MEDICAL PRACTICE

For Debate . . .

The controlled clinical trial and the advance of reliable knowledge: an outsider looks in

H A F DUDLEY

During my professional lifetime the prospective randomised clinical trial has progressed from an interesting tool devised by workers of high and subtle intellect for use in extremely well defined and limited circumstances largely outside the province of the ordinary clinician to a central concept. Controlled trials—some good in design, some bad, some indifferent—appear in every issue of reputable journals; the ethics of participation and of non-participation are endlessly debated—sometimes constructively and sometimes not; and the internal ethics, particularly the matter of seeking “informed consent,” are a current cause for concern. For those of a Kuhnian mind¹ we have seen the throwing of a massive Gestalt switch, so that “normal clinical science” has a new framework within which it can work until some other yet to be foreseen perturbation gains sufficient amplitude to throw the switch again.

The underlying belief behind the controlled trial is that it provides greatest reliability in the resolution of the problem whether or not two (or more) methods of procedure yield comparable or different results. Reliability is usually linked with the word *science*. Take, for example, two passages from a recent and important paper on informed consent.² “Randomised clinical trials are the most efficient *scientific* means of comparing alternative treatments”; and, a little further on, the random controlled trial is “. . . the preferred *scientific* method for evaluating new treatment” (my italics). In this context science means designing a test, undertaking it, and using deductive inferences from the results to arrive at certain conclusions—what in the past I have called science’s internal programme.³ The more rigorous that programme the more reliable are the results thought to be, though it must be remembered that to generalise from them to the world at large requires a leap which

some would call induction and others merely a value judgment on the applicability of conclusions from a sample. Such applicability is greater when the experiment confirms or confutes some theoretically well justified theory—for example, the experimental confirmation of Einstein’s prediction of the effect of gravity on light.

It is worth emphasising that it is *science* itself that keeps the processes of the internal programme going. Those outside science do not see the same difficulty in accepting experience and intuition as a basis for knowledge. George Steiner put it well when, discussing whether truth has a future, he said that the “scientific cultural” background of Western man has given us an “obsession with objective and abstract truth.”⁴ My purpose is gently to inquire whether that obsession has not blinded us to the idea that in some circumstances less reliability is acceptable or alternatively the pursuit of truth is a race after an illusion.

As I see it, the central argument now runs: because a controlled clinical trial is the most effective method of making a statement which contains objective scientific truth then anything less than this offends against the scientific canon and cannot be used to make recommendations about how to manage a problem. The argument is usually supported by reference to the admittedly unhappy catalogue of discredited treatments of the past,⁵ though the alternative—that prospective randomised trials may damage individuals or fail to resolve the issue they address—is not usually considered. It is not my purpose, however, to debate the issues of “*primum non nocere*” for the individual versus “the greatest happiness of the greatest number”⁶ or to discuss the matter of the sacrifice of freedom necessary to a clinical trial. These matters have been done most ably by others.⁷ ⁸ Rather, I should like to put forward the hypothesis that there is a continuous rather than a discontinuous scale of reliability not a quantum leap from none to near total reliability and that the place of the controlled trial in clinical medicine must be assessed against that scale. In doing so I have to admit bias. I have not been an extensive participant in

St Mary’s Hospital, London W2 1NY

H A F DUDLEY, CHM, FRCS, director, academic surgical unit

controlled trials. My more enthusiastic clinical colleagues will doubtless ascribe this to scientific and administrative laziness, and certainly the apparatus of clinical trials does tend to oppress me. Nevertheless, I would like to think that it is my innate "outsider" nature as described by Colin Wilson⁹ that is the background to my viewpoint: a nature that is as necessary a component of the scientific world as that which espouses the central dogma of current normal science. There is no "high ground" in science. Each individual sees the matter, as William James said, from his own particular vantage point and truth and reliability are relative to that position. How we resolve differences is a matter for continuous debate, and I look on this paper as a contribution to such a discussion not as attack or a resolutely held antiposition. Doubtless I shall be disappointed when it is construed as both.

