TABLE 4. *SORT Rating System*

Code	Definition
A	Consistent, good-quality patient-oriented evidence
B	Inconsistent or limited-quality patient-oriented evidence
C	Consensus, disease-oriented evidence, usual practice, expert opinion, or case series for studies of diagnosis, treatment, prevention, or screening

whether any new strides have been made in this area since the conclusions in the filtered resources were released. The challenge with unfiltered resources is that the onus is put on the clinician to evaluate each study to determine its validity and applicability to the query at hand. Searching for these resources efficiently and subsequently appraising what is found take more time and skill, which is why filtered information is typically considered first.

MEDLINE is considered the database of choice for the health sciences because it provides both primary and secondary literature for medicine and other allied health professionals. In these instances, RCTs, meta-analyses, and systematic reviews are considered the gold standard and should be considered first.

Ratings of Quality of Evidence: Various rating scales have been developed to help the busy clinician gauge the quality of research based on an externally applied rating before starting the critical appraisal step of practicing EBM. The Centre for Evidence-Based Medicine in Oxford provides 3 different rating scales ranging from 1 to 5, each number and occasionally added letter identify the level of evidence based on type of research design and various measures of quality, such as confidence intervals and randomization. The most updated of these detailed scales can be accessed at www.cebm.net.

Strength of Recommendation Taxonomy (SORT) (with codes A, B, and C) is a straightforward rating system,[19] shown in Table 4.

Grading of Recommendations Assessments, Developments and Evaluation (GRADE) is a rating system developed by the GRADE Working Group in 2007,[20] shown in Table 5.

Integration With Clinical Expertise Into Practice

Arguably the most important aspect of EBM, or the goal of EBM if you will, is to integrate best evidence with clinical expertise for best treatment of a patient. The ability to integrate best evidence with clinical experience into practice is 2-fold: (1) one must be comfortable and capable in utilizing EBM in his or her practice, and (2) one must be able to understand and incorporate the patient's needs and wants to establish the best course to follow in terms of treatment and management.

When using the approach to practicing EBM discussed in this chapter, it is important to recall that the goal is to combine evidence, clinical experience, and patients' rights and perspectives to determine the solution. The importance of patients' perspectives, beliefs, expectations, and goals for life and health cannot be downplayed, because the approach to care in this EBM model is patient centered. By considering how patients think about the available options and their relative benefits, harms, costs, and inconveniences when determining options through evidence and clinical expertise, we engage in shared decision making. With this approach, we can make a compromise between these 3 factors to determine the best approach to a given patient in a given context.

TABLE 5. *GRADE Rating System*

Code	Quality of Evidence	Definition
A	High	Further research is very unlikely to change our confidence in the estimate of effect. • Several high-quality studies with consistent results • In special cases, 1 large, high-quality multicenter trial
B	Moderate	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate. • One high-quality study • Several studies with some limitations
C	Low	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate. • One or more studies with severe limitations
D	Very low	Any estimate of effect is very uncertain. • Expert opinion • No direct research evidence • One or more studies with very severe limitations

There have been various resources developed to help busy clinicians to identify and integrate the best available research evidence with their clinical expertise and the patient perspective discussed above. Clinical guidelines based on best evidence have been developed in a variety of specialties, providing a good starting place for commonly encountered scenarios.

Evaluation

An important component of practicing EBM is the fifth step: self-evaluation. After working through the EBM steps on a particular clinical question, it is important to review each of the 5 steps and evaluate whether it was completed in its entirety, effectively and efficiently. By continuously self-evaluating, gaps in a clinician's EBM skill set can be identified. A complete and helpful list of important questions to ask oneself in evaluation can be found in *Evidence-Based Medicine: How to Practice and Teach EBM*.[17]

- Are my questions specific and answerable? Am I able to form questions throughout the day and save time to target them later? Is the importance of asking good questions coming across in my teaching? Am I modeling this?
- Do I know the best resources for my questions? Do I have appropriate access to these resources? Am I searching from a wide variety of resources?
- Am I critically appraising the evidence I have found? Am I accurately and efficiently applying measures introduced here (likelihood ratio [LR], number needed to treat [NNT], relative risk reduction [RRR])?
- Can I work through particular concerns about management and integration to relate this evidence to a particular patient? Can I accurately and efficiently adjust my findings to fit my unique patient?

As mentioned earlier in this chapter, one of the important concepts that has fostered an environment where EBM can blossom is the idea of lifelong learning. Alongside self-evaluation, one of the most important techniques we can use to better ourselves as clinicians is to encourage and engage in continuing professional development. Developments in how we practice EBM, identified and updated through ongoing self-evaluation, are a part of this lifelong learning, while continually aiming to increase our knowledge base of the best evidence. What good is this evidence, however, without professional wisdom? Without professional wisdom obtained through ongoing professional development, evidence cannot be adapted to specific circumstances, and circumstances where evidence is not available would present quite a challenge.

CONCLUSIONS

This section has just but scraped the surface with regard to the impact of this paradigm shift in medical practice. This new approach to clinical decision making focused around the sound application of best research evidence is becoming so common in all fields of medicine that you would be hard pressed to find a physician or surgeon not familiar with RCTs, meta-analyses, Cochrane reviews, or evidence-based guidelines. As orthopaedics moves forward with the momentum of this global EBM movement, evidence-based orthopaedics is becoming a term, concept, and way of life in the clinical setting for all in the field. As discussed, it is not only the retrieval and appraisal of evidence that are important, but also how this evidence can be applied to a specific clinical situation considering societal values, as well as each patient's individual perspective. By learning how to approach searching for evidence in an effective and efficient manner and by learning where to look, how to look, and what you are looking for, the task of using evidence in everyday clinical practice becomes less and less daunting.

Lauren E. Roberts, M.Sc.
Jón Karlsson, M.D., Ph.D.
Mohit Bhandari, M.D., Ph.D., F.R.C.S.C.

SECTION 3

Levels of Evidence

Traditionally, surgical indications and treatment decisions have been based largely on expert opinion and personal experience. Although EBM has been proclaimed as one of the greatest achievements in internal medicine over the past 150 years,[21] its influence has been slow to seep into the surgical literature

because of the unique challenges of surgical trials. In 2003 levels of evidence were first introduced into *The Journal of Bone and Joint Surgery*, reflecting increased awareness of the importance of quality in an individual study.[22] This recognition of high-level research propelled the orthopaedic community to design and accomplish better studies,[23,24] which in other areas of medicine have ultimately led to significant treatment advances.[25]

Levels of evidence are important not only in determining whether a study is of higher quality than another, but they give the reader an immediate sense of how much weight the results of the study should be given.[26,27] The Oxford Centre for Evidence-Based Medicine has created a detailed hierarchy of evidence, in which the highest level remains a meta-analysis of homogeneous, high-quality RCTs.[28] A significant proportion of current orthopaedic studies are observational studies. To ensure standards of reporting observational studies, the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement was created, which assists investigators when reporting observational studies and supports editors and reviewers when evaluating these studies.[29] More recently, Grades of Recommendation Assessment, Development, and Evaluation (GRADEs) have been introduced to allow for a transparent and comprehensive method to grade the quality of evidence and strength of recommendations about the management of patients.[30]

HOW LEVELS OF EVIDENCE ARE ASSIGNED

What Is the Primary Research Question?

For a level of evidence to be assigned, one must first assess the primary research question. The level of evidence is assigned specifically to whether the primary research question, well-defined in the purpose section of a manuscript, was aptly addressed in the results and conclusion sections. Thus, asking a focused question helps yield a more answerable question, and assignation of level of evidence is relatively straightforward.

For example, in a study comparing the use of a bioabsorbable interference screw versus washer-post construct for tibial fixation in an anterior cruciate ligament (ACL) reconstruction, it would be ideal to only manipulate a single variable. In other words the study includes the same surgeon, same technique, and all patients with the same isolated ACL injury. The outcome would be a single data point, such as Lachman examination. In this way, the primary research question is focused on answering 1 specific question: "Does tibial fixation of the graft affect the postoperative Lachman examination?" If a difference in tibial translation is found between the 2 types of fixation, then a conclusion can be made as to whether or not there was a difference.

