Introduction

Total knee arthroplasty is one of the most common orthopaedic procedures performed. In 2001 171,335 primary knee replacements and 16,895 revisions were performed.¹ Throughout this report we use the term total knee arthroplasty (TKA) in lieu of total knee replacement because the abbreviation for the latter may be readily confused with total knee revision. Because these procedures are elective and expensive (Medicare paid approximately $3.2 billion in 2000 for hip and knee joint replacements) and because the prevalence of arthritis is expected to grow substantially as the population ages,², ³ these procedures are likely to come under increasing scrutiny.

Previous reports suggest that TKAs improve functional status, relieve pain, and result in relatively low perioperative morbidity.⁴ However, based on conclusions from consensus panels or surveys of health care providers, there is considerable disagreement about the indications for the procedure; that is, which patients are most likely to benefit from TKA and, conversely, in which patients is TKA contraindicated or of low value.⁵-¹⁰ This evidence report, which was commissioned for an NIH Consensus Development Conference on Total Knee Replacement, was designed to systematically review, analyze, and discuss empirical data on Total Knee Replacement, to help inform the deliberations of the Consensus Panel.

In collaboration with the Office of Medical Applications of Research (OMAR), the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), and the TKR Planning Committee, the Agency for Healthcare Research and Quality (AHRQ) defined the work to be performed for a comprehensive evidence report on the indications for primary TKR and revisions. The scope of the project specified that it address the following key questions regarding total knee arthroplasty:

1. What are the current indications for, and outcomes from, primary total knee replacement?
2. How do specific characteristics of the patient, material and design of the prosthesis, and surgical factors, affect the short-term and long-term outcomes of primary total knee replacement?
3. Are there important perioperative interventions that influence outcomes?
4. What are the indications, approaches, and outcomes for revision total knee replacement?
5. What factors explain disparities in the utilization of total knee replacement in different populations?
6. What are the directions for future research?

Methods

Literature Review and Meta-Analysis

To address the first key question about the indications and outcomes of TKAs, the National Library of Medicine staff conducted a systematic literature review from 1995 to April 2003.

The titles and abstracts of the resulting 3,519 references were then screened, using our inclusion criteria (primary total knee arthroplasty studies; more than 100 knees per study; baseline data and
post-op outcomes data provided; experimental or quasi-experimental study design, English language, tricompartment).

All articles that appeared to meet the screening criteria were abstracted by trained abstractors. Of the original results, 611 references either met the inclusion criteria or needed further screening of the full article to determine if they met inclusion. Of these, 62 studies reported pre- and post-TKA functional data using at least one of the four established measures we relied on (Knee Society score, Hospital for Special Surgery score, WOMAC, or SF-36).11-74 All but 15 studies were conducted in the United States or Canada.

One of the problems that made summarizing this area difficult was the inconsistent use of patients and knees as the unit of analysis. The reason for this practice is related to the performance of bilateral procedures, either simultaneously or sequentially, but the result is an inconsistent count. Some studies provide both units; some only one. For some types of analysis knees seem like the best measure, but for many (including function and demographics) the data apply more reasonably to patients. Wherever feasible, we present the analysis using both patients and knees.

To address key question 2 regarding prosthesis material/design or surgical factors we analyzed studies that fell within our original search parameters. We attempted to classify a study as primarily addressing either the use of a specific type of prosthesis or testing a specific surgical procedure or technique. Specific characteristics of the patient that may affect outcomes are addressed as noted in the main analyses and reported under “Outcomes of Primary TKA.”

We limited our analysis of evidence to assess important perioperative interventions that influence outcomes (key question 3) to studies published since 1994. All were randomized controlled studies with the exception of one large cohort study. We categorized interventions as: prophylaxis for postoperative deep venous thrombosis/pulmonary embolism or infection. Several other procedures involved non-surgical elements of care.

We conducted a meta-analysis on the functional outcomes data. Because the data at baseline and followup was not consistent, we selected the model with random effects to simplify the interpretation. Because we did not have precise information from all studies, we treated each pre and post pair as if they were separate data sets.

In addressing key question 4, about the outcomes of TKA revisions, we relied heavily on the meta-analysis recently completed by one of the principals, which covered the period from 1966 through 2000.75 To update this meta-analysis, a literature search was undertaken to assess the status of the literature relating to revision total knee arthroplasty after (and including) the year 2000. The literature search was done via PubMed® using a strategy based on the search described in the previously published meta-analysis; 14 new studies were uncovered.76-89

To answer key question 5, about the evidence for access differences (disparities in utilization) related to race and gender, we conducted a literature search via PubMed from 1995 to 2003. This search resulted in 176 references. Titles and abstracts of the references were reviewed, and 23 met preliminary inclusion criteria (primary total knee arthroplasty studies; more than 100 knees per study; gender/racial data provided; experimental or quasi-experimental design). Of these, three met inclusion criteria for analysis.90-95 Additionally, reference lists from the above articles, and from articles recommended by colleagues, were searched. Three additional articles were found and included in the analysis.

