Editorial

Retrospective evaluation of implants—Is there a need?

New implant technology should deal with problems of clinical relevance and especially those related to long-term complications in young and active patients with a long expection of life. The excellent long-term outcome of THR in the less active and most commonly-operated types of patient leaves little place for experimental changes in an implant without a specific purpose (Murray et al. 1995) and a relevant outcome analysis.

During the last decade, numerous studies have been concerned with the best way of evaluating new joint implants and surgical techniques (Kärrholm at al. 1997, Ritter and Albohm 1997, Garellick et al 1999, Söderman et al. 2000, 2001). A first clinical trial of a new implant should be done in a consecutive pilot study, which is randomized against a current standard. To minimize the number of patients exposed to the potential hazard of a device or surgical method with doubtful outcome, implant fixation, wear, bone remodeling and the release of potentially toxic substances should be monitored with the most accurate methods available. In the selection of outcome parameters, emphasis should be placed on a correct evaluation of the particular problem that the new implant is supposed to address. If, for example, a new articulation that is expected to reduce wear is evaluated, accurate measurements as to wear are necessary. Other potential problems, which might be effects of necessary changes in the mechanical properties of the material to increase its wear resistance should be detected as early as possible. If these principles had been consistently applied, many inferior implants and bone cements would not have come on the market for the benefit of our patients.

A controlled introduction of a new implant is an expensive and time-consuming process. The first clinical trial will last at least for 2–3 years. If this trial speaks in favor of the new device, a multi-center trial will be done for at least 5–7 years and, in many cases, longer. Even if this extended introduction is a step forward it probably can not foresee all complications. THR is a procedure involving a number of steps. It gives a final clinical result. Even small changes in the surgical technique may affect the long-term results. Thus, evaluation of the reproducibility of the procedure at various centers is important. This limits the conclusions to be drawn from a small primary study, which may have been done in a center of excellence.

However, such an approach is very cost-effective for society. Had these principles for implant introduction been known when the Christiansen prosthesis became available in the 1970s, a conservative estimate would have obviated about 2–3,000 revisions, corresponding to a minimum cost of 22–33,000,000 Euro. The costs of disability and sick-leave should be added as well as the individual suffering of the patient. The winding road of total hip replacement from the early 1960s until today is lined with similar examples, such as the introduction of first generation cementless implants, liner fixation problems, alternative bone cements and cemented titanium stems with a matte or rough surface.

It is easy to be wise after the event, but this costly uncontrolled experimental approach recurred with the Boneloc and Capital hips only a few years ago. One explanation of these catastrophes may be that orthopedic surgeons do not always appreciate that only minor changes in an implant may dramatically change its performance. One of the first examples of this was the change in the surface finish on the Exeter stem. Today, metal-on-metal surface replacements and many THRs are introduced with modifications in the CCD-angle or the offset, but we do not yet know whether, or to what extent,

these new devices or seemingly minor changes in well-documented designs will work better than the current standards. In the United States, most surgeons have changed from using "conventionally"-sterilized polyethylene to highly cross-linked ones. On the basis of current knowledge this seems to be beneficial, but the long-term performance of these materials is still unknown. Further, the so-called "cross-linked" polyethylenes are not made of the same material. The type of radiation, doses and way to eradicate free radicals vary widely, which probably affects their long-term performance.

In 1995, when the Boneloc disaster became apparent, Lars Linder pointed out that the lesson to learn from 30 years of implant trials should have been learned (Linder 1995). At that time this was obviously not so and the examples given above may raise concern for the future. Why is it so difficult to implement a more cautious introduction of new devices?

One reason is that the appropriate introduction of new implants is expensive and time-consuming and means that a new device, provided that it has advantages over existing ones, can not be used until about one decade later. The procedure may be cost-effective for society, but the long period between its introduction and its general use may raise serious concerns in the implant industry.

Nonetheless, there is certainly some justification for stepwise and controlled introduction of new materials and implants. Reports about such studies and not least those with negative findings, and subsequent multi-center trials, should also be of particular interest for scientific journals. The question is to what extent there is a place for retrospective studies of THR, perhaps determined retrospectively, to become a "scientific" evaluation. In the ideal world such studies would rarely be done. In reality, they are very common and many contain interesting information.