Conversely, it becomes very difficult to assign a level of evidence when the primary research question is not well-defined or the conclusions do not answer the research question. Frequently, studies will make conclusions based on their results, but in fact the conclusions were not related to the primary research question. Therefore it is extremely important when designing or reviewing a study to first evaluate whether the research question for the study is well defined and then evaluate whether the conclusions of the study are related to that primary research question.

Study Designs

Once the primary question is determined, the next task is to identify the study type. Levels of evidence can be divided into 4 different study designs: therapeutic, prognostic, diagnostic, and economic or decision analyses.[31]

Therapeutic Study Type: Therapeutic studies focus on assessing the effect of a specific treatment on the outcome of a specific disease process. A practical test to determine whether a study design is considered therapeutic is if the factor being studied can be allocated in a random fashion. For example, a study of ACL reconstruction evaluating the effect of graft type (e.g., bone-tendon-bone *v* hamstring autograft) on the outcome of reconstruction would be a therapeutic study because the graft type can be randomly allocated.

Prognostic Study Type: Prognostic studies evaluate the effect of patient characteristics on the outcome of a disease process. Prognostic studies differ from therapeutic studies because the factors being evaluated cannot be randomly allocated. For example, a study of the effect of age on outcome of ACL reconstruction in 2 different study groups (e.g., patients aged <30 years *v* patients aged >30 years) would be considered a prognostic study because age cannot be randomly allocated to 2 groups of patients in the study.

Diagnostic Study Type: Diagnostic studies are designed to assess whether a specific test is related to the presence or absence of a particular pathology. For

example, in patients with femoroacetabular impingement of the hip, the anterior impingement test can be performed for assessment. A study examining the effect of the anterior impingement test and its relationship to femoroacetabular impingement is an example of a diagnostic study design. Another example would be joint-line tenderness and its ability to detect meniscus tear.

Economic Analyses: Economic analyses are designed to assess the cost-effectiveness of a certain treatment for a certain pathology. For example, in the case of a group of young patients with femoroacetabular impingement, one might compare the cost-effectiveness of open versus arthroscopic impingement surgery.

Decision Analyses: Decision analysis studies are performed to evaluate the outcome of a certain therapy to determine the ideal treatment. For example, in evaluating surgical versus nonsurgical treatment for patients aged greater than 40 years with ACL deficiency, an expected-value decision analysis, which is a systematic tool for quantitating clinical decisions, can be used to conclude ACL surgical reconstruction as a preferred treatment.[32] An inherent limitation of this study type is that actual patients are not evaluated.

LEVELS OF EVIDENCE IN DETAIL

Several systems for rating levels of evidence are available.[33] The one chosen by *The Journal of Bone and Joint Surgery* and *Arthroscopy* has 5 levels of study design for each of 4 different study types: therapeutic, prognostic, diagnostic, and economic or decision modeling.[21,22,25] Among study designs, there exists a hierarchy of evidence, with RCTs at the top (Level I), controlled observational studies in the middle, and uncontrolled studies and opinion at the bottom (Level V).[33]

Understanding the association between study design and level of evidence is important. Higher levels of evidence should be more convincing to surgeons attempting to resolve clinical dilemmas.[21] Because randomized clinical trials are not always possible, Level I evidence may not be available for all clinical situations. Therefore Level III or IV evidence can still be of great value to the practicing orthopaedic surgeon. It is important to consider that an answer to a clinical question must be based on a composite assessment of all available evidence. No single study provides a definitive answer.

Level I

Therapeutic studies

1. RCTs with (*a*) significant difference or (*b*) no significant difference but narrow confidence intervals
2. Systematic reviews of Level I RCTs (studies were homogeneous)

Prognostic studies

1. Prospective studies
2. Systematic review of Level I studies

Diagnostic studies

1. Testing of previously developed diagnostic criteria in series of consecutive patients (with universally applied reference "gold" standard)
2. Systematic review of Level I studies

Economic and decision analyses studies

1. Clinically sensible costs and alternatives; values obtained from many studies; multiway sensitivity analyses
2. Systematic review of Level I studies

Level II

Therapeutic studies

1. Prospective cohort study
2. Lesser-quality RCT (e.g., <80% follow-up, no blinding, or improper randomization)
3. Systematic review of Level II studies or Level I studies with inconsistent results

Prognostic studies

1. Retrospective study
2. Untreated controls from an RCT
3. Systematic review of Level II studies

Diagnostic studies

1. Development of diagnostic criteria on basis of consecutive patients (with universally applied reference "gold" standard)
2. Systematic review of Level I and II studies

Economic and decision analyses studies

1. Clinically sensible costs and alternatives; values obtained from limited studies; multiway sensitivity analyses
2. Systematic review of Level II studies

Level III

Therapeutic studies

1. Case-control study

2. Retrospective cohort study
3. Systematic review of Level III studies

Diagnostic studies

1. Study of nonconsecutive patients (without consistently applied reference "gold" standard)
2. Systematic review of Level III studies

Economic and decision analyses studies

1. Analyses based on limited alternatives and costs; poor estimates
2. Systematic review of Level III studies

Level IV

Therapeutic studies
Case series (no, or historical, control group)
Prognostic studies
Case series
Diagnostic studies

1. Case-control study
2. Poor reference standard

Economic and decision analyses studies
No sensitivity analyses

Level V

Therapeutic studies
Expert opinion
Prognostic studies
Expert opinion
Diagnostic studies
Expert opinion
Economic and decision analyses studies
Expert opinion

EXAMPLES OF STUDIES OF DIFFERENT LEVELS OF EVIDENCE

Level I

In a study of a consecutive series of patients with the diagnosis of internal snapping hip syndrome, patients were randomized into 2 different methods of endoscopic release of the iliopsoas tendon.[34] Patients in group 1 were treated with endoscopic iliopsoas tendon release at the lesser trochanter, and patients in group 2 were treated with endoscopic trans-scapular psoas release from the peripheral compartment. A quality randomization process included randomizing patients at the last possible time point, e.g., at the time of surgery. An a priori power analysis was performed to ensure adequate numbers of patients in each randomized group. Preoperative and postoperative clinical and imaging assessments were evaluated for all patients. No statistical difference was found between groups. Therefore this RCT, with a defined and appropriate sample size and narrow confidence intervals, is characterized as a Level I study, even though no statistically significant difference was determined.[34]

Level II

In a prospective cohort study, patients aged older than 40 years were compared with a group of patients aged younger than 40 years who underwent autologous chondrocyte implantation for isolated cartilage defects of the knee.[35] The authors' hypothesis was that the older group of patients would have inferior clinical results compared with the younger group of patients. All patients were followed up for 2 years, and validated clinical outcomes were used. The authors' hypothesis was disproved, because there was no statistically significant difference in the 2 groups of patients treated with autologous chondrocyte implantation. This prospective study does not obtain a Level I designation because it is nonrandomized.

Level III

The efficacy of open versus arthroscopic Bankart repair remains controversial; therefore the authors designed a retrospective case-control study to determine whether there is a significant difference in cost between the 2 surgical procedures. In a Level III retrospective case-control study, the authors retrospectively reviewed the medical records and billing information of consecutive patients treated for recurrent, post-traumatic anterior shoulder instability.[36] They compared 22 patients who had open Bankart repair with 20 patients who had arthroscopic Bankart repair. Total operating times and all charges were obtained from records. Patients were also clinically evaluated. This study found similar shoulder scores and rates of dislocation between the 2 groups. The arthroscopic Bankart repair had a lower cost, but if an obligatory overnight inpatient stay was taken into account, the cost difference was negligible. Because of its retrospective nature, this study was characterized as a Level III, therapeutic cohort study.[36]

Level IV

The purpose of a Level IV study is to retrospectively review the outcome of a group of patients

treated in a similar way. In a Level IV therapeutic case series study, the authors described transphyseal ACL reconstruction with hamstrings performed in 26 patients with open tibial and femoral physes.[37] Clinical and radiologic outcomes were evaluated retrospectively. Their outcomes were well defined, with validated knee scores and detection of any growth disturbance on scanograms. They concluded that their technique yielded good outcomes and no growth disturbances. Because the authors did not compare their technique with another technique, and given its retrospective nature, this represents a Level IV therapeutic case series.

