Chapter 4. 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 five and even ten 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% had osteoarthritis. No studies provided data on racial/ethnic status.
- The mean effect size (expressed as numbers of standard deviations) is considered to be large in magnitude and varies from 1.6 to 3.9 depending on the functional measure used and the duration of followup. However, these results are based on simple pre/post designs with no blinding and large attrition rates.
- There is no evidence that age, gender, or obesity is a strong predictor of functional outcomes, but the extremes of age and obesity were not actively tested.
- Patients with rheumatoid arthritis show more improvement than those with osteoarthritis, but this may be related to their poorer functional scores (or other factors) at the time of treatment and hence the potential for more improvement.
- The revision rate through five or more years is 2.0% of knees and 2.1% of patients.
- Complications were defined by each investigator and occurred in 5.4% of patients and 7.6% of knees. The vast majority were "knee related" or deep venous thrombosis. Only eight cardiovascular or pulmonary complications were 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 the choice of patients and 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 many of the designs of the 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.

The original goal of this analysis was to identify indications for TKA. To do so, we would need to review studies that compared the outcomes of persons who did and did not receive the surgery. Instead the literature was limited to studies of the outcomes of the surgery performed. If well done, this database would allow conclusions only about the effect of variables on the outcomes of surgery, not on the relative benefit of the surgery for such individuals. (There would always remain the potential for "floor" and "ceiling" effects because some patients may simply be judged too sick or too well, too young or too old to be considered candidates.)

We had initially constructed a much longer list of potential factors that we had hoped would be examined in the search for prognostic features. These included co-morbidities, x-ray evidence of joint damage, bone destruction, extensor mechanism integrity, pre-operative range of motion, alignment, tibio-femoral angle, and ligament integrity. Although these were occasionally mentioned, they were not systematically reported.

The effect of hospital and orthopedic surgeon volume on complication rates and functional outcomes has been evaluated in at least two studies. Using Medicare claims data from 1985-1990 Norton and colleagues found no benefit (in terms of lower complication rates from performing more primary TKA until at least 40 operations are performed each year and there was no further benefit of performing more once 80 TKA are being performed.[128] Heck and colleagues followed an observational cohort of 291 patients with osteoarthritis undergoing TKA for at least two years and found that the maximal improvement in the physical composite score of the SF-36 was seen in patients who had their surgery performed at institutions that performed greater than 50 knee surgeries and by surgeons who performed greater than 20 TKA per year.[72] Additionally, there was a lower likelihood of complications among these higher volume institutions and surgeons.

It is possible that our results might be change if we used a different series of study inclusion filters. For example, we only included studies if they reported at least 100 knees, were written in English, and provided pre- and post-TKS functional data using at least one of the four established measurement scales. We also excluded unicompartmental procedures. We also could not assess whether our results might be affected by potentially varying patterns of referral or access of patients to orthopaedic surgeons. For example, it is likely that primary physicians may vary in their threshold (filters) for referring a given patient for TKA and/or orthopaedic surgeons have different threshold (filters) for offering TKA. Our findings are limited to the conclusions based upon published results of patients receiving TKA. Therefore, it is not possible for a particular patient or provider who is making a decision regarding TKA to directly apply these outcomes to their situations. However, compared to the findings by Callahan and colleagues reported in 1994, subjects had similarly large improvements in symptoms and function, lower rates of complications and revisions. This may reflect differences in patient populations, reporting of outcomes or improvements/refinements in the surgical procedure.

Although there is recurring evidence that total knee arthroplasties improve function and alleviate pain, much less is known about what types of patients are most likely to benefit from this surgery. As the pressure for more informed decisions grows, this type of information will be greatly needed.[128] The search for evidence about the indications for TKA was frustrating. The literature is full of articles that compare different procedures and prostheses, but relatively little attention is paid to the characteristics of the patients. (Perhaps, not coincidentally, many of these studies are supported by manufacturers.) Typically authors describe the sample under study and then ignore these characteristics in their analyses.

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, 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 or outcomes from TKA. As pressures mount for more discrimination in identifying subjects for elective surgery, better information will be needed. The traditional approach in orthopaedics of reporting small scale case series that examine the outcomes of a specific innovation must give way to larger, more planful studies that deliberately address the areas of interest.

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. (It would be unlikely to include some provisions for sham surgery as was done with joint arthroscopic surgery.)[129] No single study could be used to test all the variations in patient characteristics and surgical techniques. 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.

Historically much of the work in joint surgery research has gone into developing outcomes measures, but at this point, more attention needs to be paid to the independent variables than to the dependent ones. It appears that the results are robust enough to be detected by any of the major outcomes measures. The second concern is to employ designs that allow for multivariate analysis, which can assess the effects of several independent variables simultaneously. This approach was encountered only rarely in our review.

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 and in whom will revision surgery likely be needed? Are there patient characteristics associated with poor outcomes such that these patients should be excluded from consideration or assigned a lower priority?
- How can one trade off the benefit of surgery against the risk of needing a revision?

- How much do outcomes vary by patient characteristics and surgical factors, including type of prosthesis, 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 been well established and may be artifactual.

Lessons Learned

Ideally, databases can be utilized to characterize practice patterns, identify and investigate prostheses failure, establish benchmarks, develop guidelines, and quantify present and future healthcare resource utilization, but incomplete data can create serious problems The literature review performed highlights some of the pitfalls that can occur in surgeon based data collection.

