A Modest Proposal for a Clinical Trial on Single-Bundle Versus Double-Bundle Anterior Cruciate Ligament Reconstruction

Having served as the Statistical Associate Editor of *Arthroscopy* since 2008,1 and as lead author of this editorial, I have had the opportunity to observe patterns in manuscripts that have been submitted and reviewed. After all, that is one of the things that statisticians are trained to do—to try and determine whether there are “signals” that are found consistently within the information that we are presented.

Reflecting on this and in consultation with the editors, we have come to the conclusion that there is a pattern that seems all too regular in the field of medicine, certainly in the field of orthopaedic surgery, and that extends well beyond the manuscripts we review for this journal of arthroscopic and related surgery.

Although a number of well-written manuscripts have been submitted (and often published) trying to compare anatomic single- and double-bundle anterior cruciate ligament (ACL) reconstruction procedures,2,3 we have noticed that these studies frequently reach conclusions that are unconvincing if not inconclusive because of limitations in study design.

Limitations often cited include inadequate time to follow-up; insufficient outcome measures; unsatisfactory outcome assessment; heterogeneous surgical indications, techniques, or rehabilitation; failure to compare similar cohorts; or deficient number of study subjects. As evidence, presented in this very issue, a meta-analysis of “Single-Bundle Versus Double-Bundle Reconstruction for Anterior Cruciate Ligament Rupture” by van Eck, Kopf, Irrgang, Blankevoort, Bhandari, Fu, and Poolman4 states, as the ultimate conclusion, “The majority of the included studies had a least one major limitation in study design that decreased the quality of the study.”

In the past, we have addressed these problems with a focus on defining bias, e.g., transfer bias as a result of inadequate time to follow-up, reporting bias as a result of insufficient outcome measures, recording bias as a result of unsatisfactory outcome assessment, performance bias as a result of heterogeneous techniques, and selection bias as a result of failure to compare similar cohorts. In general though, we have addressed only the most basic statistical concerns (i.e., beta-error as a result of a deficient number of study subjects).5-7

The goal of this editorial is to present a challenge to orthopaedists: Can we, at last, answer definitively, through a well-designed prospective study, under what conditions the single- and double-bundle ACL approaches differ in a clinically meaningful way and develop guidelines to inform the next generation of orthopaedic surgeons? We think it may be possible and will now describe the steps needed to accomplish this goal.

Before we describe our suggested design, we wish to point out to the skeptics among our readers—those who believe that the answer is already clear and that there is no need for a clinical trial—that there have been examples in other medical disciplines, where some experts were wrong and a definitive clinical trial did revolutionize our understanding.

In the early 1990s it was commonly believed that giving postmenopausal women hormone replacement therapy (HRT) might help reduce the risk of heart disease, cancer, or stroke. There was actually evidence, resulting from observational studies, that suggested this might be true.8,9 However, when randomized clinical trials were designed and conducted to address the question, it was found that there were serious risks involved with use of HRT.10,11 In 1991, the Women’s Health Initiative (WHI) was established, and a very large clinical trial (including more than 60,000 subjects) was initiated to address this and other questions.12,13 By July 2002, the WHI determined that the risks associated with HRT use (increased coronary heart disease, stroke, breast cancer, and blood clots in
the lungs) outweighed any benefits.\textsuperscript{10,11} As a result, use of HRT decreased dramatically. In 2000, before the study’s completion, the number of prescriptions of HRT were rising worldwide annually at a fast rate; however, once the results of the trial were published in 2002, the number of prescriptions dropped precipitously.\textsuperscript{14,15} From our perspective, the most important consideration required to accomplish our goal of comparing single- and double-bundle ACL procedures is the issue of how to design such a study. We thus address key elements that would be needed to design a definitive study to answer our question. To do this, we must first pose questions that need to be answered so as to properly design such a trial.

What precisely are single-bundle and double-bundle procedures? For a study to be conducted, one must define precisely the two approaches. There needs to be consensus on how each procedure should be performed and a standardized protocol developed to train surgeons in these two methods. All surgeons who agree to be part of this trial would need to confirm that they will follow this standardized protocol for each procedure after they are randomized to perform one or the other technique. This leads to a second question.

