Designing a Biomechanics Investigation: Choosing the Right Model

Steven A. Olson, MD, J. Lawrence Marsh, MD,† Donald D. Anderson, PhD,† and Loren L. Latta, PE, PhD‡*

Summary: Physical testing is commonly performed to answer important biomechanical questions in the treatment of patients with fractures and other orthopaedic conditions. However, a variety of mistakes that are made in performing such investigations can severely limit their impact. The goal of this article is to discuss important aspects of study design to consider when planning for biomechanical investigations so that the studies can provide maximal benefit to the field. The best mechanical investigations begin with a good research question, one that comes out of patient care experience, is clearly defined, and can be stated concisely. The first practical issue to be considered is often choosing the type of physical specimens to be tested to address the research question. A related issue involves determining how many specimens will be needed to answer the posed mechanical question. Cadavers are generally still the closest to the actual clinical situation, but they are limited by interspecimen variability, which often requires a matched pair design that can address only one question. Simulated bone specimens limit variability and can replicate normal and osteoporotic bone. In planning the physical testing, the critical mechanical variables involved in answering the research question must be identified and due consideration given to deciding how best to measure them. Another important issue that arises relates to whether or not single static loadings will suffice in the testing (eg, to study construct stiffness) or whether cyclic dynamic testing is necessary (eg, to study late failure likely attributable to fatigue). To summarize, experimental design should be carefully planned before initiating mechanical testing. Sample size calculations should be performed to ensure adequate power and that clinically relevant differences can be detected. This pregame analysis can save significant time and cost and greatly increase the likelihood that the results will advance knowledge.

Key Words: biomechanics, mechanics, materials testing, experimental design

(*J Orthop Trauma* 2012;26:672–677)

Accepted for publication August 31, 2012.

From the *Department of Orthopaedic Surgery, Duke University School of Medicine, Durham, NC; †Department of Orthopaedic Surgery, University of Iowa School of Medicine, Iowa City, IA; and ‡Department of Orthopaedic Surgery, University of Miami, Miami, FL.

This paper was presented in part at the Annual Meeting of the Orthopaedic Trauma Association, October 12, 2011, San Antonio, TX.

The authors have no funding or conflicts of interest to declare.

Reprints: Steven A. Olson, MD, Department of Orthopaedic Surgery, Duke University School of Medicine, DUMC 3389, Durham, NC 27710 (e-mail: Olson016@dm.duke.edu).

INTRODUCTION

Physical testing is a common way to address biomechanical questions in the clinical treatment of patients with fractures and other orthopaedic conditions. The goal of this article is to discuss aspects of study design to consider when planning for biomechanical investigations.

Consider the illustrative case shown in Figure 1. A patient older than 65 years sustains a reverse obliquity intertrochanteric fracture, classified as an OTA 33A3 injury.[1] The implant has sustained what seems to be a fatigue failure. The surgeon wishing to design an experiment to understand the mechanics of loading in this fracture and the impact on implant failure must consider the following questions, among others:

- Should the testing use cadavers or synthetic bones?
- Should the testing measure strain or stiffness?
- Should the testing use repeated measures or a single measurement per specimen?
- Should the testing load in static or cyclic mode?

The following discussion should guide the reader to answers of these questions.

SELECTING THE RIGHT SPECIMEN: DO MATERIALS MATTER?

Clinical performance in patients is the bottom-line consideration in assessing fracture fixation constructs. However, evaluating the performance of new implants or assessing the mechanical differences between various implant choices in clinical practice is hampered by many uncontrolled variables. These variables include the characteristics of the fracture, the patient, the bone, and the implantation technique and are also difficult-to-assess and control for postoperative loading conditions.

Choices: Specimen Type for Mechanical Testing

So what specimen should be tested? It helps to start by considering what question you are asking. For comparison between two or more implants or implant characteristics, where the implant is the primary issue, the specimen can and should be as generic and consistent as possible. Differences in the specimen confound the results. However, where the anatomy and bone characteristics are important, the behavior of the specimen must be as realistic as possible.