Level V

In a Level V study, the authors showed the use of the 70° arthroscope for several arthroscopic procedures and in a number of circumstances in which it offers superior visualization to a 30° arthroscope.[38] In this study the authors demonstrated their particular expertise with this arthroscopic instrument, which may be interesting for arthroscopic surgeons who are not familiar with the 70° arthroscope. However, because this study does not report any results or clinical outcomes, it is considered expert opinion.

GRADES OF RECOMMENDATION

As surgeons, we often find multiple studies to be more convincing than a single article. Although the appropriate literature on a particular clinical question can be identified in many ways, to search the literature ourselves is time-consuming and the search may not be comprehensive. Although review articles are often comprehensive in the available evidence they include, the conclusions that they contain can be uncertain. Therefore grades of recommendation have been introduced in the development of practice guidelines. In this process a reviewer or organization can gather all the appropriate literature, appraise the literature by assigning a level of evidence, and summarize the overall quality by allocating a grade of recommendation.[26,30] This helps the reader by giving definitive treatment recommendations that should definitely (grade A) or probably (grade B) guide treatment decisions for their patients. In addition, a grade of I, or insufficient or conflicting evidence not allowing a recommendation for or against intervention, advises a surgeon to treat patients based on his or her best judgment or on a case-by-case basis.

Grade A indicates good evidence (Level I studies with consistent findings) for or against recommending intervention. Grade B indicates fair evidence (Level II or III studies with consistent findings) for or against recommending intervention. Grade C indicates poor-quality evidence (Level IV or V studies with consistent findings) for or against recommending intervention. Grade I indicates that there is insufficient or conflicting evidence not allowing a recommendation for or against intervention.

CONCLUSIONS

The purpose of this chapter is to provide the orthopaedic sports medicine surgeon with a better understanding of the levels of evidence and their clinical implications. Such understanding is extremely helpful, not only from a research design standpoint but also to aid readers in understanding the importance of a particular study's conclusions.

From a design standpoint, in order for a research protocol to maximize the best possible level of evidence, it is important for the orthopaedic sports medicine researcher to consider levels of evidence when outlining the primary research question. Furthermore, it is important to recognize that performing a Level I surgical study is extremely difficult, because it requires a significant amount of preparation, time, and financial investment to allocate resources. Level II, III, and IV studies have their own worth and merit and are especially useful in the circumstances where Level I studies would not be feasible. When observational studies are being performed, the STROBE recommendations will assist the investigator in maintaining methodologic transparency and also assist the reader in comprehensively evaluating the quality of the study.

From a reader's standpoint, if a study is assigned Level I evidence, and a grade A recommendation, then the reader can feel confident that the results of the study have the highest level of validity. In this situation the reader/surgeon may choose to change clinical practice based on those recommendations, thus shaping and directing future sports medicine care. Ultimately, it is this endpoint, the best care for patients, that is our highest goal.

Aaron J. Krych, M.D.
Bruce A. Levy, M.D.
Mario Ferretti, M.D., Ph.D.

SECTION 4

Study Designs: Randomized Trials, Level I Evidence, CONSORT Checklist

The randomized clinical/controlled trial (RCT) represents the highest level of evidence or study design. In orthopaedic surgery and sport medicine, there are a multitude of questions requiring evidence-based answers. It will become increasingly more important to perform RCTs to address these questions. This chapter identifies the problems encountered by the surgeon, as well as the strategies and how to address these concerns.

WHY PERFORM RCTS?

An RCT is the most valid study design to evaluate the efficacy or effectiveness of surgical treatments. Efficacy refers to the ideal situation with optimal patient selection, well-controlled surgical technique, postoperative compliance, and so on. This type of randomized trial is sometimes referred to as an explanatory trial.[39] Effectiveness refers to a more real-world situation, where the patients have more variability in their disease state, multiple surgeons may be involved, and the postsurgical course is less well-controlled and more typical of most surgeons' practices. This type of RCT is sometimes referred to as a pragmatic trial.[30]

The RCT is prospective by definition, and therefore this term is redundant. All RCTs are prospective because the primary research question being addressed, independent variable (treatment), dependent variable (outcome), and inclusion and exclusion criteria should all be determined a priori. Patient recruitment and enrollment, consenting, randomization, data collection, and analysis are subsequently performed in a forward direction. Patients are randomly allocated to different treatment groups and are typically followed up in an identical manner with the main outcome of interest measured at a specific period of time. The groups of patients are similar with respect to known characteristics (e.g., inclusion and exclusion criteria) and unknown characteristics (those present by chance). Provided that an appropriate sample size is calculated and recruitment is achieved, the unknown characteristics are likely to be equally distributed between groups. Therefore, if a difference in outcome is identified, the findings can be directly attributed to the efficacy or effectiveness of the specific surgical treatment.

Therefore the RCT is less likely to introduce bias because treatment group assignment is randomly determined. Other biases can occur, however, despite the randomized design. These would include a lack of blinding of the patients or assessor, different follow-up times, loss to follow-up, differential exclusions, expertise-based bias, and early reporting of results before the full sample size is achieved.[40] Whatever bias is introduced must first be recognized, and then it may be accounted for.

So, if we truly want to determine the benefits of one operative technique over another, surgical versus nonsurgical treatment, different rehabilitation protocols, and so on, the RCT is the best possible study design. We must be cognizant of the fact that the design is only one component of conducting valuable research. Randomization does not compensate for poor adherence to all other important methodologic issues.

REASONS FOR NOT PERFORMING AN RCT

The Problem

Randomized clinical trials are only necessary if the clinical problem/surgical treatment is common; if the question is a significant issue to clinicians and patients; and most importantly, if the answer to the clinical question is clearly not known. It would be unnecessary to perform a randomized clinical trial in circumstances where observational studies are so compelling and/or outcomes are dramatic and life-saving (e.g., amputation compared with antibiotic treatment for clostridial gas gangrene of the foot). In other words, the treatment effect is so large, and the consequences so grave, that it is not necessary to compare the surgical procedure with existing treatment. The same can be said for parachute use compared with placebo.[41] All problems are not amenable to an RCT.[42] Until the rules of surgical engagement change to be similar to those related to medical ther-

apy it is unlikely that surgical RCTs will become the norm rather than in the minority.[43]

The Patient

Patients present with preconceived notions about what treatment is best. Whether they have talked to a friend or family member, someone in health care, a physician, or even another surgeon, there is a bias that each patient may have. Patients will typically investigate their problem on the Internet. They will identify specific operations and other surgeons who have addressed their problem with a particular procedure.

There are different cultural expectations around the world. At the same time that an RCT comparing outpatient with inpatient ACL surgery was being performed in Canada, patients were staying in the hospital for 2 weeks or more in Europe and Japan.[44]

Patients may simply want to know what procedures they will be undergoing, and any doubt leads to a lack of confidence. Some patients will consent to a trial because it is the only chance that they can undergo the latest procedure (i.e., being included in the experimental group rather than undergoing the usual technique). Some patients feel the exact opposite sentiment.

There is a permanency regarding surgery. This can affect a patient's decision, such as in a trial comparing surgical with nonsurgical treatment. If a patient ends up in the nonsurgical arm, there is a perception that if the treatment fails, then surgery may be an option. However, once a procedure is performed, there is no going back. These patient-related concerns influence whether eligible patients are open to recruitment into a trial.

The Performer (Surgeon)

The surgeon may be the greatest barrier to surgical trials! Compared with medical therapies where there are strict regulations on how a drug is released, surgical innovation can occur with little or no restraint from regulating agencies, hospitals, or local ethics committees. This is definitely the case when there is a minor variation in technique or implant used. Therefore there is no incentive whatsoever to perform a trial to determine the efficacy of a particular procedure.

Surgeons are innovators. Arthroscopic surgery may not have become the gold standard if we had the requirement of randomized clinical trials to show its benefit. Historically, arthroscopy of the knee was more expensive, took longer to perform compared with the equivalent open procedure, and was fraught with complications.

