Authors' objectives
To examine whether different prostheses for total hip replacement (THR) were associated with different medium to long term outcomes. To use evidence regarding both costs and outcomes of primary THR to model how much more effective newer prostheses must be to justify higher costs.

Searching
Journals to be handsearched were selected by identifying those with the highest rate of relevant publications. This was done by searching 3 years of MEDLINE (1985, 1990, 1994) and identifying the 11 journals that appeared with the highest frequency. A similar strategy was repeated with the EMBASE search database, and it was found that the seven journal articles identified as having the highest yield of papers had already been selected by MEDLINE. Three further journals identified via EMBASE were not included because of the very low total yield of relevant publications. Search strategies for both databases were included in the appendix. Articles were included provided that either the article itself or an abstract was available in English.

The 11 journals identified (listed in the report) were searched by three experienced handsearchers (1980-1995).

Electronic searches of MEDLINE and EMBASE were also conducted for the years 1980 to 1995 based on search strategies which were provided.

Study selection
Study designs of evaluations included in the review
The following types of study were included:

1. Randomised trials of prostheses for THR regardless of length of follow-up.
2. Comparative observational studies of prostheses with concurrent controls regardless of length of follow-up.
3. Observational studies of single prostheses with at least 5 years follow-up.

Single case studies, and observational studies not using outcome measures as defined above were excluded.

For RCTs, the mean length of follow-up was 3.9 years (the range of follow-up was 1 to 6.5 years). With the exception of 3 studies, comparative observational studies followed up patients for at least five years.

Specific interventions included in the review
Prostheses for primary total hip replacement (THR).

Types of prostheses included the meta-analysis were: Charnley, Muller, PCA, Ring, McKee-Farrar, Harris-Galante, Stanmore, Charnley-Muller, Lubinus and the Exeter. Many other types of prostheses were reported as evidence from RCTs and comparative observational studies, including: HP Garches, Spectron, Mallory cemented/cementless, Mecron cementless, Triad, Miami MOSC, and the Brunswick.

Participants included in the review
Patients with prostheses for primary THR. Patients with congenital hip problems, hip fracture, and hemiarthroplasty, and those who had received revision surgery were excluded.

Outcomes assessed in the review
Outcomes were defined as either the occurrence of revision surgery or standardised assessment of patients’ pain and function. Global ratings of success/failure were also included. Studies assessing outcome solely in terms of radiological evidence of loosening were not included.

How were decisions on the relevance of primary studies made?
The authors do not state how the papers were selected for the review, or how many of the authors performed the selection.

Assessment of study quality
A check-list was drawn up based on a previous check-list developed by one of the current investigators to assess the quality of studies appearing in an orthopaedic research journal (See Other Publications of Related Interest). The authors also drew on a commonly cited checklist for RCTs (See Other Studies of Related Interest). The resulting instrument was then piloted on four studies by three investigators and a discussion was held to draw up a revised draft in the light of the pilot.

To provide a simple standardised expression of the quality of studies, groups of studies (RCTs, comparative studies, single prosthesis studies) as a whole were scored on six areas: clarity of study question and definition of outcome, description of prosthesis and fixation, description of study sample, adequacy of randomisation, duration and completeness of follow-up, and statistical and analytical considerations. Scores were the sum of positive "yes" scores as a proportion of possible positive scores (with "not applicable" removed from the denominator).

Not all studies were rated. Twelve RCTs, 12 (of 18) comparative observational studies and 15 observational studies of single prostheses were rated. In the main study, each paper was assessed by three raters (one orthopaedic surgeon, and two non-surgeon researchers working in the field of orthopaedics). Raters were drawn from a panel of three orthopaedic surgeons and four non-surgical researchers who have worked on outcomes of orthopaedic surgery. Thus each paper was rated by at least one expert in orthopaedic surgery and two experts in research methods. Papers were scored blind to others' ratings. A paper was given a final score on the basis of a majority "vote".

Data extraction
The authors do not state how the data were extracted for the review, or how many of the authors performed the data extraction.

For the meta-analysis, a standard set of possible confounding variables were extracted for each prosthesis series (length of follow-up, age of sample, proportions of patients in study samples who were female, proportions of patients with a diagnosis of rheumatoid arthritis or osteoarthritis).