The Implant Itself

For the simplest questions, just test the implant. This might involve establishing the bending and structural stiffness

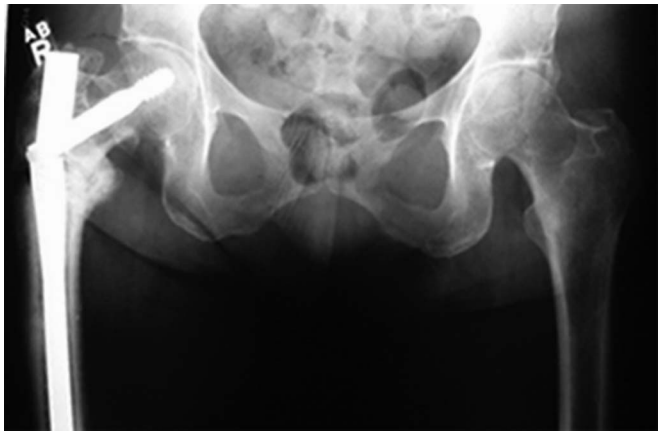

FIGURE 1. A 65-year-old woman 12 months after fixation of a reverse obliquity 33A3 proximal femur fracture.

of thick and thin plates, plates with and without cutouts, and mechanical behavior of open and closed section nails. This type of testing is usually done by the implant manufacturer and is often required for Food and Drug Administration submission. Standardized test methods can be found in American Society for Materials and Testing and International Standards Organization standards documents.

Nonanatomic Bone-Mimicking Materials

These have included polyurethane foam blocks, PVC tubing, oak blocks, Delrin, and UHMWPE blocks. Such materials are best for answering simple questions that focus on implant characteristics and interface with bones. If a test requires anatomy-specific loading or bone fit and design characteristics, this type of model is not feasible.

Surrogate Bone Models

These are synthetic and have different foam core densities to simulate both normal and decreased bone strength.[2] The advantage of surrogate bone models is that they avoid problems with the variability of size and mechanical properties of cadaver bones. They have known mechanical and geometric properties with small standard deviations; they can be made the same and that is good for experimental testing. They are particularly good for diaphyseal testing and destructive testing in multiple loading modes, and they are often used where multiple comparisons are being made.

When testing with simulated bones, the issue of osteoporotic versus strong (normal) bone specimens must be considered. Implant testing done only in strong bone may not provide information relevant to many of the osteoporotic fractures seen in clinical practice.[3] Testing in strong bone will evaluate the characteristics of the implant itself, whereas testing in osteoporotic bone evaluates the quality of the implant to bone fixation (Fig. 2). Osteoporotic surrogate bone specimens are available. The degree to which they actually characterize true osteoporotic bone is uncertain, and the extent to which they are validated is often specific to a particular outcome parameter.[2]

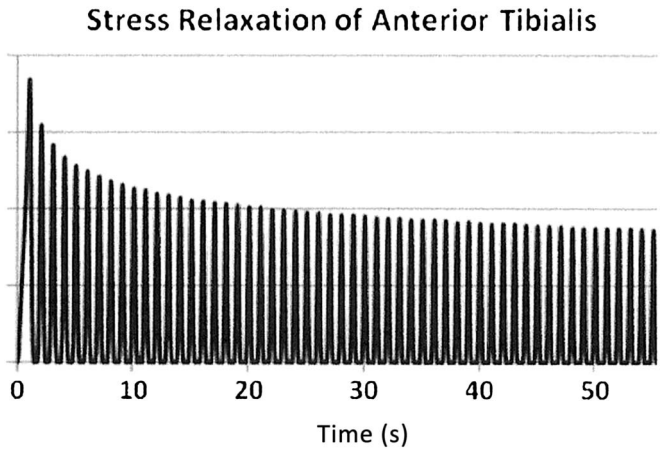

FIGURE 2. Cyclic loads were applied in stroke control to an anterior tibialis tendon sample prepared for reconstruction of an Anterior Cruciate Ligement. The peak of the cycle was set at 10% strain of the length of the specimen, and the relaxation of the load was monitored.

Cadavers

Cadaveric tissues provide the opportunity to test implants in actual bone specimens. They provide the variability in bone that is encountered in clinical practice, which is both good and bad. Cadaveric tissues are good because the structural and material properties for testing most closely approach clinical reality. A significant difference seen between 2 devices in cadaveric testing will most likely also be seen clinically.