The irony of being a surgeon is that we can perform "experimental" surgery on our patients with their consent and with little or no regulation, but if we want to perform an experiment (i.e., an RCT), then we must obtain both scientific and ethical approval.[45]

The Procedure

There are many barriers to performing an RCT regarding the surgical procedure. It is well-recognized that there is a learning curve with respect to any operation. One could argue that, if it is a minimal change, then the learning curve is shallow but a significant departure from what is usually done may in fact have a very steep curve. A certain minimum standardization is required in any trial, particularly if there is more than 1 surgeon involved to ensure consistency between surgeons. If only 1 surgeon is involved, then it may be the case that he or she is better at 1 procedure compared with another. We can take the analogy from the sport of tennis. It is well known that only a few players in history have been able to master the game on hard courts, clay, and grass all in the same year. Why would we expect surgeons to be able to perform different procedures (e.g., arthroscopic compared with open) equally as well?[46] Therefore, if more surgeons are required (to increase sample size), then strategies such as training manuals, stratification by surgeon, and matching surgeon experience are techniques that can alleviate the variability of the procedure. One recently applied method has been called the "expertise-based design."[40] In this RCT the patient is randomized to the procedure, and the surgeon with the expertise for that particular procedure carries out the surgery.[40,47]

The Process

This is probably what scares most surgeons away from performing a randomized trial: the process from start to finish is overwhelming to most surgeons. Statements include the following: It is going to take too much time! We do not have the resources! We will never get it through ethics! We do not have enough patients! I get good results with the technique I am familiar with! We do not need to do a trial; I can do a bunch of these new procedures and compare to what I have done in the past! I am too busy taking care of patients to do research! I want to do the latest techniques; it is what my patients expect! It is not my responsibility to do research; let the researchers figure this out! I do not have a research assistant! It takes too much money!

There is no doubt that conducting a randomized trial requires significant infrastructure support, time, and effort. The process is daunting at first and difficult to

implement but eventually routine. Until there are stricter rules regarding the use of new procedures, surgeons will not be compelled to be involved in appropriate trials.[45]

HOW DO WE SOLVE THESE PROBLEMS AND WHAT ARE THE PREREQUISITES?

It is much easier to identify reasons not to do something. It is easier for a medical student to answer a difficult question with a negative response. For example, when asked about the differential diagnosis of an acute hemarthrosis in the knee, he or she may say, "It is not a tumor." Although this statement is correct and easy to identify on the differential, it is not a very useful answer in evaluating the patient. The solution lies in the following concepts, the 3 C's of successful trial research: clinical equipoise, commitment, and collaboration.

Clinical Equipoise

Clinical equipoise is defined as genuine uncertainty over whether one treatment or another is beneficial. This equipoise should involve the expert clinical community where there is "honest professional disagreement among . . . clinicians."[48] Therefore, if one or more treatment options are available for a particular problem and the best option is not known, then we have reason to consider a randomized clinical trial. It is necessary to build a case for clinical equipoise, which essentially is the essence of the rationale for the RCT. A surgeon should first review the available evidence in a systematic way, analyze the results (with a meta-analysis if possible), and determine an answer to his or her question. If this answer is clear, then there is no equipoise and, therefore, no need to perform another trial.

However, clinical equipoise relates to not only uncertainty regarding treatment options but the ethics of performing a trial from all perspectives: the patient's, the surgeon's, and society's. It is necessary to ask the question, Who is uncertain, the individual surgeon or the community of surgeons? A surgeon may consider randomizing his or her patients to a particular arthroscopic fixation technique such as absorbable compared with nonabsorbable suture anchors. Whereas this trial may be easy to perform because there is little impact on the patient's decision making, it may not matter to the community of surgeons or society as a whole. Therefore, is it really worth performing a randomized clinical trial? Patients just like surgeons are influenced by their position on the equipoise spectrum. They may desire to understand in great depth the treatment options, they may want to appreciate the bigger-picture perspective of helping medical science and therefore the surgical community perspective, or they may in fact simply trust the surgeon.

Clinical equipoise requires not only the consideration of the individual surgeon's perspective but the community of surgeons that establishes standards of practice. Most surgeons have difficulty with this concept, and therefore failure of consensus of evidence within the clinical community is the usual driver for a trial. Ultimately, the ethical surgeon must do what is best for his or her individual patient. Uncertainty is a moral prerequisite for being involved in an RCT, but if we know the best and correct treatment, then we should perform it.[45,48,49]

The following example should illustrate the concept of clinical equipoise and the moral or ethical responsibility of the surgeon. A randomized clinical trial was conducted to compare 3 surgical techniques for ACL reconstruction.[50] A meta-analysis had been conducted to determine whether an autograft patellar tendon compared with autograft hamstring reconstruction resulted in improved outcomes for patients at 2 years. Not only was the meta-analysis inconclusive but it identified many concerns with respect to the available evidence.[51] At the same time, surgeons were advocating the so-called double-bundle technique for ACL reconstruction. Therefore it seemed logical to conduct a randomized clinical trial comparing the existing techniques with the newer double-bundle procedure. This would represent established clinical equipoise. However, there was a clinically identifiable subgroup of patients who have the prerequisite diagnosis of an ACL-deficient knee but whose knees on careful examination had minimal documentable translational and rotational instability. These patients, also on careful arthroscopic examination, had identifiable ACL tissue that was both biologically viable and mechanically supportive. Whether this represents a partially torn ACL, a single-bundle ACL failure, or healing of a complete ACL tear is debatable. However, the surgeon believed on moral and ethical grounds that these patients should be excluded from the trial. This was based on the principle of biological preservation of the patient's own tissue and, most importantly, the empirical evidence from his own practice that this subgroup of patients had a better outcome than those who had undergone reconstruction in the usual way. This example demonstrates the difficulty with addressing both clinical equipoise and the ethics of performing a randomized clinical trial.

Commitment

Probably the most important prerequisite for conducting an RCT is one's commitment. Commitment relates to not only being involved with respect to one's role but more importantly being committed to the question rather than finding the answer. This is particularly difficult for surgeons because we are driven to solve patients' problems through our own individual surgical skills.[30,42,43,45,52-56] Surgeons are typically characterized as innovative rather than reflective, performers rather than observers, and are interested in immediate gratification rather than long-term rewards. The RCT requires a different type of commitment that is reflective before and after the fact, requires persistence, and may lead to an answer that is not consistent with what is expected. For example, in a trial comparing surgical with nonsurgical treatment, it is inherently difficult for a surgeon to be committed to the question unless there is a perceived problem with the surgical option. Our training, experience, and rewards are derived from the outcomes of the patient's surgical treatment.

As innovators, surgeons become readily aware of the latest technique or improvements on previous procedures. It might take 2 to 5 years to recruit enough patients into a trial and then the requirement of a minimum 2-year follow-up. The process of writing the proposal, obtaining funding, and obtaining ethical approval may take at least 1 year. A meaningful trial may take anywhere from 5 to 10 years to complete. During this time, the surgical world has moved forward, techniques have been modified, case reports may have suggested complications with a particular procedure, and so on.[57]

The surgeon must therefore act and be committed in a way that is somewhat foreign to his or her normal existence. This commitment is compounded by the fact that in every respect conducting a surgical trial takes more time and effort than what is necessary to run a clinical practice. Successful surgical "trialists" are just as passionate about the research as any aspect of their clinical practice. They likely spent additional time learning how to perform research in the fields of clinical epidemiology, methodology, or public health in addition to their clinical fellowships.

Collaboration

We are not aware of any surgeon in the world who is a methodologist and biostatistician, has extra time to devote to clinical research, and also a large enough clinical practice to conduct a meaningful randomized clinical trial without help.

The collaborative infrastructure support is not only helpful but necessary. One solution is to pay for collaborative support by hiring a research organization to conduct the trial and therefore enter patients and perform the surgery. Another approach is to identify individuals with the expertise in methodology and biostatistics who will be partners/co-authors in the trial and therefore will provide their expertise without financial compensation. There will always need to be a research coordinator or assistant. This individual provides day-to-day support and keeps the trial moving forward, addressing all of the details and complexities of conducting an RCT.