Reliable knowledge

The figure gives a visual interpretation of my scale. At one end of the scale is ignorance at the other certainty. Let me move from left to right: a problem is being encountered for the first time—we are engaging something new and, at least in theory, without prior experience. Necessarily in clinical medicine we

have to act, and the outcome, good or bad, of such action becomes part of "experience." It can be argued with some force that at first encounter and given two possible ways of proceeding a random choice is reasonable. Once experience begins to accumulate, however, this can be used to illuminate with variable intensity how we will respond on the next similar occasion. In simple terms, anything we do influences how we next behave—in Heraclitean language one cannot step twice into the same river of experience and as it flows on it may or may not become richer. Experience moves us a little to the right on the reliability scale—if it did not we would be very unsuccessful in our adaptive responses to the world. In clinical medicine I would guess that 80% of what we do and do well is based on such accumulated experience refined in the ways that I shall now consider.

The next stage of the process is the conscious organisation of experience gained by repeated encounters with events which though not necessarily identical—and never so in biology—have perceived similarities. There are two components of this. Firstly, there is consolidation, generalisation, and sometimes rationalisation (in the best sense of the word) of knowledge so as to provide a corpus of understanding, perhaps at the level of concept formation, which can be fed back into our next experience. Secondly, this process may be influenced by knowledge coming in from other disciplines, especially the basic biological sciences. A particular segment of knowledge about the action of a drug on a cell or the effect of a surgical operation on the function of a system may make it *rational* to think that we can then proceed to use the drug or the operation. If that knowledge is deemed strong (because of conviction—dangerous—or on account of collateral knowledge) it has in practice appeared to move the clinical application to the right on the scale, so permitting a decision that the drug or operation will be used. Though the judgment of the quality of the knowledge is subjective, clinicians use this form of reasoning to introduce new methods of treatment. Few, for example, would regard it as necessary to carry out a controlled clinical trial to compare appendicectomy and non-operative treatment for acute appendicitis, and, though this may be regarded as a trivial example, others from daily experience are easy to cite.

Finally, in the process of increasing reliability, active intervention in the accumulation of experience takes place. An extra technique, formal verification, is introduced. Some sort of mental or physical experiment, usually the latter, is designed according to circumstances. Here again we are dealing with science's internal programme—the conscious effort to find a model which will provide for our peers and less commonly for our patients the greatest credibility. Such verification—for example, from a controlled clinical trial—puts reliability nearer to the right hand end of the scale but not fully to the right. There are two reasons for this—the instantaneous reliability of our knowledge is only as good as its underpinning in hypothesis and, in good Popperian terms, a hypothesis by definition is capable of being disproved.¹⁰ Secondly, at the level of experience, and perhaps saying the same thing differently, application of knowledge gained from a finite number of instances to a further instance in confirmatory terms is inductive inference, which I believe impairs reliability in the sense of that word as I am using it. I make this point because there is a tendency in some fields of medicine, where these matters have not been addressed with the care and attention that they deserve, to regard verified knowledge as wholly reliable in decision making, particularly when it is surrounded by a large apparatus of statistics manipulated by an authoritarian scientific figure. In particular, knowledge of how to proceed derived from trials that entail a considerable level of skill (or lack of it) from those who participate (and who can say that there are any trials in clinical medicine that do not?) may be inapplicable to other circumstances where the skill is absent.¹¹

I should like to digress for a moment to borrow an elegant concept from Paton.¹² My formal scale of reliability can be set alongside one that embraces originality and imprecision. Often the more original scientific work is, the less it will be found to be reliable and precise. As we move towards specific development of an original idea, the more precise and reliable become the procedures used. As Popper has said, "All—or very nearly all—scientific theories originated as myths"¹³—that is to the left hand of my scale. After all, if history has it right, the gravitational hypothesis was derived from the single imprecise observation of a fall of an apple. I do not think Newton would have got that published today. He would have had to make a prospective controlled study of a large number of apples from a stratified but random selection of trees. Supposing some referee had suggested that he used the behaviour of trees as a control—would not that have made the theory unsupportable because trees move up, which contradicts the idea of gravity?

