Evidence-based choice of hip prostheses

The prostheses in current use for hip replacement vary in cost by a factor of nearly ten (Murray, Carr and Bulstrode 1995). Since more than 40,000 hip arthroplasties are performed annually in the UK, it is clear that research is needed to establish whether the varying cost is related to long-term survival in patients and to the outcome as regards comfort and quality of life. In 1995, the National Health Service Research and Development Programme commissioned a systematic review which aimed at estimating the relative effectiveness of different prostheses.

In this issue of the Journal, there is a report of a long follow-up of the early Charnley and Stanmore prostheses by Britton et al (p.802-8). This suggests that the Charnley

R. W. Morris, PhD, Senior Lecturer in Medical Statistics
Department of Primary Care and Population Sciences, Royal Free Hospital
School of Medicine, Rowland Hill Street, London NW3 2PF, UK.

©1996 British Editorial Society of Bone and Joint Surgery
0301-620X/96/51280 $2.00
design, which accounts for about half of the hip prostheses implanted in the UK (Hashemi-Nejad, Birch and Goddard 1994), has a ten-year survival which is inferior by about 9%. It is of interest that the results for the Charnley implant are less good than those reported from major centres such as Wrightington and the Mayo Clinic. If this comparison is true, the difference between the implants is important in terms of the cost and clinical workload required for later revision surgery.

Orthopaedic procedures have usually been evaluated by means of case series (Morris 1988), and this is particularly true of the alternative prostheses available for hip replacement. This means that a valid comparison of series is virtually impossible: they come from different centres, with a widely differing selection of patients and measurement of outcome. There is a similar situation in general surgery, as emphasised in a recent Lancet editorial (Horton 1996).

Britton et al have compared the long-term survival of the Charnley design of hip prosthesis, most of them inserted by one consultant surgeon from 1974 to 1976, with that of the Stanmore design implanted by the same surgeon from 1977 to 1982. This is a historically controlled trial. Its advantage over the comparison of case series from different centres is that the same surgeon performed the operations and that he used the same follow-up and outcome assessment for both types of prosthesis. His selection criteria may well have changed during the whole nine-year period but the difference is likely to be much less than that between different surgeons in different centres. Another factor is that the surgeon was not involved in the design of either prosthesis; it may be claimed that his results provide an independent view. The meticulous assembly of such an extensive data base and the energetic follow-up should be applauded.

In recent years there has been increasing emphasis on evidence-based medicine, whereby all the relevant literature on a given subject is searched and critically assessed for its validity and relevance (Rosenberg and Donald 1995). The findings reported by Britton et al must therefore be appraised in conjunction with other studies which compare the Charnley and Stanmore prostheses such as that of Marston, Cobb and Bentley (1996), which reported the random allocation of a Charnley or Stanmore design of implant for 413 hip replacements in one centre from 1982 to 1987.

The main advantage claimed for truly random allocation is that the characteristics of patients assigned to each intervention are on average similar. When this is true any differences in the outcome may be attributed to the intervention rather than to the characteristics of the patients treated. Table I of the paper by Marston et al (1996) confirms that the two groups were similar in terms of preoperative diagnosis, seniority of the surgeon and surgical approach. There are many other variables which are related to surgical success, but genuinely random allocation should also balance these. For non-randomised studies, attempts may be made to allow for the known variables which may influence success, but unknown or unmeasurable variables cannot be taken into account.

The claimed advantages of randomised trials are not just academic. In a classic paper, Sacks, Chalmers and Smith (1982) compared studies in six specific areas of medicine, including coronary artery surgery, for which both randomised controlled trials and historically controlled trials had been published. Most of the randomised trials showed no benefit from a new treatment, but, in marked contrast, 80% of the historically controlled studies suggested that it was better. One apparent reason for this difference was that during the whole period of each historically controlled study selection criteria had changed; fewer high-risk patients were given the new treatment. The historical series of patients did not have the benefit of such selection. There is also, of course, a steady time trend for improvement in many other factors including operative skills and details of technique and management.

We cannot know whether patients were selected carefully for the Stanmore prosthesis in 1974 to 1976 while the surgeon became accustomed to it, or whether he reserved the more familiar Charnley prosthesis for complex cases. Could this help to explain the difference? There are other possible reasons: the authors address two of them, the use of a cement gun and the growing experience of the surgeon. Even attempting to allow for these the Stanmore design still seemed superior but to a reduced extent. It is impossible to discover whether other improvements in orthopaedic technique and general health care can explain the remaining difference. A surgeon cannot be expected to anticipate and record every conceivable factor which may influence long-term survival.