However, this variability is also bad for mechanical testing. The variability produces higher standard deviations in the data, making it harder to show differences relevant to the experimental question than it is for similar experiments in simulated bone specimens. For instance, there may be up to a 7-fold difference in strength between specimens, which could overwhelm the differences from the test question.[4] In addition, the actual anatomy varies with differences in cortical thickness across a single bone and even within a bone. There are also issues of how the cadavers are handled. Fresh frozen cadaver specimens are preferred, but properties may change based on thaw time, temperature, moisture content, and other such properties.[5]

Cadaveric tissue testing requires techniques to limit effect of specimen variability. Bone mineral densities can be controlled for by using bones from young healthy subjects, but these are difficult to find. Bone mineral density can be measured and balanced between experimental groups. The most common way to limit variability is to use matched pairs cadaver testing. However, there still may be significant right and left differences. One choice is to use cadavers to validate key results found with other testing.

SAMPLE SIZE: HOW MANY IS ENOUGH?

When we cannot measure the response of every individual, we need a tool to estimate how confident we are that the sampled data represent the behavior of the population.

Statistical tests often are used to estimate how confident we are that samples are really different. In general, we are interested in the difference between means relative to the variability of the samples.[6]

Before beginning an experiment, you should perform a sample size calculation to help ensure that you have sufficient power. Precise calculations depend on study design and analysis you will use and potentially become quite complex. Numerous software packages are available to help with these calculations. Typically, $\alpha = 0.05$ and power = 0.80 are used. That leaves you to estimate the standard deviation and size of the difference you believe to be clinically meaningful.

How do you find these data? They are generally estimated from pilot data, previous studies in the literature, or intuition. If you are uncomfortable with these calculations, consult with someone experienced in statistics and study design—it could save you considerable time and expense in the long run.[6]

Are there ways to reduce effects of large variations between individuals? Repeated measures designs, where multiple treatments are performed on the same experimental specimen, can be used to reduce variability (and hence sample size). For example, testing multiple fixation constructs in individual synthetic bone specimens allow the investigator to assess the differences between constructs from specimen to specimen; testing only one fixation construct per individual specimen requires many more specimens to compare the differences between the same number of constructs.

CRITICAL VARIABLES (STIFFNESS, STRAIN, AND SUCH VARIABLES): WHAT IS BEST?

Now you have chosen the right specimens for your study, and you have an idea of how many specimens are needed for the testing you plan to do (although a sample size estimate is difficult without first knowing what mechanical variables are best for your study). The next question has to do with what critical mechanical variables need to be considered in your biomechanical study. One must have a clear understanding of some basic mechanical principles. Then questions of relevance to the clinical application must be addressed.

Basic Mechanical Principles

There are a few basic principles that a person doing a mechanical study needs to know, and for that reason, they are briefly defined here. The discussion of these mechanical principles in a broader sense is beyond the scope of this article, so definitions are limited here.[7]

Displacement

The physical distance that an object or structure moves through under a given set of circumstances. Units of measure are generally millimeters or centimeters.

Force

The energy imparted to a system by external means effecting (or resisting) displacement or deformation of an object or structure. Units of measure are usually Newtons (N) or kilonewtons (kN).

Both displacement and force are vector quantities, which means that they act along a given direction (eg, superior–inferior and anterior–posterior).

Stiffness

The ratio of the force to the displacement in a given system (ie, stiffness = force/displacement). Units of measure are thus most commonly Newton per millimeter (N/mm) or Newton per meter (N/m). In a mechanical test where the force and displacement are being simultaneously recorded, the stiffness is simply the slope of the force-displacement curve. Because force-displacement curves are seldom truly linear, the stiffness is defined as the slope over a given portion of the curve, based on whether or not initial or overall stiffness is more or less important.

Load to Failure

The force that is measured at the time the material breaks in a test involving a continuous application of force (failure point at the ultimate load).

Fatigue Failure

Cyclical subfailure loading may result in failure due to fatigue. Load-to-failure testing data are most often reported in the orthopaedic literature: however, clinically, failure of fracture fixation constructs or implants are often because of fatigue failure.

The reader should recall that a distinction is made between material and structural parameters, whereby the former depends exclusively on a material's essential properties and the latter is further influenced by the physical details (thickness, diameter, and cross-sectional area) of a structure made from the materials. Furthermore, one should bear in mind that force and displacement are 2 parameters that can generally be measured with a very high degree of accuracy, often with fairly simple equipment. Moving from material properties now to structural properties.

Strain

Strain is the amount of deformation per given length of an object. The units of measure for strain are dimensionless, but to avoid confusion, they are often referenced as millimeter per millimeter (mm/mm) or percentage (%) strain. Because deformations act along a direction, and a material or structure deforms in multiple directions as a result, strain is more complex than deformation. In the simplest 1-dimensional sense, strain is simply the change in length divided by the original length of an object.