Will surgeons (or patients) agree to be randomized? For a trial to be optimally designed, surgeons who participate would have to agree to be randomized to perform either single- or double-bundle ACL procedures on patients who were deemed appropriate to be included. Furthermore, the patients also would have to agree to be randomized to receive one of these two procedures. This may pose a practical obstacle to conducting an optimal study. Eligible patients and surgeons both must feel that they are in equipoise to participate; namely that they must be convinced that either approach could work, that they are comfortable receiving or performing either intervention, and that pending the outcome of the investigation, they are unsure as to which technique could be superior.

Although this many criteria may seem difficult to achieve, the solution is clear delineation of study subject inclusion and exclusion criteria. Specifically, patients (subjects) would need to have some form of evaluation performed to suggest that either procedure could be performed. This evaluation would take place during the screening phase of the trial, before randomization to one of the procedures. So for example, small-framed women could be excluded from this trial if there is existing evidence to suggest that the single-bundle approach may be preferred for such patients. (Of course, each exclusion criterion limits the generalizability of a study because the research findings could not be definitely applied to excluded cohorts.) Nevertheless, assuming that a well-described list of inclusion and exclusion criteria can be determined, the surgeons who agree to participate in this study would have to agree that, for patients who meet the inclusion criteria, they would allow an external group to instruct them as to the treatment assignment (single- or double-bundle) for their next patient.

The concept of randomization is vital to clinical, comparative-effectiveness research, and the HRT research described above is a clear example. For years, clinicians had considered prescribing HRT based on observational (nonrandomized) data that suggested HRT was beneficial. It was not until clinicians (and patients) allowed themselves to be part of the large, randomized study, that the potential risks of HRT were well recognized.

What if any type of blinding can be used in a study? Assuming that we are able to address the randomization issues described above, we must confront whether there is a possibility to perform any meaningful blinding in this surgical trial. Clearly the surgeon cannot be blinded, but is it ethical to blind a patient to which procedure is performed? Our opinion is that the answer is yes. A proposed study design could be such that a study subject is informed clearly with regard to the two surgical methods, and also well informed that during the time of optimal study follow-up (a minimum of 2 years), they would not be told whether they had received the single- or double-bundle procedure; then, after the follow-up assessments were completed, their actual procedure information would be revealed. If such a blinding could occur, and were viewed as ethical by an Institutional Review Board, this would mitigate against patients themselves introducing their bias as to which procedure is “better,” which could be reflected on subjective outcome assessments.

Even if patient blinding were not possible, it is essential to have a team of clinic research staff blinded to the patient’s procedure. Specifically, the staff who perform the follow-up outcome assessment should not know which procedure was performed.

Since the primary surgeon cannot be blinded, one could imagine follow-up visits taking place in two stages. One would be the requisite follow-up with the treating surgeon. The second would be outcome assessment by the blinded, clinical research team. Finally, study data analysts would perform the analyses in a blinded fashion, and only when the modeling and analyses were completed would the true randomization be revealed. The importance of maintaining some
level of blinding in each different step is that it reduces the possibility of different forms of potential recording bias (participant, assessor, analyst) from entering into the research inferences.

This leads us to another consideration: What are the outcomes of interest and when must they be measured? We must determine, prospectively, what are the clinically meaningful outcomes to be measured and compared in the study. Furthermore, we need to determine how frequently, and for how long, the measures need to be applied.

There are a number of condition-specific and joint-or extremity-specific objective measures that can be assessed both preoperatively and postoperatively at fixed intervals. In addition, quality-of-life assessments should be measured both preoperatively and postoperatively. The importance of the preoperative measurements is that, even if we are able to randomize patients in the trial, the relevant outcomes are still to be assessed both within and between each patient. Namely, we need to determine whether improvements (within-patient changes) are different based on the procedure (between-patient changes).

A perennial research challenge is the ability to obtain medium- and long-term follow-up data on clinical research subjects. When designing a study, we are aware that the longer time we wait to gather the “final” treatment assessment, the greater the risk that patients may drop out of the trial and be considered “lost to follow-up.” This issue may be important in a study such as this because conventional wisdom suggests that a patient who is having problems is more likely to continue to return for long-term follow-up visits, whereas a patient doing well may decide to not return. Also, the opposite may occur, where a patient doing poorly may seek other treatments and choose not to follow-up with their treating physician. Regardless, recognizing prospectively that such problems could arise, researchers must work proactively and assiduously to enhance patient retention.

