Chapter 2. Methods and Analytic Framework

This review has three major components, which correspond to the questions posed in the charter. The major effort was directed at examining the indications for (or at least the outcomes of) primary TKA. The second component is a report of a meta-analysis of total knee revisions, which has already been published,[21] and an update of the literature since that work was completed to be sure no new developments had affected the initial conclusions. The third component was a review of the literature on access to care, especially the effects of gender and age.

The principal analytic framework for the first review (the outcomes of primary TKA) was based on the fundamental principles of outcomes research.[22] The underlying model can be briefly expressed as:

Outcomes = f(baseline status, clinical factors, demographic factors, treatment)

In general, the goal of outcomes research is to identify the effect of treatment on outcomes, adjusting for the other factors that might affect outcomes. In this case, however, we use the same model to address the predictive role of various patient characteristics on outcomes when all are treated similarly. Interpreting this relationship is somewhat more complex because factors associated with good outcomes are not necessarily indications for treatment. For example, a person with no problems may have a very good outcome, but one would not want to treat such a patient. The true test of an indication for surgery is a factor that gets worse without treatment and better with it. In effect, one would want to randomly assign patients with the specified condition to receive either TKA or medical management and then compare the clinical course with and without the treatment under study. Those factors that produced the greatest difference associated with treatment would be the strongest indicators for such treatment.

Where randomized clinical trials are available, many of the relevant confounding clinical and demographic factors can be assumed to be randomly distributed, or they may be controlled by elements of the study design that specific inclusion and exclusion criteria, and thus any differences between two groups can likely be attributed to the intervention. However, in the absence of RCTs, as is the case in most of the orthopaedic literature, strong quasi-experimental designs are needed, wherein multivariate analysis is employed to isolate the effect of treatment and address issues related to selection bias. The literature review was thus initially targeted at identifying those studies that had at least the rudiments of such a design. However, given the studies uncovered, we were forced to revise our criteria to assess a broader array of studies that provided at least some baseline and followup information.

Based on consultations with the technical expert panel (members are shown as Appendix B) and discussion with OMAR, AHRQ, and the Chair of the Consensus Panel, we determined that functional measures would be used as the primary outcome measures. We identified several demographic and clinical variables of primary interest: age, gender, baseline status (with regard to pain and function), arthritis type, and body mass index/obesity. The analysis for demographic factor effects, which correspond to the question about access, was conducted separately.

TKA Indicators

The literature search strategy for clinical predictors of TKAs was developed in consultation with the National Library of Medicine, which conducted the search. The literature search was done using a combination of MeSH headings, keywords, and publication types shown in Appendix C.

The search was limited to studies published between 1995 and April 2003. This start date was chosen because a previous review was published in 1994.[5] Animal studies were excluded, as were non-English language references and references on unicompartmental (unicondylar) knee replacement. Although unicondylar knee replacements (UKR) share many features with total knee replacement (tricompartment), these studies were excluded from our search because UKRs have 1) more specific indication ie, unicompartmental tibio-femoral arthritis with minimal involvement of the patello-femoral and 2) different patient demographics, primarily male population, low activity, minimal deformity, and good range of motion. Additionally, indications for UKRs appear to be in a transition phase. Surgeons have only recently gained experience with this reportedly less invasive procedure. Thus it was felt too early to adequately assess outcomes.

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 standardized symptom scale outcomes data provided; experimental or quasi-experimental study design).

All articles that appeared to meet the screening criteria were abstracted by trained abstractors. Extracted data included study and patient characteristics, baseline and followup symptom scale scores, revision rates, and perioperative complications as defined by the authors and occurring within six months of surgery. This workforce included medical students, two review staff, an orthopaedics fellow and several volunteer orthopaedic surgeons. A 10 percent subsample of all the abstracts was reviewed independently by a second abstractor to assure consistency. All of the studies that met the minimal criterion of having pre- and post-surgery data were re-reviewed independently by all three of the study principals.

The abstracting form (see Appendix D) included a long list of potential prognostic factors, developed with the assistance of our technical advisory committee. These included co-morbidities, x-ray evidence of joint destruction, bone loss, extensor mechanism integrity, pre-operative range of motion, alignment, tibio-femoral angle, and ligament integrity, as well as the characteristics of the operating surgeon, such as volume and experience.

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. The reasons for exclusions are shown in Figure 1, which traces the flow of the articles retained. 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). All but 15 studies were conducted in the US 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.

We conducted a meta-analysis on the functional outcomes data. Meta-analysis methodology assumes that to estimate the combined effect we compute the weighted mean of the results observed in different studies. In the simplest approach weights are based on the sample size but more sophisticated methods account for the precision of the studies and thus adjust for different standard deviations. The effects in this meta-analysis were normalized by dividing to combined standard deviation of two (baseline and followup) measures. Therefore the statistical results of the meta-analysis are expressed in the units of standard deviation and reported as an "effect size." An effect size greater than 1 SD is considered to be large in magnitude. An additional benefit of this approach is that various effects obtain the same measurement scale and therefore can be compared. In modeling the effects we could use either fixed or random effect models. Because the data at baseline and followup was not consistent, we selected the model with random effects to simplify the interpretation. This model assumes that all studies come from a common population. That is, if the sample size in each study were infinite, then the effect size in all studies would be identical and the standard error of the estimate would approach zero. Because we did not have precise information from all studies, we treated each pre- and post-pair as if they were separate data sets. This is a conservative approach. An analysis using pairs would have produced even more dramatic results. All calculations were implemented using the trial version of the Comprehensive Meta Analysis™ software.[23]

TKA Access

The literature search was done via PubMed using the combination of MeSH headings and keywords shown in Appendix C.

This search resulted in 176 references. Titles and abstracts of the references were reviewed, and 153 did not meet inclusion criteria (primary total knee arthroplasty studies; more than 100 knees per study; gender/racial data provided; experimental or quasi-experimental design, English language). Articles were pulled for the remaining 23 references, and, of those, three met inclusion criteria for analysis. 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 (total of six studies).

TKA Revisions

The bulk of this analysis relied on a meta-analysis recently completed by one of the principals, which covered the period from 1966 through 2000. 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.[21]

The search consisted of the combination of MeSH headings and keywords shown in Appendix C.

The original search for articles for the total knee revision meta-analysis resulted in 2,780 references. After titles and abstracts were reviewed, 2,551 did not meet the inclusion criteria of

revision knee arthroplasty studies, more than five patients per study, report of any post-operative outcomes, and use of a global knee rating scale. Articles were pulled for the remaining 229 references. In the end, 58 articles with a total of 1,965 patients met the initial inclusion criteria. Forty-two articles comprising 45 unique patient cohorts and a total of 1,515 patients had sufficient global knee score data for analysis and were used in the meta-analyses. (Descriptive tables for these studies are shown as part of the original paper reproduced in Appendix E)

The meta-analyses of global knee scores were undertaken using a fixed effects model with the assumption that the variances of each individual measurement were identical across studies. This assumption was necessary because data on variances were not provided in most studies. The variance of the overall estimate was derived under this model using the between-study variability, yielding a 95 percent confidence interval for each overall estimate. A weighted average of the values in each study based on sample size at followup was used.

The updated search was limited to articles published from 2001-2003. This search resulted in 229 references. Titles and abstracts of the references were reviewed, and 168 did not meet inclusion criteria (revision knee arthroplasty studies; more than five patients per study; report of any post-operative outcomes; use of a global knee rating scale). Articles were pulled for the remaining 61 references, and, of those, 14 met inclusion criteria for analysis.