Methods of synthesis
How were the studies combined?
Studies were synthesised narratively according to a hierarchy of evidence. Evidence from well-designed randomised controlled trials (RCTs) was considered to be strongest, followed by observational cohort studies with concurrent controls, followed by "other" observational studies.

Because the data from RCTs had produced so little clear evidence of the longer term performance of prostheses, is was decided to combine data from as many sources as possible into a form of meta-analysis.

All data (randomised and observational) were combined for any prosthesis for which at least five independent studies reporting revision surgical rates were obtained. The meta-analysis was termed "informal" because of the impossibility of controlling for numerous biases in the data and the poor quality of reporting of much of the evidence.

Revision rates for eligible prostheses were calculated, adjusted for person-years at risk. Data were also combined for meta-analysis for other outcomes (i.e. hip scores, global ratings of success, and proportion of patients pain-free). To test whether there were statistically significant differences between prostheses for relevant clinical outcomes, and whether any such differences were independent of possible confounding effects of the characteristics of study samples, analysis
of variance with covariates was performed with each study entered as a case. The dependent variables of these four analyses were, for each study: the rate of revision surgery adjusted for rate of follow-up, the mean Harris (or equivalent) follow-up score, the percentage of patients rated as "excellent" or "good", and the percentage of patients who were pain free.

Main effects were examined for prosthesis type and whether or not the prosthesis was fixed with cement. The mean age of patients in the sample and the percentage of: females patients in the sample, patients who had rheumatoid arthritis as a main diagnosis and patients who had osteoarthritis as a main diagnosis were used as covariates. The duration of follow-up was also used as a covariate on the analysis of the dependent variables other than the adjusted revision rate.

How were differences between studies investigated?
Tests for heterogeneity were not reported.

Results of the review

One hundred and ninety-one papers were included in the review, including 11 RCTs and 180 observational studies.

The informal meta-analysis involved 94 separate reports and papers, including 11 RCTs, 21 comparative observational studies and 62 reports of a single prosthesis.

Quality assessment: Ratings were given in terms of the number of studies which met various criteria as opposed to individual studies. As there were a number of criteria within each assessment item, these will not be reported here.

Overall, the evidence from RCTs in THR provided no clear evidence of the relative advantages of prostheses. Consequently, the next most robust source of evidence was also examined - observational studies with comparative data on prostheses.

Comparative observational studies: The two largest studies found that the Charnley prosthesis was favourable in both studies. The other 16 comparative studies were of a much smaller size and compared a diverse range of prostheses for different lengths of time. Eight of these included the Charnley and all except one study showed the Charnley as performing more favourably than comparison prostheses. No specific pairs of prostheses were compared sufficiently frequently to warrant combining studies. Observational data for single prostheses: These studies were reported in the meta-analysis, the results of which are described below.

Meta-analyses: When the results of all reports that included a revision rate were combined, ten prostheses met the criterion set for a meta-analysis that at least five independent studies should be available for a prosthesis to be included. Adjusted THR revision rates (revision rate per 100 person-years at risk) were calculated for each of the ten prostheses to take account of different lengths of observation. The most favourable revision rates were found for the Exeter (0.18; 95% CI: 0.14, 0.22), Lubinus (0.27; 95% CI: 0.24, 0.30) and Charnley (0.37; 95% CI: 0.35, 0.39) prostheses. Intermediate results were found for the Muller (0.68; 95% CI: 0.58, 0.78), McKee-Farrar ((0.98; 95% CI: 0.90, 1.06) and Stanmore (0.62; 95% CI: 0.53, 0.71) prostheses. The least favourable adjusted revision rates were observed for the Ring (2.04; 95% CI: 1.85, 2.23), Harris-Galante (1.40; 95% CI: 0.82, 1.98), PCA (1.31; 95% CI: 1.02, 1.60) and Charnley-Muller (1.10; 95% CI: 1.00, 1.20) prostheses.

Analysis of variance showed that there were significant differences between prostheses for the Harris hip score (F=5.42, df =4, p<0.01), and for the percentage of patients pain-free (F = 3.52, df =4, p<0.05). It was not reported which prostheses were significantly better than others.

Cost information

Costs and benefits of primary THR were assessed using Markov modelling, and calculation of costs per quality adjusted life year, with sensitivity analysis of the results. Outcomes data were taken from a prospective study of a series of patients followed up for 14 years after THR. Costs were estimated from cost-generating events for THR and unit costs from a single centre (Nuffield Orthopaedic Centre, Oxford).