Imprecision then, is not a challenge to originality and may even complement it. Let us recognise that this is so and we can liberalise our approach to clinical knowledge.

Awareness

Provided that we are aware of the reliability of the knowledge we intend to use in any situation, we are not debarred from incorporating it in our decisions. The less it is supported by either collateral evidence from basic biological science or by verification the more circumspect we must be. It is a reasonable hypothesis that lack of awareness of the position of knowledge on the reliability scale and consequent failure to think the problem through has been a more potent cause of persistence with uncertain or inappropriate treatments than has been the analysis achieved by controlled clinical trials. For example, awareness of the shaky foundation of radical mastectomy for breast cancer could have been generated from survival data equally well as from making a controlled trial between it and simple mastectomy with or without radiotherapy. The problem was a lack of awareness by surgeons (and I have no hesitation in including myself) that there was a biological fact—the continuing death of women in whom local disease had been eradicated—which should have told us that things were not as they seemed. The same argument could be applied to currently

proposed trials in the same subject: they do not address a biological as distinct from a technical problem. They are, in the interests of verification, out to establish what is already biologically obvious. There are many circumstances in which this is true. The use of streptomycin in tuberculosis is a good historical example; no one could have had serious *a priori* doubts that it would work, yet it was deemed necessary to use the randomised trial to show that it did. Again, are we not to a degree saying that such a trial is to convince our peers rather than ourselves?

It is often said by the aficionados of clinical trials that it is not justifiable to enter a prospective trial unless there is honestly complete doubt about alternative treatments²—that is, there is truly a 50-50 chance that either procedure in, say, a two arm trial is as effective or as ineffective as the other. The null hypothesis is the accepted starting point. How often can this genuinely be said to be true? The implication is that one comes to a clinical trial with all the tablets razed. In the past this may well have been so and, for certain treatments which have only an observational basis for their current justification, may still be true. Increasingly, however, there is evidence at the middle of the reliability scale before the trial is started; this being so is there not a case for proceeding sequentially? I think there is, though this brings one at once into the realm of the historical control.

Historical controls

The case against history is unanswerable: what once applied cannot be assumed to do so any longer.¹⁴ Circumstances change and comparisons with the past are invalid. High scientific reliability requires contemporaneous comparison. Yet we live by history and all comparisons must of necessity be with the past; transferring the results of a controlled trial into future practice implies using that result as historical evidence. To a certain extent this problem is answered by a rolling programme in which the better result from a previous trial is used as the control arm for a new trial, but it is not always desirable or possible to carry out such a continuous evolution. It can be done only when there is *a priori* evidence that the new treatment is beneficial rather than detrimental and this is in turn an offence against the principle that a controlled trial should be used only when there is genuine doubt about better or worse.

Historical controls are therefore not bankrupt.¹⁵ They have a basis in how we respond to the world despite the scientific imperative. In scientific terms their reliability is obviously less than contemporaneous study but with certain caveats they may be appropriate. They are in the class of organised experience. They contribute most to reliable knowledge in circumstances where there is stability of outcome with prospective recording over several epochs when each epoch is the same length as that proposed for a trial.

The other objection to the historical control is rarely formally stated but it relates once again to the scientific imperative. To produce a scientific answer from an experiment it is necessary to know what factor is truly responsible for any change observed. If historical controls are used there is always the possibility that something else other than the overt change wrought by the new treatment is responsible. This upsets clinical scientists but I wonder does it matter all that much? If, for example, surgical teams can reduce the mortality from duodenal ulcer haemorrhage by 10% between two epochs¹⁶ or quarter their rate of anastomotic leakage in colorectal cancer¹⁷ this certainly fulfils their clinical role irrespective of some slight doubt as to how they did it. Benefit also accrues to their patients. Particularly in the complex circumstances of prospective controlled clinical trials, when uncontrolled variables are rife, there will in my mind always be doubt about which factor is responsible, particularly if the results are negative. In clinical surgery at least, technical progress has been held up by well designed trials which have produced negative results almost

certainly related to the very high complication rates which occurred in the test population and which drowned out any small effect possibly present from the new or different technique being used. (I will not quote references here as it might be offensive or at least invidious. I would be happy to make them available to the interested reader—if he exists.)