Stress

The amount of internal forces acting on a given area of an object. The units of measure for stress are thus Newtons per square millimeters (N/m^2—a Pascal or Pa), but for most orthopaedic applications are more on the order of megapascal (MPa; N/mm^2). As with strain and its relationship to deformation, stress is more complex than force. In the simplest case, stress is simply the applied (or measured) force magnitude divided by the cross-sectional area (F/A).

Elastic Modulus

The ratio of the stress to the strain on a region for which the deformation is elastic—analogous to stiffness for force displacement. The units of measure are the same as for stress.

Yield Point

The point on the stress–strain curve where the deformation goes from elastic (ie, fully recoverable on unloading) to plastic (ie, unrecoverable) is the yield point.

Failure Point

At some point, the material will break; this point is the material's Failure Point.

Brittle

A material that experiences little plastic deformation (strain) before it fails is said to be brittle (eg, glass or cortical bone).

Ductile

If a material has a large plastic deformation region before it fails, it is said to be ductile (eg, copper).

Toughness

A material that can absorb more energy before failure (large area under the stress–strain curve) is said to be tougher. Unit of measure for toughness is Newton per square millimeter times millimeter per millimeter [$(N/m^2)(m/m)$] or Joules per cubic millimeter.

Strength

The highest stress on a stress–strain diagram, also known as the ultimate strength, dictates the highest stress that can be maintained by a material before it breaks. Strength and stiffness are NOT the same thing, and the difference is important.

So, given this review of mechanical principles, what variables should you measure in your testing? Load and displacement are enough in some cases, and they are much easier and more direct to measure (in general). That means that there are fewer assumptions required to quantify them. Stress most often comes into play when considering material or structural failure. In this respect, it is important to bear in mind that bone and fixation devices are together a structural construct. Strength is thus an appropriate variable for failure assessments.

In addition to basic considerations related to mechanical testing, one should remember that in the human body, biology is also at play. Sometimes, the mechanical state itself is of paramount importance, as in the case of an implant subjected to a load that exceeds its capacity. But, if you are hoping for fixation to get bone to heal, then the strains that the bone is subject to need to be consistent with bone healing. Too much strain can lead to fibrous nonunion and eventual fixation failure. Too little strain can lead to slow healing and possible nonunion. The best strain is that which is most appropriate to stimulate bone healing.

DYNAMIC VERSUS STATIC LOADING: WHAT IS THE BEST?

Terminology

Statics is that branch of mechanics that deals with bodies at rest or in equilibrium. Static loading is a term usually used to describe a "slow" continuous cycle of loading. But is also used to describe progressive incremental stepped "cycles" of load and continuous or repeated elastic cycles of load to establish steady state of behavior.[7] **Dynamics** is that branch of mechanics that deals with the action of loads affecting or producing motion or with moving loads on bodies at rest. Dynamic loading is a term usually used to describe cyclic loads.[7]

Cyclic loads are a defined change in load amount, type, and/or direction, which is repeated. Cyclic loads may be applied in single "direction" cycles, that is, from 10% to 100% of the peak in tension, compression, shear, bending, or twisting. Cyclic loads may also be applied in reversing cycles: −100% to +100% of the peaks. Load cycles can be controlled by the applied load or displacement. Load control cycles repeat the same magnitude and direction of load with each cycle. Displacement control cycles repeat the same motion of the load point or region with each cycle. If there is a progressive change in behavior with each load cycle, in load control, the displacement will change (called creep) or in displacement control, the load will change (called load relaxation).

Application to Orthopaedic Devices[7]

Fatigue Testing

Endurance limit testing is for prostheses, which are load bearing devices. Low cycle fatigue is for fixation devices which are load sharing devices that need only to survive until the damaged tissues are healed.

Practicalities of Testing

The ideal simulation of the environment would be to test all materials and constructs at 37°C, in a saline bath or at 100% humidity. The time of the test is also important because biological materials will change properties with time. The rate of these changes is very dependent on the temperature, the moisture content, and level and rate of strains on the tissues.

