British Journal of Plastic Surgery (1993), **46**, 258–262

BRITISH JOURNAL OF PLASTIC SURGERY

The role of the clinical trial in the evaluation of cleft surgery

R. P. Cole

Department of Plastic and Reconstructive Surgery, Queen Mary's University Hospital, London, UK

"What do you want?" Asked the Prisoner.
"We want information," demanded the Controller.
"You won't get it," retorted the Prisoner.

Although we need not be as pessimistic as Thomas Disch's Prisoner would have us believe, we will need to cast our net as widely as possible to gather all the information we require to evaluate cleft surgery.

In few other branches of surgery is there such a variation in treatment options and protocols as in cleft surgery. The multidisciplinary approach needed in such a complex field necessitates interactions between highly specialised individuals who inevitably focus on those aspects of the patient's care with which they have been trained to deal. Even if one just considers cleft surgery in isolation, several specialties are involved and there is even some blurring of the boundaries between them. The experience of the individual surgeon dealing with the primary treatment of the cleft various tremendously, whether one considers experience by grade and duration of training, or whether one considers number of cases performed annually. All other things being equal, perhaps the greatest variation is with the choice of operation and its timing. To address these complexities is a challenge indeed, but one which has to be faced.

Why should we evaluate cleft surgery? It is generally accepted that the more information we have about what we do (process of treatment) and our results (outcome), the better our chances of improving the service we provide for our patients. This is the main reason for the widespread acceptance of the principle of audit. However, there is no point in evaluating what we do, whether by audit or other means, if the infrastructure is such that change is difficult to accommodate or adopt. Cleft surgery at present is delivered by a spectrum of surgeons ranging from the occasional practitioner to the pure subspecialist. With such a disparate delivery system, it is likely that potential advances will be difficult to institute widely. Until the treatment of clefts becomes more centralised, the possible gains from advances will be diluted. Such a shift in the way care is delivered will impinge on the principle of clinical freedom and because of this it is likely that a compromise situation analagous to superspecialisation will occur in cleft surgery.

Evaluation of cleft surgery must include not only process and outcome, but also cost. Cost must be considered in terms of use of resources such as hospital bed stay and theatre time, total number of procedures required, consumption of parasurgical services such as speech therapy, and also the "cost" in human terms of treatment. The latter may be difficult to define and quantify but as in so many other areas of our practice, this does not mean that it can be ignored.

Finally, we should evaluate cleft surgery within the context of our overall practice as a specialty. With the recent changes introduced in the White Paper, the provider of services (at whatever level) will be expected to be aware of the impact of changing one aspect of his or her practice on the rest of his or her patients. If a new advance demands an increase in consumption of resources, this may require a reduction of resources available in another area. These implications must also be considered when we evaluate cleft surgery, by whatever means we choose.

The clinical trial and its limitations

It is widely accepted that the clinical trial is the best way to evaluate a particular intervention and that the randomised clinical trial is its best design.

One of the earliest clinical trials is related in the Book of Daniel, when Daniel asked for a special diet of pulse and water for himself and his companions so that they would not be defiled by eating the meat and wine diet of Nebuchadnezzar. He asked the chief eunuch to compare how they fared against other children in the court who were fed on meat and wine. The experimental group had a better outcome than the control group, so Daniel and his companions were allowed to continue with their diet and were not defiled.

Although it is acknowledged that clinical trials carry great authority in the evaluation of our practice, it has been estimated that only 10–20% of all medical procedures have been shown to be efficacious in clinical trials. The use of the randomised clinical trial and its alternatives will now be critically discussed in general terms before specifically dealing with cleft surgery.

The alternative design of clinical trial to the randomised clinical trial is the non-randomised clinical trial, which uses "external controls". These external controls are drawn from a population which may differ in unknown ways from the study population. External controls may be "historical controls" which are the most commonly used type, or "concurrent controls". The latter are patients drawn from a similar population to the study group but are not assigned randomly to act as controls. Because external controls are not selected at random, the assumption has to be made that the population from which they are drawn is

identical to the study population in terms of all factors which influence the outcome of the intervention. This implies that all prognostic factors influencing the outcome must be known to medical science. Inevitably this will not always be so. However, using an external control group avoids the problems associated with randomisation, not least the ethical issues which will be discussed below. In addition, fewer individuals need to be recruited into the trial and its cost is also reduced. Because patients do not need to consent to randomisation and know which treatment they are to receive before they enter the study, recruitment into the trial is greater.

Historical controls can be taken from patients treated at the same institution by the "control" method or can be taken from other reports or publications in the medical literature. Generally, medical records are not kept with the idea in mind that they will be used in the future to provide a control group. An alternative supply of external controls might be provided by data banks. These would consist of stored information from many different centres and it is not difficult to see how results generated by audit could be used in this way.