Much of the data falls short of expected standards of quality and execution.[130-135] Useful studies need: 1) clear objectives and goals; 2) a valid protocol design; 3) clear inclusion and exclusion criteria; 4) a study sample that is representative of the universal population; 5) a comprehensive collection of variables necessary to answer the project objective(s); 6) mechanisms implemented to track patients and assure complete followup; 7) mechanisms implemented to ensure high data integrity; 8) blinding of data collection personnel; and 9) a method to rectify methodological problems (such as attrition bias).

At the conception of patient and surgeon based knee arthroplasty studies it is critical to define the purpose behind the data collection effort and let this guide the development process. To help in addressing these issues it is important to ask:

- *What* questions (clinical, administrative, quality outcomes) are to be answered by the study?
- *Who* will be the consumers of this data or information—patients, surgeons, or third parties? *Who* will be held responsible for ensuring the study goals are met?
- *What* protocol design would best answer the study's objectives?
- *What* are the dependent (outcome) and independent (risk factor) variables?
- *Where* should the data be collected, i.e. patients' homes, surgeons' offices, mail packages etc? Where should the data be entered and stored?
- *Who* will collect the data?
- *When* should followup data be obtained?
- *How* will the data be used to impact clinical care?
- *How* will patient confidentiality and safety be protected? Will the data be used for quality improvement, general research or physician accountability?

Many of the studies lacked critical features of a well designed time-series protocol: a) there was no clear process in place to recruit and follow patients; b) there was extensive loss on followup; c) not every study developed a detailed set of inclusion and exclusion criteria. These measures would have ensured a more homogeneous cohort that would allow better comparisons. As a consequence, the cohorts reported were probably not representative of the universal knee joint replacement population.

Pertinent independent variables need to be identified, collected, and used in the analysis. For example, no studies addressed characteristics of the surgeons performing the procedures. Deriving a conceptual model that contains the variables that must be collected to answer the

objectives and delineating the interactions between these variables not only averts important variable omissions but also helps in developing aims and forming an analytic plan.[22]

Attrition creates potential bias. The poor followup response rate resulted from insufficient monitoring and tracking. Technical solutions can be employed to achieve this goal. The field needs to define a consistent set of postoperative followup points. What is more critical, a large number of subjects did not return for followup at all rendering the analysis and interpretation of the data difficult. Followup cannot depend on patients returning for care; it must be proactive. When a subject is no longer available or able to respond, there must be mechanisms in place to approach proxy respondents identified as the person to contact on the original hospital/contact face sheet. Based on our experience, tracking some of the patients and establishing the best proxy will take some active detective work, but it can be done. No doubt permission from the appropriate legal and governmental authorities will be needed to accomplish this task. Obtaining permission in advance can overcome many of the growing number of legal obstacles (HIPAA and others) in gaining access to patients and governmental databases (Social Security, IRS, etc.) in order to complete followup information.

Followup periods of at least five to ten years are considered necessary to allow time to test the durability of prostheses. Although some loss of sample is likely in that time frame, it is important to be able to test the effect of that attrition on the findings. In these circumstances, where decline in function is expected, intention-to-treat is not the correct technique. Statistical models will need to compensate for the selective loss to followup.

Utilizing tracking techniques as outlined by Smith and Watts[136] and carrying out traces such as the Department of Motor Vehicles traces, voter registration traces, and so on, to locate orthopaedic cases is helpful but inefficient.[137] These tracking methods are not appropriate for real time studies. They are more appropriate for collecting long-term data such as ten-year followup data, but dealing with short-term data problems needs a more proactive, pre-planned strategy. Alternative potential sources for locating patients need to be built into the initial enrollment process.

As many hospitals and clinics convert to Electronic Medical Records (EMR) it is crucial that databases be able to interact with these records. Software development to establish a common standard for collecting and annotating joint replacement followup data is critical to making this data collection process efficient. Incorporation of outcomes instruments into these products would further enhance data collection efforts and the amount of useful information collected.[116] This would also assist surgeons and physicians in completing necessary forms and submitting data. This allows for immediate submission, review of information, and can minimize errors in data entry.

To be able to test the characteristics of surgeons and hospitals, the database must be set up to identify surgeons and hospitals, in order to estimate the fraction of variance explained by these characteristics. Appropriate checks must be in place to ensure participating surgeons of confidentiality and protection from any negative impact. All of these factors will serve as risk-adjustors in analyzing time trend of functional outcomes and rate of re-operation (primary outcome measure of the database).

Feedback loops need to be set up to affect not only the data collection process (as outlined above) but the consumers of this information (patient, surgeon, hospital, and third party payers). These feedback loops should improve quality of care and streamline healthcare expenditures.

There must be obvious and compelling reasons for physicians to participate. The benefits to the orthopaedic surgeon must be clear and strategies of linking participation to getting paid or becoming credentialed or recertified must be explored.

Research Agenda

A large number of questions remain to be answered. Table 24 proposes a preliminary list. These questions illustrate the range of unanswered questions. They obviously cannot all be addressed in a single study. Indeed, it will be difficult to disentangle the effects of different aspects of treatment. For example, rehabilitation can interact with surgical technique; and both can interact with patients' characteristics.