Cyclic loading to measure endurance characteristics with nonbiological and nonviscoelastic materials can be loaded at accelerated rates to minimize the time of testing. With biological and viscoelastic materials in the construct, the strain rate is very important, so cycling at greater than 1 Hz is usually impossible, and the most realistic simulations are usually at a half or quarter Hertz. To test for endurance limit requires at least one million cycles. At quarter Hertz, the test would take more than 1000 hours; for details, see Table 1. If the properties of biological tissues are important for such a construct, it is impossible to maintain viable properties for that time even at room temperature. Fortunately, fracture fixation devices are normally expected to endure only about 200,000–250,000 cycles of mechanical loading before the tissues should be healed and the mechanical contribution by the device should no longer be needed. But still at quarter Hertz, a fracture fixation construct would require 278 hours

TABLE 1. Time (in Hours) to Complete Testing for Various Cyclic Loading Targets

Cyclic Loading Rate (Cycles/s)	10^7 Cycles	10^6 Cycles	250,000 Cycles
10 Hz	278	28	7
1 Hz	2,778	278	69
0.5 Hz	5,556	556	139
0.25 Hz	11,111	1,111	278

of cyclic testing to reach 250,000 cycles. Degradation of mechanical properties can be significant with more than 10 days of testing, even if attention to details, such as moisture content are well attended to. Constructs that are quite rigid can often be run at 1 Hz, cutting the test time to less than 3 days. But often, it is just more practical to use materials that simulate the biological components of the construct.

For simple material tests in which a single well-dimensioned piece of tissue or material of a fixation device is tested in tension, compression, or shear, a single cycle of load may be used to define a stiffness and/or strength characteristic. The rate of loading is very important for viscoelastic specimens, such as cadaveric bone. Cyclic testing within the elastic range is often the most effective way to define the stiffness characteristics of viscoelastic materials, if a single stiffness value is desired or if stress relaxation or creep characteristics are desired (Fig. 2).

Testing of single components of a fixation device (ie, a screw or pin) can usually be accomplished with a single cycle test in bending, torsion, or combination loading for stiffness and/or strength (Fig. 3). A test of a single biological structure, such as a diaphyseal bone, can be accomplished in the same manner, but the load rate becomes important, and cycling to a steady state may be better for stiffness-only measures.

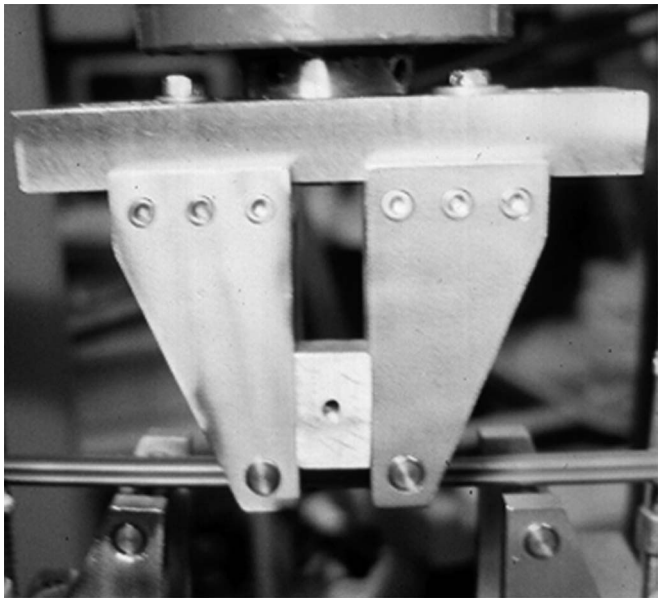

FIGURE 3. The standard test described by ASTM and International Standards Organization for bending of a component of long uniform dimensions is the 4-point bend test shown here testing an intramedullary nail.

Construct tests of assemblies of fixation device components is useful for evaluating interfaces between components and rigidity and evaluating mode and location of mechanical failures of different configurations. A single cycle of load usually provides the best loading condition unless fatigue behavior is sought.

Constructs that include anatomic portions of the musculoskeletal system will require attention to loading rate for single-cycle tests. For stiffness testing, cycling to a steady state is usually best with well-controlled loading rates. Constructs in which soft tissue components are critical are more load rate sensitive than those that only include bones and hardware. Such constructs may include a joint with ligaments and cartilage components providing some of the control of the load transfer. To measure load transfer, stiffness characteristics, and other such parameters, cyclic loading to steady state is usually the most effective because of the viscoelastic contribution of the soft tissues.

