to be on the basis of bias, of misinformation? I would wish him luck, for nothing else can save him within the limits of the material world.

In their conclusion, Burkhardt and Kienle quote Meier (8) to say that we deal in this context with a political problem, a problem concerned with what people want. I believe that he was wrong, though it is not hard to see and to sympathise with his view. It is at root a *logical* problem. (Unless, of course, one defines politics as the wants of people for what they cannot have). Do people wish to remain diseased, or do they wish to be relieved; that is the political problem. If they wish their sufferings eased then they face the logical problem, that the only way ahead, whether by clinical trials or by other kinds of experiments, is by paying the price – suffering – of acquiring new knowledge by orderly methods. Clinical assessments are not the answer, they are simply not firm enough to give a reliable basis for decisions.

Their paper ends, with seeming reasonableness, by pleading for some way to balance the interests of the individual and the group. But the very posing of this argument embodies within it that same basic logical fallacy. How on earth can an individual's interests (regarding therapeutic outcome) be balanced against the group's when no one can know what they are? Even when a trial has been made, no one knows from its findings what each individual's interests will be; but everyone has more information than ever before about the likely outcomes of treating individuals in certain ways, and can use it to plan in better ways for their care.

If medicine is to advance, sacrifices by some individuals are inevitable. But it must always be remembered that if progress is not so attained, the sacrifices required by diseased nature of us all will be far greater.

Acknowledgment

I am deeply indebted to Dr Gerfried Fischer of Georg-August Universität, Göttingen, for the references to German law, and to the members of the staff of this department for much helpful comment and criticism.

References

- (1) Vere D W. Logic and Inference in clinical trial design. *British journal of clinical pharmacology* 1982; 14: 171–176.
- (2) Popper K R. *The logic of scientific discovery*. London: Hutchinson, 1959. The 1980 impression, pages 254, 27–29 and appendix ix, pages 387–419, where the problem of induction is shown to centre around the idea that there can be a 'principle of induction'. And where it is also shown that the inductive step in statistical reasoning is acceptable, within sharp constraints, and

that 'invention' or 'discovery' may involve inductive elements, albeit subject to confirmation by testing.

- (3) Vere D W. Errors of complex prescribing. *Lancet*, 1965; 1: 370–373.
- (4) Vere D W. Scientific philosophy applies to clinical medicine. Heidelberg Symposium, 1981: the second symposium on clinical trials in early breast cancer. Ruprecht-Karls-Universität, Heidelberg Symposium Report (in press).
- (5) German law reports 1976 (i) Bundesgesetzblatt teil 1. Gesetz zur Neuordnung des Arzneimittelrechts: paragraphs 21, 22, 25, 40. (ii) Controlled trials, *ibid* para 26 (Arzneimittelprüfrichtlinien) 1971, Jun 6 published in Bundesanzeiger part IIA al 3.
- (6) Scheurlen H, Schumacher M. Planning breast cancer trials in the Federal Republic of Germany. See reference (4).
- (7) Vere D W. Controlled clinical trials, the current ethical debate. *Proceedings of the Royal Society of Medicine* 1981; 74: 85–87.
- (8) Meier P. Terminating a trial – the ethical problem. *Clinical pharmacology and therapeutics* 1979; 25: 633–640.

Response

R Burkhardt, G Kienle

Confirmative significance tests for controlled trials have never led to the discovery of drugs. This is a possible area of agreement with Dr Vere, as he speaks of controlled clinical trials *refining* decisions. It is, however, a fundamental error to make no distinction, as far as the refining of decisions is concerned, between statistical model and medical reality. Because of the practical problems of controlled trials, and also for considerations relating to the philosophy of science in general, such distinction is essential. Identity of model and reality does not exist in medicine, nor in any other empirical science. With respect to decisions on the treatment of further patients statistical decision theory, for instance, calls for random *selection* of experimental units, not random allocation. Significances based on χ^2 or t tests are therefore at most formally correct; they do not, as such, relate to medical reality. As a result it is not acceptable to consider the model result as the final arbiter in decision-making, and clinical judgment on the other hand as speculative. As it would be positivistic nonsense to do away with individual judgment, the putative knowledge would, in Dr Vere's terms, have to be judged by the doctor's subjective belief.

When Dr Vere labels clinical judgment as speculative, this, of course, effectively removes any ethical objections. It would then be impossible to identify statistical experiments in which an effective treatment is withheld from an individual patient. Then there is, of course, also no need for serious consideration of our arguments – to the detriment of trial patients.

Apart from this, practically every argument advanced by Dr Vere has already come up in the discussions in Germany. Dr Vere's paper reveals the kind of positivistic short-sightedness liable to beset anyone who is not prepared to engage in discussions on fundamental questions of human cognition.

Reply to Response

D W Vere

The comments made by Burkhardt and Kienle in response to my response are very helpful, for they reveal with exquisite clarity the nature of the gulf between us. To respond briefly, and *seriatim*, I would first deny their remark that 'confirmative significance tests for controlled trials have never led to the discovery of drugs'. Of course trials do not discover *drugs*, but they may discover *therapies*. Their statement is untrue for several therapies, hence the one ground of agreement between us which they felt they might have discovered is removed. Effective new remedies against dental extraction pain, and multiple chemotherapies afford some good examples.

The real issue, however, is the relationship between model and reality; they deny it for clinical trials, I accept it. Admittedly, no model is a full representation of that which is modelled, but correspondence between model and reality *within the defined realm of discussion* in a trial is a reasonable postulate. Otherwise no hat ever represented a head, no glove a hand, no rectangle a garden, no map a country, and no science is possible whatsoever. Nor can mathematics help us any more, for numbers model the real world. The fact that I cannot see the real world but only perceive light reflected from it makes it an acceptable aid to reading, but presumably not to those who reject models.

They are not correct to assert that 'statistical decision theory . . . calls for random *selection*'; the point was argued admirably by Geoffrey Rose (1). It is *randomisation* that is needed; random sampling is unattainable in medicine in most situations, but this does not invalidate inference.

My assertion that clinical judgment is speculative rests upon the frequent overturning of clinical dogmas and beliefs by well conducted scientific experiments, nothing more.

That my mind is closed I freely admit; closed by logical inference. If that inference is faulty, or other grounds can be shown to bear more weight, I shall be most happy to change my mind, for what we all long to do is to help our patients.

References

- (1) Rose G A. Bias. *British journal of clinical pharmacology* 1982; 13: 157–162.