Notwithstanding the advantages discussed above of using external controls, the clear benefit of the randomised clinical trial is that any differences observed between study group and control group are far more likely to be due to the treatment rather than any other difference between the groups. It has been shown that non-randomised studies (using external controls) tend to overestimate the benefit of an intervention. For these reasons, the randomised clinical trial is the preferred tool for evaluating treatment interventions. However, it is not so easy to discard the alternatives because of certain stumbling blocks of the randomised clinical trial which will be discussed below.

In addition to the relative expense and administrative complexity of the randomised clinical trial, there are major ethical and statistical pitfalls to be considered. At the core of the ethical problems is the principle that the patient has a right to receive the best treatment available and that the doctor has a duty to provide it. Assuming one treatment is better than the other, because of the study design, one group will intentionally be denied the benefit of that treatment. Professor Fincke, Professor of Criminal Law at the University of Bielefeld, has argued that a doctor participating in a randomised clinical trial would be technically guilty of manslaughter if he or she continued to use a control intervention while suspecting (correctly) that the other group had a lower mortality. In addition, there is a vague unease that the patients are being exploited regardless of any benefit they might receive.

Ethical objections to the randomised clinical trial include the way consent is obtained. The Helsinki Declaration made it clear that informed consent should be the rule in clinical trials. It is easier to obtain informed consent for a single operation than to do so for a clinical trial, when the individual has to consent first to be randomised and then in effect give consent to two different interventions which carry different complications and outcomes. In practice, when faced with the choice between two different tablets or different types of skin closure, for instance, a patient is far more likely to consent to be randomised than when facing a choice between two different operations. The recruitment of patients into two recent multicentre studies of surgical treatment of breast cancer and extracranial vascular disease illustrate the difficulties of initiating trials of surgical treatment. The multicentre Cancer Research Campaign trial of surgery for breast cancer was abandoned after only 145 patients had been recruited in 3 years. The Extracranial-Intracranial Bypass Study Group involved the entry of less than half of all eligible patients into the study and its conclusions were therefore highly questionable. These problems are specific to interventions which are major and complex and this includes many trials of surgical procedures.

Two further areas should be addressed when considering the ethical problems with randomised clinical trials. In a study which examined two methods of seeking informed consent, it was shown that subjects given a brief written explanation had a good comprehension score of the procedure and a low refusal rate to the intervention. The subjects given a much longer, more detailed written explanation had a lower comprehension score and a much higher refusal rate. A further problem with obtaining informed consent for surgery involves the drift towards involving the child in decisions regarding surgery. Parents may be reluctant to consent to their child being randomised to a surgical intervention not only because they fear that their child might not receive the best treatment, but also because of the knowledge that later on in life the child might pursue litigation against them should the outcome be unfavourable. This situation has already occurred outside the randomised clinical trial. Even if these problems do not restrict the use of the randomised clinical trial, there are certain statistical considerations which might do so.

In many controlled trials the number of cases recruited is too small to demonstrate the superiority (or inferiority) of the intervention if it exists. Failure to show the difference is known as a "Type 2" error, or a false negative. The probability of this error occurring can be expressed as the "power" of the study. A "Type 1" error is the demonstration of a difference when there is none, or a false positive. The probability of this error occurring is expressed as the statistical significance. When setting up a randomised clinical trial it is necessary to decide the sample size needed to demonstrate as statistically significant an effect of a predetermined level. This predetermined level or "expected benefit" is crucial. If the sample size is too small to show the benefit expected of the intervention, it will appear that the intervention offers no (statistically significant) benefit. However, the intervention might have been shown to confer a benefit (if less than expected) had the sample size been larger. The intervention may therefore have incorrectly been "written off" because of a Type 2 error due to inadequate sample size. The problems associated with small sample sizes and small expected benefits can usually only be addressed by the use of multicentre trials, which are not themselves without problems.

The time taken to recruit an adequate sample size is reduced by increasing the number of centres involved in the trial. A further advantage of the multicentre study is that the sample is recruited from a more heterogeneous population pool. The results are therefore more readily applicable to the population as a whole. However, multicentre trials are harder to supervise and administer. They are also usually more expensive than single centre trials. In addition to this, there is a more subtle but critical problem when multicentre trials of surgical interventions are performed. In a drug trial the dose, timing and route of administration are easy to define precisely and it is to be expected that all the individuals in the study will receive exactly the same treatment whichever centre one considers. With surgical interventions, however, exactly the same method is unlikely to be used by all the surgeons in every centre when performing what is called the same operation. Even minor variations in technique might have a significant impact on the outcome and it is impossible to control for these. This objection to the use of multicentre trials for the evaluation of surgical interventions leads into an area where rather more esoteric problems of evaluating practice exist.