Taxonomy of prior hypothesis

Even though the clinical scientist may be convinced of his genuine uncertainty or evenhandedness about outcome in a clinical trial, he is guided by the taxonomy of prior hypothesis. Clinical trials do not issue fully armed from the womb of Fortuna. They are created because there is a hypothesis, positive or negative. Their reliability in terms of being able to produce applicable results rests to a certain extent on the nature of the hypothesis. I distinguish four states.

The first is when a cause and mechanism are well established:

A causes B

A can be attacked or modified by C

Therefore C should be tried when B occurs.

The objective of such a trial is to see if the null hypothesis can be disproved in clinical circumstances given that there is *a priori* evidence that it is not true. We are already well to the right on the reliability scale, and the clinician must ask if there really is genuine persisting doubt and whether a non-randomised pilot study should be performed (there are strong arguments against a *randomised* pilot study but I will not develop them here).

The second is:

A may cause B

A can be attacked or modified by C

Therefore C should be tried when B occurs.

Here the test is of whether or not the relation between A and B is true; C is being used as a test bed. The starting point of reliability of knowledge is more to the left and the prospects of moving it to the right by the trial are, except by the operation of chance, therefore less. A more refined hypothesis could be better. A subdivision of this point in the taxonomy is when not only is there doubt about the relation between A and B but also about that between A and C.

Thirdly:

We do not know what causes B

Certain empirical manoeuvres $C_1, C_2, C_3 \dots$ are thought to influence B

Therefore a test of $C_1, C_2, C_3 \dots$ by controlled trial is appropriate.

Trials of this type are truly eliminatory and may be valuable, if tedious, to get certain things out of contention. But, as I have already indicated, experience and collateral knowledge properly analysed may eliminate some of the C s, and, further, there is a risk that such empirical trials may be swamped by either systematic error or, more important, by overwhelmingly more important factors in determining outcome. The large number axiom is usually used to counter this shortcoming but it does bring in its own problems, both ethical and logistic.

Finally, there is the circumstance that the null hypothesis is the prior one—it is thought unlikely that a difference will be found but it is necessary to convince someone that this is so. Such is often an example of the scientific imperative as I have already outlined, and it is arguable that those who need to be convinced by scientific evidence are the least capable of responding to it. To plagiarise from Patrick O'Donovan, "All the writing on the wall of science is invisible to those who most need to read it."¹⁸

If you accept this taxonomy and relate it to the scale of reliability what emerges is that clinical trials are easiest to do and most likely to provide results when reliability is already well to the right and hardest to do and least likely to provide results when reliability is well to the left. Cynics (or realists)

might translate this into: they are easiest when they are least needed and hardest when they are most required.

Conclusions

I have left out a number of other matters rarely considered by clinicians when they start either to design or to participate in a clinical trial and about which the arch exponents of the discipline are somewhat coy. These include: whether entry by accrual into a trial, particularly a multicentre one, is truly random; whether we should continue to use the formal 5% level of statistical significance and what that level actually means for decision making; and the need to consider the matter of "start-up variation." I hope, however, that I have given enough of an alternative point of view to shift what I fear might have become a central dogma—the concept of overriding need to prosecute controlled clinical trials as the only way of ensuring reliable knowledge.

I am grateful to Professor L P Fielding for discussions and to Miss Sheila Lind-Jackson for processing several drafts.

References

- ¹ Kuhn T. *The structure of scientific revolutions*. 2nd ed. Chicago: University of Chicago Press, 1970.
- ² Cancer Research Campaign Working Party on Breast Conservation.