If such studies are reported, they should be adequately interpreted. This requires a strictly scientific approach. Patient selection must be noted. There is a large difference between the expectations and amount of activity of an elderly patient with generalized joint disease and a young patient with posttraumatic degenerative disease of only one joint. The length of follow-up and number of patients available on each follow-up should be reported. The reasons for patients lost to follow-up should be clearly stated. The instruments used to evaluate the clinical outcome should be generally accepted and tested for validity, reliability and responsiveness. A questionnaire is valid if it can be shown that it measures what it is supposed to. It is reliable if it yields the same result when no true change has occurred. It has good responsiveness if it detects small changes in clinical findings.

To detect impending clinical failures, a radiographic evaluation of migration and linear (radiolucent lines) or focal osteolysis must be done. The author's definition of clinical and radiographic failure should be clearly stated and based on scientific grounds. If no specific measures are taken to improve the quality of conventional radiographic measurements, component migration less than 5 mm can rarely be detected in the individual case. When measurements of wear are of interest, the methods should be described. If new methods or methods with unknown resolution are used, their precision and preferably the true error (accuracy) should be given.

The accuracy can only be studied by comparisons with known and reliable data, but the precision concerns the reproducibility of the method. Repeated measurements of the same radiographs may be used to evaluate one of the errors in the precision. A true evaluation of precision, however, should include repetition of the whole procedure, including measurements done on—e.g., a second radiographic examination—which, in the absence of any errors, should provide the same result. In reality, there will be errors. These may be due to changes in the projection caused by different positions of the patient and the x-ray sources between the examinations in addition to the measurement errors. Unfortunately, such "true" evaluations of the precision are rarely done, probably because of ethical concerns and other reasons.

Survival diagrams should always include confidence limits and numbers at risk for each interval. Most prostheses have 2 main components, 1 on each side of the joint. Even if a report focuses on one part of the prosthesis, the outcome in the other part should be briefly mentioned, especially if it differs from that of designs with thorough documentation—e.g., in the Scandinavian registers. The outcome of bilateral THR is not independent

of each other. In randomized evaluations of 2 prosthesis designs, one patient can not be included if both hips have received the same type of implant.

If all manuscript submitted to scientific journals fulfilled these basic criteria, authors, referees and editors could spend less time on these problems and more manuscripts would probably be published. Unexpected failures, however, may also occur in patients whose case was never intended for publication. If so, all available information should be used, so far as possible, to meet the above requirements. The first report on poor clinical results with Boneloc cement was, in fact, published in *Acta Orthopaedica Scandinavica* (Suominen 1995). This finding was based on 4 failures in only 8 patients! In other cases, consideration of these guidelines before starting on a long and painstaking process to evaluate retrospective patient material may become an incentive to start another, and perhaps more rewarding, project.

The updated "Instructions to authors" in this issue of *Acta Orthopaedica Scandinavica* is one contribution to facilitate forthcoming submissions of articles. We hope that this reworked edition and information on the websites given will help our authors and readers to carry out and submit studies of good quality to our journal.

Johan Kärrholm
Department of Orthopaedics, Sahlgren Hospital, Göteborg University, SE-413 45 Göteborg, Sweden.
eva.fermen@orthop.gu.se

Garellick G, Herberts P, Malchau H. The value of clinical data scoring systems: are traditional hip scoring systems adequate to use in evaluation after total hip surgery? J Arthroplasty 1999; 14 (8): 1024-9.

Kärrholm J, Herberts P, Hultmark P, Malchau H, Nivbrant B, Thanner J. Radiostereometry of hip prostheses. Review of methodology and clinical results. Clin Orthop 1997; 344: 94-110.

Linder L. Boneloc-The Christiansen experience revisited. Acta Orthop Scand 1995;66 (3): 205-6.

Murray D W, Carr A J, Bulstrode C J. Which primary total hip replacement? J Bone Joint Surg (Br) 1995; 77 (4): 520-7.

Ritter M A, Albohm M J. Overview: maintaining outcomes for total hip arthroplasty. The past, present, and future. Clin Orthop 1997; 344: 81-7.

Suominen S. Early failure with Boneloc bone cement. 4/8 femoral stems loose within 3 years. Acta Orthop Scand 1995; 66 (1): 13.

Söderman P, Malchau H, Herberts P. Outcome after total hip arthroplasty: Part I. General health evaluation in relation to definition of failure in the Swedish National Total Hip Arthroplasty register. Acta Orthop Scand 2000; 71 (4): 354-9.

Söderman P, Malchau H, Herberts P, Zugner R, Regnér H, Garellick G. Outcome after total hip arthroplasty: Part II. Disease-specific follow-up and the Swedish National Total Hip Arthroplasty Register. Acta Orthop Scand 2001; 72 (2): 113-9.