General considerations in evaluating practice

We can easily distinguish between a great actor reading Hamlet and the average actor reading the same passage. The costume, the posture, the setting, the words and even the time of delivery may be the same and yet there is a tremendous difference. This analogy can be applied to the surgeon. We might use the randomised clinical trial to evaluate individual elements or components of our practice. We might even manage to evaluate what we consider to be all the important components. However, we must consider that in some surgeons' hands, as in some actors' hands, the whole may be greater than the sum of the parts. The nihilist might argue that although we may get all the information we think we require, we can never get the information we really need because it is too elusive and ill-defined. One doesn't have to be a nihilist to appreciate that in one person's hands a procedure may work much better than in another person's hands. We ignore this but do so at our peril.

Some of the greatest advances in medical practice have been introduced without clinical trials, for example the use of asepsis in early surgery and the use of penicillin in life-threatening infections. The most important breakthroughs have been so obviously an advance that there was no point comparing them against anything other than past experience. If we are evaluating our practice in order to improve the service we provide, we should not lose sight of the fact that a "real" advance will not need a clinical trial to evaluate it.

The truth in the apparent tautology, that you can only measure what you are measuring, must also be discussed here. This is a drawback of a rigid study design. For example in the treatment of breast cancer, for many years studies have randomised patients into clinical trials to evaluate individual components of treatment. By randomising patients, it is possible to "mask" a factor or variable which might have a far greater influence on outcome than the component under evaluation. The timing of surgery in relation to menstrual cycle has only recently been recognised as a major prognostic factor in breast cancer. Similarly, in randomised clinical trials the impact on prognosis of peri-operative blood transfusion in patients with colorectal cancer had been "controlled" out of the equation. Great care must be taken to remember that unidentified factors may be found to influence outcome more than the factor under consideration.

Evaluation of cleft surgery

There is a relative lack of published clinical trials in the evaluation of cleft surgery. The areas explored above reveal some of the potential pitfalls which might be encountered if more clinical trials are to be performed. These areas apply to clinical trials in general. In addition to these, there are further problem areas which are specific to cleft surgery and these are probably responsible for the lack of published trials to date.

It is widely accepted both in the medical literature and in the NHS management journals that one of the main problems with the implementation of the NHS reforms, and audit in particular, is the lack of information on outcomes. In cleft surgery, as in many other areas in plastic surgery, outcomes are difficult to define. Many objective measures of outcome, for example joint excursions after fasciectomy or flexor tendon repair, give only part of the picture. An overall global subjective assessment of function or "usefulness" would complete the picture. Such a subjective assessment is often little more than an impression, however, and is difficult to quantify. The tendency, therefore, is to persist with objective, quantifiable aspects of outcome which can be expressed numerically, or as "Yes/No". In cleft surgery, an example would be the assessment of outcome of speech. Velopharyngeal competence can be assessed using videofluoroscopy or nasendoscopy and is considered to be objective although, inevitably, some inter-observer variations will exist. Although velopharyngeal competence can be assessed and quantified in relatively precise terms, the main criterion of successful outcome of speech is whether or not the patient has normal voice quality and articulation as perceived by society. If the objective parameters (such as velopharyngeal competence) correlate closely with this, then one can intellectually accept their role as outcome indicators. However, if we find, even using all available objectively quantifiable data, that we consider a patient to have a good outcome but the layperson does not, then we must attempt to incorporate more subjective information into our assessment.

The problem with defining outcome of speech applies to the other areas in cleft patients. The appearance of the nose is of major importance to the

patient and therefore an area which needs to be evaluated, yet objective measurements by which outcome can be defined remain elusive. Perhaps the only outcome which matters here is the patient's own perception of his or her appearance. It is difficult to envisage how a clinical trial could address an area like this with no "hard" endpoint to measure. A clinical trial to evaluate a different approach to the cleft lip nose, for example, would also have to address the problem discussed above that different surgeons may do a "standard" procedure differently. Even a supposedly conventional rotation-advancement lip repair might be performed with subtle variations which make comparisons between individual surgeons, let alone different centres, impossible. It might be argued that clinical trials should be performed by single surgeons randomly assigning patients to either arm of the study. This still leaves a myriad of other problems as discussed above.