Collaboration may take the form of including other clinicians and surgeons. These individuals may be at the same institution or could be based out of multiple centers. In these circumstances the clinicians will need to have the same requisite ethical and clinical equipoise to the primary surgeon and the time and commitment necessary for success. It is well-recognized in multicenter trials that the host site is usually more successful in recruiting patients and conducting all aspects of the trial.[58,59] The exception to this is when the trial is funded centrally and the collaborating centers have financial incentives to recruit and follow up the patients.

CONDUCTING A RANDOMIZED CLINICAL TRIAL?

Once the prerequisites have been addressed (i.e., an important clinical concern, commitment to the question, and collaboration), the trial is ready to be implemented.

However, implicit within these prerequisites is that the research question has been carefully refined and a detailed proposal drafted, reviewed, and rewritten, along with an application for funding and ethical approval.[58,60] The implementation starts once approval and funding have been achieved.

It is necessary to engage all people and settings (hospital wards [i.e., emergency, inpatient, and outpatient], clinics, operating rooms, and so on) that may or may not be impacted by the trial. This process should ideally occur during the proposal stage but is an obligatory part of implementation. Informing and engaging everyone must occur before, during, and after conducting the trial if it is to be successful. Simple incentives such as providing refreshments to the hospital or clinic staff or giving gift vouchers to referring physi-

cians have proven to be cost-effective ways to facilitate this critical engagement.

Informing the medical community of the trial is also very important at the start. This includes presenting at rounds and business meetings, advertising the trial, and registering the trial in an international database.

Within the written proposal are specific criteria on how the patient population is to be sampled. When patients are seen, they need to be screened to determine whether they fit the inclusion and exclusion criteria. Assuming the patient is eligible, the consenting process can proceed. Informed consent is a critical time-consuming activity. There may be ethical issues with respect to who obtains consent, and this is typically regulated through each individual institution. It has been our experience with surgical trials that a surgeon is the person best suited to obtain consent for a surgical randomized clinical trial. This leads to higher recruitment of patients.

The process of randomization can take many forms, and there are different types of randomization. With respect to surgical trials with more than 1 surgeon involved, stratification by surgeon is necessary unless 2 surgeons are matched for all known characteristics such as experience, location, and so on.

One technique to help the process of randomization is called expertise-based randomization.[40] This is where the patient is randomized to the procedure before going to the operating room. This technique provides the surgeon the ability to participate in an RCT but still retain his or her independence and individual preference to perform his or her procedure of choice. We have used this expertise-based randomization technique successfully when comparing open and arthroscopic procedures in the shoulder.[46]

Irrespective of the type of randomization, there are specific requirements that must be adhered to. These include allocation concealment and adequate sequence generation, i.e., typically, computer-generated random-number sequencing.[39] Although opaque envelopes are considered an appropriate concealment technique, they can be tampered with and the sequence identified. Current standards would involve a Web-based remotely accessed computer-generated randomization process that is performed by someone independent of the surgeon or primary investigator.[39]

The randomized trial should include a primary outcome (i.e., the dependent variable), such as a validated patient-reported outcome, and the defined intervention (i.e., the independent variable), such as the standard operation compared with the new surgical procedure.

The sample size for the RCT is directly related to the primary outcome, the measured treatment effect (i.e., the expected clinical difference between the 2 treatment groups), and the variability of the outcome measured in the standard deviation. Ideally, the expected difference and variability of the outcome are known values based on previous pilot data or data from similar populations of patients.[39,53,54] Without this information, the sample size calculation becomes speculative, and therefore the trial may be underpowered to show a meaningful clinical difference between treatment groups. In general, the more precise (i.e., less variability) the outcome and the greater the expected differences between treatment groups, the smaller the sample size. In addition, those dependent variables that are measured on a scale that allows for correct statistical analysis with means and standard deviations (i.e., parametric statistics) are likely to require a smaller sample size. Trials where the primary outcome is a probability (i.e., nonparametric statistics) are more likely to require a greater sample size.

The greatest barrier to conducting a surgical trial is recruitment and therefore meeting the a priori sample size.[59] Surgeons typically overestimate the number of patients who would be eligible, and the eligible patients do not always consent to the trial.[57] Some of the strategies to improve recruitment include collaborating with more surgeons, involving multiple centers, using baseline data to recalculate the sample size (assuming that there is less variability), providing incentives to include patients, continual strategies to engage people, ensuring regular patient contact to avoid loss to follow-up, and modification of inclusion and exclusion criteria to be more inclusive with respect to eligibility.

Once all of the details of the trial are organized, carrying out the trial is arguably the easiest part. It necessitates the help of the coordinator and assistants, and it requires a time commitment; however, as people in the clinics, wards, and operating rooms become familiar with the process, the trial should move ahead smoothly.

Every strategy to maintain contact with the patients should be used. This may include regularly scheduled follow-up visits, phone communication, and use of e-mail or social media.

Once the data are collected and the patients have been followed up, the analysis will occur. The help of a biostatistician is usually necessary for randomized trials.

The results will be interpreted, presented, and subsequently published.

REPORTING RCTS: THE CONSORT CHECKLIST

The randomized clinical trial represented the gold standard for evaluating interventions, but the accuracy of such trials' reporting was not consistent and therefore bias could be introduced. A worldwide group of researchers, methodologists, and clinicians, concerned that the reporting of trials lacked lucid and complete descriptions of the critical information, created the Consolidated Standards of Reporting Trials (CONSORT) statement (1996).[61,62] This has undergone recent revision, in 2010.[63,64] The CONSORT 2010 statement, checklist, and flow diagram provide authors with guidance on how to report their trials. The flow diagram (Fig 3) illustrates the progress of the trial from the start and includes the following: (1) the enrollment phase with numbers of eligible patients and those excluded for reasons such as not meeting inclusion criteria or declining to participate or for other reasons, and the number of patients randomized; (2) the allocation phase, which includes the exact numbers of patients who were allocated to the treatment groups, whether they received the allocated treatment, and if not, why not; (3) the follow-up phase, which includes the numbers lost to follow-up and the reasons why; and (4) the analysis phase, which includes those patients in each group who were analyzed and any who were excluded and for what reasons.[63,64]

The checklist (Table 6) represents a much more detailed list of characteristics of the trial that need to be reported.[63,64] The list includes the title and structured abstract, an introduction, the methods, the results, a discussion, and a section for other information. The checklist requires the authors to identify within their manuscript the page number where the appropriate information is written. Important concepts include the background and objectives, the trial design, the patients, detailed descriptions of the interventions, whether the outcomes were completely defined and prespecified, sample size determination, blinding of