How many patients or centers would be needed to perform our proposed study? Even if all the issues above are addressed appropriately, this could be a rate-limiting step. As a statistician, I am often asked this question, and the feasibility of a clinical research proposal often hinges on the answer. In this case, the answer is always “It depends.” The answer depends, specifically, on what primary efficacy end point (outcome measure) is selected for the trial. Once this measure is determined, researchers must then determine what defines a clinically meaningful difference in the outcome. Based on my experience, the better designed a study, the fewer patients are needed. For instance, if we are able to randomize patients; include clear inclusion and exclusion criteria; blind patients, staff, and analysts; and measure objective outcomes at regular intervals with little loss-to-follow-up, then our outcomes will be assessed with a higher level of precision than if we do not achieve those goals. That being said, a study would ideally be multi-institutional. If we wish the results of this study to be definitive, it would need to include many surgeons from many locations for the broader orthopaedic community to trust the results. Ideally, a randomization scheme would need to be developed to guarantee that, within each medical center involved in the trial, there was a balance between the number of single- and double-bundle procedures performed by surgeons who are competent in both procedures. With these considerations in mind, one could imagine that many centers and surgeons would need to be invited to participate. With these centers and surgeons recruited, we can target a large number of patients for this study. I believe that the ideal number will be between 10 and 99 surgeons, and between 100 and 999 subjects. In truth, the answer to the exact numbers is, “It depends.” However, 256 patients in each group, or 512 patients in total does allow detection of a one-quarter standard deviation difference between groups, and this might be a target.

How could such a study be funded, coordinated, and monitored? The WHI study was funded by the federal government and was a multi-million dollar study with a dedicated coordinating center to help collect and analyze all of the data. In addition, there were external, funded, data-safety monitoring boards in place to monitor the trial for safety and efficacy throughout. Who would fund a single- versus double-bundle study? Perhaps the United States Agency for Healthcare Research and Quality (AHRQ) would be interested. There is a growing call for comparative-effectiveness research, and this study would fall into that category. Many great ideas may not be pursued if there are insufficient funds, so clearly this issue cannot be ignored.

In conclusion, medicine, orthopaedics, and specifically arthroscopic and related surgery continue to advance rapidly as a result of technological advances. New methods and instruments are continuing to be developed, and solid objective data are needed to determine whether the new procedures, devices, and/or instruments, really lead to better patient outcomes. The question about single- versus double-bundle ACL surgery is but one example of the many
clinical controversies with which arthroscopic and related researchers are faced.

We have a challenge: Can we conduct well-controlled clinical trials to address the important unanswered questions within our discipline? As described, there are many considerations to be addressed, but if we can tackle these challenges, we can be more confident that we know when to perform the correct procedure and use the correct technique for the appropriate patient. Clinical trials, such as the one we propose here, are performed in many other medical disciplines, and require time, patience, and funding. The WHI trial took nearly 10 years before the results were known.\textsuperscript{10,11} Orthopaedic surgical trials may not take that long, but they will take years, not months, to complete. It is time consuming to raise funds, recruit clinicians and centers, develop a protocol, standardize methods, recruit patients, and obtain a minimum of 2-year follow-up on all subjects, to say nothing of analyzing the data, writing it all up, submitting it for publication, and participating in the peer-review process. Can we wait for 5 or more years to answer a question? The sooner we decide, the sooner we may start, and if the answer is yes, then the question becomes, who is up to the challenge? Based on the history of our journal and our societies, we believe that the answer is that there are many who are!

RALPH B. D’AGOSTINO JR., PH.D.
Statistical Associate Editor

JAMES H. LUBOWITZ, M.D.
Assistant Editor-in-Chief

MATTHEW T. PROVENCHER, M.D.
Deputy Editor

GARY G. POEHLING, M.D.
Editor-in-Chief

REFERENCES

2. Lubowitz JH, Poehling GG. Techniques in double-bundle anterior cruciate ligament reconstruction: As simple as ABC, or putting the cart before the horse? \textit{Arthroscopy} 2008;24:1089-1091.
17. Lubowitz JH, Poehling GG. Comparative effectiveness research: We must lead (so as not to be misled). \textit{Arthroscopy} 2009;25:455-456.