AVOIDING COMMON PITFALLS IN STUDY DESIGN

The best research questions come out of patient-care experiences. A good research question is clearly defined; the answer to the question is not known; the answer will contribute to existing knowledge; the question can be stated concisely.[6,8]

The research question should be able to be formed into a testable hypothesis. To help define the question, turning to relevant literature is an important step. What are the inconsistent findings or gaps in knowledge in the literature? What are the common failure mechanisms? Does the fixation implant fail, or does the interface between implant and bone loosen and then fail? These observations help to guide the development of the research question and ultimately, the experimental design.

Examples of this are common. Multiple clinical reports of fixation of intra-articular distal femur fractures indicate that the implants rarely fail acutely and generally do not fail with implant breakage.[9–11] In such an instance, a study designed to evaluate the energy of load to failure of distal femoral fixation is problematic; The data will be valid biomechanically but may not be clinically relevant in any aspect except acute loading after fixation. Work by Ziran et al[8] demonstrated that cadaver bone and synthetic bone had different failure modes when testing load to failure after cyclic loading in unstable subtrochanteric fractures. Studies that seek to investigate failure modes of specific implants should consider specimen choice carefully.

Because few surgeons are trained engineers with an understanding of loading biologic tissues, in most cases, consulting an engineer trained in biomechanics is of great help in experimental design. The surgeon, in this case, must be able to describe the loading conditions clinically. The engineer then translates this interpretation into experimental designs that replicate and test relevant loading conditions.

Finally, consider how your findings compare with the literature. Do they differ in some important way? If so, can you explain why? Such observations are often the beginning

of further research questions. Occam's Razor suggests other things being equal, a simpler explanation is better than a more complex one.[12] Do not look for complex explanations when the data does not fit your hypothesis. Look for the simplest explanation that can address the complete finding of the data.

CONCLUSIONS

Fracture implant mechanical testing is important but is limited by the characteristics of the specimens and the experimental design chosen. Cadavers are still the closest to the actual clinical model but are limited by interspecimen variability often requiring a matched pair design that usually can address only one question. Simulated bone specimens limit variability and can replicate normal and osteoporotic bone. Experimental design should be carefully planned before study initiation—this pregame analysis can save significant time and cost. Sample size calculations should be performed to ensure adequate power and that clinically relevant differences can be detected.

REFERENCES

1. Marsh JL. OTA fracture classification. *J Orthop Trauma.* 2009; 23:551.
2. Sommers MB, Fitzpatrick DC, Madey SM, et al. A surrogate long-bone model with osteoporotic material properties for biomechanical testing of fracture implants. *J Biomech.* 2007;40:3297–3304.
3. Schneider E, Goldhahn J, Burckhardt P. The challenge: fracture treatment in osteoporotic bone. *Osteoporos Int.* 2005;16(suppl 2):S1–S2.
4. O'Toole RV, Andersen RC, Vesnovsky O, et al. Are locking screws advantageous with plate fixation of humeral shaft fractures? A biomechanical analysis of synthetic and cadaveric bone. *J Orthop Trauma.* 2008;22:709–715.
5. Cartner JL, Hartsell ZM, Ricci WM, et al. Can we trust ex vivo mechanical testing of fresh–frozen cadaveric specimens? The effect of postfreezing delays. *J Orthop Trauma.* 2011;25:459–461.
6. Dowrick AS, Tornetta P III, Obremskey WT, et al. Practical research methods for orthopaedic surgeons. *J Bone Joint Surg Am.* 2012;94:368–374.
7. Latta L, Milne EL, Kaimrajh DN. Building a biomechanical model. *J Orthop Trauma.* 2011;25:454–458.
8. Ziran BH, Sharkey NA, Smith TS, et al. Modified transverse locking nail fixation of proximal femoral fractures. *Clin Orthop Relat Res.* 1997;339: 82–91.
9. Merchan EC, Maestu PR, Blanco RP. Blade-plating of closed displaced supracondylar fractures of the distal femur with the AO system. *J Trauma.* 1992;32:174–178.
10. Ostrum RF, Geel C. Indirect reduction and internal fixation of supracondylar femur fractures without bone graft. *J Orthop Trauma.* 1995;9: 278–284.
11. Thomson AB, Driver R, Kregor PJ, et al. Long-term functional outcomes after intra-articular distal femur fractures: ORIF versus retrograde intramedullary nailing. *Orthopedics*. 2008;31:748–750.
12. Wikipedia. *Occam's Razor*. 2012. Available at: http://en.wikipedia.org/wiki/Occam's_razor. Accessed April 14, 2012.