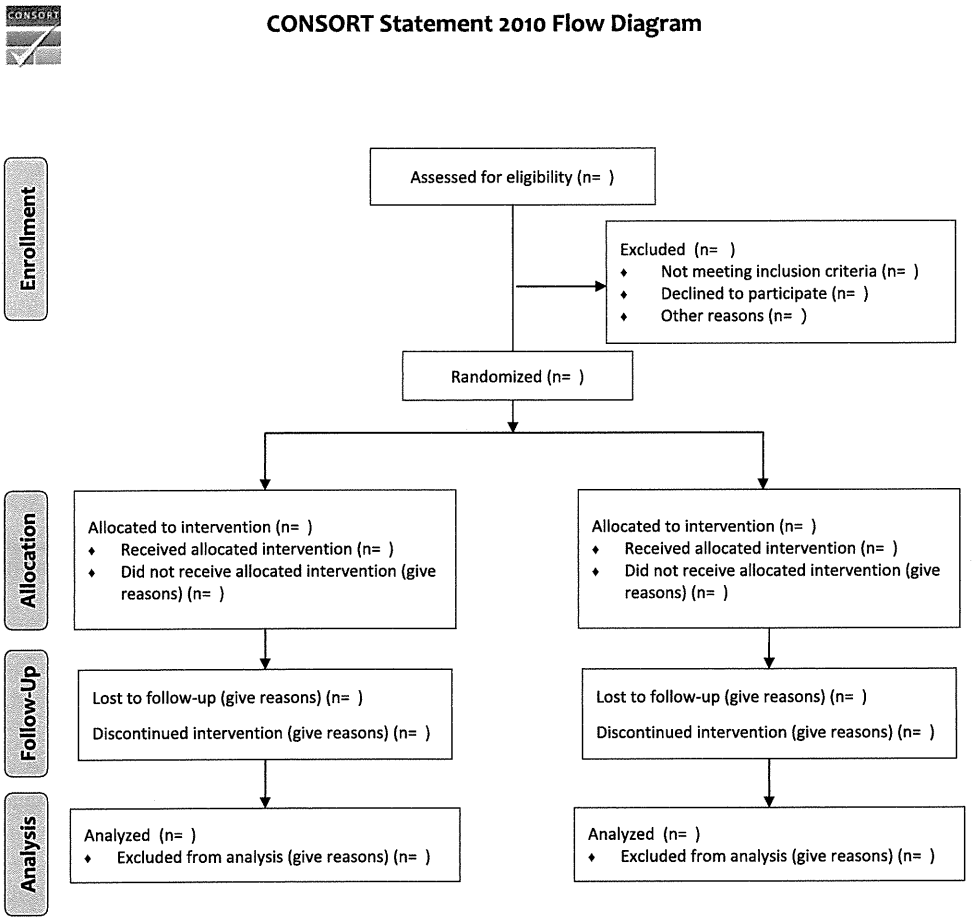

FIGURE 3. CONSORT flowchart.

TABLE 6. *CONSORT 2010 Checklist of Information to Include When Reporting a Randomized Trial**

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	
	2b	Specific objectives or hypotheses	
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	
Participants	4a	Eligibility criteria for participants	
	4b	Settings and locations where the data were collected	
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	
	6b	Any changes to trial outcomes after the trial commenced, with reasons	
Sample size	7a	How sample size was determined	
	7b	When applicable, explanation of any interim analyses and stopping guidelines	
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	
	13b	For each group, losses and exclusions after randomisation, together with reasons	
Recruitment	14a	Dates defining the periods of recruitment and follow-up	
	14b	Why the trial ended or was stopped	
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	
Other information			
Registration	23	Registration number and name of trial registry	
Protocol	24	Where the full trial protocol can be accessed, if available	
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org. Reprinted with permission.[63]

patients and investigators, the specifics of randomization, and details regarding the analysis and results, along with a discussion about the limitations and generalizability of the trial.

LIMITATIONS OF RCTS

Randomized clinical/controlled trials are really not suited to clinical problems that are uncommon or unimportant. The following examples demonstrate the limitations of the RCT design.

Example 1: To determine whether prophylactic antibiotics are required for routine arthroscopic meniscectomy of the knee in an otherwise healthy patient would be ludicrous. If the reported rate of infection is 3:1,000 patients, then to reduce the infection rate to 2:1,000 would require a sample size of nearly 40,000 patients per group.

Example 2: Several companies have developed new and innovative treatments for chondral lesions in the knee. These treatments have typically been evaluated in animal models with promising results. However, the animal model is not likely bipedal and probably uses a knee joint that is otherwise uninjured, normally aligned and the lesion is surgically prepared in one femoral condyle only. Subsequent human randomized trials would require a patient population that has an isolated chondral lesion to one femoral condyle in an otherwise stable and normally aligned knee that has failed standard treatment. In fact, these patients are very difficult to find. Assuming that the trial is completed, then the inferences from this trial and the results can only be applied to patients with similar characteristics to the original limited numbers of patients included in the trial.

Example 3: A randomized clinical trial comparing electrothermal arthroscopic capsulorrhaphy (i.e., the heat probe) versus open inferior capsular shift in patients with primary capsular redundancy was carried out in Canada.[57] This trial was hampered by several anecdotal reports of complications associated with the use of the heat probe. The trial was recently completed, and patient recruitment was the largest issue. The trial took 10 years to complete, and upon completion, the electrothermal arthroscopic capsulorrhaphy technique had been all but abandoned.

Example 4: In this hypothetical example, an RCT reports that surgical treatment is better than nonsurgical treatment but it comes to light that only 30% of the eligible patients were included in the trial. The other 70% of eligible patients may differ in several important characteristics. If this population is not accounted for, or their demographics are not compared with the trial patients, then the results may be very biased. If it turns out that there are known prognostic factors relating to patient outcome and these are dramatically different between the 2 populations, then the results of the trial are very limited.

Randomized trials are typically limited in that the strict nature of performing a trial requires specific inclusions and exclusions; consenting patients; sites with the necessary infrastructure; financial support, which may be from industry; and so on. The obvious conclusion asks the question of whether or not the results are generalizable to an individual orthopaedic surgeon.

IMPACT OF AN RCT: DOES IT CHANGE CLINICAL PRACTICE?

One of the largest problems of EBM and specifically conducting and reporting randomized clinical trials is that of knowledge translation. Peer-review funding agencies have developed strategies to ensure that the information gets to the end user, whether this is the patient or surgeon. One strategy is to provide specific funds within the grant for this purpose alone. Investigators applying for funding are therefore obligated to provide their strategies to disseminate the information. Different journals have partnered with funding agencies or organizations to provide a vehicle for authors to publish their results. Some agencies provide prizes for the best research in order to improve knowledge translation.

However, if the trial has internal and external validity and is reported widely, then there is every expectation that the results will have an impact on clinical practice. An example is the trend toward functional nonsurgical treatment of Achilles tendon ruptures based on recently published RCTs.[65] Ironically, rapid change in clinical practice is much more likely to occur if serious adverse events are reported from case series or through database surveillance rather than an RCT.[66]

CONCLUSIONS

Randomized clinical (controlled) trials in orthopaedic surgery represent the minority of published studies. The RCT is the most valid design to address clinically important questions. The question to be answered must be commonly encountered, must be important, and must show clinical equipoise. The surgeon must be committed, collaborate, and follow the necessary steps to perform a successful RCT.

Nicholas Mohtadi, M.D., M.Sc., F.R.C.S.C.
Mohit Bhandari, M.D., Ph.D., F.R.C.S.C.

SECTION 5

Study Designs: Cohort Studies, Level II Evidence, STROBE Checklist

Cohort studies allow investigators to examine a given population over a period of time for events related to the health of individuals in that population. Whereas case-control and cross-sectional studies evaluate patients at a single point in time, the defining characteristic of a cohort study is that it follows a group over time. In a typical scenario, investigators of a cohort study will collect health-related information at the beginning of the time period in question, or "baseline," and at predefined time points thereafter. The goal is to uncover, or assess the validity of, relations between a health-related variable recorded at one point in the study and an outcome recorded later. For example, all patients undergoing total knee arthroplasty at a given site are asked to fill out a questionnaire that, among other things, asks them whether they have diabetes. At the 10-year time point, some patients will be enjoying mobility whereas others may have a poor outcome. On noticing a pattern in their own patients, investigators might wish to evaluate the hypothesis that total knee arthroplasty patients with diabetes have worse outcomes than those without diabetes. Comparison of the outcomes of diabetic patients with nondiabetic patients would provide useful evidence to support or refute that hypothesis.

This example points to another important aspect of any cohort study design: the selection of the cohort. A cohort is defined by some shared characteristic. In the example above, the characteristic was that the patients underwent total knee arthroplasty. Completion of that procedure is an "eligibility criterion" for membership in the cohort. There is a broad distinction in cohort study design between a "closed cohort" and an "open cohort." A closed cohort is fixed; that is, the group is chosen at the beginning of the time period with which the study is concerned and does not change over the course of the study. The example above would be a closed cohort study if it were determined at the outset that exactly 100 patients would be included, all having undergone total knee arthroplasty at the given site during the month of July, 2010. An open cohort study follows a group whose membership changes over time. If the investigators of the arthroplasty study above aimed to continue to enroll patients indefinitely, the cohort's composition could change over time.