Results

Outcomes of Primary TKAs

On average the patients were approximately 70 years of age and very few of them were over age 85; about two-thirds were female; about one-third were considered obese (using a criterion of a BMI of 30 or higher). Nearly 90 percent of patients had osteoarthritis. We did not specifically address bilateral TKAs but did separate analyses by numbers of knees and numbers of patients.

The most commonly used functional measures were the Knee Society score (KS)96 and the Hospital for Special Surgery scale (HSS).97, 98 The WOMAC (Western Ontario and McMaster Universities) Arthritis Scale has only been used since 1991. The physical function component of the SF-36 is a generic functional outcomes measure, not specific to knees.

The KS is associated with longer followup periods, perhaps because it was in use earlier. For example, weighting for baseline patients the mean followup for KS and HSS is 66 and 67 months, compared to 45 months for WOMAC. However, weighting for baseline knees, KS has a mean followup of 90 months and WOMAC is 68 months, but HSS is only 61 months. The longest mean followup time was 90 months (KS scores weighted for baseline knees), well less than the 10 years that has been suggested in order to evaluate long term functional results. Only ten studies had a followup time of at least 10 years.

Some information on attrition rate was reported for 49 studies. Of these the median percentage of subjects lost to followup was 2 percent, the range was 0-28 percent. If death is added to the definition, the range increases to 0-56 percent with a median of 12 percent.

Although there is no formal basis for translating the size of the scores, the generally accepted rule of thumb for the KS and HSS scales is that a score of less than 60 is considered poor; 60-69 represents a fair result; 70-84 is considered a good result; 85-100 is considered an excellent result.
The functional scores after TKA are consistently higher. The mean effect size (defined as the number of standard deviations of change from baseline scores) for the HSS studies is 3.91 for those with followup up to 2 years, 3.01 for those 2-5 years, and 2.97 for those studies with more than 5 years of followup. For the studies using KS the mean effect size is 2.35 for those 0-2 years, 2.73 for those 2-5 years, and 2.67 for those 5+ years. For WOMAC studies the mean effect size for 0-2 years of followup is 1.62. The more generic SF-36 scores had the smallest mean effect size; for the studies with 0-2 years of followup it was 1.27.

When the unit of analysis was numbers of knees operated on, the perioperative complication rate (defined as occurring within 6 months of the TKA) was 5.4 percent; when the denominator was numbers of patients, the rate was 7.6 percent. The revision rate through 5 or more years is 2.0 percent of knees and 2.1 percent of patients.

We differentiated “indications for TKA” from “correlates or factors related to outcomes.” The former addresses what factors are needed to warrant a TKA (or conversely, what factors are contraindications to TKA either because the procedure is ineffective, unnecessary, or places the patient at unacceptably high perioperative risk); whereas the latter addresses whether outcomes vary according to the clinical or demographic factors. To address indicators would require a design that compared the outcomes of persons with the potential indicator with and without surgical treatment. However, it is possible to examine the potential for contraindications by examining only those who receive arthroplasties.

The number of studies that employed any analytic technique examining the functional outcome in terms of at least one independent variable of interest was limited. Only 12 of the 69 studies used any analysis that directly assessed the relationship of these patient variables to a change in functional status. Age, obesity, or gender do not seem to be significantly correlated with TKA outcomes. Whether outcomes vary according to arthritis type is unclear. Patients with rheumatoid arthritis seem to show more improvement than those with osteoarthritis but they have lower level of function preoperatively and few studies adjust for other risk factors such as obesity.

### Types of Prostheses and/or Surgical Factors

Although the sampling approach was not specifically designed to search for all outcomes associated with using different types of prostheses or different surgical approaches, we did analyze the studies that fell within the search parameters. In some cases it was difficult to classify a study as primarily addressing either the use of a specific type of prosthesis or testing a specific surgical procedure or technique. Several studies reported prostheses that were used in specific types of procedures. A number of the studies of prostheses were case series that reported generally good results. A few tested the use of a prosthesis with a specific group of patients. The studies of procedures were a mixture of case studies and comparative studies.