- Informed consent: ethical, legal, and medical implications for doctors and patients who participate in randomised clinical trials. *Br Med J* 1983;**286**:1117-21.
- ³ Dudley HAF. *Communication in medicine and biology*. Edinburgh: Churchill-Livingstone, 1977.
 - ⁴ Steiner G. Has truth a future? *Listener* 1978;**99**:42-6. (Bronowski memorial lecture.)
 - ⁵ Barnes BA. Discarded operations: surgical innovation by trial and error. In: Bunker JP, Barnes BA, Mosteller F, eds. *Costs, risks and benefits of surgery*. New York: Oxford University Press, 1977:109-23.
 - ⁶ Anonymous. Emotion and empiricism. *Br Med J* 1979;ii:288-9.
 - ⁷ Byar DP, Simon RM, Friedewald WT, et al. Randomised clinical trials. Perspectives on some recent ideas. *N Engl J Med* 1976;**295**:74-80.
 - ⁸ Schafer A. The ethics of the randomised clinical trial. *N Engl J Med* 1982;**307**:719-24.
 - ⁹ Wilson C. *The outsider*. London: Gollancz, 1956.
 - ¹⁰ Popper K. *The logic of scientific discovery*. London: Routledge and Kegan Paul, 1957.
 - ¹¹ Fielding LP, Stewart-Brown S, Dudley HAF. Surgeon-related variables and the clinical trial. *Lancet* 1978;i:778-9.
 - ¹² Paton AW. Ends, means and achievements in medical research. *Lancet* 1979;ii:512-6.
 - ¹³ Popper K. *Conjectures and refutations*. London: Routledge and Kegan Paul, 1963.
 - ¹⁴ Popper K. *The poverty of historicism*. London: Routledge and Kegan Paul, 1957.
 - ¹⁵ Cranberg L. Do retrospective controls make clinical trials inherently fallacious? *Br Med J* 1979;ii:1265-6.
 - ¹⁶ Hunt PS, Hansky TJ, Korman MG. Mortality in patients with haematemesis and melaena: a prospective study. *Br Med J* 1979;ii:1238-40.
 - ¹⁷ Matheson NA, Irving AD. Single layer anastomosis after rectosigmoid resection. *Br J Surg* 1975;**62**:239-42.
 - ¹⁸ O'Donovan P. *The Observer* 1976, Sep 12.

(Accepted 10 June 1983)

Letter from . . . Perth

Puffability People: a new way to promote health

ANDREW G DUNN, PAULINE DUNN

Puffability People, a new health promotion scheme, dominated the central precinct of Perth in Western Australia on the day it was launched, 9 April 1983. Twenty five volunteers consisting of doctors, medical students, physiotherapists, nurses, athletes, and others were assembled with a vast array of high technology spirometric equipment. The long term aims of the organisation, however, lie beyond lung function testing and public awareness of smoking problems and include promotion of health through sport.

From spirometry to promotion

The ideas for forming Puffability People originated in November 1982. During that month Western Australia held its first "smoke free day," coordinated by the National Heart Foundation. As one of several projects undertaken throughout the city

to promote awareness of the smoking problem, spirometric testing was performed in the main corridor of Fremantle Hospital, one of Perth's five large public hospitals. This testing was undertaken by a doctor, an occupational therapist, a pharmacist, and a medical secretary. The term "puffability" was coined for the percentage of predicted forced expiratory volume in one second. The predicted values were made by computerised estimation using age, sex, and height. Although overlap occurred as expected between smokers and non-smokers, the average values for smokers (92% of predicted normal) were less than for non-smokers (101% of predicted normal). These measures made considerable impact on many of the smokers, particularly when they realised that their value would drop further if they continued smoking.

The group itself started in late November 1982, when a pharmacist and one of us (AGD) were interviewed on three different commercial radio stations after the first pilot trial at Fremantle Hospital. From this point we obtained a core of about six or seven volunteers, and we then made a careful search for additional volunteers over the next few months. The group is now about 70 strong.

The effects of spirometric testing as a method of promoting increased awareness of the problem of smoking have been variable. An educated interpretation of the results is, of course, essential. Even more important, however, was the unexpected

Fremantle Hospital, Perth, Western Australia

ANDREW G DUNN, MB, CHB, medical registrar

Queen Elizabeth II Medical Centre, Perth, Western Australia

PAULINE DUNN, MB, CHB, psychiatric registrar