The study of Marston et al could perhaps be taken as the gold standard, because it is a randomised trial. The report is from the centre at which the Stanmore prosthesis was developed; one might expect that surgical expertise was optimal for this design. Despite this no superiority to the Charnley design was demonstrated. Over half of the operations for both prostheses were undertaken by trainee surgeons; the paper notes incidentally that operations carried out by such trainees were more likely to result in revision surgery. One strength of the study is that the conditions of the trial were not too narrow: this increases the extent to which useful general conclusions may be drawn. The publication of this trial may help to reduce the pessimism of those who believe that randomised trials cannot be carried out in orthopaedic surgery. The problems caused by continuously changing techniques and the difficulties of blind assessment have been discussed by Dorey, Grigoris and Amstutz (1994), but it should be emphasised that these do not disappear if other study designs are used; quite often the difficulties are simply compounded.

The sample sizes used by Marston et al were larger than those in most randomised trials of orthopaedic procedures, but may still be too small. The median follow-up was between seven and eight years, less than that for the series...
of Britton et al. The quoted 95% confidence interval of the odds ratio for revision shows that a Charnley implant could be either two-and-a-half times more likely or only one-quarter as likely to end in revision surgery as a Stanmore implant. Any extrapolation over longer periods would suggest that a widespread change in the UK from the Charnley to the Stanmore design could actually result in either a considerable increase or a considerable decrease in the overall need for revision surgery. Even if the findings of Marston et al are valid, the information which resulted cannot be said to be reliable: the sample size of 413 patients is simply not large enough when the event of interest, revision surgery, is observed in less than 5% of cases.

The main problems for the design of randomised trials are that rare outcomes require large sample sizes and long follow-up, only one aspect of care may be studied at a time, patients’ and surgeons’ preferences are ignored, participating surgeons may be unrepresentative, and the process of a trial may contaminate the effectiveness of the alternative procedures. These have recently been discussed by Black (1996), who provides support for rigorously conducted observational studies as an alternative and complementary source of evidence. The population registries of joint replacement in Sweden (Malchau, Herberts and Ahnfelt 1993) and Norway (Havelin et al 1995) are good examples of such an alternative source; a national register has now been advocated for the UK (Sochart, Long and Porter 1996).

Nevertheless, randomised trials can address the need for large sample sizes involving representative groups of surgeons, if they are carried out on a multicentre basis. Surrogate outcome measures such as roentgen stereophotogrammetric analysis, or patient-reported pain, may reduce the requirement for long-term follow-up (Bulstrode 1996). Patients’ preferences are unlikely to be a major issue with alternative prostheses, and although individual surgeons may have preferences, their collective lack of consensus is probably enough to satisfy the ethical requirements for such a trial.

In the present circumstances, no radical improvements in the outcome of hip replacements can be expected from the introduction of new designs of prosthesis. Moderate advances are certainly possible; these may be worth detecting because they apply to very large populations of patients. Such advances must be studied in large randomised trials: they will certainly be too small to show up in smaller studies or those which contain bias. In the early 1980s, there was an analogous situation concerning the treatment of patients after a myocardial infarction. Aspirin has a relatively small effect which is invisible to an individual clinician or even a small trial. A randomised multicentre trial of some 16 000 patients showed its value unequivocally: its use now has the potential to avoid 100 000 vascular deaths per annum in the developed countries (Peto, Collins and Gray 1995).

Randomised trials of orthopaedic procedures are rare; multicentre randomised trials are even less common. A frequent objection is that too many factors affect the success of a hip replacement. Researchers tend to use narrow criteria for eligibility which limit both the sample size and the general application of the results. The opposite view is better: the more factors that are known to influence success or failure, the more we need large, simple randomised trials with wide criteria for entry. Proper random allocation will produce a balance of both known and unknown influential factors so that differences in outcome may be reliably ascribed to the tested intervention. The findings of such a study will also be generally applicable. There are, of course, special issues of design for surgical research; the evaluation of alternative hip prostheses differs considerably from the evaluation of simple oral medication. The overriding need is close collaboration between surgeons and trialists to design and perform large simple, randomised studies on the wealth of clinical material which is available annually. The large population which will require hip replacements in the future and is currently paying for a health service may reasonably expect orthopaedic surgeons to choose their prostheses on the basis of valid and reliable information. Must evidence-based choice remain a far-fetched concept?

RICHARD MORRIS

REFERENCES


Horton R. Surgical research or comic opera: questions, but few answers. Lancet 1996;347:984-5.