### Perioperative Interventions

TKA studies assessing prophylaxis for postoperative deep venous thrombosis (DVT) or infection were identified by searching the 611 references meeting and not meeting inclusion criteria. The Cochrane Library was also searched back to 1994. The investigators decided a priori to include only randomized controlled trials (RCTs) with the exception of large cohort studies. Fourteen studies were identified and extracted; nine DVT, three infection, and two tourniquet studies. All included studies were randomized controlled trials with the exception of one large cohort study. One trial was identified through the Cochrane Library.

Several other procedures, which involved primarily non-surgical elements of care, were also described. Three of these addressed the use of continuous passive motion as a rehabilitative approach; two studies were positive. The other two studies tested different clinical pathways and showed mixed results.

The review of randomized trials addressing prevention of venous thrombosis and pulmonary embolus uncovered several studies that tested various approaches to anticoagulation and other preventive techniques. Two studies suggest that compression ultrasonography is not justified. Two find drug therapy better than mechanical approaches. Several studies compared anticoagulant drugs and drug regimens.

Three randomized trials addressed infection prevention. Each compared alternative antibiotic regimens. Two randomized trials tested the use of tourniquets in performing TKAs. One concluded tourniquets were safe and the other that they did not reduce surgical complications.

### Access

Six studies addressed TKA-related access issues according to race or gender. Several of these studies included both hip and knee replacement surgery. The conclusions with regard to the differential treatment of women are mixed, but the preponderance of evidence suggests that women are almost twice as likely to undergo a TKA as men. The evidence regarding non-white groups is quite consistent. Non-whites receive TKAs about half as often as whites. Most of these analyses report simply the rate at which the procedures were performed, with no attention to the actual size or nature of the population at risk. The argument that the higher rates of TKAs in women may be due to the higher prevalence of arthritis among women does not apply to the study by Wilson, which examined only persons with arthritis. However, it is possible that the severity or type of arthritis (OA vs. RA) varied.
Conversely, the lower rates of TKAs among blacks occurred despite a higher prevalence of osteoarthritis in this group, suggesting that the prevalence of OA was not a mitigating factor. Most of the studies that address access relied on large administrative data sets, which did not contain detailed clinical data on which to base the indications for knee surgery.

**Total Knee Arthroplasty Revisions**

Like all biomedical devices, total knee replacements can fail over time. The primary factors believed to cause TKA failures (and thus require consideration for TKA revision-TKAR) include trauma, chronic progressive joint disease, prosthetic loosening, and infection of the prosthetic joint. Coincident with the increased incidence of primary TKA, there has also been an increase in the number of TKAR procedures.

The primary assessment of the outcomes of TKAR for this report is derived from a systematic review of the literature published through 2000 that was done by one of the principals. It used a global knee score (GKS) measure that included the HSS and the KS, each assessed along the same range from 0-100.

There was a large improvement in GKS scores following TKAR that was both statistically and clinically significant. The preoperative combined mean KS score was 35.4 (95% CI 30.7-39.9). There was an increase of 30.8 (95% CI 26.6-35.0) points to 66.2 (95% CI 61.8-70.2) points postoperatively (p <0.0001). The preoperative mean HSS score was 51.5 (95% percent CI 48.9-54.1). There was an increase of 28.3 (95% CI 25.3-31.2) points to 79.8 (95% CI 76.4-83.1) points postoperatively (p < 0.0001).

Although there was no difference in age or gender between the multiple and single knee reports, there was a significant difference in preoperative HSS. Patients undergoing “multiple knee TKAR” had lower preoperative scores (multiple knee HSS = 49.5, 95% CI 45.9-53.2; single knee = 54.5, 95% CI 51.4-57.5; p <0.1). These results suggest that the multiple knee cohorts may have more severe disease then subjects evaluated in single knee TKAR studies. In contrast, the preoperative combined mean KS score in the multiple knees group was higher (77.0, 95% CI 64.2-89.8) than the single knee group (59.85, 95% CI 45.2-79.8), p >0.1. There was no difference in the pooled change in either the KS or HSS from pre- and postoperative scores when comparing subjects undergoing multiple vs. single TKAR.

Forty-four of 46 (95.7 percent) cohorts reported complication data on 1683 subjects who incurred 443 complications (26.3 percent). It was not possible to determine which or how many complications occurred in any given patient or patient subset. There were a total of 217 knee complications in 1,683 subjects necessitating re-revision (12.9 percent).