The STROBE checklist (Table 7)[67] was created to improve the quality of reporting of observational research, including cohort studies. It includes 22 items grouped according to the generally accepted components of a scientific article. This chapter will focus on those checklist items that are uniquely applicable to cohort study design: description of eligibility criteria and matching criteria (where applicable), explanation of loss to follow-up, summarized reporting of follow-up times, and quantitative reporting of outcome events.

STROBE: TITLE, ABSTRACT, AND INTRODUCTION

The title of any observational study should include a term denoting the design of that study, and cohort studies are no exception. An appropriate title might be "Incidence of Osteoarthritis After Meniscectomy: A Cohort Study." Not only does this allow the reader to quickly ascertain an important characteristic of the study—its design—but it provides for more effective electronic database searching because the study design can be indexed by its title.

The abstract describing a study must adhere to varying requirements set forth by individual journals, but there are certain items that should be addressed regardless of the format in which they are presented. These include, for one, background information; the impetus for the study should be explained. Authors should also state the specific objective(s) of the study and restate the study's design as indicated in the title. The setting of the study and details of patient selection, such as matching methods, are essential, in addition to the particulars of any measurements made. Naturally, authors should briefly report results and the conclusions they draw from those measurements, along with any inherent limitations of the study. A good abstract is both concise and comprehensive.

The introduction should elaborate on the context and objectives of the study as stated in the abstract. Authors should provide an overview of what is known

TABLE 7. *STROBE Statement: Checklist of Items That Should Be Included in Reports of Observational Studies*[67]

	Item No.	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	(*a*) Cohort study—Give the eligibility criteria, as well as the sources and methods of selection of participants; describe methods of follow-up Case-control study—Give the eligibility criteria, as well as the sources and methods of case ascertainment and control selection; give the rationale for the choice of cases and controls Cross-sectional study—Give the eligibility criteria, as well as the sources and methods of selection of participants (*b*) Cohort study—For matched studies, give matching criteria and number of exposed and unexposed Case-control study—For matched studies, give matching criteria and the number of controls per case
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers; give diagnostic criteria, if applicable
Data sources/measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement); describe comparability of assessment methods if there is more than 1 group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses; if applicable, describe which groupings were chosen and why
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding (*b*) Describe any methods used to examine subgroups and interactions (*c*) Explain how missing data were addressed (*d*) Cohort study—If applicable, explain how loss to follow-up was addressed Case-control study—If applicable, explain how matching of cases and controls was addressed Cross-sectional study—If applicable, describe analytic methods taking account of sampling strategy (*e*) Describe any sensitivity analyses
Results		
Participants	13*	(*a*) Report numbers of individuals at each stage of study—e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analyzed (*b*) Give reasons for non-participation at each stage (*c*) Consider use of a flow diagram
Descriptive data	14*	(*a*) Give characteristics of study participants (e.g., demographic, clinical, and social) and information on exposures and potential confounders (*b*) Indicate number of participants with missing data for each variable of interest (*c*) Cohort study—Summarize follow-up time (e.g., average and total amount)
Outcome data	15*	Cohort study—Report numbers of outcome events or summary measures over time Case-control study—Report numbers in each exposure category or summary measures of exposure Cross-sectional study—Report numbers of outcome events or summary measures
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval); make clear which confounders were adjusted for and why they were included (*b*) Report category boundaries when continuous variables were categorized (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, as well as sensitivity analyses

TABLE 7. *Continued*

	Item No.	Recommendation
Discussion		
Key results	18	Summarize key results with reference to study objectives
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision; discuss both direction and magnitude of any potential bias
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalizability	21	Discuss the generalizability (external validity) of the study results
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies. Reprinted with permission.[67]

about the topic of the study and explain how their work adds to the field. Objectives should be specific. It is important when stating the objectives of a cohort study, for example, to state precisely which populations and outcomes the study evaluates.

STROBE: METHODS

The methods section should include fundamental features of the study design early on. If the study follows a cohort, that should be stated at the outset with a justification for the choice to use that design. It is especially important for a cohort study to clearly describe the characteristics of the patients and their outcomes. The creators of the STROBE checklist emphasize a particular point regarding the use of the terms "prospective" and "retrospective." They point to the ambiguity in these terms and suggest that they simply not be used. Instead, authors should offer a careful description of the ways in which data were collected and exactly when data were collected. The methods section should also include a statement of the original purpose for which the data were collected and whether that purpose differs from the purpose for which the data are being used in the particular study at hand. As was important in the abstract and introduction sections, this provides a more accurate context for the data, which is critical when judging their limitations.

Authors should be specific when describing the setting of a study. Where geographically and at what institutions were data collected? When did data collection begin and end? At what particular intervals were data collected? The methods section should clearly address these questions.

The process by which participants were selected is particularly important to a cohort study report. The eligibility criteria should be extremely clear. In a cohort study tracking the outcomes of patients who have undergone ACL revision surgery, for example, it is not enough simply to state that the patients were included if they had this procedure performed. Were pediatric patients included? Were older patients included? What if a particular patient had undergone ACL revision at an earlier date? What if the patient had rheumatoid arthritis? It is possible that the study would include all of these patients, but it must be made clear. If there are limitations on age, existing conditions, surgical indications, or any other variable, these should be stated in full. At the very least, age, gender, comorbid conditions, and diagnosis should be addressed when setting forth eligibility criteria. The authors should also report on characteristics of the broader group from which the cohort was selected. To add to the example described above, that study might have drawn all of its cohort from the population of male persons aged between 17 and 50 years who reside in Maryland. That should be included in the description of participants. Again, the goal is to provide the reader with as much information as possible that is relevant to the evaluation of the data's limitations so that the reader may judge the validity of the conclusions drawn from those data.

The authors should report follow-up methods clearly. The authors might state that questionnaires were administered at baseline and at the 6-month, 2-year, and 5-year time points. However, sometimes, questionnaires are not completed at precise time points like these. Perhaps the baseline questionnaire for one patient was in fact filled out 3 weeks after the procedure. If this is the case, the authors should justify their inclusion of the data. They might point to a study that shows that baseline question-

naires completed within one month of a procedure still provide valid data.[68]

All variables under analysis in the study should be defined unambiguously. These include outcomes, exposures, predictors, potential confounders, and effect modifiers. Disease outcomes should be identified by specific diagnostic criteria, which should also be reported. For example, it would be necessary to describe exactly what constitutes failure of an ACL reconstruction (MRI, arthroscopic evaluation, and so on). Authors should also report the source of data for each variable and how those data were measured, along with an estimate of the reliability and validity of their measurements. Ideally, measurement methods would be identical between groups being compared. If there are variations, these should be noted. Similarly, authors should address in their report any potential sources of bias. Hopefully, steps have been taken to minimize any bias present in the results. These steps should be conveyed in full to the reader.

Quantitative variables should be explained carefully, especially in terms of their grouping and the modeling of the data representing those variables. Choices are made in the process of separating continuous quantitative data representing a given variable into ordered groups, and those choices can have a significant impact on later analysis of the data. Going one step further, this applies to the choice of statistical analysis as well. Given the possibility of choosing a particular analysis to support a particular hypothesis once all of the data have been collected, that choice should be made at the outset of the study. This is true of the methods by which interactions between subgroups were examined and missing data were accounted for, as well. If these analytic methods changed over the course of the overall analysis of the data, the changes should be reported. It should also be clear in the report how confounders were selected. The Explanation and Elaboration document[67] prepared by the creators of STROBE offers a more detailed treatment of the reporting of statistical methods, which applies not only to cohort studies but to any other type of observational study as well.

Loss to follow-up deserves particular attention when reporting on a cohort study. If the total length of follow-up time is fixed, whether in terms of age or time elapsed since the baseline time point, an assumption is made during analysis: for individuals who reach this fixed endpoint without a particular outcome, there is no relation between follow-up time and the probability of their developing that outcome. Problems arise when the distribution of loss to follow-up is uneven between groups.