Returning to outcomes of treatment, cleft surgery is unique in having opposing outcomes, such as timing of palate repair influencing speech and facial growth diametrically. A clinical trial to examine the potential advantage of a particular procedure in terms of speech may reveal that advantage, but it might be at the expense of impaired facial growth. The conclusion of such a trial would be virtually impossible to apply in practice unless a precise "weight" could be given to speech and facial development so that an overall benefit (including both parameters) could be expressed for each treatment arm. Areas for further investigation might only be revealed by studies such as the Sri Lankan project which demonstrated patterns of facial growth in the unoperated cleft, information which could not have been obtained from a modern Western clinical trial.

A further problem in evaluating cleft surgery by clinical trial is the time required for the study to run until a meaningful endpoint can be measured (assuming that to be possible, vide supra). Most outcomes in cleft surgery are not evident, whether subjectively or objectively, for many years after the initial intervention. For example, a clinical trial addressing facial growth in relation to timing and extent of muco-periosteal palatal dissection in palate repair cannot end until primary and secondary dentition are complete in order to assess transverse and anteroposterior maxillary arch development. If one can only change one component of treatment at a time in order to proceed using clinical trials, it would take decades for the impact of a few changes in intervention to be known. The problem with definition of outcomes and with the time required before outcomes can be assessed applies, of course, not only to clinical trials but any other way by which cleft surgery might be evaluated.

The use of the clinical trial in evaluating cleft surgery may be further restricted by the numbers of patients available for study.

Notwithstanding the above criticisms of multicentre studies, these might be the only way to recruit sufficient patients to avoid Type 2 errors. With a typical sample size of 50 patients, there is a probability of less than 0.4 that a randomised clinical trial will detect a difference between the two treatment arms of a 20% response rate in one group compared to a 40% response in the second group. Differences in response rates or outcomes in cleft patients are likely to be of this order. Therefore, sample sizes are likely to need to be in the hundreds if Type 2 errors (false negatives) are to be avoided. The logistic and administrative problems with the recruitment of these numbers of patients would be immense. There is no doubt, however, that if all these hurdles are negotiated, the internal validity of the randomised clinical trial will ensure that a positive result would permit the intervention to be recommended emphatically.

The way forward

Progress in cleft surgery can only be made if we have more information. This information cannot just come from randomised clinical trials. The internal validity of these trials make them the gold standard in evaluating different treatments. However, given the ethical, statistical and logistical problems they present, and those problems specific to cleft surgery, their role will be restricted. Their role may be to address what may be perceived as being matters of minor importance such as peri-operative antibiotic prophylaxis or type of sutures for closure, but which may yet have a real impact on outcome. Major areas of investigation might be more easily addressed by clinical trials using external controls. The pooling of audit information from different centres in data banks might provide control groups for comparison.

Most information will inevitably come from reports which present the results of "packages" of treatment, for example, the Zurich approach as performed by a particular team of surgeons, orthodontists and speech therapists. Most publications in the literature are of this form and generate much information for debate. The main criticism of this kind of information is that it does not assess individual aspects or components of treatment but gives an overall impression of what can be achieved with that particular "package" or philosophy. It cannot be denied that a specific philosophy applied to a particular population may produce results that are superior to what might be considered (or even be shown to be) a better philosophy used elsewhere. Ultimately, to be able to produce a good outcome using a particular philosophy reveals a better understanding of what is required of health care delivery, than the desire to find a single "best" treatment which should be adopted universally. There must be a happy medium which is based on good communication, careful documentation and open and frank discussion of results.

The future for cleft surgery may lie in widespread, centrally directed, critical and detailed audit which would generate specific ideas which a randomised clinical trial of manageable size could tackle. A few particular areas considered to have a large potential impact could then be assessed in detail. There must always be space for apparently unconventional ideas to be considered and debated. The attitude that change can only come through clinical trials may stifle creativity. After all, many great advances have come

serendipitously and been introduced without having to resort to the clinical trial. The coming decades in cleft surgery may lead to progress in prevention or perhaps intra-uterine surgery. Neither of these lend themselves to randomised clinical trial of cleft surgery, the former because it is not surgical, and the latter because of the ethical problems of randomisation.

The best service for the cleft patient will be provided if we can gather as much information as possible from whatever source, and do not become the prisoner of the clinical trial.

The Author

Mr R. P. Cole, Formerly; Registrar in Plastic Surgery, Wessex Centre for Plastic and Maxillofacial Surgery, Odstock Hospital, Salisbury SP2 8BJ. Now; Senior Registrar in Plastic and Reconstructive Surgery, Queen Mary's University Hospital, Roehampton Lane, London SW15 5PN.

Requests for reprints to the author.

Paper received 4 January 1993.
Accepted 4 January 1993.

This paper was awarded the Kay-Kilner essay prize in 1991. Ed.