**Discussion**

The basic observations can be summarized as follows:

- Both TKA and TKAR are associated with improved function. The strongest evidence exists over a followup period of up to two years, but the studies that extend to 5 and even 10 years of followup show positive results as well.
- The average age of patients undergoing TKA in these reports was 70 years with few over age 85. Two-thirds were female, one third were considered obese, and nearly 90 percent had osteoarthritis. No studies provided data on racial/ethnic status.
- The mean effect size (expressed as numbers of standard deviations) is considered large in magnitude and varies from 1.6 to 3.9 depending on the functional measure used and the duration of followup.
- There is no evidence that age, gender, or obesity is a strong predictor of functional outcomes.
- Patients with rheumatoid arthritis show more improvement than those with osteoarthritis, but this may be related to their poorer functional scores at the time of treatment and hence the potential for more improvement.
- The revision rate through five or more years is 2.0 percent of knees and 2.1 percent of patients.
- Perioperative complications as defined by the investigator occurred in 5.4 percent of patients and 7.6 percent of knees. The vast majority were “knee related” or deep venous thrombosis. There were only 8 cardiovascular or pulmonary complications reported among nearly 6,000 patients suggesting that these adverse effects were not fully addressed in this literature.
- There is reason to suspect selection effects in both the type of patients referred for TKA and those being reported in the literature as well as the attrition on followup. Hence, these findings must be interpreted with caution as the basis for clinical practice.
- TKA revisions show a similarly positive functional effect (with the same design limitations).

These conclusions are tempered by the limitations of the designs of many studies included in the analysis. Although osteoarthritis does not seem to be a predictor of outcomes, the results seem to be somewhat better for rheumatoid arthritis, but few of these studies simultaneously controlled for other aspects of the patients.

Overall, the scientific quality of the current evidence is weak. Only a handful of studies employed any form of multivariate analysis. The outcomes of orthopaedic surgery, like most other treatments, are the results of the treatments interacting with the characteristics of the patients. Real understanding will come about only when the analytic techniques can address both sets of variables simultaneously. The analyses that come from such studies will need to employ sophisticated statistical methods,

4
which can examine the effects of the patient characteristics on the outcomes of interest. Orthopaedic outcomes research has made considerable strides in the last decade. Much greater attention is now paid to using established outcomes measures. The next step in this progress is to employ more sophisticated research designs that incorporate patient characteristics into the analysis.

Because orthopaedic research will likely rely heavily on observational studies instead of RCTs, it will be important to use more robust methods of study design/analysis. Particular attention must be paid to ensuring that the cohorts remain intact. Greater efforts must be made to collect outcomes information on all participants, not just those who appear for followup visits. A substantial proportion of the studies reviewed were based on retrospective reviews of clinical records. Strong levels of evidence will require prospective designs that emphasize followup.

**Research Recommendations**

The current state of empirical work does not provide a strong basis for making clinical recommendations regarding indications for outcomes from TKA. As pressures mount for more discrimination in identifying subjects for elective surgery, better information will be needed. The ideal study design to answer questions about indications for surgery remains a randomized trial in which persons with advanced arthritis (or other potential joint problems) are randomly assigned to medical management or joint replacement. However, given the enthusiasm for joint replacement and the generally positive effects on function, it might be difficult to recruit subjects for such RCTs, even without the prospect of sham surgery. Thus, a major component of research into the effectiveness of joint replacement and the patient characteristics associated with better outcomes will be well done observational studies.

More attention needs to be paid to the independent variables (or risk factors) associated with clinically relevant outcomes. Adequate research designs will require the use of multivariate analysis. To generate the sample size needed for multivariate analysis, these studies will likely have to be cooperative ventures. Such a plan would also broaden their representation. They will require systematic collection of data on potential indicators and risk factors and active followup to maintain the cohort, even when the patients do not return for scheduled followup clinical visits.

Although many questions remain unanswered, a few major issues need to be addressed first:

- How long will the functional benefits of TKA last and when will revision surgery likely be needed?
- How much do outcomes vary by patient characteristics and surgical factors, including volume of these procedures performed? Is the volume effect related to the surgeon or the medical center? There is strong belief that volume of surgery in a center, and perhaps experience of the surgeon, is related to better outcomes, but the strength of this relationship has not yet been well established and may be artifactual.

Many of the basic questions posed for this review remain unanswered, such as:

- What are the effects of patient characteristics on outcomes?
- What is the effect of surgical technique on outcomes?
- How does the choice of prosthesis affect outcomes?
- What is the role of rehabilitation in affecting outcomes?

**Availability of the Full Report**

The full evidence report from which this summary was taken was prepared for the Agency for Healthcare Research and Quality (AHRQ) by the Minnesota Evidence-based Practice Center, Minneapolis, MN, under Contract No. 290-02-0009. It is expected to be available in December 2003. At that time, printed copies may be obtained free of charge from the AHRQ Publications Clearinghouse by calling 800-358-9295. Requesters should ask for Evidence Report/Technology Assessment No. 86, *Total Knee Replacement*. In addition, Internet users will be able to access the report and this summary online through AHRQ's Web site at www.ahrq.gov.

**References**