For example, the investigators of an ACL study hypothesize that individuals who have had an associated meniscal repair have a higher likelihood of reoperation. If 20% of the cohort is lost to follow-up, including 80% of the meniscal repair patients, it is difficult to establish a potential relationship between the meniscal repair and the outcome, a reoperation. It may appear as though there is no relation when further observation of those individuals may have established that very relation. For this reason, members of the cohort who were lost to follow-up must be distinguished from those who remain under observation until the fixed endpoint of the study. Those lost to follow-up could be excluded from the study or they could be treated as though they withdrew without the outcome in question, either at the end of the study or on the date on which they were actually lost to follow-up. When planning the study, investigators should determine how loss to follow-up will be handled. However they choose to treat those lost to follow-up when analyzing their data, they should report the exact number falling into this category.

STROBE: RESULTS

Another item on the STROBE checklist that concerns cohort studies in particular hinges on the reporting of the timing of follow-up. This should be addressed in the results section. These data can be summarized as a mean or median duration of follow-up in combination with a total of the person-years observed. For example, ACL surgery investigators might report that follow-up lasted a mean of 5.2 years in 31 patients, for a total of 161.2 person-years of follow-up. Minimum and maximum follow-up times can be used in conjunction with percentile distribution for a more comprehensive picture of follow-up duration.

Closely related is the STROBE checklist's suggestion that outcome rates should be presented by follow-up time point, each time point being associated with a certain number of outcomes and a certain number of person-years of follow-up, such that the rate can be clearly shown as a ratio between the number with a certain outcome and the total number of person-years observed at that follow-up time point. For example, the investigators might report that 6 failures had occurred by the 1-year time point, at which point 31 person-years of follow-up had been performed, for a rate of about 2 failures per 10 person-years. In addition to mean data on follow-up times,

investigators should report just how many participants were involved over time. A flowchart can be quite useful in illustrating the answers to the following questions at various points in the study's progress: Who might be eligible to participate? Who has been confirmed as eligible by examination? Who is actually participating in the study? For those eligible who are not participating, what is the reason? Who is in the process of follow-up? Whose information has already been analyzed?

The characteristics of those participating in the study must be provided in the results section just as they were in the methods section, but in tabular form. This allows the reader to judge the study's generalizability. In cohort studies, exposures and potential confounders should be reported by group. If participants are missing data for certain variables, the numbers missing should be listed by variable.

When one is reporting the main results, estimates should be given both unadjusted and adjusted for confounders. The latter estimates require a figure denoting their precision, such as a confidence interval. As mentioned earlier when explaining the reporting of quantitative variables in the methods section, any continuous variables that are converted to ordinal variables should be reported with the boundaries of those categories. Of course, all analyses should be reported, including those of subgroups and interactions.

STROBE: DISCUSSION

When composing the discussion section, authors must be careful to separate their own opinions from their rigorous and unbiased interpretation of the data. The former have no place in the discussion. In keeping with a recurring theme, the discussion section should address the limitations of the study while summarizing the most important results in terms of the original objectives of the study. In the methods section, the authors should have described the measures they took to guard against the effects of potential biases. Were these measures successful? If bias appears to have affected the results, to what extent and in what direction did this happen? These are questions that should be answered in the discussion.

All things considered—biases, statistical uncertainty, the very nature of the study—what does the study show? Every other element of the discussion section serves to help answer this question. It cannot be emphasized strongly enough that this interpretive task is easily clouded by authors' personal opinions. To successfully craft the discussion section, authors must devote attention to this tendency, such that its effects might be reduced.

Finally, the authors must broaden the scope of the discussion to explain the generalizability of their results. At this point, they have offered an interpretation of the study's findings within the realm of the study itself, but how do their findings apply to patients outside of the study cohort? It is in this part of the discussion that the study's impact on clinical practice can become clear. Previous explanation of the setting of the study, eligibility criteria for participation in the cohort, and the exposures and outcomes to be measured help readers to assess on their own the generalizability of the study findings. Naturally, authors should address these topics when offering their own argument for the ways in which the findings can be applied in other circumstances. Also important to this critical process is the disclosure of 2 more factors that may introduce bias: sources of funding for the study and any conflicts of interest the investigators may have.

THE PURPOSE OF STROBE

The reader of a cohort study report should be able to critically evaluate that report in 2 ways: in terms of its internal validity and in terms of its external significance. Do the data clearly support the conclusions reached regarding the specific domain of the study? Do those specific conclusions support the broader implications the authors suggest? After all, the ultimate goal of most cohort studies is to provide information that will make a positive impact on clinical practice.

Authors who adhere to the STROBE checklist ensure that their readers have the tools necessary to make these critical judgments. Each element of the checklist serves this purpose, some more obviously than others. Attention to this guiding principle should help authors effectively execute the particular items presented here.

While offering an overview of the whole STROBE checklist, this chapter has focused on those items with unique application to cohort study reporting, particularly in orthopaedic surgery. For a comprehensive discussion of the application of the STROBE checklist, investigators may consult the STROBE Explanation and Elaboration document referenced above.[67]

Brian W. Boyle, B.A.
Michael Soudry, M.D.
Robert G. Marx, M.D., M.Sc., F.R.C.S.C.

SECTION 6

Study Designs: Case-Control Studies, Level III Evidence, STROBE Checklist

Clinical investigations are an integral component to assessing and improving the care of our patients. Whereas prospective studies are considered the "gold standard" of clinical outcomes research (type I and type II studies), it has been estimated that 60% of surgical questions cannot be answered by these methods. Case-control studies (type III) are a type of observational study that can be quite useful in identifying risk factors associated with a specific outcome. In this type of study, subjects with the outcome of interest (cases) are compared with similar subjects who do not have that outcome (controls). Data from each subject's treatment records are then compared to identify factors common to the cases but not common to the controls by use of epidemiologic and statistical methods. These factors may be genetic, gender, or chemical or based on exposure or other comorbidities. Case-control studies are most useful when the research question addresses outcomes that are rare or take a long time to develop.[69] In these situations randomization or prospective cohort studies may not be feasible because of the required length of follow-up and expenses. Case-control studies are also indicated when randomization may be unethical (such as when investigating the fracture risk associated with the use of nonsteroidal anti-inflammatory drugs[70]).

The classic case-control study was performed by a young medical student, Ernst Wydner, who was fascinated with lung cancer. He interviewed a pool of 649 lung cancer patients (cases) and 600 patients with other cancers (controls) and found that the incidence of lung cancer was 40 times higher in smokers that in those who did not smoke.[71] Other investigators who read this index study began studies of their own to further understand the association between smoking and the development of lung cancer.

Case-control studies have been very useful in orthopaedics and sports medicine to assess the risk of a specific injury or to assess the risk of a certain outcome after injury or surgery. The results of these studies can lead to strategies to reduce the risk of injury or to improve the clinical outcomes of our treatments.

The focus of this chapter is to provide the investigator a structure to assist him or her in designing and carrying out a case-control study.

WHEN TO CONSIDER A CASE-CONTROL STUDY

When considering a case-control project, the investigator should consider several critical points. The advantages and disadvantages of this type of study should be assessed before beginning the study to be sure that the outcome will answer the research question. Table 8 provides some of the advantages and disadvantages of a case-control study.

First, the topic of study should be one that is familiar to the research team. Research questions (see be-

TABLE 8. *Strengths and Limitations of a Case-Controlled Study*

Strengths
Facilitates the study of rare outcomes
Facilitates the study of conditions with substantial time between exposure and outcome
Control groups can be matched according to known (or suspected) confounding variables
Allows for the study of multiple potential causes of an outcome of interest
Relatively inexpensive
Can be completed over relatively short time periods
Limitations
Inefficient when the exposure is rare
Information on exposure and history that is derived from interview is subject to recall bias
Selection of an appropriate control group may be challenging
Lack of randomization means that groups may suffer from an imbalance of confounding factors
Can only study one outcome of interest
Validation of exposure information is often difficult (or impossible)
Cannot provide information on prevalence of the outcome of interest
Unable to establish causality
Methodology and correct interpretation of results may be challenging

Reprinted with permission from Busse JW, Obremskey WT. Principles of designing an orthopaedic case-control study. *J Bone Joint Surg Am* 2009;91:15-20 (Suppl 3).