Economic modelling indicated that to be cost-effective the following improvements in THR outcome and revision rates
would be needed:

1. For a new prosthesis costing three times more than the standard Charnley (i.e. typical cost of a new cementless prosthesis): greater than or equal to 35-44% improvement in patients aged 50-70 years; greater than or equal to 21-27% improvement in patients aged less than 50 years.

2. For a new prosthesis costing 1.5 times more than the standard Charnley (i.e. typical cost of a new cemented prosthesis): 9-12% improvement in patients aged 50-70 years; 6-7% improvement in patients aged less than 50 years.

From the available evidence, the extent of the improvement required of new and more expensive prostheses is partly implausible for older patients. However, the new cheaper prostheses may be cost-effective because the improvements required are more likely to be achievable.

Authors' conclusions
There is a striking paucity of clear and relevant evidence on which to make well-informed choices about prostheses for primary THR. Although the basic scientific innovation continues in relation to THR, the knowledge base to inform selection of prostheses is unlikely to improve in the foreseeable future.

Of prostheses commonly used in the NHS by far the greatest volume of evidence is available for the Charnley and on the basis of that evidence the Charnley appears to perform relatively well. However, the Charnley design has changed, and it is not clear how much of the evidence is relevant to the current design.

Of other prostheses commonly used in the NHS, positive evidence (but no data from RCTs) was found in support of the Exeter prosthesis, and some positive evidence was found for the Stanmore (for example, evidence that it performed as well as the Charnley in an RCT). Positive evidence for the Lubinus IP (less widely used in the NHS) was also found. The quality of evidence for other prostheses was either poor or non-existent. No substantial evidence could be found for cementless prostheses in terms of independent observation of results from five or more studies.

None of the analyses used in this review, such as meta-analysis of evidence, could overcome the fundamental weaknesses of the available evidence. The poor quality of evidence overall does not provide a basis clearly and authoritatively to identify prostheses that could be, or should not be, recommended for use by the NHS. However, it is clear that the more expensive the prosthesis, the more difficult it is to provide justification for selection on the basis of the current evidence. On the basis of the economic analysis it seems that the use of more expensive (i.e. cementless) prostheses is hard to justify on current evidence.

CRD commentary
The review focused on a well defined question. Inclusion and exclusion criteria were appropriate and sufficient details of the individual studies were provided.

Although the search involved thorough handsearching and searches of electronic databases, no attempt was made to identify unpublished literature, thus publication bias cannot be ruled out. Although a validity assessment was carried out, it was not applied to every study. The authors stated that this was because it became clear from an early stage that few strong inferences of clear advantages between prostheses were going to emerge from the evidence. Tests for heterogeneity were not performed before the studies were combined. The conclusions follow from the results.

Implications of the review for practice and research
In terms of future research, the authors state that a substantial proportion of the evidence on outcomes of THR comes from healthcare systems quite different from the NHS and recommend that the case for a UK register should be evaluated.

They state that to detect the small but important differences that may exist between prostheses such trials must be more adequately designed and powered than those carried out previously, and should involve multicentre participation and long term follow up. Economic modelling in this review indicates that such trials might identify differences in cost-
effectiveness between cemented prostheses.

Patient based outcomes provide relevant and feasible methods to conduct large multicentre studies. To obtain unbiased assessments of outcome, the focus should be on outcomes of concern to patients, particularly pain and function, and not solely on revision surgery.

**Funding**
NHS R&D Health Technology Assessment (HTA) Programme, project number 93/11/08.

**Bibliographic details**

**PubMedID**
10103353

**Original Paper URL**
http://www.hta.ac.uk/project.asp?PjtId=890

**Other publications of related interest**

**Indexing Status**
Subject indexing assigned by NLM

**MeSH**
Arthroplasty, Replacement, Hip /economics /instrumentation; Cost-Benefit Analysis; Female; Great Britain; Humans; Male; Models, Economic; Prognosis; Prosthesis Design /economics; Prosthesis Failure; Randomized Controlled Trials as Topic; Reoperation; Treatment Outcome

**AccessionNumber**
11999008394

**Date bibliographic record published**
31/05/2000

**Date abstract record published**
31/05/2000

**Record Status**
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.